ROBUSTNESS IN CONFIRMATORY FACTOR ANALYSIS:
THE EFFECT OF SAMPLE SIZE, DEGREE OF NON-NORMALITY, MODEL, AND ESTIMATION METHOD ON ACCURACY OF ESTIMATION FOR STANDARD ERRORS

By

YOUNGKYOUNG MIN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008

To my Mom, Taesun Jeong
for her endless love, support, encouragement, and prayer

ACKNOWLEDGMENTS

First of all, I would like to thank God and my mother for every moment of my life and for providing the opportunity to gain this tremendous education.

I am thankful to Dr. James Algina and Dr. David Miller for giving me the opportunity to study in Research and Evaluation Methodology Program, and for supporting me to finish my doctoral study. I owe much gratitude to the faculty of Department of Educational Psychology for his/her excellent lecture and thoughtful consideration.

The completion of this dissertation would not be possible without the guidance of my dissertation committee. Each has helped me throughout my dissertation in his/her own ways. I especially wish to thank Dr. James Algina, my supervisory committee chair and advisor, for his precious time, professional mentoring, and valuable guidance during the last four years. I also wish to thank Dr. David Miller for his consideration and useful advice which made my confidence firm in the entire process. I would like to thank Dr. Cynthia Garvan for her constant encouragement and challenging suggestions which made me stronger academically. She was also the committee chair of my thesis for M. S. in Statistics and encouraged me to pursue Ph. D. I also want to thank Dr. Walter Leite for his useful information on references and valuable suggestions which made this dissertation rich.

I wish to give many thanks to Dr. Timothy Anderson, Associate Dean at the College of Engineering, for giving me the opportunity to work at Engineering Education Center throughout my doctoral study, and for letting me obtain valuable research experiences in engineering college education field. I wish to thank Erica Hughes for her friendly tips and technical support.

I would like to thank Dr. Mark Shermis, Chair of Department of Educational Psychology, for making the departmental process smooth and giving me practical tips for graduation. Lastly, I want to express my gratitude to Elaine Green and Linda Parsons, for their kind support and help on the official process for a graduate student during my doctoral study.

“Give Thanks to the LORD, for he is good; his love endures forever.”

Psalms 107:1

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS 4

LIST OF TABLES 8

LIST OF ABBREVIATIONS 11

ABSTRACT 12

CHAPTER

1 INTRODUCTION AND LITERATURE REVIEW 14

Introduction 14
Overview of Overall Robustness Studies in SEM 14
Covariance Structure Models and Estimation Methods in SEM 16
Robustness against Non-normal Distribution 24
Skewness and Kurtosis in Univariate Distribution 25
Skewness and Kurtosis in Multivariate Distributions 27
Robustness against Small Sample Size 29
Research Summary Characteristics 30
Findings from Overall Past Robustness Studies 32
Direction for the Present Study 33

2 DESIGN OF MONTE CARLO SIMULATION STUDY 45

Data Generation 45
Fleishman's Power Transformation Method 45
Vale and Maurelli Data Generation Method 46
Estimation Methods 47
Sample Sizes 47
Distributional Characteristics 48
Model Characteristics 48
Analytical Model Characteristics 49
Number of Replications 50
Research Summary Statistics 50
Ways to Summarize and Present Results 51

3 MONTE CARLO STUDY RESULTS 53

Preliminary ANOVA Tests 53
Welch-James Tests and Eta Squared Statistics 55
Effects on Parameter Estimates 59
Effects on Standard Error Ratio Estimates 60
Practical Problems in Estimation 67

4 SUMMARY AND CONCLUSIONS 85

Comparison of Findings with Previous Studies 85
Parameter Estimates 85
Standard Error Estimates 87
Brief Summary 90
Concluding Remarks 91

APPENDIX: STANDARD ERROR RESULTS FOR FACTOR CORRELATIONS: SCALE SET BY SPECIFYING FACTOR VARIANCES EQUAL TO ONE. 93

LIST OF REFERENCES 97

BIOGRAPHICAL SKETCH 103

LIST OF TABLES

Table page

1-1 Overview of previous robustness studies against sample size and non-normality 39

1-2 Estimation methods investigated in the past robustness studies......................................40

1-3 When are ML, GLS, and ADF equivalent? 41

1-4 Degree of non-normality examined in the past robustness studies...................................42

1-5 Range of sample size examined in the past robustness studies 43

1-6 Number of replications in the past robustness studies...44

3-1 Frequency of significance *F* tests on parameter estimates per 16 combinations of distribution and sample size: factor variances equal to one..70

3-2 Frequency of significant *F* tests on parameter estimates per 16 combinations of distribution and sample size: factor loadings equal to one. 71

3-3 Mean parameter estimate for λ = 0.6: Model 3F3I, GLS, $N = 200$, factor variances equal to one, and (0, -1.15). 72

3-4 Frequency of significant *F* tests on standard error estimates per 16 combinations of distribution and sample size: factor variances equal to one...72

3-5 Frequency of significant *F* tests on standard error estimates per 16 combinations of distribution and sample size: factor loadings equal to one. 73

3-6 Mean standard error estimates for ϕ = 0.3: model 6F3I, Robust ML, $N = 1200$, factor variances equal to one, and (2, 6). 73

3-7 *F* statistics (degrees of freedom) for WJ tests for effects on parameter estimates: scale set by specifying factor variances equal to one. 74

3-8 Eta squared for significant effects on parameter estimates: scale set by specifying factor variances equal to one..75

3-9 Marginal means by estimation for estimates of $\lambda = 0.8$ 75

3-10 Mean estimates of residual variance parameters by parameter value, estimation method, and sample size. 75

3-11 *F* statistics (degrees of freedom) for WJ tests for effects on parameter estimates: scale set by specifying factor loadings equal to one. 76

3-12 Eta squared for significant effects on parameter estimates: scale set by specifying factor loadings equal to one. 77

3-13 Grand mean estimates of parameter by parameter value 77

3-14 *F* statistics (degrees of freedom) for WJ tests for effects on standard error ratio estimates: scale set by specifying factor variances equal to one. 78

3-15 Eta squared for significant effects on standard error ratio estimates: scale set by specifying factor variances equal to one. 79

3-16 Marginal means by sample size for standard error ratio estimates of $\lambda = 0.6$ 79

3-17 Mean estimates of standard error ratio for factor loading parameters by parameter value, estimation method, and distribution 80

3-18 Mean estimates of standard error ratio for factor correlation parameters, ϕ (0.3) by model, sample size, and distribution. 80

3-19 Means estimates of standard error ratio for factor correlation parameter, ϕ (0.3) by estimation and distribution. 80

3-20 Marginal means by sample size for standard error ratio estimates of $\theta_\delta = 0.36$ 81

3-21 Means estimates of standard error ratio for residual variance parameter, θ_δ (0.36) by model and distribution. 81

3-22 Means estimates of standard error ratio for residual variance parameters by parameter value, estimation method, and distribution. 81

3-23 *F* statistics (degrees of freedom) for WJ tests for effects on standard error ratio estimates: scale set by specifying factor loadings equal to one. 82

3-24 Eta squared for significant effects on standard error ratio estimates: scale set by specifying factor loadings equal to one. 83

3-25 Mean estimates of standard error ratio for factor loading parameters by parameter value, estimation method, and distribution. 83

3-26 Marginal means by sample size for standard error ratio estimates of ϕ (i, j) = 0.108 83

3-27 Mean estimates of standard error ratio for factor correlation parameters by parameter value, estimation method, and distribution. 84

3-28 Frequency of non-convergence (NC), improper estimates (IE), and non-positive definite Hessians (NP) 84

A-1 Mean standard errors, empirical standard errors, and standard error ratios by model, distribution, and sample size for ML estimation. 93

A-2 Mean standard errors, empirical standard errors, and standard error ratios by model, distribution, and sample size for GLS estimation. 94

A-3 Mean standard errors, empirical standard errors, and standard error ratios by model, distribution, and sample size for Robust ML estimation. 95

A-4 Mean standard errors, empirical standard errors, and standard error ratios by model, distribution, and sample size for Robust GLS estimation....96

LIST OF ABBREVIATIONS

ML	Maximum Likelihood
GLS	Generalized Least Squares
ULS	Unweighted Least Squares
DWLS	Diagonally Weighted Least Squares
ADF	Asymptotically Distribution Free
RML	Robust Maximum Likelihood / Robust ML
RGLS	Robust Generalized Least Squares / Robust GLS
λ	Factor Loading
ϕ	Factor Correlation
ϕ(i, i)	Factor Variance
ϕ(i, j)	Factor Covariance / Factor inter-correlation
θ_δ	Residual Variance
SEM	Structural Equation Modeling
3F3I	3 Factors with 3 Indicators per factor
3F6I	3 Factors with 6 Indicators per factor
6F3I	6 Factors with 3 Indicators per factor
(0, 0)	Standard Normal Distribution with zero skewness and zero kurtosis
(0, -1.15)	Non-normal Distribution (platykurtic) with zero skewness and equal negative kurtosis (-1.15)
(0, 3)	Non-normal Distribution (leptokurtic) with zero skewness and equal positive kurtosis (3)
(2, 6)	Non-normal Distribution with high equal skewness (2) and high equal kurtosis (6)

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ROBUSTNESS IN CONFIRMATORY FACTOR ANALYSIS:
THE EFFECT OF SAMPLE SIZE, DEGREE OF NON-NORMALITY, MODEL, AND
ESTIMATION METHOD ON ACCURACY OF ESTIMATION FOR STANDARD ERRORS

By

Youngkyoung Min

August 2008

Chair: James Algina
Major: Research and Evaluation Methodology

A Monte Carlo approach was employed to examine the effect of model type, sample size, and characteristic of distribution on the Maximum Likelihood (ML), Generalized Least Squares (GLS), Robust ML, and Robust GLS estimates of parameters' standard errors. The LISREL program was used for estimation, and the population covariance matrix and data were generated by using the SAS program. For each of four estimation methods (ML, GLS, Robust ML, and Robust GLS) the behavior of standard error ratio estimates was examined under each combination of four distributions, four sample sizes (200, 400, 800, and 1200), three CFA models, and two scale-setting methods (set by specifying factor variances equal to one and factor loadings equal to one). In addition, the bias of the parameter estimation procedures was investigated. The effects of four factors (estimation method, distribution, model, and sample size) on parameter estimates and standard error estimates were examined within each scale-setting method with Welch-James test and eta squared.

Results for parameter estimates indicate that ML estimates were almost unbiased at all sample sizes and ML estimation had less bias than GLS estimation, although the differences were trivial for factor loadings and factor correlations. Sample size played a more critical role in

GLS estimation than in ML estimation of residual variance and, as a result, larger between-method differences in bias were observed for estimates of residual variance. When the scale was set by specifying factor loadings equal to one, there were no important effects of the factors on the factor loading, factor variance, or factor covariance estimates.

Results for standard error estimates indicate that Robust ML estimates were superior to the non-robust estimates in the bias of the standard error estimates for the non-normal distributions, and the standard error estimates were underestimated for the distribution with positive kurtosis and overestimated for the distribution with negative kurtosis.

From the results, it can be concluded that ML estimation method should be adopted for a normal distribution regardless of sample size, model, and scale-setting method to obtain less biased estimates of parameters and standard errors, and Robust ML should be used for non-normal distributions to improve estimation of standard errors. However, Robust ML estimation works very well even for a normal distribution and some cases better than GLS. It was also found that robust estimation generally worked better than non-robust estimation for the non-normal distributions regardless of the sample size and the model type. When the distribution is non-normal, Robust GLS generally performs well, although Robust ML has less bias than Robust GLS.

CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

Introduction

Structural equation modeling (SEM) has been one of the important statistical approaches in the social sciences because it can be applied to non-experimental and experimental data and offers a convenient way to differentiate between observed variables and latent variables. SEM is often conducted under the assumption of multivariate normality, which implies that all univariate distributions are normal, the joint distribution of any pair of the variables is bivariate normal, and all bivariate scatter-plots are linear and homoscedastic (Kline, 2005). However, real data in social science are rarely normal. The violation of normality assumption can cause incorrect results in SEM. Checking the degree of non-normality of the data and the use of the methods that are less reliant on normality are necessary. Over the past years, the effects of non-normality have been studied and different types of solutions have been proposed. Specifically, it was recognized that the model, the estimation method, and data characteristics such as sample size play critical roles under non-normality. Therefore, the present study used simulation methods to investigate the impact of sample size on the accuracy with which standard errors are estimated under varying degrees of non-normality, different estimation methods, and different structural equation models which include latent variables. In the present study, the effects of model complexity, sample size, estimation method, scale-setting method, and degree of non-normality on the accuracy of estimated standard errors were investigated.

Overview of Overall Robustness Studies in SEM

As the term has come to be used in statistics, robustness refers to the ability to withstand violations of theoretical assumptions (Boomsma, 1983; Harlow, 1985). A common procedure for studying robustness is to generate data sets and examine the behavior of research summary characteristics such as test-statistics, standard errors, etc. This procedure is known as the Monte Carlo simulation method and has been used in most of previous robustness studies in SEM (e.g., Anderson & Gerbing, 1984; Bearden et al., 1982; Boomsma, 1983; Browne, 1984; Muthen & Kaplan, 1985, 1992).

The issues of robustness against small sample size, distributional violations, analysis of correlation matrices, model misspecification, and nonlinear structural equations under different estimation methods have been investigated over the past decade. Hoogland (1999) set forth the following five robustness questions:

- Do the maximum likelihood (ML) and generalized least squares (GLS) estimators possess the asymptotic statistical properties predicted under normal theory when variables are non-normally distributed?
- Does a small sample size cause problems, because the statistical properties of the parameter estimators, standard error estimators and goodness-of-fit test statistics are asymptotic properties?
- Does analysis of correlation rather than covariance matrices cause problems when a model is not scale invariant?
- Does model misspecification cause the incorrect results of analyses?
- Does ignoring the nonlinear relationships between latent variables cause the wrong solution?

The problems of analyzing correlation matrices, model misspecification, and nonlinear structure were excluded in the present study because covariance matrices were used, and the structural equation models were specified to be linear.

A review of past robustness studies was necessary to set up the purpose of the current study and the design of the Monte Carlo study. Table 1-1 shows an overview of a collection of robustness studies for the effects of sample size and non-normality. (The numbers in the "Paper Number" column will be used in Tables 1-2, 1-4, 1-5, and 1-6 to identify the studies listed in Table 1-1.) Of the 47 studies, a total of 41 studies investigated the effect of sample size on robustness and a total of 34 studied the effects of non-normality on robustness in SEM. Other robustness issues (the problem of analyzing correlation matrices, model misspecification, and non-linear structure) were excluded from Table 1-1 and the present study.

Covariance Structure Models and Estimation Methods in SEM

The fundamental hypothesis in covariance structure modeling is

$$\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{\theta}),$$

where $\mathbf{\Sigma}$ $(k \times k)$ is the population covariance matrix of k observed variables, $\mathbf{\Sigma}(\mathbf{\theta})$ $(k \times k)$ is the population covariance matrix of k observed variables written as a hypothesized function of $\mathbf{\theta}$, and $\mathbf{\theta}$ $(t \times 1)$ is a vector of the model parameters. The sample estimator of the population covariance matrix $\mathbf{\Sigma}$ in a sample of size N is

$$\mathbf{S} = \mathbf{Z}'\mathbf{Z}/(N-1),$$

where $\mathbf{Z}$ is an $(N \times k)$ matrix of deviation scores of the observed variables and k is the number of the observed variables.

A general formulation of a covariance structure model for confirmatory factor analysis (CFA) with latent variables is as follows (Jöreskog & Sörbom, 1996). Using LISREL notation,

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

where $\mathbf{x}$ is a $(k \times 1)$ vector of indicators (the observed or measured variables) of the m exogenous latent variables $\mathbf{\xi}$, $\mathbf{\Lambda}$ is a $(k \times m)$ matrix of the loadings of $\mathbf{x}$ on $\mathbf{\xi}$, and $\mathbf{\delta}$ is a $(k \times 1)$ vector of measurement errors. It is assumed that the ξs and δs are random variables with zero means, δs are uncorrelated with ξs, and all observed variables are measured in deviations from their means. The measurement model represents the regression of $\mathbf{x}$ on $\mathbf{\xi}$ and the element λ_{ij} of $\mathbf{\Lambda}$ is the partial regression coefficient of ξ_j in the regression of x_i on $\xi_1, \xi_2, \ldots, \xi_m$. The model implied covariance matrix for the x variables is defined as:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta},$$

where $\boldsymbol{\Phi}$ is the covariance matrix for $\boldsymbol{\xi}$ and $\boldsymbol{\Theta}$ is the covariance matrix for $\boldsymbol{\delta}$.

Given a specified model, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, the unknown parameters of $\boldsymbol{\theta}$ are estimated so that the discrepancy between the sample implied covariance matrix $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}})$ and the sample covariance matrix, **S** is as small as possible given some criterion, where $\hat{\boldsymbol{\theta}}$ is the vector of parameter estimates. A discrepancy function $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ is needed to quantify the fit of a model to the sample data. This function should have the following properties:

1. $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ is a scalar, and $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) \geq 0$.
2. $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ = 0 if and only if $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ = **S**.
3. $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ is twice differentiable in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and **S**.

Minimizing a discrepancy function with these properties leads to consistent estimators of $\boldsymbol{\theta}$ when the model is correctly specified and some regularity conditions are satisfied (Browne, 1984; Hoogland & Boomsma, 1998).

With these definitions and properties, Browne (1982, 1984) framed the discrepancy function approach into a weighted least squares (WLS) approach and demonstrated that all existing discrepancy functions were special cases of the following WLS discrepancy function:

$$F(\boldsymbol{\theta}) = (\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}))'\mathbf{V}(\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})),$$

where $\mathbf{s} = \mathbf{K}'vec(\mathbf{S}) = (s_{11}, s_{21}, s_{22}, s_{31}, ..., s_{kk})'$ is a vector consisting of the nonduplicated elements of **S**, $vec(\mathbf{S})$ is a vector of order $k^2 \times 1$ consisting of the columns of **S** strung under each other, $\mathbf{K} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}$, where **D** is the duplication matrix which transforms **s** to $vec(\mathbf{S})$, $\mathbf{K}'$ is the generalized inverse of **D**, which transforms $vec(\mathbf{S})$ to **s**, and **V** is a specific positive definite weight matrix and is defined differently for different discrepancy functions:

Unweighted least squares (ULS): $\mathbf{V}_{ULS} = \mathbf{I}^*$

Generalized least squares (GLS): $\mathbf{V}_{GLS} = \mathbf{D}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1})\mathbf{D}$

Maximum likelihood (ML): $\mathbf{V}_{ML} = \mathbf{D}'(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})^{-1} \otimes \mathbf{\Sigma}(\hat{\boldsymbol{\theta}})^{-1})\mathbf{D}$

Asymptotically distribution free (ADF): $\mathbf{V}_{NNT} = \mathbf{W}_{NNT}^{-1}$

Diagonally weighted least squares (DWLS): $\mathbf{V}_{DWLS} = \mathbf{D}_{W_{NNT}}^{-1} = [diag\mathbf{W}_{NNT}]^{-1}$.

In theses expressions, $\mathbf{I}^* = diag(1,2,1,2,2,1,...)$, $\otimes$ is the symbol for a Kronecker product, and $\mathbf{W}_{NNT}$ is the asymptotic covariance matrix of **S**, estimated without assuming normality. DWLS can be formulated by using the diagonal elements of $\mathbf{W}_{NT} = \mathbf{V}_{GLS}$, but in this dissertation DWLS refers to estimates obtained using $\mathbf{V}_{DWLS}$ defined above. As suggested by the expressions for the weight matrices for GLS and ML estimation, several of the weight matrices are functions of population parameters and must, in practice, be estimated.

The goal of estimation is to produce $\hat{\mathbf{\Sigma}}(\hat{\boldsymbol{\theta}})$ that is as similar as possible to the sample covariance matrix, **S**, with similarity defined by the discrepancy function. The weight matrix, **V**, in the discrepancy functions above, determines the estimation method chosen. Let **U** be the sampling covariance matrix of the non-redundant elements of **S** and $\mathbf{\Delta} = \partial[\mathbf{\Sigma}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}' |_{\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}}$. The optimal choice for the weight matrix is $\mathbf{U}^{-1}$ and this matrix provides best generalized least squares estimates. For example, if the data are multivariate normal $\mathbf{V}_{GLS}$ or $\mathbf{V}_{ML}$ provide optimal weight matrices. Browne (1982) has shown that if the optimal weight matrix is used, the asymptotic sampling covariance matrix for the parameter estimates is $(N-1)^{-1}(\mathbf{\Delta}'\mathbf{U}^{-1}\mathbf{\Delta})^{-1}$. For any other choice of the weight matrix the asymptotic sampling covariance matrix is $(N-1)^{-1}(\mathbf{\Delta}'\mathbf{V}\mathbf{\Delta})^{-1}(\mathbf{\Delta}'\mathbf{V}\mathbf{U}\mathbf{V}\mathbf{\Delta})(\mathbf{\Delta}'\mathbf{V}\mathbf{\Delta})^{-1}$. In addition, Browne (1982) has shown that the diagonal elements of $(N-1)^{-1}(\boldsymbol{\Delta}'\mathbf{U}^{-1}\boldsymbol{\Delta})^{-1}$ are not smaller than the diagonal elements of $(N-1)^{-1}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-1}(\boldsymbol{\Delta}'\mathbf{V}\mathbf{U}\mathbf{V}\boldsymbol{\Delta})(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-1}$ so that the standard errors of the best generalized least squares (BGLS) estimator cannot be larger than the standard errors when a sub-optimal weight matrix is used.

