# Three-way Interactions with Latent Variables: A Maximum Likelihood Approach

Wenjing Huang

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology (Quantitative).

Chapel Hill
2008

Approved by

Patrick J. Curran, Ph.D.

Andrea M. Hussong, Ph.D.

Daniel J. Bauer, Ph.D.

© 2008

Wenjing Huang

ALL RIGHTS RESERVED

# ABSTRACT

Wenjing Huang: Three-way Interactions with Latent Variables:
A Maximum Likelihood Approach
(Under the direction of Patrick J. Curran)

Two-way interaction in latent variables has been a topic of considerable theoretical and practical interest among psychological methodologists. Since the seminal work of Kenny and Judd (1984), much research has focused on the use of product indicators for the estimation of latent moderation effects. These methods are usually difficult to use, and many popular approaches lack solid statistical justification. In recent years, the development of full-information maximum likelihood for nonlinear latent variables models provided a new approach to the estimation of latent variable interaction effects. However, a particular kind of three-way interaction, i.e., two-way latent variable interactions over an observed grouping variable, has received little attention. In this thesis, existing literature is reviewed and studied to arrive at a derivation of the full-information maximum likelihood estimator for three-way interactions in latent variables. It is also shown that this new method of estimation and testing can be implemented in Mplus (Muthén & Muthén, 1998–2007) using mixture modelling. To study the properties of this new estimation method, a simulation study is conducted, and the new method is shown to have superior performance than an existing method proposed by Marsh, Wen, and Hau (2004).

## ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Patrick Curran for his guidance throughout this project and my committee for helpful suggestions.

# TABLE OF CONTENTS

**List of Tables** . . . . . **vii**

**List of Figures** . . . . . **viii**

## CHAPTER

**1 Introduction** . . . . . **1**

1.1 Interaction in Regression . . . . . 1

1.2 The Kenny-Judd Model . . . . . 2

1.3 Fully Latent Regression with Interaction . . . . . 4

1.4 A Model with Three-way Interaction . . . . . 5

**2 Parameter Estimation** . . . . . **7**

2.1 Parameter Estimation by Discrepancy Functions . . . . . 7

2.2 Product Indicator Methods . . . . . 10

2.3 Full Information Maximum Likelihood . . . . . 13

2.4 Implementations of Full Information Methods . . . . . 17

2.5 Multiple Groups as Mixtures with Known Class Membership . . . . . 20

2.6 Non-normal Latent Variables . . . . . 21

**3 Methods** . . . . . **23**

3.1 Simulation Design . . . . . 23

3.2 Generating Model . . . . . 24

3.3 Methods for Summarizing Results . . . . . 27

**4 Results** . . . . . **29**

4.1 Convergence and Proper Solutions . . . . . 29

4.2 Likelihood Ratio Test of Interaction Effect . . . . . 31

4.3 Parameter Recovery and Standard Errors . . . . . 36

**5 Discussion** . . . . . **42**

5.1 On Model Assumptions . . . . . 42

5.2 On Relative Performance . . . . . 46

5.3 Limitations and Future Research . . . . . 47

**References** . . . . . **50**

# LIST OF TABLES

4.1 Rates of Convergence and Proper Model Estimation . . . . . 30

4.2 Likelihood Ratio Tests Within and Between Restricted and Unrestricted Models . . . . . 33

4.3 Empirical Estimates of Type I Error Rates and Statistical Power for Test of Equality Constraints on $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ . . . . . 35

4.4 Mean, Standard Deviation (*SD*), and Bias of Estimated $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ . . . . 37

4.5 Mean and Standard Deviation (*SD*) of Estimated Standard Errors (*SE*) and Empirical Rejection Rates of Univariate $z$ Tests of Estimated $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ . 38

# LIST OF FIGURES

3.1 Density Function of $\chi^2_6$ Distribution . . . . . 26

# CHAPTER 1

# Introduction

## 1.1 Interaction in Regression

Just as Aiken and West (1991) pointed out in their influential work on the statistical analysis of interaction effects in multiple regression, interactions arise naturally in many branches of psychology. For instance, in Baron and Kenny's (1986) seminal treatment of the distinction between moderators and mediators, their statistical framework for moderation analysis is linear regression with polynomial interactions. Developmental trajectory analysis is another area in which time-by-covariate interactions occur frequently when predictors are added into conditional latent curve models (Bollen & Curran, 2006). In industrial and organizational psychology, moderated regression has also received much attention (see e.g. Aguinis, 2002). In the parts of psychology where multilevel modelling is heavily used, interactions between variables measured at different levels of nesting units often provide answers to important research questions (Raudenbush & Bryk, 2002).

The interaction between two observed variables can be handled within the standard regression framework (see Aiken & West, 1991). Suppose there are three variables $x$, $y$ and $z$. Let $x$ and $z$ be the predictors, and let $y$ be the outcome variable. Consider the simple example given by Cudeck, Harring, and du Toit (in press), hereinafter CHD, in which $x$ is a measure of reasoning ability, $z$ a measure of quantitative ability, and $y$ a measure of academic achievement at school. It is natural to hypoth-

esize that the two different facets of cognitive ability interact when they are used to predict achievement. Let us assume for now that both $x$ and $z$ are free of measurement error. The regression model (for subject $i$) that contains an interaction term between $x$ and $z$ can be written as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + \varepsilon_i, \tag{1.1}$$

where $\varepsilon_i$ is an error term, $\beta_0$ the intercept, $\beta_1$ the effect of $x$, and $\beta_2$ the effect of $z$. The interaction term between $x$ and $z$ is simply the product of the two variables and $\beta_3$ is the coefficient for the interaction effect. By rearranging Equation (1.1), it is clear that the relationship between $x$ and $y$ depends on $z$,

$$y_i = (\beta_0 + \beta_2 z_i) + (\beta_1 + \beta_3 z_i) x_i + \varepsilon_i. \tag{1.2}$$

Hence $z$ can be interpreted as a *moderator* of the relationship between the *focal predictor* $x$ and the outcome $y$. By symmetry, $x$ can also be viewed as a moderator of the relationship between $z$ and $y$. The parameters in model (1.1) can be conveniently estimated with the ordinary least squares (OLS) method as shown by Cohen (1968).

## 1.2 The Kenny-Judd Model

However, when the assumption of no measurement error in the predictors is violated, it is a well known result that the estimates of regression coefficients are biased under the OLS method (Mardia, Kent, & Bibby, 1979). When multiple indicators of a variable are available, a solution is to use factor analysis in a latent variable structural equation modelling framework.

Continuing with the CHD example, suppose there are two latent predictor variables: $\xi_1$ as reasoning ability, and $\xi_2$ as quantitative ability. Now the two $\xi$'s interact to produce a nonlinear structural regression in explaining the variability in an ob-

served outcome variable $y$

$$y_i = \alpha + \gamma_1 \xi_{1i} + \gamma_2 \xi_{2i} + \gamma_3 \xi_{1i} \xi_{2i} + \zeta_i, \tag{1.3}$$

where $\zeta$ is the so-called equation disturbance.

Equation (1.3) is representative of a line of work initiated by Kenny and Judd (1984). They considered a case when there are two observed indicators for each latent variable. Let $x_1$ and $x_2$ be the indicators of $\xi_1$. In the CHD example cited above, they can be test scores of verbal reasoning and nonverbal reasoning. Let $x_3$ and $x_4$ be the indicators of $\xi_2$. In the CHD example, $x_3$ may be arithmetic skill, and $x_4$ computation skill. In Kenny and Judd's (1984) method, there is a product latent variable that is measured by four product indicator variables: $x_1x_3$, $x_1x_4$, $x_2x_3$, $x_2x_4$. They reasoned that information about the product latent variable can be obtained from exhaustive pairwise products of the original indicators. In factor analytic terms, the measurement model for $\xi_1$, $\xi_2$ and $\xi_1\xi_2$ may be written as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_1x_3 \\ x_1x_4 \\ x_2x_3 \\ x_2x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ \tau_1 \\ 0 \\ \tau_2 \\ 0 \\ \tau_3 \\ \tau_4 \\ \tau_5 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ \lambda_1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_3 \\ 0 & 0 & \lambda_4 \\ 0 & 0 & \lambda_5 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_1\xi_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{pmatrix}, \tag{1.4}$$

where the $\tau$'s are measurement intercepts, the $\lambda$'s are factor loadings, and the $\varepsilon$'s are error terms. It is clear that $x_1$, $x_3$ and $x_1x_3$ are made reference indicators (see e.g. Bollen, 1989) to set the scale for $\xi_1$, $\xi_2$, and $\xi_1\xi_2$.

However, there exist complicated nonlinear dependence among the parameters shown above. For example, let $\phi_{11}$ and $\phi_{22}$ denote the variances of $\xi_1$ and $\xi_2$ and let $\phi_{21}$ represent the covariance between them. The means for $\xi_1$, $\xi_2$ and $\xi_1\xi_2$ may be written as $\kappa_1$, $\kappa_2$ and $\kappa_3$. It can be derived that $\kappa_3$ is equal to $\phi_{21}$, which is the covariance between $\xi_1$ and $\xi_2$, and that the variance of $\xi_1\xi_2$ is equal to $\phi_{21}^2 + \phi_{11}\phi_{22}$, which completely depends on the variances and covariance of $\xi_1$ and $\xi_2$. Therefore one must be able to derive and impose those constraints in order to use methods based on product indicators, which gives rise to a number of complications in estimation and inference that will be discussed in further detail in Chapter 2.

## 1.3 Fully Latent Regression with Interaction

In recent years, methodologists have focused their attention on an extension of Kenny and Judd's (1984) model, where the outcome variable is also latent (e.g. Marsh et al., 2004). This is the model that will be discussed extensively in the sequel, so it is useful to describe it in some detail here.

The structural part of the fully latent regression model with interaction is

$$\eta_i = \alpha + \gamma_1\xi_{1i} + \gamma_2\xi_{2i} + \gamma_3\xi_{1i}\xi_{2i} + \zeta_i, \tag{1.5}$$

where $\eta$ is the latent outcome variable. The usual assumptions are made: $\xi_1$ and $\xi_2$ are jointly normally distributed; and $\zeta$ is normally distributed with mean zero and it is uncorrelated with $\xi_1$ and $\xi_2$.

Without loss of generality, let each latent variable be measured by three indicators. Instead of using all possible cross products of the original indicators as Kenny and Judd (1984) did, Yang Jonsson (1998) used matched pair indicators. That is, the first indicator $x_1$ of the first latent predictor variable $\xi_1$ is multiplied by the first indicator $x_4$ of the second latent predictor variable $\xi_2$ to form the first product indicator $x_1x_4$ of the latent interaction variable $\xi_1\xi_2$ and so on. The measurement model is an

expanded form of Equation (1.4) to accommodate the additional indicators:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_1x_4 \\ x_2x_5 \\ x_3x_6 \end{pmatrix} = \begin{pmatrix} 0 \\ \tau_1 \\ \tau_2 \\ 0 \\ \tau_3 \\ \tau_4 \\ 0 \\ \tau_5 \\ \tau_6 \\ 0 \\ \tau_7 \\ \tau_8 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ \lambda_1 & 0 & 0 & 0 \\ \lambda_2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \lambda_3 & 0 & 0 \\ 0 & \lambda_4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \lambda_5 & 0 \\ 0 & 0 & \lambda_6 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \lambda_7 \\ 0 & 0 & 0 & \lambda_8 \end{pmatrix} \begin{pmatrix} \eta \\ \xi_1 \\ \xi_2 \\ \xi_1\xi_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \end{pmatrix}. \tag{1.6}$$

Clearly, $y_1$, $x_1$, $x_4$ and $x_1x_4$ are the reference indicators. The measurement errors are jointly normal with zero means and a diagonal covariance matrix. Specifications (1.5) and (1.6) makes the model an example of a nonlinear structural equation model. To estimate the parameters in this model, one can follow the family of product indicator methods that are initiated by Kenny and Judd (1984), formalized by Jöreskog and Yang (1996), and empirically tested by Marsh et al. (2004). Alternatively, one can also use the full information maximum likelihood method without ever forming product indicators (Klein & Moosbrugger, 2000). More details of parameter estimation will be discussed in the next chapter.

