

THE PERFORMANCE OF MLR, USLMV, AND WLSMV ESTIMATION IN STRUCTURAL
REGRESSION MODELS WITH ORDINAL VARIABLES

By

Cheng-Hsien Li

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Measurement and Quantitative Methods—Doctor of Philosophy

2014

ABSTRACT

THE PERFORMANCE OF MLR, USLMV, AND WLSMV ESTIMATION IN STRUCTURAL REGRESSION MODELS WITH ORDINAL VARIABLES

By

Cheng-Hsien Li

In the educational, social, and behavioral sciences, ordered observed categorical variables are commonly used to operationalize latent constructs in structural regression models. Treating ordinal manifest variables as if they were continuous, the precision and accuracy of model parameter estimates, standard errors, and chi-square goodness of fit statistics are likely compromised, leading to invalid statistical inferences. Three robust estimators – robust maximum likelihood (MLR), robust unweighted least squares (ULSMV), and robust weighted least squares (WLSMV) – have been proposed in the literature over the past two decades, and are considered to be superior to normal theory-based maximum likelihood (ML) when ordinal observed variables are analyzed.

The purpose of this thesis was to carry out a Monte Carlo simulation study, in order to compare the performance of ML, the most widely known estimation method, with the three robust estimators (MLR, ULSMV, and WLSMV) on parameter estimates, standard errors, and chi-square goodness of fit statistics in a five-factor structural regression model with ordinal observed variables. There were 4 (level of asymmetric distributions of ordinal observed variables: symmetry, slight and moderate asymmetry, as well as bipolarization) \times 4 (number of observed variables' categories: 4, 5, 6, and 7) \times 7 (sample size: 200, 300, 400, 500, 750, 1,000, and 1,500) = 112 conditions in the study. Five hundred data sets were generated under each experimental condition. Model parameters, standard errors, chi-square goodness of fit statistics, and RMSEA were estimated for each replication using ML, MLR, ULSMV, and WLSMV. Data generation and analysis were performed with *Mplus 7*.

The results reveal that (1) the four estimators are all subjected to non-convergence

problems with 4-category, moderately asymmetric data in the smallest sample size $N = 200$; (2) WLSMV and ULSMV are likely to produce inadmissible solutions in some conditions with sample sizes $N = 200$ or 300 ; (3) WLSMV and ULSMV yield more accurate factor loading estimates than ML and MLR across all conditions in the study; (4) the estimates of structural coefficients under ML and MLR outperform WLSMV and ULSMV in all symmetric data conditions, whereas WLSMV and ULSMV surpass ML and MLR in nearly all asymmetric data conditions; (5) the robust standard errors of factor loadings obtained with ULSMV are more precise than those produced by WLSMV and MLR across all conditions; (6) the robust standard errors of structural coefficients obtained with WLSMV are more precise than those with ULSMV and MLR in all asymmetric data conditions; (7) among the three robust estimators, MLR is inferior to WLSMV and ULSMV in controlling for Type I error rates of testing overall model fit in almost every condition, unless a larger sample size is used (i.e., $N = 1,000$ in this thesis); (8) RMSEA seems to be a reliable index in the evaluation of overall model fit when the model has no specification error; (9) the benefit of using diagonal weights can be found in the estimation of factor loadings and structural coefficients as well as robust standard errors of structural coefficients, but not in the estimation of robust standard errors of factor loadings and the mean- and variance-adjusted chi-square goodness of fit statistics across all conditions; and (10) the accuracy and precision of factor loading and structural coefficient estimates and standard error estimates of factor loadings and structural coefficients improve with increasing sample size and number of observed variables' categories but decrease with a greater level of asymmetric distributions.

Collectively, the findings from this study provide a better understanding of the performance of the three robust estimators, and aim to inform the work of applied researchers with respect to the importance of attending to assumption violations and selecting an "appropriate" estimator under circumstances frequently encountered in practice. Finally, implications of the findings for structural regression models using these four estimators are discussed, as are the limitations of this study as well as potential directions for future research.

Copyright by
CHENG-HSIEN LI
2014

ACKNOWLEDGMENTS

The writing of this dissertation has been an incredible journey and a monumental accomplishment in my academic life.

First of all, I would like to express my deepest appreciation and genuine gratitude to my advisor and dissertation chair Dr. Tenko Raykov for his unflagging support and constant encouragement throughout my doctoral endeavors. I greatly appreciate his patience to review early dissertation drafts so many times and his constructive feedback, which have helped me improve the quality of this dissertation tremendously. Without his supervision, guidance, and intellectual enlightenment, this dissertation would not have been possible.

I would like to thank my dissertation committee members Dr. Mark Reckase, Dr. Richard DeShon, and Dr. Matthew Diemer, each of whom has provided insightful comments and invaluable suggestions on an earlier version of this dissertation, and has made a unique contribution to the completion of this dissertation. A special thanks is extended to Dr. Matthew Dimmer, who has been my mentor in the fields of Developmental and Educational Psychology for the past six years. Working with him has been a truly rewarding and inspirational experience along the way. I would also like to thank Dr. Konstantopoulos, Dr. Schmidt, and Dr. Kimberley for providing excellent opportunities to enrich my teaching and research experience.

I feel deeply indebted to Dr. Jing-Jyi Wu, Dr. Shu-Shen Shih, and Dr. James Tu, who have supported me emotionally and academically since I started pursuing my Ph.D. degree in

2008. I would like to express my gratitude to my colleagues Anne Traynor and Hyesuk Jang, and my friends Yun-Jia Lo, Yi-Ling Cheng, I-Chien Chen, Guan Saw, and Chi Chang, who have helped me in countless ways throughout my years at MSU.

Last but absolutely not least, a unique thanks goes to my lovely family, my parents Hsin-Hua Li and Hsiu-Kan Chen, brother Hung-Wei Li, sister-in-law Mei-Hua Lin, sister Wen-Hui Li, nieces Tzu-Ching Li and Tzu-Yi Li, and nephew Pin-Yen Li. Words cannot express how grateful I am to them for all of the sacrifices that they have made on my behalf, unwavering faith that they have had in me, and unconditional love that they have given me.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	xii
CHAPTER 1 INTRODUCTION	1
Structural Regression Models	4
Thresholds and Polychoric Correlations	7
Least Squares Estimation	11
<i>Robust Corrections to Standard Errors and Test Statistics</i>	15
Maximum Likelihood Estimation	17
<i>Robust Corrections to Standard Errors and Test Statistics</i>	19
CHAPTER 2 EMPIRICAL FINDINGS	22
Parameter Estimates	23
Standard Error Estimates	24
Chi-Square Goodness of Fit Statistics	24
CHAPTER 3 PRESENT STUDY	26
CHAPTER 4 METHOD	32
Model Specification	32
Simulation Design	34
Number of Observed Variables' Categories	35
Ordinal Observed Distributions	36
Sample Size	38
Data Generation and Analysis	40
Outcome Variables	41
CHAPTER 5 RESULTS	47
Non-Convergence and Inadmissible Solutions	47
Parameter Estimates	49
<i>Factor Loadings</i>	49
<i>Structural Coefficients</i>	51
Standard Error Estimates	53
Chi-Square Goodness of Fit Statistics	55
RMSEA	57
CHAPTER 6 DISCUSSION	58
Implications for Applied Research	63

<i>Sample Size</i>	63
<i>Estimation Methods</i>	64
<i>Response Categories and Observed Distributions</i>	67
Limitations and Directions for Future Research	68
 CHAPTER 7 SUMMARY AND CONCLUSIONS	 72
 APPENDICES	 76
Appendix A: Tables	77
Appendix B: Figures	102
Appendix C: Technical Details	110
Appendix D: <i>Mplus</i> Code for Data Generation and Analysis	112
Appendix E: Results for Sample Sizes $N = 400, 750$, and $1,500$	117
Appendix F: Results for Bipolarization Data	132
 REFERENCES	 142

LIST OF TABLES

Table 1	Overview of Six Major Simulation Studies in Ordinal CFA	77
Table 2	Robust Estimation Comparison in the Three SEM Software Packages	77
Table 3	Comparison of Two Major Estimation Approaches: Maximum Likelihood and Least Squares in <i>Mplus</i>	78
Table 4(a)	Cases of Non-Convergence	79
Table 4(b)	Cases of Inadmissible Solutions	80
Table 5	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 200$)	81
Table 6	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 300$)	82
Table 7	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 500$)	83
Table 8	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 1,000$)	84
Table 9	The Average Root Mean Squared Error (MSE_A) for the Four Structural Coefficients ($N = 1,000$)	85
Table 10	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 200$)	86
Table 11	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 300$)	88
Table 12	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 500$)	90

Table 13	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 1,000$)	92
Table 14	Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 200$)	94
Table 15	Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 300$)	96
Table 16	Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 500$)	98
Table 17	Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 1,000$)	100
Table E1	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 400$)	117
Table E2	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 750$)	118
Table E3	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 1,500$)	119
Table E4	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 400$)	120
Table E5	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 750$)	122
Table E6	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 1,500$)	124
Table E7	Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 400$)	126
Table E8	Bias and Rejection Rates of Chi-Square Statistics as well as Means and	

	Rejection Rates of RMSEA ($N = 750$)	128
Table E9	Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 1,500$)	130
Table F1	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients with Bipolarization Distribution	132
Table F2	The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients with Bipolarization Distribution	134
Table F3	Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA with Bipolarization Distribution	138

LIST OF FIGURES

<i>Figure 1</i>	The postulated five-factor structural regression model with standardized Coefficients. <i>Note.</i> Ordinal observed variables of each latent construct are not depicted for clarity.	102
<i>Figure 2</i>	Response probabilities of ordinal observed indicators	103
<i>Figure 3</i>	Average mean squared error for the factor loading estimates across the number of categories with symmetric data and the smallest sample size $N = 200$	104
<i>Figure 4</i>	Average mean squared error for the standard error estimates of factor loadings across the number of categories with slightly asymmetric data and the sample size $N = 300$	105
<i>Figure 5</i>	Average mean squared error for the standard error estimates of factor loadings across the number of categories with slightly asymmetric data and the sample size $N = 1,000$	106
<i>Figure 6</i>	Average mean squared error for the standard error estimates of structural coefficients across the number of categories with slightly asymmetric data and the sample size $N = 300$	107
<i>Figure 7</i>	P-P plots for T_{ML} , T_{MLR} , T_{WLSMV} , and T_{ULSMV} (Moderate Asymmetry and 7-category)	108
<i>Figure 8</i>	P-P plots for T_{ML} , T_{MLR} , T_{WLSMV} , and T_{ULSMV} ($N = 300$ and 7-category)	109

CHAPTER 1

INTRODUCTION

Observed variables measured with a set of ordered categories (e.g., using Likert-type scales) are commonly employed to operationalize latent constructs in the educational, social, and behavioral sciences. Unlike continuous variables, calculation of means, variances, and covariances for ordered observed categorical variables (i.e., ordinal observed variables) is in general meaningless due to the lack of substantively interpretable origins and metrics for these variables (Jöreskog, 2005). When it comes to a statistical model, one with ordinal data on outcome variables entails different parameter specification than a model with continuous response variables. By treating ordinal observed variables as if they were continuous, applied researchers may not only possibly undermine the precision and accuracy of model parameter estimates – to varying degree depending on models, data characteristics, and related circumstances – but also arrive at misleading scientific conclusions drawn from empirical data. This problem, which generally plagues applied researchers utilizing various statistical frameworks, is also inevitable when employing latent variable modeling (LVM), in particular confirmatory factor analysis (CFA) and structural equation modeling (SEM).

Over the past few decades, an extensive body of research has used structural regression models in the applied educational, behavioral, and social science literature. A structural regression model takes into account the measurement error of observed variables, and it simultaneously captures the linear relationships among latent constructs of interest. The most widely known estimator used in structural regression (SR) models is the normal theory-based maximum likelihood (ML) method. This is largely due to its optimal properties of asymptotic unbiasedness, consistency, normality, and efficiency (Bollen, 1989). Use of ML, however,

assumes that the observed variables are continuous and multivariate normally distributed in the population conditional on the covariates if included in the model (Bollen, 1989; Jöreskog, 1969). Therefore, ML is not, strictly speaking, appropriate for observed variables that are scaled ordinally. Several estimators with robust corrections to standard errors and chi-square goodness of fit statistics, such as robust ML (MLR: Muthén & Muthén, 2010), robust unweighted least squares (ULSMV: Muthén, 1993; Satorra & Bentler, 1994), and robust weighted least squares (WLSMV: Muthén, du Toit, & Spisic, 1997), have been proposed in the literature, and are considered superior to “conventional” ML when ordinal data on response variables are employed in latent variable analysis. It is noted in passing that robust ML has been suggested for use when ordinal observed variables have at least five response categories (e.g., Johnson & Creech, 1983; Rigdon, 1998; Raykov, 2012, and references therein). The robust ML estimator is also frequently used by applied researchers, based on the argument that ordinal data on response variables could be considered “approximately continuous” if the number of observed variables’ categories is sufficiently large.

A growing number of simulation studies have compared the relative performance of different estimators in ordinal CFA (Hoogland & Boomsma, 1998). However, one major limitation of previous ordinal CFA simulation studies is that researchers may have devoted less attention to inter-factor correlation estimates. Some simulation studies examined the joint performance of both factor loading and inter-factor correlation estimates (see, e.g., Lei, 2009; Yang-Wallentin, Jöreskog, & Luo, 2010). Yet, the performance of different robust estimation methods on inter-factor correlation estimates is unclear and unexplored. Although a few simulation studies compared the performance of different estimation methods in an SR model with ordinal observed variables (see, e.g., Anderson, 1996; Coender, Satorra, & Saris, 1997),

they failed to incorporate robust corrections to standard errors and chi-square statistics and excluded the effect of number of observed variables' categories. To date, no simulation study has been identified in the extant literature that has employed an SR model with ordinal observed variables to compare the performance of robust estimators since the three robust estimators were developed and made available in widely circulated computer programs. Therefore, the comparison of the performance of robust estimators on structural coefficients remains an open research question. Additionally, the performance of MLR implemented in *Mplus* has not yet been systematically evaluated in the literature. Given that robust estimators have recently received considerable attention also in applied research settings, it can be expected that findings on their performance in structural regression models with ordinal data on response variables, in particular of the three robust estimators (MLR, WLSMV, and ULSMV), would be of particular importance for empirical researchers in the educational, social, and behavioral sciences.

The central objective of this thesis is to carry out a Monte Carlo simulation study addressing gaps in the extant literature and contributing to our understanding of the impact of ordinal observed variables on parameter estimates, in particular (but not limited to) of structural regression coefficients, their associated standard errors, the chi-square goodness of fit statistics, and RMSEA in SR models. Another important objective is to compare the performance of the four different estimators (ML, MLR, ULSMV and WLSMV) in SR models with ordinal observed indicators under different experimental conditions. Findings from this study are expected (1) to inform the work of applied researchers with respect to the importance of attending to assumption violations, and (2) to translate directly into recommendations for selecting an "appropriate" estimator under empirical circumstances frequently encountered in

current research practice. Finally, implications of the findings for structural regression models using these four estimators are discussed, and the limitations of this study as well as potentially directions for future research are discussed.

The remainder of this dissertation is organized as follows. It begins by (1) delineating the parameterization of a structural regression model with ordinal observed variables, followed by (2) describing the estimation of thresholds and polychoric correlations, subsequently (3) introducing two major estimation approaches: least squares and maximum likelihood, then (4) providing a brief review of prior research that has investigated the behavior of the four estimators in applications, (5) presenting the aims of the study, (6) outlining the model specification, simulation design, and evaluation criteria, (7) reporting the results, and finally (8) concluding with a discussion of limitations of this study, recommendations for applied researchers, and directions for future research, as well as a series of brief “take-home” messages for empirical researchers.

Structural Regression Models

A structural regression model (i.e., a structural equation model with a regression relationship between some of its latent variables) permits testing hypothetical associations/relationships among latent variables measured each by a set of observed variables. A structural regression model with ordinal observed variables, in general, consists of two components: (i) the measurement models and (ii) the structural model. The measurement models can be expressed as follows (Bollen, 1989)

$$\underline{y}^* = \underline{v}_{y^*} + \underline{\Lambda}_{y^*}\underline{\eta} + \underline{\varepsilon}, \quad (1)$$

and

$$\underline{x}^* = \underline{v}_{x^*} + \mathbf{\Lambda}_{x^*}\xi + \underline{\delta}, \quad (2)$$

where \underline{v}_{y^*} is a $p \times 1$ vector of intercept terms for \underline{y}^* , \underline{v}_{x^*} a $q \times 1$ vector of intercept terms for \underline{x}^* , \underline{y}^* represents a $p \times 1$ vector of latent response variables y^* 's underlying ordinal observed, endogenous variables y s, \underline{x}^* a $q \times 1$ vector of latent response variables x^* 's underlying ordinal observed, exogenous variables x s, $\mathbf{\Lambda}_{y^*}$ a $p \times m$ matrix of factor loadings for \underline{y}^* , $\mathbf{\Lambda}_{x^*}$ a $q \times n$ matrix of factor loadings for \underline{x}^* , $\underline{\eta}$ an $m \times 1$ vector of endogenous latent variables, ξ a $n \times 1$ vector of exogenous latent variables with $E(\xi) = \underline{\kappa}$ and $\text{Cov}(\xi) = \mathbf{\Phi}$ (a $n \times n$ variance-covariance matrix of latent variables ξ), $\underline{\varepsilon}$ a $p \times 1$ vector of measurement errors in \underline{y}^* with $E(\underline{\varepsilon}) = \underline{0}$ and $\text{Var}(\underline{\varepsilon}) = \mathbf{\Theta}_{\varepsilon}$ (a $p \times p$ diagonal matrix of residual variances for \underline{y}^* , assuming measurement errors $\underline{\varepsilon}$ are uncorrelated with all other measurement errors and latent variables $\underline{\eta}$), $\underline{\delta}$ a $q \times 1$ vector of measurement errors in \underline{x}^* with $E(\underline{\delta}) = \underline{0}$ and $\text{Var}(\underline{\delta}) = \mathbf{\Theta}_{\delta}$ (a $q \times q$ diagonal matrix of residual variances for \underline{x}^* , assuming measurement errors $\underline{\delta}$ are uncorrelated with all other measurement errors and latent variables ξ). It is also assumed that $\underline{\varepsilon}$ is uncorrelated with $\underline{\delta}$.

The structural model is defined as

$$\underline{\eta} = \underline{a} + \mathbf{B}\underline{\eta} + \mathbf{\Gamma}\xi + \underline{\zeta}, \quad (3)$$

where \underline{a} is an $m \times 1$ vector of latent means for $\underline{\eta}$, \mathbf{B} an $m \times m$ matrix of structural regression

coefficients with zero diagonal elements among $\underline{\eta}$ (assuming $|\mathbf{I} - \mathbf{B}| \neq 0$), $\mathbf{\Gamma}$ an $m \times n$ matrix of structural regression coefficients between $\underline{\xi}$ and $\underline{\eta}$, $\underline{\zeta}$ an $m \times 1$ vector of disturbance terms in $\underline{\eta}$ with $E(\underline{\zeta}) = \underline{0}$ and $\text{Cov}(\underline{\zeta}) = \mathbf{\Psi}$ (an $m \times m$ diagonal matrix of residual variances for $\underline{\eta}$, assuming disturbance terms $\underline{\zeta}$ are uncorrelated with all other disturbance terms and latent variables $\underline{\xi}$). It follows that $E(\underline{\eta}) = (\mathbf{I} - \mathbf{B})^{-1}(\underline{a} + \mathbf{\Gamma}\underline{k})$ and $\text{Cov}(\underline{\eta}) = (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}' + \mathbf{\Psi})(\mathbf{I} - \mathbf{B})^{-1}$.

Let $\underline{\theta}$ denote the vector of model parameters. Then, the mean structure for the latent response variable ($\underline{y}^*, \underline{x}^*$) of a general structural regression model parameterized in $\underline{\theta}$ can be expressed as

$$\underline{\mu}(\underline{\theta}) = \begin{bmatrix} \underline{\mu}_{\underline{y}^*} \\ \underline{\mu}_{\underline{x}^*} \end{bmatrix}, \quad (4.1)$$

where $\underline{\mu}_{\underline{y}^*} = \underline{v}_{\underline{y}^*} + \mathbf{\Lambda}_{\underline{y}^*}(\mathbf{I} - \mathbf{B})^{-1}(\underline{a} + \mathbf{\Gamma}\underline{k})$ and $\underline{\mu}_{\underline{x}^*} = \underline{v}_{\underline{x}^*} + \mathbf{\Lambda}_{\underline{x}^*}\underline{k}$.

