COMPARING LATENT INTERACTION EFFECTS IN MULTI-SAMPLE STRUCTURAL EQUATION MODELING: QUASI-MAXIMUM LIKELIHOOD VERSUS THIRD MOMENT METHODS

by

Ryan T. Pohlig

B.S., Univeristy of Scranton, 2005

M.A., University of Pittsburgh, 2011

Submitted to the Graduate Faculty of

The School of Education in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

School of Education

This dissertation was presented

by

Ryan Pohlig

It was defended on

November 15, 2012

and approved by

Jeffery Shook, Associate Professor, Social Work

Clem Stone, Professor, Psychology in Education

Feifei Ye, Assistant Professor, Psychology in Education

Dissertation Advisor: Kevin Kim, Associate Professor, Psychology in Education and Business Administration

COMPARING LATENT INTERACTION EFFECTS IN MULTI-SAMPLE STRUCTURAL EQUATION MODELING: QUASI-MAXIMUM LIKELIHOOD VERSUS THIRD MOMENT METHODS

Ryan T. Pohlig, PhD

University of Pittsburgh, 2012

A simulation study was performed comparing the quasi-maximum likelihood (QML; Klein & Muthén, 2007) and third moment (Mooijaart & Bentler, 2010) methods for estimating latent interaction effects in multi-sample structural equation modeling. Both of these methods estimate latent interaction effects without the use of product indicators or the need to specify nonlinear constraints. The purpose of this study was to evaluate the power and type-I error rates for testing group differences of a latent interaction effect. This study also evaluated the parameter recovery of the two methods. A bootstrapping procedure was also proposed for the third moment method that tested the differences of empirical sampling distributions of interaction effects using a two-sample Kolmogorov-Smirnov test. There were four independent variables: i) sample size, ii) non-normality of errors, iii) effect size, and iv) estimation methods.

The QML method performed better than the third moment method. QML had lower type-I error and more power. QML had less absolute bias for estimating smaller interaction effects. For smaller sample sizes, QML had less error in estimating interaction effects, main effects and covariances than the third-moment method. The nonnormality conditions had no impact on the results. Based on the pattern of results found, it is recommended that QML method be used for testing if a latent interaction differs between groups. If the M-B method is to be used the sample size to parameter ratio should be greater than 20:1. Care should be taken in interpreting parameter estimates in the presence of a large interaction effect as both methods overestimated

an interaction coefficient as it increased in magnitude. Both methods also had more error in estimating main effects and covariances as the interaction effect increased.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA - Analysis of Variance

CFI - Comparative Fit Index

DF - Degrees of Freedom

DV - Dependent Variable

EM - Expectation Maximization

GEE - Generalized Estimating Equation

GFI - Goodness of Fit Index

GLM - General Linear Model

GLS - Generalized Least Squares

IV - Independent Variable

K-S - Kolmogorov-Smirnov

LM - Lagrange Multiplier

LMS - Latent Moderated Structural Equations

MACS - Mean and Covariance Analysis

M-B - Moojiaart and Bentler's 3rd moment Method

ML - Maximum Likelihood

MSEM - Multi-sample Structural Equation Modeling

QML - Quasi-Maximum Likelihood Method

RMSD - Root Mean Square Deviation

RMSEA - Root Mean Error of Approximation

SD - Standard Deviation

SEM - Structural Equation Modeling

SRMR - Standardized Root Mean Square Residual

ULS - Unweighted Least Squares

WLS - Weighted Least Squares

WLSA - Augmented Weighted Least Squares

ACKNOWLEDGEMENTS

I would like to thank a number of people for helping me achieve my goal. Firstly, my advisor Kevin Kim, Kevin's constant help throughout all of graduate school was amazing and without his support and friendship none of this would have been possible. His patience and willingness to help enabled me finish this document and the RM program. I am also indebted to him for all the research and learning opportunities that he has given me. I know I speak for a number of graduate students when I express gratitude for all the time he spent helping us.

I am exceedingly grateful to the members of my committee Dr. Clement Stone, Dr. Feifei Ye, and Dr. Jeffery Shook for all the support and feedback they provided. I am grateful to Dr. Ye for all her positivity and kindness throughout my time at Pitt. I enjoyed growing up together with her in the program. I want to thank Dr. Stone for his help and insight through the years. I so admired his lecture style that I have modeled my own mine after it.

I want to extend my thanks to the faculty members of the Research Methodology program, I am glad I got to take a number for courses from each of them. The faculty's friendliness is one reason I enjoyed my time in graduate school. I am grateful to Dr. Lou Pingel for all the advice he gave me when starting my graduate program. I am thankful to Dr. Suzanne Lane, and Dr. Elaine Rubenstein for the opportunities they provided me as a graduate assistant and to Dr. Carol Baker for helping to welcome me to the RM program. I want to thank Dr. Stone for introducing me to the field of IRT and his computer simulations course, which helped me to

complete my dissertation. I am especially grateful to Dr. Kim for introducing me to Structural Equation Modeling, and the statistics courses he and Dr. Ye taught that inspired me to go into the field of statistics.

I want extend my sincerest of thanks to Dr. Jeffery Shook, who stepped far outside of his comfort zone to serve on my committee. The support and guidance he provided me through the dissertation process was invaluable. Over these last four years I have come to think of Jeff as a friend and mentor. I have appreciated all the time we have spent discussing baseball, basketball, football, research, and life. I want to express my gratefulness for all of the research opportunities that Dr. Shook, and his wife Dr. Goodkind have given me. Working on the DHS project with them, Dr. Kim, and David Herring was a highlight of my graduate career.

I want to thank all my fellow Pittsburgh graduate students both inside and outside of the RM program. These years have been truly enjoyable and that is because of the people I have been able to befriend along the way. Within the RM program, I would like to specifically thank Debra, Hong, Laura, Lauren, Priya, Sean, Ting, Tom, and Yun for putting up with me in classes, projects, and around the office. Outside of the RM program, I would like to thank Anthony, Tasha, Ian, Laura, Jon, Nancy, Rich, Michelle, Steve, and anyone else who has provided me with a laugh.

There are a number of people outside of Pittsburgh I would like to thank for their support. Firstly, I want to thank Dr. Cannon who has continued to provide me with the support that started during my undergraduate days. I want to thank Chas for the year we spent in Pitt together adapting to post-undergrad life, talking sports, and always pointing out to me how much worse things could be. I am thankful to my best friends, Mak and Rich, for maintaining our friendships for the past 11 years despite the distance and time constraints. I am appreciative for my Scranton

friends Keith, Mike, Chris, and Ryan for keeping me grounded and providing some of the most memorable moments of my life. I am grateful to still maintain my childhood friendships with Art, Porter, Tom, Josh and Jerre. Lastly, I want to thank Steve, Rich, Christian and the CNR family for making the days pass quickly.

Most importantly, I want to thank my future wife Cari for her support over the last eight years, her belief in me, and willingness to enduring six years of long distance to allow me to finish. I would like to thank and dedicate this work to her, my parents, my sisters, and my extended family for all of their support throughout my schooling.

1.0 INTRODUCTION

1.1 PROBLEM STATEMENT

Quantitative researchers in education and the behavioral sciences have long been concerned with comparing groups and examining moderation. In the past decade, there has been an increased interest in applying these types of analysis to latent variables using Structural Equation Modeling (Henseler & Chin, 2010, Moojiaart & Bentler, 2010, and Preacher, Curran, & Bauer, 2006). Due to this increase in popularity, the research presented here looks to extend analyses enabling researchers to compare group differences in the moderation of latent variables more easily.

The constructs of interest in the social sciences are often complex phenomena, which are not directly measurable but can be assessed through a number of observable indicators. These latent variables and their relationships can be modeled through the use of Structural Equation Modeling (SEM). SEM uses the shared variance among responses on observed variables as indicators of an underlying latent construct (Kim & Bentler, 2006). SEM is a family of techniques that allow for variables' residuals to be modeled enabling researchers to remove measurement error from the underlying constructs of interest (Kaplan, 2009). Both group comparisons and moderation research questions can be addressed using SEM.

There are two types of research questions that can be answered when there is more than one group present in a sample. First, means can be compared allowing researchers to look at

differences in a given set of variables between the groups. Second, the relationships among the variables can be compared between the groups. Group comparisons are important for answering multiple types of research questions, for example: evaluating the effectiveness of a treatment by comparing experimental to placebo conditions, examining differences among demographic categories, or comparing the relationships among a number of factors between groups.

Moderation examines how the relationship between two variables changes in the presence of another; specifically, it tests the differential effect of an independent variable on a dependent variable in the presence of another variable, called a moderator (Barron & Kenny, 1986). For example, social support could moderate the relationship between life stress and illness; it is known that more stress leads to more illness and it is thought that having higher levels of social support could mitigate contracting illnesses (Aiken & West, 1991).

Statistical analyses use interaction effects to test moderation (Kutner, Nachtsheim, Neter, & Li, 2005). An interaction effect is created by multiplying the independent variable (IV) of interest by the hypothesized moderator and using this product term in predicting the dependent variable (DV). The interaction term is interpreted as one in which the effect of the IV on the DV depends upon the level of the moderator and finding significance indicates that the relationship between the IV and DV changes as a function of the moderator (Rosner, 2006). Statistically, it makes no difference which variable is designated as the moderator and which the independent variable (Aiken & West, 1991). This distinction should be determined prior to the analyses and be based on substantive theory.

Independent variables and moderators can be either continuous or categorical, depending upon their level of measurement. There are multiple methods of testing moderation and the procedure used depends upon the level of measurement of the IV and moderator. If both the IV

and moderator are grouping variables then the interaction term created answers the research question, “Is the pattern of differences on the DV among levels of the IV different depending on the level of the moderator?” When the DV is continuous, the easiest method for testing this interaction is through the Analysis of Variance (ANOVA) framework (Keppel & Wickens, 2004). For example, is the pattern of differences on systolic blood pressure between vegan and vegetarian diets significantly different between males and females (Rosner, 2006)? It had been hypothesized that vegans have lower blood pressure than vegetarians, and females had lower blood pressure than males. The research found a significant interaction effect; for males, being vegan led to significantly lower blood pressure, but there was no difference for females.

If either the moderator or IV is continuous then it is simpler to approach the analysis using the framework of multiple regression (Aiken & West, 1991). It should be noted that both regression and ANOVA are specific applications of the General Linear Model (GLM) (Kutner, et al., 2005). If the IV is continuous and the moderator is categorical then the research question answered by the interaction is, “Does the prediction of a DV by an IV change depending upon group membership?” If the IV and moderator are both continuous the research question answered is, “Is the prediction of a DV by an IV changing depending on the level of the moderator?”

Research investigating moderation by multiple variables simultaneously employs higher order interactions. A single product term is created using all of the moderators and the IV of interest. The research question answered by higher order interactions can be generalized from the two-way interaction, “Is the relationship between an IV and DV contingent upon the value or level of two or more other variables?” Higher order interactions are common in GLM via the ANOVA framework (Kutner, et al., 2005; Keppel & Wickens, 2004). Similar to the two-way

interaction, multiple regression is preferred when the IV or moderators are continuous (Dawson & Richter, 2006). The interaction of interest in the current research was examined in multiple regression by Aiken and West (1991) and Dawson and Richter (2006), and is created by two continuous variables and one grouping variable. This can be thought of as testing if the moderation of one continuous variable by another differs between groups.

Examining an interaction using latent instead of observed variables offers a number of advantages. Busemeyer and Jones (1983) showed that the test of an interaction effect made up of observed variables, has low power if any of the variables contain measurement error. Kenny and Judd (1984) and Jaccard and Wan (1995) argued for the use of SEM when testing interaction effects in instances where variables could have considerable measurement error. Barron and Kenny (1986) recommended using SEM for estimating the latent interaction effects.

Multi-sample SEM (MSEM) can be used when a categorical variable is hypothesized to be a moderator of a continuous latent variable, (Rigdon, Schumacker, & Wothke, 1998). These models are ubiquitous in the SEM literature (Kaplan, 1995; 2009, Kline, 2005; Schumacker & Lomax, 1996). Multi-sample models in SEM allow researchers to answer questions concerning measurement invariance, test group differences in the relationships among latent variables or structural invariance, and test if there is any group difference among the error variances and covariances. Researchers using the multi-sample SEM approach for testing interactions use nested model comparisons. A more restrictive model in which parameters are fixed to the same value across groups is tested against a model without the constraints. A significant difference is indicative of moderation (Rigdon et al., 1998). One drawback for using the multi-sample method is that large sample sizes are needed, since the model needs to be estimated for each group (Ping, 2010).

Two types of procedures have been primarily seen in the literature when investigating the interaction between a continuous observed, non-latent, variable and a continuous latent variable (Ping, 2005). The first categorizes the observed, non-latent, variable and then multi-sample SEM is used. The second is to treat the observed variable as a single indicator of a latent factor (Ping, 2005). The problem with categorizing the observed variable is the loss of power that occurs when categorizing and the increase in sample size needed for multi-sample SEM. The problem with the treating a single observed indicator as a latent variable is the assumption that the moderator was measured perfectly, with zero measurement error (Ping, 2005). Alternatively, Muthén and Asparouhov (2003) suggested a random slopes approach to estimate this type of interaction using Full-Information Maximum Likelihood. This approach has been demonstrated to lead to an increase in efficiency and power (Muthén & Asparouhov, 2003).

Kenny and Judd (1984) first formulated a way to estimate and test an interaction effect between two continuous latent variables in SEM using product indicators and non-linear constraints. Product indicators are formed by multiplying the observed indicators from the latent effects that are hypothesized to be interacting. Since Kenny and Judd's original article, a number of different methods have been proposed for examining latent interactions. Ping (2010) mentions 14 different methods that have been proposed in the last twenty years. Two newer modern approaches have been given by Klein and Muthén (2007) and Moojiaart and Bentler (2010). Both methods attempt to model the nonnormality present in interaction models directly. An advantage these two modern approaches have is that they do not rely on product indicators or non-linear constraints. A second advantage is their ease of implementation for applied researchers.

Klein and Muthén (2003; 2007) proposed a Quasi-Maximum Likelihood (QML) method. This method approximates the joint distribution of the indicators by a product of a normal and a conditionally normal distribution. Then a conditional variance model is used to estimate the parameters. Marsh, Wen, and Hau (2004; 2006) and Klein and Muthén (2003; 2007) examined the performance of QML against that of the product indicator approaches and found when the indicators were normally distributed QML had smaller standard errors, less bias, and more power in estimating the interaction term.

Moojiaart and Bentler's (2010) method (M-B) models the nonnormality seen in the data caused by the nonlinear effect by estimating third order sample moments. Third order moments are estimates of skewness and are created using products of the deviations of the indicators. The M-B method includes the third order moments along with means and covariances in the estimation process and can be thought of as an extension of Mean and Covariance Analysis (MACS). Similar to QML, it does not require researchers to create product indicators or apply non-linear constraints. Moojiaart and Bentler (2010) found that both QML and their method produced similar parameter estimates and bias, while QML produced smaller estimates of the standard error for the interaction. Moojiaart and Bentler (2010) suggest that bootstrapping the interaction coefficient could aid in its estimation.

Bootstrapping can produce more accurate estimates than those garnered from standard normal theory or obtained by asymptotic formulas (Efron & Tibshirani, 1993). Bootstrapping treats the observed data as a population and repeatedly samples with replacement from it. After each sample is taken, the statistical model is run calculating the statistic of interest. These statistics are saved and collectively form an empirical sampling distribution, enabling a confidence interval to be found. The standard deviation of the empirical sampling distribution is

an estimate of the population parameter's standard error, when bootstrap samples are of the same size as the observed sample. Bollen and Stine (1992) proposed a model based bootstrapping method that could be implemented in SEM.

Higher order interactions in SEM have been minimally investigated in published research. Ping (2004) recognized this hole in the literature, "…except for the suggestions in this monograph regarding cubics, there is no guidance for the proper specification of these variables [three way interactions] using structural equation analysis." Ping (2010) first offered a method for estimating a higher order interaction comprised of three continuous latent variables, using a single product indicator method. Ping (2010) noted that three way interactions, "have yet to appear in published SEM models, perhaps because there is little guidance for estimating them." A literature review found only one application of a three-way interaction in the applied SEM literature, an article by Bakker, Shimazu, Demerouti, Shimada, and Kawakami (2011). Bakker et al. adapted a product indicator approach proposed by Mathieu, Tannenbaum, and Salas, (1992) (as cited in Cortina, Chen, & Dunlap, 2001). The method adapted by Bakker et al. was not the preferred or recommended method as it produced estimates different than the other reviewed procedures (Cortina et al., 2001). A literature review turned up no instances of estimating a group difference in an interaction term made up of two continuous latent variables, using multi-sample SEM.

The research presented here looks to extend SEM analyses enabling researchers to compare group differences in the moderation of latent variables by extending the two modern methods. This research also adds bootstrapping to the third moment method to examine if it yields any added accuracy. This three-way interaction has not been specifically seen in the latent variable literature, but moderation of latent constructs by categorical variables has been

examined extensively through multi-sample SEM (Kaplan, 2009), as has continuous latent variable moderation (Ping, 2010). Three-way interaction effects are ubiquitous in non-latent variable statistical methods (e.g. multiple regression, and ANOVA see Aiken & West, 1991; Dawson & Richter, 2006).

The extension of these techniques to model three-way interactions within SEM will allow applied researchers to test these effects without measurement error. The impact of non-normality of main effects on the two modern methods has not been investigated. Nor has the direct comparison of power and type I error rates of the two methods been examined across different conditions. From a theoretical perspective, if examining moderation in continuous latent variables is of interest, evidenced by the multitude of methods created for testing these effects (see Ping, 2010; Schumacker & Marcoulides, 1998) it would seem inherent that comparing these estimates of moderation among groups should also be of interest.

1.2 RESEARCH QUESTIONS

Multi-sample and multi-group comparisons, as well as moderation are common in SEM. Higher order interactions have been minimally investigated in the SEM literature but are common in regression. By using latent variables any potential impact that measurement error might have on the interaction effect should be eliminated. The research presented here, looks to extend SEM analyses enabling researchers to compare group differences in a latent interaction effect and adds bootstrapping to the third moment method. By using the modern methods of examining latent interaction effects, no product indicators have to be created or non-linear constraints have to be

specified enabling applied researchers to use the methods without requiring advanced technical knowledge.

The main focus in this paper is to compare the performance of the modern methods of examining latent interaction effects using QML and the M-B method with and without bootstrapping. Specifically, the research questions to be addressed are:

1) Is the pattern of differences in Power and Type-I error rate for the interaction effect among the methods, QML, M-B, and M-B with bootstrapping, different among the levels of the factors manipulated in the study (Sample Size/Sample Size Ratio, Interaction Effect Size, and Distribution of Indicators)?
2) Is the pattern of differences in bias and absolute bias of the estimated parameters different for methods among the levels of the factors manipulated in the study (Sample Size/Sample Size Ratio, Interaction Effect Size, and Distribution of Indicators)?

2.0 REVIEW OF LITERATURE

2.1 MODERATION IN MULITPLE REGRESSION

2.1.1 Moderation and Mediation

Both moderation and mediation are important to research in the social and hard sciences. Moderation and mediation effects have historically been designated as potential threats to validity as both employ variables that could be classified as confounders (Mackinnon, Krull, & Lockwood, 2000). The distinction between them needs to be emphasized as the two phenomena answer different research questions. Stated simply, moderators influence the relationship between other variables and mediators explain the relationship between variables (Barron & Kenny, 1986).

Moderation occurs when the relationship between two variables is influenced by a third variable (*see figure 1*). The effect of the IV on the DV (path A') changes depending upon value of the moderator. Mathematically, there is no distinction between moderators and IVs. The title given to a specific variable comes from the research design and the investigator defines which variable is the moderator and which is the IV. The product term created to represent moderation does not distinguish between the role of moderator and IV. The interaction term provides an estimate that the combination of the two variables accounts for an amount of variability in the

DV above and beyond, or after controlling for, the main effects (Little, Card, Bovaird, Preacher, & Crandall, 2007). In this document both moderators and IVs will be referred to as IVs.

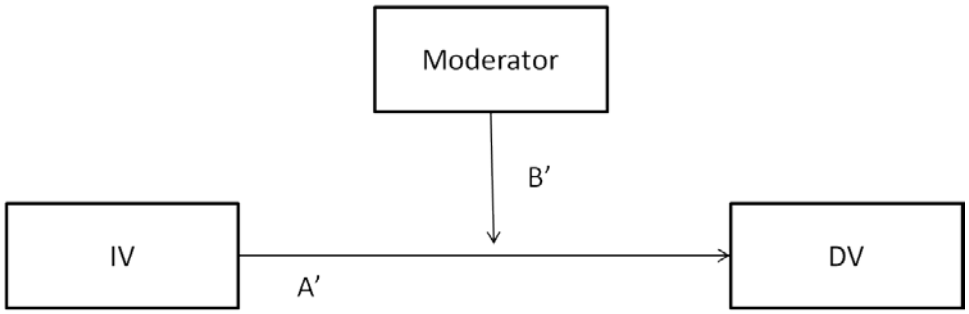

Figure 1. Path Diagram of Moderation

Mediation occurs when the effect on a DV by an IV can be explained by an intervening variable (Marsh et al., 2006) (*see figure 2*). Barron and Kenny (1986) state, "a given variable may be said to function as a mediator to the extent that it accounts for the relation between the predictor [IV] and the criterion [DV]." In this analysis, an IV predicts both the DV (path C) and the mediator (path A), and the mediator predicts the DV (path B).

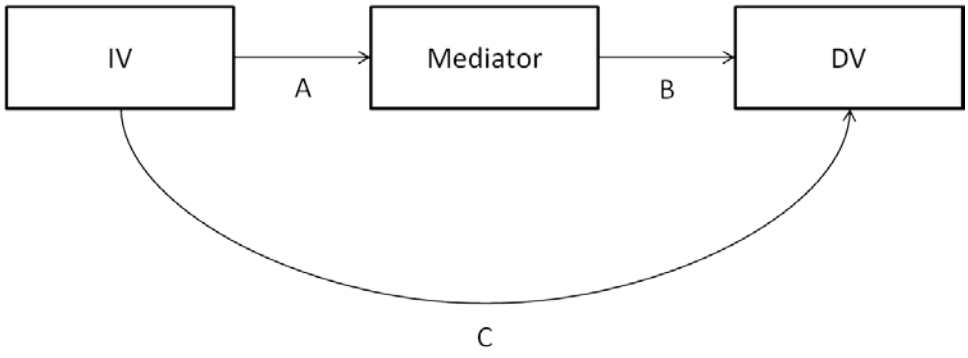

Figure 2. Path Diagram of Mediation

Mediation has occurred when the effect of the IV on the DV can at least partially be accounted for by the mediator. Four criteria should be met for mediation: 1) the IV should significantly predict the mediator, 2) the IV should significantly predict the DV, 3) the mediator should significantly predict the DV, and 4) after controlling for 1 and 3, the relationship between

the IV and the DV, becomes non-significant or significantly decreased (Barron & Kenny, 1986).

Mediation can be tested for observed variables through Path Analysis, a subset of SEM, or through a series of MR models by examining the significance of the indirect effect (Tabachnick, & Fidell, 2007). The indirect effect consists of the path from the IV to the mediator (path A) and from the mediator to the DV (path B) and its estimate is the product of these two paths. It can be tested for significance by dividing the estimate by its standard error. The standard error can be obtained using Sobel's approximation or through bootstrapping (Sobel, 1982). Shrout and Bolger (2002) argued that even if the distribution of direct effects (path A & B) are normally distributed, the indirect effect is often not normally distributed. Therefore, they recommended a bootstrap method be used to test for indirect effect, which would allow for the calculation of a non-symmetric empirical confidence interval.

2.2 REGRESSION MODEL AND INTERACTION TERM

A multiple regression model with two IVs and their interaction can be expressed as,

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon, \quad (1)$$

where the DV is Y and its predicted value is $\hat{Y}$. The IVs are X and Z. β_1, β_2, β_3 are the regression coefficients predicting Y by each IV and their interaction term and β_0 is the intercept. ε is the error term with an expected value of zero. The slopes of the coefficients are tested for significance. Both X and Z could be continuous or a dummy coded categorical indicator, corresponding to the variable's level of measurement.

2.2.1 Interpretation and Testing the Coefficients

If the interaction term is not included in the model, then the MR equation would only contain main effects and (1) becomes $Y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon$. The slopes in the model with only main effects can be interpreted as the amount of change in the DV that occurs for a one unit increase in an IV after adjusting for the relationship between X and Z. The slope estimates' interpretation becomes more complicated with the inclusion of the interaction term. The prediction of the DV by an IV is now dependent upon the value of the other IV. For example, the prediction of Y by X depends upon the value of Z; equation (1) can be reformatted to show this by isolating X,

$$Y = (\beta_0 + \beta_2 Z) + (\beta_1 + \beta_3 Z)X + \varepsilon. \quad (2)$$

In equation (2) X's simple slope is

$$(\beta_1 + \beta_3 Z).$$

This slope regressing Y on X includes X's main effect coefficient, β_1, and the interaction coefficient, β_3. It can be seen that X has a different simple slope for each value of Z, termed a conditional effect (Aiken & West, 1991). The research question addressed by the interaction effect is, "Is there a significant amount of change in X's simple slope for one unit increase in Z?" The inclusion of the interaction term also creates a simple or conditional intercept for X

$$(\beta_0 + \beta_2 Z).$$

Equation (2) could just as easily be rearranged isolating Z and its simple slope discussed, showing there is no statistical difference between a "moderator" and "IV" in this framework.

There are different strategies for interpreting the interaction depending upon the level of measurement of the IVs. If either of the IVs are categorical then it is simple to estimate a

separate regression equation for each group. For example, if Z were dichotomous then two simple regression equations could be used, one for each group. The interaction term tested would indicate if there is a significant difference in prediction between the two groups. Z would be zero for the first group and equation (2) simplifies to

$$group = 1: Y = [\beta_0 + \beta_2 Z] + [\beta_1 + \beta_3 Z]X + \varepsilon$$

$$Y = [\beta_0 + \beta_2(0)] + [\beta_1 + \beta_3(0)]X + \varepsilon$$

$$Y = \beta_0 + \beta_1 X + \varepsilon.$$

For the second group, Z would equal one and equation (2) is

$$group = 2: Y = [\beta_0 + \beta_2 Z] + [\beta_1 + \beta_3 Z]X + \varepsilon$$

$$Y = [\beta_0 + \beta_2(1)] + [\beta_1 + \beta_3(1)]X + \varepsilon$$

$$Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X + \varepsilon.$$

If both of the IVs are continuous a few different strategies have been proposed for probing and testing the interaction. The Simple-Slopes approach consists of selecting different values for the conditional variable to evaluate significance (Aiken & West, 1991; Preacher et al., 2006). This procedure uses the conditional slope as defined in equation (2), and divides that estimate by its standard error, yielding a test statistic. The standard error of the simple slope can be found by taking the square root of the variance of the simple slope. The variance of the simple slope for X can be found by

$$VAR(\beta_1) + 2ZCOV(\beta_1, \beta_3) + Z^2 VAR(\beta_3),$$

where $VAR(\cdot)$ and $COV(\cdot)$ are the variances and covariances found from the asymptotic covariance matrix (Preacher et al., 2006). The drawback to this approach is that the choice of conditional values tested is arbitrary, although industry standard is to test the mean, +1 standard deviation (SD), and -1 SD of Z (Aiken & West, 1991; Preacher et al., 2006). An alternative is the

Johnson-Neyman approach, which can be thought of as the converse of the simple slopes technique. This procedure is done by specifying the critical ratio for a given degrees of freedom (DF) and then finding the two values of the conditional variable that yield it (Preacher et al., 2006). The Johnson-Neyman technique will return a range of values of Z in which the simple slope of X is significantly different from zero.

An interaction formed from uncentered variables is correlated with the main effects that are multiplied to create it (Aiken & West, 1991). This causes arbitrary multicollinearity. The standard errors of the first order terms are inflated because they are confounded with the standard errors from the higher order effects (Aiken & West, 1991). Multicollinearity also creates instability in the estimation of the regression slopes (Little, Bovaird, & Widaman, 2006). This problem is due to the scaling of the IVs and can be fixed by centering them before creating the interaction term. Centering the variables also increases the interpretability of the regression coefficients. The estimates produced by the centered variables are more practical, since they indicate the change in the slope of the X predicting the Y at the mean of Z. Left uncentered, the estimate would indicate the change in slope of X predicting Y when Z has a raw score of zero, potentially a meaningless value of Z (Little et al,, 2006).

2.2.2 Non-normality

One problem created by using interaction terms in statistical models is that the variables created to estimate them are not normally distributed. A product term will most likely not be normal even if the variables creating it are normally distributed (Dimitruk, Schermelleh-Engel, Kelava, & Moosbrugger, 2007). Bohnstedt and Goldberger (1969) showed that the covariance of a product term (XZ) with the another variable (D), is

$$COV(XZ, D) = E(xdz) + COV(X, D)E(Z) + COV(Z, D)E(X),$$

with $d = D - E(D)$. If multivariate normality is assumed the third moment is zero, $E(xdz) = 0$ and if X and Z are centered their expected value is 0, giving $COV(X, D)E(Z) = 0$ and $COV(Z, D)E(X) = 0$. This shows that if the variables are multivariate normal, the covariance of the product term and the other variable is zero. Aiken and West (1991) point out that if the three variables are multivariate normal then by default there is no interaction effect, arguing that if there is an interaction the three variables cannot be jointly normally distributed.

2.2.3 Higher Order Interactions

Higher order interactions are those that involve three or more variables simultaneously. In MR, these variables can be any combination of categorical and continuous. Higher-order interaction effects are created by multiplying all of the IVs or moderators together, similar to the two variable interaction discussed above. This product term is then used to predict the DV, adjusting for all main effects and lower order interactions involving the variables that created the higher order interaction. Cohen (1978) pointed out that higher order interaction terms only estimate what they purport to if and only if the model is adjusted for the lower order effects. The higher order interaction term would be overestimated when the lower order terms are not included. If all the IVs are categorical, the model is simplified and the analysis becomes a factorial ANOVA. The example discussed below is relevant to the current research and is a three-way interaction involving two continuous and one categorical variable.