An important aspect of covariance structure analysis is the evaluation of the fit of the model. A typical statistic for such evaluation is $(N-1)\hat{F} \equiv (N-1)F(\hat{\boldsymbol{\theta}})$, which is the so called chi-square goodness of fit statistic (Satorra, 1990). The residual vector, $\mathbf{K}'(\mathbf{S}-\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}))$ contains the non-redundant residuals and also provides information for the fit of the model (Satorra, 1990). It may be noted that under the assumption of the asymptotic normality of the residual vector, $\left(\mathbf{K}'\left(\mathbf{S}-\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}})\right)\right)$, WLS estimation amounts to a problem in the distribution of a quadratic form in normal variables about which much is known. In fact, if one considers obtaining the symmetric square root $\mathbf{V}^{1/2}$ of **V** in the WLS discrepancy function above and multiplying the resulting matrix into the vectors on either side, it is apparent with an optimal **V** one obtains a vector of independent variates. Thus, the WLS discrepancy function can be considered to represent the sum of squares of independent normal variables which is intimately related to the chi-square distribution. In fact, with such an optimal weight matrix, the WLS estimator is a minimum chi-square estimator, or minimum modified chi-square estimator when the weight matrix is estimated (Bentler, 1983).

The matrix $\mathbf{W}_{NNT} = \mathbf{K}'\left(2\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}+\mathbf{C}\right)\mathbf{K}$, where **C** is a $k^2 \times k^2$ matrix with elements that are fourth order cumulants of **x**, the $k \times 1$ random vector for the data with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (Browne & Shapiro, 1988). That is

$$\mathbf{C} = E[vec\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'\}vec'\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'\}] - vec(\boldsymbol{\Sigma})vec'(\boldsymbol{\Sigma}) - \mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'(2\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}).$$

The matrix $(N-1)^{-1}\mathbf{W}_{NNT}$ is the asymptotic covariance matrix of the non-redundant elements of the covariance matrix **S** and its elements will be finite assuming only that the variables have finite fourth order cumulants. Because ADF and DWLS are based on $\mathbf{W}_{NNT}$ they do not make strong assumptions about the distribution of the data. ULS does not make any assumptions about the data but also uses less information about the data than do ADF or DWLS. If the data are normal, $\mathbf{C}=\mathbf{0}$ and $\mathbf{W}_{NNT}$ simplifies to $\mathbf{W}_{NT}=2\mathbf{K}'\left(\mathbf{\Sigma}\otimes\mathbf{\Sigma}\right)\mathbf{K}$. As GLS uses the inverse of $\mathbf{W}_{NT}$ as its weight matrix, GLS is based on the normality assumption.

ML estimates can be obtained by using two different discrepancy functions. First ML estimates can be obtained by minimizing the following discrepancy function which is well-known:

$$F(\boldsymbol{\theta})=\log|\mathbf{\Sigma}(\boldsymbol{\theta})|+tr\{\mathbf{S}\mathbf{\Sigma}(\boldsymbol{\theta})^{-1}\}-\log|\mathbf{S}|-k$$

In addition ML estimates can be obtained by minimizing

$$F(\boldsymbol{\theta})=(\mathbf{s}-\boldsymbol{\sigma}(\boldsymbol{\theta}))'\mathbf{V}_{ML}(\mathbf{s}-\boldsymbol{\sigma}(\boldsymbol{\theta}))\,.$$

The ML weight matrix, $\mathbf{V}_{ML}$ is derived as a function of elements of $\mathbf{\Sigma}(\boldsymbol{\theta})$. This means that the ML weight matrix is effectively updated as the estimate of $\mathbf{\Sigma}(\boldsymbol{\theta})$ changes at each iteration in the estimation process. Both of these discrepancy functions have a minimum at the same point in the parameter space, namely at the ML estimates, but the minimum value of the functions are not the same (Jöreskog et al., 1996; Satorra, 1990). The first of the two functions is referred to as the ML discrepancy function and $(N-1)\hat{F}\equiv(N-1)F(\hat{\boldsymbol{\theta}})$ is referred to as the minimum fit function chi square. The second function is referred to as the normal theory WLS discrepancy function and $(N-1)\hat{F}\equiv(N-1)F(\hat{\boldsymbol{\theta}})$ is referred to as the normal theory weighted least squares chi square. ML estimates assume the data are drawn from a multivariate normal distribution.

If an estimate of the optimal weight matrix is used, for example when the data are multivariate normal and GLS is used, then the sampling covariance matrix for $\hat{\boldsymbol{\theta}}$ is consistently estimated by $(N-1)^{-1}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-1}$ where $\boldsymbol{\Delta}$ is the Jacobian matrix. However, if the weight matrix is not correct for the distribution of the data $(N-1)^{-1}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-1}$ may not be a consistent estimator of sampling covariance matrix for $\hat{\boldsymbol{\theta}}$. Browne (1982) has shown that

$$(N-1)^{-1}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{V}\mathbf{W}_{NNT}\mathbf{V}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-1}$$

is consistent even when the weight matrix is not correct for the distribution of the data. Using this expression, replacing $\mathbf{V}$ by $\mathbf{V}_{ML} = \mathbf{D}'(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1} \otimes \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1})\mathbf{D}$ and $\mathbf{W}_{NNT}$ by a consistent estimator, to calculate standard errors of the ML estimates is referred to as Robust ML. The expression $(N-1)^{-1}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{V}\mathbf{W}_{NNT}\mathbf{V}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{V}\boldsymbol{\Delta})^{-1}$ can also be used with other weight matrices and will provide a consistent estimator of the sampling covariance matrix for the parameters. Replacing $\mathbf{V}$ by $\mathbf{V}_{DWLS}$, $\mathbf{V}_{GLS}$, and $\mathbf{V}_{ULS}$ to calculate standard errors of the DWLS, GLS, and ULS estimates, respectively, are called Robust DWLS, Robust GLS, and Robust ULS, respectively. In particular applying this procedure to the DWLS estimator may be attractive. The weight matrix in DWLS is $\mathbf{W}_{NNT}$ diagonal. Thus, like ADF it is not based on the normality assumption, but unlike ADF will not be affected by sampling errors in the off-diagonal elements of an estimator of $\mathbf{W}_{NNT}$, which is a $k\left(k+1\right)/2 \times k\left(k+1\right)/2$ matrix and is likely to have substantial sampling error.

Browne and Shapiro (1987) have shown that when the weight matrix for ML or GLS is used but the weight matrix is not optimal, under certain conditions $(N-1)^{-1}(\boldsymbol{\Delta}'\mathbf{V}_{ML}\boldsymbol{\Delta})^{-1}$ and $(N-1)^{-1}(\boldsymbol{\Delta}'\mathbf{V}_{GLS}\boldsymbol{\Delta})^{-1}$ where $\mathbf{V}_{GLS} = \mathbf{D}'\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{D}$, provides consistent estimates of the standard errors of the estimates of the elements of $\mathbf{\Lambda}$ but not of the other parameters, even when the random vectors in the model are not normally distributed. For the factor analysis model the required conditions are that (a) the various latent random vectors for a model are independent, not merely uncorrelated, (b) uncorrelated elements within a latent random vector are independent, not merely uncorrelated, and (c) other than constraints that the off-diagonal elements of a covariance matrix for a random vector are zero, the covariance matrices must be unconstrained. For the factor analysis model the elements of the vectors $\boldsymbol{\delta}$, which are specified to be uncorrelated in the CFA model, must be mutually statistically independent. Similarly the elements of the vectors $\boldsymbol{\delta}$ which are specified in the CFA model to be uncorrelated with the elements of $\boldsymbol{\xi}$ must be statistically independent of the elements of $\boldsymbol{\xi}$. In addition, with the exception of the constraints that off-diagonal elements of $\mathbf{\Theta}$ are zero and of any constraints that off-diagonal elements of $\mathbf{\Phi}$ are zero, the elements of $\mathbf{\Phi}$ and $\mathbf{\Theta}$ must be unconstrained. The assumptions in Browne and Shapiro imply that $\mathbf{C}$ is a function of the cumulant matrix for the random vectors in the model and of the matrix $\mathbf{\Lambda}$. If the assumptions are not correct standard errors should be calculated by using

$$(N-1)^{-1}(\mathbf{\Delta}'\mathbf{V}\mathbf{\Delta})^{-1}\mathbf{\Delta}'\mathbf{V}\mathbf{W}_{NNT}\mathbf{V}\mathbf{\Delta}(\mathbf{\Delta}'\mathbf{V}\mathbf{\Delta})^{-1}.$$

Table 1-2 lists the estimation methods investigated in the past robustness studies identified in the Table 1-1. Of the 47 studies, ML was investigated in a total of 43 studies, GLS in a total of 16 studies, and ADF in a total of 22 studies. Estimation methods other than ML, GLS, ULS, DWLS, their robust versions, and ADF were not considered in preparing Table 1-2.

The most popular techniques for estimating the parameters in SEM are ML and GLS and there are three major assumptions for ML and GLS (Boomsma & Hoogland, 2001):

- The sample observations are independently distributed.
- The sample observations are multivariate normally distributed.
- The hypothesized model is approximately correct.

As was mentioned above, real data in social science are rarely normally distributed. The violation of normality assumption raises a robustness question in terms of distributional violation. However, as sample size increases, the distribution of the estimator approximates a normal distribution, which is why many researchers have investigated the effect of non-normality in SEM with different sample sizes to determine the minimum required sample size for valid parameter estimates.

The restrictive characteristic of the normality assumption motivated the development of the WLS procedure, an asymptotically distribution-free (ADF) method. This method does not assume a specific distribution and can produce results which are valid under a wide variety of distributions of the data. However, these ADF methods face some practical problems in that they are computationally expensive and they lack robustness against small to moderate sample sizes (Satorra, 1990). That is, ADF methods need a sufficiently large sample size. Thus the normality assumption still plays a major role in the practice of structural equation modeling.

Several issues in regard to robustness of estimation methods against non-normality in SEM were introduced and reviewed and by Satorra (1990). He summarized several estimation methods (ML, ULS, DWLS, and ADF) and asymptotic robustness of normal theory inferences. Additionally, theoretical and empirical robustness to violation of assumptions were explained, a distinction was made between estimation methods that are either correctly or incorrectly specified for the distribution of data being analyzed, and a comparison of ML, ADF, and Robust ML was reported using a real data example about teacher stress (Bentler & Dudgeon, 1996).

Moreover, there are factors affecting the evaluation and modification of a model such as non-normality, missing data, specification error, sensitivity to sample size, etc. In regard to these aspects, general guidelines for covariance structural equation modeling have been presented over the past decades (Kaplan, 1990; MacCallum, 1990; Ullman, 2006).

Robustness against Non-normal Distribution

The parameter estimates are derived from information in the sample covariance matrix and the weight matrix. When the observations are continuous non-normal, the information in the sample covariance matrix or the weight matrix or both may be incorrect. Consequently, estimates based on the sample covariance matrix and the weight matrix may also be incorrect. The present study examined non-normality of the observed continuous variables. The variation in the measured variables is completely summarized by the sample covariances only when multivariate normality is present. If multivariate normality is violated, the variation of the measured variables will not be completely summarized by the sample covariances, so information from higher order moments is lost. In this situation, the parameter estimates do remain unbiased and consistent as sample size grows larger, but they are no longer efficient. These results suggest that theoretically important problems will occur with normal theory estimators such as ML and GLS when the observed variables do not have a multivariate normal distribution. The impacts of non-normality are that chi-square statistics, standard errors, or tests of all parameter estimates can be biased (West et al., 1995).

Most of previous studies of violation of the normality assumption are for ML, GLS, and ADF (e.g., Muthen & Kaplan, 1985, 1992; Boomsma, 1983; Hoogland, 1999; Boomsma & Hoogland, 2001; Olsson et al., 2000; Lei & Lomax, 2005). Table 1-3 provides a simple comparison of ML, GLS, and ADF, taken from Olsson, Foss, Troye, and Howell (2000). As shown in Table 1-3, when the models are incorrectly specified and the data are not multivariate normal, the methods should give different results. With multivariate normal data but a misspecified model, ADF and GLS will be equivalent (Olsson et al., 2000).

If the variables are highly non-normal, it is still an open question whether to use ML, GLS, or ADF. Previous studies have not given a clear-cut answer as to when it is necessary to use which estimation method and it is possible that standard errors produced by ML, GLS, ULS, DWLS may be underestimated when the observed variables deviate far from normality (LISREL 8 User's Reference Guide). Table 1-4 shows the degree of non-normality investigated in the past robustness studies.

The present study examined robustness against violations of the assumption of normality using Monte Carlo methods. The degree of non-normality can be specified by skewness and kurtosis. A brief review of skewness and kurtosis is presented below before discussing the range of non-normality utilized in this study.

Skewness and Kurtosis in Univariate Distribution

Univariate skewness can be viewed as how much a distribution departs from symmetry, and univariate kurtosis has been described the extent to which the height of the probability density differs from that of the normal density curve, that is, kurtosis measures the peakedness or flatness of the probability density function (Casella & Berger, 2002; Harlow, 1985; West et al., 1995). Negative skewness indicates a distribution with an elongated left-hand tail and positive skewness indicates a distribution with an elongated right-hand tail relative to the symmetrical normal distribution. Zero skewness indicates symmetry around the mean. Negative kurtosis indicates flatness and short tails relative to a normal distribution, whereas positive kurtosis indicates peakedness and long tails relative to a normal curve (Bentler & Yuan, 1999; Harlow, 1985; Olsson et al., 2000; West et al., 1995).

Measures of skewness and kurtosis can be defined by using central moments or by using cumulants. The nth central moment of X, μ_n , is

$$\mu_n = E[(X-\mu)^n],$$

or

$$\mu_n = \int_{-\infty}^{\infty} (x-\mu)^n f_X(x)dx$$

where the first non-central moment is the mean: $\mu = \mu_1' = EX$ and $f_X(x)$ is the probability density function of a continuous random variable, *X*. The second central moment is the variance:

$$\mu_2 = E[(X-\mu)^2] = VarX = E(X-EX)^2.$$

Kendall, Stuart, and Ord (1987) presented several measures of skewness and kurtosis in terms of the moments of a distribution. Skewness and kurtosis can be defined as β_1 and β_2 respectively, where

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3},$$

and

$$\beta_2 = \frac{\mu_4}{\mu_2^2}.$$

In the univariate normal case, where the variables are assumed to have symmetrical, bell-shaped distributions, β_1 = 0 and β_2= 3.

Measures of skewness and kurtosis can also be defined by using cumulants. Formally, the cumulants κ_1, κ_2, …, κ_n are defined by the identity in t

$$\exp(\sum_{n=1}^{\infty} \kappa_n t^n / n!) = \sum_{n=0}^{\infty} \mu_n' t^n / n!,$$

and it should be observed that there is no κ_0 (Kendall et al., 1987). The cumulant generating function is simply the logarithm of the moment generating function. The first order cumulant is simply the mean (expected value); the second order and third order cumulants are respectively the second and third central moments; the higher order cumulants are neither moments nor central moments, but rather more complicated polynomial functions of the moments. For example the fourth order cumulant is

$$\kappa_4 = \mu_4 - 3\mu_2^2.$$

Kendall, Stuart, and Ord (1987) used cumulants to also define alternative measures with the of skewness and kurtosis, respectively

$$\gamma_1 = \frac{\kappa_3}{\kappa_2^{3/2}} = \frac{\mu_3}{\mu_2^{3/2}},$$

$$\gamma_2 = \frac{\kappa_4}{\kappa_2^2} = \beta_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3,$$

where γ_1 = 0, and γ_2 = 0 for univariate normal distributions.

Skewness and Kurtosis in Multivariate Distributions

Examinations of the skewness and kurtosis of the univariate distributions provide only an initial check on multivariate normality. If any of the observed variables deviate from univariate normality, the multivariate distribution cannot be multinormal (West et al., 1995). While univariate measures of skewness and kurtosis are informative regarding the marginal distribution of a variable, it is also of interest to have information on the joint distribution of a set of variables (Halrow, 1985). As a result, it is important to examine multivariate measures of skewness and kurtosis developed by Mardia (1970, 1974). The Mardia measures are functions of the third order moments and the fourth order moments, which possess approximate standard normal distributions permitting tests of multivariate skewness and multivariate kurtosis (Harlow, 1985; Mardia, 1970, 1974; West et al., 1995).

Let $\mathbf{X} = (X_1,...,X_p)'$ be a random vector with mean vector $\boldsymbol{\mu} = (\mu_1,...,\mu_p)'$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{rs})$. Mardia (1970, 1974) defined the measures of multivariate skewness and kurtosis as

$$\beta_{1,p} = E\{(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\}^3,$$

and

$$\beta_{2,p} = E\{(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\}^3,$$

where **X** and **Y** are independent and identical random vectors. Mardia (1974) also suggested the alternative expression for these measures in terms of cumulants and the expressions corresponding to $\beta_{1,p}$ and $\beta_{2,p}$ are

$$\gamma_{1,p} = \sum\sum \kappa^{rr'}\kappa^{ss'}\kappa^{tt'}\kappa_{111}^{rst}\kappa_{111}^{(r's't')},$$

and

$$\gamma_{2,p} = \sum\sum \kappa^{rs}\kappa^{tu}\kappa_{1111}^{rstu},$$

where $\kappa_{i_1 \dots i_s}^{(r_1 \dots r_s)}$ denote the cumulant of order $(i_1,...,i_s)$ for the random variable $(X_{r_1},...,X_{r_s})$ where $r_1, r_2, \dots, r_s$ are s integers taking values 1, 2, …, p, and $(\kappa_{11}^{(rs)})^{-1} = (\kappa^{rs})$. The relations between $\gamma_{1,p}$ and $\beta_{1,p}$, and between $\gamma_{2,p}$ and $\beta_{2,p}$ (Mardia, 1974; Kendall et al., 1987) are

$$\gamma_{1,p} = \beta_{1,p},$$

and

$$\gamma_{2,p} = \beta_{2,p} - p(p+2).$$

In the multivariate normal case, if X_i has a normal distribution then $\gamma_{1,(i)} = 0$ and $\gamma_{2,(i)} = 0$ and if **X** has a p-variate normal distribution $\gamma_{1,p} = 0$ and also all fourth order cumulants are zero with the result that $\gamma_{2,p} = 0$ (Browne, 1982).

Robustness against Small Sample Size

In SEM or any other procedure for fitting models, inferences are made from observed data to the model believed to be generating the observations. These inferences are dependent in large part on the degree to which the information available in a sample mirrors the information in the complete population. This depends on the obtained sample size. To the extent that samples are large, more information is available and more confidence can be expressed for the model as a reflection of the population process. Thus, sample size has always been an issue in SEM. To get the correct answer to the question: "What is the minimum required sample size for each combination of estimation methods, distributional characteristics, and model characteristics?" many studies have been conducted as shown in the list of the past robustness studies in Table 1-1. The range of sample sizes examined in the past robustness studies is shown in Table 1-5.

The typical requirements for a covariance structure statistic to be trustworthy under the null hypothesis $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$ are that the sample observations are independently distributed and multivariate normally distributed in addition to identification of parameters. Identification and independence were assumed and not considered in the present study because they can often be arranged by design. Generally, real data do not meet the assumption of multivariate normality, so sample size turns out to be critical because all of the goodness of fit statistics and standard errors used in covariance structure analysis are "asymptotic" based on the assumption that sample size becomes arbitrarily large. Since this situation can rarely be obtained, it becomes important to evaluate how large the sample size must be in practice for the theory to work reasonably well.

Different data and discrepancy functions have different robustness properties with respect to sample size. Basically, sample size requirements increase as data become more non-normal, models become larger, and assumption-free discrepancy functions are used (Bentler & Dudgeon, 1996; Chou et al., 1991; Hu et al., 1992; Muthen &Kaplan, 1992; West et al., 1995; Yung & Bentler, 1994).