## 1.4 A Model with Three-way Interaction

Here I consider a further extension of the model in (1.5). By three-way interaction, I do not mean the interaction of three latent variables. Rather, the third variable is an

observed grouping variable, over which the two-way interaction effects vary. That is, in group $g = 1, \ldots, G$, there is a fully latent regression with two-way interactions:

$$\eta_i^{(g)} = \alpha^{(g)} + \gamma_1^{(g)} \xi_{1i}^{(g)} + \gamma_2^{(g)} \xi_{2i}^{(g)} + \gamma_3^{(g)} (\xi_{1i} \xi_{2i})^{(g)} + \zeta_i^{(g)}. \tag{1.7}$$

The measurement model is the same as in (1.6). In other words, strict factorial invariance, where loadings, intercepts, and unique variances are invariant over the groups, is assumed as a simplifying condition for comparisons of the latent regression coefficients. The primary focus is on testing whether the regression coefficients (especially $\gamma_3^{(g)}$'s) differ over $g$.

I propose a likelihood ratio test. Specifically, one first fits a model with the $\gamma^{(g)}$'s freely estimated, recording its log-likelihood, and then fits a model that constrains the $\gamma^{(g)}$'s equal across the $G$ groups, also recording its log-likelihood. It is a well known result (e.g. Wilks, 1938) that negative 2 times the difference in the two log-likelihoods is distributed as a central chi-square variable under the null hypothesis with $G - 1$ degrees-of-freedom.

In empirical research, models such as (1.7) can be very useful. In the CHD example, suppose one hypothesizes that the two-way interaction between reasoning ability and quantitative ability differs by gender, model (1.7) gives a direct test of this hypothesis. In developmental psychopathology for another example, suppose one is interested in predicting adolescents' drug use with internalizing and externalizing symptoms, as well as their interaction, but also suspects that the interaction effect might be different for adolescents with or without alcoholic parents. Model (1.7) can be directly applied in such situations and be of great use.

However, efficient parameter estimation for such a nonlinear latent variable model is complex and has not yet been closely studied. I will next review these issues in detail and propose a solution.

# CHAPTER 2

# Parameter Estimation

## 2.1 Parameter Estimation by Discrepancy Functions

Here I introduce standard notation and terminology for parameter estimation in structural equation modelling. Traditionally a (linear) structural equation model refers to a parametric statistical model that specifies the mean and covariance structure of a set of observed variables. Hence the term structural equation model is synonymous to *mean and covariance structure model*. Section 2.3 will extend this definition, but for now, the discussion is limited to classical discrepancy function based estimators that rely on identifying parameters from sample moments such as the sample mean vector and covariance matrix.

Generically, let there be a $p \times 1$ random vector of observed variables, say $\mathbf{y}$, whose model-implied mean vector is $\boldsymbol{\mu}(\boldsymbol{\theta})$ and covariance matrix is $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a $q$-dimensional vector of parameters. It is understood that structural equation models impose structure on the moments of the observed random vector. For a sample of size $N$, let the sample mean vector be $\mathbf{m}$ and the sample covariance matrix be $\mathbf{S}$. It is assumed that there are no missing data, although this assumption can be relaxed (Enders, 2003). To estimate $\boldsymbol{\theta}$, one tries to minimize the discrepancy between the sample moments and the model-implied moments. Different definitions of discrepancy lead to different estimators.

Under multivariate normality of $\mathbf{y}$, the most widely used estimation method is

maximum Wishart likelihood (MWL). The name Wishart comes from the fact that the distribution of sample covariance matrix is scaled Wishart if the population is normal. The MWL discrepancy function for mean and covariance structure models is often written as

$$F_{MWL}(\boldsymbol{\theta}) = \log|\boldsymbol{\Sigma}| + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log|\mathbf{S}| - q + [\mathbf{m} - \boldsymbol{\mu}]'\boldsymbol{\Sigma}^{-1}[\mathbf{m} - \boldsymbol{\mu}], \tag{2.1}$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are functions of $\boldsymbol{\theta}$ (see e.g. Bollen, 1989). If the model is correctly specified and the distributional assumptions are met, minimization of $F_{MWL}$ leads to the maximum likelihood estimate (MLE).

Another approach for defining discrepancy is the method of least squares. The following weighted least squares (WLS) discrepancy function is often used

$$F_{WLS}(\boldsymbol{\theta}) = (\mathbf{s} - \boldsymbol{\sigma})'\mathbf{W}^{-1}(\mathbf{s} - \boldsymbol{\sigma}) + [\mathbf{m} - \boldsymbol{\mu}]'\mathbf{S}^{-1}[\mathbf{m} - \boldsymbol{\mu}], \tag{2.2}$$

where $\mathbf{s} = \mathrm{vech}(\mathbf{S})$ and $\boldsymbol{\sigma} = \mathrm{vech}(\boldsymbol{\Sigma})$, and $\mathrm{vech}(\cdot)$ is the half-vectorization operator that stacks the non-duplicated elements of a symmetric matrix into a vector. The matrix $\mathbf{W}$ is a symmetric positive definite weight matrix that is equal to a consistent estimate of the asymptotic covariance matrix of $\mathbf{s}$. This discrepancy function does not assume multivariate normality of the indicator vector, but it does assume that the sample mean vector and covariance matrix are independent.

Yet another approach is the method of weighted least squares for augmented moments (WLSA; see Jöreskog & Yang, 1996). Instead of the central moments, e.g., covariances, WLSA works directly with the raw moments, e.g., cross-product moments. Specifically, an augmented moment matrix in the sample is defined as

$$\mathbf{A} = \frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N}\mathbf{y}_i\mathbf{y}_i' & \sum_{i=1}^{N}\mathbf{y}_i \\ \sum_{i=1}^{N}\mathbf{y}_i' & N \end{pmatrix}, \tag{2.3}$$

and in the population as

$$\mathbb{A} = \begin{pmatrix} \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}' & \boldsymbol{\mu} \\ \boldsymbol{\mu}' & 1 \end{pmatrix}. \tag{2.4}$$

Let $\mathbf{a} = \text{vech}(\mathbf{A})$ and $\boldsymbol{\alpha} = \text{vech}(\mathbb{A})$. The WLSA discrepancy function is defined as

$$F_{WLSA}(\boldsymbol{\theta}) = (\mathbf{a} - \boldsymbol{\alpha})'\mathbf{W}_{\mathbf{a}}^{-}(\mathbf{a} - \boldsymbol{\alpha}), \tag{2.5}$$

where $\mathbf{W}_{\mathbf{a}}$ is a consistent estimate of the covariance matrix of $\mathbf{a}$ and $\mathbf{W}_{\mathbf{a}}^{-}$ is its Moore-Penrose (generalized) inverse. A generalized inverse is required because $\mathbf{W}_{\mathbf{a}}$ is rank deficient. Note that this discrepancy function does not require the assumption of independence between the sample mean vector and the covariance matrix, as is required by WLS, which will subsequently be important when latent variable interactions are specified.

Extensions of discrepancy function based estimation methods to more than one groups is straightforward. Take MWL for example. The vector of parameters is still a $q \times 1$ vector $\boldsymbol{\theta}$. In group $g = 1, \ldots, G$, the model-implied mean vector and covariance matrix is $\boldsymbol{\mu}^{(g)}$ and $\boldsymbol{\Sigma}^{(g)}$, respectively. Both are still functions of $\boldsymbol{\theta}$. Similarly, the sample mean vector and covariance matrix in group $g$ are $\mathbf{m}^{(g)}$ and $\mathbf{S}^{(g)}$. Let the sample size in group $g$ be $N_g$ so that $N = \sum_{g=1}^{G} N_g$. The discrepancy function for all $G$ groups is defined as a linear combination of group-specific discrepancy functions

$$F_{MWL}(\boldsymbol{\theta}) = \sum_{g=1}^{G} \frac{N_g}{N} F_{MWL}^{(g)}(\boldsymbol{\theta}), \tag{2.6}$$

where

$$\begin{aligned} F_{MWL}^{(g)}(\boldsymbol{\theta}) &= \log|\boldsymbol{\Sigma}^{(g)}| + \text{tr}[\mathbf{S}^{(g)}(\boldsymbol{\Sigma}^{(g)})^{-1}] - \log|\mathbf{S}^{(g)}| - q \\ &+ [\mathbf{m}^{(g)} - \boldsymbol{\mu}^{(g)}]'(\boldsymbol{\Sigma}^{(g)})^{-1}[\mathbf{m}^{(g)} - \boldsymbol{\mu}^{(g)}]. \end{aligned}$$

## 2.2 Product Indicator Methods

Having defined the discrepancy functions, I next present the specifics of estimating models with latent variable interactions. As the name of this section suggests, the class of these methods referred to in this section requires the formation of product indicators. A distinguishing feature of product indicator methods is that standard structural equation modelling software can, in principle, be used to estimate the parameters. Many authors have worked in this area and the literature is quite substantial, sometimes with conflicting findings. A few important contributions are reviewed here.

As mentioned earlier, Kenny and Judd (1984) considered a model with four product indicators when each of the original latent variables is measured by two indicators. But with more than two indicators for each latent variable, the exhaustive pairing of the original indicators can become cumbersome. For instance, the model in (1.6) would require 9 product indicators. Therefore, one has to consider a structural equation model for the $18 \times 1$ vector $\mathbf{y} = (y_1, y_2, y_3, x_1, x_2, x_3, x_4, x_5, x_6, x_1x_4, x_1x_5, x_1x_6, x_2x_4, x_2x_5, x_2x_6, x_3x_4, x_3x_5, x_3x_6)'$. In response, Jöreskog and Yang (1996) proposed a method that uses only one product indicator. The rationale is that there is a high degree of redundancy in the full set of product indicators. Other researchers suggest the use of non-overlapping pairs of product indicators (e.g. Marsh et al., 2004).

Importantly, even if the original variables are normally distributed themselves, their products are not. This observation has important ramifications on the appropriate choice of estimation method. Estimation by MWL is, in general, inappropriate because it assumes multivariate normality of the joint vector of original indicators and product indicators. As I will show in section 2.3, it is not necessary to have a product latent variable in the model because the product is simply a nonlinear term in the latent variables, which can be handled easily if one does not confine oneself

within the linear structural equation modelling framework. Alternatively, one can choose to utilize the robustness property of maximum likelihood, and argue that even though MWL is not technically correct, it does give reasonable point estimates and sufficiently accurate standard errors (Marsh et al., 2004).

Concerned about the potential inappropriateness of MWL, Jöreskog and Yang (1996) suggested using WLSA as an alternative. WLSA does not have the multivariate normality assumption and it gives asymptotically correct chi-square values and standard errors. This method, however, is complicated and requires the model be set up in a non-standard way. Jöreskog and Yang (1996) commented that a large sample size is needed for the asymptotic distribution-free property of WLSA to be effective because of the sheer size of $\mathbf{W_a}$.

In addition to violating the assumption of multivariate normality, the use of product indicators to define a product latent variable also leads to nonlinear dependence among the parameters. This point is most clearly illustrated by the following observation. Suppose $\xi_1$ and $\xi_2$ have zero means and covariance matrix

$$cov \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} \phi_{11} & \\ \phi_{21} & \phi_{22} \end{pmatrix}.$$

Invoking the normality assumption on the $\xi$'s, it can be shown that the mean of $\xi_1\xi_2$ is equal to $\phi_{21}$, the covariance between $\xi_1$ and $\xi_2$, and that the variance of the product latent variable $\xi_1\xi_2$ is $\phi_{21}^2 + \phi_{11}\phi_{22}$. Therefore one must be able to derive and impose nonlinear equality constraints on the parameters in order to use methods based on product indicators. In addition to the relatively simple nonlinear constraints shown above, there are other far more complicated constraints on the residual variances and measurement intercepts in the full model (1.6). The illustration above also rules out the use of the WLS estimator because the sample mean vector and sample covariance matrix are clearly not independent.