In a three-way interaction, the relationship between an IV, X, and a DV, Y, is contingent not only on the other IVs, Z and W, but their interaction (ZW) as well (Dawson & Richter, 2006). The regression model takes the form of

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_5 XW + \beta_6 ZW + \beta_7 XZW + \varepsilon. \quad (3)$$

In equation (3), Z is a dummy variable that takes the value of one or zero indicating group membership. All main effects of continuous IV's should be centered and the centered variables are used to create the interaction terms. The simple slope and intercept for the regression of Y on X can be found by isolating X and rearranging equation (3) to

$$Y = (\beta_0 + \beta_2 Z + \beta_3 W + \beta_6 ZW) + (\beta_1 + \beta_4 Z + \beta_5 W + \beta_7 ZW)X + \varepsilon.$$

The regression of Y on X depends upon the values of Z and W, along with their interaction. Paralleling the procedure for the two-way interaction model, the simple slope can be tested for significance at different values of Z and W. Aiken and West (1991) suggest one standard deviation above and below the mean. Dawson and Richter (2006) provided an alternative method to testing the interaction by examining differences in simple slopes by comparing that difference relative to its standard error.

The slopes' interpretation changes when incorporating a three-way interaction term, differing from even the simple interaction model above. The centered main effects can still be thought of as the conditional effect of an IV at the mean of the other IVs (Aiken & West, 1991). In equation (3), β_1 indicates the effect of X when Z and W are equal to their mean. The two-way interaction effects are now conditional effects evaluated when the IV that was not included in the product term is equal to zero. β_4 is the conditional effect of the X^*Z interaction when W is equal to its mean. The three-way interaction coefficient, β_7, indicates that the relationship between an IV, X, and a DV, Y, varies across levels of the other IVs, Z and W, and their interaction (Dawson & Richter, 2006; Aiken & West, 1991).

It is simple to estimate a separate regression equation for each group, when one of the IVs is categorical. If Z were dichotomous, two equations would be needed, one for each group. Z would be zero for the first group and equation (3) becomes

$$group1: Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_5 XW + \beta_6 ZW + \beta_7 XZW + \varepsilon$$

$$Y = \beta_0 + \beta_1 X + \beta_2(0) + \beta_3 W + \beta_4 X(0) + \beta_5 XW + \beta_6(0)W + \beta_7 X(0)W + \varepsilon,$$

simplifying to

$$Y = \beta_0 + \beta_1 X + \beta_3 W + \beta_5 XW + \varepsilon. \quad (4)$$

Z would equal one for the second group and (3) becomes

$$group2: Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_5 XW + \beta_6 ZW + \beta_7 XZW + \varepsilon$$

$$Y = \beta_0 + \beta_1 X + \beta_2(1) + \beta_3 W + \beta_4 X(1) + \beta_5 XW + \beta_6(1)W + \beta_7 X(1)W + \varepsilon$$

$$Y = \beta_0 + \beta_2 + (\beta_1 + \beta_4)X + (\beta_3 + \beta_6)W + (\beta_5 + \beta_7)XW + \varepsilon. \quad (5)$$

The model, for the first group, is the same as the simple interaction identified in equation (1). The test of the three-way interaction would determine if there is a significant difference in the prediction of Y between groups using equations (4) and (5). This answers the research question, "Is there a significant difference in the moderation of X by W on Y between the two groups?"

2.3 STRUCTURAL EQUATION MODELING

Structural Equation Models are regression models comprised of two parts, a measurement component that relates the observed variables to hypothesized latent ones and a structural component that models the relationships among the latent variables and other covariates. SEM enables researchers to test how well a hypothesized model fits the collected data. SEM also allows researchers to test specific pathways for significance. Direct and indirect pathways

between observed and latent variables can be tested. SEM should be used to test models reflective of substantive theory (Kim & Bentler, 2006).

2.3.1 Latent Variables

When measuring any trait in social sciences it is important to remember that instruments are flawed, and measurement error will occur. Researchers hypothesize that the responses seen on a related set of items is reflective of or caused by a common underlying latent construct. Latent variables could be conceptualized as being what the observed indicators have in common, or their shared variance (MacCallum & Mar, 1995). It is important to consider more than one item, which reduces the impact of error in any one item (Kline, 2005). There are a number of ways to define what a latent variable represents, from the expected true-score value in classical test theory to the more inclusive sample-realization definition proposed by Bollen (2002). Unlike other statistical methods, SEM allows researchers to directly model errors (Kim & Bentler, 2006). This enables researchers to garner an estimate of the latent trait free of measurement error, or with perfect reliability. SEM allows researchers to measure these underlying traits or factors, and to model their relationships using a series of equations.

2.3.2 Model

The following is adapted from Bollen (1989) and Kim (2009). In SEM, there are two measurement models each expressed with its own equation in LISREL notation. One for the exogenous variables,

$$x = \tau_x + \Lambda_x \xi + \delta, \tag{6}$$

and one for the endogenous variables,

$$y = \tau_y + \Lambda_y \eta + \varepsilon. \quad (7)$$

x is a $q \times 1$ vector of exogenous indicator variables, and y is a $p \times 1$ vector of endogenous indicator variables. Λ_x is $q \times n$ matrix of regression coefficients predicting x by ξ and Λ_y is $p \times m$ matrix of regression coefficients predicting y by η. τ_x is a $q \times 1$ vector of x-intercepts and τ_y is a $p \times 1$ vector of y-intercepts. δ is a $q \times 1$ vector of measurement errors of x, and ε is a $p \times 1$ vector of measurement errors of y. Lastly, there are two covariance matrices associated with the measurement part of the LISREL model, Θ_δ which is a $q \times q$ covariance matrix of δ, $E(\delta\delta')$ and Θ_ε which is a $p \times p$ covariance matrix of ε, $E(\varepsilon\varepsilon')$.

The structural model is

$$\eta = \alpha + B_0 \eta + \Gamma_1 \xi + \zeta. \quad (8)$$

η is an $m \times 1$ vector of latent endogenous variables and ξ is an $n \times 1$ vector of latent exogenous variables. B_0 is an $m \times m$ matrix of regression coefficients for the latent endogenous variables with zeros on the diagonal and Γ_1 is an $m \times n$ matrix of regression coefficients, γ, predicting η by ξ. α is an $m \times n$ matrix of intercepts. ζ is a vector of disturbances. There are two covariance matrices associated with the structural model, Φ, which is a $n \times n$ covariance matrix of ξ, which is $E(\xi\xi')$, and Ψ, which is a $m \times m$ covariance matrix of ζ, which is $E(\zeta\zeta')$. The parameters estimated in SEM comprise the non-redundant elements of the following vectors and matrices, $\theta = \{\alpha,\ \tau_x,\ \tau_y,\ \kappa,\ B_0,\ \Gamma_1,\ \Lambda_x,\ \Lambda_y,\ \Theta_\delta,\ \Theta_\varepsilon,\ \Phi,\ \Psi\}$.

There are three statistical assumptions required for SEM. First, the expected values of the errors and disturbances are zero, $E(\zeta) = E(\delta) = E(\varepsilon) = 0$. Second, the covariances of errors and disturbances with other parameters are zero, $COV(\zeta, \xi) = 0$, $COV(\varepsilon, \eta) = COV(\varepsilon, \xi) = COV(\varepsilon, \delta) = 0$, and $COV(\delta, \eta) = COV(\delta, \xi) = 0$. Lastly, B is non-singular, where $B = I - B_0$.

The means and covariances of the indicators can be found using the model equations. The means of x are given by $x = \tau_x + \Lambda_x\kappa$, with κ being an $n\ x\ 1$ vector of the means of the latent exogenous variables, $E(\xi) = \kappa$. The means of y are given by $E(y) = \tau_y + \Lambda_y B^{-1}(\alpha + \Gamma_1\kappa)$, and the means of η are $E(\eta) = B^{-1}(\alpha + \Gamma_1\kappa)$. The model covariance matrix, $\Sigma(\theta)$, is made up for four sub-matrices

$$\Sigma(\theta) = \begin{pmatrix} \Sigma_{xx}(\theta) & \Sigma_{yx}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{yy}(\theta) \end{pmatrix}.$$

Each of the sub matrices can be expressed using the measurement and structural models given in (6) thru (8), by centering the variables and applying covariance algebra:

$$\Sigma_{xx}(\theta) = E(xx') = E[(\Lambda_x\xi + \delta)(\Lambda_x\xi + \delta)']$$

$$= \Lambda_x\Phi{\Lambda_x}' + \Theta_\delta$$

$$\Sigma_{yx}(\theta) = E(yx') = E\left[\left(\Lambda_y\eta + \varepsilon\right)(\Lambda_x\xi + \delta)'\right]$$

$$= \Lambda_y B^{-1}\Gamma_1\Phi{\Lambda_x}'$$

$$\Sigma_{xy}(\theta) = E(xy') = E\left[(\Lambda_x\xi + \delta)\left(\Lambda_y\eta + \varepsilon\right)'\right]$$

$$= \Lambda_x\Phi{\Gamma_1}'B^{-1}\Lambda_y$$

$$\Sigma_{yy}(\theta) = E(yy') = E\left[\left(\Lambda_y\eta + \varepsilon\right)\left(\Lambda_y\eta + \varepsilon\right)'\right],$$

substituting, $\eta = B^{-1}\Gamma_1\xi + \zeta$

$$= \Lambda_y B^{-1}(\Gamma_1\Phi{\Gamma_1}' + \Psi)(B^{-1})'{\Lambda_y}' + \Theta_\varepsilon.$$

The goal of SEM is to model the relationships between observed and latent variables in such a way that the observed covariance matrix is reproduced by the model covariance matrix. The null hypothesis is the covariance matrix produced by the model is equal to the population covariance matrix, $H_0: \Sigma = \Sigma(\theta)$. Here Σ is the true population covariance matrix, and Σ is unknown without specifically measuring each unit in the population, because of which *S* is used

as its proxy. S is the unbiased sample covariance matrix found using the observed data. An implicit assumption of SEM is that observed covariance matrix is an adequate estimate of the population covariance matrix and this can be met when the sample accurately represents the population you want to generalize to. In addition, null hypotheses concerning individual pathways and parameters can be investigated by directly testing them.

2.3.3 Model Fit Functions

A variety of fit functions have been proposed to test the null hypothesis, $H_0: \Sigma = \Sigma(\theta)$, which is assessed by testing $\Sigma(\theta) = S$. They can be expressed using the Generalized Least Squares (GLS) estimation function

$$F\big(s, \sigma(\theta)\big) = F_{GLS} = \Big[\big(s - \sigma(\theta)\big)' W \big(s - \sigma(\theta)\big)\Big] + \Big[\big(\bar{z} - u(\theta)\big)' S^{-1} \big(\bar{z} - u(\theta)\big)\Big], \qquad (9)$$

with $s = vech(S)$ and $\sigma(\theta) = vech\ (\Sigma(\theta))$. $F(\cdot)$ is a discrepancy function and $vech(\cdot)$ creates a vector of unique elements contained in a matrix (Magnus & Neudecker, 1999). $z = (y', x')'$ is a joint vector of all the indicators, with the vector of means being given by $\bar{z} = (\bar{y}', \bar{x}')'$. $u(\theta)$ is the vector of estimated means obtained from the model. A null matrix is returned if the model fits perfectly. The weight matrix, W, is changed to reflect the estimation procedure being used (Bentler & Dijkstra, 1985). If an identity matrix is chosen for the weight matrix, $W = I$, then the estimation procedure used is the Unweighted Least Squares (ULS) function. ULS is known to calculate unbiased parameter estimates, but requires the assumption that the disturbances are homoscedastic; otherwise incorrect standard errors are produced (Kaplan, 2009).

Alternatively, the Weighted Least Squares (WLS) fit function could be chosen, where the weight matrix is the inverse of the sample covariance matrix, $W = S^{-1}$ (Browne, 1984). The

weight matrix in WLS has elements that are a function of fourth order sample moments from the residuals of the indicators (Browne, 1984). Using the WLS function addresses the problem of nonnormality. Accurate standard errors and chi-square values are obtained regardless of the shape of the joint distribution of the indicators (Kaplan, 2009). This estimation procedure does not require the assumption of normality, but it does require that the mean vector be independent from the covariance matrix, which is not met in interaction models if the variables are not centered (Jonsson, 1998). When the data are multivariate normal, WLS has been shown to have the same properties as Maximum Likelihood (ML), asymptotic normality and efficiency (Kaplan, 2009). If normality is met the fit function can be expressed as

$$F_{WLS} = \frac{1}{2} tr\left[I - \left(S^{-1}\,\Sigma(\theta)\right)\right]^2$$

(Kaplan, 2009).

Augmented Weighted Least Squares (WLSA) estimation is used in MACS models since means and covariances could be correlated. An augmented moment matrix, A, is a matrix of moments with its elements including the covariance matrix, the vector of indicator means, and a constant. The observed sample augmented moment matrix is

$$A = \begin{pmatrix} S & \bar{z} \\ \bar{z}' & 1 \end{pmatrix},$$

and the model estimated augmented moment matrix is

$$\hat{\alpha}_m = \begin{pmatrix} \Sigma(\hat{\theta}) & \hat{u} \\ \hat{u}' & 1 \end{pmatrix}.$$

The fit function changes equation (9) and is now expressed as $F_{WLSA} = (a - \alpha)' W_a^- (a - \alpha)$, with $a = vech(A)$ and $\alpha = vech\,(\alpha_m)$ and W_a^- is the generalized inverse, which is needed since the last row is a constant and the matrix is singular (Jöreskog & Yang, 1996).

The ML method updates the weight matrix after each iteration as the estimates of the model covariance matrix, $\Sigma(\theta)$, change. The ML discrepancy function can be found directly by

$$F_{ml} = \log|\Sigma(\theta)| - \log|S| + \text{trace}(S\Sigma(\theta)^{-1}) - (p+q) + \big(\bar{z} - u(\theta)\big)' S^{-1}\big(\bar{z} - u(\theta)\big).$$

This fit function relies on the assumption that z is multivariate normal (Kaplan, 2009).

For all fit functions, the model chi-square value will approximate a central chi-square distribution if the null hypothesis is true, indicating that the model fits. The model chi-square will follow a noncentral chi-square distribution if the null hypothesis is false and the model does not fit. In general the fit function's test statistic is calculated by $\chi^2 = (N-1)F_{GLS}$, which is asymptotically distributed as a noncentral chi-square with degrees of freedom equal to the number of free parameters minus the number of estimated parameters, $df = \frac{1}{2}(p+q+1)(p+q) - t$, with t being the total number of parameters in the model. The noncentrality parameter, δ_{np}, is $(N-1)F\big(\Sigma, \Sigma(\theta_0)\big)$, where θ_0 is chosen to minimize $F(\Sigma, \Sigma(\theta))$ over choices of θ (Nevitt & Hancock, 2001). The noncentrality parameter, δ_{np}, is an estimate of an error of approximation that specifies the degree of falsehood of the null hypothesis (Hu & Bentler, 1995). A perfect fitting model, or one that is just-identified should produce a $\chi^2 = 0$, which would indicate perfect fit and its $\delta_{np} = 0$. As the discrepancy between the observed and modeled covariance matrices increase both the χ^2 and δ_{np} will increase, indicating a worse fit.

There are a few assumptions that need to be satisfied for the fit tests to be considered accurate. The indicators should be jointly normally distributed for ML and normal theory estimates to be accurate (Kaplan, 2009). If data is missing, then it should missing at random or missing completely at random (Kaplan, 2009). There should be no specification error, or important omitted covariates. Omitting important covariates could cause the exogenous latent variables to be correlated with error terms (Kaplan, 2009). ML estimation is typically

inappropriate for models that include an interaction, as the chi-square test fails to detect the misfit of nonlinear models (Mooijaart & Satorra, 2009).

2.3.4 Fit Indices

Model chi-square goodness-of-fit tests are known to be biased depending upon sample size (Kim & Bentler, 2006). Large samples lead researchers to conclude that models do not fit, by producing a significant chi-square value, potentially making a type-I error (Hu & Bentler, 1995). Small samples will often lead researchers to conclude that their hypothesized model fits well when it should not, a type-II error (Hu & Bentler, 1995). Given these well known inadequacies of the chi-square test a multitude of fit indices that are not as influenced by sample size have been proposed (Kaplan, 2009; Kline, 2005).

Kline (2005) recommends reporting Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Comparative Fit Index (CFI), and Goodness of Fit Index (GFI). RMSEA takes into account sample size and is an estimate of the amount of error in the model per one degree of freedom (Kline, 2005). Good fit using the RMSEA is between.05 to .08; a value greater than .1 would indicate poor fit. SRMR is an index that calculates the mean absolute value of the correlation residuals (Kline, 2005). SRMR is a standardized metric, so both the observed covariance matrix and the model covariance matrix have to be transformed into correlation matrices. A good fit using SRMR is a value of .08 or less. CFI is an incremental fit index, which compares the noncentrality parameter obtained from the model covariance matrix to that of a baseline model (Kline, 2005). The CFI will indicate the relative improvement of the researcher's hypothesized model over a null model, and a good fit using this index would be a value of .95 or greater. GFI is an absolute fit index that uses the

difference between the observed covariance matrix, S, and modeled covariance matrix, $\Sigma(\theta)$, relative to the model covariance matrix (Hu & Bentler, 1995). A value of .90 or greater signifies a good fit using the GFI.

2.4 PRODUCT INDICATOR METHODS FOR ESTIMATING LATENT INTERACTION EFFECTS

SEM allows researchers to model not only observed interaction effects like multiple regression and ANOVA, but allows for interactions involving latent variables. There are three types of interactions in SEM: between two observed variables, between a latent and an observed variable, and between two latent variables. Similar to regression, the problem of nonnormality is present. Interaction terms are not normally distributed even if the latent variables creating them are. The interpretation of the parameters produced in a latent interaction effect model is the same as that in regression. Barron and Kenny (1986) suggested using the Kenny and Judd method to test interactions via latent variables in order to eliminate measurement error. Jaccard & Wan (1995) found that using a latent variable approach controlled Type I and Type II error better than the multiple regression approach, when testing an interaction effect with measurement error present. Similarly, Bispe, Coenders, Saris and Batista-Foguet (2006) concluded that methods that correct for measurement error in interaction models perform better than those that do not, recommending an SEM approach for large samples.

2.4.1 Kenny-Judd's Method

Kenny and Judd (1984) first formulated a way to estimate and test latent interaction effects in SEM. Their model creates product indicators, which are the products of the observed indicators from the latent effects that are interacting. These product indictors load onto a new latent variable, which represents the interaction. By applying certain nonlinear constraints, the loadings of the product indicators are formulated by multiplying structural parts of the model so no new loadings have to be estimated (Kenny & Judd, 1984). Kenny and Judd expressed their regression equation as

$$y = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta, \tag{10}$$

where y is a centered observed dependant variable. ξ_1 and ξ_2 are latent exogenous variables which have been centered, $\xi_1\xi_2$ is the latent interaction term between ξ_1 and ξ_2. ζ is the disturbance. In the original Kenny and Judd model all the latent variables had two indicators, with the first loading on each fixed to one in order to identify the model. The indicators were expressed using

$$\begin{aligned} x_1 &= \lambda_1\xi_1 + \delta_1, \\ x_2 &= \lambda_2\xi_1 + \delta_2, \\ x_3 &= \lambda_3\xi_2 + \delta_3, \text{ and} \\ x_4 &= \lambda_4\xi_2 + \delta_4. \end{aligned} \tag{11}$$

Kenny and Judd suggested using all four cross-products of observed variables as indicators of the latent interaction. The product indicators would be: $x_1x_3, x_1x_4, x_2x_3,$, and x_2x_4, and load on the latent interaction $\xi_1\xi_2$. Kenny and Judd (1984) had to impose nonlinear constraints on the product indicators in order to estimate the model, an example of which is

$$x_1 x_3 = \lambda_1 \lambda_3 \xi_1 \xi_2 + \lambda_1 \xi_1 \delta_3 + \lambda_3 \xi_2 \delta_2 + \delta_1 \delta_3. \quad (12)$$

This method requires that the observed and latent variables be centered and nonlinear constrains are applied to the factor loadings and variances. The latent variables are assumed to be normally distributed (Kenny & Judd, 1984; Rigdon et al., 1998). Due to the fact that the product indicators are nonnormal ML estimation is inappropriate. Kenny and Judd (1984) proposed using WLS estimation.

This method can be generalized to a latent endogenous variable, η, with multiple observed predictors (*see figure 3*). The generalized Jöreskog and Yang (1996) specification for the original Kenny and Judd measurement model can be expressed using LISREL notation,

$$y_1 = \tau_{y1} + \lambda_{y1} \eta + \varepsilon_1,$$

$$y_2 = \tau_{y2} + \lambda_{y2} \eta + \varepsilon_2,$$

the first loading for each latent variable is fixed at one, $\lambda_1 = \lambda_3 = 1$, for identification,

$$x_1 = \lambda_1 \xi_1 + \delta_1,$$

$$x_2 = \lambda_2 \xi_1 + \delta_2,$$

$$x_3 = \lambda_3 \xi_2 + \delta_3, \text{ and}$$

$$x_4 = \lambda_4 \xi_2 + \delta_4.$$

The structural model is

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

In order to estimate the model, the error variances for the product indicators must be constrained. For example,

$$VAR(\delta_{13}) = \lambda_1^2 VAR(\xi_1) VAR(\delta_3) + \lambda_3^2 VAR(\xi_2) VAR(\delta_1) + VAR(\delta_1) VAR(\delta_3).$$

Joreskog and Yang (1996) changed equation (12) to:

$$x_1 x_3 = \lambda_1 \lambda_3 \xi_1 \xi_2 + \delta_{13}.$$

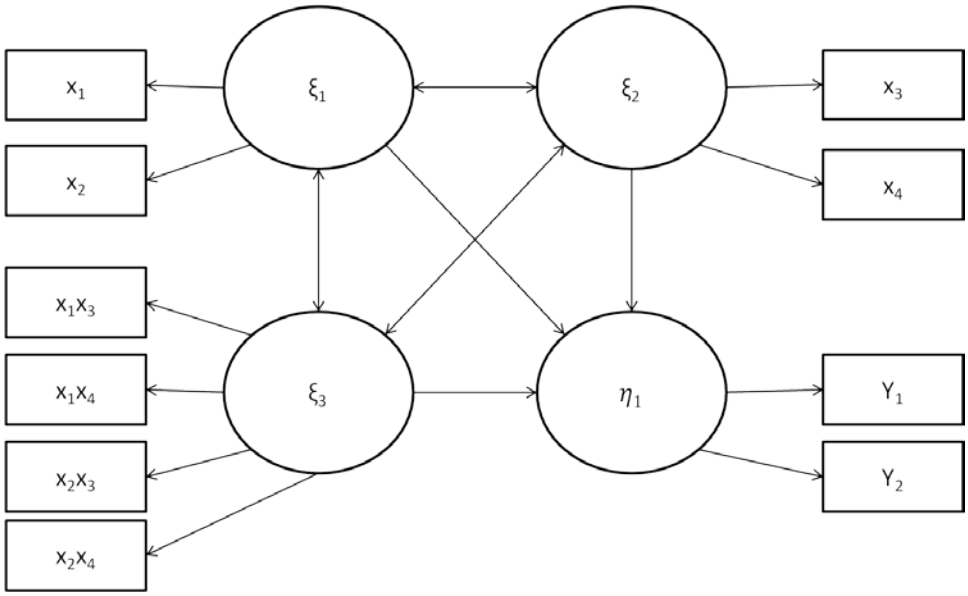

Figure 3. Generalized Kenny-Judd Product Indicator Path Diagram, errors and disturbances not shown.

2.4.2 Jöreskog –Yang's Single Indicator Method

Further, Jöreskog and Yang (1996) presented a model where observed variables are not mean centered, including intercepts, and only one product indicator is needed to identify the model. The Kenny and Judd model did not include an intercept term and Jöreskog & Yang (1996) have shown this to be inappropriate unless the covariance of $\xi_1$with ξ_2 is small or zero.

Jöreskog and Yang estimate their model from the covariance matrix created from the uncentered observed variables, the product indicators, and the observed variables' means. This method places nonlinear constraints upon the mean vector and covariance matrix. Jöreskog and Yang (1996) conclude that an Augmented Weighted Least Squares (WLSA) estimation procedure should be used. It does not rely on the assumption of normality, which is violated if an

interaction is present. WLS also does not require normality but does need the mean vector and covariance matrix to be independent, an assumption that is not met when product indicators are used (Jöreskog & Yang, 1996). The weight matrix suggested was the same as in Browne's Asymptotic Distribution Free estimation (Jonsson, 1998). Two drawbacks to this method are that it requires a large sample and is subject to sampling error.

Jöreskog and Yang (1996) and Jonsson's (1998) model does not require the observed variables to be mean centered and includes their intercepts in a vector, α, equation (10) becomes

$$\eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta.$$

Equations (11) now includes an intercept, for example

$$x_1 = \tau_1 + \lambda_1\xi_1 + \delta_1.$$

Equation (12) changes to

$$x_1x_3 = (\tau_1 + \lambda_1\xi_1 + \delta_1)(\tau_3 + \lambda_3\xi_2 + \delta_3).$$

Even if y, ξ_1, ξ_2, and ζ are all centered α will not be zero (Jonsson, 1998). The means of observed indicators are functions of other parameters and need to be jointly estimated with those parameters. The Jöreskog and Yang model has six assumptions. First, ξ_1 and ξ_2 are bivariate normal and centered. Second, ζ is distributed as $N(0, \psi)$. Third δ_i is distributed as $N(0, \theta_i)$ for all *x*. Fourth, δ_{ij} is independent of $\delta_{i'j'}$ for $ij \neq i'j'$. Fifth, δ_i is independent of ξ_i for all *x* and latent IVs. Sixth, ζ is independent of δ_i and ξ_i for all *x* and latent IVs. Jöreskog and Yang (1996) also point out that assumptions 4, 5, and 6 are untestable while 1, 2, and 3 can be assessed.

Jöreskog and Yang (1996) showed three consequences of using product indicators. First, the distribution of y is not normal even if ξ_1 and ξ_2 are, indicating that the joint distribution of $(y, x_1, \ldots, x_q)'$ is not multivariate normal even if the joint distribution of $(x_1, \ldots, x_q)'$ is. Second, the mean of y is a function of the x-intercepts, the coefficient of the latent interaction term, and

the covariance of the main effects; therefore, a mean vector needs to be estimated along with the covariance matrix. Lastly, if a researcher decides to use the Kenny and Judd method with all product indicators some nonlinear constraints are placed upon the mean vector and covariance matrix (Jöreskog & Yang, 1996).

In the Jöreskog and Yang model all cross-products do not have to be used to estimate the interaction, and in order to identify the model only one is needed. The Kenny and Judd model had five observed variables, y, x_1, x_2, x_3, x_4 and the first and third variable's factor loadings are fixed to one for identification purposes. The mean vector is

$$v' = (\alpha + \gamma_3\phi_{21}, \tau_1, \tau_2, \tau_3, \tau_4), \tag{13}$$

and the corresponding covariance matrix is

$$\Sigma(\hat{\theta}) = \begin{pmatrix} \sigma_{yy} & & & & \\ \gamma_1\phi_{11} + \gamma_2\phi_{21} & \phi_{11} + \theta_1 & & & \\ \gamma_1\lambda_2\phi_{11} + \gamma_2\lambda_2\phi_{21} & \lambda_2\phi_{11} & \lambda_2^2\phi_{11} + \theta_2 & & \\ \gamma_1\phi_{21} + \gamma_2\phi_{22} & \phi_{21} & \lambda_2\phi_{21} & \phi_{22} + \theta_3 & \\ \gamma_1\lambda_4\phi_{21} + \gamma_2\lambda_4\phi_{22} & \lambda_4\phi_{21} & \lambda_4\lambda_2\phi_{21} & \lambda_4\phi_{22} & \lambda_4^2\phi_{22} + \theta_4 \end{pmatrix}.$$

If the two main effects, ξ_1 and ξ_2 are bivariate normal then

$$\begin{aligned} VAR(\xi_1\xi_2) &= E(\xi_1^2\xi_2^2) - [E(\xi_1\xi_2)]^2 \\ &= \phi_{11}\phi_{22} + 2\phi_{21}^2 - \phi_{21}^2 \\ &= \phi_{11}\phi_{22} + \phi_{21}^2 \end{aligned}$$

and the variance of y can be expressed using $\phi_{11}\phi_{22} + \phi_{21}^2$

$$\sigma_{yy} = \gamma_1^2\phi_{11} + \gamma_2^2\phi_{22} + 2\gamma_1\gamma_2\phi_{21} + \gamma_3^2(\phi_{11}\phi_{22} + \phi_{21}^2) + \psi_{11}.$$

The model is identified as long as none of the other assumptions in SEM are violated. The product term of x_1x_3 is added to the model, with $\lambda_1 = \lambda_3 = 1$, and can be expressed as,

$$\begin{aligned} x_1x_3 &= (\tau_1 + \xi_1 + \delta_1)(\tau_3 + \xi_2 + \delta_3) \\ &= \tau_1\tau_3 + \tau_1\xi_2 + \tau_3\xi_1 + \xi_1\xi_2 + \delta_{x_1x_3}. \end{aligned}$$

The error term for the product indicator is

$$\delta_{x_1x_3} = \tau_1\delta_3 + \xi_1\delta_3 + \tau_3\delta_1 + \xi_2\delta_3 + \delta_1\delta_3 \, .$$

Including this product term adds seven more parameters that need to be estimated. The five covariances of the product indicators with the other variables are

$$COV(y, x_1x_3) = \tau_3\gamma_1\phi_{11} + \tau_3\gamma_2\phi_{21} + \tau_1\gamma_1\phi_{21} + \tau_1\gamma_2\phi_{22} + \gamma_3^2(\phi_{11}\phi_{22} + \phi_{21}^2), \quad (14)$$

$$COV(x_1, x_1x_3) = \tau_3\phi_{11} + \tau_1\phi_{21} + \tau_1\theta_1,$$

$$COV(x_2, x_1x_3) = \tau_3\lambda_2\phi_{11} + \tau_1\lambda_2\phi_{21},$$

$$COV(x_3, x_1x_3) = \tau_3\phi_{21} + \tau_1\phi_{22} + \tau_1\theta_3,$$

$$COV(x_4, x_1x_3) = \tau_3\lambda_4\phi_{21} + \tau_1\lambda_4\phi_{22}.$$

The product indicator's variance is

$$VAR(x_1x_3, x_1x_3) = \tau_3^2\phi_{11} + \tau_1^2\phi_{22} + \phi_{12}^2 + \phi_{11}\phi_{22} + \tau_1^2\theta_3 + \tau_3^2\theta_1 + \phi_{11}\theta_3 + \phi_{22}\theta_1 + \theta_1\theta_3.$$

Lastly, the product indicator's mean transforms v_1' from (13) into

$$v_1' = (\alpha + \gamma_3\phi_{21}, \tau_1, \tau_2, \tau_3, \tau_4, \tau_3 + \phi_{21}).$$

Seven new parameters were created by adding the product indicator. Only the covariance of the y and the product indicator, $COV(y, x_1x_3)$, contains new information about γ_3, while all the other parameters have already been identified. Equation (14) can be rearranged to solve for γ_3 in terms of the covariance of the y and product indicator. This allows γ_3 to be identified. The endogenous variable's variance, ψ_{11}, is estimated from the variance of the latent interaction, $Var(\xi_1\xi_2)$, and α from v_1.