Similarly, the covariance structure implied by this model can be expressed as

$$\mathbf{\Sigma}^*(\underline{\theta}) = \begin{bmatrix} \mathbf{\Sigma}_{\underline{y}^*\underline{y}^*} & \mathbf{\Sigma}_{\underline{y}^*\underline{x}^*} \\ \mathbf{\Sigma}_{\underline{x}^*\underline{y}^*} & \mathbf{\Sigma}_{\underline{x}^*\underline{x}^*} \end{bmatrix}, \quad (4.2)$$

where $\mathbf{\Sigma}_{\underline{x}^*\underline{x}^*} = \mathbf{\Lambda}_{\underline{x}^*}\mathbf{\Phi}\mathbf{\Lambda}_{\underline{x}^*}' + \mathbf{\Theta}_{\delta}$, $\mathbf{\Sigma}_{\underline{y}^*\underline{y}^*} = \mathbf{\Lambda}_{\underline{y}^*}(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}' + \mathbf{\Psi})(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Lambda}_{\underline{y}^*}' + \mathbf{\Theta}_{\varepsilon}$, and $\mathbf{\Sigma}_{\underline{y}^*\underline{x}^*} = \mathbf{\Lambda}_{\underline{y}^*}(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Lambda}_{\underline{x}^*}'$. Unlike a structural regression model with continuous observed variables, the variances of measurement errors (i.e., the diagonal elements of $\mathbf{\Theta}_{\delta}$ and $\mathbf{\Theta}_{\varepsilon}$) are not identified here. These variances can be identified by either standardizing the latent response

variables \underline{y}^* and \underline{x}^* or standardizing the measurement errors $\underline{\delta}$ and $\underline{\epsilon}$. The former is the default given by the *Delta* parameterization in *Mplus*; the latter is referred to as *Theta* parameterization (Muthén & Muthén, 2010). In order to introduce metrics for the latent response variables, the variances of the latent response variables \underline{y}^* and \underline{x}^* have been assumed for convenience to be equal to 1 when ordinal observed variables are observed. Therefore, Θ_{δ} has to be constrained as

$$\Theta_{\delta} = \mathbf{I} - \text{diag}(\Lambda_{\underline{x}^*} \Phi \Lambda'_{\underline{x}^*}), \quad (4.3)$$

and Θ_{ϵ} has to be constrained accordingly as

$$\Theta_{\epsilon} = \mathbf{I} - \text{diag}(\Lambda_{\underline{y}^*} (\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Psi) (\mathbf{I} - \mathbf{B})^{-1'} \Lambda'_{\underline{y}^*}). \quad (4.4)$$

As a consequence, the $\Sigma^*(\underline{\theta})$ has unit diagonal elements and therefore reduces as a correlation matrix implied by the model under consideration. Next, the relationships between the latent constructs (η and ξ) and underlying latent response variables (\underline{y}^* and \underline{x}^*) are estimated via analysis of the correlation matrix among the latent response variables \underline{y}^* and \underline{x}^* , using the ordinal observed data.

Thresholds and Polychoric Correlations

A correlation between two normal, latent response variables is referred to as a polychoric correlation, for which the two ordinal observed indicators have at least three response categories. A polychoric correlation is typically estimated using a two-stage procedure proposed by Olsson (1979; also see Bollen, 1989; Jöreskog, 2005): (i) the

estimation of thresholds from the univariate marginal distributions, and (ii) the estimation of polychoric correlations through the bivariate marginal distributions for given the threshold estimates.

A continuous, normal, latent response variable y^* underlies an ordinal observed variable y in the population:

$$y = c, \text{ if } \tau_{c-1} < y^* < \tau_c, \quad c = 1, 2, \dots, g, \quad (5)$$

where c defines the observed value of an ordinal variable y , τ is the threshold ($-\infty = \tau_0 < \tau_1 < \tau_2 \dots < \tau_{g-1} < \tau_g = \infty$), and g is the number of ordered categories. In the educational, behavioral, and social sciences, many latent constructs of interest are “conceptually” continuous, and therefore assuming an underlying continuous y^* is a reasonable approach (Coenders, Satorra, & Saris, 1997). For example, a respondent tends to endorse the k^{th} response category when her/his latent response value y^* lies between τ_{k-1} and τ_k . The ordinal observed data only provide an approximation of the underlying continuous, latent response variable because ordered observed categorical data in nature are discrete. A standard normal distribution is selected for the latent response variable y^* with a probability density function $\phi_1(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, -\infty < u < \infty$ and a cumulative distribution function $\Phi_1(u)$. The probability of the i^{th} category response is obtained as

$$\pi_i = p(y = i) = p(\tau_{i-1} < y^* < \tau_i) = \int_{\tau_{i-1}}^{\tau_i} \phi_1(u) du = \Phi_1(\tau_i) - \Phi_1(\tau_{i-1}), \quad (6)$$

and it follows that

$$\tau_i = \Phi_1^{-1}(\pi_1 + \pi_2 + \dots + \pi_i), \quad i = 1, 2, \dots, g-1, \quad (7)$$

where Φ_1^{-1} is the inverse of the standard normal cumulative distribution function. Next, τ_i can be estimated as

$$\hat{\tau}_i = \Phi_1^{-1}(p_1 + p_2 + \dots + p_i), \quad i = 1, 2, \dots, g-1, \quad (8)$$

where p_i is the sample proportion of responses in category i . It is noted that the threshold estimation model is saturated. Namely, the number of threshold parameters (i.e., $m-1$) is equal to the number of non-redundant sample proportions.

Let each ordinal observed variable y_1 and y_2 have g_1+1 ($\tau_{1,1}, \tau_{1,2}, \dots, \tau_{1,g_1}$) and g_2+1 ($\tau_{2,1}, \tau_{2,2}, \dots, \tau_{2,g_2}$) categories, respectively. Assume that underlying variables y_1^* and y_2^* are both standard normal distributions with zero means, unit variances, and a correlation ρ . A standard bivariate normality of y_1^* and y_2^* is also assumed with its probability density function

$\phi_2(u, v, \rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} e^{-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}}, -\infty < u, v < \infty$. This correlation ρ between y_1^* and y_2^* defines a polychoric correlation. The likelihood function of yielding the observed bivariate sample can be defined as

$$L = C \prod_{i=1}^{g_1+1} \prod_{j=1}^{g_2+1} \pi_{ij}^{n_{ij}}, \quad (9)$$

where C is a constant and n_{ij} is the frequency in cell (i, j) of a bivariate contingency table. π_{ij} is the probability in cell (i, j) defined as

$$\begin{aligned}\pi_{ij} &= p(y_1 = i, y_2 = j) = p(\tau_{1,i-1} < y_1^* < \tau_{1,i}, \tau_{2,j-1} < y_2^* < \tau_{2,j}) \\ &= \int_{\tau_{1,i-1}}^{\tau_{1,i}} \int_{\tau_{2,j-1}}^{\tau_{2,j}} \phi_2(u, v, \rho) du dv,\end{aligned}\tag{10}$$

which can then be rewritten as

$$\pi_{ij} = \Phi_2(\tau_{1,i}, \tau_{2,j}, \rho) - \Phi_2(\tau_{1,i}, \tau_{2,j-1}, \rho) - \Phi_2(\tau_{1,i-1}, \tau_{2,j}, \rho) + \Phi_2(\tau_{1,i-1}, \tau_{2,j-1}, \rho),\tag{11}$$

where Φ_2 is the standard bivariate normal cumulative distribution function with the correlation coefficient ρ . Take the natural logarithm of the likelihood function L and the partial derivative on $\ln L$ with respect to ρ :

$$\ln L = \ln C + \sum_{i=1}^{g_1+1} \sum_{j=1}^{g_2+1} n_{ij} \ln \pi_{ij},\tag{12}$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \rho} &= \sum_{i=1}^{g_1+1} \sum_{j=1}^{g_2+1} \frac{n_{ij}}{\pi_{ij}} [\Phi_2(\widehat{\tau}_{1,i}, \widehat{\tau}_{2,j}, \rho) - \Phi_2(\widehat{\tau}_{1,i}, \widehat{\tau}_{2,j-1}, \rho) \\ &\quad - \Phi_2(\widehat{\tau}_{1,i-1}, \widehat{\tau}_{2,j}, \rho) + \Phi_2(\widehat{\tau}_{1,i-1}, \widehat{\tau}_{2,j-1}, \rho)],\end{aligned}\tag{13}$$

Threshold estimates are obtained using sample cumulative marginal proportions of the bivariate contingency table, for example,

$$\widehat{\tau}_{1,i} = \Phi_1^{-1}(\sum_{k=1}^i \sum_{j=1}^{g_2+1} p_{kj}),\tag{14}$$

where p_{kj} is the sample proportion in cell (k, j) . Next, solve the equation $\frac{\partial \ln L}{\partial \rho} = 0$ (i.e., maximizing $\ln L$) using the given threshold estimates to obtain the polychoric correlation estimate $\hat{\rho}$. It is noted that a Pearson product-moment correlation between two ordinal observed variables is generally attenuated because the underlying continuum is coarsely categorized to obtain ordinal observed variables. A greater amount of attenuation in Pearson product-moment correlation estimates occurs when ordinal observed variables have only a few alternatives, and/or opposite skewed and increasingly leptokurtic distributions (Bollen, 1989; Olsson, 1979; Muthén & Kaplan, 1992). In the next section, two estimation families used in SEM with ordinal observed variables to obtain model parameters, standard errors, and chi-square goodness of fit statistics are introduced in turn: least squares and maximum likelihood approaches.

Least Squares Estimation

Muthén (1984) made a substantial breakthrough in analyzing a structural equation model with ordinal observed variables using a weighted least squares (WLS) approach. The thresholds and polychoric correlations are first estimated using two-stage ML estimation in the preceding paragraph. Parameter estimates are then obtained using a consistent estimator of the asymptotic covariance matrix of the polychoric correlation and threshold estimates (denoted as $\hat{\mathbf{V}}$) in a weight matrix \mathbf{W} , to minimize the weighted least squares fit function (Muthén, 1984):

$$F_{\text{WLS}} = [\underline{s} - \underline{\sigma}(\underline{\theta})]' \mathbf{W}^{-1} [\underline{s} - \underline{\sigma}(\underline{\theta})], \quad (15)$$

where $\underline{\theta}$ is the vector of model parameters, $\underline{\sigma}(\underline{\theta})$ is the model-implied vector consisting of the non-duplicated, vectorized elements of $\Sigma^*(\underline{\theta})$ (i.e., $\text{vech}[\Sigma^*(\underline{\theta})]$), and \underline{s} is the vector containing the non-duplicated, vectorized elements of sample statistics (i.e., threshold and polychoric correlation estimates). Note that the $\text{vech}(\cdot)$ operator strings out non-redundant matrix elements by stacking them up into a column vector, leaving out the upper-diagonal elements. The weight matrix includes variability of threshold and polychoric correlation estimates and interrelationships among polychoric correlation estimates. This procedure only incorporates univariate and bivariate margins into the estimation of model parameters, and it often has been termed as *limited information* estimation, in contrast to *full information* that uses subjects' complete multivariate response pattern, typically paralleling the item response theory (IRT) framework (see, e.g., Forero & Maydeu-Olivares, 2009; Wirth & Edwards, 2007, for a full discussion of limited information vs. full information).

Standard errors are given by the square roots of the diagonals of the asymptotic covariance matrix of the parameter estimates $\hat{\underline{\theta}}$ from a Taylor expansion (see, e.g., Browne, 1984; Satorra, 1989):

$$\widehat{\text{aCov}}(\hat{\underline{\theta}})_{\text{WLS}} = N^{-1}(\hat{\Delta}'\mathbf{W}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\mathbf{W}^{-1}\hat{\mathbf{V}}\mathbf{W}^{-1}\hat{\Delta}(\hat{\Delta}'\mathbf{W}^{-1}\hat{\Delta})^{-1}, \quad (16)$$

and because of $\mathbf{W} = \hat{\mathbf{V}}$, it reduces to

$$\widehat{\text{aCov}}(\hat{\underline{\theta}})_{\text{WLS}} = N^{-1}[\hat{\Delta}'\hat{\mathbf{V}}^{-1}\hat{\Delta}]^{-1}, \quad (17)$$

where N represents the sample size, $\hat{\mathbf{A}} = \frac{\partial \mathbf{g}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}$ is the so-called Jacobian matrix of first derivatives when evaluating at the parameter estimates $\hat{\boldsymbol{\theta}}$, and $\hat{\mathbf{V}}$ is the estimated asymptotic covariance matrix of $\underline{\mathbf{s}}$. The chi-square goodness of fit statistic is defined as

$$T_{\text{WLS}} = (N - 1) F_{\text{WLS}}(\hat{\boldsymbol{\theta}}, \underline{\mathbf{s}}), \quad \text{df} = s - t, \quad (18)$$

where s = the number of unique elements in $\underline{\mathbf{s}}$ and t = the number of independent model parameters. That is, degrees of freedom are the difference between the number of parameters in the unrestricted model and the number of parameters in the estimated model. However, the performance of WLS deteriorates with small sample sizes and/or model complexity, mainly because of the size and the invertibility of the weight matrix $\mathbf{W} = \hat{\mathbf{V}}$. Specifically, WLS has been subject to non-convergence problems with small sample sizes and/or complex models in simulation studies (Flora & Curran, 2004; Oranje, 2003). As the number of ordinal observed variables increases, the size of $\hat{\mathbf{V}}$ grows exponentially, leading to demanding computations and numerical problems in the process of estimation. In addition, when sample sizes are small, the estimated asymptotic covariance matrix $\hat{\mathbf{V}}$ has much sampling variation, and the inversion of $\hat{\mathbf{V}}$ is typically infeasible as well (Browne, 1984; Jöreskog & Sörbom, 1996; Muthén, 1993). These weaknesses render the WLS estimator less attractive for applications.

Empirical research has also suggested that WLS is inferior to other Least-Squares-family estimators (e.g., WLSMV or ULSMV) in CFA models when the sample size is small and/or the model becomes complicated (Flora & Curran, 2004; Oranje, 2003; Yang-Wallentin, Jöreskog, & Luo, 2010). Flora and Curran (2004) found that (1) parameter

estimates were less overestimated by WLSMV than WLS; (2) standard errors were less negatively biased by WLSMV than WLS, relative to the standard deviation of parameter estimates across replications; and (3) chi-square statistics were less inflated by WLSMV than WLS. Yang-Wallentin, Jöreskog, & Luo (2010) revealed that the performance of WLS was uniformly worse in terms of parameter estimates, standard errors, and chi-square statistics, than WLSMV and ULS with robust corrections.

One possible way to circumvent the troubling features and ease the computational burden is to choose a simple weight matrix, such as the identity matrix \mathbf{I} , or a reduced and invertible form of $\hat{\mathbf{V}}$ (e.g., retaining diagonal elements of $\hat{\mathbf{V}}$ only). The former choice simplifies WLS to unweighted least squares (ULS: Muthén, 1993), and the latter reduces to diagonally weighted least squares (DLS (two-step weighted least squares): Christofferson, 1977; DWLS: Jöreskog & Sörbom, 1996; robust WLS or WLSMV: Muthén, du Toit, & Spisic, 1997). The fit function for each can be represented as follows

$$F_{\text{ULS}} = [\underline{s} - \underline{\sigma}(\underline{\theta})]' (\mathbf{I})^{-1} [\underline{s} - \underline{\sigma}(\underline{\theta})], \quad (19)$$

and

$$F_{\text{D-WLS}} = [\underline{s} - \underline{\sigma}(\underline{\theta})]' (\mathbf{W}_D)^{-1} [\underline{s} - \underline{\sigma}(\underline{\theta})], \quad (20)$$

where $\mathbf{W}_D = \text{diag}(\hat{\mathbf{V}})$ contains only diagonal elements of the estimated asymptotic covariance matrix of the polychoric correlation and threshold estimates. Throughout this dissertation, D-WLS is used to represent diagonally weighted least squares due to various terms used in

currently circulated computer programs. More specifically, D-WLS only weights the residual vector $[\underline{s} - \underline{g}(\theta)]$ using the asymptotic “variances” of polychoric correlation and threshold estimates, and ULS weights all elements of the residual vector “equally” using the identity matrix \mathbf{I} (Bollen, 1989; Muthén & Muthén, 2010).

Robust Corrections to Standard Errors and Test Statistics

Unlike the aforementioned full weight matrix \mathbf{W} in WLS, \mathbf{I} and \mathbf{W}_D only contain limited or reduced/partial information in the weight matrix. A disadvantage of ULS is that the weight matrix \mathbf{I} makes obtained parameter estimates less sensitive to differences in the elements of the residual vector (Bolt, 2005). Although improvement can be expected while using the diagonal weight matrix \mathbf{W}_D , the asymptotic covariances between polychoric correlation estimates are still left outside the estimation procedure. The parameter estimates obtained by ULS and D-WLS are therefore not asymptotically efficient (i.e., smaller sampling error), resulting in potentially inaccurate standard error estimates. That is, the WLS parameter estimates have the smallest variances within the class of least squares estimators. Because both ULS and D-WLS are less efficient than WLS, upward corrections applied to standard errors are suggested. Underestimation of standard errors may affect statistical inferences for parameter estimates. Robust correction to standard errors are implemented in the estimated asymptotic covariance matrix of the parameter estimates $\hat{\theta}$ for ULS estimation (Muthén, 1993; Satorra & Bentler, 1994):

$$\widehat{\text{aCov}}(\hat{\theta})_{\text{ULS}} = N^{-1}(\hat{\mathbf{A}}'\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}'\hat{\mathbf{V}}\hat{\mathbf{A}}(\hat{\mathbf{A}}'\hat{\mathbf{A}})^{-1}, \quad (21)$$

and for D-WLS estimation (Muthén, du Toit, & Spisic, 1997):

$$\widehat{\text{aCov}}(\hat{\theta})_{\text{D-WLS}} = N^{-1}(\hat{\Delta}'\mathbf{W}_D^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\mathbf{W}_D^{-1}\hat{\mathbf{V}}\mathbf{W}_D^{-1}\hat{\Delta}(\hat{\Delta}'\mathbf{W}_D^{-1}\hat{\Delta})^{-1}. \quad (22)$$

Likewise, because of using a consistent estimator of the asymptotic covariance matrix of the polychoric correlation and threshold estimates ($\hat{\mathbf{V}}$) as the full weight matrix, T_{WLS} is asymptotically chi-square distributed. However, the standard test statistics T_{ULS} and T_{WLS} are not appropriate for model fit evaluation because the test statistics produced by ULS and D-WLS are no longer asymptotically chi-square distributed. This robust correction entails adjusting both the mean and variance of the test statistics. Therefore the mean- and variance-adjusted chi-square statistic can each be implemented in the ULS estimator (Asparouhov & Muthén, 2010):

$$T_{\text{ULSMV}} = aT_{\text{ULS}} + b, \quad \text{df} = s - t, \quad (23)$$

where $T_{\text{ULS}} = (N - 1) F_{\text{ULS}}(\hat{\theta}, \underline{s})$, $\hat{\mathbf{V}}$ is the estimated asymptotic covariance matrix of \underline{s} , $a =$

$\sqrt{\frac{\text{df}}{\text{trace}(\hat{\mathbf{U}}\hat{\mathbf{V}}\hat{\mathbf{U}}\hat{\mathbf{V}})}}$ is a scale factor, $b = \text{df} - \sqrt{\frac{\text{df}[\text{trace}(\hat{\mathbf{U}}\hat{\mathbf{V}})]^2}{\text{trace}(\hat{\mathbf{U}}\hat{\mathbf{V}}\hat{\mathbf{U}}\hat{\mathbf{V}})}}$ is a shift parameter, and $\hat{\mathbf{U}} = \mathbf{I} - \hat{\Delta}(\hat{\Delta}'\hat{\Delta})^{-1}\hat{\Delta}'$;

and in the D-WLS estimator (Asparouhov & Muthén, 2010):

$$T_{\text{D-WLSMV}} = \check{a}T_{\text{D-WLS}} + \check{b}, \quad \text{df} = s - t, \quad (24)$$

where $T_{\text{D-WLS}} = (N - 1) F_{\text{D-WLS}}(\hat{\theta}, \underline{s})$, $\hat{\mathbf{V}}$ is the estimated asymptotic covariance matrix of \underline{s} , $\check{a} =$

$\sqrt{\frac{\text{df}}{\text{trace}(\check{\mathbf{U}}\hat{\mathbf{V}}\check{\mathbf{U}}\hat{\mathbf{V}})}}$, $\check{b} = \text{df} - \sqrt{\frac{\text{df}[\text{trace}(\check{\mathbf{U}}\hat{\mathbf{V}})]^2}{\text{trace}(\check{\mathbf{U}}\hat{\mathbf{V}}\check{\mathbf{U}}\hat{\mathbf{V}})}}$, and $\check{\mathbf{U}} = \mathbf{W}_D^{-1} - \mathbf{W}_D^{-1}\hat{\Delta}(\hat{\Delta}'\mathbf{W}_D^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\mathbf{W}_D^{-1}$. Unlike WLS, $\hat{\mathbf{V}}$

need not be inverted (i.e., a positive definite matrix) in the computation of robust standard

errors and adjusted chi-square test statistics using ULS and D-WLS. Both T_{ULSMV} and $T_{D-WLSMV}$ result in smaller test statistics in comparison to T_{WLS} . That is, chi-square statistics in the robust estimators are downwardly adjusted to compensate for the effect of only including limited or reduced/partial information in the weight matrix. This correction can help control for the probability of Type I error (i.e., rejecting a correctly specified model by chance). Furthermore, this new second order chi-square correction has been implemented in *Mplus* 6 and later versions. For the Satterthwaite (1941) type correction prior to *Mplus* 6, refer to Satorra & Bentler (1994) and Muthén, du Toit, & Spisic (1997).