Even if the nonlinear constraints can be imposed, it is important to note that most of these constraints require normality of the latent variables. Equation (1.5) implies that $\eta$ is a linear combination of $\xi_1$ (normal), $\xi_2$ (normal), their product (nonnormal), and $\zeta$ (normal). Thus the assumption of latent (and hence observed) variable normality is not met.

To fix the problem of having far too many and also far too complicated nonlinear constraints, Marsh et al. (2004) proposed an unconstrained approach, wherein the parameters that would have been nonlinear functions of other parameter in the model were simply left freely estimated. After an extensive set of simulations, Marsh and colleagues concluded that their unconstrained MWL approach was not only robust but also much easier to implement. I do not argue against the unconstrained approach based on existing empirical evidence about its usefulness. What I find unappealing about this approach is that it essentially tries to mask important problems by over-parameterizing a model. There is also little theoretical evidence that the good performance of the unconstrained approach will generalize to conditions other than those covered by simulations.

There are many other similar approaches that uses product indicators (see e.g. Cortina, Chen, & Dunlap, 2001 for a survey) and a whole book discussing similar approaches (Schumacker & Marcoulides, 1998). Other recent developments that use the MWL discrepancy function include methods by Ping (1996a, 1996b), Algina and Moulder (2001), and Batista-Foguet, Coenders, and Saris (2004). There is also the 2-stage least squares (T2LS) estimator by Bollen and Paxton (1998). According to a number of studies (Coenders, Batista-Foguet, & Saris, 2006; Moulder & Algina, 2002; Schermelleh-Engel, Klein, & Moosbrugger, 1998), T2LS has the disadvantage of having low power and also being a limited information estimator, even though it does not require multivariate normality. Wall and Amemiya (2001) proposed a partially constrained generalized appended product indicator (GAPI) approach and they

showed that it was much more effective than the traditional constrained procedures for non-normal data. However their new GAPI approach still required complicated nonlinear constraints on the parameters, which made it difficult to implement in applied research.

One issue that has been rarely discussed in the literature on product indicator methods is missing data. When one of the original observed variables contains missing values, all product indicators that involve this variable are affected. Related to the missing data problem is the multiple group analysis that is necessary for three-way models of the kind in Equation (1.7). To my knowledge there has been no systematic investigation, empirical or theoretical, of three-way interactions, whether or not one focuses on product indicators or not. For instance, a salient question relates to the performance of the likelihood ratio test statistic that is central to the three-way interaction hypothesis. Is the statistic sufficiently close to being a central chi-square variable under the null hypothesis? Does the test have enough power to be useful in empirical research? Theoretically, answers to the above questions are all positive, so long as one abandons the product indicator approach for the full-information maximum likelihood method that will be discussed in the next section.

## 2.3 Full Information Maximum Likelihood

The full information maximum likelihood (FIML) approach identifies the parameters directly from raw data, without product indicators. Although technically more complicated when compared with the product indicator methods, FIML provides a far more flexible, coherent, and theoretically justified statistical framework for estimating and testing such nonlinear structural equation models. Furthermore, FIML can accommodate the presence of missing data. The main drawback of FIML is its computational complexity, which is becoming less of an issue with the availability of high speed computers of today.

To describe the FIML estimator in detail, it is useful to make some notational

simplifications based on Equations (1.5), (1.6), and (1.7). For subject $i$ in group $g = 1, \ldots, G$ that consists of $N_g$ subjects, let $\mathbf{y}_i^{(g)}$ be the $9 \times 1$ vector of observed variables, i.e., consisting of observations on $y_1$, $y_2$, $y_3$, $x_1$, $x_2$, $x_3$, $x_4$, $x_5$, and $x_6$ (the left hand side of 1.6). Let $\boldsymbol{\tau}$ be the vector of measurement intercepts in (1.6) and let $\boldsymbol{\Lambda}$ be the factor loading matrix in (1.6). Let $\boldsymbol{\varepsilon}_i^{(g)}$ be the $9 \times 1$ vector of measurement error terms for that subject, whose covariance matrix is $\boldsymbol{\Delta}$, a diagonal matrix. The parameter matrices for the measurement model are $\boldsymbol{\tau}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Delta}$. They are invariant over the groups to provide a stable measurement model for meaningful comparisons at the latent level. For the purpose of identification and user-defined constraints, these parameter matrices are really functions of a parameter vector in matrix-forms $\boldsymbol{\theta} = (\tau_1, \ldots, \tau_6, \lambda_1, \ldots, \lambda_6, \delta_{11}, \ldots, \delta_{99})$, i.e., $\boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{\theta})$, $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\theta})$, $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\theta})$.

The derivation of the FIML estimator begins with the specification of the *conditional model* that defines the distribution of the observed variables given the latent variables. In the case of (1.6), the conditional distribution of $\mathbf{y}_i^{(g)}$ is multivariate normal with mean vector

$$\boldsymbol{\mu}_i^{(g)} = \boldsymbol{\tau} + \boldsymbol{\Lambda} \begin{pmatrix} \eta_i^{(g)} \\ \xi_{1i}^{(g)} \\ \xi_{2i}^{(g)} \end{pmatrix},$$

and conditional covariance matrix $\boldsymbol{\Delta}$. Therefore one can write down the conditional density function as

$$f(\mathbf{y}_i^{(g)} | \eta_i^{(g)}, \xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\theta}) = |2\pi\boldsymbol{\Delta}|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{y}_i^{(g)} - \boldsymbol{\mu}_i^{(g)})' \boldsymbol{\Delta}^{-1} (\mathbf{y}_i^{(g)} - \boldsymbol{\mu}_i^{(g)}) \right\}. \quad (2.7)$$

The next step in the derivation is to realize that conditional on $\xi_{1i}^{(g)}$ and $\xi_{2i}^{(g)}$, $\eta_i^{(g)}$ is also normally distributed, with mean

$$\nu_i^{(g)} = \alpha^{(g)} + \gamma_1^{(g)} \xi_{1i}^{(g)} + \gamma_2^{(g)} \xi_{2i}^{(g)} + \gamma_3^{(g)} \xi_{1i}^{(g)} \xi_{2i}^{(g)},$$

and conditional standard deviation $\sigma^{(g)}$. Let us next stack the following parameters into a vector: $\boldsymbol{\omega}^{(g)} = (\alpha^{(g)}, \gamma_1^{(g)}, \gamma_2^{(g)}, \gamma_3^{(g)}, \sigma^{(g)})'$. The conditional density of $\eta_i^{(g)}$ given $\xi_{1i}^{(g)}$ and $\xi_{2i}^{(g)}$ can be written as

$$f(\eta_i^{(g)}|\xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\omega}^{(g)}) = \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left\{-\frac{1}{2}\left(\frac{\eta_i^{(g)} - \nu_i^{(g)}}{\sigma^{(g)}}\right)^2\right\}. \tag{2.8}$$

Next, the distribution of $\boldsymbol{\xi}_i^{(g)} = (\xi_{1i}^{(g)}, \xi_{2i}^{(g)})'$ must be specified. According to the assumptions in section 1.3, the unconditional distribution of $\boldsymbol{\xi}_i^{(g)}$ is multivariate normal with mean vector $\boldsymbol{\kappa}^{(g)} = (\kappa_1^{(g)}, \kappa_2^{(g)})'$, and covariance matrix

$$\boldsymbol{\Phi}^{(g)} = \begin{pmatrix} \phi_{11}^{(g)} & \\ \phi_{21}^{(g)} & \phi_{22}^{(g)} \end{pmatrix}.$$

Let $\boldsymbol{\varrho}^{(g)} = (\kappa_1^{(g)}, \kappa_2^{(g)}, \phi_{11}^{(g)}, \phi_{21}^{(g)}, \phi_{22}^{(g)})'$ be the vector of parameters that are involved in specifying the distribution of $\boldsymbol{\xi}_i^{(g)}$. The density function of $\boldsymbol{\xi}_i^{(g)}$ is

$$f(\xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\varrho}^{(g)}) = |2\pi\boldsymbol{\Phi}^{(g)}|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\xi}_i^{(g)} - \boldsymbol{\kappa}^{(g)})'(\boldsymbol{\Phi}^{(g)})^{-1}(\boldsymbol{\xi}_i^{(g)} - \boldsymbol{\kappa}^{(g)})\right\}. \tag{2.9}$$

Finally, Equations (2.7), (2.8), and (2.9) are assembled together. The joint distribution of the observed and the latent variables can be written as the product of three density functions

$$\begin{aligned} & f(\mathbf{y}_i^{(g)}, \eta_i^{(g)}, \xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\theta}, \boldsymbol{\omega}^{(g)}, \boldsymbol{\varrho}^{(g)}) \\ = \; & f(\mathbf{y}_i^{(g)}|\eta_i^{(g)}, \xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\theta}) \times f(\eta_i^{(g)}|\xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\omega}^{(g)}) \times f(\xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\varrho}^{(g)}). \end{aligned}$$

However, the latent variables are not observable. Therefore, one must integrate

the latent variables out of the joint distribution to obtain the marginal density of $\mathbf{y}_i^{(g)}$:

$$f(\mathbf{y}_i^{(g)}; \boldsymbol{\theta}, \boldsymbol{\omega}^{(g)}, \boldsymbol{\varrho}^{(g)}) = \tag{2.10}$$
$$\int_{\mathbb{R}^2} \left[ \int_{\mathbb{R}} f(\mathbf{y}_i^{(g)} | \eta_i^{(g)}, \xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\theta}) f(\eta_i^{(g)} | \xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\omega}^{(g)}) d\eta_i^{(g)} \right] f(\xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\varrho}^{(g)}) d\boldsymbol{\xi}_i^{(g)},$$

where the outer integral is over two dimensions: $\xi_{1i}^{(g)}$ and $\xi_{2i}^{(g)}$. This three-fold integral can be reduced to a two-dimensional integral in a fairly straightforward way. Note that the inner integral over $\eta$ essentially corresponds to a normal-normal mixture, conditional on $\boldsymbol{\xi}$. The assumption of $\boldsymbol{\Delta}$ being diagonal as well as the fact that $\boldsymbol{\Lambda}$ is assumed to be block-diagonal (perfect simple structure) in Equation (1.6) can be utilized to solve the inner integral in closed-form. However, the outer integral over $\boldsymbol{\xi}$ is not so easy to simplify. Numerical integration methods must be used instead.

Several features of Equation (2.10) need additional comments. First, the marginal distribution of $\mathbf{y}$ is specified as a continuous mixture density, where the conditional density corresponding to the measurement model is first integrated over $\eta$, conditional on $\boldsymbol{\xi}$, and then integrated over the distribution of $\boldsymbol{\xi}$. Due to the presence of nonlinear terms in the mixture distribution, analytical simplifications of the integrals beyond what are outlined above are difficult. Second, though I wrote the conditional density $f(\mathbf{y}_i^{(g)} | \eta_i^{(g)}, \xi_{1i}^{(g)}, \xi_{2i}^{(g)}; \boldsymbol{\theta})$ for the full vector of observed variables, missing data in $\mathbf{y}$ pose no significant difficulty. In that case, one simply takes the standard approach in full-information estimation by using all available information in $\mathbf{y}$. Third, the current derivation easily permits extensions to the case of categorical observed variables (or mixture of continuous-categorical variables). One changes the form of the conditional density associated with the measurement model to an appropriate one according to the type of the observed variables. This point will not be pursued further, but it is in principle a solvable problem.