Adding one product indicator to the model increases the number of elements that need to be estimated including a factor loading, an error variance, and a mean. All of these added elements can be estimated as functions of already existing parameters; therefore the number of parameters is the same. This is not an argument to use all product indicators to increase model

degrees of freedom. Increasing the number of product indicators creates more rank deficiency in the weight matrix. The degrees of freedom for the chi-square test uses the rank of the weight matrix minus the number of independent parameters. Therefore increasing the number of product indicators used does not create more DF for a model (Jöreskog & Yang, 1996; Jonsson, 1998).

Jöreskog and Yang (1996) examined their procedures performance by comparing SE and chi-square goodness-of-fit among the three estimation procedures ML, WLS and WLSA. WLS and ML had incorrect chi-square and SE values, and they concluded it was because of the violation of the assumptions of normality and independence. The WLSA method did provide the correct asymptotic SE and goodness-of-fit (Jonsson, 1998). The drawback of using WLSA is that it leads to more convergence problems and can provide multiple solutions. WLSA also requires larger sample sizes (Jöreskog & Yang, 1996).

2.4.3 Modifications to Product Indicator Method

A number of modifications to the Jöreskog and Yang model have been proposed, attempting to make latent interaction models easier to implement in practice. Algina and Moulder (2001) suggested centering the observed variables. While not different in its estimation, the centering alleviates convergence issues that Jöreskog and Yang (1996) model can have. A partially constrained approach developed by Wall and Amemiya (2001) removes the assumptions that ξ_1 and ξ_2 are normally distributed with a covariance of zero. This approach allows for all covariances between exogenous latent variables to be nonzero and different, but still requires nonlinear constraints. Algina and Moulder (2001) and Moulder and Algina (2002) compared a number of methods using simulation studies and found that their constrained approach with mean centering had less convergence problems, less bias, and more power.

An unconstrained approach was developed by Marsh et al. (2004), which removed the nonlinear constraints placed on the product indicators and the latent interaction. It requires at least two product indicator terms in order to be identified. Marsh et al. (2004) found that this unconstrained approach was comparable to or better than the partially constrained approach in goodness-of-fit, convergence, and bias in estimating the main and interaction effects for models that violate the assumption of normality. Another advantage of the unconstrained approach is that it is easy to implement, but it has less power when the assumption of normality is met and with smaller sample sizes (Marsh et al., 2004). A second simulation by Marsh et al. (2006) led them to conclude that ML estimation was best when used with either the partially constrained or unconstrained approach, concluding the parameter estimates are robust to the violation of normality.

2.4.4 Selection of Product indicators

Jöreskog and Yang provided a method where only one product indicator was necessary, but did not specify criteria for choosing items to use in creating it. All product indicator procedures require the selection of at least one item from each latent variable and Marsh et al.'s (2004) unconstrained method requires two product indicators. Jonsson (1998) found that using all product indicators led to less bias, but increased convergence issues and estimation problems. Jonsson's (1998) recommendation was that only one product indicator be used. Saris, Batista-Foguet, and Coenders (2007) found that the best indicators to use were those with the largest factor loadings. The product indicator created would then have the highest reliability, given that the indicators from the latent variables are congeneric (Saris, et al., 2007). Congeneric in this context means that all indicators for one latent variable are unidimensional and load on that

factor. One problem found was that the method was unreliable if there was a correlation between the error terms of the product indicators with the indicators of the endogenous variable (Saris, et al., 2007).

Marsh et al. (2004) investigated choosing optimal product indicators for the interaction effect concluding that all matched-pairs should be used. They agreed with Jonsson (1998) that all product indicators are not necessary. Each indicator is only used once and matched with an item from the other latent variable in the interaction. This satisfies two optimal properties. First, an item is not used in more than one product indicator so there is no redundancy. Second, information contained in the instrument is maximized because each item is used (Marsh et al., 2004). The problem with this recommendation is when measures have a different number of indicators. Marsh et al. (2004) suggested combining items, creating composite scores, to even the number of indicators for each latent variable. There is no consensus on which indicators to use if the items are not congeneric, if the number of indicators for the latent variables is uneven, or if the factor loadings are equal (Marsh et al., 2004).

2.5 MODERN METHODS FOR ESTIMATING A LATENT INTERACTION EFFECTS

2.5.1 Latent Moderated Structural Equation Models

A procedure for estimating latent interactions that does not rely on choosing product indicators or applying nonlinear constraints is Latent Moderated Structural Equations (LMS; Klein & Moosbrugger, 2000). It uses the Expectation-Maximization (EM) algorithm to derive the

distributions of the exogenous and endogenous variables, as a function of all model parameters including the interaction effect (Coenders, Batista-Foguet, & Saris, 2007). This method was proposed to avoid the problems of nonnormality caused by the nonlinear effect that afflict ML estimation.

The distribution of the indicators is estimated by a finite mixture distribution, which takes into account the nonnormality that interaction models have (Klein & Moosbrugger, 2000). Then the log-likelihood function is maximized using the EM algorithm (Dempster, Laird, & Rubin, 1977). EM is an iterative estimation procedure that converges on ML estimates for model parameters (Klein & Muthén, 2003). Initial values are specified for the model parameters and the EM algorithm is run producing new estimates for those parameters. These estimates are then used as the starting values for the next iteration. In the expectation step the posterior probabilities for parameter estimates are found, and in the maximization step the new parameters are chosen to maximize the sum of the weighted posterior probabilities. LMS is a full-information approach, and the nonnormal joint density function of the indicators is explicitly found. This enables standard errors (SE) to be calculated from the Fisher Information matrix (Schermelleh-Engel, Klein, & Moosbrugger, 1998).

Klein and Moosbrugger (2000) conducted a simulation study and concluded that LMS method has less bias than two-stage least squares, the Jöreskog and Yang method using WLSA estimation, and the Jöreskog and Yang method using ML estimation. Similar to ML, the estimates provided by LMS are asymptotically efficient, unbiased, and consistent (Moosbrugger, Schermelleh-Engel, Kelava, & Klein, 2009). LMS also provided accurate standard errors. Klein and Moosbrugger (2000) examined the likelihood ratio test for an interaction with LMS finding type I error rates twice the nominal alpha level, indicating that this method could erroneously

cause researchers to conclude there is an interaction effect. LMS assumes that the main effect latent variables and error terms of the indicators are normally distributed (Moosbrugger et al., 2009). This method cannot be used in complicated models due to the computational burden involved in modeling the joint distribution. LMS does have the advantage of not needing product indicators or non-linear constraints making it easier to implement in practice.

2.5.2 Quasi-Maximum Likelihood

A Quasi-Maximum Likelihood (QML) procedure was developed for complex models where LMS was not possible or difficult due to the amount of computation necessary (Klein & Muthén, 2003). It results in nearly identical parameter estimates as LMS, but is more practical (Klein & Muthén, 2007). QML takes into account the nonnormality in the data by modeling it. It is based on approximating the nonnormal distribution for the indicator vector, $f(x, y) = (y', x')'$, using a nonnormal density function, $f^*(x, y)$, where $f^*(x, y)$ is the product of a normal and a conditionally normal density. This method is based on the fact that the conditional distribution of $y_1|\mathrm{x} = x, \mathrm{u} = u$, can be approximated by a normal distribution even though the joint vector of indicators is not normally distributed. A transformation is applied making all elements of the transformed vector normally distributed but one, y_{1t},

$$(y_t, x_t)' \rightarrow (x_t, y_{1t}, u_t = \mathrm{R}y_t)$$

(Klein & Muthén, 2007).

For approximating the normal distribution the mean function

$$E(y_{1t}|x_t = x, u_t = u),$$

and variance function

$$VAR(y_{1t}|x_t = x, u_t = u),$$

are used. $\beta = (\lambda_{y21}, \dots, \lambda_{yp1})'$ is a $p\ -1$ vector of free parameters from Λ_y, and

$$\mathrm{R} = (-\beta, \mathrm{I}),$$

in this case I is $(p\ -1)$ by $(p\ -1)$. Let $u_t = \mathrm{R}y_t$, so $u_t = \mathrm{R}\varepsilon_t$. Then the Krickeberg and Ziezold (1988) transformation is used (as cited in Klein & Muthén, 2003; 2007). This transformation gives an approximation of $f(x, y)$ by the density function $f^*(x, y)$ for the indicator vector,

$$\begin{aligned} f(x, y) &= f_1(x, y_1, \mathrm{R}y) \\ &= f_2(x, \mathrm{R}y) f_3(y_1 | x_t = x, u_t = \mathrm{R}y) \\ &\approx f_2(x, \mathrm{R}y) f_3^*(y_1 | x_t = x, u_t = \mathrm{R}y) \\ &= f^*(x, y), \end{aligned}$$

where $f_1(x, y_1, \mathrm{R}y)$ is the density function of the transformed indicator vector. $f_2(x, u)$ is the normal density function of $(x_t, u_t)'$. $f_3^*(y_1 | x_t = x, u_t = \mathrm{R}y)$ is the conditional density of y_{1t} under the condition $x_t = x$ and $u_t = u$ and is univariate normal with a mean and variance equal to

$$E(y_{1t} | x_t = x, u_t = \mathrm{R}y),$$

and

$$VAR(y_{1t} | x_t = x, u_t = \mathrm{R}y).$$

$(y_1 | x_t = x, u_t = u)$ can be expressed as the sum of its mean function and residual variable, e_t:

$$(y_1 | x_t = x, u_t = u) \propto E(y_{1t} | x_t = x, u_t = u) + \ e_t(x, u).$$

The conditional expectation of y_{1t} given $x_t = x$ and $u_t = u$ is $y_{1t} | x_t = x, u_t = u$ (adapted from Klein & Muthén, 2003; 2007).

After this transformation a variance function model is used (see Carroll, Rupert, and Stefanski, 1995), with the mean and variance a function of the y conditional on x's (Klein &

Muthén, 2003; 2007). Quasi-likelihood estimation is then applied, and the joint density function of the indicator vector is approximated by the product of a conditionally normal and an unconditionally normal distribution (Klein & Muthén, 2007). The QML parameter estimates are found by maximizing the quasi-log-likelihood function derived from the approximated density function, $f^*(x, y)$ (Klein & Muthén, 2003). Compared to LMS, QML is less sensitive to the violation of normality, because it only has the assumption that the conditional distribution of the latent endogenous variable can be approximated as a normal distribution (Klein & Muthén, 2007). QML only assumes that the conditional distribution of the latent endogenous variable given the exogenous indicators can be approximated by normal distributions (Moosbrugger, et al. 2009).

Klein and Muthén (2003; 2007) compared QML to LMS using the Kenny and Judd model as well as another simple interaction model. Parameter estimates for both methods were equivalent when the assumption of normality was met and QML had less bias when the indicator variables were skewed. Standard errors were smaller for LMS when data were normal, which was expected since QML uses an approximation. QML's estimates of standard errors were more accurate when the indicators were skewed. Klein and Muthén (2007) examined the likelihood ratio test for the interaction to assess type I error rates and power. LMS had a type I error rate over twice the nominal alpha and QML's type I error rate was one percent higher than alpha. Power was similar between the two methods in the interaction conditions (Klein & Muthén, 2007). A model with three main effects and three interaction terms was used to test bias between the methods in a complex model. The results showed that LMS underestimated one main and one interaction effect, while QML had no sizeable bias (Klein & Muthén, 2007). Klein and Muthén

(2003, 2007) also found that QML was more efficient than the product indicator methods when factor loadings were low.

Marsh et al. (2004, 2006) examined the performance of QML against that of the constrained approaches, for a model similar to Kenny and Judd's but with three indicators for each latent variable. They examined three different indicator distributions, normal, uniform, and skewed. Three sample sizes were simulated and strength of the correlation between the main effects was varied. Klein and Muthén (2007) repeated the Marsh et al. studies using an updated version of Mplus, which uses an observed not an estimated Fisher Information matrix to find standard errors. Marsh et al (2004, 2006) and Klein and Muthén (2007) found that when the indicators were normally distributed QML had greater precision, smaller SE, less bias, and more power than the constrained approaches, while sample size and correlation strength had no effect. For normally distributed data, Marsh et al. (2004) found that QML had type I error rates that were twice alpha and the robust standard errors were underestimated. Klein and Muthén (2007) found no inflation of type I error when using the newer software, and did not report robust standard errors.

QML assumes that the latent main effects are normally distributed. This assumption is probably violated when the observed indicators are not normally distributed (Mooijaart & Bentler, 2010; Mooijaart & Satorra, 2009). Type I error rates and bias were worse for QML than for the constrained approaches in the presence of nonnormality (Klein & Muthén, 2007; Marsh et al., 2004). In these simulations, the indicators were simulated to be nonnormal over and above the nonnormality caused by the nonlinear effect. The bias that QML produces for nonnormal conditions was not consistent; for skewed data the bias was positive and for uniform data the bias was negative. QML severely underestimated a main effect when the indicators were skewed and

the nonlinear effect was large (Klein & Muthén, 2007). QML also had greater power than the product indicator methods (Klein & Muthén, 2007). This led Marsh et al. (2006) to conclude that when data are nonnormal QML may give more precise estimates that are incorrect. In a limited simulation study, Moosbrugger et al. (2009) concluded that LMS and QML are both better methods than the product indicator approaches, as they have greater power and lower type I error rates. The only effects manipulated in the Moosbrugger et al. (2009) simulation were strength of the interaction and size of the correlation between the interaction and the main effects.

2.5.3 Mooijaart -Bentler Method

An alternative method for estimating latent interaction effects without product indicators is to estimate higher order sample moments along with the covariance matrix and mean vector (Mooijaart & Bentler, 2010). Like LMS and QML, this procedure eliminates the burden of specifying nonlinear constraints and choosing product indicators. Unlike LMS and QML, it removes the need to estimate or approximate the joint density function of the indicators. The M-B method models the nonnormality in the data caused by the nonlinear effect using third order moments, which are estimates of skewness. Meijer and Mooijaart (1994) found that when observed indicators are not normally distributed, it is preferable to model more than just covariances, and recommended including the mean vector. Bentler (1983) concluded that it is preferable to model third order moments when nonnormality is present.

Mooijaart and Bentler (2010) fit third order moments of the observed indicators along with their means and covariances for estimating latent interaction effects. A third order moment, s_{ijk}, can be defined as a product of three deviation scores,

$$s_{ijk} = n^{-1} \sum_1^n (z_{it} - \bar{z}_i)\,(z_{jt} - \bar{z}_j)(z_{kt} - \bar{z}_k),$$

the deviations can be from the same or different indicator variables. Mooijaart and Bentler (2010) recommended choosing a third moment that reflects the nonnormality seen in the data. Like QML, latent main effect variables are assumed to be normally distributed. This assumption is required so the covariances and third order moments of the interaction can be found as products from deviations of the latent main effects (Mooijaart & Bentler, 2010). The latent main effects need to be centered, and the interaction effect is created as a product of the centered main effects.

The ML chi-square goodness-of-fit test has been shown to be insensitive to nonlinear effects, like those due to interaction terms (Mooijaart & Satorra, 2009). It is currently unknown if the chi-square tests are proper for ML estimation when nonnormality is present (Mooijaart & Bentler, 2010; Mooijaart & Satorra, 2009).The nonnormality of variables in the interaction term means there is no model that fits the data perfectly for an ML ratio test. Applied researchers may be incorrectly concluding that their hypothesized model fits the data, when a nonlinear effect or interaction term is needed (Mooijaart & Bentler, 2010). ML goodness-of-fit difference tests assume that there is a saturated model that fits the data perfectly. This calls into question the validity of the using a likelihood ratio test for an interaction effect, for methods that rely on ML like QML. Mooijaart and Bentler's procedure provides a goodness-of-fit test that is based on the residuals between the observed sample moments and those estimated from the model. Fitting higher order moments allows for there to be a true saturated model that can reproduce the nonnormality caused by the nonlinear effect.

A structural model with an interaction term could be alternatively expressed in partial matrix notation as,

$$\eta = \alpha + B_0\eta + \Gamma_1\xi + \Gamma_2(\xi \otimes \xi) + \zeta.$$

Here $\xi \otimes \xi$ is a vector of interaction effects, and Γ_2 is a vector of coefficients for the interaction effects. Γ_1 is the vector of coefficients for the main effects, and $z = (y', x')'$ is the joint vector of all indicators. In this format, $\xi \otimes \xi$ could represent interaction and quadratic effects. The model reduces to the form given in (8), $\eta = \alpha + B_0\eta + \Gamma_1\xi + \zeta$, if Γ_2 has all zero elements.

This method assumes that the covariances, and 3rd moments can be found as products of deviation scores of the indicators. z can be found as a function of model parameters, if this assumption is met (Mooijaart & Satorra, 2011). The means, or the first order vector of moments, of z is σ_1, the vector of non-redundant covariances, or second order moments of moments, of z is σ_2, and the vector of selected 3rd order moments of z is σ_3. σ_1, σ_2, and σ_3 can be expressed as a function of model parameters by the following,

$$\sigma_1 \equiv D_p^+ E(z),$$

$$\sigma_2 \equiv D_p^+ E[(z - \sigma_1) \otimes (z - \sigma_1)],$$

$$\sigma_3 \equiv T_p^+ E\left[\left((z - \sigma_1) \otimes (z - \sigma_1) \otimes (z - \sigma_1)\right)\right]$$

(Mooijaart and Bentler, 2010). D_p^+ is the Moore-Penrose generalized inverse of the duplication matrix, D_p , and T_p^+ is the Moore-Penrose generalized inverse of the triplication matrix, T_p. σ is a vector with the chosen means, covariances, and 3rd order moments of z and s is the vector of sample estimates of σ based on an independently and identically distributed sample of z. With $\sigma = \sigma(\theta)$, where $\sigma(\theta)$ is a continuously differentiable function of model parameters, θ, estimation can be made by minimizing the GLS fit function given in equation (9), $F(s, \sigma) = \left(s - \sigma(\theta)\right)' W (s - \sigma(\theta))$, except that s and $\sigma(\theta)$ are vectors containing third order moments as well as means and covariances. The weight matrix, W, contains the weights for the first, second, and third order moments. Mooijaart and Bentler (2010) and Mooijaart and Satorra (2011) suggest using GLS, where W is an identity matrix, to reduce convergence problems.

If the model is correctly specified, the test statistic, $T_{GLS} = (N-1)F(s,\sigma)$, is central chi-square distributed; if the model is not correctly specified the test statistic follows a noncentral chi-square distribution with noncentrality parameter λ_{GLS} (Mooijaart & Bentler, 2010). For a correctly specified model,

$$T_{GLS} = (N-1)\big(s - \sigma(\theta)\big)' \left[\hat{\Gamma}^{-1} - \hat{\Gamma}^{-1}\hat{\dot{\sigma}}\big(\hat{\dot{\sigma}}'\hat{\Gamma}^{-1}\hat{\dot{\sigma}}\big)^{-1}\hat{\dot{\sigma}}'\hat{\Gamma}^{-1}\right]\big(s - \sigma(\theta)\big).$$

Γ is the asymptotic covariance matrix of s and $\hat{\Gamma}$ being its consistent estimator. $\hat{\dot{\sigma}}$ is the Jacobian Matrix of $\sigma(\theta)$ with respect to $\hat{\theta}$ and T_{GLS} is asymptotically chi-square distributed when $\sigma = \sigma(\theta)$ (Bentler & Mooijaart, 2010; Mooijaart & Satorra, 2011). The expression of the noncentrality parameter for misspecified models is given in Bentler and Mooijaart (2010) or Mooijaart and Satorra (2011).

Mooijaart and Bentler (2010) hypothesized that fitting more than one third order moment could aid in the estimation process and increase power, but did not specify how many or in what circumstances. Mooijaart and Satorra (2011) presented a theorem for a moment test that does not require a model to be fitted. It uses the raw data and returns the most appropriate 3rd moment to choose. Mooijaart and Satorra (2011) also showed that including more 3rd moments lessened power and increased parameter bias, leading them to state, "We therefore recommend researchers to refrain from adding higher-order moments much beyond the strictly necessary for identification purposes" (Mooijaart & Satorra, 2011).

The degrees of freedom for this method are equal to the number of elements of s minus the number of estimated parameters. This is equal to the number of free parameters subtracted from the total number of means, covariances, and third order moments. For model comparisons a baseline model would leave free all of the moments. It should be noted that the saturated model

in this case is dependent upon the choice of the third-order moment or moments included in the analysis (Mooijaart & Satorra, 2011).

Mooijaart and Bentler (2010) developed a goodness-of-fit test for an interaction term. This test is based on the Lagrange multiplier (LM) principle. In SEM, an LM test examines all parameters, either individually or simultaneously, that have been fixed to zero and tests whether including these in the model will significantly improve fit (Kim & Bentler, 2006). The Mooijaart and Bentler (2010) method provides an accurate LM test. To show this a model with a known interaction effect was misspecified and fitted without the interaction term. ML estimation concluded the incorrect model fit the data, providing evidence to suggest that ML estimation was not sensitive to finding the interaction effect. Mooijaart and Bentler (2010) then used GLS with a third moment and estimated the model. The LM test results indicated that adding an interaction effect would significantly increase model fit.

Mooijaart and Bentler (2010) compared their method's performance to QML, in a simulation study using a latent growth model consisting of four time points. The slope was a function of initial status, a covariate, and the interaction between them. Mooijaart and Bentler (2010) found no differences in bias between the methods for any of the estimated parameters. Standard error estimates were equivalent for all parameters except for the interaction effect where QML had smaller standard errors. Mooijaart and Bentler (2010) concluded this was appropriate because it is based on large sample theory and indicators were normally distributed.

To test type I error rates and power of their method, Mooijaart and Bentler (2010) simulated data using the Kenny and Judd model with five different effect sizes. The results showed that type I error rates were close to alpha. The power of their method performed as expected, increasing as the interaction effect size increased. The power of this method was not

directly compared to QML's or the product indicator methods'. Mooijaart and Bentler (2010) then examined the impact of nonnormality of the indicators for their method and QML, using a CFA model with three factors and three interactions. The three factors had three indicators loading on them and each indicator also loaded onto one of the three interaction effects. The results parallel that of the growth model simulation, the two methods produce similar parameter estimates and standard errors except for the interaction, where QML's standard errors were smaller (Mooijaart & Bentler, 2010). If the nonnormality in the indicators is caused by the modeled nonlinear effect, they concluded that it does not adversely impact either method. Nonnormality over and above that caused by the nonlinear effect was not investigated.

There are a few disadvantages for using the Mooijaart and Bentler method. This method produces slightly larger SE for the interaction term than QML, meaning it could be less powerful. The latent main effect variables have to be centered and are assumed to be normally distributed. Lastly, sample sizes need to be large enough to model third order moments. Conversely, the advantages are that it is just an extension of typical SEM models and no specialized or technical knowledge is needed. It gives consistent estimates for parameters that are comparable to QML's but is computationally less demanding. M-B's procedure provides a correct goodness-of-fit test, including Satorra-Bentler corrected chi-square values. Their procedure provides an accurate LM test, allowing researchers to see if an interaction effect is needed. Along with QML, no product indicators are needed and nonlinear constraints do not have to be specified. Lastly, it has been shown to provide unbiased estimates if the nonnormality in the data is caused by the nonlinear effect (Mooijaart and Bentler, 2010).

2.6 GROUP COMPARISONS IN SEM

2.6.1 Multi-Sample Structural Equation Modeling

In MSEM, the model equations are modified to allow for parameters to differ among the subsamples. Allowing the parameters to differ enables researchers to test group differences in both the measurement and structural components of a model. The structural model specified in equation 8 becomes:

$$\eta^{(g)} = \alpha + B_0^{(g)}\eta^{(g)} + \Gamma_1^{(\mathrm{g})}\xi^{(g)} + \zeta^{(g)},$$

with the corresponding measurement equations (6 and 7) similarly altered,

$$x^{(g)} = \tau_{\mathrm{x}} + \Lambda_{\mathrm{x}}^{(\mathrm{g})}\xi^{(g)} + \delta^{(g)}, \text{ and}$$

$$y^{(g)} = \tau_{\mathrm{y}} + \Lambda_{\mathrm{y}}^{(\mathrm{g})}\eta^{(g)} + \varepsilon^{(g)}.$$

The g superscript indicates that parameters are group specific. Researchers can test for the invariance of parameters between groups by using chi-square tests of nested models. A baseline or freed model is first estimated without the parameter or parameters of interest constrained across groups, and the chi-square goodness-of-fit value and degrees of freedom are then compared to the values from the constrained model (Kline, 2005). The chi-square difference value is then examined for significance, with the degrees of freedom equal to the number of parameters that were constrained. Alternatively, a likelihood ratio test can be performed, where the difference between the log-likelihood for the constrained and unconstrained models is compared to evaluate fit using a chi-square test.

Measurement invariance can be tested for the endogenous variables, $\Lambda_{\mathrm{y}}^{(1)} = \Lambda_{\mathrm{y}}^{(2)} \ldots = \Lambda_{\mathrm{y}}^{(\mathrm{g})}$, the exogenous variables, $\Lambda_{\mathrm{x}}^{(1)} = \Lambda_{\mathrm{x}}^{(2)} \ldots = \Lambda_{\mathrm{x}}^{(\mathrm{g})}$ or for the intercepts. Factor means can be

tested for relative differences, but in order to identify the model one group's means have to be fixed to zero, making them the reference group (Sörbom, 1974). The means of the other groups are estimated and these estimates provide a relative mean difference in the latent factors from the reference group's (Kline, 2005). MSEM also allows for the testing of invariance of factor variances and covariance, $\Phi^{(1)} = \Phi^{(2)} ... = \Phi^{(g)}$, or for invariance of residual matrices among groups, $\Psi^{(1)} = \Psi^{(2)} ... = \Psi^{(g)}$, while the latter is not often done in practice (Chen, 2007).

Researchers are typically interested in the differences among the structural parameters in the models, $B_0^{(1)} = B_0^{(2)} ... = B_0^{(g)}$and $\Gamma_1^{(1)} = \Gamma_1^{(2)} ... = \Gamma_1^{(g)}$ (Chen, Sousa, & West, 2005, and Kim, 2009). Tests of invariance in structural parameters should be performed if the measurement model is shown to be invariant, or constrained to be equal (Chen et al., 2005). Rigdon et al. (1998) recommend the opposite approach for testing interactions using MSEM. A more restrictive model in which parameters are fixed to the same value across groups is tested against a model freeing the parameter that is hypothesized to be interacting with group membership. Then a Chi-square difference test is used and if significance is found it is indicative of moderation.

2.7 BOOSTRAPPING AND BOOSTRAPPING IN SEM

2.7.1 Naïve Bootstrapping

The bootstrap can be thought of as a method for evaluating parameter estimates (Stine, 1989) and a data-based approach for producing inferences (Efron & Tibshirani, 1993). Most commonly it is used to estimate quantities associated with a sampling distribution of an estimate or test statistic

(Boos, 2003). Conceptually, bootstrapping treats the observed data as a pseudo-population and repeatedly randomly samples with replacement from it. A random sample is defined as a selection of units of size *a*, selected at random with each unit's probability of being chosen equal to $\frac{1}{a}$. Sampling with replacement returns a unit to the population after being chosen, making the unit eligible to be selected again. The population from which the bootstrap samples (*B*) are taken come from the resample space (*R*), which is the original observed distribution (Efron & Tibshirani, 1993). The distribution of the test statistic from the bootstrap samples will form an empirical sampling distribution ($\hat{F}_{ED}$) (Rodgers, 1999). This empirical sampling distribution is an approximation of the statistic's true population distribution. When bootstrap samples are of the same size as the observed sample, n, then the standard deviation of $\hat{F}_{ED}$ is an estimate of the population parameter's standard error.

Bootstrapping has two main assumptions. The first pertains to generalizing the findings; for accurate inferences to be made from the analysis the original observed sample must be representative of the population of interest. Second, the statistical assumption of bootstrapping is that the relationship between a population and its sample can be modeled by the relationship between *R* and *B* (Yung & Bentler, 1996; Efron & Tibshirani,1993). Bickel and Friedman (1981) conclude that the success of all bootstrap methods depend upon the assumption that the sampling behavior of a statistic is the same when it is taken from the empirical distribution and when it is taken from the original population.

Outside of these assumptions there are a few limitations for the bootstrap. Using a naïve bootstrap with dependent data, such as those seen in autoregressive models is inappropriate. To alleviate this problem a semi-parametric or parametric bootstrap can be used if the lack of independence is modeled correctly. There are some statistics for which the typical bootstrapping

procedure fails; an example is when a variable is uniformly dichotomously distributed. It is possible to use a bootstrap method with $m < n$, where m is bootstrap sample size, taken without replacement which fixes this problem and can be thought of similar to a jackknife procedure (Bickel, Gotze, & van Zwet, 1997). Lastly, if the object of interest is a narrow feature of the original observed sample such as the largest (maximum) or smallest (minimum) order statistic then bootstrapping does not work (Bickel & Freedman, 1981).

2.7.2 Model Based Bootstrapping

In significance testing the conclusion reached, rejecting or failing to reject the null hypothesis, is based on finding a test statistic that is considered extreme or highly unlikely given that the null hypothesis is true. It is important to make sure the test statistic comes from a distribution where the null hypothesis is true. Naïve bootstrapping in SEM causes the bootstrap samples to be taken from a population in which the null hypothesis is not true (Bollen & Stine, 1992).

Yung and Bentler (1996) and Efron and Tibshirani (1993) give the steps for naïve bootstrapping in SEM. For example, if the normal theory ML discrepancy function for goodness-of-fit in an SEM model is used, the test statistic, $T_{ML} = (N-1)F_{GLS}$, is asymptotically distributed as a noncentral chi-square with DF equal to the number of free parameters minus the number of estimated parameters. First, resample from the original observed data set to find S^*, which is the observed covariance matrix for the bootstrap sample. Second, $\Sigma(\theta)$ and S^* are used, minimizing $F_{ml}\left(S^*, \Sigma(\theta^*)\right)$ for choices of θ^*. Third, compute the bootstrap test statistic T^*

$$T^* = (n-1)F_{ml}\left(S^*, \Sigma\left(\hat{\theta}^*\right)\right),$$

here $\hat{\theta}^*$ is the value of θ^* that minimizes $F_{ml}\left(S^*, \Sigma(\theta^*)\right)$. This process is repeated b times and the test statistic from the original sample, T_{ML}, is compared to the distribution of bootstrap test statistic values, $T_1^*, T_2^*, T_3^*, \ldots, T_b^*$ for significance (Yung & Bentler, 1996). When the asymptotic expected mean value and variances of T_{ML} and T^* are compared, the problem becomes apparent.