It is worth reiterating that (1) the aim of the robust corrections to standard errors in the already available ULS and D-WLS estimators is to compensate for the loss of efficiency (i.e., smaller variability of parameter estimates) when the full weight matrix is not performed; and (2) the mean- and variance-adjustments for test statistics in ULS and D-WLS estimators are targeted to make the shape of test statistics be approximately close to the reference chi-square distribution with the associated degrees of freedom. Note that the mean-adjusted chi-square statistic in diagonally weighted least squares estimation is not presented here (i.e., ESTIMATOR = WLSM, see Appendix C for details).

Maximum Likelihood Estimation

When the assumption of multivariate normality is considered tenable in a SEM model with “*continuous*” observed variables, parameter estimates can be obtained by maximizing the likelihood of the observed data; that is, the minimization of the maximum likelihood fit function (Bollen, 1989):

$$F_{ML} = \ln|\mathbf{\Sigma}(\underline{\Theta})| + \text{trace}[\mathbf{S}\mathbf{\Sigma}^{-1}(\underline{\Theta})] - \ln|\mathbf{S}| - r, \quad (25)$$

where $\underline{\Theta}$ denotes the vector of model parameters, $\mathbf{\Sigma}(\underline{\Theta})$ is the model-implied “covariance” matrix, \mathbf{S} is the sample-based “covariance” matrix, and $r (= p + q)$ is the total number of continuous observed variables in the model. Under the multivariate normality assumption, standard errors are the square roots of the diagonal elements of the estimated asymptotic covariance matrix for $\hat{\underline{\Theta}}$ from F_{ML} :

$$\widehat{\text{aCov}}(\hat{\underline{\Theta}})_{ML} = \left(\frac{2}{N-1} \right) \left\{ E \left[\frac{\partial^2 F_{ML}}{\partial \underline{\Theta} \partial \underline{\Theta}'} \right] \right\}^{-1}. \quad (26)$$

The test statistic that uses Wishart-based likelihood is defined as

$$T_{ML} = (N - 1) F_{ML}(\hat{\underline{\Theta}}, \mathbf{S}), \quad \text{df} = s - t, \quad (27)$$

where s = the number of unique elements in \mathbf{S} and t = the number of independent model parameters (Bollen, 1989; Muthén & Muthén, 2010). However, it is generally not advisable to use ML for ordinal observed variables with only a few response categories. In order to use ML, one may assume that a given set of ordinal observed variables are “approximately continuous” if they have more than five response alternatives, and further one treats them as if they were continuous. The normality of ordinal observed variables due to categorization is typically not plausible. The superiority of the robust ML method (MLR) over the normal theory-based ML method has proved manifested in the extant literature when modeling ordinal observed variables.

Robust Corrections to Standard Errors and Test Statistics

In order to accommodate ordinal data on response variables (i.e., approximately continuous), standard errors and chi-square goodness of fit statistics are corrected in the MLR estimation to enhance robustness against the presence of non-normality. Ordinal observed variables are rarely normally distributed but often exhibit non-normality in the form of asymmetry to some degree (Micceri, 1989). Acquiescence (or disacquiescence) response style may introduce both skewed and leptokurtic distributions, whereas extreme response style may result in slightly skewed and platykurtic distributions (Weijters, Geuens, & Schillewaert, 2010). The parameter estimates obtained with ML are not asymptotically efficient, provided that the normality assumption is not tenable. The obtained $\widehat{aCov(\hat{\theta})}_{ML}$ in equation (26) is no longer consistent for the asymptotic covariance matrix of $\hat{\theta}$, leading to inaccurate standard error estimates (Yuan, Bentler, & Zhang, 2005; Yuan & Hayashi, 2006). Rather, a consistent estimator of the asymptotic covariance matrix of the parameter estimates $\hat{\theta}$ for MLR can be estimated using the pseudo maximum likelihood (PML) approach (Asparouhov & Muthén, 2005; Savalei, 2010; Yuan & Schuster, 2013):

$$\widehat{aCov(\hat{\theta})}_{MLR} = N^{-1}(\widehat{\Delta}'\widehat{I}_{OB}\widehat{\Delta})^{-1}\widehat{\Delta}'\widehat{I}_{OB}\widehat{V}\widehat{I}_{OB}\widehat{\Delta}(\widehat{\Delta}'\widehat{I}_{OB}\widehat{\Delta})^{-1}, \quad (28)$$

and

$$\widehat{I}_{OB} = \mathbf{D}'\{\mathbf{\Sigma}^{-1}(\hat{\theta}) \otimes [(\mathbf{\Sigma}^{-1}(\hat{\theta})\mathbf{S}\mathbf{\Sigma}^{-1}(\hat{\theta}) - \frac{1}{2}\mathbf{\Sigma}^{-1}(\hat{\theta}))]\}\mathbf{D}, \quad (29)$$

where $\tilde{\mathbf{A}} = \frac{\partial \underline{\mathbf{g}}(\underline{\hat{\theta}})}{\partial \underline{\hat{\theta}}}$ is the matrix of model first derivatives evaluated at the parameter estimates $\underline{\hat{\theta}}$, $(\tilde{\mathbf{A}}' \widehat{\mathbf{I}}_{\text{OB}} \tilde{\mathbf{A}})$ is the estimated “observed” information matrix, and $\tilde{\mathbf{V}}$ is the estimated asymptotic covariance matrix of \mathbf{S} . The “duplication” matrix \mathbf{D} is of order $r^2 \times \frac{1}{2}r(r+1)$ (r = the number of observed variables in $\mathbf{Z}(\underline{\theta})$, see Magnus & Neudecker, 1986, p. 172) and \otimes denotes a Kronecker product. Note that \mathbf{D} is utilized to transform a $r^2 \times r^2$ symmetric matrix, $\mathbf{\Sigma}^{-1}(\underline{\hat{\theta}}) \otimes (\mathbf{\Sigma}^{-1}(\underline{\hat{\theta}}) \mathbf{S} \mathbf{\Sigma}^{-1}(\underline{\hat{\theta}}) - \frac{1}{2} \mathbf{\Sigma}^{-1}(\underline{\hat{\theta}}))$, into a $\frac{1}{2}r(r+1) \times \frac{1}{2}r(r+1)$ symmetric matrix, $\widehat{\mathbf{I}}_{\text{OB}}$. The middle matrix $\tilde{\mathbf{A}}' \widehat{\mathbf{I}}_{\text{OB}} \tilde{\mathbf{V}} \widehat{\mathbf{I}}_{\text{OB}} \tilde{\mathbf{A}}$ contains the sample estimates of skewness and kurtosis of observed variables in order to correct the possible violation of normality assumption (Yuan, Bentler, & Zhang, 2005). While modeling non-normal data, the ML standard error estimates in general are deflated, whereas the robust standard errors obtained with MLR are therefore adjusted upward to alleviate some underestimation of standard error estimates.

As is well known, non-normality of observed variables could lead to substantial overestimation of chi-square goodness of fit statistics. Similar to the two variants of the Yuan-Bentler (1997, 1998) and the Satorra & Bentler (1994) robust chi-square statistics, a modification of chi-square statistics proposed by Asparouhov & Muthén (2005) using the pseudo maximum likelihood (PML) estimator is defined as

$$T_{\text{MLR}} = \tilde{a} T_{\text{ML}}, \quad \text{df} = s - t, \quad (30)$$

where $\tilde{a} = \frac{\text{df}}{\text{trace}[(\tilde{\mathbf{V}} \widehat{\mathbf{I}}_{\text{OB}})] - \text{trace}[\widehat{\text{aCov}}(\underline{\hat{\theta}})_{\text{MLR}} (\tilde{\mathbf{A}}' \widehat{\mathbf{I}}_{\text{OB}} \tilde{\mathbf{A}})]}$ is a scale factor, $T_{\text{ML}} = (N - 1) F_{\text{ML}}(\underline{\hat{\theta}}, \mathbf{S})$, T_{MLR}

denotes the robust ML chi-square test statistic using MLR estimation in *Mplus*, $\tilde{\mathbf{V}}$ is the estimated asymptotic covariance matrix of \mathbf{S} , s = the number of unique elements in \mathbf{S} , and t =

the number of total model parameters. The scale factor \tilde{a} is used to remove the effect of skewness and kurtosis of observed data in order to adjust for deviation from normality. T_{MLR} was found to perform well under a variety of conditions investigated by Asparouhov & Muthén (2005). It is worth noting that the downward adjustments for test statistics in MLR can yield the distributional behavior of test statistics that more closely follows a central chi-square in the presence of non-normality. Note that different robust corrections to standard errors and chi-square statistics in maximum likelihood estimation computations are also available but outside the scope of this study (i.e., ESTIMATOR = MLM or MLMV, see Appendix C for details).

CHAPTER 2

EMPIRICAL FINDINGS

A review of simulation studies across six high-impact journals was conducted to determine whether a Monte Carlo simulation study examined ordinal confirmatory factor analysis or structural equation modeling with ordinal observed variables over 20 years (between the years 1994 and 2013) in *Structural Equation Modeling*, *Psychological Methods*, *Multivariate Behavioral Research*, *Psychometrika*, *Educational and Psychological Measurement*, and *Applied Psychological Measurement*. I have identified a total of 13 studies carrying out structural equation modeling with ordinal observed variables (4 articles) or ordinal confirmatory factor analysis (9 articles). The two studies using structural regression models with ordinal observed indicators examined the effect of parceling methods for categorical variable methodology, which is less relevant to the goals of current research. For the other two studies, Anderson (1996) mainly focused on an evaluation of distributional misspecification corrections applied to the McDonald Fit Index that was rarely used in empirical studies and typically not provided in software programs. Coenders, Satorra, and Saris (1997) examined the performance of three correlation estimation methods in an SR model, and their attention was only restricted to point estimates of model parameters using the normal-theory maximum likelihood method and the weighted least squares procedure. The eight studies associated with ordinal confirmatory factor analysis typically compared the relative performance of different estimators on parameter estimates (i.e., factor loadings, inter-factor correlations if any), standard errors, and chi-square goodness of fit statistics. The empirical findings, using ML and the three robust estimators (MLR, ULSMV, and WLSMV), can be briefly summarized below. Table 1 lists 6 major simulation studies that have investigated the performance of the three robust estimators in ordinal CFA models. Because MLR has not been systematically studied in

the previous simulation literature across the six aforementioned journals, a review of robust corrections to standard errors and chi-square goodness of fit statistics in least squares and maximum likelihood estimators is included with all other robust methods in *Mplus*, EQS, and LISREL (see Table 2 for comparison of the three robust estimators in the 3 different SEM software programs; see Table 3 for comparison of the two major estimation approaches in *Mplus*). While these robust standard errors and chi-square statistics may exhibit very slight differences across varying adjustments in a finite sample, they should be asymptotically equivalent as the sample size approaches infinity.

Parameter Estimates

Factor loading estimates were less biased by WLSMV than ML and MLR, even with more than five response alternatives (Beauducel & Herzberg, 2006). Relative bias in factor loading estimates from ULSMV was equal to or smaller than WLSMV (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009) across the conditions (i.e., varying distributions of ordinal observed variables, numbers of observed variables' categories) investigated, and relative bias in factor loading estimates from ULSMV was smaller than ML and MLR even with more than seven response alternatives (Rhemtulla, Brosseau-Liard, & Savalei, 2012), irrespective of the level of asymmetric distributions of ordinal observed variables. Inter-factor correlations were, generally, less overestimated by ML and MLR than WLSMV (Beauducel & Herzberg, 2006) and ULSMV (Rhemtulla, Brosseau-Liard, & Savalei, 2012) across varying numbers of observed variables' categories from two to seven, except under extremely asymmetric distributions of ordinal observed indicators. However, Yang-Wallentin, Jöreskog, & Luo (2010) gave empirical evidence that parameter estimates (consisting of factor loadings and inter-factor correlations jointly) were essentially unbiased for ULSMV, WLSMV, MLR, and ML, regardless of the number

of observed variables' categories and the level of asymmetric distributions of ordinal observed variables. Lei (2009) found that relative bias in parameter estimates (including both factor loadings and inter-factor correlations) was generally negligible for WLSMV, MLR, and ML across different distributions of ordinal observed variables. Oranje (2003) concluded that ML, MLR, and WLSMV produced equally accurate parameter estimates across different numbers of observed variables' categories.

Standard Error Estimates

The "uncorrected" standard errors of factor loadings produced by ML were higher than the robust standard errors of those obtained by WLSMV across different numbers of observed variables' categories (Beauducel & Herzberg, 2006). However, the "uncorrected" standard errors of factor loadings produced by ULS were more accurate, in terms of the standard deviation of parameter estimates over replication, than the robust standard errors of factor loadings produced by WLSMV (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009). The robust standard errors of parameter estimates (including both factor loadings and inter-factor correlations) produced by ULSMV and WLSMV were generally less biased than those obtained by robust ML, regardless of the number of observed variables' categories and the level of asymmetric distributions of ordinal observed variables (Yang-Wallentin, Jöreskog, & Luo, 2010; Lei, 2009). More specifically, Rhemtulla, Brosseau-Liard, and Savalei (2012) revealed that ULSMV produced less biased standard errors of factor loadings than MLMV, whereas ULSMV produced more biased standard errors of inter-factor correlations than MLMV, consistently across different numbers of observed variables' categories.

Chi-Square Goodness of Fit Statistics

The “uncorrected” chi-square statistics produced by ML tended to over-reject the proposed models compared to the robust chi-square statistics obtained by WLSMV, when the number of observed variables’ categories was less than 4 (Beauducel & Herzberg, 2006). The “mean-adjusted” chi-square statistics obtained by MLM provided the most correct rejection rates compared to those obtained by WLSMV across varying numbers of observed variables’ categories (Oranje, 2003). On the contrary, the “mean- and variance-adjusted” chi-square statistics obtained by WLSMV have shown to be slightly more powerful than the “mean-adjusted” chi-square statistics produced by MLM across different levels of asymmetric distributions of ordinal observed variables (Lei, 2009). On the other hand, the “mean- and variance-adjusted” chi-square statistics were comparably good for MLMV and ULSMV when the number of observed variables’ categories ranged from four to six. Furthermore, when the number of observed variables’ categories was two or three, the “mean- and variance-adjusted” chi-square statistics produced by MLMV tended to over-reject the proposed models, and those obtained by ULSMV were likely to under-reject the proposed models (Rhemtulla, Brosseau-Liard, & Savalei, 2012). Finally, the “mean-adjusted” chi-square statistics were essentially equal for MLM, ULSM, and WLSM (Yang-Wallentin, Jöreskog, & Luo, 2010), regardless of the number of observed variables’ categories and the level of asymmetric distributions of ordinal observed variables.

CHAPTER 3

PRESENT STUDY

The present study was designed to address gaps in the literature and to advance our understanding of the impact of ordinal observed variables on parameter estimates, standard errors, and chi-square goodness of fit statistics in a structural regression (SR) model using ML, MLR, ULSMV, and WLSMV. An SR model was selected to broaden the scope of methodological perspectives beyond any previous study in terms of model complexity, because numerous simulation studies have been conducted with ordinal CFA models under extensive conditions. Two literature reviews reported that the median number of latent factors for a CFA model was 3, and the median number of observed indicators in total was 16, indicating that a CFA model in the areas of scale development and item analysis is in general smaller than an SR model in applied settings (Jackson, Gillaspay, & Purc-Stephenson, 2009; DiStefano & Hess, 2005). One major limitation of previous ordinal CFA simulation studies is that researchers have devoted excessive attention to factor loading estimates instead of inter-factor correlation estimates. More specifically, they have (1) simply excluded the inter-factor correlations (e.g., Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009); (2) used homogeneous values for the population inter-factor correlations (e.g., Beauducel & Herzberg, 2006; DiStefano, 2002); or (3) examined the joint performance of both factor loading and inter-factor correlation estimates (e.g., Lei, 2009; Yang-Wallentin, Jöreskog, & Luo, 2010), leaving the undetermined performance of inter-factor correlation estimates.

When a researcher employs an SR model to study relational phenomena among latent constructs of interest, the ultimate goal is to identify successfully the structural coefficients (inter-factor correlations, structural regression coefficients, and possibly mediating effects),

given a tenable measurement model. The SR model proposed here has its practical advantage of allowing applied researchers to study the inter-factor correlations, direct effects, and mediating/indirect effects, which are not uncommon in published research. Heterogeneity of structural regression coefficients and inter-factor correlations in the proposed SR model is more realistic in applied settings and can also assess the effect of structural coefficient magnitude.

Hoogland and Boomsma (1998) systematically reviewed 34 studies in SEM from 1984 to 1994. They found that 89% used CFA models and 11% employed SR models. The aforementioned literature search from 1994 to 2013 that I conducted reflects a growing interest in SR models (about 30% of 59 studies). Researchers also arrived at the same recommendation that future research on a more complex SR model is needed (see, e.g., Bandalos, 2006; Beauducel & Herzberg, 2006; Flora & Curran, 2004; Rhemtulla, Brosseau-Liard, & Savalei, 2012). In this study, an effort was undertaken to extend the existing literature (Anderson, 1996; Coenders, Satorra, & Saris, 1997; Ethington, 1987) on sample size, ordinal observed distributions, and number of observed variables' categories to a broader set of structural regression models with ordinal observed variables. This study aims to address several important limitations of generalizability applied to the work by Anderson (1996) and Coenders, Satorra, and Saris (1997), in which (1) both studies failed to incorporate robust corrections to standard errors and chi-square statistics due to unavailability of computer programs; (2) one merely used two ordinal observed variables for each latent construct in the SR model, not generally reflecting realistic applications; and (3) both left the effect of number of observed variables' categories outside the simulation design. Therefore, the proposed model design in this study first attempts to complement the related prior

research, and findings are expected to assist applied researchers in making more informed decisions while analyzing an SR model with ordinal observed indicators.

Second, although MLR is not designed specifically for ordinal data on response variables, one may assume that data are “approximately continuous” if the number of observed variables’ categories is sufficiently large. In practice, empirical researchers have congruously performed MLR in ordinal CFA or CFA-based models when the number of categories for each observed variable is more than five. Yet, unlike other robust ML estimators, MLR implemented in *Mplus* has not been systematically evaluated by means of a Monte Carlo simulation study in the literature, although its robust correction is similar but not equivalent to other robust ML estimators (e.g., MLM in *Mplus* or ML, ROBUST in EQS). The inclusion of WLSMV and ULSMV in the study also contributes to the existing literature because (1) MLR and WLSMV are very often regarded as the most common estimators in an SR model with ordinal observed indicators due to the violation of normality assumption; and (2) ULSMV has been shown to have some relative superiority over ML with robust corrections in the analysis of ordinal confirmatory factor models (Yang-Wallentin, Jöreskog, & Luo, 2010; Rhemtulla, Brosseau-Liard, & Savalei, 2012), although it has less appeared in applied research.

Comparison of WLSMV and ULSMV can shed some light on the effectiveness of the two weight matrices. More specifically, by looking into the weight matrices of ULSMV and WLSMV, it seems that using the identity matrix **I** essentially makes the parameter estimates consistent, and adding diagonal weights may possibly bring about a small improvement on parameter estimates (Muthén and Muthén, 2010). This study’s specificity in evaluating the effectiveness of diagonal weights can contribute to scholarly understanding of how the diagonal weight

matrix improves the accuracy and precision of parameter and standard error estimates. Finally, as clearly explicated in the existing literature, it is generally not recommended to use the normal theory-based maximum likelihood (ML) method when ordinal observed variables are analyzed. However, ML estimation in this study served as a baseline to explore the differences between ML and the three robust estimators. Therefore, it is worthwhile to investigate the performance of the four estimators in an SR model with ordinal data on response variables.

Third, several simulation studies have also examined the impact of the number of observed variables' categories on ML and other least squares estimators in ordinal CFA (see, e.g., Rhemtulla, Brosseau-Liard, & Savalei, 2012; Yang-Wallentin, Jöreskog, & Luo, 2010). However, what has not yet been known is the impact of the number of observed variables' categories on the overall quality of parameter estimates, especially structural regression coefficients, robust standard error estimates, and the sensitivity of adjusted chi-square statistics using MLR, ULSMV, and WLSMV in an SR model. Additionally, this study compared the behavior of the MLR, ULSMV, and WLSMV estimators under varying degrees of normality violation in an SR model, which extends the literature by the inclusion of asymmetric distributions of ordinal observed variables (Beauducel & Herzberg, 2006). MLR has been developed to permit modeling non-normal (approximately) continuous variables, whereas ULSMV and WLSMV have been implemented to deal with non-normal data because both estimators make no distributional assumption germane to the shape of observed variables in the population from which samples are drawn. When ordinal observed variables exhibit different levels of asymmetric distributions, the standard error estimates and chi-square statistics produced from these estimators are different. Without better understanding of the robustness of these estimators against non-normality, researchers are unlikely able to settle

upon an appropriate estimation method under suboptimal conditions in applications (e.g., Boomsma, 2013). The choice of estimation methods thus depends on the continuity (concerning the number of observed variables' categories) and the distribution of the ordinal observed measures.