At this point, the likelihood function becomes straightforward to derive. The

marginal likelihood for one observation is

$$L(\boldsymbol{\theta}, \boldsymbol{\omega}^{(g)}, \boldsymbol{\varrho}^{(g)} | \mathbf{y}_i^{(g)}) = f(\mathbf{y}_i^{(g)}; \boldsymbol{\theta}, \boldsymbol{\omega}^{(g)}, \boldsymbol{\varrho}^{(g)}). \tag{2.11}$$

Let the full vector of parameters be $\boldsymbol{\vartheta} = (\boldsymbol{\theta}, \boldsymbol{\omega}^{(1)}, \ldots, \boldsymbol{\omega}^{(G)}, \boldsymbol{\varrho}^{(1)}, \ldots, \boldsymbol{\varrho}^{(G)})'$. Let $\mathbf{Y}^{(g)}$ be the matrix of observations from group $g$. The log-likelihood for group $g$ is

$$\log L(\boldsymbol{\theta}, \boldsymbol{\omega}^{(g)}, \boldsymbol{\varrho}^{(g)} | \mathbf{Y}^{(g)}) = \sum_{i=1}^{N_g} \log L(\boldsymbol{\theta}, \boldsymbol{\omega}^{(g)}, \boldsymbol{\varrho}^{(g)} | \mathbf{y}_i^{(g)}). \tag{2.12}$$

The log-likelihood for the full sample is a sum of the contributions from each group

$$\log L(\boldsymbol{\vartheta} | \mathbf{Y}) = \sum_{g=1}^{G} \log L(\boldsymbol{\theta}, \boldsymbol{\omega}^{(g)}, \boldsymbol{\varrho}^{(g)} | \mathbf{Y}^{(g)}), \tag{2.13}$$

where $\mathbf{Y} = (\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(G)})$.

Numerical maximization of $\log L(\boldsymbol{\vartheta} | \mathbf{Y})$ leads to the full information maximum likelihood estimate $\hat{\boldsymbol{\vartheta}}$ of all parameters. At the same time, the inverse of the negative of the second derivative matrix of $\log L(\boldsymbol{\vartheta} | \mathbf{Y})$, evaluated at the MLE, gives the large sample covariance matrix of $\hat{\boldsymbol{\vartheta}}$. The square root of the diagonal elements of the covariance matrix of $\hat{\boldsymbol{\vartheta}}$ provides estimated standard errors for all parameters. Finally, the maximized value of the log-likelihood function itself is obtained as a by-product with which likelihood ratio tests can be conducted.

## 2.4 Implementations of Full Information Methods

The implementation of FIML estimation can take various forms in practice. Five important implementations are reviewed here. They are called implementations rather than estimators because in principle, they all lead to the same set of estimates, which is the MLE.

The latent moderated structural equations LMS is the first approach without

using product indicators or non-linear constraints on the parameters (Klein & Moosbrugger, 2000; Schermelleh-Engel et al., 1998). The original derivation by Klein and Moosbrugger (2000) considered a slightly more general structural equation model with quadratic effects. A distinguishing feature of the LMS approach is that it is explicitly based on approximating a continuous mixture distribution by a finite mixture of normal densities. Referring back to the derivations in section 2.3, this is equivalent to using numerical integration methods for approximating the integrals in Equation (2.10). An EM algorithm is developed by Klein and Moosbrugger (2000) to obtain the parameter estimates. Currently there is no commercially available software for LMS.

Klein (2007) later developed the Quasi Maximum Likelihood (QML) estimation method, which is computationally less intensive than LMS. This approach is available in the form of free software QML (Klein, 2007). The QML method provides approximate maximum likelihood estimates of model parameters. The QML software itself can handle multiple latent interaction and quadratic effects. However, the software is a prototype version that has limitations. The number of observed indicators is limited to 10; the number of latent exogenous variables ($\xi$'s) is limited to 4; only one latent endogenous variable ($\eta$) is permitted; and the maximum sample size is limited to 2000. QML does not support multi-sample analysis, nor equality constraints on the parameters that are both crucial to the three-way interaction model that I consider here.

The availability of flexible software for nonlinear mixed-effect modelling in SAS (SAS Institute Inc., 2004) provides a more "brute-force" approach. That is, one first converts the latent variable model into an equivalent nonlinear mixed model, and then estimate the parameters using PROC NLMIXED. There are two reasons why it is a brute-force method. First, because of the generality of NLMIXED, it does not have hard-coded derivatives for the log-likelihood. Numerical derivatives (finite difference) are used instead. This results in many more likelihood function evaluations in

the estimation procedure and a dramatic decrease in efficiency that takes a long time for the program to converge, assuming that it converges at all. Second, unlike some of the other more specialized methods that will be discussed next, NLMIXED does not take advantage of the possibility of further analytical simplifications to the integrals in Equation (2.10). It never reduces the dimensions of integration and always performs strictly three-dimensional integration for this model. Even with modern adaptive numerical quadrature methods implemented in NLMIXED, the total run time is often several hours when compared with more specialized methods which takes just a few seconds. Furthermore, the user must be able to write down the likelihood function (2.11) completely as NLMIXED programming statements before one can start fitting the model. This may be a dauntingly high requirement for a typical researcher, which in turn made this approach nearly inaccessible.

Recently Cudeck et al. (in press) proposed a new method that is specifically targeted at reducing the dimension of integration for latent variable interaction models. With an ingenious conditioning argument often found in the nonlinear regression literature, they were able to simplify the three dimensional integral to a one dimensional integral. Their development is based on the assumption that all indicators are continuous, and they provide SAS/IML programs that implement the method. However, these programs are not easy to use and require significant modifications before they can be adapted to fit models a researcher would want with three-way interactions. Though somewhat limited due to software distribution, it is a truly original and promising method for investigating nonlinear effects in latent variable models.

Finally, the implementation that I have chosen to pursue is the one currently implemented in Mplus (Muthén & Muthén, 1998–2007). The Mplus command language provides a convenient way to specify product interactions via the XWITH keyword. The program seems to be efficient. Because this feature is embedded within a general statistical modelling package that supports multi-sample analysis, models with

three-way interactions can be tested. However, few details are given in the software manual or technical white papers. Given the fact that Mplus reports the use of adaptive numerical quadrature for approximating two-dimensional integrals (as opposed to three-dimensional) in the model-fitting process, I suspect that a dimension reduction technique like the one that was outlined in section 2.3 was employed. In addition, I noticed that when the outcome variable becomes categorical, Mplus turns back to three-dimensional numerical integration. This is in accordance with what was discussed earlier, dimension reduction by direct solution of the inner integral in Equation (2.10) is only possible when the outcome variables are continuous (conditionally normal to be exact).

### 2.5 Multiple Groups as Mixtures with Known Class Membership

Due to current restrictions in Mplus, multiple group analysis cannot be specified in conjunction with maximum likelihood estimation. However, I have developed a work around through the use of the mixture modelling option. Maximizing the multiple group log-likelihood in (2.13) is equivalent to maximizing a mixture model log-likelihood with known class membership.

To show this, consider the following mixture density for observed variables $\mathbf{y}$:

$$f(\mathbf{y};\boldsymbol{\vartheta}) = \sum_{g=1}^{G} \pi^{(g)} f^{(g)}(\mathbf{y};\boldsymbol{\vartheta}), \tag{2.14}$$

where $\pi^{(g)}$ is the mixing probability and $f^{(g)}(\mathbf{y};\boldsymbol{\vartheta})$ is the within-class density. The fact that $\boldsymbol{\vartheta}$ is shared across classes means that cross-class (i.e., cross-group) constraints are permitted. With unknown class membership, the mixing probabilities must be estimated. However, if the class membership is known, as is the case here, the vector $(\pi^{(1)}, \ldots, \pi^{(G)})$ becomes a Bernoulli vector with zeroes everywhere except on the

location for the class to which the observations in $\mathbf{y}$ belong. This is equivalent to

$$f(\mathbf{y}_i^{(g)}; \boldsymbol{\vartheta}) = f^{(g)}(\mathbf{y}_i^{(g)}; \boldsymbol{\vartheta}),$$

where $\mathbf{y}_i^{(g)}$ denotes observation $i$ from group $g$, so that the $\pi^{(g)}$'s that are zero drop out of the mixture density (details on mixture modeling can be found in Bauer & Curran, 2004). In Mplus, this approach is implemented with the KNOWNCLASS option for mixture models. It is important to note that this is currently the only practical method available for estimating the model discussed above.

### 2.6 Non-normal Latent Variables

Based on the theory developed in Chapters 1 and 2, FIML performs well with normally distributed latent variables (Lee & Zhu, 2002). However, its performance when the exogenous latent variables are non-normal is still unknown. Existing knowledge is mostly on the robustness of product indicator methods. For example, Marsh et al. (2004) compared four methods using product indicators, the constrained MWL approach, the unconstrained MWL approach, the GAPI method (Wall & Amemiya, 2001), and QML (Klein, 2007). They pointed out that all four approaches were relatively unbiased for normally distributed latent variables with large sample sizes. When the latent variables were non-normal, the unconstrained and GAPI approaches were more robust.

Three general research questions arise. First, the quality of the FIML estimates under both normality and non-normality must be studied. Second, the performance of FIML, both in terms of parameter recovery and accuracy of the chi-square difference test, should be compared with product indicator based methods. Third, to show that FIML is applicable in substantive research, one must investigate its statistical power, preferably also relative to product indicator methods. To investigate these questions, a simulation study is necessary. I chose the unconstrained maximum

likelihood approach by Marsh et al. (2004) (hereinafter UML) as representative of the product indicator approaches because it is the easiest for an applied researcher to implement and is among the best performing methods in existing simulation studies.

# CHAPTER 3

# Methods

## 3.1 Simulation Design

I designed the simulation study to achieve two goals. The first is parameter recovery and accuracy of standard error estimates under different conditions. The second is the behavior of the chi-square difference statistic for testing the three-way interaction hypothesis.

Particular emphasis is placed on finite sample performance. Maximum likelihood performs optimally under large sample, but for realistic sample size in psychological research, the large sample results depend also on the degree of complexity of the model under investigation (van der Vaart, 2000). It may be the case that a much larger sample size than is usually needed would be required for the asymptotic results to hold in the three-way interaction model. It may also be the case that the power of the proposed likelihood ratio test may be too low at the typical sample size encountered in real data analysis where the thee-way model may be used.

In this simulation study, four design factors are manipulated: estimator (FIML and UML), distribution of latent variables (normal and non-normal), sample size ($N = 150$ and $N = 500$ per group) and three-way interaction effect (present and absent). To retain focus, I only consider models with two groups (labelled as group 0 and group 1 hereinafter). However, the extension to more than two groups is straightforward.

Crossing the two estimators (FIML and UML) with two conditions of latent variables (normal and non-normal), two conditions of three-way effect (absent and present), and the two sample sizes (150 and 500 per group) results in 16 simulation conditions.

### 3.2 Generating Model

Under all conditions, for either group 0 or group 1, the generating measurement model is

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1.4 \\ 0 \\ 1 \\ 1.4 \\ 0 \\ 1 \\ 1.4 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0.8 & 0 & 0 \\ 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0.8 \\ 0 & 0 & 0.5 \end{pmatrix} \begin{pmatrix} \eta \\ \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{pmatrix}, \tag{3.1}$$

where $\varepsilon_1 \sim \mathcal{N}(0, 0.7)$, $\varepsilon_2 \sim \mathcal{N}(0, 0.8)$, $\varepsilon_3 \sim \mathcal{N}(0, 1.5)$, $\varepsilon_4 \sim \mathcal{N}(0, 0.7)$, $\varepsilon_5 \sim \mathcal{N}(0, 0.8)$, $\varepsilon_6 \sim \mathcal{N}(0, 1.5)$, $\varepsilon_7 \sim \mathcal{N}(0, 0.7)$, $\varepsilon_8 \sim \mathcal{N}(0, 0.8)$, $\varepsilon_9 \sim \mathcal{N}(0, 1.5)$.