Sample size also affects the likelihood of non-convergence and improper solution. Non-convergence occurs when a minimum of the discrepancy function cannot be obtained. Improper solutions refer to estimates that are outside of their proper range, for example estimated variances that are less than zero. The likelihood of non-convergence decreases with larger sample size; N >200 is generally safer. Non-convergence rates also decrease with larger factor loadings and larger k/m ratio (Boomsma & Hoogland, 2001; Marsh et al., 1998), where k/m is ratio of the number of observed variables to the number of common factors. The likelihood of improper solutions is reduced with a larger k/m ratio and increased N (Boomsma & Hoogland, 2001; Marsh et al., 1998). Thus sample size plays a crucial role in robustness, non-convergence, and improper solutions.

Research Summary Characteristics

Research summary characteristics determine how the quality of the simulation results is assessed. The following research summary statistics have been used in the previous robustness studies of the Table 1-1.

- Bias of parameter estimates
- Bias of standard error estimates
- Standard deviation of parameter estimates
- Percentage of replications that lead to non-convergence

- Percentage of convergent replications that give improper solutions
- Rejection rate, mean, and standard deviation of a chi-square statistic
- *p*-value of the Kolmogorov-Smirnov test for a chi-square distribution Statistics

The relative bias of parameter estimators is defined as

$$Bias(\hat{\theta}_i) = \frac{\overline{\hat{\theta}}_i - \theta_i}{\theta_i},$$

where θ_i is the population value of the i^{th} parameter ($\theta_i \neq 0$) and $\overline{\hat{\theta}_i}$ is the mean of the estimates for the i^{th} parameter across the total number of replications. The mean absolute relative bias for parameter estimation is

$$\frac{1}{t}\sum_{i=1}^{t}\left|Bias(\hat{\theta}_i)\right| ,$$

where *t* is the number of parameters in the model. According to Boomsma and Hoogland (2001) the mean absolute relative bias for parameter estimation should be less than 0.025.

The relative bias of estimators for the standard error of parameter estimates θ_i is defined as

$$Bias(se(\hat{\theta}_i)) = \frac{\overline{se(\hat{\theta}_i)} - sd(\hat{\theta}_i)}{sd(\hat{\theta}_i)},$$

where $sd(\hat{\theta}_i)$ is the standard deviation of the estimates for parameter *i* and $\overline{se(\hat{\theta}_i)}$ is the mean of the standard error estimates regarding parameter *i* across the total number of replications. The mean absolute relative bias for standard error estimation is

$$\frac{1}{t}\sum_{i=1}^{t}\left|Bias(se(\hat{\theta}_i))\right|,$$

which is considered to be acceptable, if it is smaller than 0.05 (Boomsma & Hoogland, 2001; Harlow, 1985; Hoogland, 1999).

Specially, standard errors of parameter estimates are important in many applications to judge the significance of the parameter estimates. LISREL gives an estimated asymptotic standard error for each parameter. It is well known that these asymptotic standard errors for ML may be incorrect when the observed variables are not multivariate normal (Browne, 1984; Jonsson, 1997).

Findings from Overall Past Robustness Studies

Findings from Boomsma (1983) and Muthen and Kaplan (1985) suggest that estimated standard errors did not show bias when using ML, GLS, and ADF with approximately normal data. In non-normal samples, there is some evidence of negative bias in estimated standard errors when using ML with continuous data (Browne, 1984; Tanaka, 1984) as well as with ADF in sample sizes of 100 (Tanaka, 1984). In studies with 400 or more subjects, estimated standard errors also showed some bias, relative to empirical standard errors, with non-normal samples using ML with categorical data (Boomsma, 1983; Muthen & Kaplan, 1985) and ADF with categorical data (Muthen & Kaplan, 1984).

Hoogland summarized (1999) his conclusions about parameter estimators in his research with the following points:

- An important finding is that the ML estimator of the model parameters (5 factors with 3 indicators per factor, and 4 factors with 3 indicators per factor) is almost unbiased when the sample size is at least 200.
- In the case of a small sample the GLS parameter estimator has a much larger bias than the ML parameter estimator when the model has at least twelve observed variables.
- The bias of the ADF parameter estimator increases when the kurtosis increases.
- With positive kurtosis the bias of the ADF parameter estimator is larger than that of the GLS parameter estimator.

Hoogland made (1999) the following points about the standard error estimators based on the results of his research and the previous studies:

- The ML and GLS estimators of the standard errors are biased when the average kurtosis of the observed variables deviates from zero.
- The standard errors are underestimated in the case of a positive average kurtosis and overestimated in the case of negative average kurtosis.
- The ADF estimation method results in a large underestimation of the standard errors when the sample size is small relative to the number of observed variables in the model.
- The robust ML standard error estimator has a smaller bias than the other standard error estimations when the average kurtosis is at least 2.0 and the sample size is at least 400.

From the past robustness studies, the following points can be set forth:

- The sample sizes 200 and 400 were often investigated and were minimum sample sizes, though it depends on the models investigated.
- If the variables are highly non-normal it is still an open question whether to use ML, GLS, ADF, or other methods in regard to bias of standard errors.
- Robust ML standard errors were rarely investigated for the effect of sample size and non-normality compared with robustness studies in ML.
- DWLS estimation method was not examined for continuous non-normal distribution.
- Standard errors for Robust DWLS, Robust GLS, and Robust ULS were not investigated in previous studies, though GLS and ML estimation methods have quite often been studied.
- The numbers of replications, 100, 200, and 300 have often been used in previous robustness studies (Table 1-6). The number of replications is chosen to be reasonable because it is a trade off between precision and the amount of information to be handled (Hoogland, 1999).
- Effects of different scale setting methods on standard error estimates in robust procedures were not investigated.

Direction for the Present Study

The present study examined the effect of sample size on the accuracy of estimation of standard errors under varying degrees of non-normality with four estimation methods (ML, GLS, Robust ML, and Robust GLS) and three CFA models. The reasons ML, GLS, Robust ML, and

Robust GLS were chosen as estimation methods are that standard errors of Robust ML and Robust GLS have not been very thoroughly investigated for effect of sample size, non-normality, and scale-setting method. One of the main purposes of the present study is to answer the question: "Which estimation method is better to get asymptotically correct standard errors?"

A review of the previous studies of Robust ML standard errors provides a clearer idea for the simulation design of the present study. Chou et al. (1991) investigated the effect of non-normality on standard errors varying estimation methods (ML, Robust ML, and ADF) with two different versions of the CFA model, a sample size of 400, and 100 replications. For non-normality condition, they chose four non-normal conditions: (a) symmetric with equal negative kurtosis, (b) symmetric with unequal kurtosis, (c) unequal negative skewness with zero kurtosis, and (d) unequal skewness with unequal kurtosis to study the behavior of the robust standard errors. For the CFA model with two factors and three indicators per factor, there were two different versions. The first version contained 13 parameters, six factor loadings, six measurement error variances, and one factor covariance. The second version only included the measurement error variances and the factor covariance as free parameters while fixing the factor loadings at the true values (Chou et al., 1991). In case of the first version of the model, for non-normality conditions (a) and (c), all three types of estimated standard errors were very similar, mean ADF and robust standard errors were closer to the expected values. Under condition (b) and (d), all three types of estimated standard errors were negatively biased with robust standard errors performing slightly better than ADF and ML. Similar conclusions were drawn from the results for the second version of the CFA model. Robust standard errors provided better estimates than both ML and ADF, although some robust standard errors were not as close to the expected values for the second version of the model as they were for the first version of the model.

Finch et al. (1997) studied the effect of sample size and non-normality on the estimation of mediated effects in latent variable models. They investigated the mediated effects varying the structural regression coefficients in a model with three factors with three items per factor and concluded that the magnitude of the observed bias for estimating standard errors varied little across differing ratios of direct to indirect effects for ML, Robust ML, and ADF. First, ML, ADF, and Robust ML standard errors were examined to determine the range of non-normality conditions under which these standard errors are accurate, varying sample size (150, 250, 500, and 1000), the population values for the hybrid model parameter, and the degree of non-normality (normal [0,0], moderate non-normality [2,7], and extreme non-normality [3, 21]) with 200 replications. The partially mediated structural model included three factors and three indicators per factor and had four different mediated effect sizes. For Robust ML, weaker effects of non-normality on the standard errors were observed. With regard to the robust estimates of the standard error of the indirect effect, there was minimal bias under normality, and bias increased under severe non-normality. Weak effects of samples size were also observed with some decrease in relative bias associated with larger sample sizes. Under non-normality, the Robust ML standard errors performed much better at all sample sizes. For all three methods of estimating standard errors, the magnitude of the observed bias varied little across differing ratios of direct to indirect effects. The pattern of bias in the standard errors of direct and indirect effects was also not influenced by variation in the population values of the factor loadings. In the second study, they extended the generality of the findings by examining bias under conditions in which the degree of non-normality differed across observed (manifest) variables. The population values of skewness and kurtosis were specified in the three distributional conditions. In the first distributional condition, the nine measured variables were mildly to moderately non-normal; in the second the nine manifest variables were moderately to severely non-normal; and in the third the measured indicators of the final outcome construct were severely non-normal, whereas the other six variables were normally distributed. They adopted one specific latent variable model for the second study with 400 replications and the same conditions for sample size that they used in their first study. Under the first distributional condition, there were no appreciable effects of sample size on the estimated standard errors of the indirect effect using ML or Robust ML. Under the second distributional condition, Robust ML estimated standard errors of the structural coefficients showed a general tendency to become more accurate with increasing sample size. Under the third distributional condition, there were large effects of non-normality on the ML and Robust ML standard errors of the latent variable coefficients. The Robust ML standard errors provided more accurate estimates of sampling variability than the normal theory standard errors as non-normality increased. The Robust ML standard errors were also more accurate than the ADF standard errors under all distributional conditions at the smaller sample sizes (i.e., 150 or 250).

Hoogland (1999) investigated the effect of four factors on relative bias of standard error estimates: sample size, size of factor loadings, model complexity, and number of indicators per factor. Model complexity was varied as follows: four factors with three items, four factors with four items, and five factors with three items per factor, and one hybrid model with five factors and three items per factor. For $N = 200$, the mean absolute relative bias of the Robust ML standard error estimator was larger for each parameter type when the indicators were poor measures of the latent variables. For $N = 400$, the mean absolute relative bias of the Robust ML standard errors was also larger for smaller factor loadings. For $N \geq 800$, the effect of the size of the factor loadings on the mean absolute relative bias of the Robust ML standard error estimator was mostly small. In case of model complexity, there was hardly an effect on the mean absolute relative bias of the Robust ML standard errors as the model complexity increases for $N \geq 200$. The mean absolute relative bias of the Robust ML standard errors decreased somewhat as the number of indicators per factor increased when the sample size was small. The difference between the CFA model and the hybrid model was negligible for the mean absolute relative bias of the Robust ML standard errors.

According to the previous studies of Robust ML (Chou et al., 1991; Finch et al., 1997; Hoogland et al., 1999), Robust ML generally showed better performance on standard errors rather than other estimation methods investigated over all sample sizes and degrees of non-normality.

For Robust GLS, no Monte Carlo study has been conducted to investigate the effect of sample size on accuracy of estimated standard error. Thus, a Monte Carlo study of the effect of sample size on accuracy of estimated standard errors under varying degrees of non-normality with three CFA models using ML, GLS, Robust ML, and Robust GLS will provide more precise guidelines in regard to sample size in SEM for a non-normally distributed data and a complex model.

The scale of measurement for a factor can be changed without affecting the fit of model to the data, which is called ‘scale indeterminacy’. The data cannot determine the values of the parameters due to scale indeterminacy and therefore this is an identification problem. Arbitrary restrictions are required to remove scale indeterminacy. A parameter’s standard error is sensitive to how the model is identified (i.e., how scale is set). The different scale-setting methods to identify a model may yield different standard errors and reach different conclusion about a parameter's significance although equivalent models on the same data are tested (Gonzalez & Griffin, 2001).

There are two common methods to remove scale indeterminacy: 1) Factor variances are set equal to one, and 2) Factor loadings are set equal to one. Changing the method for setting the scale of a factor multiplies the factor by a constant. Estimates of parameters related to the factor are affected as one would expect for a scale change. For example if the method of setting the scale multiplies the factor by *k,* then the factor loading is multiplied by k^{-1}. However, the standard errors of the factor loading are not simply multiplied by k^{-1}. As a result it is possible that the method for setting the scales may affect the accuracy of estimating of standard errors for factor loadings and this question has not been investigated in the previous robust studies. Note that the scale-setting method does not affect the size of the residual variances at all, whereas the scale-setting method affects the size of the factor loading, factor variances, and factor covariances.

Table 1-1. Overview of previous robustness studies against sample size and non-normality

Paper Number	Author(s)	Year	Sample size	Non-normality
1	Anderson & Gerbing	1984	O	
2	Andreassen, Lorentzen, & Olsson	2006		O
3	Babakus, Ferguson, & Joreskog	1987	O	O
4	Baldwin	1986	O	
5	Beardeu, Sharma, & Teel	1982	O	
6	Beauducel & Herzberg	2006	O	
7	Benson & Fleishman	1994	O	O
8	Bentler & Yuan	1999	O	
9	Boomsma	1983	O	O
10	Brown	1990		O
11	Browne	1984		O
12	Chou, Bentler, & Satorra	1991	O	O
13	Curran, West, & Finch	1996	O	O
14	Dolan	1994	O	O
15	Ethington	1987		O
16	Fan, Thompson, & Wang	1999	O	
17	Finch, J. F.	1992	O	O
18	Finch, West, & Mackinnon	1997	O	O
19	Gallini & Mandeville	1984	O	
20	Gerbing & Anderson	1985	O	
21	Harlow	1985	O	O
22	Hau & Marsh	2004	O	O
23	Henly	1993	O	O
24	Hoogland	1999	O	O
25	Hu, Bentler, & Kano	1992	O	O
26	Huba & Harlow	1987		O
27	Jaccard & Wan	1995	O	O
28	Jackson	2001	O	
29	Kaplan	1989	O	
30	Lee, Poon, & Bentler	1995	O	O
31	Lei & Lomax	2005	O	O
32	Muthen & Kaplan	1985		O
33	Muthen & Kaplan	1992	O	O
34	Nevitt & Hancock	2000	O	O
35	Nevitt & Hancock	2001	O	O
36	Nevitt & Hancock	2004	O	O
37	Olsson, Foss, Troye, & Howell	2000	O	O
38	Ping	1995	O	
39	Potthast	1993	O	O
40	Reddy	1992	O	
41	Sharma, Durvasula, & Dillon	1989	O	O
42	Sivo, Fan, Witta, & Willse	2006	O	O
43	Tanaka	1984	O	O
44	Tanaka	1987	O	
45	Yang Jonsson	1997	O	O
46	Yuan & Bentler	1998	O	O
47	Yung & Bentler	1994	O	O

Table 1-2. Estimation methods investigated in the past robustness studies

Paper number	Estimation methods
1	ML
2	ML, GLS, ADF
3	ML
4	ML
5	ML
6	ML, DWLS*
7	ML, ADF
8	ML
9	ML
10	ML
11	ML, ADF
12	ML, ADF, Robust ML
13	ML, ADF
14	ML, ADF, GLS
15	ML, ULS
16	ML, GLS
17	ML
18	ML, ADF, Robust ML
19	ML
20	ML
21	ML, ADF
22	ML, ADF
23	ML, GLS, ADF
24	ML, GLS, ADF, Robust ML
25	ML, GLS, ADF
26	ML, GLS, ADF, ULS
27	ML, GLS, ADF
28	ML
29	ML, GLS
30	GLS
31	ML, GLS
32	ML, GLS, ADF
33	GLS, ADF
34	ML
35	ML
36	ML, ADF
37	ML, GLS, ADF
38	ML, ADF
41	ML, GLS
43	ML, ADF
44	ML
45	ML, ADF
46	ADF
47	ML, GLS, ADF

* In this article, DWLS means WLSMV (WLS means and variance adjusted) which is weighted least square parameter estimates using a diagonal weight matrix with standard errors and mean- and variance- adjusted chi-square test statistic that use a full weight matrix.

Table 1-3. When are ML, GLS, and ADF equivalent?

Model/Distribution	Normal	Non-normal
Correct model	ML ←→ GLS ←→ ADF asymptotically equivalent	ML ←→ GLS asymptotically equivalent
Misspecified model	GLS ←→ ADF asymptotically equivalent	No equivalence

Table 1-4. Degree of non-normality examined in the past robustness studies

Paper number	Degree of non-normality (Skewness, Kurtosis)
2	Non-normal real data
3	Combinations of skewness = 0.0, 0.5, 1.5
7	(1.0, 2.0), (2.0, 6.0)
9	Various combinations of skewness and kurtosis
10	(0.0, 0.0), (0.25, 3.75), (1.0, 3.75), (1.75, 3.75)
11	Multivariate normal distribution vs. rescaled multivariate chi-square distribution with 2 degree of freedom and all marginal relative kurtosis coefficients equal to 3.0.
12	Various combinations of skewness and kurtosis
13	(0.0, 0.0), (2.0, 7.0), (3.0, 21.0)
14	(0.0, 3.0), (0.0, 1.0)
15	Combinations of skewness = 0.0, 1.5, -1.5
17	(0.0, 0.0), (2.0, 7.0), (3.0, 21.0)
18	(0.0, 0.0), (2.0, 7.0), (3.0, 21.0)
19	Possible combinations with skewness = 0.0, 1.0, 2.0 and kurtosis = -1.0, 0, 1.0, 3.0, 6.0 with the restriction of $(\text{skewness})^2 < 0.629576(\text{kurtosis}) + 0.717247$
21	13 combinations of skewness and kurtosis
22	(0.0, 0.0), (0.5, 0.5), (1.0, 1.5), (1.5, 3.25), Balanced (± 0.5, ± 0.5)
23	Multivariate normal, elliptical, non-elliptical
24	11 combinations of skewness and kurtosis
25	7 conditions of the range of kurtoses -0.010 ~ 0.010, -0.502 ~ 3.098, -0.502 ~ 3.908, -0.262 ~ 0.989 4.658 ~ 6.827, 4.635 ~ 9.659, 3.930 ~ 20.013
26	Non-normal real data
30	Continuous and polytomous variable
31	(0.0, 0.0), (0.3~0.4, 1.0), (0.7, >3.5)
33	(0.0, 0.0), (-0.74, -0.33), (-1.22, 0.85), (-2.03, 2.90), (0.0, 2.79), (0.0, -1.3)
34	(0.0, 0.0), (2.0, 7.0), (3.0, 21.0)
35	(0.0, 0.0), (2.0, 7.0), (3.0, 21.0)
36	(0.0, 0.0), (0.0, 6.0), (3.0, 21.0)
37	11 different levels of kurtosis in the latent variables ranging from mild (-1.2) to severe peakedness (+25.45)
39	Negative kurtosis(1.9, -1.12), zero skewness and kurtosis(0.0, 0.0), positive kurtosis(0.0, 2.79), high skewness and kurtosis(2.52, 5.8)
41	Combinations of skewness = -1.0, 0.0, 1.0 and kurtosis = 2.0, 4.0, 6.0
43	Zero skewness and various kurtosis
45	After data generation and estimation procedure, checking the degree of non-normality (kurtosis and skewness with p-value)

Table 1-5. Range of sample size examined in the past robustness studies

Paper number	Sample sizes
1	50, 75, 100, 150, 300
2	1042
3	100, 500
4	100, 200
5	25, 50, 75, 100, 500, 1000, 2500, 5000, 10000
6	250, 500, 750, 1000
7	200, 400
8	60, 70, 80, 90, 100, 110, 120
9	25, 50, 100, 200, 400, 800
10	500
11	500
12	200, 400
13	100, 200, 500, 1000
14	200, 300, 400
15	500
16	50, 100, 200, 500, 1000
17	100, 200, 1000
18	150, 250, 500, 1000
19	50, 100, 500
20	50, 75, 100, 150, 300
21	200, 400
22	50, 100, 250, 1000
23	75, 150, 300, 600, 1200, 2400, 4800, 9600
24	200, 400, 800, 1600
25	150, 250, 500, 1000, 2500, 5000
26	257
27	175, 400
28	50, 100, 200, 400, 800
29	100, 500
30	100, 200, 500
31	100, 250, 500, 1000
32	1000
33	500, 1000
34	100, 200, 500, 1000
35	100, 200, 500, 1000
36	n:q = 1:1, 2:1, 5:1, 10:1
	(n: sample size, q: number of parameter estimates
37	100, 250, 500, 1000, 2000
38	100, 300
39	500, 1000
40	100, 500
41	150, 300, 500
42	150, 250, 500, 1000, 2500, 5000
43	100, 500, 1500
44	50, 1200
45	100, 200, 400, 800, 1600, 3200
46	150, 200, 300, 500, 1000, 5000
47	250, 500

Table 1-6. Number of replications in the past robustness studies

Paper number	Number of replications per each condition
3, 4, 23, 24	300
8, 31	500
9, 21	100, 300
10, 12, 14, 19, 20, 38, 39	100
11, 43	20
13, 16, 17, 18, 25, 28, 34, 35, 42, 47	200
15, 40	50
18	400
22	100, 400, 1000, 2500
32	25
33	1000
36	2000
37	16500
41	120
45	600

CHAPTER 2
DESIGN OF MONTE CARLO SIMULATION STUDY

Monte Carlo simulation relates to or involves the use of random sampling techniques and often the use of computer simulation to obtain approximate solutions to mathematical or physical problems especially in terms of a range of values each of which has a calculated probability of being the solution (Fan et al., 2002; Paxton et al., 2001). Monte Carlo simulation simulates the sampling process from a defined population repeatedly by using a computer instead of actually drawing multiple samples to estimate the sampling distributions of the events of the interest.