Finally, this study was designed to examine the effect of sample size while utilizing these four estimators, because researchers have noted that a desirable sample size is known to be an important factor in SR models. Sample size is almost universally an experimental factor in a Monte Carlo simulation study (Paxton, Curran, Bollen, Kirby, & Chen, 2001). Sample size has been shown to interact with the characteristics of the data (e.g., non-normality). A small sample size may not only cause inaccurate parameter estimates and unreliable standard errors, but can also give problems of non-convergence and improper or inadmissible solutions. In addition, for a small sample size, the test statistic is likely not asymptotically chi-square distributed. Applied researchers are therefore interested in determining the smallest sample size (i.e., the sufficient sample size) at which the accuracy of parameter estimates, the stability of standard error estimates, and the robustness of chi-square statistics can be fulfilled.

The four estimators were evaluated by the quality of parameter estimates (i.e., factor loadings, inter-factor correlations, and structural regression coefficients) and standard errors, and by the performance of chi-square goodness of fit statistics, detailed further in the Outcome Variables section of this thesis. In summary, this study builds on previous simulation studies and mixed findings in pursuing the following two research questions:

1. Are any of the four estimators (ML, MLR, ULSMV, and WLSMV) consistently better or worse than the others in the estimation of model parameters, standard errors, and chi-square goodness of fit statistics across the experimental conditions investigated?
2. Are there any effects of the number of observed variables' categories, the level of asymmetric distributions of ordinal observed variables, and sample size on the performance of ML, MRL, ULSMV, and WLSMV estimates in an SR model?

CHAPTER 4

METHOD

A Monte Carlo simulation study was carried out to determine what effects of different configurations of the number of observed variables' categories, the level of asymmetric distributions of ordinal observed variables, and sample size have on parameter estimates, standard errors, and chi-square goodness of fit statistics in a five-factor structural regression model with ordinal observed variables.

Model Specification

A five-factor structural regression model (SRM) with ordinal data on response variables is depicted in Figure 1. A five-factor structural regression model with each factor having 4 ordinal observed variables was examined as the representative of the "medium-sized" SEM model specification frequently encountered in applications. To ensure representativeness of the model design from an applied standpoint, I conducted another review of 29 empirical studies using structural equation modeling from journals published by the American Psychological Association, the APA Educational Publishing Foundation, and the Canadian Psychological Association (through the PsycARTICLES database) during 2013, and 7 empirical studies that appeared in *Structural Equation Modeling* since 1994. In terms of the size of model being tested, the median number of latent factors across 36 studies was 5 (with 38% of the models tested), and the median number of total observed variables was 18 (with 15 and 24 representing the 25th and 75th percentiles, respectively).

It is critical to choose a number of observed indicators per factor that is not too small (e.g., 2 indicators per factor; see Coenders, Satorra, & Saris, 1997; Ethington, 1987), yet remains

practical in the context of a simulation study. In structural equation modeling applications, the number of indicators per factor typically falls within the range of 2 to 5 (Ding, Velicer, & Harlow, 1995), and five or more indicators per factor have rarely appeared in the literature (Gerbing & Anderson, 1985). I chose 4 ordinal observed indicators per factor, resulting in 20 ordinal observed variables in total, which represents a reasonable number of observed variables in the reviews of both Monte Carlo simulation studies and the applied literature, but apparently this number is smaller than some impressive studies (more than 40 ordinal observed indicators in total, e.g., see Beauducel & Herzberg, 2006; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009). Prior research has shown that the performance of parameter estimates and standard errors improves with increasing the number of observed variables per factor conditional on a set of good quality indicators (Forero & Maydeu-Olivares, 2009; Forero & Maydeu-Olivares, & Gallardo-Pujol, 2009; Gagné and Hancock, 2006; Gerbing & Anderson, 1985; Velicer & Fava, 1998). Marsh, Hau, Balla, and Grayson (1998) noted that the maximum accuracy of parameter estimates appeared to be reached when the number of observed variables per factor was 4, and trivially improved as the number of observed variables for each factor increased.

Four different estimation procedures that are given by ML, MLR, ULSMV, and WLSMV in *Mplus* were used. For the first and second estimation procedures, each factor was measured by four ordinal observed indicators that were treated as if they were continuous variables. The parameter estimates, standard errors, and chi-square goodness of fit statistic were obtained using ML and MLR. Since the analyzed ordinal observed indicators were assumed to be approximately continuous in this case, data analysis for the ML and MLR estimators was based on a sample-based covariance matrix. Regarding the third and fourth estimation procedures, each ordinal observed indicator was instead determined by its continuous, normal, latent

response distribution. The asymptotic covariance matrix of the polychoric correlation and threshold estimates was used for data analysis in ULSMV and WLSMV estimators to obtain the parameter estimates, standard errors, and chi-square goodness of fit statistic. The analyzed ordinal observed indicators in this case were specified as categorical variables in *Mplus*.

Simulation Design

For the sake of simplicity, homogeneous factor loadings are commonly used in simulation studies (see, e.g., Anderson, 1996; Flora & Curran, 2004; Forero & Maydeu-Olivares, 2009), which may not be representative of real-world conditions. In this study, four factor loadings (Λ_{y*} and Λ_{x*}) were held at .8, .7, .6, and .5, with corresponding residual variances (Θ_ϵ and Θ_δ) automatically set to .36, .51, .64, and .75 under a standardized solution (according to equations (4.3) and (4.4)) in the population model across all exogenous and endogenous latent variables. The common standardized factor loadings range from .4 to .9 in research practice and simulation studies (Bandalos, 2006; Ethington, 1987; Hoogland & Boomsma, 1998; Paxton, Curran, Bollen, Kirby, & Chen, 2001). The variance-covariance matrix of two exogenous latent variables (Φ) consists of two components: (1) the one inter-factor correlation was set to .3 in the population, reflecting a reasonable and empirical inter-factor correlation value in the reviews of both simulation studies (66% using .3) and applied literature (about 50% between .2 and .4); and (2) the two exogenous factor variances were set equal to 1. The two matrices of structural regression coefficients \mathbf{B} and $\mathbf{\Gamma}$ were each set up as

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ .3 & 0 & 0 \\ .2 & .5 & 0 \end{bmatrix} \text{ and } \mathbf{\Gamma} = \begin{bmatrix} .4 & .6 \\ .4 & .2 \\ .1 & .1 \end{bmatrix}.$$

The residual variances of the three endogenous latent variables (Ψ) were designated at .336, .436, and .379, based on the computation of the equation (4.4), in order to obtain standardized structural regression coefficients. The common standardized solutions ranged from .1 to .7 for structural regression coefficients, and from .2 to .8 for residual variances (i.e., $1-R^2$) in practice and simulation studies. Structural regression coefficients below .1 were, in general, not statistically and practically significant in applied research (Bandalos, 2006; Ethington, 1987; Hoogland & Boomsma, 1998; Paxton, Curran, Bollen, Kirby, & Chen, 2001). Note that the structural model is saturated (i.e., no unspecified relationships among the exogenous and endogenous latent variables).

Number of Observed Variables' Categories

Of the 157 psychometric measures in the SEM applications search that I conducted, the greatest percentage of response category was five (39.4%), followed by seven (29.9%), four (10.2%), and six (8.3%). Odd-numbered Likert scales with the middle response category seem to occur more frequently in empirical studies. Prior simulation studies in SR models with ordinal observed indicators did not fully examine the effect of number of observed variables' categories (e.g., Anderson, 1996; Coenders, Satorra, & Saris, 1997; Ethington, 1987). However, MLR has been congruously considered "appropriate" in the majority of published studies when ordinal observed variables have more than five response categories without piling or flooring effects. The chief goal here is to examine whether this general recommendation is empirically valid in an SR model. In order to explore the impact of categorization, four, five, six, and seven categories were generated for each ordinal observed indicator within different levels of ordinal observed distributions; details are in the next

section.

Ordinal Observed Distributions

Micceri (1989) found that non-normality in the form of asymmetry for psychometric distributions (due to categorization) was very usual in applied studies. Only about 3% of the 125 distributions he examined were close to normal and near symmetric, and over 80% exhibited at least slight or moderate asymmetry. Micceri attempted to provide an empirical base from which a simulation study could be closely related to the real-world data. Therefore, four ordinal observed distributions that vary in symmetry and response style were manipulated in this thesis: (1) a symmetric distribution, (2) a slightly asymmetric distribution, (3) a moderately asymmetric distribution, and (4) a bipolarized distribution. When responding to Likert-type items in the educational, behavioral, and social sciences, respondents vary in their endorsement and exhibit different response styles. Distribution (1) can be considered as middle-category response style (reference pattern), Distributions (2) and (3) as acquiescence response style (disacquiescence if going toward the opposite direction), and Distribution (4) as extreme response style (Weijters, Geuens, & Schillewaert, 2010). For a symmetric distribution, the middle categories had the highest probabilities; for slightly and moderately asymmetric distributions, the probabilities increased from low to high categories to different degrees; and for a bipolarized distribution, the higher probabilities were placed on the both end-points.

For the sake of simplification, a standard normal distribution was selected for each latent response variable in the data generation (i.e., with zero mean and variance at one), which led to a zero mean structure. Random draws of the vector \underline{y}^* and \underline{x}^* were made from a multivariate normal distribution with a zero mean vector (i.e., $\underline{\mu} = \underline{0}$) and a correlation matrix Σ^* (see

equations (4.1) and (4.2)). The multivariate normally distributed data were first generated, then ordinally scaled using prior thresholds to induce the desired asymmetric distributions and response probabilities along the standard normal distributions (Muthén & Muthén, 2010). Sixteen sets of thresholds (z-scores) were used to categorize the continuous response distributions into ordinal observed data. That is, the response probability for each category is the area under the standard normal density function between a pair of thresholds through integral calculus. In order to limit the complexity of the simulation, the underlying normal distribution was not manipulated in the study, because it requires an additional factor with several distributions of interest that would multiply the number of experimental design conditions beyond practical manageability. More importantly, the polychoric correlation estimates have been proved robust against violation of the latent normality assumption (Coenders, Statorra, & Saris, 1996; Flora & Curran, 2004; Micceri, 1989; Quiroga, 1992).

Response probabilities of ordinal observed indicators used in the study are displayed in Figure 2. Note that 1(a) to 1(d) represent a symmetric distribution with zero skewness and kurtosis from $-.49$ to $-.48$; 2(a) to 2(d) represent a slightly asymmetric distribution with skewness from $-.92$ to $-.91$ and kurtosis from $.80$ to $.84$; 3(a) to 3(d) represent a moderately asymmetric distribution with skewness from -1.39 to -1.38 and kurtosis from 1.14 to 1.19 ; and 4(a) to 4(d) represent a bipolarized distribution with skewness from $-.32$ to $-.31$ and kurtosis from -1.58 to -1.57 .

In the symmetry condition, the threshold values were $[-1.282, 0, 1.282]$ for four categories with 10%, 40%, 40%, and 10% falling into each category; $[-1.282, -.524, .524, 1.282]$ for five categories with 10%, 20%, 40%, 20%, and 10%; $[-1.645, -.806, 0, .806,$

1.645] for six categories with 5%, 16%, 29%, 29%, 16%, and 5%; and $[-1.645, -.954, -.385, .385, .954, 1.645]$ for seven categories with 5%, 12%, 18%, 30%, 18%, 12%, and 5%. In the slight asymmetry condition, the threshold values were $[-1.645, -1.08, .412]$ for four categories with 5%, 9%, 52%, and 34% falling into each category; $[-1.751, -1.341, -.524, .706]$ for five categories with 4%, 5%, 21%, 46%, and 24%; $[-1.751, -1.341, -1.08, 0, .878]$ for six categories with 4%, 5%, 5%, 36%, 31%, and 19%; and $[-1.751, -1.341, -1.036, -.613, .496, 1.341]$ for seven categories with 4%, 5%, 6%, 12%, 42%, 22%, and 9%. In the moderate asymmetry condition, the threshold values were $[-1.645, -1.08, -.253]$ for four categories with 5%, 9%, 26%, and 60% falling into each category; $[-1.751, -1.282, -.842, .05]$ for five categories with 4%, 6%, 10%, 32%, and 48%; $[-1.751, -1.341, -1.036, -.674, .202]$ for six categories with 4%, 5%, 6%, 10%, 33%, and 42%; and $[-1.751, -1.341, -1.126, -.878, -.553, .279]$ for seven categories with 4%, 4%, 5%, 6%, 10%, 32%, and 39%. In the bipolarization condition, the threshold values were $[-.524, -.253, .253]$ for four categories with 30%, 10%, 20%, and 40% falling into each category; $[-.583, -.332, -.151, .332]$ for five categories with 28%, 9%, 7%, 19%, and 37%; $[-.674, -.385, -.253, 0, .385]$ for six categories with 25%, 10%, 5%, 10%, 15%, and 35%; and $[-.842, -.468, -.305, -.176, .100, .524]$ for seven categories with 20%, 12%, 6%, 5%, 11%, 16%, and 30%. A check on the generated data sets was made to ensure that the response probabilities of observed variables approximated the four pre-specified targets (i.e., symmetry, slight and moderate asymmetry, and bipolarization).

Sample Size

Sample size in SEM has been shown to interact with the size of model complexity (e.g., number of observed variables). The guideline for an adequate sample size in a CFA model is

commonly a function of the number of observed variables. For example, Jöreskog and Sörbom (1986; 1996) recommended a sample size requirement of $1.5p(p+1)$, where p is the number of observed variables. The five-factor structural regression model with 20 ordinal observed indicators in this study requires a minimum sample size of 630. Additionally, if a sample size is too small, polychoric correlation estimates may be unstable. Several reviews of published applications of SEM and CFA have appeared. Breckler (1990) reviewed 72 studies in both CFA and SEM between 1977 and 1987, and reported that the median sample size was 198. Only 25% of the models were tested on samples of more than 200. Medsker, Williams, and Holohan (1994) identified 28 studies in both CFA and SEM between 1988 and 1993, and reported that the average sample size was 299. DiStefano and Hess (2005) reviewed 101 studies in CFA from 1990 to 2002, and reported that the median sample size was 377, and about 19% of the models were tested on samples of less than 200. Jackson, Gillaspay, and Purc-Olivares (2009) systematically reviewed 194 studies in CFA from 1998 to 2006. They reported that the median sample size was 389, and about 20% of the models were tested on samples of less than 200. The SEM applications search that I conducted showed that the sample size ranged from 110 to 2,512, with a mean of 518, across 36 studies. The median sample size was 341, with the 25th and 75th percentiles of 245 and 603 respectively. About 14% of the models were tested on samples of less than 200. Overall, there seems to be a strong consensus to increase sample size for SEM and CFA models over the past 35 years.

Seven different sample sizes commonly encountered in empirical investigations were employed in this study: $N = 200, 300, 400, 500, 750, 1,000, \text{ and } 1,500$ (see, e.g., Beauducel & Herzberg, 2006; Flora & Curran, 2004; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009). In the case of a five-factor SR model with 20 ordinal indicators, a sample size $N = 200$ and 300 is

considered as small, a sample size between 400 and 750 as medium, and a sample size $N = 1,000$ or $1,500$ as large. The corresponding ratio of N and p (N/p) are 10, 15, 20, 25, 37.5, 50, and 75, which reaches to the minimum recommendation of having sample size N at least 10 times the number of variables p (Nunnally, 1978; DiStefano & Hess, 2005) and covers a wide range of N/P values (about 94%, from 7.41 to 49.27 that were observed in the aforementioned literature review). The values selected for this simulation study were intended to provide information across a broad array of sample sizes, and the N/P values reflected what has appeared in applied work.

Data Generation and Analysis

There are 4 (ordinal observed distributions) \times 4 (number of observed variables' categories) \times 7 (sample size) = 112 experimental conditions in the study. A random seed was set up across experimental conditions for the random draws on the process of data generation. The advantages of this decision are that the data can be regenerated, and the results can be reproduced by other researchers. Five hundred data sets were generated per experimental condition, yielding a total of 56,000 data sets. The choice of 500 replications was made with consideration to sampling variance reduction, adequate power, and practical manageability (Muthén, 2002). Note that this study did not consider the possible effects of missing data on the performance of the four estimation methods but only focused on complete case analysis. Model parameters, standard errors, and the chi-square goodness of fit statistic were estimated for each replication using ML, MLR, ULSMV, and WLSMV. Data generation and analysis were performed with *Mplus* 7 (Muthén & Muthén, 2010), unless explicitly noted otherwise. Appendix D includes *Mplus* code for data generation and analysis.

Outcome Variables

Seven outcomes were empirically studied in the study: (1) average relative bias of parameter estimates, (2) average mean squared error of parameter estimates, (3) average relative bias of standard error estimates, (4) average mean squared error of standard error estimates, (5) relative bias of chi-square goodness of fit statistics, (6) the model rejection rate associated with the chi-square goodness of fit statistic at an alpha level of .05, and (7) the model rejection rate judging by the 90% confidence interval for the RMSEA.

The difference between the estimated and the true values of a parameter (i.e., the bias) was used to evaluate the performance of the four different estimators. Since bias is highly dependent on the magnitude of the true parameter value, and a great number of parameter estimates were involved in each experiment being planned, the relative bias (RB) over the replications and average relative bias (RB_A) across the total number of parameter estimates were calculated, in tandem, by

$$RB(\hat{\theta}_i) = \frac{1}{n_r} \sum_j \left[\frac{\hat{\theta}_{ij} - \theta_i}{\theta_i} \right] \times 100\%, \quad i = 1, 2, \dots, n_p; \quad j = 1, 2, \dots, n_r, \quad (31)$$

and

$$RB_A(\hat{\theta}) = \frac{1}{n_p} \sum_i RB(\hat{\theta}_i), \quad (32)$$

where $RB(\hat{\theta}_i)$ denotes the relative bias of the parameter estimate $\hat{\theta}_i$ over the replications, $\hat{\theta}_{ij}$ is the parameter estimate of the i^{th} population parameter estimate θ_i in the j^{th} replication, n_r

is the number of replications in each experimental condition, and n_p is the total number of parameter estimates. The formulae can be applied to model parameter estimates of interest, such as factor loadings (λ), inter-factor correlations (ϕ), and structural regression coefficients (β or γ). A RB_A value less than 5% can be interpreted as a trivial bias, between 5% and 10% as a moderate bias, and greater than 10% as a substantial bias (Curran, West, & Finch, 1996). Note that RB_A should be interpreted with caution, since it is used to describe an “overall” picture of average bias, i.e., lumping bias in a positive and negative direction together.

To quantify the overall quality of parameter estimates, the mean squared error is commonly used in simulation studies because it accounts for both the amount of bias and the sampling variability of parameter estimates (i.e., efficiency). The mean squared error (MSE) and average mean squared error (MSE_A) can be defined as

$$MSE(\hat{\theta}_i) = \frac{1}{n_r} \sum_j \left[\frac{\hat{\theta}_{ij} - \theta_i}{\theta_i} \right]^2, \quad (33)$$

and

$$MSE_A(\hat{\theta}) = \frac{1}{n_p} \sum_i MSE(\hat{\theta}_i), \quad (34)$$

where $MSE(\hat{\theta}_i)$ denotes the mean squared error of the parameter estimate $\hat{\theta}_i$ over the replications; and other notations have been defined. A small MSE_A value is suggested as favorable because it indicates better overall quality of parameter estimates, that is, less biased and more precise.

To obtain accurate and precise standard error estimates is also a primary concern in applied and simulation studies. In a similar way, the bias formulations can be used for standard error estimates, relative to the standard deviation of the parameter estimates over the replications (also referred to as the empirical standard error). That is, the standard deviation of the parameter estimates over the replications is used as a proxy for the population standard error. The RB and RB_A for standard error estimates are formulated as

$$RB[\widehat{SE}(\hat{\theta}_i)] = \frac{1}{n_r} \sum_j \left[\frac{\widehat{SE}(\hat{\theta}_i)_j - SD(\hat{\theta}_i)}{SD(\hat{\theta}_i)} \right] \times 100\%, \quad (35)$$

and

$$RB_A[\widehat{SE}(\hat{\theta})] = \frac{1}{n_p} \sum_i RB[\widehat{SE}(\hat{\theta}_i)], \quad (36)$$

where $\widehat{SE}(\hat{\theta}_i)_j$ is the estimated standard error of parameter $\hat{\theta}_i$ in the j^{th} replication, and $SD(\hat{\theta}_i)$ is the standard deviation of parameter $\hat{\theta}_i$ over the replications. The mean squared error (MSE) and average mean squared error (MSE_A) can also be defined as

$$MSE[\widehat{SE}(\hat{\theta}_i)] = \frac{1}{n_r} \sum_j \left[\frac{\widehat{SE}(\hat{\theta}_i)_j - SD(\hat{\theta}_i)}{SD(\hat{\theta}_i)} \right]^2, \quad (37)$$

and

$$\text{MSE}_A[\widehat{\text{SE}}(\widehat{\theta})] = \frac{1}{n_p} \sum_i \text{MSE}[\widehat{\text{SE}}(\widehat{\theta}_i)], \quad (38)$$

where $\text{MSE}[\widehat{\text{SE}}(\widehat{\theta}_i)]$ denotes the mean squared error of the estimated standard error of parameter estimate $\widehat{\theta}_i$ over the replications.