The structural equation for group 0 is

$$\eta = -1.5 + 0.7\xi_1 + 0.5\xi_2 + \zeta, \tag{3.2}$$

where $\zeta \sim \mathcal{N}(0, 0.6)$. The interaction coefficient is set to zero in this group so that the product term drops out. In group 1, when data are generated under the null hypothesis (i.e., absence of three-way interaction effect), the structural equation is exactly identical as Equation (3.2). Under the alternative hypothesis (i.e., when there

is three-way effect), the structural equation for group 1 is

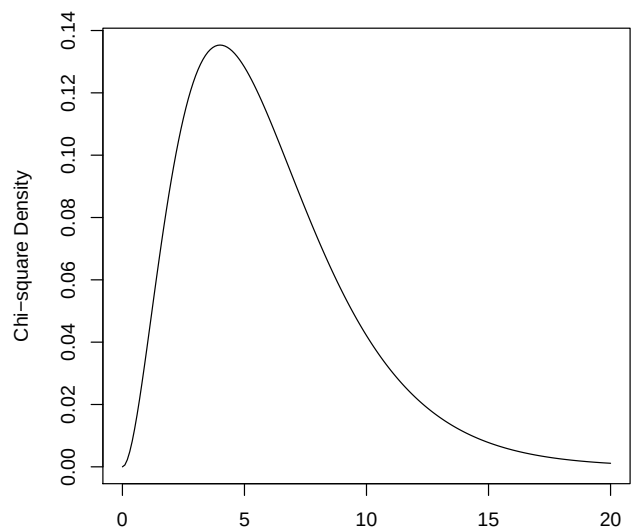
$$\eta = -1.5 + 0.7\xi_1 + 0.5\xi_2 + 0.125\xi_1\xi_2 + \zeta, \tag{3.3}$$

where $\zeta$ still has mean zero and error variance .6. In all conditions, $\xi_1$ and $\xi_2$ are uncorrelated, with $\xi_1$ having a mean of 2 and variance of 1.5, and $\xi_2$ having a mean of 1 and variance of .7. The squared multiple correlation for the structural equation (3.3) is .70. The effect of $\xi_1$, $\xi_2$ and the interaction each account for 14.95%, 2.37% and 0.79% of the variance of $\eta$, as measured by squared semi-partial correlations. These effect sizes are commonly seen in applied research. Considering the fact that the equation disturbance variance is equal to 0.6, the magnitude of the three-way effect (i.e., the difference in the $\gamma_3$'s between group 0 and 1) corresponds to a small effect size ($d = 0.16$), when expressed in terms of Cohen's (1992) $d$. Keeping the three-way effect small is important because it will enable the simulation to provide rough guidelines for the appropriate sample size in applied research.

The data were simulated in SAS (SAS Institute Inc., 2004). When the exogenous latent variables $\xi_1$ and $\xi_2$ were normally distributed, the SAS RANNOR function was used. Under the non-normal conditions, $\xi_1$ and $\xi_2$ were generated as independent central chi-square variates each with 6 degrees-of-freedom, just as Marsh et al. (2004) did in one of their studies. They argued that this chi-square distribution is a reasonable representation of skewed data and has been used in Wall and Amemiya (2001)'s simulations for the GAPI approach too.

Specifically, the RANGAM function was used to generate the chi-square variates. The random numbers were then standardized by subtracting their mean (6 for chi-square with 6 degrees-of-freedom) and dividing by their standard deviation ($\sqrt{12}$) to produce standardized variates that are subsequently transformed to have the same means and variances as the normal conditions. Figure 3.1 illustrates the density func-

Figure 3.1: Density Function of $\chi^2_6$ Distribution




tion of a chi-square distribution with 6 degrees-of-freedom. Note that its skewness is equal to 1.15 and its kurtosis is 2.

The nine indicators ($x_1$ to $x_6$ and $y_1$ to $y_3$) were then generated with non-normal latent variables and normal residuals. It is important to note that even though the latent variable were distributed non-normally, the marginal distributions of some of the indicators were only mildly non-normal. More will be said about this later. To test the univariate normality of the nine indicators in group 0 where there is no interaction effect and all the non-normality of the indicators comes from the non-normal latent variables, I conducted the Shapiro-Wilk test (Shapiro & Wilk, 1965) for one simulated data set. Univariate normality was rejected for all indicators in group 0 both at $N = 500$ and $N = 150$ ($p < .001$) except for $x_3$, $x_6$, $y_2$ and $y_3$. For example, $x_3$ only has a skewness of 0.10 and excess kurtosis of 0.03. As a result, the distributions of those indicators with non-significant univariate normality test results may look very similar to a normal distribution. However, univariate normality alone does

not imply multivariate normality. To test the multivariate normality of the indicator vector, I conducted Mardia's multivariate skewness and kurtosis tests (Mardia et al., 1979). The results indicated that at $N = 500$ the multivariate skewness was 3.69 and the kurtosis was 104.38, both being highly significant ($p < .001$). Even under small sample size at $N = 150$, multivariate normality was still rejected with highly significant multivariate skewness of 10.39 and kurtosis of 109.38 ($p < .001$).

### 3.3 Methods for Summarizing Results

In the evaluation of potential biases of estimation, both the absolute bias and the relative bias of the parameter estimates are examined. Let $M$ be the number of replications in a condition. Absolute bias (or raw bias) of point estimates, defined as the Monte Carlo average of the point estimates minus the true parameter value,

$$B(\theta) = M^{-1} \sum_{i=1}^{M} (\hat{\theta}_i - \theta),$$

can tell the difference between the true parameter value and the mean of parameter estimates across replications. Relative bias of point estimates, defined as raw bias divided by the true parameter value (when the ratio is well-defined),

$$B_r(\theta) = M^{-1} \sum_{i=1}^{M} \frac{(\hat{\theta}_i - \theta)}{\theta},$$

is the proportion of absolute bias relative to the true parameter value.

To evaluate the standard errors, the mean of the estimated standard errors for that parameter across the replications should be compared with the Monte Carlo standard deviation of a given parameter estimate. Let

$$SE(\hat{\theta}) = M^{-1} \sum_{i=1}^{M} SE(\hat{\theta}_i),$$

be the mean of the estimated standard errors, where $SE(\hat{\theta}_i)$ is the estimated standard error from replication $i$, and let

$$SD(\hat{\theta}) = \left( \frac{1}{M-1} \sum_{i=1}^{M} (\hat{\theta}_i - \bar{\hat{\theta}}) \right)^{1/2}$$

be the Monte Carlo standard deviation of the point estimates, where $\bar{\hat{\theta}}$ is the mean of point estimates. As a comparable measure of accuracy of the standard error estimates, the relative bias of the standard errors is also reported. It is defined as:

$$B_r(SE) = \frac{SE(\hat{\theta}) - SD(\hat{\theta})}{SD(\hat{\theta})}.$$

Root Mean Square Error (RMSE), which can provide information of both the distance of each parameter estimate from the true value and the variability of such distances is also calculated. For a generic parameter $\theta$, RMSE is defined as

$$RMSE = \left( \frac{1}{M} \sum_{i=1}^{M} (\hat{\theta}_i - \theta)^2 \right)^{1/2},$$

where $\hat{\theta}_i$ is the estimate from replication $i$, $\theta$ is the true value.

# CHAPTER 4

# Results

## 4.1 Convergence and Proper Solutions

After the raw data were generated in SAS and saved, all model fitting was conducted in Mplus Version 5 (Muthén & Muthén, 1998–2007). For FIML, 15 quadrature points per dimension of integration was used for all conditions. Due to non-convergence and improper solution problems using UML, 800 to 1000 replications per condition were conducted to ensure that there were at least 500 converged and proper solutions (i.e., no negative variances or non positive definite covariance matrices) in each condition. Default starting values were used.

FIML converged with proper solutions for all replications under all conditions, while UML had significant convergence issues. Such problems were more severe when the sample size is small ($N = 150$). The results for the rates of convergence and proper solutions are summarized in Table 4.1. To make the results directly comparable across the two estimators, the first 500 replications in each condition where both FIML and UML properly converged were used for all subsequent comparisons of parameter estimates, standard errors, and likelihood ratio test statistics. In other words, in any given condition, the results for FIML and UML were based on precisely the same data set.

The results in Table 4.1 is striking on its own. For example, in the worst case scenario, UML had only 57% proper solution rate when there is no interaction effect

Table 4.1: Rates of Convergence and Proper Model Estimation

| | | | Converged Solutions | | | | Converged and Proper Solutions | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Restricted | | Unrestricted | | Restricted | | Unrestricted | | Both | |
| | $(\gamma_3^{(0)}, \gamma_3^{(1)})$ | $N$ | UML | FIML | UML | FIML | UML | FIML | UML | FIML | UML | FIML |
| Normal | $(0, 0)$ | 150 | 76% | 100% | 77% | 100% | 63% | 100% | 64% | 100% | 57% | 100% |
| | $(0, 0)$ | 500 | 92% | 100% | 90% | 100% | 82% | 100% | 80% | 100% | 77% | 100% |
| | $(0, 0.125)$ | 150 | 79% | 100% | 81% | 100% | 65% | 100% | 68% | 100% | 60% | 100% |
| | $(0, 0.125)$ | 500 | 93% | 100% | 95% | 100% | 85% | 100% | 88% | 100% | 82% | 100% |
| Non-normal | $(0, 0)$ | 150 | 77% | 100% | 77% | 100% | 64% | 100% | 66% | 100% | 58% | 100% |
| | $(0, 0)$ | 500 | 90% | 100% | 89% | 100% | 80% | 100% | 78% | 100% | 73% | 100% |
| | $(0, 0.125)$ | 150 | 77% | 100% | 80% | 100% | 63% | 100% | 66% | 100% | 57% | 100% |
| | $(0, 0.125)$ | 500 | 93% | 100% | 95% | 100% | 85% | 100% | 87% | 100% | 82% | 100% |

*Notes*. 1. UML = Unconstrained Maximum Likelihood; FIML = Full Information Maximum Likelihood

2. Number of total replications varies from 800 to 1000 across cells.

3. For restricted model $\gamma_3^{(0)} = \gamma_3^{(1)}$; for unrestricted model $\gamma_3^{(0)} \neq \gamma_3^{(1)}$

under normal condition at $N = 150$. This is in sharp contrast with the optimistic results reported by Marsh et al. (2004), wherein UML converged properly in 88.8% to 100% of the cases. Despite software difference (LISREL was used in their paper), one plausible explanation is that Marsh et al. only considered one group. Another plausible explanation is that the generating model used in the present investigation had a much weaker interaction effect. The proportion of variance explained by the interaction term in Marsh et al.'s (2004) model is over 5 times larger than the effect size in the current simulation.

To further investigate the high rate of non-convergence and improper solutions observed in the simulations, I conducted logistic regression analysis of the rate of converged and proper solutions for UML (represented in Table 4.1 in the second column but last), using the design factors (sample size, interaction effect, and latent variable normality) as the predictors. The overall likelihood ratio chi-square for all regression coefficients against the intercept only model is $\chi^2(df = 7) = 385.23$, $p < .0001$. The results showed that the single most important predictor for the rate of convergence is sample size. For instance, the odds of non-convergence is 3.38 times higher ($\beta = 1.22$, $p < .001$) for $N = 150$ than $N = 500$. The results also showed that the rate of non-convergence can be partly attributed to the size of the interaction effect. When the interaction effect is not present, the odds of non-convergence is 1.39 times higher ($\beta = 0.33$, $p < .01$) than the conditions where the interaction effect is present. All subsequent analysis are based on 500 properly converged solutions for both FIML and UML.

### 4.2 Likelihood Ratio Test of Interaction Effect

To test the significance of the three-way interaction effect, a constrained model was fitted to each data set, where $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ were freely estimated but constrained to be equal. Next, an unconstrained model was fitted to the same data, where $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ were not constrained to be equal. This was done for both FIML and UML.