Data Generation

There are several methods for generating non-normal multivariate data for simulation (Mattson, 1997; Reinartz et al., 2002; Vale & Maurelli, 1983). In some of the past robustness studies of SEM the performance of the Vale and Maurelli data generation method was assessed. This method has received much attention due to its flexibility (Harlow, 1985; Mattson, 1997). The Vale and Maurelli method is an extension of Fleischman's power function method for univariate case (Vale & Maurelli, 1983).

Fleishman's Power Transformation Method

Fleishman (1978) introduced a method to simulate univariate non-normal conditions with desired degrees of skewness and kurtosis. This method uses polynomial transformation to transform a normally distributed variable to a variable with specified degrees of skewness and kurtosis. The polynomial transformation shown in Fleishman (1978) takes the form:

$$Y = a + bZ + cZ^2 + dZ^3,$$

where Y is the transformed non-normal variable with specified population skewness and kurtosis, Z is a unit normal variate (i.e., a normally distributed variable with population mean of zero and variance of one), and a, b, c, and d are coefficients needed for transforming the unit normal variate to a non-normal variable with specified degrees of population skewness and kurtosis. Of the four coefficients, $a = -c$.

Vale and Maurelli Data Generation Method

Vale and Maurelli (1983) presented a procedure for generating multivariate non-normal data that is related to Fleishman's power transformation method. By using the Fleishman's power transformation method, two normal variates, Z_1 and Z_2 can be transformed to two non-normal variables X_1 and X_2, each with known skewness and kurtosis:

$$X_1 = a_1 + b_1 Z_1 + c_1 Z_1^2 + d_1 Z_1^3,$$

and

$$X_2 = a_2 + b_2 Z_2 + c_2 Z_2^2 + d_2 Z_2^3.$$

Once the degrees of skewness and kurtosis are known, the coefficients are available. In addition to these coefficients, the modeled population correlation between the two non-normal variables X_1 and X_2 can be specified as R_{12}. Once R_{12} is set, and the Fleishman coefficients for the specified skewness and kurtosis conditions of the two variables (X_1 and X_2) are available, Vale and Maurelli (1983) showed that the following relation exists:

$$R_{12} = \rho(b_1 b_2 + 3b_1 d_2 + 3d_1 b_2 + 9d_1 d_2) + \rho^2(2c_1 c_2) + \rho^3(6d_1 d_2),$$

where ρ is the correlation between the two normal variates, Z_1 and Z_2, and it is called intermediate correlation. All elements of the equation above are known except the bivariate intermediate correlation, ρ. This bivariate intermediate correlation coefficient, ρ must be solved for all possible pairs of variables involved. These intermediate correlation coefficients are then assembled in proper order into an intermediate correlation matrix and multivariate normal data are generated by using

$$\mathbf{y} = \mathbf{F}\mathbf{z}$$

where $\boldsymbol{\Sigma} = \mathbf{F}\mathbf{F}'$, $\mathbf{z} \sim MVN(\mathbf{0}, \mathbf{I})$ is $k \times 1$.

In three previous robustness studies with Robust ML, Chou et al. (1991), Finch et al. (1997), and Hoogland (1999) used the Vale and Maurelli method for data generation and problems were not reported. The Vale and Maurelli method based on Fleishman formulas was used to generate the non-normal data in the present study.

The simulation study had 384 cells (the combination of four sample sizes, four distributional characteristics, three models, four estimation methods, and two methods for setting the scale).

Estimation Methods

As stated in Chapter 1, four estimation methods were studied:

- Maximum Likelihood method (ML)
- Robust Maximum Likelihood method (Robust ML)
- Generalized Least Squares method (GLS)
- Robust Generalized Least Squares method (Robust GLS)

Sample Sizes

As stated in Chapter 1, 200 was chosen as the minimum sample size to avoid problems of non-convergence and improper solution. The sample sizes that were studied are 200, 400, 800, and 1200. The sample size, 200 is considered to be small, 400 and 800 are considered to be medium, and 1200 is judged to be large. The sample sizes 200 and 400 were often investigated in past robustness studies and are also often in this range in observational research according to Table 1-1, Chou et al. (1991), and Hoogland (1999). The sample size of 1200 was chosen to investigate whether there will be important improvements from sample size 800 to 1200.

Distributional Characteristics

The degrees of non-normality are reflected by various skewnesses and kurtoses of the variables. In previous studies, several classifications for categorizing the degree of non-normality for simulation were introduced (Harlow, 1985; Curran et al., 1996; Muthen & Kaplan, 1985, 1992; Potthast, 1993; Hoogland, 1999). In the present study, four different levels of non-normality were selected to reflect distributions thought to be common in the social sciences:

- Zero skewness and zero kurtosis (normal distribution)
- Zero skewness and equal negative kurtosis (platykurtic): Each variable's kurtosis is -1.15.
- Zero skewness and equal positive kurtosis (leptokurtic): Each variable's kurtosis is 3.
- High equal skewness and high equal kurtosis: Each variable's skewness is 2 and kurtosis is 6.

The baseline case of zero skewness and zero kurtosis affords opportunity for comparison and contrast with zero skewness but negative kurtosis (-1.15) , representing almost uniform distribution (platykurtic), a peaked (leptokurtic) distribution with zero skewness but positive kurtosis (3), and a distribution with both high skewness (2) and kurtosis (6). The effects of non-normality on standard errors have not been investigated in Robust GLS. Thus, these non-normal distributional conditions are studied in the present study.

Model Characteristics

Two factors with three indicators per factor, three factors with three indicators per factor for hybrid model analysis, four factors with three or four indicators per factor, and five factors with three indicators per factor for CFA and hybrid models were studied in previous robustness studies of Robust ML (Chou et al., 1991; Finch et al., 1997; Hoogland, 1999).Three CFA models were chosen as population models in the present study. The CFA model 3F3I consists of three factors with three indicators per factor, the CFA model 3F6I consists of three factors with six indicators per factor, and the CFA model 6F3I includes six factors with three indicators per factor. Three-factor and six-factor CFA models were chosen, because these numbers of factors cover the range typically found in CFA studies. For the models 3F3I and 6F3I, the three factor loadings for each factor were 0.6, 0.7, and 0.8. For the model 3F6I, the six factor loadings for each factor were 0.6, 0.6, 0.7, 0.7, 0.8, and 0.8. In addition, for each population model the factor correlations were 0.3, because this is most often chosen in the previous studies and also often found in CFA studies. This study focuses on increasing the number of factors and indicators for model characteristic and excludes the effects of factor loadings and factor correlation. The average of the factor loadings in each CFA model is 0.7.

Analytical Model Characteristics

The data matrix typically used for computations in SEM programs is a covariance matrix, and also the general rule is that a covariance matrix should be analyzed. However, in many social/behavioral sciences applications the scales of the observed variables are often arbitrary or irrelevant. Therefore in many cases a correlation matrix is analyzed instead of a covariance matrix for convenience and interpretational purposes. The analysis of a correlation matrix may cause the following problems (Cudeck, 1989; Jöreskog & Sörbom, 1996): (a) modification of the model being analyzed, (b) incorrect values of test statistic, and (c) incorrect standard errors.

Boomsma (1983) concluded that the analysis of correlation matrices lead to imprecise values for the parameter estimates in a structural equation model. She specifically found a problem with the estimation of standard errors for the parameter estimates. Browne (1982) pointed out that parameter estimates and chi-square statistic are unaffected when a model is scale invariant under the analysis of a correlation matrix. Suggested corrections for the standard errors when correlations for standardized coefficients are used have been recommended by Browne

(1982). For these reasons, a covariance matrix was analyzed and the same unit of observed variables across factors was assumed in this study.

The scale of measurement for a factor can be changed without affecting the fit of the model to the data, which is called ‘scale indeterminacy’. The data cannot determine the values of the parameters due to scale indeterminacy and therefore this is an identification problem. Arbitrary restrictions are required to remove scale indeterminacy. Two methods to solve the scale indeterminacy were investigated: factor variances are set equal to one, and factor loadings are set equal to one. In all conditions the first observed variable has its loading set equal to one for the scale-setting method: factor loadings equal to one.

Number of Replications

The number of replications determines the accuracy with which sampling distributions of parameters estimates are approximated. The larger number of replications the more closely the sampling distributions will be approximated. The previous robustness studies on SEM have often used 100, 200, and 300 replications, according to Table 1-6. Four-hundred replications is the largest value in the past robustness studies of Robust ML (Finch et al., 1997). In the present study, the number of replications, 1000 was chosen to obtain precise simulation results.

Research Summary Statistics

In the present study, the following research summary statistics were used:

- Empirical standard error - The empirical standard error is the standard deviation of the parameter estimates.
- Standard errors - LISREL gives estimates of the asymptotic standard errors of parameter estimates in each sample. These can be compared with the empirical standard errors to see if standard errors are overestimated or underestimated.
- Standard error ratio – More detailed information about the standard errors can be obtained by studying standard error ratio, i.e., the ratios of the standard errors to the empirical standard error (Jonsson, 1997).

Standard error ratios were used instead of mean absolute relative bias for standard error estimation because mean absolute relative bias cannot detect underestimation or overestimation of standard error estimates nor can it detect variation in bias in the standard errors across different parameters and/or parameter values. Thus standard error ratio estimates were compared for all combinations of sample size, degree of non-normality, model, estimation method, and scale-setting method.

In addition, although all methods provide consistent estimates of model parameters, at some sample sizes estimation bias may vary over estimation methods and there may be an interactive effect of estimation method with one or more of the other factors in the study. For this reason, the bias of the parameter estimation was also investigated. For the reasons cited in regard to standard errors, parameter estimates were compared instead of mean absolute relative bias for parameter estimation. Thus, parameter estimates were compared for all combinations of sample size, degree of non-normality, model, estimation method, and scale-setting method.

Ways to Summarize and Present Results

Because Monte Carlo simulations can produce a great deal of data, data summary and presentation of results are important features in simulation studies. There are three main ways to present results: descriptive, graphical, and inferential (Paxton et al., 2001), though clearly inferential procedures can be used to supplement descriptive and graphical presentation of results. Descriptive statistics (i.e., mean, variance, mean relative bias, correlation, etc.) present information concisely and simply. Graphical techniques include using figures such as box plots, scatter plots, or power curves. Inferential techniques allow for the formal testing of the significance of the design factors as well as the computation of various effect size estimates.

The inferential approach has the following advantages:

- The statistical significance and importance of the main and interaction effects of the factors in the design can be determined, and the presentation can focus on effects that are both statistically significant and important effects.

- Designation of effects as significant and important is based on objective criteria.

- Selecting effects that are significant and important for presentation can allow for concise summarization of a large amount of data.

In the present study, inferential techniques such as Welch-James test and eta squared statistic were used because of the advantages above.

CHAPTER 3
MONTE CARLO STUDY RESULTS

A Monte Carlo simulation study was conducted to examine the effects of five factors on parameter estimates and standard error estimates for CFA model. These factors were:

1. CFA model complexity, with three levels: a model with three factors and three indicators per factor (Model 3F3I), a model with three factors and six indicators per factor (Model 3F6I), and a model with six factors and three indicators per factor (Model 6F3I).

2. Sample size with four levels: 200, 400, 800, and 1200.

3. Distribution with four levels identified by the skewness (γ_1) and kurtosis (γ_2) of the distribution used for all indicators within a model: $(\gamma_1, \gamma_2) =$ (0, 0), (0, -1.15), (0, 3) and (2, 6). The (0, 0) distribution is a standard normal distribution and the multivariate distribution is also normal. The (0, -1.15) distribution is a short-tailed distribution, the (0, 3) distribution is a long-tailed distribution, and the (2, 6) distribution is a skewed and long-tailed distribution.

4. Estimation method with two levels (GLS and ML) that could affect estimates and four levels (GLS, Robust GLS, ML, and Robust ML) that could affect standard errors. The robust procedure and a non-robust procedure (GLS and Robust GLS; ML and Robust ML) within the same estimation method produce exactly the same parameter estimates, whereas GLS, Robust GLS, ML, and Robust ML generate different standard error estimates.

5. Scale-setting method with two levels: setting factor variances to unity and setting factor loadings to unity.

The first three factors are between-subjects factors. For each combination of levels of the between-subjects factors 1000 replications were conducted.

Preliminary ANOVA Tests

For Model 3F3I, for example, the population value for three of the factor loadings is 0.6. Evidence that the results are similar for the three indicators for which the factor loading is 0.6 would permit simplification of the analyses and reporting, as an analysis of the estimates and standard errors for only one of the three parameters would be sufficient. This approach can also be applied to the three indicators for which the factor loadings are 0.7, to the three indicators for which the factor loadings are 0.8, and also to other parameters such as factor correlations and residual variances with the same population value. Due to the likelihood that estimates have a non-zero sampling correlation for a particular parameter, repeated measures ANOVA tests were conducted to investigate differences among the means of parameter estimates with the same population value for each of the combinations of the factors; ANOVA tests were also conducted for the means of standard error estimates.

The results of the ANOVA tests for parameter estimates are presented in Table 3-1 and 3-2. The results are the frequency of significance differences among mean parameter estimates per 16 combinations of distribution and sample size within each combination of parameter value, model, and estimation method. In all cases the frequency was three or smaller. Even when the ANOVA test was significant, the means of the parameter estimated were very similar. For example, when the population factor loading was 0.6 in Model 3F3I, GLS was used, and factor variances were set equal to one, differences among the three estimates were significant in three of the 16 combinations. The smallest p value [$F(1.99, 1989.81] > 3.51$, $p = 0.0303$] occurred for the sample size of 200 in combination with the (0, -1.15) distribution. Mean parameter estimates for this condition are presented in Table 3-3. Inspection of the results in Table 3-3 indicates that the means of parameter estimates with the same population value of 0.6 are similar although the ANOVA test resulted in significant differences.

The results of the ANOVA tests for standard error estimates are presented in Table 3-4 and 3-5. The number of significant tests of mean equality of standard error estimates per 16 combinations of distribution and sample size in each combination of parameter value, model, and estimation method is two or fewer. The means of standard error estimates for parameters with the same population value were similar even when the F test was significant. For standard error estimates of factor correlation in Model 6F3I, Robust ML, and setting factor variances to one as an example, the numbers of significance are two in Table 3-6. The combination of a (2, 6) distribution and a 1200 sample size resulted in the smallest p value [$F(13.64, 11033.76) > 2.34$, p = 0.0034]. Mean parameter estimates for this combination of distribution and sample size are reported in Table 3-6. Inspection of the results in Table 3-6 indicates that the means of standard error estimates with the same population value of 0.3 are quite similar, differing only at the thousandths place even though the ANOVA test resulted in a significant difference.

In sum, the results indicate that estimation by a particular estimation method (e.g., ML) of the same parameter type (e.g. a factor loading), with the same value results in mean estimates and standard errors that are very similar. Therefore, in the following section, results are presented for just one of the parameters of the same type and value.

Welch-James Tests and Eta Squared Statistics

Using results from Johansen (1980), Keselman et al. (1993) showed how to apply the Welch-James (WJ) test to repeated measures designs and presented evidence that this method provides better control of Type I error rates than does MANOVA for repeated measures when covariance matrices are unequal across levels of the between-subjects factors. In the present study covariance matrix heterogeneity for the parameter estimates would certainly occur for the various sample sizes and therefore the WJ test was used to test hypotheses about parameter estimates. In testing hypotheses using the WJ method, a saturated means model was used and all possible main effects and interactions were tested.

Consider the linear model (Lix and Keselman, 1995):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi},$$

where $\mathbf{Y}$ is an $N \times p$ matrix of scores on p dependent variables or p repeated measurements, N is the total sample size, $\mathbf{X}$ is an $N \times r$ design matrix consisting entirely of zeros and ones with rank(**X**) = r, $\boldsymbol{\beta}$ is an $r \times p$ matrix of population means, and $\boldsymbol{\xi}$ is an $N \times p$ matrix of random error components. Let $\mathbf{Y}_j$ ($j = 1, \ldots, r$) denote the $n_j \times p$ submatrix of **Y** containing the scores associated with the n_j subjects in the j^{th} group (cell). In the present it is typically assumed that the rows of $\mathbf{Y}_j$ are independently and normally distributed, with mean vector $\boldsymbol{\beta}_j$ and variance-covariance matrix $\boldsymbol{\Sigma}_j$ [i.e., $N(\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j)$], where $\boldsymbol{\beta}_j = (\mu_{j1} \ldots \mu_{jp})$, the j^{th} row of $\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}_j \neq \boldsymbol{\Sigma}_{j'}$ ($j \neq j'$). Specific formulas for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_j$, as well as an elaboration of $\mathbf{Y}_j$ are provided in Lix and Keselman (1995).

In the present study, there were three between-subjects factors: sample size with four levels, distribution with four levels, and model with thee levels. Each combination of these factors constitutes a cell. Consequently there were 48 cells. Each cell was replicated 1000 times. So $N = 48000$. In each WJ analysis, estimation method was the only repeated measure. For analysis of bias of parameter estimates, ML and GLS estimates were analyzed. So $p = 2$. For analysis of standard errors, ML, Robust ML, GLS, and Robust GLS estimates were analyzed. So $p = 4$. In all WJ analyses all possible main effects and interactions of between-subjects factors were included in the model. So $r = 48$.

The general linear hypothesis is

$$H_0 : \mathbf{R}\boldsymbol{\mu} = \mathbf{0},$$

where $\mathbf{R} = \mathbf{O} \otimes \mathbf{U}^{\mathrm{T}}$, **O** is a $df_O \times r$ matrix which controls contrasts on the independent groups (between-subjects) effect(s) with rank(**O**) = $df_O \leq r$, **U** is a $p \times df_U$ matrix which controls contrasts on the within-subjects effect(s) with rank(**U**) = $df_U \leq p$, $\otimes$ is the Kronecker or direct product function, and superscript, 'T' is the transpose operator. For multivariate independent groups designs, **U** is frequently an identity matrix of dimension *p* (i.e., $\mathbf{I}_p$). The **R** contrast matrix has $(df_O)\times(df_U)$ rows and $r\times p$ columns. In the equation of the hypothesis above $\boldsymbol{\mu} = \text{vec}(\boldsymbol{\beta}^{\mathrm{T}}) = [\boldsymbol{\beta}_1 \ldots \boldsymbol{\beta}_{\mathrm{r}}]^{\mathrm{T}}$. In other words, $\boldsymbol{\mu}$ is the column vector with $r\times p$ elements obtained by stacking the columns of $\boldsymbol{\beta}^{\mathrm{T}}$. The zero ($\mathbf{0}$) column vector of the equation of the hypothesis is of order $df_O \times df_U$ [See Lix & Keselman (1995) for illustrative examples].

The generalized test statistic given by Johansen (1980) is

$$\mathbf{T}_{\mathrm{WJ}} = (\mathbf{R}\hat{\boldsymbol{\mu}})^{\mathrm{T}}(\mathbf{R}\hat{\boldsymbol{\Sigma}}\mathbf{R}^{\mathrm{T}})^{-1}(\mathbf{R}\hat{\boldsymbol{\mu}}),$$

where $\hat{\boldsymbol{\mu}}$ estimates $\boldsymbol{\mu}$, and $\hat{\boldsymbol{\Sigma}} = \text{diag}[\hat{\boldsymbol{\Sigma}}_1 / \mathrm{n}_1 \ldots \hat{\boldsymbol{\Sigma}}_{\mathrm{r}} / \mathrm{n}_{\mathrm{r}}]$, a block matrix with diagonal elements $\hat{\boldsymbol{\Sigma}}_{\mathrm{j}} / \mathrm{n}_{\mathrm{j}}$. This statistic, divided by a constant, *o* (i.e., $\mathrm{T}_{\mathrm{WJ}} / o$), approximately follows an *F* distribution with $v_1 = df_O \times df_U$ and $v_2 = v_1(v_1 + 2) / (3\mathrm{A})$, where $o = v_1 + 2\mathrm{A} - (6\mathrm{A}) / (v_1 + 2)$. The formula for the statistic A is given by (Lix and Keselman, 1995)

$$\mathrm{A} = \frac{1}{2}\sum_{\mathrm{j}=1}^{\mathrm{r}}[\mathrm{tr}\{\hat{\boldsymbol{\Sigma}}\mathbf{R}^{\mathrm{T}}(\mathbf{R}\hat{\boldsymbol{\Sigma}}\mathbf{R}^{\mathrm{T}})^{-1}\mathbf{R}\mathbf{Q}_{\mathrm{j}}\}^2 + \{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{R}^{\mathrm{T}}(\mathbf{R}\hat{\boldsymbol{\Sigma}}\mathbf{R}^{\mathrm{T}})^{-1}\mathbf{R}\mathbf{Q}_{\mathrm{j}})\}^2] / (\mathrm{n}_{\mathrm{j}} - 1),$$

where '*tr*' is the trace operator and $\mathbf{Q}_{\mathrm{j}}$ is a symmetric block matrix of dimension $r\times p$ associated with $\mathbf{X}_{\mathrm{j}}$, such that the (*s*, *t*) th diagonal block of $\mathbf{Q}_{\mathrm{j}} = \mathbf{I}_{\mathrm{p}}$ if $s = t = j$ and is zero matrix (**0**) otherwise.