Likewise, the performance of chi-square statistics can be assessed by the relative bias. Because of the expected value of a chi-square distribution equal to its degrees of freedom, the relative bias of chi-square statistics over the replications can be expressed as

$$\text{RB}(\widehat{\chi^2}_j) = \left[\frac{\widehat{\chi^2}_j - \text{df}}{\text{df}} \right] \times 100\%, \quad (39)$$

and

$$\text{RB}(\widehat{\chi^2}) = \frac{\sum_j \text{RB}(\widehat{\chi^2}_j)}{n_r}, \quad j = 1, 2, \dots, n_r, \quad (40)$$

where $\widehat{\chi^2}_j$ is the estimate of the chi-square statistic in the j^{th} replication, df is the model degrees of freedom, and n_r is the number of replications in each experimental condition.

Alternatively, chi-square test statistics have been examined often through the calculation of the rejection rate at a given nominal alpha level of .05 in simulation studies. The rejection rate equals the number of replications for which the chi-square value is greater than the critical value divided by the number of replications (successfully analyzed). The rejection rate of the proposed model should, therefore, approximate 5% specified in the population

model. The obtained rejection rates lying between .025 and .075 can be considered acceptable at a nominal alpha level of .05 (Bradley, 1978). A high rate of rejection suggests an inflated Type I error rate of testing overall model fit, reflecting that chi-square tests may have been over-rejected; a low rate of rejection otherwise indicates that chi-square test statistics may have been underestimated. Moreover, a high rate of rejection implies increased likelihood against the null hypothesis, whereas a low rate of rejection may indicate a potential compromise of the power of rejecting the hypothesized model.

Finally, applications of ad hoc fit indices have been less common in the extant literature (e.g., Flora & Curran, 2004; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Rhemtulla, Brosseau-Liard, & Savalei, 2012; Yang-Wallentin & Jöreskog, & Luo, 2010). However, the root mean square error of approximation (RMSEA) has received the most attention, and it recently has been recognized as one of the most informative and trustworthy indices of model fit in applied research. RMSEA is a function of the sample estimate of the noncentrality parameter, $\hat{\lambda}$:

$$RMSEA = \sqrt{\max \left(0, \frac{\hat{\lambda}}{(N-1) \times d} \right)}, \quad (41)$$

where $\hat{\lambda} = T - d$, T is the estimated chi-square test statistic, and d is the degrees of freedom. One can replace T as T_{ML} , T_{MLR} , T_{ULSMV} , or $T_{D-WLSMV}$ in the equation (41) to obtain $RMSEA_{ML}$, $RMSEA_{MLR}$, $RMSEA_{ULSMV}$, or $RMSEA_{D-WLSMV}$. RMSEA not only takes into account model complexity, as reflected in the degrees of freedom, but also it is least sensitive to sample size among ad hoc fit indices. It has been suggested that an RMSEA value of less than or equal to .05 is

indicative of a model of close fit (Browne & Cudeck, 1993). Because RMSEA, unlike chi-square statistics, does not have a known sampling distribution to assess its behavior, the performance of RMSEA was therefore assessed by the calculation of the rejection rate, judging by the 90% confidence interval. The upper and lower bounds of a 90% confidence interval for the RMSEA can be calculated as (Browne & Cudeck, 1993):

$$\text{RMSEA}_{\text{low}} = \sqrt{\max \left(0, \frac{\hat{\lambda}_{.05}}{(N-1) \times d} \right)}, \quad (42)$$

and

$$\text{RMSEA}_{\text{upp}} = \sqrt{\max \left(0, \frac{\hat{\lambda}_{.95}}{(N-1) \times d} \right)}. \quad (43)$$

$\hat{\lambda}_{.05}$ is the value that T is the 95th percentile of the noncentral chi-square distribution $\chi^2(d, \hat{\lambda}_{.05})$, and $\hat{\lambda}_{.95}$ is the value that T is the 5th percentile of the noncentral chi-square distribution $\chi^2(d, \hat{\lambda}_{.95})$. Likewise, a 90% confidence interval for the RMSEA for ML, MLR, ULSMV, or D-WLSMV can be obtained by replacing T as T_{ML} , T_{MLR} , T_{ULSMV} , or $T_{\text{D-WLSMV}}$ in the noncentral chi-square distribution. The rejection rate is determined as the number of replications for which the lower bound of a 90% confidence interval for the RMSEA is greater than the “practical” guideline of cutoff value of .05 divided by the number of replications (successfully analyzed). Also, means of RMSEA over the replications are reported to illustrate the practical relevance of the findings.

CHAPTER 5

RESULTS

Due to an overwhelming amount available output, this results section needs reduction to accomplish a concise and attractive, though informative, presentation. The tables and figures are collapsed across several conditions, and the results for certain conditions not presented here are available in Appendix E. More specifically, results for sample sizes of $N = 400$, 750 , and $1,500$ are not presented here but are appended as supplemental materials in Appendix E, mainly because a similar pattern of the results for $N = 400$ and $N = 750$ was observed to those for $N = 500$; and results with $N = 1,500$ were comparable to $N = 1,000$, in terms of the performance of the parameter, standard error estimates, test statistics, and RMSEA. Furthermore, an exhaustive report of bipolarized data was not undertaken here, as the effect of bipolarization on model results fell between that of slight asymmetry and moderate asymmetry across most conditions; the effect of bipolarization on chi-square goodness of fit statistics fell between that of symmetry and slightly asymmetry across many conditions. However, Appendix F contains all results from the bipolarized data conditions. Because ML and MLR produced the same rates of non-convergence and inadmissible solutions, and the same values of parameter estimates (including both factor loadings and structural coefficients), these results were combined within the estimator denoted by "ML/MLR" in some tables. However, uncorrected standard errors and unadjusted chi-square goodness of fit statistics obtained with ML were different from MLR, so they were reported separately in the pertinent result tables.

Non-Convergence and Inadmissible Solutions

Non-convergence was defined as the iterative estimation process that failed to converge

because the maximum number of iterations (by *Mplus* default) exceeded or because there were difficulties in optimizing the fit function before the maximum number of iterations had been reached (Muthén & Muthén, 2010). An inadmissible solution (i.e., Heywood cases) was defined as a statistically converged solution that, however, produced unbounded parameter estimates (i.e., an estimated inter-factor correlation larger than 1 in absolute value) or negative residual variances.

Tables 4(a) and 4(b) show the number of cases that failed to converge or produced inadmissible solutions. Note that ML and MLR produced the same number of cases of non-convergence and inadmissible solutions, so results were combined within the estimator denoted by “ML/MLR” in Tables 4(a) and 4(b). As shown in Table 4(a), estimation that failed to converge most likely occurred with 4-category, moderately asymmetric data in the smallest sample size $N = 200$ for all four estimators (in boldface). Convergence failures disappeared for all four estimators when sample size increased to $N = 300$, except for 2 cells (in boldface). Regarding inadmissible solutions in Table 4(b), ML and MLR did not yield any inadmissible solution across all conditions in the study. WLSMV and ULSMV tended to produce inadmissible solutions particularly when sample size was small. Among the four estimators, ULSMV had a higher probability of producing inadmissible solutions than the other three estimators across many conditions with sample size $N = 200$. The highest rate of inadmissible solutions obtained with ULSMV was 1.2% (6 cases), and it appeared with four-category, moderately asymmetric data, and sample size $N = 200$. However, there were no inadmissible solution across all but three conditions (in boldface) as sample size increased to $N = 300$ or more. In general, the four estimators (ML, MLR, WLSMV, and ULSMV) all resulted in convergence failures when data were 4-category, moderately asymmetric in the smallest sample size $N = 200$. ULSMV and WLSMV

were more likely subjected to inadmissible solutions across many conditions with sample sizes $N = 200$ or 300 .

In order to inform research practice and maximize external validity, the replications that were classified as non-convergence or inadmissible solutions were considered invalid empirical observations and were excluded from studying the impact of experimental factors on the performance of the four estimators and evaluating the parameter and standard error estimates, test statistics, and RMSEA (cf. Boomsma, 2013; Chen, Bollen, Paxton, Curran, & Kirby, 2001; Flora & Curran, 2004; Forero & Maydeu-Olivares, 2009). Note that additional analyses that included the inadmissible solutions were conducted. The analyses indeed brought about minor changes in outcome variables, but the conclusions remained unchanged.

Parameter Estimates

Factor Loadings

Tables 5–8 display average relative bias (RB_A) and average mean squared error (MSE_A) of factor loadings and structural coefficients by number of observed variables' categories and ordinal observed distributions for all four estimators. Note that ML and MLR produced the same parameter estimates (both factor loadings and structural coefficients), so results were combined within the estimator denoted by "ML/MLR" in Tables 5–8. Factor loadings were, on average, underestimated by ML and MLR. They were moderately or substantially downward-biased across all sample size conditions, except for symmetric data with 5 categories or more. The magnitude of this negative bias was reduced with increasing the number of observed variables' categories but increased with increasing the level of asymmetric distributions of ordinal observed variables. Conversely, factor loading estimates

obtained by WLSMV and ULSMV appeared to be negligibly unbiased on average, irrespective of the number of observed variables' categories, the shape of ordinal observed distributions, and sample size. Overall, WLSMV and ULSMV were consistently superior to ML and MLR for factor loading estimation in all investigated conditions, indicating that WLSMV and ULSMV yield more accurate factor loading estimates than ML and MLR.

In order to quantify the overall quality of parameter estimates, both the amount of bias and the sampling variability of parameter estimates (i.e., efficiency) should be considered simultaneously. An index that combines both squared bias and sampling variance is the mean squared error (MSE). A small MSE value is suggested as favorable because it indicates better overall quality of parameter estimates, that is, less biased and more precise. Regarding the overall quality of estimated factor loadings, the average mean squared error (MSE_A) decreased with increasing sample sizes and the number of observed variables' categories but increased with a greater level of asymmetric distributions. That is, the performance of factor loading estimates became better when sample size and the number of observed variables' categories increased but turned worse when the level of asymmetric distributions increased. MSE_A was most pronounced in the conditions where RB_A was appreciable; in particular, it was noticeably large with four-category ordinal data. In general, MSE_A obtained with WLSMV and ULSMV were smaller than ML and MLR across nearly all conditions.

However, there were few cells where MSE_A obtained with ML and MLR was smaller than WLSMV and ULSMV when data were symmetric. In order to get a deeper understanding of this scenario, the MSE_A was then partitioned into two components: squared bias and sampling variance in a stacked histogram. The lower portion in the stacked histogram is the squared bias,

whereas the upper portion represents the sampling variance. Figure 3 clarifies that ML and MLR displayed higher bias for categories 5, 6, and 7 than WLSMV and ULSMV in the combination of symmetric data and $N = 200$, despite generally lower MSE_A . Specifically, ML and MLR produced more biased, but less variable, factor loading estimates, indicating that the estimates obtained in any given replications are likely to be close to each other but too far from the population value. Such observation disappeared as sample size increased, reflecting that a large sample size can wash out the advantage of symmetric data in ML and MLR estimation.

It is of particular interest that MSE_A obtained with WLSMV was consistently slightly smaller than ULSMV across all cells, suggesting that the diagonal weights indeed contribute a small improvement on the overall quality of factor loading estimates. That is, factor loading estimates obtained with WLSMV were less biased and more precise than those produced by ULSMV. Uniformly, WLSMV and ULSMV are considered better than ML and MLR on the performance of factor loading estimates across nearly all conditions. However, the performance of ULSMV fell between that of WLSMV and ML/MLR.

Structural Coefficients

The overall bias in structural coefficients (including both structural regression coefficients and the inter-factor correlation) obtained with the four estimators was, on average, trivially biased (either positively or negatively). Averaging over the structural coefficient estimates, the bias obtained with WLSMV and ULSMV appeared to be consistently trivial, rarely leading to the amount of bias greater than 1%. ML and MLR however introduced the amount of slightly marked bias into the estimates of structural coefficients with moderately asymmetric data (about -3%) across all sample sizes. In comparison, there was no remarkable distinction

among the four estimators, in terms of the absolute value of RB_A . However, ML and MLR produced slightly larger bias than WLSMV and ULSMV in many conditions, except for certain symmetric data conditions.

With respect to the overall quality of estimated structural coefficients, the MSE_A decreased with increasing sample size and the number of observed variables' categories but increased with increasing the level of asymmetric distributions of ordinal observed variables. Similarly, the performance of structural coefficient estimates improved when sample size and the number of observed variables' categories increased but dropped when the level of asymmetric distributions increased. Unlike factor loading estimates, the benefit of using diagonal weights to bring about a small improvement on the overall quality of structural coefficient estimates did not accrue until a medium sample size ($N = 500$) was reached. In general, there was no remarkable evidence suggesting that one of the four estimators is inferior to another one, in terms of MSE_A . However, ML and MLR produced smaller MSE_A than WLSMV and ULSMV in all conditions of symmetric data, while WLSMV and ULSMV produced smaller MSE_A in nearly all asymmetric data conditions.

In addition to overall bias and quality of estimated structural coefficients, examination of each structural coefficient was also employed to gain further insight into which type(s) of structural coefficients performed better. In terms of MSE_A , the performance of estimated structural coefficients became better as the magnitude of coefficients increased for all the four estimators. The estimates of the inter-factor correlation and structural regression coefficients in Γ generally were less biased and more precise than those of structural regression coefficients in B , provided with the same magnitude of coefficient. Table 9 shows that MSE_A of

ϕ_{12} was about 4 to 5 times smaller than MSE_A of β_{21} , and MSE_A of γ_{32} was consistently smaller than MSE_A of β_{21} across the four estimators in the $N = 1,000$ conditions. Again, it was observed that ML and MLR produced smaller MSE_A than WLSMV and ULSMV in all the conditions of symmetric data, while WLSMV and ULSMV produced smaller MSE_A in all asymmetric data conditions.

Standard Error Estimates

The RB_A and MSE_A for standard errors of factor loadings and structural coefficients are presented in Tables 10–13. Standard errors exhibited, on average, moderately downward bias for both WLSMV and ULSMV with the smallest sample size ($N = 200$), reflecting that robust standard errors were not upward adjusted enough to compensate the loss of efficiency caused by WLSMV and ULSMV estimation in the sample size $N = 200$ conditions. Not surprisingly, standard error underestimation improved when sample size increased. In contrast, the amount of negatively moderate-to-substantial bias was observed in ML estimation across most simulation conditions, except for all cells of symmetric data. The amount of trivial bias (essentially unbiased) was produced by MLR across most conditions. That is, this moderate-to-substantial underestimation of ML standard errors was significantly attenuated when robust corrections to standard errors in MLR estimation were employed. As soon as the sample size reached to $N = 500$ or more, the three robust estimators performed comparably well for estimating standard errors of parameter estimates, in terms of RB_A . Uncorrected standard errors produced by ML still remained moderately-to-substantially biased in all asymmetric data conditions. Overall, the performance of MLR surpassed that of ML, WLSMV, and ULSMV across most conditions, in terms of RB_A . The performance of ML was the worst in all asymmetric data conditions. However, there was no remarkable distinction between ML and

MLR in the conditions with symmetric data and sample size $N = 300$ or more.

Much as with the overall quality of parameter estimates, MSE_A associated with standard error estimates decreased with increasing sample size but increased with increasing the level of asymmetric distributions of ordinal observed variables. MSE_A obtained with ML and MLR demonstrated little sensitivity to the number of observed variables' categories, whereas in estimating structural coefficients, MSE_A obtained with WLSMV and ULSMV diminished as the number of observed variables' categories increased. The advantage of incorporating diagonal weights into the estimated asymptotic covariance matrix of the parameter estimates was only sustained with robust standard errors of structural coefficient estimates, not with that of factor loading estimates.

In general, ULSMV produced more precise estimated standard errors of factor loadings than WLSMV and MLR in all conditions, whereas WLSMV produced more precise estimated standard errors of structural coefficients than ULSMV and MLR across all asymmetric data conditions. The MSE_A was partitioned into two components: squared bias and sampling variance in a stacked histogram, as described in the preceding section. As depicted in Figure 4, due to lower sampling variance, ULSMV displayed the lowest MSE_A in the combination of slightly asymmetric data and $N = 300$, despite slightly higher bias. As sample size increased to $N = 1,000$, the bias produced by ULSMV was essentially equal to the other two robust estimators (see Figure 5), and ULSMV still had the lowest MSE_A among the 4 estimators. Because of the amount of moderate bias, ML exhibited the highest MSE_A among the 4 estimators, although some lower sampling variances were observed. In Figure 6, WLSMV displayed lowest MSE_A due to lower sampling variance, despite trivial bias. Although MLR

produced less biased standard error estimates, it had relatively high sampling variance, illustrating that the standard error estimates obtained in any given replications are usually not far from the empirical standard error (i.e., the standard deviation of parameter estimates over replications), but are widely spread out. In addition, ML produced higher biased standard errors but smaller sampling variances than MLR.

Chi-Square Goodness of Fit Statistics

Tables 14–17 present findings for chi-square goodness of fit statistics and RMSEA with ML, MLR, ULSMV, and WLSMV estimators: (1) relative bias of chi-square goodness of fit statistics, (2) rejection rates associated with the Likelihood Ratio (LR) Test, (3) mean of RMSEA, and (4) rejection rates associated with the 90% CI for the RMSEA. The rejection rates associated with the LR test equal the number of replications for which the chi-square value is greater than the critical value divided by the number of successfully analyzed replications, and the rejection rates associated with the 90% CI for the RMSEA are determined as the number of replications for which the lower bound of a CI is greater than the practical cutoff value of .05 divided by the number of successfully analyzed replications. The boldface numbers in these tables indicate unacceptable rejection rates, implying that acceptable difference rates in the tables are within the range [2.5%, 7.5%] (Bradley, 1978).

For WLSMV and ULSMV, the empirical Type I error rates of testing overall model fit were almost all within the range of .025 and .075, very close to the nominal Type I error ($\alpha = .05$), except for the smallest sample size ($N = 200$). When the sample size was too small, WLSMV seemed to reject the hypothesized model too frequently, whereas the corresponding chi-square statistics of ULSMV were too conservative, as evidenced by $N = 200$. On the other

hand, MLR appeared to be systematically inferior in controlling for Type I error rates of testing overall model fit across nearly all conditions, unless a larger sample size was used (e.g., $N = 1,000$). ML performed worse than MLR across most conditions, except for some cells of symmetric data. When data were slightly or moderately asymmetric, ML seemed to reject the hypothesized model much beyond expectation (more than 10 times the nominal Type I error), indicating that uncorrected chi-square statistics may have been substantially inflated in the presence of non-normality.

Among the three robust estimators, the corrected chi-square test statistics tend to be positively biased across all experimental conditions, with MLR correction being particularly unstable. The degree of positive bias diminished as sample size increased. Also, the number of observed variables' categories and the level of asymmetric distributions of ordinal observed variables had an increasing effect on the inflation of chi-square statistics, but this effect was more pronounced for small sample sizes (e.g., $N = 200$ or 300). In general, MLR estimation was prone to yield moderately inflated chi-square statistics in the conditions of moderately asymmetric data and the small sample size (e.g., $N = 200$ or 300). Compared to WLSMV estimation, the positive bias was seen to be slightly smaller with ULSMV estimation.

Graphical comparisons of the observed distributions of the test statistics to the expected chi-square distributions are further provided to visualize some information from Tables 14 through 17 using Probability-Probability (P-P) plots. Figures 7 and 8 demonstrated extremely bad distributional behavior of T_{ML} and also evidently disclosed the effects of small sample size and asymmetric distributions on the inflation of chi-square statistics. The plots for the moderately asymmetric data with seven-category were shown for the worst scenario where N

= 200 for all four estimators. Remarkably, the overall behavior of T_{ML} is clearly deviant across most conditions unless the data were symmetric and sample size increased. Overall, T_{ULSMV} had the closest approximation to the reference chi-square distribution, followed by T_{WLSMV} , T_{MLR} , and then T_{ML} in all conditions.

RMSEA

As seen in Tables 14–17, rejection rates associated with the 90% CI for the RMSEA were not sensitive to the conditions of the study. This may be attributed to the population SR model being correctly specified in data analysis. However, means of RMSEA were minimally positively biased for all four estimators. It is not surprising that this inflation was reduced monotonically with increasing sample size, regardless of the number of observed variables' categories and the shape of ordinal observed distributions.

CHAPTER 6

DISCUSSION

This study sought to compare the performance of ML, MLR, WLSMV, and ULSMV in regard to parameter estimates, standard errors, and chi-square goodness of fit statistics in a five-factor structural regression model with ordinal observed variables under different experimental configurations of ordinal observed distributions, number of observed variables' categories, and sample size, resulting in 112 conditions. The conditions were chosen to highlight the differences among the four estimators, as well as to investigate a wide variety of empirical circumstances frequently encountered in research practice. Several general findings are discussed as follows.