T_{ML} is distributed as a noncentral chi-square with

$$E(T_{ml}) = df + \delta_{np}\ , \tag{15}$$

$$VAR(T_{ml}) = 2df + 4\delta_{np}, \tag{16}$$

(Bollen & Stine, 1992). The noncentrality parameter is defined, $\delta_{np} = (N-1)F\left(\Sigma, \Sigma(\theta_0)\right)$ (Nevitt & Hancock, 2001). When the hypothesized model recreates the covariance structure of the data perfectly the noncentrality parameter is zero, $\delta_{np} = 0$, and T_{ml} is distributed as a central chi-square distribution with an expected mean value equal to the degrees of freedom. The expected value for the variance would be twice the degrees of freedom.

The distribution for T^* is approximately a noncentral chi-square distribution where the noncentrality parameter is T_{ML}. Analogous to T_{ML}, the expected mean value and variance of T^* are,

$$E^*(T^*) \approx df + T_{ML}, \tag{17}$$

$$VAR^*(T^*) \approx 2df + 4T_{ML}. \tag{18}$$

Here the noncentrality parameter for bootstrapping is T_{ML} (Bollen & Stine, 1992). If expectation of equations (17) and (18) are taken with respect to the parent distribution, the expected value for the mean of T_{ML} is simply inserted into equation (17) giving,

$$E(E^*(T^*)) \approx df + \left[df + \delta_{np}\right],$$

$$E(E^*(T^*)) \approx 2df + \delta_{np} \tag{19}$$

and for the variance, using equation (18)

$$E(VAR^*(T^*)) \approx 2df + 4[df + \delta_{np}]$$

$$E(VAR^*(T^*)) \approx 2df + 4df + 4\delta_{np}$$

$$E(VAR^*(T^*)) \approx 6df + 4\delta_{np}. \quad (20)$$

By comparing equations (15) and (19), it is seen that $E(T_{ml}) \neq E^*(T^*)$. The expected value for the bootstrap estimate is larger than the ML method's by a value equal to the degrees of freedom. This inequality also holds for comparing the variances, (16) with (20); the expected value for the variance of bootstrap test statistic is larger than its ML counterpart, $VAR(T_{ml}) < VAR(T^*)$ (Bollen & Stine, 1992). The null hypothesis fails when using naïve bootstrapping, regardless of whether it is true in the population or not, due to the difference in the expected values. This violates the assumption of bootstrapping; the empirical distribution created by resampling is different than the population distribution.

Bollen and Stine (1992) proposed a semi-parametric bootstrap method as a solution. The sample covariance matrix is transformed to have a covariance structure specified by the null hypothesis, and the bootstrap samples are taken from the transformed data. Applying the transformation enables the observed sample test statistic to be compared to an empirical sampling distribution taken from an R where the null hypothesis is true. Y is the centered observed data matrix, $Y = (Y_p - \bar{Y}_p)$, of size $n \times p$ and S is the sample covariance matrix found by $Y'Y/(n-1) = S$ and the model covariance matrix is noted as before, $\Sigma(\hat{\theta})$. Bollen and Stine (1992) suggest a transformation to S_T, where

$$S_T = YS^{-1/2}\Sigma(\hat{\theta})^{1/2}.$$

The covariance matrix of S_T is found by $S_T{}'S_T/(n-1)$, and is equal to the model covariance matrix $\Sigma(\hat{\theta})$,

$$S_T{}'S_T/(n-1) = \Sigma(\hat{\theta})^{\frac{1}{2}}S^{-1/2}Y'YS^{-1/2}\Sigma(\hat{\theta})^{\frac{1}{2}}/(n-1).$$

The interior of $Y'Y$ can be divided by the n-1 constant giving *S*,

$$= \Sigma(\hat{\theta})^{\frac{1}{2}}S^{-1/2}SS^{-1/2}\Sigma(\hat{\theta})^{\frac{1}{2}}.$$

The interior can be expanded with $S = S^{1/2}S^{1/2}$,

$$= \Sigma(\hat{\theta})^{\frac{1}{2}}S^{-1/2}S^{1/2}S^{1/2}S^{-1/2}\Sigma(\hat{\theta})^{\frac{1}{2}},$$

$$= \Sigma(\hat{\theta})^{\frac{1}{2}}\Sigma(\hat{\theta})^{\frac{1}{2}} = \Sigma(\hat{\theta}).$$

Bootstrapping is done using the transformed data matrix in lieu of the observed data matrix. This forces T_{ML} to be 0 since the transformed covariance matrix is equal to $\Sigma(\hat{\theta})$. The resampled data from the transformed matrix are used to form S_m^*, which is the covariance matrix for the bootstrap sample taken from the modified data. The ML fit equation is used to fit $\Sigma(\theta)$ to S_m^* minimizing $F_{ml}(S_m^*, \Sigma(\theta_m^*))$ for choices of θ_m^*. A bootstrap test statistic T_m^* is computed using, $T_m^* = (n-1)F_{ml}\left(S_m^*, \Sigma(\hat{\theta}_m^*)\right)$, and this is repeated *B* times (Bollen & Stine, 1992).

The mean and variance for T_m^* are $E^*(T_m^*) \approx df$ and $Var^*(T_m^*) \approx 2df$, respectively. By taking the expected values with respect to the parent population for the mean and variance the Bollen and Stine method gives $E(E^*(T_m^*)) \approx df$ and $E(Var^*(T_m^*)) \approx 2df$. After this transformation, the empirical sampling distribution of T_m^* behaves like the sampling distribution of T_{ML}, meeting the bootstrap assumption (Bollen & Stine, 1992). The empirical sampling distribution can be used as a reference distribution for giving T_{ML} a *p*-value (Finney & Distefano, 2006). The probability value is simply the proportion of T_m^* greater than T_{ML}. This model based bootstrapping method can also be used to test nested models in SEM (Bollen, & Stine, 1992).

Nevitt and Hancock (2001) examined the performance of model based bootstrapping by comparing it to T_{ML}, and the Satorra-Bentler scaled chi-square statistic (S-B). Both normal and nonnormal data were examined, as well as correctly and incorrectly specified models for a range of sample sizes. The T_{ML} performed well for correctly specified models when the assumption of normality was met. Regardless of sample size T_{ML} performed poorly for nonnormal data. The Satorra-Bentler correction controled type-I error rates for moderately nonnormal data well as long as sample size to parameter ratio was at least 10:1. The S-B performed well in the presence of severe nonnormality when sample sizes were large, a 20:1 ratio. The model based bootstrapping performed well regardless of normality and sample size conditions (Nevitt & Hancock, 2001).

Type I error rates follow a similar pattern as standard errors, with ML method having the largest estimates and bootstrapping having the least with S-B in between. For misspecified models ML had the best rejection rates but the authors surmise this was due to the fact the design was crossed with the nonnormality conditions; bootstrapping had the least and S-B in between (Nevitt & Hancock, 2001). Fouladi (2000) and Nevitt and Hancock (2001) agreed that the model based bootstrap method is conservative with respect to type I error, but less powerful. The S-B had more power but higher type I error rates (Finney & Distefano, 2006).

3.0 METHODS

3.1 DESIGN

3.1.1 Model

The model used in this simulation is similar to the Kenny-Judd one, except with three indicators for each latent variable (*see figure 4*), employed by Marsh et al. (2004) and comparable to Moosbrugger et al.'s (2009) model without the quadratic effects. The addition of the third indicator for each latent factor makes the model more realistic for applied users (Marsh et al., 2006).

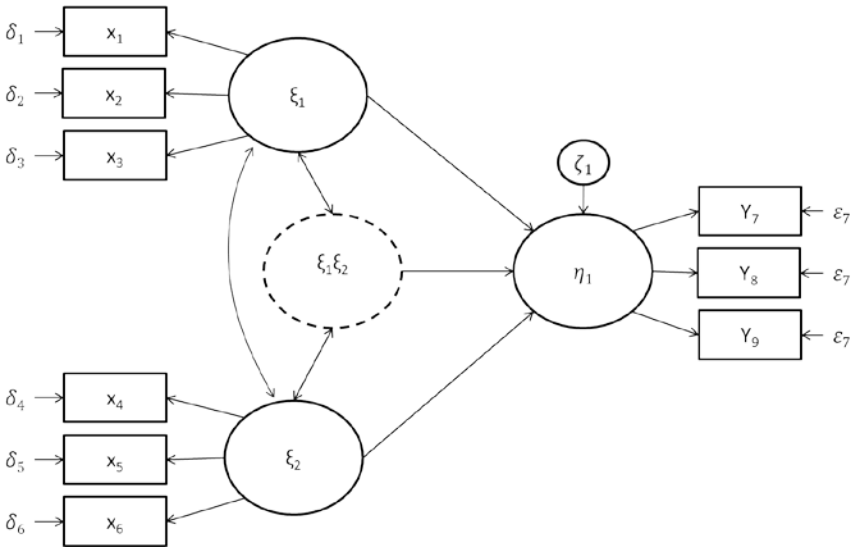

Figure 4. Path Diagram of Simulation Model

There are nine observed variables. Six indicators, $x_1, \ldots, x_6$, for the two latent exogenous variables, $\xi_1$and ξ_2. Three observed indicators, $y_7, \ldots, y_9$, for the latent endogenous variable, η_1. The observed variable covariance matrix contains 45 unique elements. The model contains 23 parameters to be estimated: six of the nine factor loadings, nine error variances, four factor variances, the covariance between main effects, and three structural regression coefficients. It should be noted that the model technically contains 24 parameters for the M-B method, including the 3rd moment.

3.1.2 Estimation Method

The two modern methods, M-B and QML, for estimating latent interaction effects were evaluated along with bootstrapping the M-B method in a multi-sample SEM. The two methods are contained in two different SEM software programs. QML is contained in Mplus and M-B in EQS, which also has the bootstrapping built into the software. For multi-sample models the QML analysis estimates the latent interaction effect through mixture modeling with known class membership. This is an alternative way of estimating multi-sample models in Mplus (Muthén & Muthén, 2010). The 3rd moment chosen in the M-B method was the combination that produced the largest magnitude (Pohlig & Kim, 2010; Mooijaart & Satorra, 2011). The Bollen-Stine bootstrap procedure was used to find the empirical sampling distribution of the interaction coefficient for each group separately using the M-B method. The difference between these distributions was tested for significance. Type I error rate and power for these methods were assessed across a number of conditions. Aside from convergence failure, cases in which extreme or unreasonable results were produced were excluded from analyses. This included interaction

effects larger in absolute value than two, which is over three times larger than the largest simulated interaction value.

3.1.3 Sample Size

Sample size was manipulated by changing the sample size in group two, n_2, in six different conditions. The parameter to sample size ratios chosen followed Kline's (2005) ideal ratio of 20:1, the realistic ratio of 10:1, and the minimum standard of 5:1. The sample size needed to be large enough to estimate the model for each group; therefore, the first group's sample size was fixed close to the realistic ratio of 10:1, $n_1 = 250$. The second group's sample size was manipulated using the ratios of $3{:}\,1, 2{:}\,1, 1{:}\,1, 1{:}\,2, 1{:}\,3,$ and $1{:}\,5$ resulting in $n_2 = (84, 125, 250, 500, 750, 1250)$. The total sample size ranges from 334 to 1500.

3.1.4 Interaction Effect Size

The three-way interaction effect was manipulated in this study by varying the two-way interaction between groups. The first group's interaction effect was fixed to a small effect, $\gamma_3^{(1)} = .1$. The second group's interaction was manipulated to create no difference, a small difference, a moderate difference, and a large difference, $\Delta\gamma_3 = (0, .1, .3, .5)$, which is associated with $R^2 = (.091, .095, .164, .314)$, respectively. The second group's interaction effect was $\gamma_3^{(2)} = (.1, .2, .4, .6)$ producing $R_\eta^2 = (.244, .274, .394, .594)$ for the latent DV, η_1. Jaccard and Wan (1995) and subsequently, Moulder and Algina (2002) and Marsh et al. (2004), used interaction effects equal to .2 and .4, arguing that these are typical of interaction effects seen in the literature. Klein and Moosbrugger (2000) used interaction sizes ranging from .1 to .7.

3.1.5 Distribution of Indicators

QML and M-B have been shown to be robust to the violation of normality, when the non-normality was caused by the non-linear effect (Mooijaart & Bentler, 2010). The performance of the M-B method has not been examined in the presence of non-normality over and above that caused by the interaction effect. QML had been previously shown to be more biased than product indicator approaches for nonnormal data (Marsh et al,, 2004; 2006). Yet Klein and Muthén (2007) found QML to be more robust and outperformed product indicator methods with updated software. Previous research that examined skewed distributions for indicators, used a $\chi^2_{df=6}$, with skewness ≈ 1.15 (Marsh et al., 2004) or directly manipulated skewness $= 1.50$ (Klein & Muthén, 2007). Skewness in this study will be ≈ 1.63 and simulated using a $\chi^2_{df=3}$.

Five different indicator distribution conditions were examined. For each condition, the three indicators for a latent variable were simulated from the same distribution (*see table 1*). The first condition had all the indicators, for both the latent IVs and DV, drawn from a normal distribution, $N(0,1)$. The second condition had skewed indicators for one latent IV using the $\chi^2_{df=3}$ distribution, while the other IV and DV's indicators were normal, $N(0,1)$. Third, both of the latent IVs indicators were $\chi^2_{df=3}$ distributed, and the latent DV was $N(0,1)$. Fourth, the IVs indicators were $N(0,1)$ distributed and the DV's indicators were skewed, $\chi^2_{df=3}$. Lastly, all the indicators for the latent variables were skewed using $\chi^2_{df=3}$.

Table 1. Indicator Distributions for the five conditions.

	Latent Variable		
Condition	ξ_1	ξ_2	η_1
1	$N(0,1)$	$N(0,1)$	$N(0,1)$
2	$\chi^2_{df=3}$	$N(0,1)$	$N(0,1)$
3	$\chi^2_{df=3}$	$\chi^2_{df=3}$	$N(0,1)$
4	$N(0,1)$	$N(0,1)$	$\chi^2_{df=3}$
5	$\chi^2_{df=3}$	$\chi^2_{df=3}$	$\chi^2_{df=3}$

3.1.6 Effects Held Constant

This simulation had a number of effects that were held constant across conditions (*see figure 5)*. The strengths of the correlations, main effects, and loadings chosen were based upon previous simulation research investigating interaction effects for latent variables (Marsh et al., 2004; 2006; Klein and Muthén, 2003; 2007; Moosbrugger et al., 2009; Mooijaart & Bentler, 2010). The correlations between the interaction and the exogenous main effects, Φ_{13}, and Φ_{23}, are constrained by the model,

$$\Phi_{13} = COV(\xi_1, \xi_1\xi_2)$$

$$= E\big[\big(\xi_1 - E(\xi_1)\big)\big(\xi_1 - E(\xi_1)\big)\big(\xi_2 - E(\xi_2)\big)\big] + VAR(\xi_1)E(\xi_2) + COV(\xi_1, \xi_2)E(\xi_1).$$

By centering ξ_1 and ξ_2, the expected value of both becomes zero, $E(\xi_1) = 0$ and $E(\xi_2) = 0$

$$= E\big[\big(\xi_1 - E(\xi_1)\big)\big(\xi_1 - E(\xi_1)\big)\big(\xi_2 - E(\xi_2)\big)\big] + VAR(\xi_1)(0) + COV(\xi_1, \xi_2)(0)$$

$$= E\big[\big(\xi_1 - E(\xi_1)\big)\big(\xi_1 - E(\xi_1)\big)\big(\xi_2 - E(\xi_2)\big)\big].$$

and $COV(\xi_1, \xi_1\xi_2)$ is reduced to only the first term, $E\left[\left(\xi_1 - E(\xi_1)\right)\left(\xi_1 - E(\xi_1)\right)\left(\xi_2 - E(\xi_2)\right)\right]$. If ξ_1 and ξ_2 are bivariate normally distributed then all 3rd moments are zero and the $COV(\xi_1, \xi_1\xi_2) = 0$ (see Bohnstedt & Goldberger 1969; Aiken & West, 1991).

The correlation between the exogenous main effects was fixed, $\Phi_{12} = .3$. Previous research has examined this correlation ranging from .2 to .4 (Jaccard, & Wan, 1995; Marsh et al., 2004) and 0, .3 and .7 (Moosbrugger et al., 2009). The main effects of the latent IVs were fixed, $\gamma_1 = \gamma_2 = .3$. Prior research has examined the values ranging from .2 to .4 (Jonsson, 1998; Schermelleh-Engle et al., 1998). Indicators of each latent variable had factor loading values randomly chosen from a uniform distribution, $\lambda \sim U(.5, .7)$, these values have been used in previous literature (Marsh et al., 2006). All indicators for a given latent variable were simulated following the same distribution. The first factor loading for each latent variable is fixed to one for identification. Two hundred fifty replications were used for each cell in the design and 250 bootstrap replications were used.

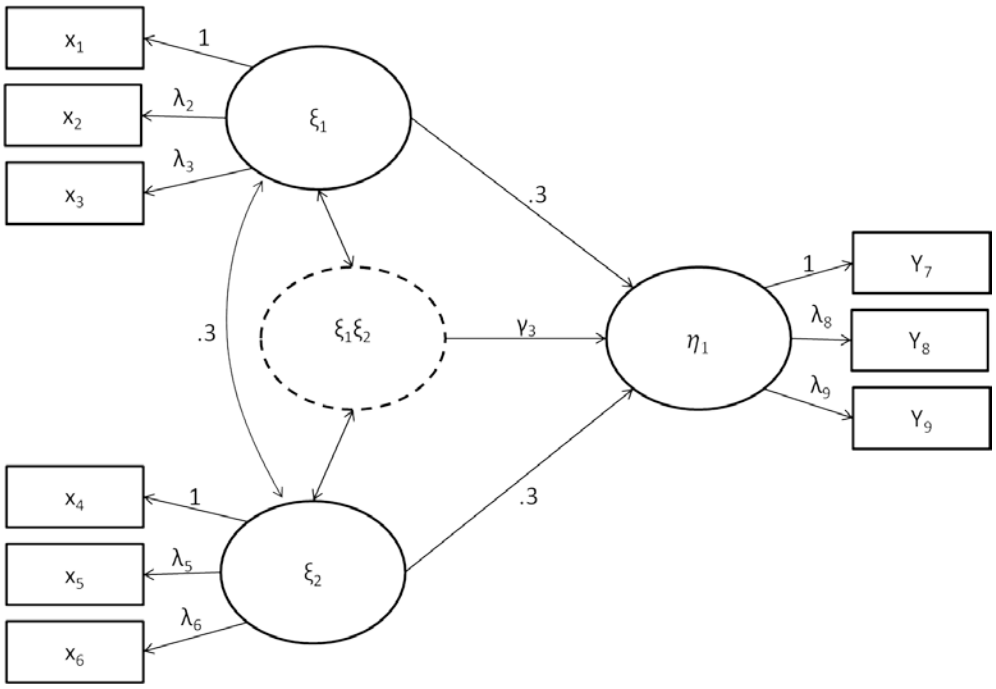

Figure 5. Constant effects indicated numerically in path diagram, errors and disturbances not shown.

3.2 PROCEDURES

3.2.1 Parameters

The simulation model's structural equation, $\eta = \Gamma_1\xi + \Gamma_2\xi_3 + \zeta$, consists of the components:

$$\Gamma_1 = \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix},$$

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix},$$

$$\Gamma_2 = \gamma_3,$$

$$\xi_3 = \xi_1 \times \xi_2 \,,$$

and

$$\zeta = \zeta_1.$$

Fully expressed the structural equation is

$$\eta_1 = \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + (\gamma_3)(\xi_3) + (\zeta_1)$$

$$= \begin{pmatrix} .3 & 0 \\ 0 & .3 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + (\gamma_3)(\xi_3) + (\zeta_1),$$

with γ_3 and ζ_1 being dependent upon the interaction effect condition.

Two correlation matrices are associated with the structural model, the correlation matrix for the exogenous latent variables is

$$\Phi = \begin{pmatrix} \phi_{11} & & \\ \phi_{12} & \phi_{22} & \\ \phi_{13} & \phi_{23} & \phi_{33} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & & \\ .3 & 1 & \\ \phi_{13} & \phi_{23} & \phi_{33} \end{pmatrix},$$

and with one endogenous latent variable $\Psi = (1)$, and the $VAR(\psi_{11}) = \left(1 - R_\eta^2\right)$. ϕ_{13}, ϕ_{23}, and ϕ_{33} are a function of model parameters dependent upon the simulation conditions.

The simulation model has two measurement equations. One for the exogenous latent variables, $x = \Lambda_x \xi + \delta$, which consists of the components:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix},$$

$$\Lambda_x = \begin{pmatrix} 1 & 0 \\ \lambda_2 & 0 \\ \lambda_3 & 0 \\ 0 & 1 \\ 0 & \lambda_5 \\ 0 & \lambda_6 \end{pmatrix},$$

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix},$$

and

$$\delta = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix}.$$

Fully expressed,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda_2 & 0 \\ \lambda_3 & 0 \\ 0 & 1 \\ 0 & \lambda_5 \\ 0 & \lambda_6 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix},$$

with λ_i's dependent upon the simulation conditions, and $\delta_i = 1 - \lambda_i^2$. The second measurement equation is for the exogenous variables, $y = \Lambda_y \eta + \varepsilon$, and consists of the components:

$$y = \begin{pmatrix} y_7 \\ y_8 \\ y_9 \end{pmatrix},$$

$$\Lambda_y = \begin{pmatrix} 1 \\ \lambda_8 \\ \lambda_9 \end{pmatrix},$$

$$\eta = (\eta_1),$$

and

$$\varepsilon = \begin{pmatrix} \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{pmatrix}.$$

Fully expressed,

$$\begin{pmatrix} y_7 \\ y_8 \\ y_9 \end{pmatrix} = \begin{pmatrix} 1 \\ \lambda_8 \\ \lambda_9 \end{pmatrix} \eta_1 + \begin{pmatrix} \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{pmatrix},$$

with ε_i's dependent upon the simulation conditions, and $\varepsilon_i = 1 - \lambda_i^2$.

The two covariance matrices associated with the measurement models are

$$\Theta_\delta = \begin{pmatrix} 1-\lambda_1^2 & & & & & \\ 0 & 1-\lambda_2^2 & & & & \\ 0 & 0 & 1-\lambda_3^2 & & & \\ 0 & 0 & 0 & 1-\lambda_4^2 & & \\ 0 & 0 & 0 & 0 & 1-\lambda_5^2 & \\ 0 & 0 & 0 & 0 & 0 & 1-\lambda_6^2 \end{pmatrix},$$

and

$$\Theta_\varepsilon = \begin{pmatrix} 1-\lambda_7^2 & & \\ 0 & 1-\lambda_8^2 & \\ 0 & 0 & 1-\lambda_9^2 \end{pmatrix}.$$

3.2.2 Data Generation & Simulation Outline

1. Data Simulation in R
    a. The two latent main effects were generated, correlated and centered.
    b. Their interaction term was computed.
    c. The disturbance generated and dependent variable was computed.
    d. The measurement error generated and indicators were computed.
2. The Estimation of models
    a. In EQS (M-B)
        a) Constrained & Unconstrained models estimated
        b) Their fit test, and all parameters were saved
        c) Bootstrapping
            1) Bootstrap run for each group separately & estimates for $\gamma_3$saved.
            2) Kolmogorov-Smirnov test run on γ_3 distributions (using R) and p-value saved.
    b. In Mplus (QML)
        a) Constrained & Unconstrained models estimated
        b) The likelihood ratio statistic, and all parameters from models were saved
3. Repeated for all replications across all conditions.

3.3 MEASURES

3.3.1 Model Difference Tests

There are a number of nested models that could be used in this analysis, each with an associated chi-square fit and log-likelihood statistic. The baseline model freely estimates every parameter, allowing them to differ between groups. One possible partial structural invariant model fixes the measurement parameters, the covariance between exogenous factors and the main effects to be equal but the interaction parameter is allowed to differ between groups, referred to here as the unconstrained model. The structural invariant, or constrained model, fixes all measurement and structural parameters to be equal.

Chi-square difference and log-likelihood tests were used to assess type I error rates and power for the M-B and QML methods. The unconstrained and constrained models were estimated and the difference in the fit was tested. A significant chi-square difference is indicative of a better fit by the model with fewer constraints (Kline, 2004). The reason the structural and partially structural invariant models were used in this analysis is because the data was simulated with measurement invariance and equal main effects. The structural component of interest was the interaction coefficient, and was allowed to vary between the groups in the unconstrained model. The difference in fit was recorded for the constrained and unconstrained models and a chi-square test was performed with the results dichotomized at the nominal alpha level, $\alpha = .05$.

3.3.2 Bootstrap Distribution Comparison

In order to test the efficacy of the bootstrapping method, the Bollen-Stine model based bootstrap procedure was performed on the interaction coefficient, γ_3, for each group. The empirical sampling distributions created were saved and the difference between them was tested using a two-sample Kolmogorov-Smirnov test (K-S) (Conover, 1999). The K-S tests if two samples were drawn from the same parent population by testing the maximum vertical distance between two distributions to see if a significant difference exists (Conover, 1999). A significant difference found between the two γ_3 bootstrap distributions would indicate that the method detected a difference in the interaction coefficient between the two groups. The *p*-value resulting from the test was saved and dichotomized as significant or not using the nominal alpha level, $\alpha = .05$. Comparing the bootstrap distributions for the parameter estimate removes any error that bootstrapping the chi-square difference test might introduce due to the non-normality present in the models (Mooijaart & Satorra, 2009).

3.3.3 Parameter Recovery

The parameter recovery for the estimates was evaluated for each group using Bias and Absolute Bias. Bias is calculated by subtracting the known population effect from the estimate produced, $Bias = (\hat{\theta}_i - \theta_i)$. The absolute bias provides an indication of the average error produced from a model after removing the direction of the error, $Absolute\ Bias = \sqrt{\frac{\sum_i^n (\hat{\theta}_i - \theta_i)^2}{n}}$.

Four absolute bias and nine bias measures were examined. Absolute bias for the interaction effects was analyzed separately, $\gamma_3^{(1)}$ and $\gamma_3^{(2)}$. The error for the main effects was created from the average absolute error resulting in estimating both main effects, γ_1 and γ_2. Lastly, error of the exogenous covariances was created from averaging the absolute bias in

estimating the factor variances and covariance, $\phi_{11}^{(1)}$, $\phi_{11}^{(2)}$, $\phi_{22}^{(1)}$, $\phi_{22}^{(2)}$and ϕ_{12}. Bias was examined for each of the nine parameter estimates individually, γ_1, γ_2, $\gamma_3^{(1)}$, $\gamma_3^{(2)}$, $\phi_{11}^{(1)}$, $\phi_{11}^{(2)}$, $\phi_{22}^{(1)}$, $\phi_{22}^{(2)}$and ϕ_{12}.

3.4 ANALYTIC PLAN

Both descriptive and inferential statistics were used to assess the methods' performance across all conditions. Descriptive statistics are reported detailing estimation failures and analyses that produce extreme estimates. A repeated measures Logistic Regression was performed on the dichotomized *p*-values for the model difference tests using a Generalized Estimating Equation (GEE). A GEE is an extension of the Generalized Linear Model for longitudinal or clustered data (Liang & Zeger, 1986). The advantage of using a GEE model is that it provides unbiased marginal regression coefficients regardless of the correlation structure of the errors (Ghisletta & Spini, 2004). The procedure allows for the specification of a working correlation matrix to account for the lack of independence in the observations (Ballinger, 2004). Only significant effects with an $R^2 \geq .01$ were interpreted. A separate analysis was performed for the type I error and power conditions, results were unchanged from analyzing all the $\Delta\gamma_3$ conditions simultaneously. For ease of interpretation the separate analyses are presented.

A series of $2 \times 6 \times 4 \times 5$ mixed design ANOVAs were performed to examine the IVs' impact on bias and absolute bias for the parameter estimates obtained from the unconstrained model. The within subject factor was method. The between subject factors were

sample size, difference in interaction effect size, and indicator distribution. Only effects with moderate effect sizes, $\eta_p^2 > .06$, were interpreted.

3.5 DATA VERIFICATION

Data for two conditions were generated and examined with exceedingly large sample sizes, $n_1 = n_2 = 10{,}000$, to verify that the data generation was functioning correctly. The first was a type I error condition, where the main effects were simulated to be equal, $\gamma_1 = \gamma_2 = .3$, and the interaction coefficient was the same for both groups, $\gamma_3^{(1)} = \gamma_3^{(2)} = .1$. The second data generation was a moderate interaction effect condition, the main effects for both groups remained equal to .3, group one's interaction coefficient was $\gamma_3^{(1)} = .1$ and group two's interaction coefficient was $\gamma_3^{(2)} = .4$. All errors simulated for both of these verifications were normally distributed, and factor loadings were randomly chosen from a uniform distribution, $\lambda \sim U(.5, .7)$.

The results of the verification analyses show the main effects were generated and estimated correctly for the type I error condition (*see table 2*). One main effect was underestimated in the moderate effect condition for both M-B and QML, $\gamma_2 = .223$ and $\gamma_2 = .231$, respectively. The interaction effects were over estimated in both conditions for the M-B method, $\gamma_3 = .175$, $\gamma_3^{(1)} = .135$, and $\gamma_3^{(2)} = .538$. Similarly, the interaction effects were over estimated in both conditions for QML, $\gamma_3 = .199$, $\gamma_3^{(1)} = .164$, and $\gamma_3^{(2)} = .524$.

Table 2. Data verification with random loadings.

Condition	Effect	M-B	QML
Type I Error	γ_1	0.291	0.304
	γ_2	0.302	0.286
	γ_3	0.175	0.199
Moderate Effect	γ_1	0.320	0.318
	γ_2	0.223	0.231
	$\gamma_3^{(1)}$	0.135	0.164
	$\gamma_3^{(2)}$	0.538	0.524

To make sure that these differences were not due to a problem with the generation process, data were then generated with less measurement error. All the factor loadings were simulated to be .9, and errors were normally distributed. The interaction estimates for the QML method were slightly high $\gamma_3 = .120$, $\gamma_3^{(1)} = .115$, and $\gamma_3^{(2)} = .432$, but the main effects in both conditions were correct (*see table 3*). For the M-B method all estimates were as expected. The results of the verification analyses indicated that the generation process was functioning correctly.

Table 3. Data verification with high factor loadings.