The scale-setting method is known to affect the size of the factor loading, factor variances, and factor covariances. As an example, consider the effect of scale-setting method on the factor loadings. Recall that the factor loadings were 0.6, 0.7, or 0.8 when the scale was set by specifying factor variances equal to unity. When the scale was set by specifying unit factor loadings, the loading for the first factor (i.e., the factor that had a 0.6 loading when unit factor variances were specified) was set equal to one. Thus the other two factor loadings became

0.7/0.6 and 0.8/0.6, respectively. Including scale-setting method as a factor in the analysis would results in artificial effect of scale setting method on the estimates and perhaps artificial interactions of the other factors with scale setting method. Therefore the data obtained by setting factor variances equal to one and factor loadings equal to one were analyzed separately. However, the scale-setting method does not affect the size of the residual variances at all and the results obtained by setting factor variance equal to one and factor loadings equal to one are exactly the same. Thus only the data obtained by setting factor variances equal to one was analyzed and presented in the present study.

For the data when factor variances were set equal to one, seven WJ analyses were conducted: three were for results estimating the 0.6, 0.7, and 0.8 factor loadings, respectively; one was for the results estimating the 0.3 factor correlation, and three were for the results estimating the 0.64, 0.51, and 0.36 residual variances, respectively. For the data when factor loadings were set equal to one, four WJ analyses were conducted: two were for results estimating the 0.7/0.6 and 0.8/0.6 factor loadings, respectively, one was for the results estimating the 0.36 factor variance, and one was for results estimating the 0.108 factor covariances. Residual variances are not affected by the scale-setting method, so analyses of the residual variance estimates were not conducted again.

Because of the large sample size, it was expected that many effects would be significant even if the effects were quite small. To address this problem an effect-size measure was adopted. An effect-size measure is a standardized index, estimates a parameter that is independent of sample size, and quantifies the magnitude of the difference between populations or the relationship between explanatory and response variables (Olejnik & Algina, 2003). Two broad categories of effect size are standardized mean differences and measures of association or the proportion of variance explained (Olejnik & Algina, 2003). In the present study eta squared was used as an effect size. Eta squared is the proportion of the total sums of squares that is attributed to an effect:

$$\hat{\eta}^2 = \frac{SS_{Effect}}{SS_{Total}} \quad ,$$

where SS_{Effect} is the sum of squares for the effect for which the effect size is being estimated and SS_{Total} is the total sum of squares. These sums of squares were calculated by using an ANOVA for repeated measures. The coefficient $\hat{\eta}^2$ was used rather than $\hat{\omega}^2$. (Omega squared is an estimate of the dependent variable population variability accounted for by the independent variable.) Because the latter is based on the assumption of covariance homogeneity and compound symmetry but the former is not. Compound symmetry is an assumption about the variances and covariance of the repeated measures and all correlations are equal under compound symmetry.

Effects on Parameter Estimates

Tables 3-7 and 3-8 contain results for estimation when the scale was set by specifying factor variances equal to one. Table 3-7 presents the *F* statistics for the WJ tests and Table 3-8 presents eta squared statistics for all effects with significant *F* statistics in Table 3-7. Inspection of the results in Table 3-8 indicates that most effects were quite small, with only the effect of estimation method on estimates of the residual variance accounting for as much as approximately 5% of the total sum of squares. To select effects for further examination the following rules were applied. For each column, effects which did not enter into higher order interactions and for which eta squared was at least 0.01 were selected.

Applying the rule, no factors or interactions had important effects on estimation of the factor correlation and only estimation procedure had an important effect on estimates of the factor loadings. This effect occurred only when the population factor loading was 0.8. Table 3-9 contains marginal means of factor loading estimates, classified by estimation method, for a population factor loading equal to 0.8. Inspection of Table 3-9 suggests the effects of estimation method on estimates of the factor loadings were trivial in size, although GLS estimates have more bias than ML estimates.

Several factors had important effects on estimates of the residual variance. Table 3-10 contains the means by estimation and sample size for estimates of population residual variances. The results show that ML estimates have minimal bias at all sample sizes; whereas GLS estimates have more substantial bias when sample size is 200 or 400 and suggest the GLS estimates are not unbiased even when the sample size is 1200.

Table 3-11 presents the *F* statistics for the WJ tests and Table 3-12 presents eta squared statistics for all effects with significant *F* statistics in Table 3-11. Both tables contain results for estimation when the scale was set by specifying factor loadings equal to one. Inspection of the results in Table 3-12 indicates that most effects on parameter estimates of factor loadings and factor correlations were quite small and none of them did accounted for even 1% of the total sum of squares.

Because none of the effect accounted for even 1% of the total sum of squares, Table 3-13 presents the grand mean estimates of parameter by population parameter value. The results indicate that the grand means are very similar to their population values.

Effects on Standard Error Ratio Estimates

LISREL provides estimates of the asymptotic standard errors of parameter estimates in each sample. These asymptotic standard errors are called standard errors in the present study. An empirical standard error is the standard deviation of the distribution of the parameter estimates. For each population parameter value, empirical standard errors were calculated for each combination of the five factors. An empirical standard error provides an estimate of the actual standard error for a particular combination of population parameter value and the five factors, an estimate that does not rely on large sample theory for its validity. The ratio of the standard errors to the empirical standard error can be used to determine if the standard error overestimates or underestimates the sampling variability of the estimates. The WJ tests for the standard error ratios were conducted and eta squared statistics for the standard error ratios were calculated in the present study.

Tables 3-14 and 3-15 contain results for standard error ratios when the scale was set by specifying factor variances equal to one. Table 3-14 presents the F statistics for the WJ tests and Table 3-15 presents eta squared statistics for all effects with significant F statistics in Table 3-14. For standard error ratio estimates, every effect was significant in the WJ tests as shown in Table 3-14. To select effects for further examination the same rules used for the effects on parameter estimates were applied.

Table 3-15 indicates that when the population factor loading was 0.6, sample size accounted for more than 1% of the sums of squares. Table 3-16 contains marginal means of standard error ratios for factor loading estimates, classified by sample size, for a population factor loading equal to 0.6. As the sample size increases the bias of standard error estimates of the factor loadings decreases. Although not included, tables of marginal means when the factor loading was 0.7 or 0.8 show a similar pattern. Inspection of Table 3-16 suggests the effects of sample size on standard error estimates of the factor loadings is consistent with theory that standard error estimates will converge to their true values as the sample size increases without bound.

As shown in Tables 3-14 and 3-15, for each factor loading population value, the interaction of estimation method and distribution is significant and accounts for a substantial portion of the total sums of squares for the standard error ratio. Table 3-17 contains mean standard error ratios of factor loading estimates by estimation method and distribution. The results show that standard errors of ML estimates have minimal bias only when the distribution is normal, whereas Robust ML standard error estimates have a relatively small bias regardless of the distribution. GLS standard error estimates have about the same bias as ML estimates, with slightly more bias for the normal distribution and slightly less for the short-tailed distribution (0, -1.15). Robust GLS estimates of standard errors tend to have slightly more bias than do robust ML estimates. For each estimation method, the standard error ratio varies as a function of the distribution. The mean standard error ratio tends to be the highest for the (0, -1.15) distribution and indicates overestimation when ML or GLS was used and slight underestimation when the robust variants were used. With the (2, 6) distribution, the mean standard error ratios tend to be the smallest. In particular GLS and ML standard errors are gross underestimates with this distribution. When the distribution is long tailed (0, 3) estimates also tend to be too small and the ML and GLS estimates strongly underestimate the standard errors.

The results of the WJ tests and eta squared statistics indicated a significant three-way interaction among sample size, model, and the distribution on the standard error ratio for the factor correlation, accounting for 5% of the total sum of squares. The means for the standard error ratios of factor correlation as a function of sample size, model, and distribution are presented in Table 3-18. The standard error ratio varies as a function of the sample size within each model and distribution. The bias of the standard error estimates tends to decrease as the sample size increases for the (2, 6) distribution and the standard error estimates in the (2, 6) distribution tend to be underestimated regardless of the model. Inspection of Table 3-18 indicates that for the other distributions the bias tends to decrease as the sample size increases from 200 to 400, but the effect of further increases of sample size on bias is irregular, particularly in the normal distribution and (0, -1.15) distribution. Shown in Appendix A are empirical standard errors and mean standard errors and mean standard error ratios, by model, sample size, and distributions. Results are presented in four tables, one for each estimation method. The results show that the empirical standard errors and mean standard errors decreased as the sample size increased from 200 to 1200. However as the ratio of the mean standard error to empirical standard error approached 1.0 as a function of sample size, relatively small differences between the mean standard error and empirical standard error resulted in ratios that varied around 1.0. For example, with ML estimation, Model 3F3I and a normal distribution, the mean standard error and empirical standard error were 0.0607274 and 0.0598677 $(N = 400)$, 0.0430910 and 0.0448250 $(N = 800)$, and 0.0351752 and 0.0341341 $(N = 1200)$, resulting in standard error ratios of 1.014, 0.961, and 1.031, respectively. Thus, although at first glance the behavior of the standard error ratio does not seem to be consistent with the expectation that the bias of standard errors would decline as the sample size increased, it seems more likely that the results mean that for all distributions except the (2, 6) distribution, the standard error became nearly unbiased when the sample size was 400 and the variation of the standard error ratio around 1.0, is due to sampling error.

The interaction of distribution and estimation method accounted for approximately 8% of the total sums of squares in the standard error ratio for the factor correlation. The means for the standard error ratios of factor correlations as a function of distribution and estimation method are presented in Table 3-19. When the distribution was normal, (0, -1.15) , or (0, 3), ML and Robust ML showed better performances than GLS and Robust GLS in the bias of standard error estimates, whereas the robust estimation procedures showed better performances than the non-robust estimation procedures for the non-normal distributions, with a strong advantage for the (2, 6) distribution.

Tables 3-14 and 3-15 indicate a significant main effect of sample size on the standard error ratio that accounts for 1% of the total sums of squares when the residual variance was 0.36. Table 3-20 contains the marginal means for this main effect. Inspection of Table 3-20 suggests the effects of sample size on standard error estimates of the residual variance are consistent with theory that estimates will converge to their true values as the sample size increases without bound. That is, the bias of standard error estimates decreases as the sample size increases. Similar results, not tabled in the dissertation, were found when the population residual variance was 0.51 or 0.64.

When the residual variance was 0.36, there was a significant Model × Distribution interaction that accounted for 1% of the total sums of squares. The mean standard error ratios are presented in Table 3-21. Inspection of the result in Table 3-21 indicates that the bias of the standard error estimates increase as the degree of non-normality increases regardless of model type, but the effect of non-normality appears to be stronger when there are more indicators. Similar results, not tabled in the dissertation, were found when the population residual variance was 0.51 or 0.64.

The results of WJ tests indicated a significant two-way interaction between the estimation method and distribution accounting for 20% of the total sum of squares when the residual variance was 0.64, 14% when the residual variance was 0.51, and 10% when the residual variance was 0.36. The means for the standard error ratios of residual variances as a function of estimation method and distribution are presented in Table 3-22. For the normal distribution, ML and Robust ML showed better performances than GLS and Robust GLS in the bias of standard error estimates, whereas the robust estimation procedures showed better performances than the non-robust estimation procedures for the non-normal distributions. Specially, ML has the minimal bias of standard error estimates for the normal distribution, whereas Robust ML has the smallest bias of standard error estimates for the non-normal distributions. For the distributions with positive kurtosis (0, 3), and high skewness and high kurtosis (2, 6), the bias of standard error estimates tends to decrease as the population parameter values (0.64, 0.51, and 0.36) decrease.

Tables 3-23 and 3-24 contain results for standard error ratios when the scale was set by specifying factor loadings equal to one. Table 3-23 presents the *F* statistics for the WJ tests, and Table 3-24 presents eta squared statistics for all effects with significant *F* statistics in Table 3-23. Inspection of the results in Table 3-24 indicates that most effects on standard error ratio estimates of factor loadings and factor correlations were quite small compared with the effects of distribution and the two-way interaction between estimation and distribution. To select effects for further examination the rules used when factor variances were set equal to one were applied. For the residual variance, there is no difference between both scale-setting methods because the scale-setting methods do not affect the estimates of residual variances.

Tables 3-23 and 3-24 indicate that the Estimation × Distribution interaction was significant for both factor loading population values and accounted for 11 % and 9 % of the total sum of squares when the factor loading was $0.7/0.6$ and $0.8/0.6$, respectively. The means for the standard error ratios of factor loadings as a function of estimation and distribution, classified by parameter value, are presented in Table 3-25. Differences between the robust estimation procedures and the non-robust estimation procedures are very small for the normal distribution. For the non-normal distribution, the robust estimation procedures showed better performance rather than the non-robust procedures. For the (0, -1.15) distribution (negative kurtosis), the standard error estimates tend to be overestimated by the non-robust estimation procedures and underestimated by the robust estimation procedures. For the other distributions, the standard error estimates tend to be underestimated regardless of the estimation procedure.

Tables 3-23 and 3-24 indicate that the sample size effect was significant for the factor covariance and accounted about 1.5 % of the total sum of squares. Table 3-26 contains marginal means of standard error ratio estimates for factor covariance parameter, classified by sample size, for the factor covariance. Inspection of Table 3-26 suggests the effects of sample size on standard error estimates of the factor covariance consistent with theory that bias will decrease the sample size increases and 1200 sample size had the minimal bias of standard error estimates. Although means by sample size are not reported for the other parameters since eta squared statistics are less than 1% of the total sum of squares, results show the same pattern of results as in Table 3-26: less bias with increasing sample size accounting for rounding 1% of the total sum of squares.

Tables 3-23 and 3-24 indicate that the Estimation × Distribution interaction was significant for both elements of the factor covariance matrix and accounted 14% of the total sum of squares for the factor variance and 5% of the total sum of squares for the factor covariance. The means for the standard error ratios of factor correlations as a function of estimation method and distribution are presented in Table 3-27. For the normal distribution, ML and Robust ML showed better performances than GLS and Robust GLS in the bias of standard error estimates, whereas the robust estimation procedures showed better performances than the non-robust estimation procedures for the non-normal distributions. Robust ML has the smallest bias of standard error estimates for the non-normal distributions. There are the consistent patterns between factor variances and factor covariances with different parameter values (0.108 and 0.36), that is, the means of standard error ratio estimates in factor variances and factor covariances showed similar behaviors.

Practical Problems in Estimation

As noted in Chapter 1, estimates in SEM are found by minimizing a discrepancy function. Because a direct algebraic solution to finding the minimizer is not available for all models, estimation in SEM is generally carried out by iteration. As a result, non-convergence can be encountered when SEM is used. The iterative process is said to have converged when the criterion for convergence is met. In ML estimation in LISREL, the criterion for convergence is a change in the likelihood from one iteration to the next that is smaller than 0.000001 (Jöreskog and Sörbom, 1996). The maximum number of iterations is set a priori and non-convergence in LISREL occurs when the change in the likelihood next to last and last iteration is larger than 0.000001. Another possible problem is a non-positive definite matrix of second-order partial derivatives of the discrepancy function, a matrix called the Hessian matrix. The Hessian matrix is used to diagnose identification problems and to compute standard errors. If the Hessian matrix is not positive definite the solution is classified as non-identified and, in addition, standard errors cannot be computed. In addition, unless the iterative procedure is constructed to keep variance and covariance estimates within the bounds of proper estimates, estimates outside these bound can occur, for example, negative residual variances. Such estimates are called improper estimates in the present study. For the details on the problems, potential solutions, and limitations to these potential solutions, see Chen et al (2001), Bentler and Chou (1987), and Wothke (1993).

Previous studies reported non-convergence mainly for the ML estimation method. Non-convergence of ML estimation method decreased when the sample size, the number of indicators per factor, or the size of the factor loadings increased (Anderson & Gerbing, 1984; Boomsma, 1985). Also Hoogland (1999) reported that none of the estimation methods had substantial convergence problems for $N \geq 200$, and for $N = 200$, the percentage of improper solutions was considerable. This percentage increased when the size of the factor loadings or the number of indicators per factor decreased.

While conducting the simulation the following problems occurred in some replications of some conditions: (a) non-convergence in 10,000 iterations, (b) improper estimates (e.g., a negative residual variance), and (c) non-positive definite Hessian matrix (so standard errors could not be computed). The frequency of non-convergence, improper estimates, and non-positive definite Hessian matrix for certain combinations of model, distribution, sample size, scale-setting method, and estimation is presented in Table 3-28. Non-convergence did not occur for the normal distribution at all and mainly occurred for sample sizes such of 200 and 400 for the non-normal distributions. For the 3F3I model, non-convergence occurred only for ML and Robust ML, but occurred at a very low frequency and never when the sample size was 800 or 1200. For the 3F6I model, non-convergence, improper estimates, and non-positive definite Hessians occurred for all estimation procedures, but only when the scale was set by specifying factor loadings equal to one, and again at a very low frequency and never when the sample size was 800 or 1200. Increasing the number of indicators per factor caused more non-convergence, improper estimates, and non-positive definite Hessian matrices. However, setting the scale by specifying factor variances equal to one reduced the problems of non-convergence, improper estimates, and non-positive definite Hessians. For the model 6F3I, non-convergence, improper estimates, and non-positive definite Hessians occurred only for GLS and Robust GLS and at a very low frequency. There was no non-convergence problem at all when the scale was set by specifying factor variances equal to one. As has been reported in previous studies (e.g., Boomsma & Hoogland, 2001) degree of non-normality, the number of indicators per factor, and the sample size played important roles in whether non-convergence, improper estimates, and non-positive definite Hessians occurred. In addition, the present study suggests that the scale-setting method has an important effect under non-normality.

Table 3-1. Frequency of significance *F* tests on parameter estimates per 16 combinations of distribution and sample size: factor variances equal to one.

Model	Parameter value	Estimation method ML	Estimation method GLS
3F3I	λ (0.6)	2	3
	λ (0.7)	1	1
	λ (0.8)	1	0
	ϕ (0.3)	1	1
	θ_δ (0.64)	0	0
	θ_δ (0.51)	1	1
	θ_δ (0.36)	0	0
3F6I	λ (0.6)	1	2
	λ (0.7)	0	0
	λ (0.8)	0	0
	ϕ (0.3)	1	2
	θ_δ (0.64)	1	1
	θ_δ (0.51)	1	1
	θ_δ (0.36)	0	0
6F3I	λ (0.6)	0	1
	λ (0.7)	1	0
	λ (0.8)	0	0
	ϕ (0.3)	1	1
	θ_δ (0.64)	1	1
	θ_δ (0.51)	1	0
	θ_δ (0.36)	1	1

Note. λ indicates factor loading, ϕ: factor correlation, and θ_δ: residual variance.

Table 3-2. Frequency of significant *F* tests on parameter estimates per 16 combinations of distribution and sample size: factor loadings equal to one.

Model	Parameter Values	Estimation method	
		ML	GLS
3F3I	λ (0.7/0.6)	0	0
	λ (0.8/0.6)	0	1
	ϕ (i,i) = (0.36)	1	0
	ϕ (i,j) = (0.108)	1	1
3F6I	λ (0.7/0.6)	2	2
	λ (0.8/0.6)	2	2
	ϕ (i,i) = (0.36)	1	2
	ϕ (i,j) = (0.108)	1	2
6F3I	λ (0.7/0.6)	2	2
	λ (0.8/0.6)	0	0
	ϕ (i,i) = (0.36)	0	1
	ϕ (i,j) = (0.108)	1	1

Table 3-3. Mean parameter estimate for λ = 0.6: Model 3F3I, GLS, N = 200, factor variances equal to one, and (0, -1.15).

Three factor loadings: λ (0.6)	Means of parameter estimates
1	0.593
2	0.593
3	0.586

Table 3-4. Frequency of significant F tests on standard error estimates per 16 combinations of distribution and sample size: factor variances equal to one.