First, the four estimators all results in convergence failures when data were 4-category, moderately asymmetric in the smallest sample size $N = 200$. Furthermore, ULSMV and WLSMV were more likely subject to inadmissible solutions when small sample sizes $N = 200$ or 300 were analyzed. The small sample degradation of the ML and three robust estimators is consistent with previous simulation studies in which non-convergence or inadmissible solutions more frequently occur with small sample sizes (Herzog, Boomsma, & Reinecke, 2007; Rhemtulla, Brosseau-Liard, & Savalei, 2012; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009). However, it is important to note that increasing sample size N to 300 apparently can help reduce the degree of this unfavorable outcome.

Second, this study replicated previous results that factor loadings were typically underestimated by ML and MLR but were essentially unbiased with WLSMV and ULSMV (Beauducel & Herzberg, 2006; Flora & Curran, 2004; Forero, Maydeu-Olivares, Gallardo-Pujol,

2009; Rhemtulla, Brosseau-Liard, & Savalei, 2012). The accuracy and precision of estimated factor loadings with WLSMV and ULSMV were better than that of estimated factor loadings with ML and MLR across nearly all conditions, in terms of MSE_A . Interestingly, on the basis of this simulation study, a clear superiority of WLSMV and ULSMV over ML and MLR in factor loading estimates across all simulation conditions was confirmed in this study, irrespective of the number of observed variables' categories, the shape of ordinal observed distributions, and sample size. Even when the number of observed variables' categories in the data reached to seven, ML and MLR still led to moderately biased factor loading estimates (Beauducel & Herzberg, 2006; Rhemtulla, Brosseau-Liard, & Savalei, 2012), suggesting that prior studies with ordinal observed indicators using ML and MLR underestimated associations between ordinal observed variables and latent constructs. In turn, the estimates of reliability for composite scores on Likert-type scales may have been undermined, which is particularly more appreciable with increasing the level of asymmetric distributions of ordinal observed variables.

ML and MLR led to moderately biased factor loading estimates but only produced a small amount of bias in structural coefficients across all conditions. More specifically, ML and MLR displayed mind robustness against violation of normality in estimating structural coefficients but rather factor loadings. This "unique" finding contributes to the literature by demonstrating that a combination effect of categorization and asymmetric observed distributions is larger on the measurement model parameters than on the structural model parameters. Such observation is similar to that of Coenders, Satorra, and Saris (1997), who concluded that Pearson product-moment correlations between ordinal observed indicators (through maximum likelihood estimation) perform badly in the estimation of factor loadings but such lower measurement quality estimates can lead to approximately correct point estimates of

structural coefficients.

Moreover, in terms of the accuracy and precision of estimated structural coefficients, a clear superiority of ML and MLR over WLSMV and ULSMV was found in all symmetric data conditions, whereas the advantage shifted to WLSMV and ULSMV in nearly all asymmetric data conditions. This may be attributed to the symmetric data being analyzed and the desirable estimation properties of maximum likelihood, such as unbiasedness and maximal efficiency, can therefore be retained. It is readily apparent that the accuracy and precision of parameter estimates (including factor loadings, the inter-factor correlation, and structural regression coefficients) became better with increasing sample size and the number of observed variables' categories but decreased with a greater level of asymmetric distributions of ordinal observed variables.

In addition, increasing the magnitude of population structural coefficients was associated with higher accuracy and precision of structural coefficient estimation. A finding worth noting is that given the same magnitude of structural coefficients, the inter-factor correlation and structural regression coefficients between exogenous and endogenous latent variables generally performed better than structural regression coefficients among endogenous latent variables across all conditions. Prior simulation studies rarely explored the effect of ordinal observed variables on structural coefficients, but these findings complement the existing literature by showing the performance of structural coefficient estimates in an SR model. The specification of heterogeneous structural coefficients also highlights the potential weakness in an SR model – estimating a small structural regression coefficient among endogenous latent variables is likely compromised.

Third, it was observed that MLR gave more accurate, but less precise, standard error estimates than WLSMV and ULSMV across most conditions. Despite the slightly higher amount of bias, among the three robust estimators, ULSMV produced more precise estimated standard errors of factor loadings across all conditions, whereas WLSMV produced more precise estimated standard errors of structural coefficients, due to smaller sampling variation, in all asymmetric data conditions. These findings resonate with the existing literature, in showing that the performance of standard error estimates for WLSMV and ULSMV is better than for MLR (Rhemtulla, Brosseau-Liard, & Savalei, 2012; Yang-Wallentin, Jöreskog, & Luo, 2010). In addition, ML produced moderately-to-substantially biased standard error estimates across all conditions, except for the conditions of symmetric data, congruent with previous studies (e.g., Beauducel & Herzberg, 2006; Kaplan, 2009). Likewise, the accuracy and precision of standard error estimates improved with increasing sample size and the number of observed variables' categories but was reduced with a greater level of asymmetric distributions of ordinal observed variables. However, standard error estimates obtained by ML and MLR were less sensitive to the number of observed variables' categories.

Fourth, in the evaluation of overall model fit using chi-square goodness of fit statistics, WLSMV and ULSMV had empirical rejection rates within the acceptable range of .025 and .075, closed to the nominal Type I error $\alpha = .5$. However, when the sample size was too small (e.g., $N = 200$), WLSMV was likely to over-reject the hypothesized model more often than expected, echoing the high tendency of WLSMV rejection rates (Beauducel & Herzberg, 2006; Flora & Curran, 2004). In contrast, ULSMV tended to under-reject the hypothesized model less than the alpha level with a small sample. As with previous studies, ML produced unacceptable

rejection rates and T_{ML} exhibited extremely deviant distributional behavior in all asymmetric data conditions (e.g., Kaplan, 2009; Muthén & Kaplan, 1992). Among the three robust estimators, MLR was systematically inferior to WLSMV and ULSMV in controlling for Type I error rates of testing overall model fit across many conditions, due to moderate-to-substantial inflation of chi-square goodness of fit statistics. The deviant distributional behavior of T_{MLR} occurred with the moderately asymmetric data having seven categories in the smallest sample size $N = 200$. The finding also suggests that T_{MLR} is subjected to sizeable overestimation in a small sample. Until the sample size increased to 1,000, an acceptable rejection rate associated with MLR chi-square statistics was consequently observed.

With respect to the supplemental fit index, RMSEA, rejection rates judging by 90% confidence intervals revealed less sensitivity to the correctly specified SR model in this study. Although RMSEA showed promise for assessing the adequacy of a hypothesized model, means of RMSEA were slightly positively biased for all four estimators. This is in line with Curran et al. (2002) and Herzog & Boomsma (2009), who found that RMSEA is upward biased in smaller sample size conditions. Overall, RMSEA seems to be a reliable index in the evaluation of overall model fit when the model has no specification error.

Fifth, this study also aimed to evaluate the performance of the two weight matrices (\mathbf{I} versus \mathbf{W}_D) in the estimation of parameters, robust standard errors, and test statistics. An interesting finding of this study is that the benefit of incorporating diagonal weights into the least squares fit function and estimated asymptotic covariance matrix was observed with parameter estimates and robust standard error estimates across all conditions, except for robust standard errors of factor loadings. That is, the diagonal weights contributed a small

improvement upon the performance of parameters and robust standard errors estimates. Additionally, this advantage was not sustained with chi-square statistic corrections because T_{ULSMV} appeared to have the closest approximation to the reference chi-square distribution. However, these findings make a distinct contribution to the existing literature in which the effectiveness of the diagonal weights is not very clear. In sum, not only can this diagonal weight matrix get around computational troubles in the conditions of small sample sizes or/and complex models but also yield relatively accurate parameter and standard error estimates, and a well-behaved distribution of test statistics that is approximately close to a central chi-square.

Implications for Applied Research

Sample Size

There are several specific implications of the findings with respect to fitting an SR model with ordinal observed variables using these four estimators in practice. First, applied researchers are concerned with the rates of non-convergence and inadmissible solutions. A non-converged or inadmissible solution often plagues applied researchers, and it is of no use for substantive interpretation. Evidence suggests that the four estimators all resulted in convergence failures when data were 4-category, moderately asymmetric in the smallest sample size $N = 200$, but only WLSMV and USLMV were subjected to inadmissible solutions in many conditions with sample sizes $N = 200$ or 300 . In addition, a small sample size is often problematic because parameter and standard error estimates can be biased seriously and less precise. Not surprisingly, increasing sample size not only protects against convergence failures and inadmissible solutions but also improves the performance of model estimation.

One of the relevant implications for applied researchers is the presence of conditions for

which none of the three robust estimators yields adequate results. WLSMV and ULSMV do not need a large sample size for the recovery of population parameters and to evaluate overall model fit via the mean- and variance-adjusted chi-square goodness of fit statistics, but a medium sample ($N = 500$ or more) is needed to obtain better standard error estimates. On the other hand, MLR does not require a large sample to produce stable structural coefficient and standard error estimates, but may need a quite large sample ($N = 1,000$ or more) to control for Type I error rates of testing overall model fit, despite the existence of moderate underestimation in factor loading estimates.

Taken together, a sample size less than 500 should be avoided to use when fitting a medium-size model with ordinal observed variables in practice. In this case, the ratio of sample size (N) and the number of observed variables (p) is 25 which much more exceeds the recommendation of having N at least 10 times p (Nunnally, 1978). Additionally, the ratio of N and the number of free parameters (q) is 10 which just meets the minimum requirement of having at least $N : q = 10 : 1$ with non-normal data when using maximum likelihood estimation (Bentler & Chou, 1987; Hu, Bentler, & Kano, 1992). Note that the number of free parameters in MLR estimation was 50 because there were 20 factor loadings, 20 error variances, and 10 structural coefficients.

Estimation Methods

Regarding estimation method selection, if the structural relationships are of primary concern in a research setting, the use of MLR can be recommended when fitting an SR model with ordinal observed variables on this ground. The biases of MLR estimates remained quite small and were typically less than .01 in the standardized structural coefficient metric,

although a substantial amount of bias in estimating factor loadings is inevitable (about 5% to 10%). Given this recommendation, a word of caution is warranted. In a small sample, the robust chi-square goodness of fit statistic obtained with MLR is likely compromised, and RMSEA can be regarded as another alternative to evaluate the plausibility of overall model fit. This fit index could be of particular benefit to applied researchers in evaluating overall model fit. In general, it is not advisable to use ML in an SR model with ordinal observed variables unless data are symmetric and the desired sample size $N = 500$ is reached. In this case, structural coefficient and standard error estimates are considered reliable but the factor loadings are slightly underestimated and the uncorrected chi-square statistic is still slightly inflated, and RMSEA should be used to evaluate overall model fit. Generally speaking, the moderate-to-substantial underestimation of standard errors and considerable inflation of chi-square statistics make ML less attractive and favorable in practice, particularly when data moderately deviate from normality. This study also supports the argument that the performance of ML is generally unacceptable in the presence of non-normality.

It seems that WLSMV and ULSMV compensate more effectively than MLR for the bias and model fit evaluation measures due to the observed indicators by virtue of being ordinal rather than continuous in the SR model. Furthermore, the benefit of using diagonal weights makes WLSMV superior to ULSMV in many conditions; however, in a very rare scenario, when WLSMV is subject to the non-convergent issue, ULSMV may serve as another alternative for applied researchers. It is worth noting that once applied researchers confront the problem of missing data, ML or MLR with full information estimation is considered as a promising approach to handling missing data without employing (single or multiple) data imputation. Yet, the treatment of missing data in WLSMV and ULSMV estimators remains technically

underdeveloped, providing its bivariate orientation (pairwise deletion as the default in *Mplus*, Muthén & Muthén, 2010). Additionally, some applied researchers may be limited in the choice of software programs or by estimation availability of certain software programs that they are familiar with. For instance, diagonally weighted least squares estimation is only implemented in *Mplus*, LISREL, SAS PROC CALIS, and the R package 'lavaan' but currently unavailable in EQS, Amos, and STATA.

Finally, another relevant implication for applied researchers is related to practical differences among the three robust estimators in this study. Take the model inference for example, the robust chi-square goodness of fit statistics obtained with MLR may tend to over-reject the true model about 5-20% more in the conditions with small sample sizes (e.g., $N = 200, 300$, or 500) than WLSMV and ULSMV. Specifically, applied researchers are very likely to reach completely different conclusions by rejecting the true model if they employ MLR rather than WLSMV or ULSMV in data analysis. Additionally, applied researchers occasionally use the RMSEA estimates to evaluate the model misfit, instead of the 90% confidence interval of the RMSEA. They reject the hypothesized model if the RMSEA estimate is greater than the "practical" cutoff value of .05. Another observation of possibly misleading conclusions drawn from empirical data is that a slightly higher bias of the RMSEA estimates makes MLR a little vulnerable in the evaluation of overall model fit with the smallest sample size $N = 200$. Some replications ended up with rejecting the true model based on the RMSEA estimates when MLR was employed in the analysis, and this situation became even worse when ML was used. Regarding the model parameter inference, estimating a small structural regression coefficient of .1 in the population model is very likely challenging as well, in particular of the conditions with asymmetric data and/or small sample sizes. The parameter estimates obtained with

WLSMV and ULSMV had higher rates of statistical significance than those obtained with MLR across all asymmetric data conditions, regardless of number of observed variables' categories and sample size. Namely, applied researchers have higher likelihood of detecting these small relationships (i.e., 0.1 in the standardized regression coefficient metric) between latent constructs if they employ WLSMV or ULSMV in data analysis. For instance, the statistical significant rates of WLSMV and ULSMV were about 5% higher than those of MLR in the conditions with sample size $N = 1000$ when data were slightly or moderately asymmetric.

Therefore, advocates of robust estimation methods take the view that if standard errors and chi-square goodness of fit statistics are statistically corrected, then the power of uncovering the relationships between observed variables and/or latent variables can be enhanced, and the overall model hypothesis testing is able to maintain the type I error rate close to the nominal level in the evaluation of overall model fit. These statistical properties directly translate into substantive and practical advantages – applied researchers are likely to detect genuine relationships with precision and have more reliable model inference.

Response Categories and Observed Distributions

The accuracy and precision of parameter and standard error estimates improved as the number of observed variables' categories increased. These findings support the recommendation that applied researchers are encouraged to use 7-category ordinal observed indicators in a measurement design whenever possible. As was stated in the preceding section, ML and MLR did not fare well on factor loading estimation even when the number of observed variables' categories was seven across all sample sizes. Although the point of superiority of ML and MLR over WLSMV and ULSMV may probably be reached with a larger number of observed

variables' categories (e.g., 9 or 10), the implication for applied research for this further investigation is limited because ordinal observed indicators with more than 9 categories are rarely used in practice. Of the 157 psychometric measures in the SEM applications search I conducted, there were only 6 cases (3.8%) in which ordinal observed indicators had more than 7 categories. Previous studies appear to support the desirability of a larger number of observed variables' categories (e.g., higher psychometric qualities), but increasing the number of response categories may also affect respondents' cognitive capability to process the meaning of each response category (see, e.g., Cook, Heath, & Thompson, 2001; Lietz, 2010). The number of response categories is also closely related to the distributions of ordinal observed variables. Statisticians have agreed that none of the real-world data is perfectly symmetric or/and normal (Gartside, 2001; Nester, 1996). Given the pervasiveness of non-normal data in practice, a general guideline for applied researchers is to examine the extent to which normality violation in the distributions of ordinal observed variables occurs before conducting data analysis. If ordinal observed indicators with moderately asymmetric and leptokurtic distributions are present, interpretation should be with much caution in structural coefficients, factor loadings, standard errors, and chi-square goodness of fit statistics. A cross-validated study can also help replicate the findings.

Limitations and Directions for Future Research

There are innumerable combinations to manipulate in a single simulation study, but one can only focus on certain factors of particular interest to make the research design feasible and manageable. One drawback of carrying out a Monte Carlo study is that results are conditional on the simulation design. This study shares the same limitation as all Monte Carlo simulation studies, in that generalizations are constrained by the specification of the experimental

conditions employed in this study. Several limitations embedded in this study can be considered as potentially fruitful directions for future research.

First, a thorough examination of the effects of violation of the latent normality assumption on WLSMV and ULSMV is beyond the scope of this study. However, the polychoric correlation estimates have been proved to be robust against moderate violations of normality assumption in the latent response variables (Coenders, Satorra, & Saris, 1996; Flora & Curran, 2004; Quiroga, 1992). Thoughtful consideration of a given construct is necessary to judge whether the underlying normality is tenable. The underlying distribution of the frequency of aggressive behaviors per day, for example, is unlikely normally distributed in the population. Besides, a test of the underlying bivariate normality assumption is available with LISREL's processor PRELIS. The assumption of underlying bivariate normality is needed to calculate the polychoric correlation. Future research may investigate ordinal observed indicators with non-normal underlying distributions, or a mixture of underlying normality and non-normality on the same factor.

Although this study did not empirically examine the effects of violation of the underlying normality distributions, some predictions could be made with caution while selecting the three robust estimators. For example, the effects of violation of the underlying normality distributions would likely be more saliently on the performance of WLSMV and ULSMV than that of MLR, holding other conditions constant. In addition, given the condition of multiple underlying distributions across several groups of interest, it could be expected that the situation of heterogeneous underlying distributions would exacerbate the effects of violation of the underlying normality distributions on the performance of WLSMV and ULSMV than that of

MLR with all remaining conditions being equal. Although I only considered ordinal observed variables for each latent construct in this study, I would predict that MLR would likely perform better in the estimation of factor loadings under the condition of a mixture of continuous observed variables and ordinal observed variables, compared to the condition of all ordinal observed variables.

Second, the 5-factor SR model in this study was selected to be the representative of the medium-sized SEM model specification, which is beyond any prior studies documented in the SEM literature. However, further investigation tailored to various applications of SEM is suggested, in which models approximate real-world situations likely to be encountered in empirical studies: (1) a latent growth curve model, with the aim to capture individual trajectories, (2) a multiple-group structural regression model to possibly study group similarities and differences, or (3) a multilevel structural equation model in consideration of clustering effects. Additionally, this study was limited to a saturated structural model; therefore a natural extension of this study is the investigation of a non-saturated structural model by manipulating the number of structural coefficients.

Third, due to the population SR model being correctly specified, the present study does not pursue the possible effects of model misspecification. In fact, applied researchers have to recognize that they may not always work with models without specification errors. The popular supplemental fit index, RMSEA, showed promise for assessing the adequacy of a hypothesized model without specification error in this study, but an interesting avenue of further investigation would examine the power of both corrected chi-square goodness of fit statistics and RMSEA to detect model misspecification. Although previous simulation studies have

suggested that the ordinal CFA models are robust to slight model misspecification (e.g., Flora & Curran, 2004; Maydeu-Olivares, 2006), a worthy topic for future research is to compare the performance of these robust estimators on parameter and standard error estimates, chi-square goodness of fit statistics, and ad hoc fit indices when an SR model with ordinal observed variables under different levels of model specification errors. For example, applied researchers may omit structural regression coefficients or cross-factor loadings, or include certain misspecified structural regression coefficients that are not actually in the population model.

CHAPTER 7

SUMMARY AND CONCLUSIONS

The conclusions of this study can be summarized as follows:

- (1) the four estimators are all subjected to non-convergence problems with 4-category, moderately asymmetric data in the smallest sample size $N = 200$;
- (2) WLSMV and ULSMV are likely to produce inadmissible solutions in some conditions with sample sizes $N = 200$ or 300 ;
- (3) WLSMV and ULSMV yield more accurate factor loading estimates than ML and MLR across all conditions in the study;
- (4) the estimates of structural coefficients under ML and MLR outperform WLSMV and ULSMV in all symmetric data conditions, whereas WLSMV and ULSMV surpass ML and MLR in nearly all asymmetric data conditions;
- (5) the robust standard errors of factor loadings obtained with ULSMV are more precise than those produced by WLSMV and MLR across all conditions;
- (6) the robust standard errors of structural coefficients obtained with WLSMV are more precise than those with ULSMV and MLR in all asymmetric data conditions;
- (7) among the three robust estimators, MLR is inferior to WLSMV and ULSMV in controlling for Type I error rates of testing overall model fit in almost every condition, unless a larger sample size is used (i.e., $N = 1,000$ in this thesis);
- (8) RMSEA seems to be a reliable index in the evaluation of overall model fit when the model has no specification error;
- (9) the benefit of using diagonal weights can be found in the estimation of factor loadings and structural coefficients and robust standard errors of structural coefficients, but not in the estimation of robust standard errors of factor loadings

and the mean- and variance-adjusted chi-square goodness of fit statistics across all conditions; and

- (10) the accuracy and precision of factor loadings and structural coefficients, and standard error estimates of both factor loadings and structural coefficients improve with increasing sample size and number of observed variables' categories but decrease with a greater level of asymmetric distributions.