Condition	Effect	M-B	QML
Type I Error	γ_1	0.291	0.291
	γ_2	0.296	0.298
	γ_3	0.108	0.120
Moderate Effect	γ_1	0.303	0.303
	γ_2	0.303	0.301
	$\gamma_3^{(1)}$	0.109	0.115
	$\gamma_3^{(2)}$	0.403	0.432

4.0 RESULTS

4.1 CONVERGENCE RATES

QML had no convergence failures while the M-B method failed to converge 18% of the time. Sample size impacted M-B model convergence rates most, with smaller n_2 leading to more failures (*see table 4*). After removing convergence failures, both methods were screened for extreme estimates of γ_3, using box plots and histograms. This screening resulted in excluding models with $\gamma_3 > |2|$, which is over three times larger than the largest simulated interaction effect. The M-B method resulted in extreme estimates in 3.1% of the cases. The QML method only produced extreme estimates in .3% of the cases. Total exclusion rate after removing extreme estimates and convergence failures for the M-B method was 20.1% and for QML method was .3%.

Table 4. EQS convergence rates by n_2.

n_2	Rate
84	76%
125	78%
250	80%
500	85%
750	85%
1250	88%

4.2 MODEL DIFFERENCE TESTS

4.2.1 GEE Results for Type I error

A GEE was performed on the results from the model difference test from the two methods to investigate the impact of the IVs on type I error rate. QML ($M = .06, SE = .003$) had a significantly lower rejection rate than the M-B method ($M = .13, SE = .005$), Wald $\chi^2(1) = 139.11$, $p < .001, R^2 = .01$ (*see table 5*). The M-B method's rejection rate was more than twice the nominal alpha level, while the QML method's was approximately equal to the alpha level. The three way interaction was unable to be estimated in the GEE.

Table 5. GEE Effects table for Type I Error.

Effect	Df	Wald χ^2	p	R^2
n_2	5	3.55	.615	.000
Distribution	4	1.66	.165	.001
n_2*Distribution	5	14.16	.822	.001
Method	1	139.11	<.001	.010
Method*n_2	5	5.16	.396	.000
Method*Distribution	4	2.27	.687	.000
Method*n_2*Distribution		Not estimable		

4.2.2 GEE Results for Power

A GEE was performed on the results from the model difference tests from the two methods to investigate the impact of the IVs on power for detecting a different interaction coefficient, γ_3, between groups. The pattern of differences in rejection rate between methods was significantly different among the $\Delta\gamma_3$, Wald $\chi^2(2) = 856.01$, $p < .001, R^2 = .015$ (*see figure 6*). QML ($M = .44, SE = .004$) had a significantly higher rejection rate than the M-B method ($M =$

$.23, SE = .003$), Wald $\chi^2(1) = 1114.48,\ p < .001, R^2 = .067$ (*see table 6*). There was a difference among the $\Delta\gamma_3$ averaged across the other IVs, as $\gamma_3^{(2)}$ increased (and the $\Delta\gamma_3$ increased), rejection rates increased, Wald $\chi^2(2) = 3636.87,\ p < .001, R^2 = .188$. Rejection rates increased as n_2 increased averaged across the other IVs, Wald $\chi^2(5) = 723.92,\ p < .001, R^2 = .037$ (*see table 7*). The different error distributions did not explain any variability in rejection rate, Wald $\chi^2(4) = 15.31,\ p < .001, R^2 = .000$.

In order to find the pattern of differences between methods in the rejection rate among $\Delta\gamma_3$ the simple main effect for method was compared at each $\gamma_3^{(2)}$ condition (*see table 8)*. There was minimal difference between the two methods in error when $\Delta\gamma_3 = .1$. QML had greater power than the M-B method when there was a moderate difference, $\Delta\gamma_3 = .3$, and large difference, $\Delta\gamma_3 = .5$, in the interaction effect.

Table 6. GEE Effects table for Power.

Effect	Df	Wald χ^2	p	R^2
n_2	5	723.92	<.001	.037
$\Delta\gamma_3$	2	3636.87	<.001	.188
Distribution	4	15.31	.004	.000
n_2*$\Delta\gamma_3$	10	207.13	<.001	.005
n_2*Distribution	20	24.75	0.211	.000
$\Delta\gamma_3$*Distribution	8	15.57	0.049	.000
n_2*$\Delta\gamma_3$*Distribution	40	59.57	.024	.001
Method	1	1114.48	<.001	.067
Method*n_2	5	45.87	<.001	.001
Method*$\Delta\gamma_3$	2	856.01	<.001	.015
Method*Distribution	4	39.59	<.001	.000
Method*n_2*$\Delta\gamma_3$	10	37.23	<.001	.000
Method*$\Delta\gamma_3$*Distribution	8	13.44	0.098	.000
Method*n_2*Distribution	20	18.29	.569	.000
Method*n_2*$\Delta\gamma_3$*Distribution	40	35.97	.652	.000

Table 7. Rejection Rates by n_2.

n_2	84	125	250	500	750	1250
Rejection Rate	21%	24%	30%	36%	41%	45%

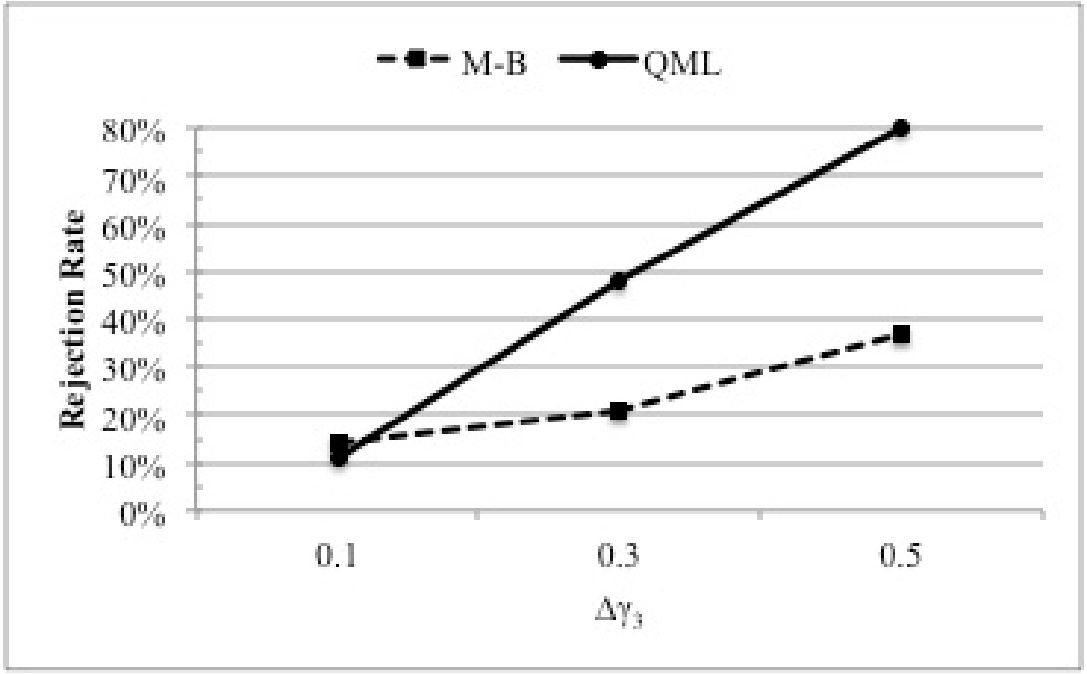

Table 8. Rejection Rates of Methods by $\Delta\gamma_3$.

$\Delta\gamma_3$	0.1	0.3	0.5
M-B	14%	21%	37%
QML	11%	48%	80%

4.3 PARAMETER RECOVERY

A series of $2 \times 6 \times 4 \times 5$ mixed design ANOVAs were performed to investigate the recovery of the parameter estimates, as measured by absolute bias and bias. The within subject factor was method (QML vs. MB). The between subject factors were group two's sample size, difference in the interaction effect size, and indicator distribution.

4.3.1 Absolute Bias

4.3.1.1 Absolute Bias of the Interaction Regression Coefficients

The impact of the IVs on the absolute bias resulting from estimating $\gamma_3^{(1)}$ and $\gamma_3^{(2)}$ was examined separately. QML ($M = .178, SE = .001$) had significantly less error in estimating $\gamma_3^{(1)}$ than the M-B method ($M = .294, SE = .002$), $F(1, 19020) = 2667.61, p < .001$, $\eta_p^2 = .123$ (*see table 9*).

Table 9. Absolute Bias for $\boldsymbol{\gamma_3^{(1)}}$ ANOVA table.

Effect	df	F	p	η_p^2
n_2	5	43.23	<.001	.011
$\Delta\gamma_3$	3	8.88	<.001	.001
Distribution	4	2.13	.075	.000
n_2*$\Delta\gamma_3$	15	1.93	.016	.002
n_2*Distribution	20	.91	.570	.001
$\Delta\gamma_3$* Distribution	12	.37	.974	.000
n_2*$\Delta\gamma_3$* Distribution	60	1.34	.040	.004
Method	1	2667.61	<.001	.123
Method*n_2	5	4.86	<.001	.001
Method*$\Delta\gamma_3$	3	8.80	<.001	.001
Method* Distribution	4	6.67	<.001	.001
Method*n_2*$\Delta\gamma_3$	15	1.64	.056	.001
Method*n_2* Distribution	20	.59	.925	.001
Method*$\Delta\gamma_3$* Distribution	12	.35	.979	.000
Method*n_2*$\Delta\gamma_3$* Distribution	60	.89	.721	.003

The pattern of differences on the absolute bias in estimating $\gamma_3^{(2)}$ between the two methods differed among sample size conditions averaged across the other IVs, $F(5, 19020) = 250.06, p < .001$, $\eta_p^2 = .062$ *(see figure 7*). QML had less error ($M = .258, SE = .001$) than the M-B method ($M = .338, SE = .001$), $F(1, 19020) = 1272.50, p < .001$, $\eta_p^2 = .063$ (*see table 10)*. As n_2 increased there was less error on the estimate of $\gamma_3^{(2)}$, $F(5, 19020) =$

$501.50, p < .001,\ \eta_p^2 = .116$. The absolute bias of $\gamma_3^{(2)}$ increased as $\gamma_3^{(2)}$ increased (and the $\Delta\gamma_3$ increased), $F(3, 19020) = 655.69, p < .001,\ \eta_p^2 = .094$ (*see table 11).*

In order to find the pattern of differences between methods on the error of $\gamma_3^{(2)}$ among sample sizes, the simple main effect of method was compared at each n_2 (*see table 12*). The difference on absolute bias of $\gamma_3^{(2)}$ between QML and M-B decreased as n_2 increased. For smaller sample sizes, $n_2 < 500$, QML had less error than the M-B method. There was minimal to no difference between the methods when the sample sizes were large, $n_2 \geq 500$.

Table 10. Absolute Bias for $\boldsymbol{\gamma_3^{(2)}}$ ANOVA table.

Effect	**df**	***F***	***p***	$\boldsymbol{\eta_p^2}$
n_2	5	39.50	<.001	.116
$\Delta\gamma_3$	3	51.64	<.001	.094
Distribution	4	2.08	.080	.000
n_2*$\Delta\gamma_3$	15	40.89	<.001	.031
n_2*Distribution	20	1.39	.116	.001
$\Delta\gamma_3$* Distribution	12	.96	.490	.001
n_2*$\Delta\gamma_3$* Distribution	60	1.81	.040	.006
Method	1	1272.50	<.001	.063
Method*n_2	5	250.06	<.001	.062
Method*$\Delta\gamma_3$	3	34.30	<.001	.005
Method* Distribution	4	12.81	<.001	.003
Method*n_2*$\Delta\gamma_3$	15	1.37	.154	.001
Method*n_2* Distribution	20	1.79	.017	.002
Method*$\Delta\gamma_3$* Distribution	12	2.84	.001	.002
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.70	..001	.005

Table 11. Absolute Bias for $\boldsymbol{\gamma_3^{(2)}}$ for $\boldsymbol{\Delta\gamma_3}$ conditions.

$\boldsymbol{\Delta\gamma_3}$	**0**	**.1**	**.3**	**.5**
RMSD	.225	.250	.329	.387

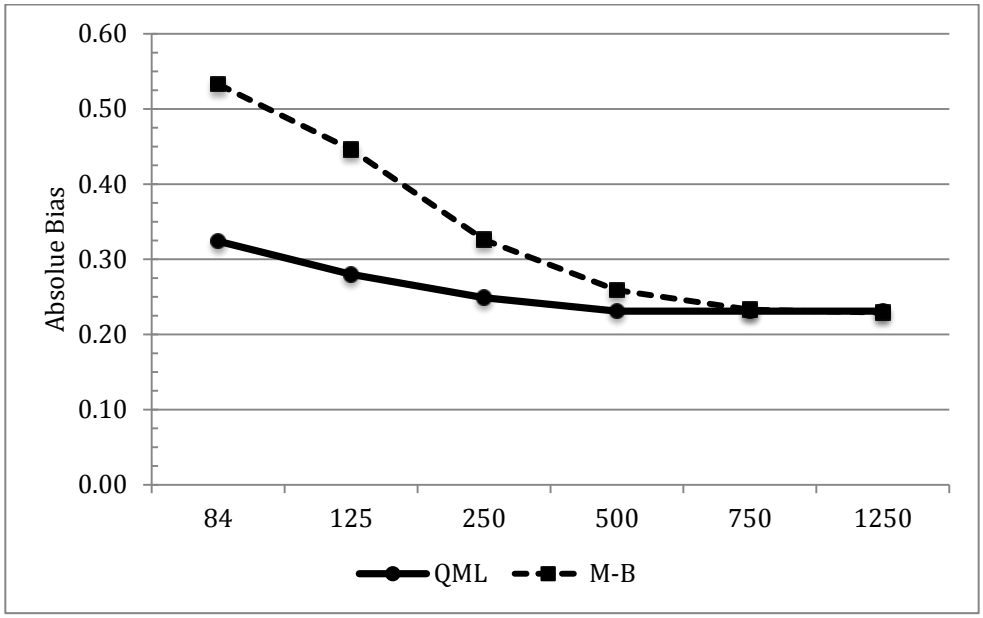

Figure 7. Absolute Bias for $\gamma_3^{(2)}$ for Method by n_2.

Table 12. Absolute Bias for $\boldsymbol{\gamma_3^{(2)}}$ for $\boldsymbol{n_2}$ by Method.

$\boldsymbol{n_2}$	**84**	**125**	**250**	**500**	**750**	**1250**
QML	.324	.280	.249	.231	.231	.231
M-B	.533	.446	.326	.259	.233	.229

4.3.1.2 Absolute Bias of Main Effects' Regression Coefficients

The pattern of differences on the average absolute bias of estimating $\gamma_1$and γ_2 between the two methods differed among sample sizes averaged across the other IVs, $F(5, 19020) = 313.68, p < .001,\ \eta_p^2 = .076$ *(see figure 8*). QML had less error ($M = .197, SE = .001$) than the M-B method ($M = .259, SE = .001$), $F(1, 19020) = 1628.42, p < .001,\ \eta_p^2 = .079$ (*see table 13).* As n_2 increased there was less error on the estimation of γ_1 and γ_2, $F(5, 19020) =$

$648.71, p < .001,\ \eta_p^2 = .146$. The absolute bias of $\gamma_1$and γ_2 increased as $\gamma_3^{(2)}$ increased (and the $\Delta\gamma_3$ increased), $F(3, 19020) = 592.60, p < .001,\ \eta_p^2 = .085$ (*see table 14).*

In order to find the pattern of differences between methods on the average error of $\gamma_1$and γ_2 among sample sizes, the simple main effect of method was compared at each n_2 (*see table 15*). The difference on absolute bias of γ_1 and γ_2 between QML and M-B decreased as n_2 increased. For smaller sample sizes, $n_2 < 500$, QML had less error than the M-B method. There was minimal to no difference between the methods when the sample sizes were large, $n_2 \geq 500$.

Table 13. Absolute Bias for $\boldsymbol{\gamma_1}$and $\boldsymbol{\gamma_2}$ ANOVA table.

Effect	df	F	p	η_p^2
n_2	5	648.71	<.001	.146
$\Delta\gamma_3$	3	592.60	<.001	.085
Distribution	4	1.44	.218	.000
n_2*$\Delta\gamma_3$	15	39.59	<.001	.030
n_2*Distribution	20	1.26	.196	.001
$\Delta\gamma_3$* Distribution	12	.89	..558	.001
n_2*$\Delta\gamma_3$* Distribution	60	1.74	<.001	.005
Method	1	1628.42	<.001	.079
Method*n_2	5	313.68	<.001	.076
Method*$\Delta\gamma_3$	3	31.93	<.001	.005
Method* Distribution	4	12.99	<.001	.003
Method*n_2*$\Delta\gamma_3$	15	1.52	.089	.001
Method*n_2* Distribution	20	1.71	.025	.002
Method*$\Delta\gamma_3$* Distribution	12	2.55	.002	.002
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.60	.002	.005

Table 14. Absolute Bias for $\boldsymbol{\gamma_1}$and $\boldsymbol{\gamma_2}$ for $\boldsymbol{\Delta\gamma_3}$ conditions.

$\boldsymbol{\Delta\gamma_3}$	**0**	**.1**	**.3**	**.5**
RMSD	.181	.197	.248	.288

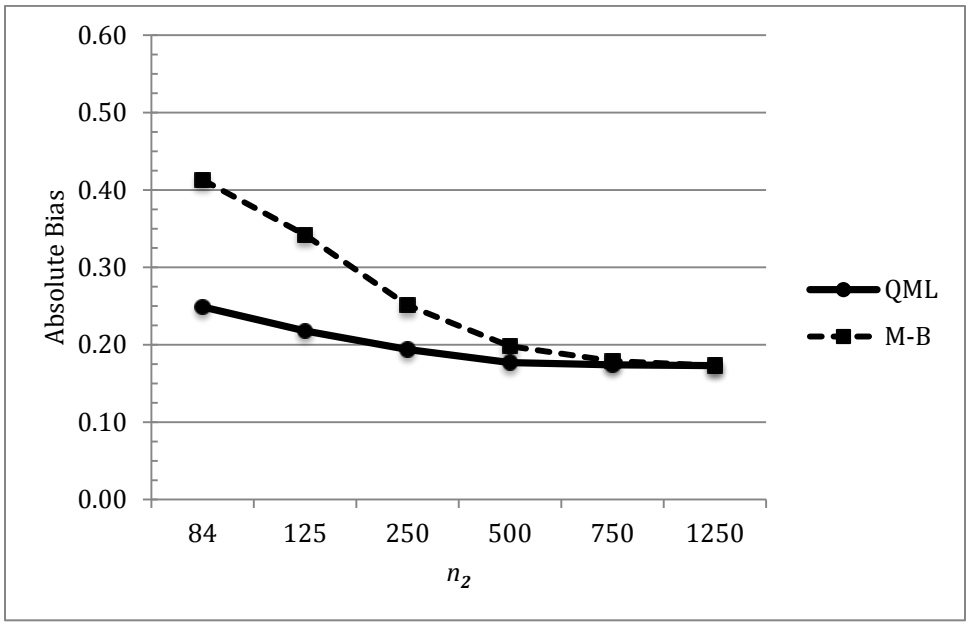

Figure 8. Absolute Bias for γ_1 and γ_2 for Method by n_2.

Table 15. RMSD for $\boldsymbol{\gamma_1}$and $\boldsymbol{\gamma_2}$ by Method by Sample Size.

$\boldsymbol{n_2}$	**84**	**125**	**250**	**500**	**750**	**1250**
QML	.249	.218	.194	.177	.174	.173
M-B	.413	.342	.251	.198	.179	.173

4.3.1.3 Absolute Bias of Variances and Covariance of the Main Effects

The average absolute bias of the exogenous covariances, $\phi_{11}^{(1)}$, $\phi_{11}^{(2)}$, $\phi_{22}^{(1)}$, $\phi_{22}^{(2)}$, and ϕ_{12}, was examined for the methods among the IVs. The pattern of differences on the absolute bias for the exogenous covariances between the two methods differed among sample sizes averaged across the other IVs, $F(5, 19020) = 321.62, p < .001$, $\eta_p^2 = .078$ (*see figure 9)*. QML had less error $(M = .085, SE = .001)$ than the M-B method $(M = .193, SE = .001)$, $F(1, 19020) = 8041.30$, $p < .001$, $\eta_p^2 = .279$ (*see table 16)*. As n_2 increased there was less error on the

estimation of the exogenous covariances averaged across the other IVs, $F(5, 19020) = 648.71, p < .001,\ \eta_p^2 = .146$.

In order to find the pattern of differences on error for the exogenous covariances between the methods among sample sizes, the simple main effect for method was evaluated at each n_2 (*see table 17*). The differences on the absolute bias of the exogenous covariance estimates between QML and M-B decreased as n_2 increased. The smaller the sample size the better the QML method performed when compared to the M-B method.

Table 16. ANOVA table for Exogenous Covariances.

Effect	**df**	F	p	η_p^2
n_2	5	621.06	<.001	.140
$\Delta\gamma_3$	3	1.24	.292	.000
Distribution	4	11.68	<.001	.002
n_2*$\Delta\gamma_3$	15	1.73	.038	.001
n_2*Distribution	20	1.57	.052	.002
$\Delta\gamma_3$* Distribution	12	.58	.864	.000
n_2*$\Delta\gamma_3$* Distribution	60	1.09	.287	.003
Method	1	8041.30	<.001	.279
Method*n_2	5	321.62	<.001	.078
Method*$\Delta\gamma_3$	3	1.98	.114	.000
Method* Distribution	4	14.01	<.001	.003
Method*n_2*$\Delta\gamma_3$	15	1.52	.087	.001
Method*n_2* Distribution	20	1.40	.111	.001
Method*$\Delta\gamma_3$* Distribution	12	.42	.957	.000
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.02	.439	.003

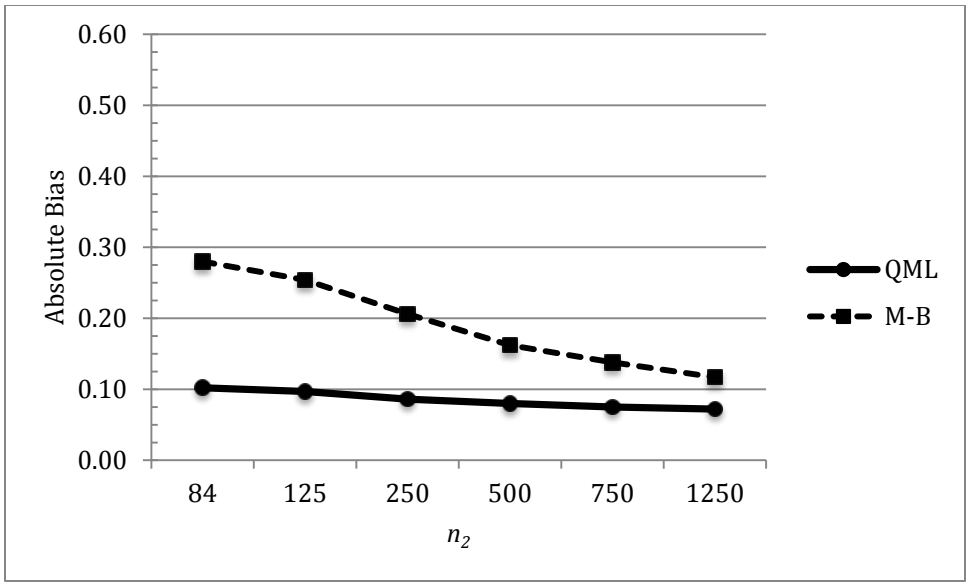

Figure 9. Absolute Bias for Exogenous Covariances for Method by n_2.

Table 17. Absolute Bias for Exogenous Covariances by Method by Sample Size.

n_2	84	125	250	500	750	1250
QML	.102	.097	.086	.080	.075	.072
M-B	.280	.254	.206	.162	.138	.117

4.3.2 Bias

4.3.2.1 Interaction Regression Coefficients

Biases of the latent interaction effects, $\gamma_3^{(1)}$ and $\gamma_3^{(2)}$, were examined separately for each group across the IVs. There were no meaningful differences on the bias for the first exogenous variable's regression coefficient, $\gamma_3^{(1)}$ (*see table 18*). The positive bias of $\gamma_3^{(2)}$ increased as the difference between $\gamma_3^{(1)}$ and $\gamma_3^{(2)}$ increased averaged across the other IVs (*see table 19*), $F(3, 19020) = 584.84, p < .001,\ \eta_p^2 = .084$ (*see table 20*).

Table 18. ANOVA table for Bias of $\boldsymbol{\gamma_3^{(1)}}$.

Effect	df	F	p	η_p^2
n_2	5	1.50	.187	.000
$\Delta\gamma_3$	3	5.49	.001	.001
Distribution	4	37.89	<.001	.008
n_2*$\Delta\gamma_3$	15	1.27	.209	.001
n_2*Distribution	20	1.83	.013	.002
$\Delta\gamma_3$* Distribution	12	.48	. 930	.000
n_2*$\Delta\gamma_3$* Distribution	60	.91	.668	.003
Method	1	160.92	<.001	.008
Method*n_2	5	2.49	<.029	.001
Method*$\Delta\gamma_3$	3	2.73	.042	.000
Method* Distribution	4	13.64	<.001	.031
Method*n_2*$\Delta\gamma_3$	15	1.59	.068	.001
Method*n_2* Distribution	20	2.33	.001	.002
Method*$\Delta\gamma_3$* Distribution	12	1.18	.287	.001
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.09	.298	.003

Table 19. ANOVA table for Bias for $\boldsymbol{\gamma_3^{(2)}}$.

Effect	df	F	p	η_p^2
n_2	5	144.05	<.001	.036
$\Delta\gamma_3$	3	584.84	<.001	.084
Distribution	4	30.48	<.001	.006
n_2*$\Delta\gamma_3$	15	22.44	<.001	.017
n_2*Distribution	20	3.29	<.001	.004
$\Delta\gamma_3$* Distribution	12	1.67	. 067	.001
n_2*$\Delta\gamma_3$* Distribution	60	1.18	.164	.004
Method	1	862.39	<.001	.043
Method*n_2	5	103.20	<.001	.026
Method*$\Delta\gamma_3$	3	56.34	<.001	.009
Method* Distribution	4	115.37	<.001	.024
Method*n_2*$\Delta\gamma_3$	15	7.58	<.001	.006
Method*n_2* Distribution	20	6.20	<.001	.006
Method*$\Delta\gamma_3$* Distribution	12	1.45	.134	.001
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.03	.419	.003

Table 20. Bias for $\boldsymbol{\gamma_2}$ across $\boldsymbol{\gamma_3^{(2)}}$ conditions.

$\Delta\boldsymbol{\gamma_3}$	**.0**	**.1**	**.3**	**.5**
Bias	.060	.096	.195	.271

4.3.2.2 Main Effects' Regression Coefficients

Bias of the main effects, γ_1 and γ_2, were examined separately across the IVs. There were no meaningful differences on the bias of γ_1 across the IVs (*see table 21*). Similarly, there were no meaningful differences on the bias of γ_2 across the IVs (*see table 22*).

Table 21. ANOVA table for Bias of $\boldsymbol{\gamma_1}$.

Effect	**df**	$\boldsymbol{F}$	$\boldsymbol{p}$	$\boldsymbol{\eta_p^2}$
n_2	5	6.24	<.001	.000
$\Delta\gamma_3$	3	.80	.495	.002
Distribution	4	2.08	.080	.000
n_2*$\Delta\gamma_3$	15	1.09	.364	.001
n_2*Distribution	20	1.41	.106	.001
$\Delta\gamma_3$* Distribution	12	1.87	. 033	.001
n_2*$\Delta\gamma_3$* Distribution	60	.77	.902	.002
Method	1	80.49	<.001	.004
Method*n_2	5	10.46	<.001	.003
Method*$\Delta\gamma_3$	3	.63	.596	.000
Method* Distribution	4	5.46	<.001	.002
Method*n_2*$\Delta\gamma_3$	15	1.54	.083	.001
Method*n_2* Distribution	20	1.35	.135	.001
Method*$\Delta\gamma_3$* Distribution	12	.85	.596	.001
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.04	.395	.003

Table 22. ANOVA table for Bias of $\boldsymbol{\gamma_2}$.

Effect	df	F	p	η_p^2
n_2	5	9.23	<.001	.002
$\Delta\gamma_3$	3	.16	.926	.000
Distribution	4	10.12	<.001	.002
n_2*$\Delta\gamma_3$	15	1.14	.316	.001
n_2*Distribution	20	1.77	.018	.002
$\Delta\gamma_3$* Distribution	12	1.69	.063	.001
n_2*$\Delta\gamma_3$* Distribution	60	1.31	.054	.004
Method	1	383.71	<.001	.020
Method*n_2	5	17.53	<.001	.005
Method*$\Delta\gamma_3$	3	1.52	.208	.000
Method* Distribution	4	8.99	<.001	.002
Method*n_2*$\Delta\gamma_3$	15	.96	.501	.001
Method*n_2* Distribution	20	1.76	.019	.002
Method*$\Delta\gamma_3$* Distribution	12	.97	.468	.001
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.27	.077	.004

4.3.2.3 Main Effects' Variances & Covariance of the Main Effects

Bias of the exogenous variables' variances for each group, $\phi_{11}^{(1)}$, $\phi_{11}^{(2)}$, $\phi_{22}^{(1)}$, and $\phi_{22}^{(2)}$, and the covariance between them, ϕ_{12}, was examined separately among the IVs. There were no meaningful differences on the bias of $\phi_{11}^{(1)}$, $\phi_{11}^{(2)}$,and ϕ_{12} across IVs (*see tables 23 to 25*).