		Estimation method			
Model	Parameter value	ML	RML	GLS	RGLS
3F3I	λ (0.6)	0	0	0	0
	λ (0.7)	1	0	0	0
	λ (0.8)	1	0	1	0
	ϕ (0.3)	0	0	0	0
	θ_δ (0.64)	0	0	0	1
	θ_δ (0.51)	1	1	0	1
	θ_δ (0.36)	2	1	1	1
3F6I	λ (0.6)	1	1	1	1
	λ (0.7)	0	0	1	0
	λ (0.8)	1	0	1	0
	ϕ (0.3)	1	2	1	1
	θ_δ (0.64)	1	2	1	2
	θ_δ (0.51)	1	1	1	1
	θ_δ (0.36)	0	1	0	0
6F3I	λ (0.6)	0	0	1	0
	λ (0.7)	0	0	0	0
	λ (0.8)	2	1	2	1
	ϕ (0.3)	2	2	1	2
	θ_δ (0.64)	1	2	0	1
	θ_δ (0.51)	0	2	0	2
	θ_δ (0.36)	2	1	2	1

Table 3-5. Frequency of significant F tests on standard error estimates per 16 combinations of distribution and sample size: factor loadings equal to one.

Model	Parameter value	Estimation method ML	RML	GLS	RGLS
3F3I	λ (0.7/0.6)	1	1	1	2
	λ (0.8/0.6)	1	1	1	2
	ϕ (i,i) (0.36)	1	2	1	2
	ϕ (i,j) (0.108)	0	1	1	2
3F6I	λ (0.7/0.6)	1	1	2	2
	λ (0.8/0.6)	2	1	1	2
	ϕ (i,i) (0.36)	2	2	2	2
	ϕ (i,j) (0.108)	1	0	1	2
6F3I	λ (0.7/0.6)	1	1	1	1
	λ (0.8/0.6)	1	0	0	0
	ϕ (i,i) (0.36)	0	0	0	0
	ϕ (i,j) (0.108)	1	0	0	0

Table 3-6. Mean standard error estimates for ϕ = 0.3: model 6F3I, Robust ML, N = 1200, factor variances equal to one, and (2, 6).

Fifteen factor correlations: ϕ (0.3)	Means of standard error estimates
SE19	0.0400
SE20	0.0399
SE21	0.0398
SE22	0.0396
SE23	0.0396
SE24	0.0399
SE25	0.0397
SE26	0.0397
SE27	0.0395
SE28	0.0396
SE29	0.0395
SE30	0.0397
SE31	0.0397
SE32	0.0397
SE33	0.0396

Table 3-7. *F* statistics (degrees of freedom) for WJ tests for effects on parameter estimates: scale set by specifying factor variances equal to one.

	Parameter						
	λ (0.6)	λ (0.7)	λ (0.8)	ϕ (0.3)	θ_δ (0.64)	θ_δ (0.51)	θ_δ (0.36)
Between-subjects							
S	46.02 (3, 22506.63)	61.53 (3, 21599.19)	72.73 (3, 21198.98)	1.83 (3, 24825.76)	620.72 (3, 20239.98)	446.12 (3, 21093.75)	312.53 (3, 20896.37)
D	1.66 (3, 17775.07)	5.64 (3, 17700.60)	3.12 (3, 17406.09)	0.62 (3, 17283.86)	14.52 (3, 18013.56)	9.82 (3, 17424.44)	15.52 (3, 15742.61)
M	23.31 (2, 19266.16)	38.43 (2, 18800.59)	45.71 (2,18258.67)	7.72 (2, 20687.19)	219.81 (2, 17766.82)	194.44 (2, 18747.50)	67.91 (2, 18928.08)
S*D	1.0296 (9, 18427.51)	0.82 (9, 18228.70)	0.72 (9, 18073.82)	1.05 (9, 18432.73)	2.32 (9, 18206.46)	1.53 (9, 17812.25)	3.52 (9, 16505.48)
S*M	6.12 (6, 18082.12)	7.33 (6, 17408.11)	8.10 (6, 17138.26)	1.78 (6, 20066.18)	26.30 (6, 16326.06)	23.27 (6, 17217.24)	8.05 (6, 17829.92)
D*M	0.84 (6, 14311.52)	0.46 (6, 14305.69)	0.57 (6, 14157.82)	0.75 (6, 13965.46)	2.21 (6, 14588.77)	2.70 (6, 14366.58)	0.13 (6, 13566.43)
S*D*M	1.16 (18, 16850.26)	0.69 (18, 16726.97)	0.47 (18, 16682.06)	0.99 (18, 16940.70)	0.92 (18, 16705.38)	0.95 (18, 16628.99)	0.96 (18, 16075.00)
Within-subjects							
E	12178.18 (1, 16540.83)	17764.71 (1, 16636.72)	25686.78 (1, 16649.31)	4520.48 (1, 15325.14)	152399.97 (1, 17241.46)	108604.31 (1, 16748.83)	52581.77 (1, 15320.01)
E*S	1524.95 (3, 20451.17)	2281.22 (3, 20465.43)	3231.08 (3, 20501.19)	471.46 (3, 18942.21)	18111.71 (3, 20302.35)	13166.86 (3, 20372.53)	6457.34 (3, 19094.80)
E*D	30.04 (3, 9590.41)	45.51 (3, 9692.51)	85.56 (3, 9709.39)	2.09 (3, 8897.09)	297.00 (3, 10375.78)	274.18 (3, 9880.68)	141.91 (3, 9052.65)
E*M	1298.57 (2, 11765.34)	1884.21 (2, 11864.30)	2434.13 (2, 11985.07)	242.89 (2, 11306.97)	12160.34 (2, 12197.84)	8728.98 (2, 11987.15)	4810.11 (2, 11892.42)
E*S*D	4.93 (9, 15655.19)	5.85 (9, 15702.67)	12.01 (9, 15683.89)	0.62 (9, 14463.13)	29.90 (9, 15870.74)	29.11 (9, 15730.18)	15.06 (9, 14699.42)
E*S*M	162.49 (6, 17485.34)	238.26 (6, 17568.41)	308.24 (6, 17691.47)	25.60 (6, 17204.20)	1410.96 (6, 17274.57)	1040.98 (6, 17475.85)	572.56 (6, 17603.68)
E*D*M	1.57 (6, 8198.83)	3.84 (6, 8276.55)	6.43 (6, 8424.32)	0.71 (6, 7874.98)	37.14 (6, 8788.30)	35.89 (6, 8603.60)	19.99 (6, 8483.09)
E*S*D*M	0.78 (18, 15209.78)	0.79 (18, 15292.40)	1.22 (18, 15427.25)	0.69 (18, 14955.98)	6.36 (18, 15340.25)	5.13 (18, 15485.72)	2.36 (18, 15525.14)

Note. S indicates Sample size, D: distribution, M: model, and E: estimation method. λ indicates factor loading, ϕ: factor correlation, and θ_δ: residual variance.

Table 3-8. Eta squared for significant effects on parameter estimates: scale set by specifying factor variances equal to one.

	Parameter						
	λ (0.6)	λ (0.7)	λ (0.8)	ϕ (0.3)	θ_δ (0.64)	θ_δ (0.51)	θ_δ (0.36)
Between-subjects							
S	0.0033	0.0044	0.0055		0.0397	0.0300	0.0221
D		0.0005	0.0003		0.0010	0.0007	0.0011
M	0.0010	0.0015	0.0018	0.0003	0.0079	0.0074	0.0030
S*D					0.0003		0.0010
S*M	0.0009	0.0010	0.0010		0.0039	0.0034	0.0012
D*M					0.0003	0.0004	
S*D*M							
Within-subjects							
E	0.0053	0.0073	0.0104	0.0019	0.0693	0.0540	0.0286
E*S	0.0024	0.0033	0.0046	0.0005	0.0293	0.0233	0.0123
E*D	0.0001	0.0001	0.0001		0.0005	0.0005	0.0003
E*M	0.0009	0.0012	0.0015	0.0002	0.0085	0.0067	0.0039
E*S*D	0.0000	0.0000	0.0001		0.0002	0.0002	0.0001
E*S*M	0.0004	0.0006	0.0007	0.0001	0.0034	0.0027	0.0016
E*D*M		0.0000	0.0000		0.0001	0.0001	0.0001
E*S*D*M							

Note. S indicates Sample size, D: distribution, M: model, and E: estimation method. λ indicates factor loading, ϕ: factor correlation, and θ_δ: residual variance.

Table 3-9. Marginal means by estimation for estimates of $\lambda = 0.8$

Estimation	Marginal means
GLS	0.788
ML	0.800

Table 3-10. Mean estimates of residual variance parameters by parameter value, estimation method, and sample size.

		Sample size			
Parameter value	Estimation	200	400	800	1200
0.36	GLS	0.310	0.334	0.347	0.351
	ML	0.353	0.356	0.359	0.359
0.51	GLS	0.442	0.475	0.492	0.498
	ML	0.506	0.508	0.509	0.509
0.64	GLS	0.554	0.596	0.617	0.625
	ML	0.634	0.637	0.638	0.639

Table 3-11. *F* statistics (degrees of freedom) for WJ tests for effects on parameter estimates: scale set by specifying factor loadings equal to one.

	Parameter			
	λ (0.7/0.6)	λ (0.8/0.6)	ϕ (i,i) = (0.36)	ϕ (i,j) = (0.108)
Between-subjects				
S	30.28	38.49	8.57	19.33
	(3, 22125.35)	(3, 22189.49)	(3, 22208.53)	(3, 24267.26)
D	4.77	7.66	2.40	0.21
	(3, 16883.50)	(3, 16473.47)	(3, 17582.66)	(3, 17633.72)
M	2.41	4.19	26.54	7.75
	(2, 18254.97)	(2, 18080.90)	(2, 18840.57)	(2, 20592.78)
S*D	1.93	1.41	1.14	0.82
	(9, 18240.52)	(9, 17942.50)	(9, 18356.55)	(9, 18449.22)
S*M	1.69	0.79	6.37	0.95
	(6, 17835.21)	(6, 18077.52)	(6, 17881.18)	(6, 19593.93)
D*M	1.08	0.51	0.86	0.42
	(6, 13645.46)	(6, 13494.24)	(6, 14207.07)	(6, 14255.28)
S*D*M	0.63	1.09	1.19	1.17
	(18, 16747.68)	(18, 16671.06)	(18, 16820.97)	(18, 16942.48)
Within-subjects				
E	2.62	3.20	11807.82	1207.51
	(1, 15632.45)	(1, 14317.90)	(1, 16404.87)	(1, 15341.86)
E*S	1.61	1.90	1444.36	214.33
	(3, 20172.44)	(3, 19755.23)	(3, 20283.60)	(3, 19103.04)
E*D	0.21	0.90	31.49	9.39
	(3, 9092.03)	(3, 8681.75)	(3, 9653.42)	(3, 9082.70)
E*M	1.51	0.44	1243.90	425.01
	(2, 11208.28)	(2, 10793.27)	(2, 11644.45)	(2, 11327.47)
E*S*D	0.89	0.99	5.29	2.45
	(9, 15496.59)	(9, 15387.27)	(9, 15680.82)	(9, 14733.99)
E*S*M	0.91	1.10	151.13	68.39
	(6, 17372.35)	(6, 17429.38)	(6, 17338.56)	(6, 17007.79)
E*D*M	0.54	1.42	1.55	3.37
	(6, 7845.83)	(6, 7776.98)	(6, 8244.30)	(6, 8050.43)
E*S*D*M	0.87	0.78	0.76	1.00
	(18, 15166.03)	(18, 15366.89)	(18, 15233.44)	(18, 14969.96)

Note. S indicates Sample size, D: distribution, M: model, and E: estimation method. λ indicates factor loading, and ϕ: factor correlation.

Table 3-12. Eta squared for significant effects on parameter estimates: scale set by specifying factor loadings equal to one.

	Parameter			
	λ (0.7/0.6)	λ (0.8/0.6)	ϕ (i,i) = (0.36)	ϕ (i,j) = (0.108)
Between-subjects				
S	0.0024	0.0030	0.0005	0.0018
D	0.0003	0.0005		
M		0.0002	0.0011	0.0003
S*D	0.0004			
S*M			0.0009	
D*M				
S*D*M				
Within-subjects				
E			0.0050	0.0005
E*S			0.0022	0.0004
E*D			0.0001	0.0000
E*M			0.0008	0.0002
E*S*D			0.0000	0.0000
E*S*M			0.0003	0.0002
E*D*M				0.0000
E*S*D*M				

Note. S indicates Sample size, D: distribution, M: model, and E: estimation method. λ indicates factor loading, and ϕ: factor correlation.

Table 3-13. Grand mean estimates of parameter by parameter value

Parameter	Grand mean
λ (0.7/0.6 = 1.167)	1.176
λ (0.8/0.6 = 1.333)	1.345
ϕ (i,i) = (0.36)	0.358
ϕ (i,j) = (0.108)	0.107

Table 3-14. *F* statistics (degrees of freedom) for WJ tests for effects on standard error ratio estimates: scale set by specifying factor variances equal to one.

	Parameter						
	λ (0.6)	λ (0.7)	λ (0.8)	ϕ (0.3)	θ_δ (0.64)	θ_δ (0.51)	θ_δ (0.36)
Between-subjects							
S	595.90 (3, 15782.04)	467.62 (3, 16318.87)	379.78 (3, 17343.85)	3603.70 (3, 23369.23)	198.16 (3, 19385.97)	242.22 (3, 20429.42)	363.96 (3, 22444.21)
D	20946.67 (3, 19027.12)	26581.99 (3, 18729.86)	22200.70 (3, 18881.47)	4595.39 (3, 19551.74)	10741.40 (3, 19649.16)	7329.45 (3, 17648.10)	3889.99 (3, 18945.82)
M	225.66 (2, 15794.47)	42.24 (2, 16166.66)	116.43 (2, 17599.89)	273.00 (2, 22385.22)	15.44 (2, 18824.20)	169.66 (2, 19600.56)	218.83 (2, 21405.65)
S*D	48.99 (9, 17638.50)	218.89 (9, 17553.80)	101.65 (9, 17538.81)	194.24 (9, 18642.07)	49.07 (9, 17772.69)	48.22 (9, 17301.84)	96.55 (9, 17901.15)
S*M	87.91 (6, 12666.42)	198.36 (6, 13141.15)	57.97 (6, 13948.56)	68.55 (6, 13948.56)	34.21 (6, 15528.44)	31.56 (6, 16452.99)	30.47 (6, 18161.94)
D*M	135.74 (6, 15268.67)	127.56 (6, 15117.87)	276.67 (6, 15383.29)	263.29 (6, 15693.62)	80.08 (6, 15761.30)	111.48 (6, 14712.34)	234.54 (6, 15264.45)
S*D*M	82.00 (18, 16102.75)	107.39 (18, 16146.65)	137.13 (18, 16225.82)	627.60 (18, 17017.36)	77.87 (18, 16322.40)	42.42 (18, 16228.62)	29.03 (18, 16583.77)
Within-subjects							
E	77257.59 (3, 15303.02)	69198.35 (3, 15692.24)	36958.85 (3, 14287.87)	23486.29 (3, 14896.53)	22848.71 (3, 18411.56)	14151.02 (3, 17556.47)	11625.93 (3, 15108.66)
E*S	10195.94 (9, 19079.13)	7328.54 (9, 19516.57)	2784.84 (9, 19235.63)	2703.90 (9, 18792.67)	932.12 (9, 22540.47)	497.97 (9, 22243.10)	620.77 (9, 21133.74)
E*D	21183.88 (9, 14286.31)	27486.10 (9, 14395.97)	27151.98 (9, 13363.58)	1147.50 (9, 13708.24)	15919.85 (9, 14747.70)	8033.84 (9, 13891.66)	5491.45 (9, 13233.02)
E*M	4459.91 (6, 12911.48)	5099.82 (6, 13239.75)	3952.74 (6, 12553.37)	1887.56 (6, 12425.43)	1215.18 (6, 15405.97)	706.68 (6, 15007.13)	473.90 (6, 13573.18)
E*S*D	1497.61 (27, 21971.93)	1621.64 (27, 22161.95)	529.65 (27, 21948.32)	99.75 (27, 21536.83)	980.36 (27, 21907.35)	509.83 (27, 21498.40)	236.39 (27, 21159.85)
E*S*M	1247.12 (18, 18044.97)	927.37 (18, 18551.41)	540.04 (18, 18499.14)	404.69 (18, 17733.29)	459.54 (18, 21227.91)	222.73 (18, 21268.16)	272.15 (18, 20822.39)
E*D*M	1360.22 (18, 13539.20)	1946.81 (18, 13654.05)	2745.12 (18, 13203.38)	53.13 (18, 13049.66)	1163.97 (18, 14049.75)	389.33 (18, 13741.50)	247.23 (18, 13283.14)
E*S*D*M	499.25 (54, 21984.87)	1017.96 (54, 22257.57)	583.71 (54, 22170.23)	157.36 (54, 21639.51)	268.91 (54, 22006.12)	150.17 (54, 22095.98)	124.57 (54, 21887.70)

Note. S indicates Sample size, D: distribution, M: model, and E: estimation method. λ indicates factor loading, ϕ: factor correlation, and θ_δ: residual variance.

Table 3-15. Eta squared for significant effects on standard error ratio estimates: scale set by specifying factor variances equal to one.

	Parameter						
	λ (0.6)	λ (0.7)	λ (0.8)	ϕ (0.3)	θ_δ (0.64)	θ_δ (0.51)	θ_δ (0.36)
Between-subjects							
S	0.0107	0.0065	0.0067	0.0965	0.0041	0.0070	0.0152
D	0.2918	0.3201	0.2882	0.1195	0.2300	0.1783	0.1165
M	0.0023	0.0004	0.0010	0.0039	0.0002	0.0026	0.0045
S*D	0.0008	0.0046	0.0022	0.0145	0.0019	0.0021	0.0053
S*M	0.0017	0.0051	0.0015	0.0023	0.0014	0.0010	0.0018
D*M	0.0051	0.0027	0.0057	0.0099	0.0030	0.0049	0.0101
S*D*M	0.0037	0.0043	0.0057	0.0526	0.0038	0.0038	0.0054
Within-subjects							
E	0.0612	0.0572	0.0602	0.0333	0.1077	0.1120	0.1008
E*S	0.0028	0.0024	0.0020	0.0070	0.0017	0.0018	0.0020
E*D	0.2290	0.2542	0.3009	0.0817	0.1960	0.1371	0.1008
E*M	0.0004	0.0005	0.0004	0.0018	0.0002	0.0010	0.0043
E*S*D	0.0013	0.0011	0.0013	0.0014	0.0012	0.0012	0.0012
E*S*M	0.0003	0.0002	0.0002	0.0007	0.0003	0.0003	0.0002
E*D*M	0.0009	0.0024	0.0064	0.0002	0.0008	0.0017	0.0029
E*S*D*M	0.0002	0.0005	0.0004	0.0007	0.0002	0.0003	0.0004

Note. S indicates Sample size, D: distribution, M: model, and E: estimation method. λ indicates factor loading, ϕ: factor correlation, and θ_δ: residual variance.

Table 3-16. Marginal means by sample size for standard error ratio estimates of $\lambda = 0.6$

Sample size	Marginal means of standard error ratio
200	0.907
400	0.936
800	0.937
1200	0.949

Table 3-17. Mean estimates of standard error ratio for factor loading parameters by parameter value, estimation method, and distribution

Parameter value	Estimation	Distribution (0, 0)	(0, -1.15)	(0, 3)	(2, 6)
0.6	GLS	0.971	1.072	0.828	0.681
	ML	0.994	1.102	0.838	0.683
	Robust GLS	0.965	0.968	0.952	0.950
	Robust ML	0.988	0.992	0.967	0.965
0.7	GLS	0.967	1.102	0.800	0.650
	ML	0.989	1.138	0.808	0.652
	Robust GLS	0.961	0.963	0.955	0.943
	Robust ML	0.982	0.992	0.967	0.958
0.8	GLS	0.972	1.109	0.787	0.634
	ML	0.991	1.142	0.795	0.633
	Robust GLS	0.967	0.941	0.970	0.953
	Robust ML	0.985	0.965	0.983	0.965

Table 3-18. Mean estimates of standard error ratio for factor correlation parameters, ϕ(0.3) by model, sample size, and distribution.

Model	Sample size	Distribution (0, 0)	(0, -1.15)	(0, 3)	(2, 6)
3F3I	200	0.916	0.948	0.933	0.856
	400	1.010	0.992	0.988	0.883
	800	0.955	1.009	0.973	0.911
	1200	1.027	0.974	0.958	0.970
3F6I	200	0.941	0.932	0.946	0.898
	400	1.010	1.012	0.976	0.917
	800	0.942	1.007	1.003	0.923
	1200	1.004	1.041	1.021	0.920
6F3I	200	0.939	0.928	0.982	0.858
	400	0.990	0.986	0.954	0.933
	800	1.038	0.967	0.957	0.951
	1200	0.989	1.009	0.989	0.975

Table 3-19. Means estimates of standard error ratio for factor correlation parameter, ϕ(0.3) by estimation and distribution.