Although WLSMV and ULSMV can be generally recommended to use when fitting an SR model with ordinal observed variables, it is worthwhile to point out that each estimator considered in this thesis has its own advantages and disadvantages. This study provides evidence that WLSMV and ULSMV perform better than MLR, and that MLR does so than ML in many conditions. WLSMV and ULSMV do not need a large sample size for the recovery of population factor loadings and structural coefficients, and to evaluate overall model fit using the mean- and variance-adjusted chi-square goodness of fit statistics, but a medium sample (e.g., $N = 500$ or more) is required to obtain stable standard error estimates of both factor loadings and structural coefficients. In addition, the benefit of using diagonal weights in WLSMV can be observed in the estimation of factor loadings and structural coefficients as well as robust standard errors of structural coefficients. Compared to ML and MLR, WLSMV and ULSMV have more reliable model inference in small sample sizes and are more likely to detect small structural relationships with precision when data were slightly or moderately asymmetric.

On the other hand, MLR has its own strengths – e.g., generally less biased standard error estimates of factor loadings and structural coefficients, and accurate and precise

structural coefficient estimates in the conditions of symmetric data. MLR does not require a large sample to produce stable structural coefficient estimates and standard error estimates of factor loadings and structural coefficients, but may need a quite large sample (e.g., $N = 1,000$ or more) to control for Type I error rates of testing overall model fit, despite the existence of moderate underestimation in factor loading estimates. However, the small amount of bias in structural coefficient estimates makes MLR practically recommendable when applied researchers are primarily concerned with structural relationships among latent constructs. Consistent with asymptotic theory, ML can perform pretty well in a relatively large sample when data are near symmetric (or close to normal). Generally speaking, the moderate-to-substantial underestimation of standard errors for both factor loadings and structural coefficients, and considerable inflation of chi-square goodness of fit statistics make ML less attractive and favorable in practice, particularly when data moderately deviate from normality. However, ML and MLR with full information estimation can be considered a promising approach in research practice when applied researchers have to deal with missing data because the treatment of missing data in WLSMV and ULSMV estimators remains technically underdeveloped.

It is important to keep in mind that any working recommendations provided herein are based on the current model configurations. This study did not consider the possible effects of violation of the underlying normality distributions. However, it can be expected that the effects of the underlying normality assumption violation would be more saliently on the performance of WLSMV and ULSMV than that of MLR on model estimation. Furthermore, it is unclear that the performance of the four estimators on parameter and standard error estimates, chi-square goodness of fit statistics, and RMSEA in an SR model with ordinal observed variables under

varying levels of model misspecification. Future investigations into these simulation design characteristics would likely render informative suggestions and more fine-grained recommendations. Applied researchers still have to weigh the pros and cons of different estimators, in order to make better-informed decisions while analyzing an SR model with ordinal observed indicators.

APPENDICES

Table 1. Overview of Six Major Simulation Studies in Ordinal CFA

	Studies					
	Oranje	Beauducel & Herzberg	Lei	Forero, Maydeu-Olivares, & Gallardo-Pujol	Yang-Wallentin, Joreskog, & Luo	Rhemtulla, Brosseau-Liard, & Savalei
Year	2003	2006	2009	2009	2010	2012
Sample Size	200, 500, 1000	250, 500, 750, 1000	100, 250, 1000	200, 500, 2000	100, 200, 400, 800, 1600	100, 150, 350, 600
No. Factors	1 & 3	1, 2, 4, 8	2 & 3	1 & 3	2 & 4	2
No. Variables	5, 10, 15, 30, 45	5, 10, 20, 40	6 & 9	9, 21, 42	6 & 16	10 & 20
No. Categories	2, 3, 5	2, 3, 4, 5, 6	5	2 & 5	2, 5, 7	2, 3, 4, 5, 6, 7
Item Asymmetry	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	ML* & WLSMV	ML & WLSMV	ML, ROBUST & WLSMV	ULS & WLSMV	ML*, ULS*, DWLS*	ULSMV & MLMV
Software	LISREL & <i>Mplus</i>	<i>Mplus</i>	EQS & <i>Mplus</i>	<i>Mplus</i>	LISREL	<i>Mplus</i>

Note. *Polychoric correlation estimates and estimated asymptotic covariance matrix need to compute from PRELIS before performing LISREL.

Table 2. Robust Estimation Comparison in the Three SEM Software Packages

Estimation	SEM Software Programs		
	<i>Mplus</i>	EQS	LISREL
Robust Maximum Likelihood	MLR	ML, ROBUST	ML*
Robust Unweighted Least Squares	ULSMV	LS, ROBUST	ULS*
Robust Weighted Least Squares	WLSMV	x	DWLS*

Note. *Polychoric correlation estimates and estimated asymptotic covariance matrix need to compute from PRELIS before performing LISREL. Robust weighted least squares estimation is currently unavailable in EQS.

Table 3. Comparison of Two Major Estimation Approaches: Maximum Likelihood and Least Squares in *Mplus*

Estimators	Parameters	Standard Errors	Chi-square
Maximum Likelihood			
ML	ML	ML	ML
MLM	MLM = ML	MLM = MLMV \neq ML	MLM \neq ML
MLMV	MLMV = ML	MLMV = MLM \neq ML	MLMV \neq ML
MLR	MLR = ML	MLR \neq ML	MLR \neq ML
Least Squares			
ULS	ULS	ULS	ULS
ULSMV	ULSMV = ULS	ULSMV \neq ULS	ULSMV \neq ULS
WLS	WLS	WLS	WLS
WLSM	WLSM = WLSMV \neq WLS	WLSM = WLSMV \neq WLS	WLSM \neq WLS
WLSMV	WLSMV = WLSM \neq WLS	WLSMV = WLSM \neq WLS	WLSMV \neq WLS

Note. ML = maximum likelihood, MLM = maximum likelihood with a mean-adjusted chi-square statistic, MLMV = maximum likelihood with a mean- and variance-adjusted chi-square statistic; ULS = unweighted least squares, ULSMV = unweighted least squares with a mean- and variance-adjusted chi-square statistic; WLS = weighted least squares, WLSM = weighted least squares with a mean-adjusted chi-square statistic, WLSMV = weighted least squares with a mean- and variance-adjusted chi-square statistic.

Table 4(a). Cases of Non-Convergence

Est.	N	Dis.	Symmetry				Slight Asymmetry				Moderate Asymmetry				Bipolarization			
		Cat.	4	5	6	7	4	5	6	7	4	5	6	7	4	5	6	7
ML/MLR	200		0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
	300		0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	400		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	750		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1000		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WLSMV	200		0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0
	300		0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	400		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	750		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1000		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ULSMV	200		0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
	300		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	400		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	750		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1000		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Note. Est. = Estimators, Dis. = distribution type, and Cat. = number of categories. ML/MLR = maximum likelihood/robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweighted least squares. N = sample sizes.

Table 4(b). Cases of Inadmissible Solutions

Est.	N	Dis.	Symmetry				Slight Asymmetry				Moderate Asymmetry				Bipolarization			
		Cat.	4	5	6	7	4	5	6	7	4	5	6	7	4	5	6	7
ML/MLR	200		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	300		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	400		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	750		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1000		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WLSMV	200		0	1	0	0	0	1	0	0	0	0	0	2	3	1	1	0
	300		0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	400		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	750		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1000		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ULSMV	200		0	1	0	0	1	2	0	0	6	0	0	1	4	1	1	1
	300		0	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0
	400		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	750		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1000		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1500		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Note. Est. = Estimators, Dis. = distribution type, and Cat. = number of categories. ML/MLR = maximum likelihood/robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweighted least squares. N = sample sizes.

Table 5. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 200$)

Dis.	Cat.	ML/MLR				WLSMV				ULSMV			
		RB_A		MSE_A		RB_A		MSE_A		RB_A		MSE_A	
		FL	SC	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC
sym	4	-7.00	-2.13	0.0142	0.9570	0.30	-1.73	0.0118	0.9716	-0.08	-1.89	0.0122	0.9454
	5	-4.44	-1.33	0.0107	0.7460	0.22	-0.74	0.0108	0.7847	-0.13	-0.82	0.0112	0.7637
	6	-3.20	-0.90	0.0095	0.6817	0.22	-0.52	0.0103	0.7338	-0.11	-0.82	0.0107	0.7123
	7	-2.49	-0.42	0.0089	0.6734	0.13	-0.17	0.0100	0.7222	-0.19	-0.20	0.0104	0.7002
slight	4	-10.10	-2.79	0.0216	1.3215	0.20	-0.65	0.0136	1.1989	-0.23	-0.97	0.0140	1.1492
	5	-6.92	-2.17	0.0154	0.8693	0.21	-1.02	0.0118	0.9212	-0.18	-1.00	0.0122	0.8813
	6	-6.04	-2.35	0.0138	0.8267	0.15	-1.50	0.0111	0.8199	-0.20	-1.69	0.0115	0.7826
	7	-5.43	-1.40	0.0132	0.7546	0.18	-0.31	0.0105	0.7231	-0.16	-0.41	0.0109	0.7017
mod	4	-11.86	-3.23	0.0291	1.2265	0.06	0.50	0.0161	1.4529	-0.55	-0.02	0.0168	1.4251
	5	-9.26	-3.03	0.0218	1.1744	0.11	-0.25	0.0133	1.1013	-0.37	-0.69	0.0138	1.1302
	6	-8.78	-3.18	0.0212	1.0577	0.10	0.11	0.0126	1.0693	-0.32	-0.35	0.0131	1.0445
	7	-8.73	-3.03	0.0213	1.0262	0.11	0.25	0.0121	0.9686	-0.30	-0.16	0.0126	0.9564

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML/MLR = maximum likelihood/robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. FL represents factor loadings and SC is structural coefficients.

Table 6. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 300$)

		ML/MLR				WLSMV				ULSMV			
		RB_A		MSE_A		RB_A		MSE_A		RB_A		MSE_A	
Dis.	Cat.	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC
sym	4	-6.91	-0.05	0.0108	0.4427	0.23	0.27	0.0076	0.4781	-0.03	0.18	0.0079	0.4693
	5	-4.44	-0.26	0.0076	0.4325	0.12	0.23	0.0069	0.4365	-0.12	0.09	0.0072	0.4316
	6	-3.19	0.04	0.0065	0.3915	0.10	0.44	0.0065	0.4135	-0.11	0.35	0.0068	0.4155
	7	-2.43	0.08	0.0060	0.3824	0.08	0.47	0.0064	0.3997	-0.12	0.35	0.0067	0.3995
slight	4	-9.96	-0.89	0.0174	0.6313	0.13	0.55	0.0088	0.6205	-0.15	0.32	0.0091	0.6097
	5	-6.87	-0.81	0.0117	0.5484	0.11	0.57	0.0076	0.5144	-0.15	0.42	0.0079	0.5208
	6	-5.88	-0.64	0.0101	0.4766	0.14	0.48	0.0071	0.4545	-0.10	0.33	0.0074	0.4618
	7	-5.36	-0.67	0.0097	0.4908	0.06	0.40	0.0069	0.4304	-0.17	0.25	0.0072	0.4297
mod	4	-11.72	-3.51	0.0238	0.7830	0.07	-0.23	0.0107	0.8631	-0.32	-0.64	0.0112	0.7905
	5	-9.12	-2.62	0.0173	0.6345	0.07	0.08	0.0088	0.6117	-0.24	0.05	0.0092	0.6212
	6	-8.63	-2.81	0.0164	0.7001	0.10	0.34	0.0081	0.5747	-0.19	0.18	0.0084	0.5856
	7	-8.65	-2.68	0.0166	0.6522	0.04	0.51	0.0078	0.5210	-0.23	0.32	0.0082	0.5271

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML/MLR = maximum likelihood/robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. FL represents factor loadings and SC is structural coefficients.

Table 7. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 500$)

		ML/MLR				WLSMV				ULSMV			
		RB_A		MSE_A		RB_A		MSE_A		RB_A		MSE_A	
Dis.	Cat.	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC
sym	4	-6.95	-0.78	0.0084	0.2751	0.12	-0.48	0.0045	0.2953	-0.05	-0.53	0.0046	0.3000
	5	-4.40	-0.59	0.0053	0.2451	0.09	-0.28	0.0041	0.2618	-0.06	-0.33	0.0043	0.2676
	6	-3.18	-0.89	0.0043	0.2371	0.06	-0.69	0.0039	0.2522	-0.08	-0.73	0.0041	0.2560
	7	-2.43	-0.38	0.0038	0.2211	0.06	-0.13	0.0038	0.2366	-0.08	-0.19	0.0039	0.2413
slight	4	-9.85	-1.27	0.0141	0.3203	0.14	-0.18	0.0052	0.3231	-0.04	-0.26	0.0053	0.3279
	5	-6.78	-1.70	0.0087	0.2976	0.10	-0.68	0.0045	0.3035	-0.06	-0.75	0.0047	0.3082
	6	-5.86	-1.35	0.0073	0.2775	0.08	-0.62	0.0041	0.2794	-0.08	-0.68	0.0043	0.2865
	7	-5.27	-1.26	0.0067	0.2640	0.07	-0.24	0.0040	0.2422	-0.08	-0.29	0.0042	0.2475
mod	4	-11.56	-3.65	0.0192	0.3944	0.11	-0.69	0.0062	0.4510	-0.14	-0.91	0.0065	0.4686
	5	-9.00	-3.48	0.0133	0.3458	0.09	-1.35	0.0052	0.3485	-0.10	-1.52	0.0054	0.3566
	6	-8.56	-3.44	0.0126	0.3428	0.03	-1.15	0.0048	0.3113	-0.15	-1.28	0.0050	0.3162
	7	-8.49	-3.39	0.0125	0.3428	0.04	-0.90	0.0047	0.3017	-0.13	-1.03	0.0049	0.3084

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML/MLR = maximum likelihood/robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. FL represents factor loadings and SC is structural coefficients.

Table 8. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 1,000$)

Dis.	Cat.	ML/MLR				WLSMV				ULSMV			
		RB_A		MSE_A		RB_A		MSE_A		RB_A		MSE_A	
		FL	SC	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC
sym	4	-6.89	-0.44	0.0065	0.1363	0.08	-0.23	0.0023	0.1433	0.01	-0.27	0.0024	0.1452
	5	-4.34	-0.60	0.0035	0.1246	0.09	-0.34	0.0020	0.1284	0.01	-0.37	0.0021	0.1299
	6	-3.07	-0.54	0.0025	0.1133	0.11	-0.36	0.0019	0.1177	0.04	-0.39	0.0020	0.1190
	7	-2.34	-0.59	0.0021	0.1082	0.09	-0.37	0.0019	0.1126	0.03	-0.39	0.0020	0.1142
slight	4	-9.78	-1.18	0.0117	0.1632	0.15	-0.27	0.0026	0.1622	0.05	-0.24	0.0027	0.1647
	5	-6.69	-1.51	0.0065	0.1484	0.11	-0.87	0.0022	0.1431	0.03	-0.92	0.0023	0.1446
	6	-5.79	-1.20	0.0053	0.1345	0.10	-0.51	0.0021	0.1248	0.03	-0.54	0.0022	0.1263
	7	-5.20	-1.08	0.0046	0.1377	0.10	-0.47	0.0020	0.1228	0.04	-0.50	0.0021	0.1246
mod	4	-11.49	-3.04	0.0162	0.1840	0.14	-0.16	0.0031	0.2001	0.02	-0.25	0.0033	0.2021
	5	-8.94	-2.47	0.0106	0.1694	0.11	-0.16	0.0026	0.1648	0.02	-0.21	0.0027	0.1660
	6	-8.46	-2.57	0.0098	0.1699	0.11	-0.32	0.0024	0.1498	0.03	-0.38	0.0025	0.1524
	7	-8.44	-2.47	0.0098	0.1709	0.11	-0.32	0.0023	0.1476	0.03	-0.36	0.0024	0.1495

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML/MLR = maximum likelihood/robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. FL represents factor loadings and SC is structural coefficients.

Table 9. The Average Root Mean Squared Error (MSE_A) for the Four Structural Coefficients ($N = 1,000$)

		ML/MLR				WLSMV				ULSMV			
		Structural Coefficients											
Dis. & Cat.		$\gamma_{22} = .2$	$\beta_{31} = .2$	$\beta_{21} = .3$	$\phi_{12} = .3$	$\gamma_{22} = .2$	$\beta_{31} = .2$	$\beta_{21} = .3$	$\phi_{12} = .3$	$\gamma_{22} = .2$	$\beta_{31} = .2$	$\beta_{21} = .3$	$\phi_{12} = .3$
sym	4	0.1288	0.2066	0.0820	0.0176	0.1323	0.2227	0.0866	0.0179	0.1349	0.2283	0.0884	0.0178
	5	0.1203	0.1769	0.0805	0.0172	0.1241	0.1855	0.0846	0.0176	0.1258	0.1893	0.0864	0.0176
	6	0.1045	0.1596	0.0688	0.0175	0.1086	0.1668	0.0721	0.0179	0.1105	0.1701	0.0735	0.0179
	7	0.1038	0.1473	0.0676	0.0169	0.1090	0.1565	0.0719	0.0172	0.1105	0.1602	0.0730	0.0171
slight	4	0.1541	0.2350	0.1009	0.0227	0.1576	0.2353	0.1031	0.0217	0.1588	0.2414	0.1044	0.0216
	5	0.1440	0.2010	0.0987	0.0207	0.1392	0.1965	0.0967	0.0193	0.1407	0.2018	0.0981	0.0193
	6	0.1211	0.1904	0.0821	0.0194	0.1148	0.1795	0.0761	0.0180	0.1163	0.1841	0.0777	0.0180
	7	0.1324	0.1863	0.0895	0.0197	0.1238	0.1677	0.0816	0.0175	0.1265	0.1714	0.0836	0.0175
mod	4	0.1890	0.2554	0.1197	0.0310	0.2014	0.2926	0.1350	0.0264	0.2048	0.2995	0.1370	0.0262
	5	0.1551	0.2488	0.0999	0.0266	0.1433	0.2446	0.0959	0.0226	0.1453	0.2494	0.0978	0.0226
	6	0.1614	0.2319	0.1088	0.0271	0.1384	0.2069	0.0952	0.0210	0.1402	0.2137	0.0970	0.0210
	7	0.1521	0.2381	0.1005	0.0273	0.1285	0.2112	0.0883	0.0199	0.1304	0.2169	0.0899	0.0199

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML/MLR = maximum likelihood/robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweight least squares.