Table 23. ANOVA table Bias for $\boldsymbol{\phi_{11}^{(1)}}$

Effect	df	F	p	η_p^2
n_2	5	31.94	<.001	.008
$\Delta\gamma_3$	3	1.40	.242	.000
Distribution	4	4.17	.002	.001
n_2*$\Delta\gamma_3$	15	1.86	.022	.001
n_2*Distribution	20	.58	.931	.001
$\Delta\gamma_3$* Distribution	12	1.08	.371	.001
n_2*$\Delta\gamma_3$* Distribution	60	1.33	.045	.004
Method	1	7.27	.007	.000
Method*n_2	5	115.19	<.001	.029
Method*$\Delta\gamma_3$	3	.06	.982	.000
Method* Distribution	4	.85	.493	.000
Method*n_2*$\Delta\gamma_3$	15	.67	.821	.001
Method*n_2* Distribution	20	1.35	.135	.001
Method*$\Delta\gamma_3$* Distribution	12	.80	.650	.001
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.31	.055	.004

Table 24. ANOVA table Bias for $\boldsymbol{\phi_{11}^{(2)}}$

Effect	df	F	p	η_p^2
n_2	5	19.29	<.001	.005
$\Delta\gamma_3$	3	1.32	.267	.000
Distribution	4	2.04	.086	.000
n_2*$\Delta\gamma_3$	15	1.41	.131	.001
n_2*Distribution	20	1.24	.208	.001
$\Delta\gamma_3$* Distribution	12	.95	.499	.001
n_2*$\Delta\gamma_3$* Distribution	60	1.61	.002	.005
Method	1	185.21	<.001	.010
Method*n_2	5	75.87	<.001	.020
Method*$\Delta\gamma_3$	3	5.72	.001	.001
Method* Distribution	4	2.01	.090	.000
Method*n_2*$\Delta\gamma_3$	15	.999	.452	.001
Method*n_2* Distribution	20	2.15	.002	.002
Method*$\Delta\gamma_3$* Distribution	12	1.43	.143	.001
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.45	.013	.005

Table 25. ANOVA table Bias for $\boldsymbol{\phi_{12}}$

Effect	df	F	p	η_p^2
n_2	5	8.37	<.001	.002
$\Delta\gamma_3$	3	.66	.578	.000
Distribution	4	3.48	.008	.001
n_2*$\Delta\gamma_3$	15	.80	.681	.001
n_2*Distribution	20	.99	.477	.001
$\Delta\gamma_3$* Distribution	12	.51	.910	.000
n_2*$\Delta\gamma_3$* Distribution	60	1.33	.046	.004
Method	1	36.84	<.001	.002
Method*n_2	5	17.36	<.001	.005
Method*$\Delta\gamma_3$	3	.44	.727	.000
Method* Distribution	4	8.56	<.001	.002
Method*n_2*$\Delta\gamma_3$	15	1.19	.274	.001
Method*n_2* Distribution	20	1.99	.005	.002
Method*$\Delta\gamma_3$* Distribution	12	.74	.718	.000
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.38	.027	.004

QML ($M = .002, SE = .001$) had significantly less bias on estimating $\phi_{22}^{(1)}$ than the M-B method ($M = .139, SE = .003$), $F(1, 19020) = 1557.94,\ p < .001,\ \eta_p^2 = .076$ (*see table 26*). Similarly, QML ($M = .003, SE = .001$) had significantly less bias in estimating $\phi_{22}^{(2)}$ than the M-B method ($M = .064, SE = .001$), $F(1, 19020) = 2416.83,\ p < .001,\ \eta_p^2 = .113$ (*see table 27*).

Table 26. ANOVA table for Bias of $\boldsymbol{\phi}_{22}^{(1)}$.

Effect	df	F	p	η_p^2
n_2	5	107.53	<.001	.027
$\Delta\gamma_3$	3	.12	.951	.000
Distribution	4	13.14	<.001	.003
n_2*$\Delta\gamma_3$	15	1.44	.117	.001
n_2*Distribution	20	1.57	.050	.002
$\Delta\gamma_3$* Distribution	12	.27	.993	.000
n_2*$\Delta\gamma_3$* Distribution	60	1.07	.335	.003
Method	1	1557.94	<.001	.076
Method*n_2	5	103.17	<.001	.026
Method*$\Delta\gamma_3$	3	.43	.729	.000
Method* Distribution	4	12.01	<.001	.003
Method*n_2*$\Delta\gamma_3$	15	1.29	.200	.001
Method*n_2* Distribution	20	1.62	.039	.002
Method*$\Delta\gamma_3$* Distribution	12	.45	.944	.000
Method*n_2*$\Delta\gamma_3$* Distribution	60	1.00	.475	.003

Table 27. ANOVA table Bias for $\boldsymbol{\phi}_{22}^{(2)}$.

Effect	df	F	p	η_p^2
n_2	5	70.29	<.001	.018
$\Delta\gamma_3$	3	1.72	.160	.000
Distribution	4	6.41	<.001	.001
n_2*$\Delta\gamma_3$	15	.94	.518	.001
n_2*Distribution	20	1.15	.289	.001
$\Delta\gamma_3$* Distribution	12	1.16	.310	.001
n_2*$\Delta\gamma_3$* Distribution	60	1.26	.082	.004
Method	1	2416.83	<.001	.113
Method*n_2	5	126.51	<.001	.032
Method*$\Delta\gamma_3$	3	4.90	.002	.001
Method* Distribution	4	13.02	<.001	.003
Method*n_2*$\Delta\gamma_3$	15	.812	.665	.001
Method*n_2* Distribution	20	2.36	.001	.002
Method*$\Delta\gamma_3$* Distribution	12	1.07	.385	.001
Method*n_2*$\Delta\gamma_3$* Distribution	60	.93	.627	.003

4.4 BOOTSTRAPPING

The bootstrap procedure outlined above was performed for the M-B method. The K-S test for the procedure proposed here proved to be too sensitive, so it was not included in the GEE analyses. The M-B method failed to converge in 18% of the cases; therefore the bootstrap estimates for these cases were excluded. The same decision rule for extreme estimates was applied; the case was removed if the mean of the empirical sample distribution for the interaction effect was estimated to be greater than the absolute value of two. This further eliminated 5.7% of the cases; the total exclusion rate after removing unrealistic estimates for bootstrapping the M-B method was 22.7%.

The rejection rates for the bootstrapping were high for every condition (*see table 28*). The difference in the interaction effect between the groups had no impact on rejection rate nor did the other two IVs, distribution of the error or sample size. Even in the type I error condition the rejection rate was 96% averaged across other IVs. Meaning this procedure had an extremely high false positive rate, incorrectly concluding that the two groups had different interaction coefficients when they did not.

Table 28. Rejection Rates of Bootstrapping by $\Delta\boldsymbol{\gamma_3}$.

$\Delta\boldsymbol{\gamma_3}$	**0**	**0.1**	**0.3**	**0.5**
Rejection Rate	96%	96	96	97

In order to check whether the model based bootstrapping was failing or if the K-S test was not functioning correctly, bias of the empirical sampling distributions was examined. Bias in this case was calculated for the mean and median of the empirical sampling distributions. If the model based bootstrapping produced large amounts of bias, it could be concluded that the bootstrapping was not functioning properly. If the bias produced is small then it could be

concluded that the K-S test was too sensitive. The magnitude of the bias seen for both the mean and median was smaller than the bias produced from each of the methods (*see table 29).* Across all IV conditions and for both groups the median of the empirical sampling distribution had bias less than the mean. Bias was impacted most by the $\Delta\gamma_3$. Similar to the results for each method, the larger the difference between the interaction effects the more positive bias seen. Sample size and error distribution conditions had no impact on the bias found from the model based bootstrapping.

Table 29. Mean Bootstrap Bias for $\boldsymbol{\gamma_3}$ by $\boldsymbol{\Delta\gamma_3}$ conditions.

$\boldsymbol{\Delta\gamma_3}$	**Mean Group 1**	**Mean Group 2**	**Median Group 1**	**Median Group 2**
0.0	0.01	0.04	0.01	0.02
0.1	0.01	0.04	0.00	0.02
0.3	0.01	0.07	0.01	0.05
0.5	0.02	0.13	0.01	0.11

5.0 DISCUSSION

5.1 SUMMARY OF FINDINGS

QML performed better than M-B across different conditions, especially with small sample size. The QML method had no convergence failures, and produced extreme estimates less than 1% of the time. The M-B method failed to converge in over 17% of the cases and produced extreme estimates in an additional 3%. Group two's sample size had the largest impact on convergence failure, with smaller sample sizes resulting in higher failure rates.

QML's rejection rate for the type-I error conditions was close to the nominal alpha level, while the M-B method's rejection rate was over twice the nominal level. QML had increasingly more power than the M-B method when the difference in the interaction effect between groups increased. Both methods, as expected, had an increase in power as sample size increased and when the difference between $\gamma_3^{(1)}$ and $\gamma_3^{(2)}$ increased. Rejection rates of the methods were not affected by the distribution of the indicator's error.

The QML method had less absolute bias than the M-B method in estimating the parameters, especially for smaller sample sizes. The M-B method had twice the error that QML had for group one's interaction coefficient, $\gamma_3^{(1)}$. For group two's interaction effect, $\gamma_3^{(2)}$, QML had less error when n_2 was small. For n_2 sizes equal to or greater than 500 the methods had about the same amount of error. Absolute bias for the main effects of the exogenous latent

variables, γ_1 and γ_2, differed between methods among sample sizes. For smaller sample sizes, QML had less absolute bias than the M-B method. Both methods had equal amount of error when n_2 was 500 or larger. QML had a less absolute error for the exogenous variances and covariance than the M-B method. QML was more accurate than the M-B method when sample size was small, the difference in error between the M-B method and QML decreased as n_2 increased. The error in estimating the exogenous covariances was smaller in magnitude than the error from the regression coefficients' estimates.

Other then group two's interaction coefficient, the biases of the estimates were largely unaffected by the IVs evaluated. The larger $\Delta\gamma_3$ the more positive bias that was seen in the estimate of $\gamma_3^{(2)}$. There were no differences in bias between methods or among the IVs in the estimates of $\gamma_3^{(1)}$, γ_1, γ_2, $\Phi_{11}^{(1)}$, $\Phi_{11}^{(2)}$, and Φ_{12}. QML had significantly less bias than the M-B method in estimating the second factor's variance in both groups, $\Phi_{22}^{(1)}$ and $\Phi_{22}^{(2)}$. Bias estimates were positive, indicating that the methods tended to overestimate parameters. The magnitude of bias for estimating the exogenous variances and covariance was smaller than the bias in estimating the regression coefficients.

The bootstrap method proposed here, testing the differences in empirical sampling distributions generated for each group using the K-S test did not function well. The rejection rates were high for all conditions, concluding that the two groups had different interaction effects even in the type-I error conditions. Although using bootstrapping for finding a difference in the interaction effect between groups did not work, bootstrapping the interaction parameter functioned well. The bias for the mean and median of the empirical sampling distribution was smaller than the bias produced from either the M-B or QML methods.

The findings in this study were consistent with previous research using these two methods. This study provides more evidence to support the notion that both methods are robust to nonnormality over and above that caused by an interaction effect (Marsh et al., 2004; Klein & Muthén, 2007). The distribution of the indicators' errors conditions, whether positively skewed or normally distributed, had no impact on rejection rate or parameter recovery. Moojiaart and Bentler (2010) and Klein and Muthén (2003; 2007) found an increase in bias for the interaction coefficient as it increased in magnitude. Similar to these previous findings, this study found an increase in bias and absolute bias in estimating the interaction coefficient when the strength of the interaction effect increased. Klein and Muthén (2007) found that the bias of the main effects increased as the interaction increased, the research reported here found an increase in the absolute bias of the main effects, γ_1 and γ_2, as the interaction effect increased. It is unknown if the error seen in estimating the exogenous covariances is consistent with previous research, as it has not been previously reported.

While the patterns of bias and absolute bias are consistent with previous findings, the present research found a larger magnitude in bias. Reliability of the latent variables was simulated to be between .25 and .49. Klein and Muthén (2007) and Moojiaart and Bentler (2010) used similar levels of reliability, and found smaller magnitudes of bias. It should be noted that their dependent variables were single indicators of an endogenous latent variable measured without error, and were single sample analyses. Jaccard and Wan (1993) found bias of similar magnitude for the product indicator methods when reliability was less than .7. Marsh et al (2004) had an endogenous latent variable with error and found less bias than seen here but reliability was simulated to be greater than .7 and was a single group analysis.

It is hypothesized that the poorer performance of the M-B method compared to the QML could be due to the chosen sample sizes being too small. This may have caused higher model convergence failure and more error in estimation of the parameters. This study used Kline's (2005) standard SEM recommendation of sample size to parameter ratio of 10:1. Bentler (1983) pointed out that estimating higher order moments in small samples can be unstable, and recommended sample sizes larger than traditionally used when estimating them. Gillard and Iles (2005) state, "Moreover sample sizes needed to identify third order moments with a practically useful degree of precision are somewhat larger than is the case for first and second order moments." The research here found when n_2 was less than 500, the M-B method exhibited more error in the estimates of the parameters and had higher non-convergence rates. This suggests that when using the M-B method, the sample size to parameter ratio should be 20:1 or greater. In the article proposing the method, Moojiaart & Bentler (2010) used a parameter to sample size ratio greater than 20:1, and found less bias and error then the research reported here.

5.2 LIMITATIONS

The conditions evaluated in this research were chosen to try and replicate ones found in real data. Yet there are a number of limitations, ranging from the IVs selected to the factors that were fixed. First, the model chosen was based upon previous research, and is a simple structural model involving only three latent variables and one interaction effect. The simulation model was used by Marsh et al. (2004 and 2006). Originally based on the Kenny and Judd model (1984), except that a latent endogenous variable was used instead of an observed DV and all the latent variables had three indicators. This structural model consisted of two main effects and one interaction

effect that were all positive. It is plausible that using a more complex model with more than one interaction effect could cause more estimation problems, and an increase error.

This research did not consider any heterogeneity in effects (e.g., combination of negative and positive effects). The impact of having a negative interaction was not investigated, nor was the scenario where the main effects were in opposite directions. This research only examined differences between groups in the strength of one interaction effect in the same direction. The potential situation where an interaction is present in only one group or where interactions are in the opposite directions was not investigated. Quadratic effects were not examined here, only an interaction between two exogenous latent variables. An interaction between an exogenous and endogenous or two endogenous variables was not investigated.

Secondly, the IVs chosen were those that had been previously examined and impacted the performances of the two methods sample size, interaction effect size, and nonnormality. Total sample size was not directly manipulated but only through changing the ratio of group sizes. The first group had a sample size to parameter ratio of 10:1, and this did not vary. Therefore the impact of both groups having small sample sizes or large ones simultaneously was not investigated. For examining robustness to nonnormality of the indicators, the non-normal distribution chosen was limited to one, having positive skew. All indicators for a latent factor were simulated from the same distribution. Other types of non-normal distributions, or situations where indicators for a given latent variable come from different distributions were not investigated.

Lastly, three factors that could potentially affect the results of the methods were held constant: measurement error, correlation between the exogenous latent variables, and strength of the exogenous latent variables' regression coefficients. The amount of measurement error in the

latent variables was not systematically manipulated; factor loadings were fixed between .5 and .7. The covariance between the two main effects was fixed. Finally, varying the strength of the main effects was not investigated.

5.3 CONCLUSION AND RECOMMENDTIONS

It is the conclusion of the author that the QML procedure, contained in the Mplus software, is the preferred method for testing if a two way interaction comprised of two continuous latent variables differs between groups. QML outperformed the M-B method across conditions, especially for the smaller sample sizes. QML had no model convergence problems and did not produced unrealistic estimates often. It better controlled for false positives having a type-I error rate close to the nominal level. The M-B method had a type-I error rate that was twice alpha. QML had more power than the M-B method when the interaction effect differed between groups by larger amounts.

QML exhibited less error in estimating the parameters than the M-B method. QML had less absolute bias for the interaction effect in group one. For smaller sample sizes, QML had less error in estimating the interaction effect in group two, the exogenous regression coefficients, and the exogenous covariances. Based on the pattern of results found in this research it is recommended that the sample size to parameter ratio be greater than 20:1 when using the M-B method.

Care should be taken in interpreting the parameter estimates resulting from these analyses. The resulting interaction estimates may be biased. Both methods overestimated the interaction coefficient as the difference in the interaction effect between the groups increased and

the interaction effect increased in magnitude. Both methods also had an increase in error in estimating the main effect's regression coefficients as the difference in the interaction effects increased. It is the author's opinion that QML can accurately determine if the interaction effect does differ between the two groups, but may not be accurate in estimating the exact numerical difference.

It is not recommended to use the bootstrap procedure outlined here because the K-S test was too sensitive. Bootstrapping to obtain more accurate parameters is a viable option. While QML is the recommended procedure, Mplus does not currently allow for an automated bootstrapping procedure when estimating a latent interaction. This software limitation is not specified in the user's manual. Mplus also necessitates the use of a mixture modeling framework with known group membership when generalizing the QML to MSEM.

5.4 FUTURE DIRECTIONS

A number of future directions have be illuminated from this research. First and most importantly, a practical application of these methods should be performed. This would show the utility in generalizing the methods for examining differences in a continuous latent variable interaction in a multi-sample analysis. An application of the methods would also prove useful for applied researchers in acting as a guide on how to conduct the analysis and interpret the results. Secondly, research should consider larger sample sizes for more stable and accurate estimates when using the M-B method and modeling a third moment.

Both Klein and Muthén (2007) and Moojiaart and Bentler (2010) used a variety of models to demonstrate their methods. Specifically, Klein and Muthén (2007) show the

effectiveness of their method when a number of latent interactions are present in a model. Moojiaart and Bentler (2010) present an example of using the 3rd moment method in latent growth modeling. Research should be performed to show that these methods can be generalized for comparing group differences in interaction effects in both of those situations. Future research should consider examining differences in an interaction effect in more than two groups.

Simulation research should further investigate a number of factors. The influence of measurement error on these methods in a MSEM model should be systematically examined. From the analysis for the accuracy of the data generation process, it was seen that an increase in measurement error adversely impacted both methods' estimates. Simulation research should examine conditions where larger interaction effects are not confounded with larger differences between the groups. Other types of error term distributions should be examined to test if these methods are robust to non-normality, over and above that caused by the interaction effect, beyond just positively skewed distributions. Having indicators for one latent variable that come from different distributions should also be examined. Lastly, simulation research should examine the effect of having categorical or Likert-type indicators on these methods in a MSEM.

Future research should consider other types of SEM models and real world data situations. Most of the research examining latent interactions is based on the Kenny and Judd model, or one similar. Various combinations in the direction of the effects whether positive or negative, for the interaction or main effects, should be examined to see the impact on bias and error. Specifically, the cases where an interaction exists in one group but not another and when the interaction effects are in opposite directions. Different kinds of latent interactions, such as between two endogenous variables or between an exogenous and endogenous variable should be considered. The impact of missing data should be investigated for both of these methods,

especially when comparing group differences in an interaction effect. These methods could be further generalized to Mixture and Multi-Level SEM models. The three-way interaction tested was among two continuous latent variables and one categorical variable through MSEM, future research should investigate a 3-way interaction effect of three latent variables.

Bootstrapping SEM models with an interaction effect should be further investigated to see if it provides utility in estimating the interaction and other parameters in an SEM model. The biases seen by bootstrapping the M-B method were smaller than the biases seen from either the QML or M-B estimate for the interaction effect. For bootstrapping in MSEM models, in lieu of testing the distributional differences between groups with a K-S test, a test of medians or another type of test could prove more effective in controlling type-I error.

APPENDIX A

DESCRIPTIVE SATISTICS FOR INTERACTION EFFECTS BY INDEPENDENT VARIABLES

A.1 ALL LATENT VARIABLES NORMALLY DISTRIBUTED

Table 30. Parameter estimates for $\boldsymbol{\gamma_3^{(1)}}$ and $\boldsymbol{\gamma_3^{(2)}}$ by $\boldsymbol{n_2}$ and $\boldsymbol{\Delta\gamma_3}$ for first distribution condition.

		M-B		QML	
$\boldsymbol{n_2}$	$\boldsymbol{\Delta\gamma_3}$	$\boldsymbol{\gamma_3^{(1)}}$	$\boldsymbol{\gamma_3^{(2)}}$	$\boldsymbol{\gamma_3^{(1)}}$	$\boldsymbol{\gamma_3^{(2)}}$
84	0.0	0.11 (0.39)	0.15 (0.69)	0.18 (0.19)	0.23 (0.37)
84	0.1	0.14 (0.40)	0.17 (0.62)	0.17 (0.23)	0.33 (0.39)
84	0.3	0.10 (0.38)	0.35 (0.71)	0.13 (0.18)	0.57 (0.37)
84	0.5	0.11 (0.56)	0.47 (0.75)	0.19 (0.25)	0.86 (0.35)
125	0.0	0.12 (0.32)	0.15 (0.53)	0.16 (0.19)	0.20 (0.30)
125	0.1	0.13 (0.38)	0.14 (0.59)	0.21 (0.22)	0.28 (0.32)
125	0.3	0.16 (0.40)	0.48 (0.63)	0.17 (0.22)	0.60 (0.32)
125	0.5	0.20 (0.52)	0.68 (0.52)	0.19 (0.24)	0.88 (0.31)
250	0.0	0.11 (0.44)	0.16 (0.38)	0.17 (0.22)	0.20 (0.21)
250	0.1	0.07 (0.37)	0.26 (0.41)	0.17 (0.19)	0.32 (0.20)
250	0.3	0.14 (0.45)	0.60 (0.42)	0.16 (0.22)	0.66 (0.26)
250	0.5	0.15 (0.35)	0.84 (0.39)	0.19 (0.22)	0.91 (0.25)
500	0.0	0.19 (0.38)	0.19 (0.22)	0.14 (0.21)	0.19 (0.14)
500	0.1	0.17 (0.40)	0.30 (0.23)	0.19 (0.20)	0.32 (0.15)
500	0.3	0.18 (0.39)	0.61 (0.35)	0.20 (0.23)	0.66 (0.20)
500	0.5	0.22 (0.37)	0.89 (0.32)	0.20 (0.22)	0.94 (0.23)
750	0.0	0.17 (0.36)	0.16 (0.20)	0.19 (0.20)	0.17 (0.12)
750	0.1	0.16 (0.38)	0.33 (0.20)	0.15 (0.16)	0.33 (0.12)
750	0.3	0.21 (0.38)	0.63 (0.25)	0.21 (0.23)	0.66 (0.17)
750	0.5	0.12 (0.34)	0.93 (0.28)	0.20 (0.23)	0.94 (0.19)
1250	0.0	0.14 (0.36)	0.16 (0.12)	0.15 (0.19)	0.17 (0.09)
1250	0.1	0.13 (0.43)	0.33 (0.15)	0.16 (0.20)	0.34 (0.10)
1250	0.3	0.20 (0.35)	0.68 (0.20)	0.20 (0.19)	0.68 (0.15)
1250	0.5	0.14 (0.37)	0.99 (0.23)	0.20 (0.26)	0.98 (0.18)

A.2 ONE EXOGENOUS VARIABLE POSITIVELY SKEWED

Table 31. Parameter estimates for $\boldsymbol{\gamma_3^{(1)}}$ and $\boldsymbol{\gamma_3^{(2)}}$ by $\boldsymbol{n_2}$ and $\boldsymbol{\Delta\gamma_3}$ for second distribution condition.

		M-B		QML	
$\boldsymbol{n_2}$	$\boldsymbol{\Delta\gamma_3}$	$\boldsymbol{\gamma_3^{(1)}}$	$\boldsymbol{\gamma_3^{(2)}}$	$\boldsymbol{\gamma_3^{(1)}}$	$\boldsymbol{\gamma_3^{(2)}}$
84	0.0	0.11 (0.30)	0.17 (0.65)	0.15 (0.20)	0.15 (0.34)
84	0.1	0.20 (0.41)	0.23 (0.71)	0.17 (0.21)	0.30 (0.39)
84	0.3	0.20 (0.40)	0.36 (0.70)	0.19 (0.20)	0.62 (0.37)
84	0.5	0.19 (0.43)	0.62 (0.66)	0.20 (0.21)	0.88 (0.36)
125	0.0	0.15 (0.45)	0.25 (0.60)	0.16 (0.19)	0.18 (0.30)
125	0.1	0.13 (0.38)	0.28 (0.63)	0.16 (0.22)	0.33 (0.29)
125	0.3	0.14 (0.38)	0.43 (0.66)	0.13 (0.22)	0.66 (0.31)
125	0.5	0.15 (0.39)	0.73 (0.58)	0.20 (0.23)	0.87 (0.31)
250	0.0	0.15 (0.37)	0.19 (0.41)	0.14 (0.20)	0.15 (0.20)
250	0.1	0.17 (0.36)	0.29 (0.32)	0.15 (0.22)	0.32 (0.20)
250	0.3	0.21 (0.41)	0.54 (0.46)	0.17 (0.21)	0.62 (0.22)
250	0.5	0.15 (0.47)	0.89 (0.43)	0.17 (0.25)	0.94 (0.26)
500	0.0	0.22 (0.36)	0.15 (0.24)	0.15 (0.21)	0.13 (0.14)
500	0.1	0.20 (0.40)	0.33 (0.26)	0.17 (0.19)	0.35 (0.15)
500	0.3	0.22 (0.37)	0.65 (0.32)	0.17 (0.22)	0.66 (0.20)
500	0.5	0.17 (0.37)	0.93 (0.33)	0.18 (0.24)	0.96 (0.22)
750	0.0	0.21 (0.35)	0.17 (0.15)	0.16 (0.23)	0.16 (0.12)
750	0.1	0.14 (0.35)	0.33 (0.18)	0.17 (0.18)	0.32 (0.13)
750	0.3	0.20 (0.38)	0.65 (0.25)	0.15 (0.20)	0.67 (0.16)
750	0.5	0.23 (0.37)	0.96 (0.30)	0.17 (0.24)	0.98 (0.22)
1250	0.0	0.21 (0.38)	0.17 (0.14)	0.15 (0.18)	0.16 (0.08)
1250	0.1	0.15 (0.39)	0.34 (0.15)	0.17 (0.20)	0.34 (0.10)
1250	0.3	0.21 (0.32)	0.64 (0.20)	0.19 (0.22)	0.65 (0.15)
1250	0.5	0.21 (0.35)	0.97 (0.27)	0.21 (0.27)	1.00 (0.23)

A.3 BOTH EXOGENOUS VARIABLES POSITIVELY SKEWED

Table 32. Parameter estimates for $\boldsymbol{\gamma_3^{(1)}}$ and $\boldsymbol{\gamma_3^{(2)}}$ by $\boldsymbol{n_2}$ and $\boldsymbol{\Delta\gamma_3}$ for third distribution condition.

		M-B		QML	
$\boldsymbol{n_2}$	$\boldsymbol{\Delta\gamma_3}$	$\boldsymbol{\gamma_3^{(1)}}$	$\boldsymbol{\gamma_3^{(2)}}$	$\boldsymbol{\gamma_3^{(1)}}$	$\boldsymbol{\gamma_3^{(2)}}$
84	0.0	0.22 (0.47)	0.18 (0.77)	0.16 (0.19)	0.17 (0.34)
84	0.1	0.23 (0.85)	0.32 (0.63)	0.14 (0.22)	0.36 (0.48)
84	0.3	0.22 (0.62)	0.46 (0.67)	0.14 (0.22)	0.57 (0.42)
84	0.5	0.18 (0.36)	0.66 (0.61)	0.15 (0.18)	0.81 (0.31)
125	0.0	0.17 (0.39)	0.23 (0.52)	0.15 (0.19)	0.12 (0.32)
125	0.1	0.19 (0.47)	0.32 (0.60)	0.16 (0.20)	0.32 (0.30)
125	0.3	0.22 (0.38)	0.55 (0.51)	0.13 (0.22)	0.64 (0.30)
125	0.5	0.20 (0.37)	0.74 (0.68)	0.13 (0.21)	0.82 (0.28)
250	0.0	0.20 (0.37)	0.22 (0.35)	0.15 (0.20)	0.15 (0.22)
250	0.1	0.23 (0.43)	0.34 (0.40)	0.15 (0.23)	0.32 (0.21)
250	0.3	0.24 (0.42)	0.70 (0.43)	0.17 (0.22)	0.68 (0.24)
250	0.5	0.28 (0.40)	0.94 (0.47)	0.15 (0.22)	0.94 (0.26)
500	0.0	0.27 (0.40)	0.19 (0.23)	0.17 (0.20)	0.16 (0.14)
500	0.1	0.24 (0.43)	0.38 (0.29)	0.14 (0.20)	0.34 (0.16)
500	0.3	0.21 (0.39)	0.66 (0.29)	0.17 (0.20)	0.65 (0.19)
500	0.5	0.26 (0.42)	0.95 (0.33)	0.18 (0.24)	0.95 (0.25)
750	0.0	0.26 (0.38)	0.18 (0.18)	0.17 (0.19)	0.17 (0.13)
750	0.1	0.22 (0.34)	0.34 (0.18)	0.15 (0.21)	0.31 (0.12)
750	0.3	0.23 (0.33)	0.65 (0.23)	0.16 (0.20)	0.68 (0.17)
750	0.5	0.23 (0.33)	0.96 (0.30)	0.16 (0.26)	0.98 (0.19)
1250	0.0	0.19 (0.36)	0.19 (0.15)	0.13 (0.20)	0.16 (0.10)
1250	0.1	0.22 (0.36)	0.35 (0.14)	0.19 (0.24)	0.32 (0.11)
1250	0.3	0.19 (0.41)	0.68 (0.21)	0.18 (0.26)	0.70 (0.16)
1250	0.5	0.22 (0.40)	0.97 (0.22)	0.21 (0.30)	0.97 (0.16)

A.4 ENDOGENOUS VARIABLE POSITIVELY SKEWED

Table 33. Parameter estimates for $\boldsymbol{\gamma_3^{(1)}}$ and $\boldsymbol{\gamma_3^{(2)}}$ by $\boldsymbol{n_2}$ and $\boldsymbol{\Delta\gamma_3}$ for fourth distribution condition.

		M-B		QML	
$\boldsymbol{n_2}$	$\boldsymbol{\Delta\gamma_3}$	$\boldsymbol{\gamma_3^{(1)}}$	$\boldsymbol{\gamma_3^{(2)}}$	$\boldsymbol{\gamma_3^{(1)}}$	$\boldsymbol{\gamma_3^{(2)}}$
84	0	-0.01 (0.37)	-0.04 (0.77)	0.20 (0.19)	0.16 (0.37)
84	0.1	0.07 (0.35)	0.01 (0.65)	0.17 (0.18)	0.35 (0.40)
84	0.3	0.09 (0.44)	0.16 (0.77)	0.18 (0.20)	0.72 (0.43)
84	0.5	-0.03 (0.41)	0.45 (0.66)	0.22 (0.23)	0.92 (0.34)
125	0	0.09 (0.38)	-0.07 (0.61)	0.17 (0.24)	0.17 (0.27)
125	0.1	0.14 (0.58)	0.05 (0.56)	0.22 (0.22)	0.31 (0.30)
125	0.3	0.03 (0.38)	0.27 (0.67)	0.23 (0.22)	0.62 (0.33)
125	0.5	0.10 (0.35)	0.45 (0.64)	0.19 (0.24)	0.89 (0.32)
250	0	0.11 (0.41)	0.08 (0.41)	0.20 (0.21)	0.24 (0.22)
250	0.1	0.10 (0.53)	0.19 (0.43)	0.19 (0.23)	0.36 (0.21)
250	0.3	0.04 (0.35)	0.38 (0.44)	0.18 (0.21)	0.65 (0.26)
250	0.5	0.05 (0.43)	0.71 (0.46)	0.23 (0.24)	0.95 (0.27)
500	0	0.04 (0.40)	0.07 (0.26)	0.19 (0.21)	0.21 (0.15)
500	0.1	0.02 (0.38)	0.25 (0.27)	0.19 (0.21)	0.36 (0.15)
500	0.3	0.08 (0.48)	0.53 (0.27)	0.19 (0.19)	0.67 (0.21)
500	0.5	0.04 (0.39)	0.85 (0.34)	0.21 (0.28)	0.95 (0.24)
750	0	0.03 (0.38)	0.13 (0.18)	0.21 (0.23)	0.19 (0.13)
750	0.1	0.00 (0.39)	0.27 (0.22)	0.19 (0.23)	0.37 (0.14)
750	0.3	0.04 (0.34)	0.57 (0.20)	0.20 (0.22)	0.67 (0.17)
750	0.5	0.03 (0.41)	0.92 (0.29)	0.24 (0.29)	1.01 (0.23)
1250	0	0.05 (0.37)	0.15 (0.13)	0.17 (0.18)	0.18 (0.09)
1250	0.1	0.01 (0.37)	0.30 (0.16)	0.17 (0.21)	0.36 (0.12)
1250	0.3	0.00 (0.40)	0.64 (0.20)	0.20 (0.24)	0.72 (0.17)
1250	0.5	0.03 (0.36)	0.94 (0.23)	0.27 (0.24)	0.98 (0.18)

A.5 ALL LATENT VARIABLES POSITIVELY SKEWED

Table 34. Parameter estimates for $\boldsymbol{\gamma_3^{(1)}}$ and $\boldsymbol{\gamma_3^{(2)}}$ by $\boldsymbol{n_2}$ and $\boldsymbol{\Delta\gamma_3}$ for fifth distribution condition.