Estimation	Distribution (0, 0)	(0, -1.15)	(0, 3)	(2, 6)
GLS	0.974	0.974	0.968	0.856
ML	0.991	0.992	0.985	0.875
Robust GLS	0.969	0.975	0.962	0.956
Robust ML	0.986	0.993	0.979	0.977

Table 3-20. Marginal means by sample size for standard error ratio estimates of θ_δ = 0.36

Sample size	Marginal means of standard error ratio
200	0.864
400	0.905
800	0.913
1200	0.920

Table 3-21. Means estimates of standard error ratio for residual variance parameter, θ_δ (0.36) by model and distribution.

	Distribution			
Model	(0, 0)	(0, -1.15)	(0, 3)	(2, 6)
3F3I	0.959	0.947	0.934	0.811
3F6I	0.991	0.909	0.867	0.770
6F3I	0.956	0.922	0.914	0.827

Table 3-22. Means estimates of standard error ratio for residual variance parameters by parameter value, estimation method, and distribution.

		Distribution			
Parameter value	Estimation	(0, 0)	(0, -1.15)	(0, 3)	(2, 6)
0.36	GLS	0.962	0.903	0.839	0.664
	ML	0.982	0.910	0.840	0.653
	Robust GLS	0.956	0.940	0.966	0.946
	Robust ML	0.975	0.949	0.975	0.949
0.51	GLS	0.973	0.951	0.766	0.601
	ML	0.992	0.972	0.760	0.592
	Robust GLS	0.967	0.950	0.955	0.938
	Robust ML	0.985	0.970	0.959	0.948
0.64	GLS	0.964	1.027	0.719	0.567
	ML	0.986	1.057	0.715	0.556
	Robust GLS	0.957	0.960	0.954	0.948
	Robust ML	0.978	0.982	0.962	0.953

Table 3-23. *F* statistics (degrees of freedom) for WJ tests for effects on standard error ratio estimates: scale set by specifying factor loadings equal to one.

	Parameter			
	λ (0.7/0.6)	λ (0.8/0.6)	ϕ (i,i) = (0.36)	ϕ (i,j) = (0.108)
Between-subjects				
S	220.62 (3, 25086.80)	178.34 (3, 24879.45)	231.28 (3, 17868.11)	249.85 (3, 21326.61)
D	4013.23 (3, 18064.26)	2570.75 (3, 17484.25)	6248.21 (3, 18068.64)	1829.03 (3, 18537.10)
M	34.15 (2, 21571.44)	103.02 (2, 21102.18)	75.84 (2, 16417.49)	68.43 (2, 19090.67)
S*D	19.71 (9, 18614.53)	11.82 (9, 18401.97)	9.60 (9, 17774.22)	60.09 (9, 18451.59)
S*M	28.90 (6, 20170.81)	13.17 (6, 20142.66)	23.57 (6, 14309.88)	5.54 (6, 17071.21)
D*M	42.28 (6, 14530.87)	57.75 (6, 14186.86)	44.35 (6, 14467.44)	13.43 (6, 14855.56)
S*D*M	22.67 (18, 17021.57)	18.33 (18, 16949.97)	24.57 (18, 16181.22)	56.38 (18, 16805.40)
Within-subjects				
E	14355.94 (3, 16937.01)	13654.15 (3, 15881.56)	12849.56 (3, 13520.35)	6768.33 (3, 15521.07)
E*S	477.94 (9, 20805.45)	402.60 (9, 20286.70)	1146.01 (9, 18049.14)	551.36 (9, 19104.87)
E*D	11663.55 (9, 14485.95)	10329.60 (9, 13932.84)	14093.12 (9, 13417.50)	2848.15 (9, 15184.48)
E*M	196.35 (6, 14082.87)	195.18 (6, 13308.35)	546.87 (6, 11479.07)	425.92 (6, 12913.51)
E*S*D	133.87 (27, 21842.24)	96.65 (27, 21851.62)	241.73 (27, 21595.30)	29.38 (27, 22533.00)
E*S*M	51.77 (18, 19681.63)	46.87 (18, 19066.28)	171.71 (18, 17130.85)	92.77 (18, 17868.57)
E*D*M	135.32 (18, 13767.40)	242.49 (18, 13309.31)	250.44 (18, 12822.42)	53.21 (18, 14338.56)
E*S*D*M	36.30 (54, 22217.71)	22.02 (54, 22209.29)	71.59 (54, 21709.12)	19.90 (54, 22375.98)

Note. S indicates Sample size, D: distribution, M: model, and E: estimation method. λ indicates factor loading, and ϕ: factor correlation.

Table 3-24. Eta squared for significant effects on standard error ratio estimates: scale set by specifying factor loadings equal to one.

	Parameter			
	λ (0.7/0.6)	λ (0.8/0.6)	ϕ (i,i) = (0.36)	ϕ (i,j) = (0.108)
Between-subjects				
S	0.0094	0.0086	0.0093	0.0150
D	0.1576	0.1139	0.1789	0.0910
M	0.0009	0.0029	0.0016	0.0020
S*D	0.0012	0.0014	0.0005	0.0049
S*M	0.0020	0.0013	0.0008	0.0004
D*M	0.0036	0.0045	0.0035	0.0013
S*D*M	0.0041	0.0035	0.0031	0.0086
Within-subjects				
E	0.0421	0.0364	0.0385	0.0153
E*S	0.0021	0.0025	0.0015	0.0019
E*D	0.1098	0.0898	0.1415	0.0486
E*M	0.0003	0.0005	0.0002	0.0005
E*S*D	0.0012	0.0013	0.0005	0.0005
E*S*M	0.0002	0.0002	0.0002	0.0002
E*D*M	0.0008	0.0019	0.0006	0.0003
E*S*D*M	0.0003	0.0003	0.0002	0.0001

Note. S indicates Sample size, D: distribution, M: model, and E: estimation method. λ indicates factor loading, and ϕ: factor correlation.

Table 3-25. Mean estimates of standard error ratio for factor loading parameters by parameter value, estimation method, and distribution.

		Distribution			
Parameter value	Estimation	(0, 0)	(0, -1.15)	(0, 3)	(2, 6)
0.7/0.6	GLS	0.964	1.040	0.812	0.670
	ML	0.989	1.073	0.824	0.673
	Robust GLS	0.959	0.964	0.944	0.937
	Robust ML	0.983	0.993	0.960	0.953
0.8/0.6	GLS	0.959	1.019	0.825	0.685
	ML	0.987	1.048	0.839	0.692
	Robust GLS	0.954	0.953	0.946	0.937
	Robust ML	0.981	0.978	0.964	0.957

Table 3-26. Marginal means by sample size for standard error ratio estimates of ϕ (i, j) = 0.108

Sample size	Marginal means of standard error ratio
200	0.921
400	0.961
800	0.964
1200	0.975

Table 3-27. Mean estimates of standard error ratio for factor correlation parameters by parameter value, estimation method, and distribution.

Parameter value	Estimation	Distribution			
		(0, 0)	(0, -1.15)	(0, 3)	(2, 6)
ϕ (i, j) = 0.108	GLS	0.976	1.022	0.921	0.795
	ML	0.994	1.043	0.938	0.807
	Robust GLS	0.971	0.982	0.960	0.945
	Robust ML	0.988	1.003	0.978	0.964
ϕ (i, i) = 0.36	GLS	0.969	1.072	0.821	0.672
	ML	0.993	1.100	0.832	0.675
	Robust GLS	0.963	0.966	0.950	0.948
	Robust ML	0.986	0.990	0.965	0.964

Table 3-28. Frequency of non-convergence (NC), improper estimates (IE), and non-positive definite Hessians (NP)

Model	Distribution	Sample size	Scale set	Estimation	NC	IE or NP
3F3I	(0, 3)	200	Loading	ML/RML	0.1%	0.0%
	(2, 6)	200	Variance	ML/RML	0.2%	0.0%
		400	Loading	ML/RML	0.1%	0.0%
3F6I	(0, -1.15)	200	Loading	ML/RML	0.2%	0.0%
		200	Loading	GLS/RGLS	0.0%	0.1%
	(0, 3)	200	Loading	ML/RML	0.8%	0.5%
		200	Loading	GLS/RGLS	0.2%	0.1%
		400	Loading	ML/RML	0.1%	0.1%
	(2, 6)	200	Loading	ML/RML	0.4%	0.3%
		200	Loading	GLS/RGLS	0.1%	1.3%
		400	Loading	ML/RML	0.1%	0.1%
6F3I	(2, 6)	200	Variance	GLS/RGLS	0.0%	0.5%
		200	Loading	GLS/RGLS	0.1%	0.2%

CHAPTER 4
SUMMARY AND CONCLUSIONS

For each of four estimation methods (ML, GLS, Robust ML, and Robust GLS) the behavior of standard error ratio estimates was examined under each combination of four distributions ([0, 0], [0, -1.15], [0, 3], and [2, 6]), four sample sizes (200, 400, 800, and 1200), three CFA models (3F3I, 3F6I, and 6F3I), and two scale-setting methods (set by specifying factor variances equal to one and factor loadings equal to one). In addition, the bias of the parameter estimation was investigated since estimation bias might have varied over estimation methods at some sample sizes and there might have be an interactive effect of estimation method with other factors. The effects of four factors (estimation method, distribution, model, and sample size) on parameter estimates and standard error estimates were examined within each scale-setting method.

Comparison of Findings with Previous Studies

Parameter Estimates

Recall that ML and Robust ML produce the same parameter estimates as do GLS and Robust GLS, so for bias only a comparison of the ML and GLS estimation procedures was necessary. Important effects of factors were defined as those that were significant by the WJ test and accounted for at least 1% of the total sum of squares in a repeated measures ANOVA of the data. When the scale was set by specifying factor variances equal to one, no factor had an important effect on estimation of the factor correlation and only estimation method had an important effect on estimates of the factor loading and then only when the population factor loading was 0.8. ML estimates had less bias than GLS estimates, but the effect was very small. The interaction of estimation method and sample size had an important effect on estimates of the residual variance. The results showed that ML estimates had minimal bias at all sample sizes.

Therefore, although the bias of the ML estimates declined as sample size increased, the effect of sample size on bias of ML estimates was quite small. The bias of GLS estimates also declined as sample size increases. Compared to the bias of ML estimates, the bias of GLS estimates was larger with larger differences between the methods when the sample size was 200 or 400 and a trivial difference when the sample size was1200.

When the scale was set by specifying factor loadings equal to one, most effects on parameter estimates of factor loading and factor correlation were quite small and none of them accounted for even 1% of the total sum of squares. Recall that estimates of residual variance are the same for both scale-setting methods. Therefore effects of the factors on the residual variance were the same for both methods of setting the scale.

Boomsma (1983) indicated that, when the distribution is normal and the scale is set by specifying factor loadings equal to one, the bias of ML parameter estimates is small for models with more indicators per factor and higher factor loadings compared with the bias of ML estimates for models with fewer indicators per factor and smaller factor loadings. Boomsma only investigated models for which the scale was set by specifying factor loadings equal to one. Muthen and Kaplan (1985) found that there was not much difference between ML and GLS parameter estimates. Henly (1993) also pointed out the similarity between ML and GLS, that is, equivalent conclusion was drawn about the behavior of the ML and GLS parameter estimates in the distribution and sample size conditions studied: Like the ML parameter estimates, the GLS estimates appeared to produce consistent parameter estimates that are unbiased when $N \geq 600$ for samples from the multivariate normal and asymmetric multivariate populations (Henly, 1993). According to Hoogland and Boomsma (1998), the bias of ML parameter estimates increases when the levels of univariate skewness and kurtosis deviate increasingly from those of a normal distribution and a larger sample size is a remedy to obtain unbiased parameter estimates. Hoogland (1999) reported that ML parameter estimates were almost unbiased when the sample size was at least 200 and GLS had a much larger bias than ML.

For parameter estimates, the overall results of the present study are consistent with the results of the previous studies: ML estimates were almost unbiased at all sample sizes and ML estimation had less bias than GLS estimation, although the differences were trivial for factor loadings. Sample size played more a critical role in GLS estimation than in ML estimation of residual variance and, as a result, larger between-method differences in bias were observed for estimates of residual variance. When the scale was set by specifying factor loadings equal to one, there were no important effects of the factors on the factor loading, factor variance, or factor covariance estimates.

Standard Error Estimates

When the scale was set by specifying factor variances equal to one, the bias of standard error estimates for the factor loadings decreased as the sample size increased, showing consistency with theory that standard error estimates will converge to their true values as the sample size increases without bound. For factor loadings, the ML standard error estimates had minimal bias for the normal distribution and Robust ML estimates had a relatively small bias regardless of the distribution. Also, there was relatively little difference between the biases of ML and GLS estimates of standard errors, and relatively little difference between the biases of Robust ML and Robust GLS estimates of standard errors regardless of the distribution. ML and GLS estimates of standard errors of factor loading were overestimated when the distribution was short-tailed (0, -1.15) and strongly underestimated when the distributions were long-tailed (0, 3) and long-tailed and skewed (2, 6).

A long tail to the distribution tended to result in underestimation of standard errors of the inter-factor correlation and underestimation was particularly large for the (2, 6) distribution. For the (2, 6) distribution, bias of the standard errors decreased systematically as the sample size increased. For the other distributions, estimates of the standard error ratio did not systematically decline as the sample size increased because the empirical standard error estimates and standard error estimates for the factor correlation decreased as the sample size increased but not the same rate (please see Appendix A). When the distribution was normal, (0, -1.15) , and (0, 3), ML and Robust ML estimation of standard errors were less biased than were GLS and Robust GLS estimation of standard errors. The robust estimation procedures showed substantially better performance than the non-robust estimation procedures for the (2, 6) distribution.

The bias of the standard error estimates for residual variance decreased as the sample size increased regardless of the population value. The bias increased as the degree of non-normality increased regardless of the model type. For a normal distribution, more indicators per factor resulted in less bias, but resulted in more bias for the non-normal distributions. For a normal distribution, ML and Robust ML estimation performed better than GLS and Robust GLS estimation of the standard error estimates of residual variances, whereas the robust procedures performed better than the non-robust procedures for the non-normal distributions. The bias of the standard error estimates for residual variances tended to decrease as the population value (0.64, 0.51, and 0.36) decreased.

When the scale was set by specifying factor loadings equal to one and the distribution was normal, between-estimation method differences in the bias of the standard error estimates for the factor loadings were very small. For the non-normal distributions, robust estimation of standard errors of factor loading performed better than did non-robust estimation. When the distribution was (0, -1.15), the standard error estimates for the factor loadings tended to be overestimated by the non-robust estimation procedures and underestimated by the robust estimation procedures. For the other distributions, the standard error estimates of the factor loadings tended to be underestimated regardless of the estimation procedures.

The bias of standard error estimates for the factor covariances (ϕ (i, j) = 0.108) tended to decrease as the sample size increased. For the normal distribution, ML and Robust ML estimates of the standard errors for the factor covariance had less bias than did the GLS and Robust GLS estimates. The robust procedures had less bias than the non-robust procedures for the non-normal distributions. Robust ML estimates of the standard errors for the factor correlations had the minimal bias for the non-normal distributions.

Chou, Bentler, and Satorra (1991) studied the performance of Robust ML estimates of standard errors and found that when the distribution had excessive kurtosis, the robust estimates of standard errors were superior to ML estimates. Finch, West, and MacKinnon (1997) indicated that the Robust ML estimates of standard errors provided more accurate estimates of sampling variability than ML and GLS as non-normality increased, and the standard error estimates generated by ML and GLS were likely to be too small. Hoogland (1999) pointed out that the ML and GLS estimates of the standard errors were biased when the average kurtosis of the observed variables deviates from zero, the standard error estimates were underestimated in the case of a positive average kurtosis and overestimated in the case of negative average kurtosis. Also Hoogland (1999) reported that the Robust ML standard error estimates had a smaller bias than ML and GLS standard error estimates when the average kurtosis was at least 2.0 and the sample size was at least 400.

The results of the present study generally showed consistency with the results of the previous studies in that Robust ML estimates were superior to the non-robust estimates in the bias of the standard error estimates for the non-normal distributions, and the standard error estimates were underestimated for the distribution with positive kurtosis and overestimated for the distribution with negative kurtosis. However, the present study gives more general results because Robust GLS was added, two scale-setting methods were compared, and results were analyzed statistically.

Brief Summary

The main purpose of the present study was to answer the question: "Which estimation method provides better standard errors?" A Monte Carlo simulation study was conducted to analyze and investigate the effects of four factors (estimation, sample size, distribution, and model) on the bias of standard error estimates and the bias of parameter estimates in each scale-setting method.

First, from the findings of the study, the following conclusions can be set forth in regard to for the bias of parameter estimates:

- When the scale is set by specifying factor variances equal to one, ML estimates of the factor loading have less bias than do the GLS estimates.
- The bias of ML and GLS estimates of residual variance decreases as the sample size increases and the bias of GLS estimation is more affected by sample sizes between 200 and 1200 than is ML estimation.
- The ML estimates of residual variance have minimal bias at all sample sizes whereas the GLS estimates have more substantial bias when sample size is 200 or 400.
- When the scale is set by specifying factor loadings equal to one, none of the effects on the factor loading or the factor correlation parameter estimates account for even 1% of the total sum of squares.

Second, the following points can be noted about the bias of standard error estimates:

- In regard to bias of the standard error estimators for factor loadings: Regardless of scale-setting method, for a normal distribution, ML estimation is superior to the other estimation procedures, whereas for non-normal distributions, Robust ML is superior to the other estimation procedures.

- In regard to bias of the standard error estimators for the factor correlation:

    1. When the scale is set by specifying factor variances equal to one, the mean standard error estimates and the empirical standard error estimates decrease as the sample size increases, regardless of the model type, but not the same rate.

    2. When the scale is set by specifying factor variances equal to one, ML and Robust ML are superior to GLS and Robust GLS for the (0, 0), (0, -1.15), (0, 3) distributions, whereas Robust ML and Robust GLS are superior to ML and GLS for the (2, 6) distribution.

    3. When the scale is set by specifying factor loadings equal to one, ML and Robust ML are superior to GLS and Robust GLS for the (0, 0) distribution, whereas Robust ML and Robust GLS are superior to ML and GLS for the (0, -1.15), (0, 3), and (2, 6) distributions

- In regard to the bias of standard error estimates for residual variance: ML and Robust ML are superior to GLS and Robust GLS for a normal distribution, whereas Robust ML and Robust GLS are superior to ML and GLS for non-normal distributions

- The bias of standard error estimates for residual variance increases as the number of indicators per factor increases for non-normal distributions, whereas the bias of these estimates decreases as the number of indicators per factor increases for a normal distribution.

Concluding Remarks

Based on the findings presented, it can be concluded that ML estimation method should be adopted for a normal distribution regardless of sample size, model, and scale-setting method to obtain less biased estimates of parameters and standard errors, and Robust ML should be used for non-normal distributions to improve estimation of standard errors. However, Robust ML estimation works very well even for normal distributions and some cases better than GLS. It has also been found that robust estimation generally worked better than non-robust estimation for the non-normal distributions regardless of the sample size and the model type. When the distribution is non-normal, Robust GLS generally performs well, although Robust ML has less bias than Robust GLS.

Problems of non-convergence, improper estimates, and non-positive definite Hessian matrices were more common when the scale was set by specifying factor loadings equal to one and the distribution is non-normal, particularly for the 3F6I model with 200 and 400 sample sizes. However, as a practical matter these problems occurred infrequently. Setting the scale by specifying factor variances equal to one should be chosen to avoid these problems.

Generalization from the findings of Monte Carlo simulation studies is limited by the design of the simulation. Although the present study evaluated the empirical behavior of parameter estimates and standard error estimates under a range of sample sizes, distributions, and models which were chosen considering the results of the previous studies, the study of the robust estimation for standard errors still needs to be investigated further under the various conditions of models (e.g., misspecified models, or hybrid models), and the other robust estimation procedures (e.g., Robust DWLS, or Robust ULS).

APPENDIX

STANDARD ERROR RESULTS FOR FACTOR CORRELATIONS: SCALE SET BY SPECIFYING FACTOR VARIANCES EQUAL TO ONE.

Table A-1. Mean standard errors, empirical standard errors, and standard error ratios by model, distribution, and sample size for ML estimation.