Then the three-way interaction effect was tested with the likelihood ratio test statistic in reference to a chi-square distribution with 1 degree-of-freedom (i.e. a chi-square change test).

Table 4.2 summarizes the results for the likelihood ratio test as well as additional model fit information for UML. Note that because FIML identifies the parameters directly from the raw data and a nonlinear model is fitted, there does not exist a natural saturated model for FIML. Therefore, the chi-square test of the model is undefined for FIML. In contrast, UML identifies the parameters from mean vectors and covariance matrices, so a saturated model is clearly defined. For the restricted model, the degrees-of-freedom for UML is 147, and for the unrestricted model, the degrees-of-freedom for UML is 146.

Under all null conditions, i.e., where $\gamma_3^{(1)} = 0$, the means of the estimated model fit chi-squares are all sightly larger than the degrees-of-freedom. For example, under normal condition at $N = 150$, the means of the chi-squares for the restricted and unrestricted models are 155.93 and 154.72 respectively, both being larger than the degrees-of-freedom (147 and 146 respectively). The standard deviations for the two means are 22.14 and 22.04 respectively, and both are larger than the expected value, which are equal to the square root of 2 times the degree-of-freedom (17.15 and 17.09 for restricted and unrestricted models, respectively).

Turning attention to the likelihood ratio statistic for the comparison of the restricted and the unrestricted models, it can be seen from the last two columns in Table 4.2 that under the null hypothesis, both UML and FIML provide essentially unbiased chi-square statistics, at least when latent variables are normal. For example, at $N = 150$, the mean of the FIML chi-squares (0.95) is close to the the expected value of 1. In the same condition, the mean of the UML chi-square (1.21) is similarly close. Under all normal and non-normal null conditions, the means of FIML chi-squares are consistently closer to 1 than the UML chi-squares. The exception

Table 4.2: Likelihood Ratio Tests Within and Between Restricted and Unrestricted Models

| | | | Restricted Model $\chi^2$ (*SD*) | | Unrestricted Model $\chi^2$ (*SD*) | | $\chi^2_\Delta$ (*SD*) between Restricted and Unrestricted Model | |
|---|---|---|---|---|---|---|---|---|
| | $(\gamma_3^{(0)}, \gamma_3^{(1)})$ | $N$ | UML | FIML | UML | FIML | UML | FIML |
| Normal | $(0, 0)$ | 150 | 155.93 (22.14) | – | 154.72 (22.04) | – | 1.21 (1.53) | 0.95 (1.35) |
| | $(0, 0)$ | 500 | 155.94 (23.13) | – | 154.89 (23.01) | – | 1.05 (1.60) | 1.12 (1.64) |
| | $(0, 0.125)$ | 150 | 165.36 (21.69) | – | 163.32 (21.72) | – | 2.03 (2.79) | 7.87 (5.55) |
| | $(0, 0.125)$ | 500 | 178.47 (24.96) | – | 175.38 (25.13) | – | 3.09 (3.81) | 24.23 (9.55) |
| Non-normal | $(0, 0)$ | 150 | 165.02 (25.23) | – | 163.80 (25.20) | – | 1.21 (1.64) | 1.16 (1.59) |
| | $(0, 0)$ | 500 | 165.79 (24.89) | – | 164.58 (24.81) | – | 1.21 (1.68) | 1.10 (1.62) |
| | $(0, 0.125)$ | 150 | 174.29 (26.78) | – | 171.87 (26.26) | – | 2.42 (3.41) | 7.83 (5.59) |
| | $(0, 0.125)$ | 500 | 189.53 (28.34) | – | 186.19 (27.87) | – | 3.33 (4.17) | 24.11 (9.48) |

*Notes*. 1. UML = Unconstrained Maximum Likelihood; FIML = Full Information Maximum Likelihood

2. Results are based on the first 500 properly converged solutions for both UML and FIML.

3. For UML, degrees-of-freedom is 147 for the restricted model and 146 for the unrestricted model.

4. for FIML, the within model $\chi^2$ and the degrees-of-freedom are undefined.

is normal data and $N = 500$, where the mean of FIML chi-square (1.12) is slightly further from the expected value of 1 than the mean of UML chi-square (1.05). This may be due to the fact that more quadrature points are needed as $N$ increases so that the error of integral approximation does not contaminate the calculation of the marginal log-likelihood. Recall that there is one integral per subject and the marginal log-likelihood is accumulated over the number of subjects. Minor inaccuracies for individual integral approximations may become amplified as $N$ increases.

When the latent variables are non-normal, both UML and FIML lead to biased chi-square values, but the bias is more pronounced with UML. For example, at $N = 150$ and $N = 500$, the FIML chi-squares are 1.16 and 1.10 respectively, which are closer to 1 than the UML chi-squares (1.21 in both sample sizes).

Under the alternative hypothesis, the mean and variance of chi-squares from FIML are much larger than those of UML. This suggests that UML has low power to detect the interaction effect than FIML. This point is clearly illustrated in Table 4.3, where the empirical Type I error rates and power estimates for the likelihood ratio test of the restrictions on $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ are tabulated. Table 4.3 shows that both UML and FIML have calibrated Type I error rates when latent variables are normal. FIML's type I error rates are closer to the alpha level than UML. For example, at the traditional alpha level of .05, the type I error rate was .06 for UML and was .05 for FIML. When latent variables are non-normal, both methods have slightly elevated Type I error rates. For example, the type I error rates for UML and FIML are .07 and .08 respectively at $N = 150$, and are .08 and .07 at at $N = 500$ respectively.

In Table 4.3 under the alternative hypothesis, the equality constraint rejection rate represents power to detect the interaction effect. For both normal and non-normal conditions, UML has substantially lower power than FIML. The difference in the power between UML and FIML are striking for all conditions. For example, at alpha level of .05 when $N = 500$, UML has a power of .28 under normal conditions

Table 4.3: Empirical Estimates of Type I Error Rates and Statistical Power for Test of Equality Constraints on $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$

| | | | Rejection Rate at .01 Level | | Rejection Rate at .05 Level | | Rejection Rate at .10 Level | |
|---|---|---|---|---|---|---|---|---|
| | $(\gamma_3^{(0)}, \gamma_3^{(1)})$ | $N$ | UML | FIML | UML | FIML | UML | FIML |
| Normal | $(0, 0)$ | 150 | .02 | .01 | .06 | .05 | .13 | .09 |
| | $(0, 0)$ | 500 | .02 | .02 | .06 | .05 | .11 | .11 |
| | $(0, 0.125)$ | 150 | .06 | .51 | .18 | .76 | .26 | .84 |
| | $(0, 0.125)$ | 500 | .14 | .99 | .28 | 1.00 | .39 | 1.00 |
| Non-normal | $(0, 0)$ | 150 | .02 | .01 | .07 | .08 | .14 | .13 |
| | $(0, 0)$ | 500 | .02 | .02 | .08 | .07 | .15 | .12 |
| | $(0, 0.125)$ | 150 | .10 | .52 | .21 | .73 | .30 | .80 |
| | $(0, 0.125)$ | 500 | .16 | .99 | .31 | 1.00 | .42 | 1.00 |

*Notes*. 1. UML = Unconstrained Maximum Likelihood; FIML = Full Information Maximum Likelihood

2. Results are based on the first 500 properly converged solutions for both UML and FIML.

3. Rejection rates represent Type I error for $(0, 0)$ conditions and statistical power for $(0, 0.125)$ conditions.

and a power of .31 under non-normal conditions, whereas FIML's power is 1 under both normal and non-normal conditions. Note that when $N = 150$ under normal condition, FIML had a power of .76 to detect the interaction effect at the conventional alpha level of .05. In the same condition, UML's power is .18. This is a practically significant difference and implies that researchers who use FIML are more than four times more likely to find a significant effect at this sample size.

It is interesting to see in Table 4.3 that UML's power under non-normal conditions is higher than its power under normal conditions. This is probably due to the higher type I error rate at non-normal conditions than normal conditions. In comparison, there is not much change in power for FIML when latent variables go from normal to non-normal . At alpha level of .05, FIML's power is 1 at $N = 500$ in both normal and non-normal conditions. Even at $N = 150$, FIML's power is still maintained above .73.

### 4.3 Parameter Recovery and Standard Errors

In the evaluation of potential biases of estimation, I will concentrate on parameter estimates and standard errors of the interaction effect $\gamma_3$, because all other parameters are set to be equal across the two groups. Table 4.4 shows the means and Monte Carlo standard deviations of the point estimates of $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ for both UML and FIML in the first two columns. The raw bias, relative bias and RMSE of the point estimates are also reported.

Several features in Table 4.4 are immediately evident. First, FIML recovers parameters better than UML, and in many cases, much better. The last two columns in Table 4.4 present values of RMSE which provide aggregated information on parameter recovery of $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ for both UML and FIML. The results indicate that there is not a single condition where UML has a smaller RMSE than FIML. In several conditions, the RMSE of UML is over 4 times the RMSE of FIML.

Second, UML's variability is much larger than FIML. Again, the Monte Carlo standard deviations of FIML are smaller than those of UML in all conditions. For

Table 4.4: Mean, Standard Deviation (*SD*), and Bias of Estimated $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$

| | | | Mean and (*SD*) of $\begin{pmatrix}\gamma_3^{(0)}\\ \gamma_3^{(1)}\end{pmatrix}$ | | Raw Bias of $\begin{pmatrix}\gamma_3^{(0)}\\ \gamma_3^{(1)}\end{pmatrix}$ | | Relative Bias of $\begin{pmatrix}\gamma_3^{(0)}\\ \gamma_3^{(1)}\end{pmatrix}$ | | RMSE of $\begin{pmatrix}\gamma_3^{(0)}\\ \gamma_3^{(1)}\end{pmatrix}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $(\gamma_3^{(0)}, \gamma_3^{(1)})$ | $N$ | UML | FIML | UML | FIML | UML | FIML | UML | FIML |
| Normal | $(0, 0)$ | 150 | -.01 (.21) | -.01 (.10) | -.01 | -.01 | – | – | .21 | .10 |
| | | | .00 (.28) | -.01 (.10) | .00 | -.01 | – | – | .28 | .10 |
| | $(0, 0)$ | 500 | -.01 (.09) | -.01 (.05) | -.01 | -.01 | – | – | .09 | .05 |
| | | | -.00 (.09) | -.01 (.05) | -.00 | -.01 | – | – | .09 | .05 |
| | $(0, 0.125)$ | 150 | .00 (.49) | .01 (.10) | .00 | .01 | – | – | .49 | .10 |
| | | | .19 (.58) | .13 (.10) | .07 | .01 | .55 | .07 | .58 | .10 |
| | $(0, 0.125)$ | 500 | -.02 (.10) | -.00 (.05) | -.02 | -.00 | – | – | .10 | .05 |
| | | | .15 (.12) | .12 (.05) | .03 | -.00 | .23 | -.01 | .12 | .05 |
| Non-normal | $(0, 0)$ | 150 | .01 (.47) | .01 (.11) | .01 | .01 | – | – | .47 | .11 |
| | | | .04 (.34) | .01 (.11) | .04 | .01 | – | – | .34 | .11 |
| | $(0, 0)$ | 500 | .00 (.10) | .01 (.06) | .00 | .01 | – | – | .10 | .06 |
| | | | .00 (.10) | .01 (.06) | .00 | .01 | – | – | .10 | .06 |
| | $(0, 0.125)$ | 150 | .03 (.40) | .02 (.11) | .03 | .02 | – | – | .39 | .11 |
| | | | .18 (.48) | .14 (.11) | .06 | .01 | .46 | .11 | .48 | .11 |
| | $(0, 0.125)$ | 500 | -.02 (.41) | .02 (.06) | -.02 | .02 | – | – | .41 | .06 |
| | | | .18 (.44) | .14 (.06) | .05 | .02 | .41 | .15 | .44 | .06 |