		M-B		QML	
$\boldsymbol{n_2}$	$\boldsymbol{\Delta\gamma_3}$	$\boldsymbol{\gamma_3^{(1)}}$	$\boldsymbol{\gamma_3^{(2)}}$	$\boldsymbol{\gamma_3^{(1)}}$	$\boldsymbol{\gamma_3^{(2)}}$
84	0	0.12 (0.38)	0.06 (0.59)	0.18 (0.25)	0.19 (0.38)
84	0.1	0.20 (0.82)	0.00 (0.71)	0.18 (0.24)	0.38 (0.4)
84	0.3	0.12 (0.38)	0.34 (0.65)	0.18 (0.22)	0.63 (0.37)
84	0.5	0.12 (0.63)	0.47 (0.73)	0.14 (0.23)	0.88 (0.39)
125	0.0	0.09 (0.38)	0.14 (0.57)	0.20 (0.24)	0.23 (0.36)
125	0.1	0.07 (0.44)	0.16 (0.51)	0.19 (0.23)	0.35 (0.3)
125	0.3	0.09 (0.47)	0.34 (0.62)	0.19 (0.24)	0.64 (0.38)
125	0.5	0.10 (0.31)	0.67 (0.57)	0.20 (0.23)	0.90 (0.32)
250	0.0	0.11 (0.44)	0.14 (0.35)	0.15 (0.22)	0.18 (0.21)
250	0.1	0.10 (0.42)	0.26 (0.43)	0.21 (0.22)	0.37 (0.23)
250	0.3	0.12 (0.39)	0.45 (0.45)	0.22 (0.24)	0.67 (0.29)
250	0.5	0.11 (0.57)	0.77 (0.44)	0.19 (0.26)	0.93 (0.28)
500	0.0	0.05 (0.36)	0.12 (0.24)	0.18 (0.20)	0.16 (0.13)
500	0.1	0.15 (0.46)	0.27 (0.26)	0.17 (0.19)	0.33 (0.15)
500	0.3	0.14 (0.37)	0.60 (0.31)	0.18 (0.23)	0.70 (0.21)
500	0.5	0.10 (0.39)	0.94 (0.30)	0.23 (0.28)	1.00 (0.25)
750	0.0	0.03 (0.35)	0.17 (0.19)	0.18 (0.22)	0.17 (0.12)
750	0.1	0.05 (0.33)	0.28 (0.17)	0.15 (0.19)	0.32 (0.13)
750	0.3	0.11 (0.36)	0.66 (0.23)	0.18 (0.24)	0.72 (0.18)
750	0.5	0.10 (0.37)	0.94 (0.26)	0.25 (0.29)	1.03 (0.24)
1250	0.0	0.09 (0.38)	0.14 (0.16)	0.16 (0.20)	0.16 (0.10)
1250	0.1	0.11 (0.34)	0.30 (0.15)	0.18 (0.22)	0.33 (0.11)
1250	0.3	0.05 (0.40)	0.67 (0.19)	0.22 (0.24)	0.73 (0.16)
1250	0.5	0.09 (0.35)	0.96 (0.22)	0.25 (0.26)	1.02 (0.20)

APPENDIX B

DESCRIPTIVE SATISTICS FOR PARAMETER RECOVERY

B.1 ABSOLUTE BIAS BY IVS

Table 35. Mean Absolute Bias for $\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$ by IVs.

n_2	M-B	QML	**Distribution**	M-B	QML	$\Delta\gamma_3$	M-B	QML
84	0.414	0.259	Normal	0.252	0.197	0.0	0.212	0.143
125	0.343	0.228	IV skew	0.253	0.198	0.1	0.224	0.166
250	0.251	0.200	Both IV skew	0.258	0.199	0.3	0.267	0.228
500	0.198	0.180	DV Skew	0.250	0.209	0.5	0.306	0.276
750	0.179	0.178	All Skew	0.245	0.213			
1250	0.172	0.175						

Table 36. Mean Absolute Bias for Main Effects' Exogenous Covariances by IVs.

n_2	M-B	QML	**Distribution**	M-B	QML	$\Delta\gamma_3$	M-B	QML
84	0.281	0.102	Normal	0.189	0.086	0.0	0.192	0.085
125	0.254	0.097	IV skew	0.192	0.085	0.1	0.188	0.086
250	0.206	0.086	Both IV skew	0.175	0.085	0.3	0.187	0.085
500	0.162	0.080	DV Skew	0.200	0.085	0.5	0.184	0.086
750	0.138	0.075	All Skew	0.182	0.085			
1250	0.117	0.072						

B.2 BIAS RESULTS BY IV

Table 37. Mean Bias of $\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$ by IVs.

IV		$\boldsymbol{\gamma_1}$		$\boldsymbol{\gamma_2}$	
		M-B	QML	M-B	QML
$\boldsymbol{n_2}$	84	-0.011	0.011	-0.034	0.007
	125	-0.010	0.008	-0.029	0.007
	250	0.005	0.012	-0.016	0.009
	500	0.002	0.007	-0.010	0.009
	750	0.004	0.008	-0.005	0.008
	1250	0.001	0.006	-0.001	0.009
$\boldsymbol{\Delta\gamma_3}$	0.0	0.000	0.009	-0.015	0.008
	0.1	-0.002	0.009	-0.014	0.006
	0.3	-0.002	0.007	-0.014	0.009
	0.5	0.001	0.010	-0.016	0.009
Distribution	Normal	-0.003	0.007	-0.020	0.006
	IV skew	-0.001	0.007	-0.016	0.009
	Both IV skew	0.001	0.006	-0.001	0.010
	DV Skew	0.003	0.013	-0.027	0.006
	All Skew	-0.004	0.010	-0.010	0.010

Table 38. Mean Bias of $\boldsymbol{\gamma_3^{(1)}}$ and $\boldsymbol{\gamma_3^{(2)}}$ by IVs.

IV		$\boldsymbol{\gamma_3^{(1)}}$		$\boldsymbol{\gamma_3^{(2)}}$	
		M-B	QML	M-B	QML
$\boldsymbol{n_2}$	84	0.035	0.075	-0.046	0.191
	125	0.033	0.081	0.025	0.193
	250	0.041	0.075	0.121	0.212
	500	0.058	0.082	0.167	0.214
	750	0.040	0.084	0.185	0.222
	1250	0.031	0.087	0.203	0.227
$\boldsymbol{\Delta\gamma_3}$	0.0	0.033	0.071	0.046	0.075
	0.1	0.037	0.074	0.062	0.142
	0.3	0.046	0.080	0.141	0.268
	0.5	0.042	0.098	0.223	0.354
Distribution	Normal	0.050	0.080	0.127	0.203
	IV skew	0.081	0.070	0.149	0.201
	Both IV skew	0.122	0.057	0.185	0.196
	DV Skew	-0.052	0.104	0.030	0.224
	All Skew	-0.001	0.092	0.094	0.225

Table 39. Mean Bias for $\boldsymbol{\phi_{12}}$ by IVs.

IV		ϕ_{12}	
		M-B	QML
$\boldsymbol{n_2}$	84	0.002	0.006
	125	-0.002	0.007
	250	0.000	0.006
	500	0.006	0.007
	750	0.008	0.006
	1250	0.009	0.006
$\boldsymbol{\Delta\gamma_3}$	0.0	0.004	0.006
	0.1	0.005	0.006
	0.3	0.003	0.006
	0.5	0.005	0.007
Distribution	Normal	0.006	0.007
	IV skew	0.005	0.006
	Both IV skew	0.006	0.005
	DV Skew	0.003	0.009
	All Skew	0.002	0.005

Table 40. Mean Bias for $\boldsymbol{\phi_{11}^{(1)}}$ and $\boldsymbol{\phi_{22}^{(1)}}$ by IVs.

IV		$\phi_{11}^{(1)}$		$\phi_{22}^{(1)}$	
		M-B	QML	M-B	QML
$\boldsymbol{n_2}$	84	0.040	0.005	0.222	0.005
	125	0.031	0.005	0.240	0.009
	250	0.008	0.000	0.180	0.001
	500	0.005	0.005	0.107	0.003
	750	-0.001	0.002	0.060	0.002
	1250	-0.004	0.004	0.023	-0.001
$\boldsymbol{\Delta\gamma_3}$	0.0	0.010	0.003	0.138	0.002
	0.1	0.013	0.005	0.130	0.004
	0.3	0.010	0.002	0.132	0.003
	0.5	0.013	0.004	0.129	0.004
Distribution	Normal	0.016	0.006	0.132	0.002
	IV skew	0.010	0.004	0.137	0.003
	Both IV skew	0.015	0.006	0.093	0.002
	DV Skew	0.011	0.003	0.170	0.006
	All Skew	0.007	-0.001	0.128	0.003

Table 41. Mean Bias of $\boldsymbol{\phi_{11}^{(2)}}$ and $\boldsymbol{\phi_{22}^{(2)}}$ by IVs.

IV		$\phi_{11}^{(2)}$		$\phi_{22}^{(2)}$	
		M-B	QML	M-B	QML
$\boldsymbol{n_2}$	84	0.045	0.001	0.106	0.001
	125	0.027	0.004	0.104	0.007
	250	0.011	0.002	0.073	0.004
	500	0.009	0.003	0.050	0.004
	750	0.008	0.002	0.034	0.005
	1250	0.009	0.005	0.019	0.002
$\boldsymbol{\Delta\gamma_3}$	0.0	0.017	0.001	0.068	0.002
	0.1	0.020	0.004	0.061	0.004
	0.3	0.016	0.002	0.057	0.003
	0.5	0.016	0.005	0.061	0.007
Distribution	Normal	0.020	0.004	0.066	0.003
	IV skew	0.014	0.003	0.068	0.004
	Both IV skew	0.020	0.004	0.047	0.003
	DV Skew	0.015	0.002	0.071	0.005
	All Skew	0.015	0.001	0.056	0.004