			Sample size			
Model	Distribution	Statistics	200	400	800	1200
3F3I	(0, 0)	SE	0.0852035	0.0607274	0.0430910	0.0351752
		Emp. SE	0.0916527	0.0598677	0.0448250	0.0341341
		SE Ratio	0.9296346	1.0143586	0.9613180	1.0305009
	(0, -1.15)	SE	0.0853402	0.0605011	0.0429852	0.0351710
		Emp. SE	0.0894783	0.0606789	0.0425763	0.0361427
		SE Ratio	0.9537530	0.9970703	1.0096033	0.9731165
	(0, 3)	SE	0.0850728	0.0605791	0.0429177	0.0351232
		Emp. SE	0.0892047	0.0609162	0.0441473	0.0365462
		SE Ratio	0.9536806	0.9944662	0.9721477	0.9610636
	(2, 6)	SE	0.0843846	0.0604525	0.0428893	0.0351204
		Emp. SE	0.1015078	0.0719724	0.0499687	0.0385356
		SE Ratio	0.8313114	0.8399404	0.8583233	0.9113744
3F6I	(0, 0)	SE	0.0747718	0.0529857	0.0375588	0.0306787
		Emp. SE	0.0770873	0.0517938	0.0395937	0.0305296
		SE Ratio	0.9699637	1.0230110	0.9486042	1.0048813
	(0, -1.15)	SE	0.0747955	0.0529123	0.0375141	0.0306708
		Emp. SE	0.0784540	0.0517873	0.0368942	0.0293325
		SE Ratio	0.9533679	1.0217236	1.0168020	1.0456226
	(0, 3)	SE	0.0748360	0.0530562	0.0375188	0.0306476
		Emp. SE	0.0764260	0.0534427	0.0372220	0.0299660
		SE Ratio	0.9791960	0.9927687	1.0079729	1.0227563
	(2, 6)	SE	0.0745467	0.0528650	0.0374607	0.0305994
		Emp. SE	0.0844731	0.0599450	0.0428135	0.0352090
		SE Ratio	0.8824907	0.8818920	0.8749737	0.8690774
6F3I	(0, 0)	SE	0.0855209	0.0606582	0.0430008	0.0351380
		Emp. SE	0.0887259	0.0602660	0.0411450	0.0353648
		SE Ratio	0.9638773	1.0065075	1.0451038	0.9935871
	(0, -1.15)	SE	0.0850098	0.0606301	0.0430029	0.0351318
		Emp. SE	0.0891051	0.0608826	0.0440900	0.0346944
		SE Ratio	0.9540398	0.9958530	0.9753441	1.0126064
	(0, 3)	SE	0.0850900	0.0606005	0.0429920	0.0351811
		Emp. SE	0.0843096	0.0625306	0.0445236	0.0356247
		SE Ratio	1.0092562	0.9691320	0.9655982	0.9875455
	(2, 6)	SE	0.0847380	0.0603836	0.0429598	0.0351398
		Emp. SE	0.1011257	0.0672843	0.0477766	0.0382489
		SE Ratio	0.8379476	0.8974397	0.8991809	0.9187121

Note. SE indicates standard error, and Emp. Indicates empirical.

Table A-2. Mean standard errors, empirical standard errors, and standard error ratios by model, distribution, and sample size for GLS estimation.

			Sample size			
Model	Distribution	Statistics	200	400	800	1200
3F3I	(0, 0)	SE	0.0865475	0.0611873	0.0432570	0.0352676
		Emp. SE	0.0947246	0.0604901	0.0454183	0.0343848
		SE Ratio	0.9136754	1.0115267	0.9524125	1.0256748
	(0, -1.15)	SE	0.0866679	0.0609604	0.0431456	0.0352652
		Emp. SE	0.0916833	0.0618860	0.0429062	0.0362790
		SE Ratio	0.9452964	0.9850431	1.0055798	0.9720555
	(0, 3)	SE	0.0863269	0.0610873	0.0430876	0.0352161
		Emp. SE	0.0932935	0.0616207	0.0444169	0.0369437
		SE Ratio	0.9253258	0.9913442	0.9700712	0.9532371
	(2, 6)	SE	0.0860443	0.0610668	0.0431184	0.0352416
		Emp. SE	0.1066839	0.0734757	0.0504100	0.0387986
		SE Ratio	0.8065352	0.8311150	0.8553538	0.9083230
3F6I	(0, 0)	SE	0.0784797	0.0542636	0.0380305	0.0309424
		Emp. SE	0.0852081	0.0541515	0.0405504	0.0308609
		SE Ratio	0.9210350	1.0020709	0.9378569	1.0026409
	(0, -1.15)	SE	0.0785564	0.0542692	0.0379842	0.0309353
		Emp. SE	0.0859818	0.0541645	0.0381844	0.0299032
		SE Ratio	0.9136399	1.0019347	0.9947568	1.0345156
	(0, 3)	SE	0.0785063	0.0543838	0.0379884	0.0309098
		Emp. SE	0.0840802	0.0563051	0.0380301	0.0303428
		SE Ratio	0.9337076	0.9658763	0.9989033	1.0186850
	(2, 6)	SE	0.0790310	0.0545171	0.0380504	0.0309130
		Emp. SE	0.0952944	0.0638307	0.0443069	0.0361298
		SE Ratio	0.8293349	0.8540890	0.8587920	0.8556083
6F3I	(0, 0)	SE	0.0878715	0.0615771	0.0433324	0.0353174
		Emp. SE	0.0950906	0.0628897	0.0419638	0.0357739
		SE Ratio	0.9240818	0.9791284	1.0326145	0.9872395
	(0, -1.15)	SE	0.0874593	0.0615880	0.0433357	0.0353145
		Emp. SE	0.0965776	0.0631581	0.0452786	0.0352822
		SE Ratio	0.9055857	0.9751398	0.9570903	1.0009153
	(0, 3)	SE	0.0876133	0.0615888	0.0433481	0.0353710
		Emp. SE	0.0897761	0.0654029	0.0457940	0.0357338
		SE Ratio	0.9759086	0.9416839	0.9465895	0.9898479
	(2, 6)	SE	0.0880338	0.0615664	0.0433858	0.0353801
		Emp. SE	0.1090550	0.0706136	0.0486242	0.0390347
		SE Ratio	0.8072426	0.8718775	0.8922679	0.9063758

Table A-3. Mean standard errors, empirical standard errors, and standard error ratios by model, distribution, and sample size for Robust ML estimation.

			Sample size			
Model	Distribution	Statistics	200	400	800	1200
3F3I	(0, 0)	SE	0.0841564	0.0603712	0.0429578	0.0350993
		Emp. SE	0.0916527	0.0598677	0.0448250	0.0341341
		SE Ratio	0.9182100	1.0084095	0.9583457	1.0282767
	(0, -1.15)	SE	0.0850832	0.0605774	0.0431090	0.0352981
		Emp. SE	0.0894783	0.0606789	0.0425763	0.0361427
		SE Ratio	0.9508808	0.9983270	1.0125115	0.9766321
	(0, 3)	SE	0.0838108	0.0599901	0.0430542	0.0351824
		Emp. SE	0.0892047	0.0609162	0.0441473	0.0365462
		SE Ratio	0.9395330	0.9847970	0.9752397	0.9626822
	(2, 6)	SE	0.0919380	0.0673081	0.0482695	0.0397790
		Emp. SE	0.1015078	0.0719724	0.0499687	0.0385356
		SE Ratio	0.9057236	0.9351928	0.9659951	1.0322664
3F6I	(0, 0)	SE	0.0740826	0.0526849	0.0374271	0.0306684
		Emp. SE	0.0770873	0.0517938	0.0395937	0.0305296
		SE Ratio	0.9610227	1.0172036	0.9452783	1.0045460
	(0, -1.15)	SE	0.0745119	0.0529434	0.0375916	0.0307174
		Emp. SE	0.0784540	0.0517873	0.0368942	0.0293325
		SE Ratio	0.9497528	1.0223245	1.0189048	1.0472134
	(0, 3)	SE	0.0731563	0.0526214	0.0374305	0.0306659
		Emp. SE	0.0764260	0.0534427	0.0371785	0.0299660
		SE Ratio	0.9572180	0.9846324	1.0067770	1.0233568
	(2, 6)	SE	0.0818413	0.0587417	0.0422960	0.0346940
		Emp. SE	0.0844731	0.0599450	0.0428135	0.0352090
		SE Ratio	0.9688451	0.9799270	0.9879128	0.9853713
6F3I	(0, 0)	SE	0.0845748	0.0602601	0.0428938	0.0350724
		Emp. SE	0.0887259	0.0602660	0.0411450	0.0353648
		SE Ratio	0.9532151	0.9999020	1.0425049	0.9917328
	(0, -1.15)	SE	0.0846887	0.0606585	0.0431039	0.0352690
		Emp. SE	0.0891051	0.0608826	0.0440900	0.0346944
		SE Ratio	0.9504365	0.9963187	0.9776355	1.0165624
	(0, 3)	SE	0.0833226	0.0604504	0.0431205	0.0352033
		Emp. SE	0.0843096	0.0625306	0.0445236	0.0356247
		SE Ratio	0.9882931	0.9667324	0.9684846	0.9881700
	(2, 6)	SE	0.0922164	0.0671313	0.0483098	0.0399671
		Emp. SE	0.1011257	0.0672843	0.0477766	0.0382489
		SE Ratio	0.9118993	0.9977262	1.0111612	1.0449214

Table A-4. Mean standard errors, empirical standard errors, and standard error ratios by model, distribution, and sample size for Robust GLS estimation.

			Sample size			
Model	Distribution	Statistics	200	400	800	1200
3F3I	(0, 0)	SE	0.0854879	0.0608352	0.0431266	0.0351923
		Emp. SE	0.0947246	0.0604901	0.0454183	0.0343848
		SE Ratio	0.9024894	1.0057061	0.9495434	1.0234863
	(0, -1.15)	SE	0.0864308	0.0610456	0.0432699	0.0353919
		Emp. SE	0.0916833	0.0618860	0.0429062	0.0362790
		SE Ratio	0.9427107	0.9864204	1.0084753	0.9755476
	(0, 3)	SE	0.0850932	0.0605307	0.0432372	0.0352852
		Emp. SE	0.0932935	0.0616207	0.0444169	0.0369437
		SE Ratio	0.9121026	0.9823105	0.9734396	0.9551081
	(2, 6)	SE	0.0939100	0.0681430	0.0485903	0.0399571
		Emp. SE	0.1066839	0.0734757	0.0504100	0.0387986
		SE Ratio	0.8802636	0.9274210	0.9639022	1.0298613
3F6I	(0, 0)	SE	0.0776360	0.0539350	0.0378868	0.0309388
		Emp. SE	0.0852081	0.0541515	0.0405504	0.0308609
		SE Ratio	0.9111333	0.9960018	0.9343122	1.0025226
	(0, -1.15)	SE	0.0782640	0.0542898	0.0380718	0.0309850
		Emp. SE	0.0859818	0.0541645	0.0381844	0.0299032
		SE Ratio	0.9102398	1.0023141	0.9970510	1.0361790
	(0, 3)	SE	0.0768061	0.0539844	0.0379251	0.0309261
		Emp. SE	0.0840802	0.0563051	0.0380641	0.0303428
		SE Ratio	0.9134865	0.9587832	0.9963499	1.0192230
	(2, 6)	SE	0.0867332	0.0607091	0.0430780	0.0350903
		Emp. SE	0.0952944	0.0638307	0.0443069	0.0361298
		SE Ratio	0.9101606	0.9510957	0.9722629	0.9712282
6F3I	(0, 0)	SE	0.0869908	0.0611649	0.0432345	0.0352456
		Emp. SE	0.0950906	0.0628897	0.0419638	0.0357739
		SE Ratio	0.9148202	0.9725734	1.0302802	0.9852329
	(0, -1.15)	SE	0.0870643	0.0616383	0.0434450	0.0354532
		Emp. SE	0.0965776	0.0631581	0.0452786	0.0352822
		SE Ratio	0.9014962	0.9759358	0.9595031	1.0048462
	(0, 3)	SE	0.0858012	0.0614686	0.0434688	0.0354037
		Emp. SE	0.0897761	0.0654029	0.0457940	0.0357338
		SE Ratio	0.9557239	0.9398458	0.9492261	0.9907638
	(2, 6)	SE	0.0954363	0.0682084	0.0486965	0.0401757
		Emp. SE	0.1090550	0.0706136	0.0486242	0.0390347
		SE Ratio	0.8751212	0.9659379	1.0014864	1.0292304

LIST OF REFERENCES

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155-173.

Andreassen, T. W., Lorentzen, B. G., & Olsson, U. H. (2006). The impact of non-normality and estimation methods in SEM on satisfaction research in marketing. *Quality & Quantity*, *40*, 39-58.

Arminger G., & Schoenberg R. J. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika*, *54*, 409-425.

Babakus, E., Ferguson, C. E., & Joreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, *24*, 222-228.

Baldwin, B. O. (1986). *The effects of structural model misspecification and sample size on the robustness LISREL maximum likelihood estimates*. Doctoral dissertation, Louisiana State University.

Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi-square and other statistics used in evaluating causal models. *Journal of Marketing Research*, *19*, 425-430.

Beauducel, A. B., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, *13*, 186-203.

Benson, J., & Fleishman, J. A. (1994). The robustness of maximum likelihood and distribution-free estimators to non-normality in confirmatory factor analysis. *Quality & Quantity*, *28*, 117-136.

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

Bentler, P. M., & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*, 78-117.

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review Psychology*, *47*, 563-92.

Bentler, P. M., & Yuan, K. (1999). Structural equation modeling with small samples: Test Statistics. *Multivariate Behavioral Research*, *34*, 181-197.

Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality*. Amsterdam: Sociometric Research Foundation. Doctoral dissertation, Rijksuniversiteit Groningen.

Boomsma, A., & Hoogland, J. J. (2001). The Robustness of LISREL Modeling Revisited. In R. Cudeck, S. Du Toit, & D. Sörbom (Eds), *Structural equation modeling: Present and future. A Festschrift in honor of Karl Jöreskog* (pp. 139-168). Lincolnwood, IL: Scientific Software International.

Brown, R. L. (1990). The robustness of 2SLS estimation of non-normality distributed confirmatory factor analysis model. *Multivariate Behavioral Research*, *25*, 455-466.

Browne, M. W. (1982). Covariance structure. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). Cambridge, UK: Cambridge University.

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

Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika*, *74*, 375-384.

Browne, M. W. & Shapiro A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, *41*, 193-208.

Casella, G., & Berger, R. L. (2002). *Statistical Inference*, (2nd Ed.), Pacific Grove, CA: Duxbury.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models. *Sociological Methods & Research*, *29*, 468-508.

Chou, C., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, *44*, 347-357.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, *105*, 317-327.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16-29.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5, and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, *47*, 309-326.

Ethington, C. A. (1987). The robustness of LISREL estimates in structural equation models with categorical variables. *Journal of Experimental Education*, *55*, 80-88.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, *6*, 56-83.

Finch, J. F. (1992). *The robustness of maximum likelihood parameter estimates in structural equation models with nonnormal variables*. Doctoral dissertation, Arizona State University.

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling*, *4*, 87-107.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521-532.

Gallini, J. K. (1984). An investigation of the effect of sample size and specification error on the fit of structural equation models. *Journal of Experimental Education*, *53*, 9-19.

Gerbing, D. W., & Anderson, J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. *Multivariate Behavioral Research*, *20*, 255-271.

Gonzalez, R., & Griffin D. (2001). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods, 6*, 258-269.

Harlow, L. L. (1985). *Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study*. Doctoral dissertation, University of California, Los Angeles.

Hau, K., & Marsh, H. W. (2004). The use of item parcels in structural equation modeling: Non-normal data and small sample sizes. *British Journal of Mathematical Statistical Psychology*, *57*, 327-351.

Henly, S. J. (1993). Robustness of some estimators for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *46*, 313-338.

Hoogland, J. J. (1999). *The robustness of estimation methods for covariance structure analysis.* Doctoral dissertation, Groningen University.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods & Research*, *26*, 329-367.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted?. *Psychological Bulletin*, *112*, 351-362.

Huba, G. J., & Harlow, L. L. (1987). Robust structural equation models: Implications for developmental psychology. *Child Development*, *58*, 147-166.

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

Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling*, *8*, 205-223.

Johansen, S. (1980). The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika*, *67*, 85-92.

Jonsson, F. Y. (1997). *Non-linear structural equation models: Simulation Studies of the Kenny-Judd Model*. Doctoral dissertation, Uppsala University.

Jöreskog, K. G., & Sörbom, D., (1996). *LISREL 8: User's Reference Guide*. Chicago, IL: Scientific software International Inc.

Jöreskog, K. G., Sörbom, D., Toit, S. D., & Toit, M. D. (1996). *LISREL 8: New Statistical Features.* Chicago, IL: Scientific software International Inc.

Kaplan, D. (1989). A study of the sampling variability and z-values of parameter estimates from misspecified structural equation models. *Multivariate Behavioral Research*, *24*, 41-57.

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, *25*, 137-155.

Kendall M., Stuart, A. & Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics: Distribution theory*, (5th Ed.). Vol. 1. New York, NY: Oxford University Press

Keselman, H. J., Carriere, K. C. H., & Lix, L. M. (1993). Testing repeated measures hypotheses when covariance matrices are heterogeneous. *Journal of Education Statistics*, *18*, 305-319.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd Ed.). New York, NY: The Guilford Press.

Lee, S., Poon, W., & Bentler, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology*, *48*, 339-358.

Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling*, *12*, 1-27.

Lix, L. M., & Keselman, H. J. (1995). Approximate degrees of freedom tests: A unified perspective on testing for mean equality. *Psychological Bulletin, 117*, 547-560.

MacCallum, R. C. (1990). The need for alternative measures of fit in covariance structure modeling. *Multivariate Behavioral Research*, *25*, 157-162.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519-530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya: The Indian Journal of Statistics*, *36*, 115-128.

Marsh, H. W., Hau, K., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory. *Multivariate Behavioral Research*, *33*, 181-220.

Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research*, *32*, 355-373.

Muthen, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171-189.

Muthen, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19-30.

Nevitt, J., & Hancock, G. R. (2000). Improving the root mean square error of approximation for nonnormal conditions in structural equation modeling. *The Journal of Experimental Education*, *68*, 251-268.

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

Nevitt, J., & Hancock, G. R. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate Behavioral Research*, *39*, 439-478.

Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, *8*, 434-447.

Olsson, U., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling*, *7*, 557-595.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, *8*, 287-312.

Ping, R. A. (1995). A parsimonious estimating technique for interaction and quadratic latent variables. *Journal of Marketing Research*, *32*, 336-347.

Potthast, M. J. (1993). Confirmatory factor analysis of ordered categorical variables with large models. *British Journal of Mathematical and Statistical Psychology*, *46*, 273-286.

Reddy, S. K. (1992). Effects of ignoring correlated measurement error in structural equation models. *Educational and Psychological Measurement*, *52*, 549-570.

Reinartz, W. J., Echanmbadi, R., & Chin, W. W. (2002). Generating non-normal data for simulation of structural equation models using Mattson's method. *Multivariate Behavioral Research*, *37*, 227-244.

Satorra, A. (1990). Robustness issues in structural equation modeling: A review of recent developments. *Quality & Quantity*, *24*, 367-386.

Sharma, S., Durvasula, S., & Dillon, W. R. (1989). Some results on the behavior of alternate covariance structure estimation procedures in the presence of non-normal data. *Journal of Marketing Research*, *26*, 214-221.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for "Optimal" cutoff properties: Fit index criteria in structural equation modeling. *The Journal of Experimental Education*, *74*, 267-288.

Tanaka, J. S. (1984). *Some results on the estimation of covariance structure models*. Doctoral dissertation, University of California, Los Angeles.

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, *58*, 134-146.

Ullman, J. B. (2006). Structural equation modeling: Reviewing the basics and moving forward. *Journal of personality assessment*, *87*, 35-50.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465-471.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation modeling: concepts, issues and applications. In R. Hoyle (Eds.), *Structural equation models with nonnormal variables* (pp. 56-75). Newbury Park, CA: Sage.

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256-293). Newbury Park, CA: Sage.

Yuan, K., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, *51*, 289-309.

Yung, Y., & Bentler, P. M. (1994). Bootsrtap-corrected ADF test statistics in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, *47*, 63-84.

BIOGRAPHICAL SKETCH

Youngkyoung Min was born in Seoul, South Korea. She received her bachelor's and master's degrees in polymer science and engineering from Chungnam National University, South Korea. She was an engineering researcher of the research and development center in Kumho Chemical Inc. In August 2001, she completed her M. S. in Management at the University of Florida, and she also achieved her M. S. in Statistics at the University of Florida in August 2004. In the fall of 2004, she enrolled for graduate studies in research and evaluation methodology program of Educational Psychology Department at the University of Florida and completed her Ph.D. in research and evaluation methodology in August 2008. Youngkyoung Min conducted researches as a research assistant at the Department of Health and Clinical Psychology at the University of Florida from August 2004 to August 2005, and in Engineering Education Center of College of Engineering at the University of Florida from August 2005 to June 2008.