*Notes*. 1. UML = Unconstrained Maximum Likelihood; FIML = Full Information Maximum Likelihood

2. Results are based on the first 500 properly converged solutions for both UML and FIML.

3. Relative bias cannot be calculated for all conditions because true values in such cases are zero.

Table 4.5: Mean and Standard Deviation (*SD*) of Estimated Standard Errors (*SE*) and Empirical Rejection Rates of Univariate $z$ Tests of Estimated $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$

| | | | Mean and (*SD*) of Estimated *SE* of $\begin{pmatrix}\gamma_3^{(0)}\\ \gamma_3^{(1)}\end{pmatrix}$ | | *SD* of Estimated $\begin{pmatrix}\gamma_3^{(0)}\\ \gamma_3^{(1)}\end{pmatrix}$ | | Relative Bias of Estimated *SE* of $\begin{pmatrix}\gamma_3^{(0)}\\ \gamma_3^{(1)}\end{pmatrix}$ | | Rejection Rate at .01 Level for $\begin{pmatrix}\gamma_3^{(0)}\\ \gamma_3^{(1)}\end{pmatrix}$ | | Rejection Rate at .05 Level for $\begin{pmatrix}\gamma_3^{(0)}\\ \gamma_3^{(1)}\end{pmatrix}$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $(\gamma_3^{(0)}, \gamma_3^{(1)})$ | $N$ | UML | FIML | UML | FIML | UML | FIML | UML | FIML | UML | FIML |
| Normal | $(0,0)$ | 150 | .20 (.12) | .09 (.02) | .21 | .10 | -0.05 | -0.10 | .00 | .02 | .02 | .08 |
| | | | .20 (.16) | .09 (.02) | .28 | .10 | -0.29 | -0.10 | .01 | .01 | .04 | .06 |
| | $(0,0)$ | 500 | .09 (.03) | .05 (.00) | .09 | .05 | 0.00 | 0.00 | .01 | .01 | .04 | .06 |
| | | | .09 (.03) | .05 (.00) | .09 | .05 | 0.00 | 0.00 | .00 | .01 | .03 | .04 |
| | $(0,0.125)$ | 150 | .29 (.76) | .09 (.02) | .49 | .10 | -0.41 | -0.10 | .01 | .01 | .05 | .07 |
| | | | .31 (.91) | .09 (.02) | .58 | .10 | -0.47 | -0.10 | .02 | .16 | .10 | .34 |
| | $(0,0.125)$ | 500 | .09 (.03) | .05 (.00) | .10 | .05 | -0.10 | 0.00 | .00 | .01 | .04 | .04 |
| | | | .10 (.03) | .05 (.00) | .12 | .05 | -0.17 | 0.00 | .10 | .47 | .27 | .73 |
| Non-normal | $(0,0)$ | 150 | .30 (.91) | .09 (.02) | .47 | .11 | -0.36 | -0.18 | .00 | .01 | .01 | .07 |
| | | | .28 (.84) | .09 (.02) | .34 | .11 | -0.18 | -0.18 | .00 | .02 | .03 | .10 |
| | $(0,0)$ | 500 | .09 (.03) | .05 (.01) | .10 | .06 | -0.10 | -0.17 | .01 | .03 | .05 | .11 |
| | | | .09 (.03) | .05 (.01) | .10 | .06 | -0.10 | -0.17 | .01 | .03 | .03 | .09 |
| | $(0,0.125)$ | 150 | .31 (1.01) | .09 (.02) | .40 | .11 | -0.23 | -0.18 | .00 | .02 | .05 | .08 |
| | | | .33 (.98) | .09 (.02) | .48 | .11 | -0.31 | -0.18 | .04 | .16 | .12 | .34 |
| | $(0,0.125)$ | 500 | .12 (.43) | .05 (.01) | .41 | .06 | -0.71 | -0.17 | .01 | .01 | .04 | .09 |
| | | | .13 (.44) | .05 (.01) | .44 | .06 | -0.70 | -0.17 | .13 | .65 | .33 | .80 |

*Notes*. 1. UML = Unconstrained Maximum Likelihood; FIML = Full Information Maximum Likelihood

2. Results are based on the first 500 properly converged solutions for both UML and FIML.

example, under the alternative conditions with normal latent variable, the standard deviations for $\gamma_3^{(1)}$ using UML are .58 and .12 at $N = 150$ and $N = 500$ respectively, whereas using FIML, they are only .10 and .05 at $N = 150$ and $N = 500$ respectively. Under the non-normal conditions, the standard deviations for $\gamma_3^{(1)}$ using UML (.48 and .44 at $N = 150$ and $N = 500$ respectively) are still much larger than the standard deviations using FIML (.11 and .06 at $N = 150$ and $N = 500$ respectively).

Third, the bias of FIML is also uniformly smaller than UML. For example, for UML estimation under multivariate normality, the interaction effect $\gamma_3^{(1)}$ was overestimated by 55% at $N$=150 and was also overestimated by 23% at $N$=500. In comparison, for FIML under the same multivariate normality condition, the interaction effect was only overestimated by 7% at $N$=150 and was underestimated by 1% at $N$=500. Under non-normal conditions, the interaction effect in UML was overestimated by 46% and 41% at $N$=150 and $N$=500 respectively, whereas in FIML, it was only overestimated by 11% and 15% at $N$=150 and $N$=500 respectively. Thus it can be safely concluded that FIML provides superior point estimates.

Table 4.5 examines another aspect of parameter recovery, namely, the accuracy of the standard error estimates. For a standard error estimate to be valid, its mean across Monte Carlo replications must be close enough to the empirically observed Monte Carlo standard deviation of the point estimates. The first and second columns in Table 4.5 shows the means and standard deviations of the standard error estimates for both UML and FIML. These entries should be compared with the Monte Carlo standard deviations reported in columns 3 and 4. It is clear that overall under both normal and non-normal conditions, FIML standard errors are closer to the expected Monte Carlo standard deviations. For example, for UML under the normal alternative conditions at $N$=150, the mean of the standard error estimates for $\gamma_3^{(1)}$ across Monte Carlo replications was .31, whereas the empirically observed Monte Carlo standard deviation of the point estimates is .58. In comparison, under the

same conditions at $N$=150, the difference between the mean of the standard error estimates and the observed standard deviation of the point estimates is much smaller for FIML, whose mean of the standard error estimates for $\gamma_3^{(1)}$ is .09, and the corresponding observed standard deviation of the point estimates is .10. At large sample size $N$=500, such difference is not that big as in the small sample size for UML, in which the mean of the standard error estimates for $\gamma_3^{(1)}$ is .10 and is not so far from the empirically observed Monte Carlo standard deviation of the point estimate (.12). In comparison for FIML at $N$=500, the mean of the standard error estimates and the observed standard deviation of the point estimates are both .05.

Table 4.5 also presents the relative bias of the standard error estimates for $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$. Under normal alternative conditions, FIML standard errors have a slight downward bias when $N = 150$ (10% underestimated for $\gamma_3^{(1)}$), but the bias goes away completely at $N = 500$. The downward bias of UML is more pronounced (47% underestimated for $\gamma_3^{(1)}$) and does not go away completely at $N = 500$ (17% underestimated for $\gamma_3^{(1)}$). For the non-normal alternative conditions, both UML and FIML have biased standard errors, but the bias of FIML is much smaller. For example, at $N = 150$, standard errors for $\gamma_3^{(1)}$ in UML was underestimated by 31%, but in FIML, it was only underestimated by 18%. At $N = 500$, standard errors for $\gamma_3^{(1)}$ was underestimated by 70% in UML, but was still only underestimated by 17% in FIML.

The observed rejection rates of the univariate $z$ tests of $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ are reported in the final columns in Table 4.5. Under normal conditions at a traditional alpha level of .05, UML could detect the effect of $\gamma_3^{(1)}$ only 10% of the time at $N = 150$ and 27% of the time at $N = 500$, whereas FIML could detect the effect 34% of the time at $N = 150$ and 73% of the time at $N = 500$. Under non-normal conditions at the traditional alpha level of .05, UML could detect the effect of $\gamma_3^{(1)}$ 12% of the time at $N = 150$ and 33% of the time at $N = 500$, whereas FIML could still detect the effect 34% of the time at $N = 150$ and 80% of the time at $N = 500$.

These results are in accordance with the empirical estimates of statistical power for test of equality constraints on $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ represented in Table 4.3. Under both the normal and non-normal conditions at alpha level of .05, power increases as sample size increases. For UML at $N = 150$ and $N = 500$, power is higher for non-normal conditions (.21 and .31 respectively) than normal conditions (.18 and .28 respectively). For FIML at $N = 150$ and $N = 500$, power is .76 and 1 respectively for normal conditions and is .73 and 1 respectively for non-normal conditions. Similar trend can be found in the observed rejection rates of the univariate $z$ test of $\gamma_3^{(1)}$ in Table 4.5. Overall, FIML has higher power than UML under both normal and normal conditions, and this is verified by the observed rejection rates of the univariate $z$ tests of $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ reported in Table 4.5.

The empirical rejection rates for univariate $z$ test of estimated $\gamma_3^{(1)}$ under the alternative conditions are based on the Wald test. Comparing the power of the Wald test represented in Table 4.5 and the power of the likelihood ratio test shown in Table 4.3, one can find the difference in the powers of the two tests. Specifically, the power of the Wald test is lower than the likelihood ratio test. Such difference is due to the testing of a different null hypothesis. The likelihood ratio test is evaluating a null hypothesis that $\gamma_3^{(0)}$ and $\gamma_3^{(1)}$ are equal, where as the Wald test is evaluating that each individual coefficient is equal to zero.

# CHAPTER 5

# Discussion

The goal of this research is to propose a new estimation method for three-way interactions with latent variables based on full information maximum likelihood. Specifically, the three-way interactions under this study refer to the two-way interaction of latent variables with multiple groups. There has been little discussion about models with three-way interactions in the literature and this model aims to fill that gap. The FIML likelihood is derived and it is shown that the model can be fitted in commercially available software using mixture modelling. The performance of FIML is compared with the UML method (Marsh et al., 2004) in a simulation study focusing on chi-squaredness of the likelihood ratio test of interaction effect, parameter recovery and standard error estimation, and violations of the assumption of latent variable normality. In the conditions covered by the simulation, FIML outperformed UML in virtually all meaningful aspects of the comparison. Several notable features emerged from the present research and will be addressed in the remainder of this chapter.

## 5.1 On Model Assumptions

To date, most of the existing work on parameter estimation for models with latent variable interactions has focused on methods involving product indicators. These traditional methods attempt to identify the parameters of a (nonlinear) model from the means and covariance matrix alone. UML is one of the most successful candidates

among this class of methods. The key benefit of UML, from the standpoint of its proponents, is its simplicity. Any standard structural equation modelling software can be used to implement UML and very little programming is involved. In contrast with earlier product indicator based methods such as those studied by Jöreskog and Yang (1996), UML does not require complex nonlinear restrictions be placed on parameters. Marsh et al. (2004) argued that the nonlinear restrictions are based on the assumption of latent variable normality, and by not imposing the restrictions, "there are no such assumptions" in the UML approach (p. 295). The first part of their statement is correct, but the second part is not.

The derivations presented earlier in Chapter 2 show that as long as the MWL discrepancy function is used for estimating the parameters, the joint vector of observed variables and product indicators has to be multivariate normal, regardless of whether the interaction-induced nonlinear restrictions are imposed or not. Multivariate normality of the indicators, which is at the heart of the MWL theory, does not distinguish between product indicators and directly observed indicators. By definition, the product indicators cannot even be univariately normal, and thus multivariate normality collapses. One has to appeal to the robustness properties of the MWL estimator to justify the use of MWL under such assumption violations.