APPENDIX C

SIMULATION CODE

```
rm(list=ls());

setwd("C:/Temp/ryan");
library(MASS);
library(stringr);

ptm <- proc.time();

####################Data Generation Group 2####################.
bootrep <-250;
sample2 <- matrix(c(84,125,250,500,750,1250));

b1 <- .3;
b2 <- .3;
g2effect <- matrix(c(.1,.2,.4,.6));

for (l in 1:length(sample2)){
n2 <- sample2[l];

for (k in 1:length(g2effect)){
b3 <- g2effect[k];
psi11 <- 1 - (b1^2*(1) + b2^2*(1) + 2*b1*b2*(.3)*(1)*(1) + b3^2*(1));

for (diste in 1:5) {
for (replication in 1:200){

seed1 <-  replication*10000000 + n2*1000 + diste*100 + b3*10;
set.seed(seed1);
```

```
lambdaxy <- runif(9, min = .5, max = .7);
lambda <- matrix (c(lambdaxy),9,1);

Zeta1 <- rnorm(n2,0,sqrt(psi11));

Xi <- matrix(c(1,.3,
            .3,1),2,2);
kappa <- matrix(c(0,0),2,1);

f1f2 <- mvrnorm(n2,kappa,Xi);
Xi1 <- f1f2[,1] - mean(f1f2[,1]);
Xi2 <- f1f2[,2] - mean(f1f2[,2]);
Xi3 <- Xi1 * Xi2;

Eta1 <- b1*Xi1 + b2*Xi2 + b3*Xi3 + Zeta1;

if (diste == 1) {
d1 <- rnorm(n2,0,sqrt(1-lambda[1]^2));
d2 <- rnorm(n2,0,sqrt(1-lambda[2]^2));
d3 <- rnorm(n2,0,sqrt(1-lambda[3]^2));
d4 <- rnorm(n2,0,sqrt(1-lambda[4]^2));
d5 <- rnorm(n2,0,sqrt(1-lambda[5]^2));
d6 <- rnorm(n2,0,sqrt(1-lambda[6]^2));
e1 <- rnorm(n2,0,sqrt(1-lambda[7]^2));
e2 <- rnorm(n2,0,sqrt(1-lambda[8]^2));
e3 <- rnorm(n2,0,sqrt(1-lambda[9]^2));
}

if (diste == 2) {
d1s <- rchisq(n2,3);
d2s <- rchisq(n2,3);
d3s <- rchisq(n2,3);
d1 <- ((d1s - 3)/2.449) * sqrt(1-lambda[1]^2);
d2 <- ((d2s - 3)/2.449) * sqrt(1-lambda[2]^2);
d3 <- ((d3s - 3)/2.449) * sqrt(1-lambda[3]^2);
d4 <- rnorm(n2,0,sqrt(1-lambda[4]^2));
d5 <- rnorm(n2,0,sqrt(1-lambda[5]^2));
d6 <- rnorm(n2,0,sqrt(1-lambda[6]^2));
e1 <- rnorm(n2,0,sqrt(1-lambda[7]^2));
e2 <- rnorm(n2,0,sqrt(1-lambda[8]^2));
e3 <- rnorm(n2,0,sqrt(1-lambda[9]^2));
}

if (diste == 3) {
d1s <- rchisq(n2,3);
d2s <- rchisq(n2,3);
```

```
d3s <- rchisq(n2,3);
d1 <- ((d1s - 3)/2.449) * sqrt(1-lambda[1]^2);
d2 <- ((d2s - 3)/2.449) * sqrt(1-lambda[2]^2);
d3 <- ((d3s - 3)/2.449) * sqrt(1-lambda[3]^2);
d4s <- rchisq(n2,3);
d5s <- rchisq(n2,3);
d6s <- rchisq(n2,3);
d4 <- ((d4s - 3)/2.449) * sqrt(1-lambda[4]^2);
d5 <- ((d5s - 3)/2.449) * sqrt(1-lambda[5]^2);
d6 <- ((d6s - 3)/2.449) * sqrt(1-lambda[6]^2);
e1 <- rnorm(n2,0,sqrt(1-lambda[7]^2));
e2 <- rnorm(n2,0,sqrt(1-lambda[8]^2));
e3 <- rnorm(n2,0,sqrt(1-lambda[9]^2));
}

if (diste == 4) {
d1 <- rnorm(n2,0,sqrt(1-lambda[1]^2));
d2 <- rnorm(n2,0,sqrt(1-lambda[2]^2));
d3 <- rnorm(n2,0,sqrt(1-lambda[3]^2));
d4 <- rnorm(n2,0,sqrt(1-lambda[4]^2));
d5 <- rnorm(n2,0,sqrt(1-lambda[5]^2));
d6 <- rnorm(n2,0,sqrt(1-lambda[6]^2));
e1s <- rchisq(n2,3);
e2s <- rchisq(n2,3);
e3s <- rchisq(n2,3);
e1 <- ((e1s - 3)/2.449) * sqrt(1-lambda[7]^2);
e2 <- ((e2s - 3)/2.449) * sqrt(1-lambda[8]^2);
e3 <- ((e3s - 3)/2.449) * sqrt(1-lambda[9]^2);
}

if (diste == 5) {
d1s <- rchisq(n2,3);
d2s <- rchisq(n2,3);
d3s <- rchisq(n2,3);
d1 <- ((d1s - 3)/2.449) * sqrt(1-lambda[1]^2);
d2 <- ((d2s - 3)/2.449) * sqrt(1-lambda[2]^2);
d3 <- ((d3s - 3)/2.449) * sqrt(1-lambda[3]^2);
d4s <- rchisq(n2,3);
d5s <- rchisq(n2,3);
d6s <- rchisq(n2,3);
d4 <- ((d4s - 3)/2.449) * sqrt(1-lambda[4]^2);
d5 <- ((d5s - 3)/2.449) * sqrt(1-lambda[5]^2);
d6 <- ((d6s - 3)/2.449) * sqrt(1-lambda[6]^2);
e1s <- rchisq(n2,3);
e2s <- rchisq(n2,3);
e3s <- rchisq(n2,3);
```

```
e1 <- ((e1s - 3)/2.449) * sqrt(1-lambda[7]^2);
e2 <- ((e2s - 3)/2.449) * sqrt(1-lambda[8]^2);
e3 <- ((e3s - 3)/2.449) * sqrt(1-lambda[9]^2);
}


x1 <- lambda[1] * Xi1 + d1;
x2 <- lambda[2] * Xi1 + d2;
x3 <- lambda[3] * Xi1 + d3;
x4 <- lambda[4] * Xi2 + d4;
x5 <- lambda[5] * Xi2 + d5;
x6 <- lambda[6] * Xi2 + d6;
y7 <- lambda[7] * Eta1 + e1;
y8 <- lambda[8] * Eta1 + e2;
y9 <- lambda[9] * Eta1 + e3;

datasetg2 <- cbind(x1,x2,x3,x4,x5,x6,y7,y8,y9,1,n2,b3,diste);

####################Data Generation Group 1####################.

n1 <- 250;
g1psi11 <- .756;
g1b3 <- .1;

g1Zeta1 <- rnorm(n1,0,sqrt(g1psi11));

g1f1f2 <- mvrnorm(n1,kappa,Xi);
g1Xi1 <- g1f1f2[,1] - mean(g1f1f2[,1]);
g1Xi2 <- g1f1f2[,2] - mean(g1f1f2[,2]);
g1Xi3 <- g1Xi1 * g1Xi2;

g1Eta1 <- b1*g1Xi1 + b2*g1Xi2 + g1b3*g1Xi3 + g1Zeta1;

if (diste == 1) {
g1d1 <- rnorm(n1,0,sqrt(1-lambda[1]^2));
g1d2 <- rnorm(n1,0,sqrt(1-lambda[2]^2));
g1d3 <- rnorm(n1,0,sqrt(1-lambda[3]^2));
g1d4 <- rnorm(n1,0,sqrt(1-lambda[4]^2));
g1d5 <- rnorm(n1,0,sqrt(1-lambda[5]^2));
g1d6 <- rnorm(n1,0,sqrt(1-lambda[6]^2));
g1e1 <- rnorm(n1,0,sqrt(1-lambda[7]^2));
g1e2 <- rnorm(n1,0,sqrt(1-lambda[8]^2));
g1e3 <- rnorm(n1,0,sqrt(1-lambda[9]^2));
}

if (diste == 2) {
```

```
g1d1s <- rchisq(n1,3);
g1d2s <- rchisq(n1,3);
g1d3s <- rchisq(n1,3);
g1d1 <- ((g1d1s - 3)/2.449) * sqrt(1-lambda[1]^2);
g1d2 <- ((g1d2s - 3)/2.449) * sqrt(1-lambda[2]^2);
g1d3 <- ((g1d3s - 3)/2.449) * sqrt(1-lambda[3]^2);
g1d4 <- rnorm(n1,0,sqrt(1-lambda[4]^2));
g1d5 <- rnorm(n1,0,sqrt(1-lambda[5]^2));
g1d6 <- rnorm(n1,0,sqrt(1-lambda[6]^2));
g1e1 <- rnorm(n1,0,sqrt(1-lambda[7]^2));
g1e2 <- rnorm(n1,0,sqrt(1-lambda[8]^2));
g1e3 <- rnorm(n1,0,sqrt(1-lambda[9]^2));
}

if (diste == 3) {
g1d1s <- rchisq(n1,3);
g1d2s <- rchisq(n1,3);
g1d3s <- rchisq(n1,3);
g1d1 <- ((g1d1s - 3)/2.449) * sqrt(1-lambda[1]^2);
g1d2 <- ((g1d2s - 3)/2.449) * sqrt(1-lambda[2]^2);
g1d3 <- ((g1d3s - 3)/2.449) * sqrt(1-lambda[3]^2);
g1d4s <- rchisq(n1,3);
g1d5s <- rchisq(n1,3);
g1d6s <- rchisq(n1,3);
g1d4 <- ((g1d4s - 3)/2.449) * sqrt(1-lambda[4]^2);
g1d5 <- ((g1d5s - 3)/2.449) * sqrt(1-lambda[5]^2);
g1d6 <- ((g1d6s - 3)/2.449) * sqrt(1-lambda[6]^2);
g1e1 <- rnorm(n1,0,sqrt(1-lambda[7]^2));
g1e2 <- rnorm(n1,0,sqrt(1-lambda[8]^2));
g1e3 <- rnorm(n1,0,sqrt(1-lambda[9]^2));
}

if (diste == 4) {
g1d1 <- rnorm(n1,0,sqrt(1-lambda[1]^2));
g1d2 <- rnorm(n1,0,sqrt(1-lambda[2]^2));
g1d3 <- rnorm(n1,0,sqrt(1-lambda[3]^2));
g1d4 <- rnorm(n1,0,sqrt(1-lambda[4]^2));
g1d5 <- rnorm(n1,0,sqrt(1-lambda[5]^2));
g1d6 <- rnorm(n1,0,sqrt(1-lambda[6]^2));
g1e1s <- rchisq(n1,3);
g1e2s <- rchisq(n1,3);
g1e3s <- rchisq(n1,3);
g1e1 <- ((g1e1s - 3)/2.449) * sqrt(1-lambda[7]^2);
g1e2 <- ((g1e2s - 3)/2.449) * sqrt(1-lambda[8]^2);
g1e3 <- ((g1e3s - 3)/2.449) * sqrt(1-lambda[9]^2);
}
```

```
if (diste == 5) {
g1d1s <- rchisq(n1,3);
g1d2s <- rchisq(n1,3);
g1d3s <- rchisq(n1,3);
g1d1 <- ((g1d1s - 3)/2.449) * sqrt(1-lambda[1]^2);
g1d2 <- ((g1d2s - 3)/2.449) * sqrt(1-lambda[2]^2);
g1d3 <- ((g1d3s - 3)/2.449) * sqrt(1-lambda[3]^2);
g1d4s <- rchisq(n1,3);
g1d5s <- rchisq(n1,3);
g1d6s <- rchisq(n1,3);
g1d4 <- ((g1d4s - 3)/2.449) * sqrt(1-lambda[4]^2);
g1d5 <- ((g1d5s - 3)/2.449) * sqrt(1-lambda[5]^2);
g1d6 <- ((g1d6s - 3)/2.449) * sqrt(1-lambda[6]^2);
g1e1s <- rchisq(n1,3);
g1e2s <- rchisq(n1,3);
g1e3s <- rchisq(n1,3);
g1e1 <- ((g1e1s - 3)/2.449) * sqrt(1-lambda[7]^2);
g1e2 <- ((g1e2s - 3)/2.449) * sqrt(1-lambda[8]^2);
g1e3 <- ((g1e3s - 3)/2.449) * sqrt(1-lambda[9]^2);
}

g1x1 <- lambda[1] * g1Xi1 + g1d1;
g1x2 <- lambda[2] * g1Xi1 + g1d2;
g1x3 <- lambda[3] * g1Xi1 + g1d3;
g1x4 <- lambda[4] * g1Xi2 + g1d4;
g1x5 <- lambda[5] * g1Xi2 + g1d5;
g1x6 <- lambda[6] * g1Xi2 + g1d6;
g1y7 <- lambda[7] * g1Eta1 + g1e1;
g1y8 <- lambda[8] * g1Eta1 + g1e2;
g1y9 <- lambda[9] * g1Eta1 + g1e3;


datasetg1 <- cbind(g1x1,g1x2,g1x3,g1x4,g1x5,g1x6,g1y7,g1y8,g1y9,0,n2,b3,diste);

write.table(datasetg2, file="temp2", quote=F, row.names=F, col.names=F, sep=" ")
write.table(datasetg1, file= "temp1",quote=F, row.names=F, col.names=F, sep=" ")
write.table(datasetg1, file= "temp",quote=F, row.names=F, col.names=F, sep=" ")
file.append("temp", "temp2");

################### M-B (EQS) analysis ###################.
eqssyntax1 <- "/Title";
eqssyntax2 <- "constrained model group1";
eqssyntax3 <- "/Specifications";
eqssyntax4 <- paste("data= 'C:\\Temp\\ryan\\temp1';");
eqssyntax5 <- paste("Variables=10; Cases= ",n1,";");
```

```
eqssyntax6 <- "Method=ALS; Analysis=moment; Matrix=raw;" ;
eqssyntax7 <- "group = 2;"
eqssyntax8 <- "/Interaction";
eqssyntax9 <- "F3= F1&F2;";
eqssyntax10 <- "/Equations";
eqssyntax11 <- "v1 = 1f1 + e1;";
eqssyntax12 <- "v2 = *f1 + e2;";
eqssyntax13 <- "v3 = *f1 + e3;";
eqssyntax14 <- "v4 = 1f2 + e4;";
eqssyntax15 <- "v5 = *f2 + e5;";
eqssyntax16 <- "v6 = *f2 + e6;";
eqssyntax17 <- "v7 = 1f4 + e7;";
eqssyntax18 <- "v8 = *f4 + e8;";
eqssyntax19 <- "v9 = *f4 + e9;";
eqssyntax20 <- "f4 = *f1 + *f2 + *f3 + d4;";
eqssyntax21 <- "f3 = *v999 + d3;";
eqssyntax22 <- "/Variances";
eqssyntax23 <- "f1 to f2 = *;";
eqssyntax24 <- "e1 to e9 = *;";
eqssyntax25 <- "d4 = *;";
eqssyntax26 <- "d3 = 0;";
eqssyntax27 <- "/Covariance"
eqssyntax28 <- "f1,f2 = *;";
eqssyntax29 <- "/Print";
eqssyntax30 <- "Fit = all;";
eqssyntax31 <- "Table = compact;";
eqssyntax32 <- "cov=yes;";
eqssyntax33 <- "/Technical"
eqssyntax34 <- "itr = 1;";
eqssyntax35 <- "aitr = 5000;";
eqssyntax36 <- "/Output";
eqssyntax37 <- "data = 'C:\\Temp\\ryan\\eqsconstrained.ets';";
eqssyntax38 <- "parameters;";
eqssyntax39 <- "rsquare;";
eqssyntax40 <- "standard error;";
eqssyntax41 <- "/end"
eqssyntax42 <- ""
eqssyntax43 <- "/Title";
eqssyntax44 <- "constrained model group2";
eqssyntax45 <- "/Specifications";
eqssyntax46 <- paste("data= 'C:\\Temp\\ryan\\temp2';");
eqssyntax47 <- paste("Variables=10; Cases= ",n2,";");
eqssyntax48 <- "Method=ALS; Analysis=moment; Matrix=raw;";
eqssyntax49 <- "/Interaction";
eqssyntax50 <- "F3= F1&F2;";
eqssyntax51 <- "/Equations";
```

```
eqssyntax52 <- "v1 = 1f1 + e1;";
eqssyntax53 <- "v2 = *f1 + e2;";
eqssyntax54 <- "v3 = *f1 + e3;";
eqssyntax55 <- "v4 = 1f2 + e4;";
eqssyntax56 <- "v5 = *f2 + e5;";
eqssyntax57 <- "v6 = *f2 + e6;";
eqssyntax58 <- "v7 = 1f4 + e7;";
eqssyntax59 <- "v8 = *f4 + e8;";
eqssyntax60 <- "v9 = *f4 + e9;";
eqssyntax61 <- "f4 = *f1 + *f2 + *f3 + d4;";
eqssyntax62 <- "f3 = *v999 + d3;";
eqssyntax63 <- "/Variances";
eqssyntax64 <- "f1 to f2 = *;";
eqssyntax65 <- "e1 to e9 = *;";
eqssyntax66 <- "d4 = *;";
eqssyntax67 <- "d3 = 0;";
eqssyntax68 <- "/Covariance"
eqssyntax69 <- "f1,f2 = *;";
eqssyntax70 <- "";
eqssyntax71 <- "/Constraint";
eqssyntax72 <- "(1,V2,F1) = (2,V2,F1);";
eqssyntax73 <- "(1,V3,F1) = (2,V3,F1);";
eqssyntax74 <- "(1,V5,F2) = (2,V5,F2);";
eqssyntax75 <- "(1,V6,F2) = (2,V6,F2);";
eqssyntax76 <- "(1,V8,F4) = (2,V8,F4);";
eqssyntax77 <- "(1,V9,F4) = (2,V9,F4);";
eqssyntax78 <- "(1,F4,F3) = (2,F4,F3);";
eqssyntax79 <- "(1,F4,F1) = (2,F4,F1);";
eqssyntax80 <- "(1,F4,F2) = (2,F4,F2);";
eqssyntax81 <- "(1,F1,F2) = (2,F1,F2);";
eqssyntax82 <- "/end";

eqssyntax <- {};
for (i in 1:82){
temp <- (paste("eqssyntax",i,sep=""));
eqssyntax <- rbind(eqssyntax,eval(as.name(temp)));
}
write(eqssyntax,file="eqsconstrained.eqs",append=F);
system("wineqs.exe in=eqsconstrained.eqs out=eqsconstrained.out length=2000000", wait = T,
invisible = T);


## EQS Unconstrained
eqssyntax1 <- "/Title";
eqssyntax2 <- "unconstrained model group1" ;
eqssyntax3 <- "/Specifications";
```

```
eqssyntax4 <- paste("data= 'C:\\Temp\\ryan\\temp1';");
eqssyntax5 <- paste("Variables=10; Cases= ",n1,";");
eqssyntax6 <- "Method=ALS; Analysis=moment; Matrix=raw;" ;
eqssyntax7 <- "group = 2;"
eqssyntax8 <- "/Interaction";
eqssyntax9 <- "F3= F1&F2;";
eqssyntax10 <- "/Equations";
eqssyntax11 <- "v1 = 1f1 + e1;";
eqssyntax12 <- "v2 = *f1 + e2;";
eqssyntax13 <- "v3 = *f1 + e3;";
eqssyntax14 <- "v4 = 1f2 + e4;";
eqssyntax15 <- "v5 = *f2 + e5;";
eqssyntax16 <- "v6 = *f2 + e6;";
eqssyntax17 <- "v7 = 1f4 + e7;";
eqssyntax18 <- "v8 = *f4 + e8;";
eqssyntax19 <- "v9 = *f4 + e9;";
eqssyntax20 <- "f4 = *f1 + *f2 + *f3 + d4;";
eqssyntax21 <- "f3 = *v999 + d3;";
eqssyntax22 <- "/Variances";
eqssyntax23 <- "f1 to f2 = *;";
eqssyntax24 <- "e1 to e9 = *;";
eqssyntax25 <- "d4 = *;";
eqssyntax26 <- "d3 = 0;";
eqssyntax27 <- "/Covariance"
eqssyntax28 <- "f1,f2 = *;";
eqssyntax29 <- "/Print";
eqssyntax30 <- "Fit = all;";
eqssyntax31 <- "Table = compact;";
eqssyntax32 <- "cov=yes;";
eqssyntax33 <- "/Technical"
eqssyntax34 <- "itr = 1;";
eqssyntax35 <- "aitr = 5000;";
eqssyntax36 <- "/Output";
eqssyntax37 <- "data = 'C:\\Temp\\ryan\\eqsunconstrained.ets';";
eqssyntax38 <- "parameters;";
eqssyntax39 <- "rsquare;";
eqssyntax40 <- "standard error;";
eqssyntax41 <- "/end"
eqssyntax42 <- ""
eqssyntax43 <- "/Title";
eqssyntax44 <- "unconstrained model group2";
eqssyntax45 <- "/Specifications";
eqssyntax46 <- paste("data= 'C:\\Temp\\ryan\\temp2';");
eqssyntax47 <- paste("Variables=10; Cases= ",n2,";");
eqssyntax48 <- "Method=ALS; Analysis=moment; Matrix=raw;";
eqssyntax49 <- "/Interaction";
```

```
eqssyntax50 <- "F3= F1&F2;";
eqssyntax51 <- "/Equations";
eqssyntax52 <- "v1 = 1f1 + e1;";
eqssyntax53 <- "v2 = *f1 + e2;";
eqssyntax54 <- "v3 = *f1 + e3;";
eqssyntax55 <- "v4 = 1f2 + e4;";
eqssyntax56 <- "v5 = *f2 + e5;";
eqssyntax57 <- "v6 = *f2 + e6;";
eqssyntax58 <- "v7 = 1f4 + e7;";
eqssyntax59 <- "v8 = *f4 + e8;";
eqssyntax60 <- "v9 = *f4 + e9;";
eqssyntax61 <- "f4 = *f1 + *f2 + *f3 + d4;";
eqssyntax62 <- "f3 = *v999 + d3;";
eqssyntax63 <- "/Variances";
eqssyntax64 <- "f1 to f2 = *;";
eqssyntax65 <- "e1 to e9 = *;";
eqssyntax66 <- "d4 = *;";
eqssyntax67 <- "d3 = 0;";
eqssyntax68 <- "/Covariance"
eqssyntax69 <- "f1,f2 = *;";
eqssyntax70 <- "";
eqssyntax71 <- "/Constraint";
eqssyntax72 <- "(1,V2,F1) = (2,V2,F1);";
eqssyntax73 <- "(1,V3,F1) = (2,V3,F1);";
eqssyntax74 <- "(1,V5,F2) = (2,V5,F2);";
eqssyntax75 <- "(1,V6,F2) = (2,V6,F2);";
eqssyntax76 <- "(1,V8,F4) = (2,V8,F4);";
eqssyntax77 <- "(1,V9,F4) = (2,V9,F4);";
eqssyntax78 <- "(1,F4,F1) = (2,F4,F1);";
eqssyntax79 <- "(1,F4,F2) = (2,F4,F2);";
eqssyntax80 <- "(1,F1,F2) = (2,F1,F2);";
eqssyntax81 <- "/end";


eqssyntax <- {};
for (i in 1:81){
temp <- (paste("eqssyntax",i,sep=""));
eqssyntax <- rbind(eqssyntax,eval(as.name(temp)));
}
write(eqssyntax,file="eqsunconstrained.eqs",append=F);
system("wineqs.exe in=eqsunconstrained.eqs out=eqsunconstrained.out length=2000000", wait =
T, invisible = T);

temp1eqsconres <- scan("eqsconstrained.ets", na.strings=c("ANALYSIS", "USING",
"METHOD", "=", "LS", "AGLS"));
temp2eqsconres <- matrix(temp1eqsconres, nrow=2, ncol=292, byrow=T);
```

```
temp3eqsconres <- matrix(temp2eqsconres[temp2eqsconres[,6]!=1,],nrow=1,ncol=292);
#write.table(temp3eqsconres, file="tempeqsconres", append=T, quote=F, row.names=F,
col.names=F, sep=" ");

temp1eqsunconres <- scan("eqsunconstrained.ets", na.strings=c("ANALYSIS", "USING",
"METHOD", "=", "LS", "AGLS"));
temp2eqsunconres <- matrix(temp1eqsunconres, nrow=2, ncol=292, byrow=T);
temp3eqsunconres <- matrix(temp2eqsunconres[temp2eqsconres[,6]!=1,],nrow=1,ncol=292);
#write.table(temp3eqsunconres, file="tempeqsunconres", append=T, quote=F, row.names=F,
col.names=F, sep=" ");

tempeqsres <- cbind(replication,n2,b3,diste,temp3eqsunconres,temp3eqsconres);
write.table(tempeqsres, file="alleqsresults", append=T, quote=F, row.names=F, col.names=F,
sep=" ");


####################M-B (EQS) Bootstrapping Analysis###################.
#EQS Boot group 1#.
eqsboot1syntax1 <- "/Title";
eqsboot1syntax2 <- "model group1 boot" ;
eqsboot1syntax3 <- "/Specifications";
eqsboot1syntax4 <- paste("data= 'C:\\Temp\\ryan\\temp1';");
eqsboot1syntax5 <- paste("Variables=10; Cases= ",n1,";");
eqsboot1syntax6 <- "Method=ALS; Analysis=moment; Matrix=raw;" ;
eqsboot1syntax7 <- "";
eqsboot1syntax8 <- "/Interaction";
eqsboot1syntax9 <- "F3= F1&F2;";
eqsboot1syntax10 <- "/Equations";
eqsboot1syntax11 <- "v1 = 1f1 + e1;";
eqsboot1syntax12 <- "v2 = *f1 + e2;";
eqsboot1syntax13 <- "v3 = *f1 + e3;";
eqsboot1syntax14 <- "v4 = 1f2 + e4;";
eqsboot1syntax15 <- "v5 = *f2 + e5;";
eqsboot1syntax16 <- "v6 = *f2 + e6;";
eqsboot1syntax17 <- "v7 = 1f4 + e7;";
eqsboot1syntax18 <- "v8 = *f4 + e8;";
eqsboot1syntax19 <- "v9 = *f4 + e9;";
eqsboot1syntax20 <- "f4 = *f1 + *f2 + *f3 + d4;";
eqsboot1syntax21 <- "f3 = *v999 + d3;";
eqsboot1syntax22 <- "/Variances";
eqsboot1syntax23 <- "f1 to f2 = *;";
eqsboot1syntax24 <- "e1 to e9 = *;";
eqsboot1syntax25 <- "d4 = *;";
eqsboot1syntax26 <- "d3 = 0;";
eqsboot1syntax27 <- "/Covariance"
eqsboot1syntax28 <- "f1,f2 = *;";
```

```
eqsboot1syntax29 <- "/Print";
eqsboot1syntax30 <- "Fit = all;";
eqsboot1syntax31 <- "Table = compact;";
eqsboot1syntax32 <- "cov=yes;";
eqsboot1syntax33 <-  "/Output";
eqsboot1syntax34 <- paste("data= 'C:\\Temp\\ryan\\group1boot.ets';");
eqsboot1syntax35 <- "parameters;";
eqsboot1syntax36 <- "standard error;";
eqsboot1syntax37 <- "/simulation";
eqsboot1syntax38 <- paste ("MBB=",n1,";");
eqsboot1syntax39 <- paste ("replications=",bootrep,";");
eqsboot1syntax40 <- "save = no;";
eqsboot1syntax41 <- "/Technical";
eqsboot1syntax42 <- "ITR = 1;";
eqsboot1syntax43 <- "AITR = 5000;";
eqsboot1syntax44 <- "/end"


eqsboot1syntax <- {};
for (i in 1:44)
{
temp <- (paste("eqsboot1syntax",i,sep=""));
eqsboot1syntax <- rbind(eqsboot1syntax,eval(as.name(temp)));
}
write(eqsboot1syntax,file="group1boot.eqs",append=F);
system("wineqs.exe in=group1boot.eqs out=group1boot.out length=2000000", wait = T,
invisible = T);

#EQS Boot group 2#.
eqsboot2syntax1 <- "/Title";
eqsboot2syntax2 <- "model group2 boot" ;
eqsboot2syntax3 <- "/Specifications";
eqsboot2syntax4 <- paste("data= 'C:\\Temp\\ryan\\temp2';");
eqsboot2syntax5 <- paste("Variables=10; Cases= ",n2,";");
eqsboot2syntax6 <- "Method=ALS; Analysis=moment; Matrix=raw;" ;
eqsboot2syntax7 <- "";
eqsboot2syntax8 <- "/Interaction";
eqsboot2syntax9 <- "F3= F1&F2;";
eqsboot2syntax10 <- "/Equations";
eqsboot2syntax11 <- "v1 = 1f1 + e1;";
eqsboot2syntax12 <- "v2 = *f1 + e2;";
eqsboot2syntax13 <- "v3 = *f1 + e3;";
eqsboot2syntax14 <- "v4 = 1f2 + e4;";
eqsboot2syntax15 <- "v5 = *f2 + e5;";
eqsboot2syntax16 <- "v6 = *f2 + e6;";
eqsboot2syntax17 <- "v7 = 1f4 + e7;";
```

```
eqsboot2syntax18 <- "v8 = *f4 + e8;";
eqsboot2syntax19 <- "v9 = *f4 + e9;";
eqsboot2syntax20 <- "f4 = *f1 + *f2 + *f3 + d4;";
eqsboot2syntax21 <- "f3 = *v999 + d3;";
eqsboot2syntax22 <- "/Variances";
eqsboot2syntax23 <- "f1 to f2 = *;";
eqsboot2syntax24 <- "e1 to e9 = *;";
eqsboot2syntax25 <- "d4 = *;";
eqsboot2syntax26 <- "d3 = 0;";
eqsboot2syntax27 <- "/Covariance"
eqsboot2syntax28 <- "f1,f2 = *;";
eqsboot2syntax29 <- "/Print";
eqsboot2syntax30 <- "Fit = all;";
eqsboot2syntax31 <- "Table = compact;";
eqsboot2syntax32 <- "cov=yes;";
eqsboot2syntax33 <-  "/Output";
eqsboot2syntax34 <- paste("data= 'C:\\Temp\\ryan\\group2boot.ets';");
eqsboot2syntax35 <- "parameters;";
eqsboot2syntax36 <- "standard error;";
eqsboot2syntax37 <- "/simulation";
eqsboot2syntax38 <- paste ("MBB=",n2,";");
eqsboot2syntax39 <- paste ("replications=",bootrep,";");
eqsboot2syntax40 <- "save = no;";
eqsboot2syntax41 <- "/Technical";
eqsboot2syntax42 <- "ITR = 1;";
eqsboot2syntax43 <- "AITR = 5000;";
eqsboot2syntax44 <- "/end"

eqsboot2syntax <- {};
for (i in 1:44)
{
temp <- (paste("eqsboot2syntax",i,sep=""));
eqsboot2syntax <- rbind(eqsboot2syntax,eval(as.name(temp)));
}
write(eqsboot2syntax,file="group2boot.eqs",append=F);
system("wineqs.exe in=group2boot.eqs out=group2boot.out length=2000000", wait = T,
invisible = T);

temp1boot1res <- scan("group1boot.ets", na.strings=c("ANALYSIS", "USING", "METHOD",
"=", "LS", "AGLS", "IN", "REPLICATION"));
l1boot <- length(temp1boot1res)/193;
temp2boot1res <- matrix(temp1boot1res, nrow=l1boot, ncol=193, byrow=T);
temp3boot1res <- matrix(temp2boot1res[temp2boot1res[,9]!=1,], ncol=193);
#write.table(temp3boot1res, file="eqsboottemp1", quote=F, row.names=F, col.names=F, sep="
");
```

```
temp1boot2res <- scan("group2boot.ets", na.strings=c("ANALYSIS", "USING", "METHOD",
"=", "LS", "AGLS", "IN", "REPLICATION"));
l2boot <- length(temp1boot2res)/193;
temp2boot2res <- matrix(temp1boot2res, nrow=l2boot, ncol=193, byrow=T);
temp3boot2res <- matrix(temp2boot2res[temp2boot2res[,9]!=1,], ncol=193);
#write.table(temp3boot2res, file="eqsboottemp2", quote=F, row.names=F, col.names=F, sep="
");

kstemp <- ks.test (temp3boot1res[,145],temp3boot2res[,145]);
kstemp2 <- matrix(kstemp);
ksteststat <- kstemp2[1,];
kspvalue <- kstemp2[2,];
meang1 <- mean(temp3boot1res[,145]);
meang2 <- mean(temp3boot2res[,145]);
mediang1 <- median(temp3boot1res[,145]);
mediang2 <- median(temp3boot2res[,145]);
sdg1 <- sd(temp3boot1res[,145])
sdg2 <- sd(temp3boot2res[,145]);
ci95g1lb <- meang1 - 1.96*(sdg1/bootrep);
ci95g1ub <- meang1 + 1.96*(sdg1/bootrep);
ci95g2lb <- meang2 + 1.96*(sdg1/bootrep);
ci95g2ub <- meang2 - 1.96*(sdg1/bootrep);

bootresults <- cbind(replication,n2,b3,diste, meang1, meang2, mediang1, mediang2, sdg1, sdg2,
ci95g1lb, ci95g1ub, ci95g2lb, ci95g2ub , ksteststat,kspvalue);
write.table(bootresults, file="allbootres", quote=F, row.names=F, append=T, col.names=F, sep="
");

####################QML (MPLUS) Analysis####################.
#mplus unconstrained
mplussyntax1 <- "Title:";
mplussyntax2 <- "model unconstrained";
mplussyntax3 <- "data:";
mplussyntax4 <- "file is C:\\Temp\\ryan\\temp;";
mplussyntax5 <- "variable:";
mplussyntax6 <- "names are x1-x6 y7-y9 g n effect dist;";
mplussyntax7 <- "usevariables are x1-x6 y7-y9;";
mplussyntax8 <- "Classes = c(2);";
mplussyntax9 <- "Knownclass = c (g = 0 g = 1);";
mplussyntax10 <- "analysis:";
mplussyntax11 <- "type = mixture random;";
mplussyntax12 <- "algorithm = integration;";
mplussyntax13 <- "";
mplussyntax14 <- "model:";
mplussyntax15 <- "%overall%";
mplussyntax16 <- "f1 by x1-x3;";
```

```
mplussyntax17 <- "f2 by x4-x6;";
mplussyntax18 <- "f4 by y7-y9;";
mplussyntax19 <- "f4 on f1 f2;";
mplussyntax20 <- "f1xf2 | f1 xwith f2;";
mplussyntax21 <- "f4 on f1xf2;";
mplussyntax22 <- "[f1@0 f2@0];";
mplussyntax23 <-"%c#2%";
mplussyntax24 <- "f4 on f1xf2;";
mplussyntax25 <- "[x1-x6];";
mplussyntax26 <- "[y7-y9];";
mplussyntax27 <- "x1-x6;";
mplussyntax28 <- "y7-y9;";
mplussyntax29 <- "f1;";
mplussyntax30 <- "f2;";
mplussyntax31 <- "savedata:";
mplussyntax32 <- "results are C:\\Temp\\ryan\\mplusunconstrained.dat;";

mplussyntax <- {};
for (m in 1:32) {
temp <- (paste("mplussyntax",m,sep=""));
mplussyntax <- rbind(mplussyntax,eval(as.name(temp)));
}
write(mplussyntax,file="mplusunconstrained.inp",append=F);
system("Mplus.exe mplusunconstrained.inp", show.output.on.console=F);


#mplus constrained
mplussyntax1 <- "Title:";
mplussyntax2 <- "model constrained";
mplussyntax3 <- "data:";
mplussyntax4 <- "file is C:\\Temp\\ryan\\temp;";
mplussyntax5 <- "variable:";
mplussyntax6 <- "names are x1-x6 y7-y9 g n effect dist;";
mplussyntax7 <- "usevariables are x1-x6 y7-y9;";
mplussyntax8 <- "Classes = c(2);";
mplussyntax9 <- "Knownclass = c (g = 0 g = 1);";
mplussyntax10 <- "analysis:";
mplussyntax11 <- "type = mixture random;";
mplussyntax12 <- "algorithm = integration;";
mplussyntax13 <- "";
mplussyntax14 <- "model:";
mplussyntax15 <- "%overall%";
mplussyntax16 <- "f1 by x1-x3;";
mplussyntax17 <- "f2 by x4-x6;";
mplussyntax18 <- "f4 by y7-y9;";
mplussyntax19 <- "f4 on f1 f2;";
```

```
mplussyntax20 <- "f1xf2 | f1 xwith f2;";
mplussyntax21 <- "f4 on f1xf2;";
mplussyntax22 <- "[f1@0 f2@0];";
mplussyntax23 <-"%c#2%";
mplussyntax24 <- "[x1-x6];";
mplussyntax25 <- "[y7-y9];";
mplussyntax26 <- "x1-x6;";
mplussyntax27 <- "y7-y9;";
mplussyntax28 <- "f1;";
mplussyntax29 <- "f2;";
mplussyntax30 <- "savedata:";
mplussyntax31 <- "results are C:\\Temp\\ryan\\mplusconstrained.dat;";

mplussyntax <- {};
for (m in 1:31) {
temp <- (paste("mplussyntax",m,sep=""));
mplussyntax <- rbind(mplussyntax,eval(as.name(temp)));
}
write(mplussyntax,file="mplusconstrained.inp",append=F);
system("Mplus.exe mplusconstrained.inp", show.output.on.console=F);

checkunconstrained <- readLines("mplusunconstrained.out");
converge1 <- "THE MODEL ESTIMATION TERMINATED NORMALLY" %in%
checkunconstrained ;
if (converge1 == "TRUE"){
unconstrainedconverge = 0;
temp1mplusunconres <- scan("mplusunconstrained.dat");
temp2mplusunconres <- matrix(temp1mplusunconres, nrow=1, ncol=115, byrow=T);
}
if (converge1 == "FALSE"){
unconstrainedconverge = 1;
temp2mplusunconres <- matrix(0, nrow=1, ncol=115, byrow=T);
}

checkconstrained <- readLines("mplusconstrained.out");
converge2 <- "THE MODEL ESTIMATION TERMINATED NORMALLY" %in%
checkconstrained ;
if (converge2 == "TRUE"){
constrainedconverge = 0;
temp1mplusconres <- scan("mplusconstrained.dat");
temp2mplusconres <- matrix(temp1mplusconres, nrow=1, ncol=113, byrow=T);
}
if (converge2 == "FALSE"){
constrainedconverge = 1;
temp2mplusconres <- matrix(0, nrow=1, ncol=113, byrow=T);
}
```

```
tempmplusres <-
cbind(replication,n2,b3,diste,temp2mplusunconres,temp2mplusconres,unconstrainedconverge,
constrainedconverge);
write.table(tempmplusres, file="allmplusres", append=T, quote=F, row.names=F, col.names=F,
sep=" ");


}
}
}
}
proc.time()- ptm;
```

BIBLIOGRAPHY

Aiken, L. S., & West, S. G. (1991). *Multiple regression Testing and interpreting interactions.* Newbury Park: SAGE Publications.

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

Bakker, A. B., Shimazu, A., Demerouti, E., Shimada, K., and Kawakami, N. (2011). Crossover of work engagement among Japanese couples: Perspective taking by both partners. *Journal of Occupational Health Psychology*, 16, p112-125.

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods, 7,* 127-150.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika, 48,* 493-517.

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research, strategic, and statistical considerations.*Journal of Personality and Social Psychology*, 51, 1173-1182.

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. *Multivariate Analysis*, 6, 9-42.

Bentler, P. M. (2010). SEM with simplicity and accuracy. *Journal of Consumer Psychology*, 20, 215-220.

Bickel, P. J. & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The annals of statistics*, 9, 1196-1217.

Bickel, P. J., Gotze, F, van Zwet, W. R. (1997). Resampling fewer than n observations: Gains, losses and remedies for losses. *Statistica Sinica*, 7, 1-31.

Bisbe, J., Coenders G., Saris, W. E., Batista-Foguet, J. M., (2006). Correcting measurement error bias interaction models with small samples. *Metodoloski Zvezki*, 3: 267-287.

Bollen, K. A. & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Psychological Methods & Research*, 21, 205-229.

Bollen, K.A. (1989), *Structural Equations with Latent Variables*, New York: John Wiley & Sons, Inc.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605-634.

Boos, D. D. (2003). Introduction to the bootstrap world. *Statistical Science*, 18 168-174.

Bohnstedt, G. W., & Goldberger, A. S. (1969). On the exact covariance of products of random variables. *Journal of the American Statistical Association*, 64, 325-328.

Browne, M. W. (1984) Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83

Busemeyer, J. R., & Jones, L.E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. Psychological Bulletin, 93, 549-562.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman & Hall.

Chen, F.F., Sousa, K.H., & West, S.G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, 12, 471-492.

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464-504.

Coenders G., Batista-Foguet, J. M., Saris, W. E. (2007). Simple, efficient, and distribution-free approach to interaction effects in complex structural equation models. *Quality & Quantity*, 42: 369-396.

Cohen, J. (1978). Partialed products *are* interactions; partialed vectors *are* curve components. *Psychological Bulletin*, 85, 858-866.

Conover, W. J. (1999) *Practical Nonparametric Statistics*. New York: John Wiley & Sons, Inc.

Cortina, J. M., Chen, G., and Dunlap, W. P., (2001) Testing interaction effects in LISREL: Examination and illustration of available procedures. *Organizational Research Methods*, 4, 324-360.

Dawson, J. F., & Richter, A. W., (2006). Probing three-way interactions in moderated multiple regression: Development and application of a slope difference test. *Journal of Applied Psychology*, 91, 917-926.

Dempster, A.P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*. Series B, 39 pp1-38.

Dimitruk, P., Schermelleh-Engel, K., Kelava, A., Moosbrugger, H. (2007). Challenges in nonlinear structural equation modeling. *Methodology*, 3: 100-114.

Efon, B. & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York, NY. Chapman Hall.

Finney, S. J. & Distefano, C. (2006). Nonnormal and categorical data in structural equation modeling. In: G. R. Hancock & R. O. Mueller (Eds.) *Structural equation modeling: A second course*. pp 269-314.

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling, 7*, 356–410.

Ghisletta, P., & Spini, D. (2004). An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational and Behavioral Statistics,* 29, 421-437.

Gillard, J. W. & Iles, T. C. (2005). Method of moments estimation in linear regression with errors in both variables. [on-line paper]. http://cf.ac.uk/maths/resources/Iles_Gillard_tech_report.pdf

Henseler J. & Chin, W. W. (2010). A comparison of approaches for the analysis of latent interaction effects between latent variables using partial least squares path modeling. *Structural Equation Modeling*, 17, 82–109.

Hu, L.T., & Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural Equation Modeling. Concepts, Issues, and Applications* (pp.76-99). London: Sage.

Jaccard, J. & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Quantitative Methods in Psychology*, 117, pp 348-357.

Jonsson, F. (1998). Modeling interaction and nonlinear effects: A step-by-step LISREL example. In: R. E. Schumacker & G. A. Marcoulides (eds.), *Interactions and Nonlinear Effects in Structural Equation Models.* Mahwah, NJ: Lawrence Erlbaum, pp. 17–42.

Joreskög , K. & Yang, F. (1996). Non-linear 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 195-226.

Kaplan, D. (2009). *Structural equation modeling foundations and extensions. 2nd Ed.* Thousand Oaks CA: Sage Publications, Inc.

Kaplan, D. (1995). Statistical power in structural equation modeling. In R. H. Hoyle (Ed.), *Structural Equation Modeling. Concepts, Issues, and Applications* (pp.100-117). London: Sage.

Kenny, D. A. & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin, 96, 201-210.

Keppel G., & Wickens, T. D., (2004) *Design and analysis a researcher's handbook 4th ed.* Upper Saddle River, NJ: Pearson/Prentice Hall.

Kim, K. H., & Bentler, P. M. (2006). Data modeling: structural equation modeling. In J. L. Green, G. Camilli & P. B. Elmore (Eds.), *Complementary methods for research in education* (3rd ed., pp. 161-175). Mahwah, NJ: Lawrence Erlbaum Associates.

Kim, K. H. (2009, Spring Semester). Lecture on General SEM presented in Structural Equation Modeling. University of Pittsburgh, Pittsburgh, PA.

Klein, A. G. & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65: 457-474.

Klein, A.G. & Muthén, B. (2003). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. Unpublished Manuscript, Graduate School of Education, University of California Los Angeles.

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

Kline, R. (2005). *Principles and practice of structural equation modeling. 2nd Ed.* New York, NY: The Guilford Press.

Kutner, Nachtsheim, Neter, & Li, (2005). *Applied linear statistical models. 5th ed.* New York, NY:McGraw-Hill Irwin.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13-22.

Little, T. D., Bovaird, J. A., 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.

Little, T. D., Card, N. A., Bovaird, J. A, Preacher, K. J., & Crandall, C. S, (2007). Structural equation modeling of mediation and moderation with contextual factors. In T.D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling Contextual Effects in Longitudinal Studies* (pp. 207-230). Mahwah, NJ: Lawrence Erlbaum Associates.

MacCallum, R. C., & Mar, C. M. (1995). Disinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin, 118*(3), 405-421.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding, and suppression effect. *Prevention Science,* 1, 173-181.

Magnus, J.R. & H. Neudecker (1999). *Matrix Differential Calculus: With Applications in Statistics and Econometrics*. Revised Edition (Wiley, New York).

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 interaction and quadratic effects. In: G. R. Hancock & R. O. Mueller (Eds.) *Structural equation modeling: A second course*. pp 269-314.

Meijer, E. & Mooijaart, A. B. (1994). The use of third-order moments in structural models. *Questiio,* 18: 75-84.

Mooijaart, A. & Bentler, P. M. (2010). An alternative approach for modeling nonlinear latent variable models. *Structural Equation Modeling* 17: 357–373.

Mooijaart, A. & Satorra, A. (2009). On the insensitivity of the chi-square model test to nonlinear misspecifition in structural equation models. *Psychometrika,* 74, pp 443-455.

Mooijaart A. & Satorra, A. (2011). Moment testing for interaction terms in structural equation modeling. *Psychometrika,* 77, 65-84.

Moosbrugger, H., Schermelleh-Engel, K., Kelava; A. & Klein, A. G. (2009). Testing Multiple Nonlinear Effects in Structural Equation Modeling: A Comparison of Alternative Estimation Approaches. Invited Chapter in T. Teo & M. S. Khine (Eds.), *Structural Equation Modeling in Educational Research: Concepts and Applications*. Rotterdam, NL: Sense Publishers.

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

Muthén, B. O. & Asparouhov, T. (2003). Modeling interactions between latent and observed continuous variables using maximum-likelihood estimation in mplus. Mplus Web Note #6.

Muthén, L.K. & Muthén, B. O. (2010). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Nevitt, J. & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling,* 8, 353-377.

Ping, R.A. (2004). *Testing latent variable models with survey data, 2nd edition.* [on-line monograph]. http://www.wright.edu/~robert.ping/lv1/toc1.htm

Ping, R.A. (2005). "What is the "validity" of a latent variable interaction (or quadratic)?" [on-line paper]. http://home.att.net/~rpingjr/validity.doc

Ping, R.A. (2010). "How does one estimate categorical variables in theoretical model tests using structural equation analysis?" [on-line paper]. http://www.wright.edu/~robert.ping/categorical3.doc

Pohlig R. T. & Kim, K. H., (2010, May) *A Comparison of methods to estimate latent interaction effects.* Poster session presented at the annual convention of the American Psychological Science Convention, Washington, DC.

Preacher, K. J., Curran,P. J., & Bauer, D. J. (2006) Computational tools for probing interactions in multiple linear regression, multilevel modeling and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437-448.

Rigdon, E. E., Schumacker, R. E. & Wothke, W. (1996). A comparative review interaction and nonlinear modeling. In: G. A. Marcoulides & R. E. Schumacker (Eds.) *Advanced Structural Equation Modeling Issues and Techniques*. pp 1-16.

Rodgers, J. L. (1999). The Bootstrap, the Jackknife, and the Randomization Test: A Sampling Taxonomy. *Multivariate Behavioral Research*, 34(4), 441-456.

Rosner, B. (2006). *Fundamentals of biostatistics*. Belmont, CA: Thomson-Brooks/Cole.

Saris, W. E., Batista-Foguet, J. M., & Coenders, G. (2007). Selection of indicators for the interaction term in structural equation models with interaction. Quality & Quantity, 21, pp 55-72.

Schermelleh-Engel, K., Klein, A. & Moosbrugger, H. (1998). Estimating non-linear effects using a latent moderated structural equations approach. In: R. E. Schumacker & G. A. Marcoulides (eds.), *Interactions and Nonlinear Effects in Structural Equation Models.* Mahwah, NJ: Lawrence Erlbaum, pp. 203–238.

Schumacker, R. E., & Lomax, R. G. (1996). *A Beginner's Guide to Structural Equation Modeling*. Mahwah, NJ: Lawrence Erlbaum.

Schumacker, R. E. & Marcoulides, G. A. (1998). *Interactions and Nonlinear Effects in Structural Equation Models*. Mahwah, NJ: Lawrence Erlbaum.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, pp 422-445.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology 1982* (pp. 290-312). Washington DC: American Sociological Association.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 27, 229-239.

Stine, R. (1989). An introduction to bootstrap methods: Examples and ideas. *Sociological Methods & Research*, 18, 243-291.

Tabachnick, B. G., and Fidell, L. S. (2007). *Using Multivariate Statistics, 5th ed.* Boston: Allyn and Bacon.

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

Yung, Y-F., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In: G. A. Marcoulides & R. E. Schumacker (Eds.) *Advanced Structural Equation Modeling Issues and Techniques*. pp 195-226.