Current asymptotic robustness theory (e.g. Amemiya & Anderson, 1990) on MWL for mean and covariance structure analysis suggests that for linear structural models, the MWL is robust (in large samples) under latent variable non-normality, with the important qualification that the measurement errors (unique factors) must be normally distributed. Thus, the technical conditions set forth in asymptotic robustness theory depend both on the observed data and on characteristics of the model that are hard to verify. Existing robustness results that are based on asymptotic arguments and derived under linear structural equations do not shed light on the extent to which the model in UML is robust against assumption violations.

Essentially, the UML approach is attempting to identify nonlinear relations in the data that are only identifiable from the raw data points by forcing a linear model upon the means and covariances. In this regard, UML is doubly misspecified. The criticism in Marsh et al. (2004) against Jöreskog and Yang's (1996) use of nonlinear restrictions is therefore misguided. One cannot cover up one kind of potential misspecification (exogenous latent variable non-normality) by introducing another kind of misspecification (dropping nonlinear restrictions on the parameters) whose impact is even harder to understand than the first kind of misspecification. Furthermore, leaving the nonlinear restrictions free leads to an over-parameterized model, i.e., there are more parameters than is needed to describe the data. This is fundamentally against the principle of parsimony in model building. The (consequently weak) argument about UML's ease of implementation is therefore unfounded. As Jöreskog and Yang (1996) pointed out, the nonlinear restrictions help define the nonlinear model, so the they are part of a model. Potential violations of the latent normality assumption must be addressed, but it must be addressed directly, as opposed to indirectly as in Marsh et al. (2004) by masking it with over-fitting.

This leads to a general question. When an existing modelling framework becomes incompatible with the data generating process, should we change the data so that they fit into our familiar modelling framework, or should we change our modelling framework? Latent variable interaction leads to nonlinearity. To model it using standard linear tools, product indicators must be formed, so data must be changed in the first place. Data can be changed in an infinite number of different ways. This perhaps helps explain the abundance of seemingly contradictory findings in the product indicator literature. Every change (no matter how small it may seem) in the formation of product indicators, or the parameterization of the restrictions, or the choice of discrepancy functions, leads to potentially important differences in the quality of statistical inference regarding the interaction terms. At the same time,

important contributions by Klein and Moosbrugger (2000) have been pushed to the sideline due to the lack of software support and the lack of understanding of the full generality of nonlinear structural equation models, of which the latent variable interaction model is a special case.

As is shown in section 2.4, FIML can be implemented easily without fitting a misspecified model or violating the multivariate normality assumption of the indicators. The fact that FIML still requires that the exogenous latent variables, measurement errors and the equation disturbance terms be normal is a software restriction rather than a restriction due to the lack of theoretical support. As noted earlier, FIML can support arbitrary mixtures of measurement models by simply replacing the conditionally normal measurement model density in Equation (2.7) with, say, a logistic measurement model for Bernoulli data, or a cumulative logistic model for ordinal data. To handle latent non-normality, one can conveniently replace the multivariate normality assumption of the the disturbance terms mentioned above with multivariate $t$ assumption, or as Woods and Thissen (2006) showed in a slightly different context, latent non-normality can be captured semi-parametrically using spline densities or any other semi-parametric curve systems. These changes to FIML do not alter the basic set up of the modelling framework and that is: a latent nonlinear structural equation model can be directly identified from the raw data by forming a marginal likelihood function and maximizing it over just as many parameters as it is necessary to describe the nonlinearity. The resultant solution is asymptotically unbiased and most efficient, and is based on the celebrated likelihood principle.

The only traditional approach that is "statistically correct" is the WLSA method Jöreskog and Yang (1996) proposed. It is a true weighted least squares method that does not require multivariate normality, and it gives asymptotically unbiased estimates. However, it is still attempting to identify the parameters from the sample moments, and it requires a very large sample size to be stable (Jöreskog & Yang,

1996). To give due credit to Marsh et al. (2004), WLSA is extremely difficult to implement in practice, perhaps more so than FIML, thus leaving FIML the only practical and theoretically justified candidate for use in applied research.

### 5.2 On Relative Performance

Given the fact that the present implementation of FIML still requires latent exogenous variable normality, its performance must be checked in comparison with UML under both normality and non-normality. Via a simulation study, it is shown that FIML significantly outperforms UML which has the best performance to-date in simulations. In both normal and non-normal latent variable conditions, not only was FIML able to recover the parameters more accurately, more honestly maintain Type I error rates, but it also had much higher power than UML.

As expected from standard maximum likelihood theory, FIML's performance becomes better with larger sample size, when normality of latent exogenous variables is satisfied. Under normal latent variable conditions, it has a perfect proper convergence record in the current study. The type I error rates are maintained at the traditional .05 nominal alpha level. Its power for detecting the interaction effect is estimated to be .76 at $N = 150$ (for alpha level of .05), and close to 1 at $N = 500$. It only overestimates the coefficients of the interaction effect by 7% at $N = 150$, and is almost unbiased at $N = 500$. In comparison, under the same normal latent variable conditions, UML's performance was unacceptable. It converged 57% to 82% of the time with proper solutions; it has inflated type I error rate and low power (.18 to .28); it overestimates the regression coefficients for the interaction effect by 23% to 55%.

Under non-normal latent variable conditions, bias resulted from assumption violations, which is expected, but unexpectedly UML performed worse than FIML, despite claims made by Marsh et al. (2004) about its insensitivity to non-normality. Specifically, FIML still converged 100% with proper solutions, but UML only converged 57% to 82% of the time with proper solutions. Both FIML and UML have

inflated type I error rates, but UML has much lower power (.21 and .31 at $N = 150$ and $N = 500$ respectively) than FIML (.73 and 1 at $N = 150$ and $N = 500$ respectively) at the traditional alpha level of .05 for detecting the interaction effect. Regarding the bias in the point estimates of the interaction effect, both FIML and UML are biased. However, UML's bias is much more severe (overestimated by 46% at $N = 150$ and 41% at $N = 500$) when compared with FIML (overestimated by 11% at $N = 150$ and 15% $N = 500$ respectively).

In retrospect, the lackluster performance of UML may not be completely unexpected. Marsh et al. (2004) only considered UML in one group, and the interaction effect in their simulation study (.2 when interaction effect is present) is much higher than the present study (.125 for the interaction effect in group 1). It is reasonable to expect that UML's performance appears better in their study when the interaction effect is more pronounced and when there is no multiple groups involved. Above all, UML involves a misspecified model. Convergence problems become more problematic as one increases the complexity of a model already misspecified. In this case, adding more groups clearly increases the complexity, so that the previous problems such as low rate of convergence become more serious.

### 5.3 Limitations and Future Research

As with all simulation studies, my findings only directly generalize to the conditions studied here, and these may not fully represent those commonly found in applied research settings. I will thus describe the potential limitations, attempt to address the extent to which they threaten the external validity of the current study, and outline directions for future research.

In the current study, the indicators were generated with non-normal common factors and normally distributed unique factors, resulting in some univariate marginal distributions appearing to be only mildly non-normal, despite the fact that the vector of indicators is still jointly multivariate non-normal. Under the current set of condi-

tions, FIML performed better than UML, but its success requires further investigation when both the common factors and the unique factors are more severely non-normal. However, under a high degree of non-normality, perhaps the more fruitful avenue to take is to utilize the inherent flexibility of the FIML method and implement it with semi-parametric latent variable densities such as the ones discussed by Woods and Thissen (2006).

Another limitation in the current study is that the indicators are all continuous. In an applied setting, the observed variables could be of mixed types, such as binary or ordinal. This points to some more directions for future research. As mentioned in section 2.3, FIML can be readily extended to the case of categorical observed indicators, or where there is mixed type indicators. Much is already known in standard structural equation modelling for non-normal indicators, but relatively little has been done for non-linear models. The proposed method laid out in this research should be helpful in those future extensions.

A further limitation in the current design is that the covariance between the two latent exogenous variables is set to zero in the generating model. The fitted model does contain a parameter for the covariance, but it is rare in applied settings to have two predictors completely uncorrelated. Future simulations are needed to investigate whether the FIML estimator would perform differently when the covariance is non-zero.

In the current generating model, all parameters except the interaction effect in the second group are equal across the two groups. This also may be an unrealistic condition in some areas of research. The unconditional effects of the two latent variables could be different across groups. When more parameters are allowed to be different across groups, the model becomes more complicated, but more realistic and more generalizable at the same time. Future research will reveal how much more complicated a model FIML could handle.

Marsh et al. (2004) investigated much larger interaction effects than the current study. They reported satisfactory rate of proper solutions (above 94%) as well as power for UML. The small effect size in the current study is specifically chosen so as to ensure a fair comparison between FIML and UML. As discussed in Chapter 4, FIML's power sometimes approaches 1.0 even with a small effect size. Choosing a larger effect size could lead to no variability for FIML results, i.e. power consistently equal to 1.0. In future simulations, effect size and sample size should be chosen carefully to obtain a more detailed comparison of UML and FIML. The combination should not result in lack of variability for one or both estimators.

In an applied research setting, one can easily extend the non-linear structural equation model demonstrated here by adding more covariates. In the current model, the set of covariates only contain latent variables. It would be advantageous in applied research to add more observed predictor variables into the structural equation. Furthermore, the observed predictor variables can also potentially interact with other observed or latent predictor variables. The resulting model would be far more complex and realistic. FIML is ideally suited for such extensions because the addition of more predictors does not change the scheme of parameter estimation already laid out. The product indicator methods on the other hand, would have a more difficult time handling more observed predictor variables in the structural model.

Despite these potential limitations, these simulation results provide a new and unique insight into the promising method of FIML for estimating latent variable interactions. Further, the implementation of FIML in applied research can be demonstrated in future explorations with appropriate data sets where the model could be more flexible and realistic as shown above.

## References


Aguinis, H. (2002). Estimation of interaction effects in organization studies. *Organizational Research Methods*, *5*, 207–211.

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

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

Amemiya, Y., & Anderson, T. W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *Annals of Statistics*, *18*, 1453–1463.

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

Batista-Foguet, J. M., Coenders, G., & Saris, W. E. (2004). A parsimonious approach to interaction effects in structural equation models: An application to consumer behavior. *Working Papers d'ESADE*, *183*, 1–28.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent latent variable models: Potential problems and promising opportunities. *Psychological Methods*, *9*, 3–29.

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

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

Bollen, K. A., & Paxton, P. (1998). Two-stage least squares estimation of interaction effects. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling* (p. 125-151). Mahwah, NJ: Erlbaum.

Coenders, G., Batista-Foguet, J. M., & Saris, W. E. (2006). Simple, efficient and distribution-free approach to interaction effects in complex structural equation models. *Quality & Quantity*, *?*, ?-?

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*, 426-443.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

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

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

Enders, C. K. (2003). Using the EM algorithm to estimate coefficient alpha for scales with item level missing data. *Psychological Methods*, *8*, 322–337.

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

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

Klein, A. G. (2007). QuasiML (Version 3.10) [Computer software]. Champaign, IL: Author

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

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

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. New York: Academic Press.

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

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

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Ping, R. A. (1996a). Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psychological Bulletin*, *119*, 166-175.

Ping, R. A. (1996b). Latent variable regression: A technique for estimating interaction and quadratic coefficients. *Multivariate Behavioral Research*, *31*, 95-120.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

SAS Institute Inc. (2004). *SAS/STAT® 9.1 users guide*. Cary, NC: SAS Institute Inc.

Schermelleh-Engel, K., Klein, A. G., & Moosbrugger, H. (1998). Estimating nonlinear effects using a latent moderated structural equations approach. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling* (p. 203-238). Mahwah, NJ: Erlbaum.

Schumacker, R. E., & Marcoulides, G. A. (Eds.). (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah, NJ: Erlbaum.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591–611.

van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge, UK: Cambridge University Press.

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

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, *9*, 60–62.

Woods, C. M., & Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika*, *71*, 281–301.

Yang Jonsson, F. (1998). Modeling interaction and nonlinear effects: A step-by-step LISREL example. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling* (p. 17-42). Mahwah, NJ: Erlbaum.