Table 10. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 200$)

Dis.	Cat.	ML				MLR			
		RB_A		MSE_A		RB_A		MSE_A	
		SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}
sym	4	-1.98	-4.28	0.0093	0.0963	-0.66	-2.56	0.0143	0.0990
	5	-2.34	-2.82	0.0103	0.0451	-1.14	-1.17	0.0152	0.0505
	6	-2.53	-2.61	0.0103	0.0399	-1.37	-1.19	0.0150	0.0452
	7	-2.79	-4.27	0.0114	0.0358	-1.59	-2.71	0.0162	0.0415
slight	4	-7.32	-8.21	0.0143	0.3196	0.31	-1.93	0.0166	0.4033
	5	-8.13	-5.80	0.0159	0.0751	-0.80	0.64	0.0170	0.0898
	6	-6.86	-5.14	0.0143	0.0643	-0.10	0.59	0.0167	0.0738
	7	-8.46	-4.81	0.0167	0.0527	-0.48	1.57	0.0162	0.0637
mod	4	-16.97	-11.63	0.0374	0.3579	0.33	2.71	0.0191	0.4418
	5	-14.85	-12.33	0.0310	0.2261	0.30	-0.04	0.0186	0.2428
	6	-16.07	-12.02	0.0348	0.1406	-0.09	1.01	0.0178	0.1590
	7	-16.75	-12.47	0.0370	0.1186	0.02	1.18	0.0176	0.1314

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table 10 (cont'd)

		WLSMV				ULSMV			
		RB _A		MSE _A		RB _A		MSE _A	
Dis.	Cat.	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}
sym	4	-5.77	-6.62	0.0157	0.1017	-3.81	-4.27	0.0120	0.1088
	5	-5.96	-5.91	0.0158	0.0528	-4.67	-3.87	0.0122	0.0514
	6	-7.45	-8.11	0.0173	0.0544	-6.31	-5.86	0.0139	0.0543
	7	-8.10	-9.96	0.0187	0.0522	-7.23	-7.80	0.0150	0.0475
slight	4	-6.14	-6.32	0.0169	0.1399	-4.97	-4.37	0.0134	0.1251
	5	-7.15	-7.34	0.0176	0.0816	-6.11	-5.15	0.0140	0.0709
	6	-7.38	-7.82	0.0175	0.0551	-6.38	-5.49	0.0142	0.0546
	7	-8.05	-7.54	0.0180	0.0581	-6.93	-5.27	0.0144	0.0514
mod	4	-6.66	5.13	0.0188	0.2212	-5.06	-3.56	0.0142	0.2649
	5	-6.84	-7.11	0.0172	0.1199	-5.30	-5.64	0.0134	0.1272
	6	-7.86	-10.01	0.0182	0.0971	-7.06	-8.28	0.0146	0.0918
	7	-7.72	-9.52	0.0184	0.0936	-7.14	-8.05	0.0154	0.0916

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table 11. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 300$)

Dis.	Cat.	ML				MLR			
		RB_A		MSE_A		RB_A		MSE_A	
		SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}
sym	4	-0.74	2.52	0.0065	0.0392	-0.02	3.95	0.0098	0.0451
	5	-1.10	-0.08	0.0068	0.0273	-0.43	1.18	0.0098	0.0314
	6	-0.71	0.95	0.0065	0.0278	0.05	2.06	0.0096	0.0316
	7	-1.25	1.38	0.0069	0.0274	-0.54	2.56	0.0100	0.0324
slight	4	-7.26	-4.63	0.0114	0.0557	-0.43	1.50	0.0103	0.0679
	5	-7.98	-6.75	0.0144	0.0460	-0.87	-0.62	0.0108	0.0501
	6	-7.08	-4.13	0.0115	0.0401	-0.78	1.45	0.0105	0.0509
	7	-8.94	-5.46	0.0147	0.0348	-1.51	0.66	0.0107	0.0414
mod	4	-16.82	-14.01	0.0346	0.2006	-0.69	-0.35	0.0121	0.2541
	5	-15.08	-10.93	0.0292	0.0867	-0.84	1.46	0.0114	0.1692
	6	-16.06	-12.60	0.0325	0.1475	-0.99	0.18	0.0115	0.0519
	7	-16.68	-12.79	0.0345	0.0623	-0.96	0.49	0.0112	0.0640

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table 11 (cont'd)

		WLSMV				ULSMV			
		RB _A		MSE _A		RB _A		MSE _A	
Dis.	Cat.	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}
sym	4	-3.04	-0.22	0.0101	0.0435	-1.97	1.41	0.0080	0.0456
	5	-3.03	-1.06	0.0097	0.0325	-2.68	0.63	0.0080	0.0334
	6	-4.02	-3.04	0.0100	0.0341	-3.50	-1.78	0.0079	0.0341
	7	-4.69	-2.51	0.0105	0.0345	-4.49	-0.97	0.0086	0.0335
slight	4	-3.97	-1.56	0.0106	0.0550	-3.06	0.33	0.0083	0.0598
	5	-3.91	-3.73	0.0103	0.0460	-3.72	-2.44	0.0084	0.0522
	6	-4.28	-3.25	0.0102	0.0361	-3.50	-1.82	0.0082	0.0374
	7	-5.33	-3.78	0.0113	0.0337	-5.02	-2.16	0.0093	0.0339
mod	4	-5.34	-4.21	0.0135	0.1137	-4.89	-1.26	0.0110	0.2204
	5	-4.73	-2.70	0.0116	0.0913	-4.09	-1.51	0.0092	0.1000
	6	-4.49	-3.27	0.0113	0.0555	-4.12	-1.91	0.0091	0.0692
	7	-4.89	-3.08	0.0114	0.0382	-4.74	-1.73	0.0093	0.0407

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table 12. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 500$)

Dis.	Cat.	ML				MLR			
		RB_A		MSE_A		RB_A		MSE_A	
		SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}
sym	4	0.74	-1.45	0.0041	0.0150	0.88	-0.59	0.0060	0.0166
	5	0.08	-0.47	0.0048	0.0120	0.23	0.50	0.0066	0.0137
	6	-0.18	-0.97	0.0046	0.0119	0.16	-0.25	0.0065	0.0134
	7	-0.29	-0.35	0.0051	0.0113	0.09	0.48	0.0070	0.0130
slight	4	-5.91	-3.55	0.0084	0.0249	0.49	2.38	0.0075	0.0295
	5	-6.10	-5.77	0.0089	0.0186	0.42	0.16	0.0080	0.0196
	6	-5.04	-5.05	0.0075	0.0163	0.96	0.23	0.0077	0.0175
	7	-7.00	-4.72	0.0097	0.0159	0.04	1.26	0.0072	0.0175
mod	4	-15.22	-12.88	0.0276	0.0402	0.38	0.23	0.0080	0.0331
	5	-13.61	-10.90	0.0228	0.0275	0.17	1.05	0.0074	0.0224
	6	-14.59	-11.81	0.0259	0.0310	-0.03	0.76	0.0074	0.0246
	7	-15.14	-12.79	0.0276	0.0323	0.12	0.15	0.0074	0.0227

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table 12 (cont'd)

		WLSMV				ULSMV			
		RB _A		MSE _A		RB _A		MSE _A	
Dis.	Cat.	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}
sym	4	-0.43	-3.00	0.0065	0.0169	0.25	-2.05	0.0054	0.0171
	5	-1.23	-1.70	0.0069	0.0145	-1.36	-1.04	0.0055	0.0149
	6	-1.74	-2.76	0.0068	0.0139	-1.56	-1.83	0.0055	0.0141
	7	-1.91	-2.60	0.0069	0.0135	-2.06	-1.80	0.0056	0.0138
slight	4	-1.64	-0.14	0.0070	0.0239	-1.46	0.49	0.0057	0.0249
	5	-1.80	-3.20	0.0074	0.0183	-2.14	-2.51	0.0062	0.0185
	6	-0.94	-4.00	0.0066	0.0155	-1.32	-3.92	0.0053	0.0160
	7	-1.99	-1.15	0.0069	0.0147	-2.09	-0.54	0.0058	0.0152
mod	4	-2.03	-2.60	0.0080	0.0295	-1.80	-2.48	0.0060	0.0324
	5	-1.70	-2.07	0.0070	0.0202	-0.97	-1.82	0.0061	0.0212
	6	-2.32	-1.69	0.0071	0.0201	-1.96	-0.87	0.0060	0.0207
	7	-2.58	-3.04	0.0071	0.0180	-2.49	-2.42	0.0062	0.0187

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table 13. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 1,000$)

Dis.	Cat.	ML				MLR			
		RB_A		MSE_A		RB_A		MSE_A	
		SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}
sym	4	0.56	-1.09	0.0028	0.0072	0.29	-0.55	0.0036	0.0079
	5	0.36	-2.31	0.0026	0.0060	0.14	-1.74	0.0034	0.0066
	6	1.18	-0.82	0.0028	0.0056	1.16	-0.35	0.0037	0.0063
	7	1.51	-0.76	0.0030	0.0054	1.48	-0.26	0.0038	0.0062
slight	4	-5.50	-5.64	0.0055	0.0118	0.45	-0.15	0.0037	0.0107
	5	-6.38	-8.22	0.0067	0.0135	-0.24	-2.74	0.0038	0.0094
	6	-5.34	-5.04	0.0058	0.0090	0.26	0.13	0.0041	0.0084
	7	-6.24	-7.11	0.0066	0.0113	0.49	-1.51	0.0038	0.0083
mod	4	-14.72	-13.60	0.0239	0.0262	0.28	-0.70	0.0039	0.0111
	5	-12.31	-11.68	0.0178	0.0205	1.17	-0.08	0.0043	0.0100
	6	-13.46	-13.17	0.0208	0.0244	0.71	-1.14	0.0041	0.0102
	7	-13.67	-13.55	0.0213	0.0257	1.19	-1.08	0.0041	0.0105

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table 13 (cont'd)

		WLSMV				ULSMV			
		RB _A		MSE _A		RB _A		MSE _A	
Dis.	Cat.	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}
sym	4	-0.73	-1.91	0.0035	0.0082	0.23	-1.47	0.0031	0.0084
	5	-0.10	-2.32	0.0034	0.0070	0.29	-1.75	0.0031	0.0071
	6	-0.45	-1.82	0.0033	0.0066	-0.32	-1.25	0.0029	0.0068
	7	0.05	-1.89	0.0034	0.0065	-0.02	-1.31	0.0029	0.0066
slight	4	-0.26	-1.36	0.0038	0.0098	0.38	-1.01	0.0032	0.0103
	5	-0.84	-3.48	0.0035	0.0088	-0.58	-2.79	0.0031	0.0089
	6	-0.80	-0.62	0.0034	0.0070	0.02	-0.18	0.0028	0.0074
	7	-0.69	-2.35	0.0036	0.0072	0.13	-1.76	0.0032	0.0073
mod	4	-1.41	-1.93	0.0042	0.0115	-1.17	-1.31	0.0035	0.0119
	5	-0.41	-0.57	0.0037	0.0093	-0.33	-0.14	0.0033	0.0095
	6	-0.57	-1.89	0.0040	0.0085	-0.87	-1.56	0.0036	0.0087
	7	-0.07	-2.34	0.0038	0.0081	-0.16	-1.88	0.0031	0.0083

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table 14. Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 200$)

		ML				MLR			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	5.14	9.60	0.015	0.00	5.99	11.60	0.016	0.00
	5	5.69	13.80	0.015	0.00	6.431	15.80	0.016	0.00
	6	5.61	14.00	0.015	0.00	6.37	15.00	0.016	0.00
	7	5.77	12.40	0.015	0.00	6.54	14.60	0.016	0.00
slight	4	13.52	32.20	0.023	0.00	9.58	21.20	0.019	0.00
	5	14.33	34.40	0.024	0.00	9.57	22.20	0.019	0.00
	6	13.53	31.40	0.023	0.00	9.33	19.80	0.019	0.00
	7	14.79	36.60	0.025	0.00	9.59	23.00	0.019	0.00
mod	4	28.44	74.49	0.036	0.00	11.70	26.11	0.022	0.00
	5	25.90	68.40	0.034	0.00	11.37	26.20	0.021	0.00
	6	27.41	67.74	0.035	0.00	11.62	27.66	0.021	0.00
	7	28.92	74.50	0.036	0.00	11.78	26.91	0.021	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. M = mean.

Table 14 (cont'd)

		WLSMV				ULSMV			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	2.88	4.80	0.011	0.00	0.80	2.60	0.009	0.00
	5	3.45	6.21	0.012	0.00	1.50	3.61	0.010	0.00
	6	4.65	7.20	0.013	0.00	2.55	4.80	0.011	0.00
	7	5.23	8.20	0.014	0.00	2.99	5.04	0.012	0.00
slight	4	4.04	4.82	0.013	0.00	1.83	1.61	0.010	0.00
	5	4.73	6.02	0.014	0.00	2.62	3.22	0.011	0.00
	6	5.25	7.01	0.014	0.00	3.26	4.81	0.012	0.00
	7	5.78	9.02	0.015	0.00	3.45	5.81	0.012	0.00
mod	4	5.19	4.89	0.014	0.00	3.36	2.86	0.012	0.00
	5	5.58	8.22	0.015	0.00	3.76	5.20	0.012	0.00
	6	5.94	9.02	0.015	0.00	4.12	6.81	0.013	0.00
	7	6.47	9.70	0.016	0.00	4.65	7.46	0.013	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. M = mean.

Table 15. Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 300$)

		ML				MLR			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	2.83	9.80	0.010	0.00	3.24	10.80	0.010	0.00
	5	3.59	10.00	0.011	0.00	3.89	10.80	0.011	0.00
	6	3.53	10.80	0.010	0.00	3.92	11.40	0.011	0.00
	7	3.82	10.40	0.011	0.00	4.17	11.20	0.011	0.00
slight	4	11.82	27.60	0.018	0.00	6.12	12.80	0.013	0.00
	5	13.59	33.87	0.019	0.00	7.24	18.04	0.014	0.00
	6	11.95	27.66	0.018	0.00	6.14	13.63	0.013	0.00
	7	13.76	33.40	0.019	0.00	6.91	16.60	0.013	0.00
mod	4	26.40	67.74	0.028	0.00	7.37	17.43	0.014	0.00
	5	24.11	64.20	0.027	0.00	7.40	18.60	0.014	0.00
	6	26.62	69.00	0.028	0.00	8.39	20.00	0.015	0.00
	7	27.29	68.60	0.029	0.00	7.87	19.80	0.014	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. M = mean.

Table 15 (cont'd)

		WLSMV				ULSMV			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	1.65	4.00	0.009	0.00	0.23	3.40	0.007	0.00
	5	2.13	5.00	0.009	0.00	0.80	4.40	0.008	0.00
	6	3.12	6.20	0.010	0.00	1.71	4.60	0.008	0.00
	7	3.44	7.60	0.010	0.00	1.98	5.60	0.008	0.00
slight	4	2.57	4.01	0.009	0.00	1.17	2.61	0.008	0.00
	5	3.61	6.61	0.010	0.00	2.11	4.61	0.009	0.00
	6	3.53	6.81	0.010	0.00	2.15	5.21	0.009	0.00
	7	4.06	7.60	0.010	0.00	2.51	6.20	0.009	0.00
mod	4	3.44	5.24	0.010	0.00	2.39	4.05	0.009	0.00
	5	3.67	6.40	0.010	0.00	2.58	5.20	0.009	0.00
	6	4.00	5.80	0.011	0.00	2.85	4.41	0.009	0.00
	7	4.12	5.60	0.011	0.00	2.95	3.60	0.010	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. M = mean.

Table 16. Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 500$)

		ML				MLR			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	2.84	9.20	0.008	0.00	2.92	8.80	0.008	0.00
	5	2.56	9.20	0.008	0.00	2.54	9.20	0.007	0.00
	6	2.37	8.40	0.007	0.00	2.45	8.60	0.008	0.00
	7	2.68	9.60	0.008	0.00	2.70	9.40	0.008	0.00
slight	4	10.79	24.00	0.013	0.00	3.86	9.40	0.009	0.00
	5	11.11	25.40	0.013	0.00	3.59	9.20	0.008	0.00
	6	11.05	27.05	0.013	0.00	3.98	11.02	0.009	0.00
	7	11.38	26.80	0.014	0.00	3.46	9.60	0.008	0.00
mod	4	25.65	66.53	0.022	0.00	5.00	10.06	0.009	0.00
	5	23.57	60.32	0.021	0.00	5.15	12.22	0.009	0.00
	6	24.48	63.40	0.021	0.00	4.72	12.80	0.009	0.00
	7	25.88	67.40	0.022	0.00	4.94	12.40	0.009	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. M = mean.

Table 16 (cont'd)

		WLSMV				ULSMV			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	1.55	4.80	0.007	0.00	0.64	4.00	0.006	0.00
	5	1.46	4.60	0.006	0.00	0.60	3.80	0.006	0.00
	6	1.86	5.60	0.007	0.00	0.86	4.40	0.006	0.00
	7	2.30	7.40	0.007	0.00	1.34	5.80	0.006	0.00
slight	4	1.36	3.00	0.006	0.00	0.34	2.60	0.006	0.00
	5	1.99	3.80	0.007	0.00	1.01	3.80	0.006	0.00
	6	2.13	5.61	0.007	0.00	1.21	5.20	0.006	0.00
	7	1.91	3.80	0.007	0.00	1.03	3.20	0.006	0.00
mod	4	2.31	4.62	0.007	0.00	1.56	3.80	0.007	0.00
	5	2.41	6.83	0.007	0.00	1.65	4.80	0.007	0.00
	6	2.28	4.80	0.007	0.00	1.48	3.60	0.006	0.00
	7	2.66	6.60	0.007	0.00	1.88	5.20	0.007	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. M = mean.

Table 17. Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 1,000$)

		ML				MLR			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	0.81	5.60	0.005	0.00	0.70	5.40	0.005	0.00
	5	0.88	6.00	0.005	0.00	0.67	5.80	0.005	0.00
	6	1.28	7.20	0.005	0.00	1.12	7.20	0.005	0.00
	7	1.10	5.80	0.005	0.00	0.93	5.80	0.005	0.00
slight	4	10.06	22.20	0.009	0.00	2.22	6.60	0.005	0.00
	5	9.51	22.00	0.009	0.00	1.12	7.80	0.005	0.00
	6	9.36	19.80	0.009	0.00	1.45	7.40	0.005	0.00
	7	9.73	21.20	0.009	0.00	0.93	6.80	0.005	0.00
mod	4	23.43	63.20	0.014	0.00	1.78	7.80	0.005	0.00
	5	21.09	56.40	0.014	0.00	1.69	5.40	0.005	0.00
	6	22.43	58.00	0.014	0.00	1.60	5.60	0.005	0.00
	7	23.35	59.80	0.014	0.00	1.50	7.40	0.005	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. M = mean.

Table 17 (cont'd)

		WLSMV				ULSMV			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	0.28	3.00	0.004	0.00	-0.21	2.80	0.004	0.00
	5	0.30	3.60	0.004	0.00	-0.13	3.01	0.004	0.00
	6	1.14	3.60	0.005	0.00	0.74	3.80	0.004	0.00
	7	0.90	4.40	0.004	0.00	0.56	3.80	0.004	0.00
slight	4	1.04	3.60	0.005	0.00	0.56	3.41	0.004	0.00
	5	0.62	4.80	0.005	0.00	0.19	4.80	0.004	0.00
	6	0.91	4.40	0.004	0.00	0.49	4.40	0.004	0.00
	7	0.79	4.60	0.004	0.00	0.30	3.80	0.004	0.00
mod	4	0.61	5.00	0.004	0.00	0.30	4.41	0.004	0.00
	5	0.89	4.20	0.004	0.00	0.55	3.60	0.004	0.00
	6	0.71	3.60	0.004	0.00	0.35	2.60	0.004	0.00
	7	0.55	4.20	0.004	0.00	0.19	3.60	0.004	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. M = mean.

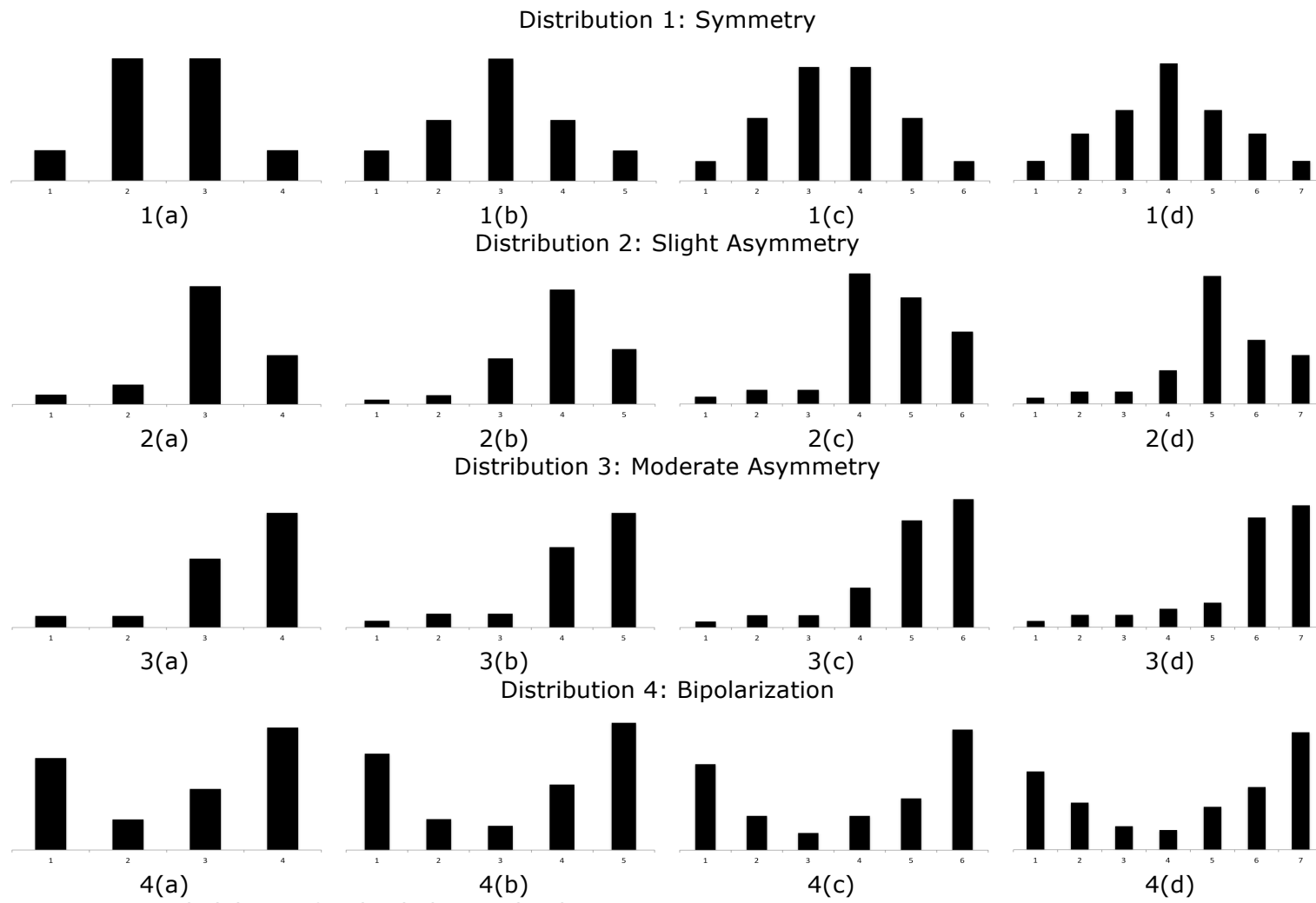


Figure 2. Response probabilities of ordinal observed indicators.

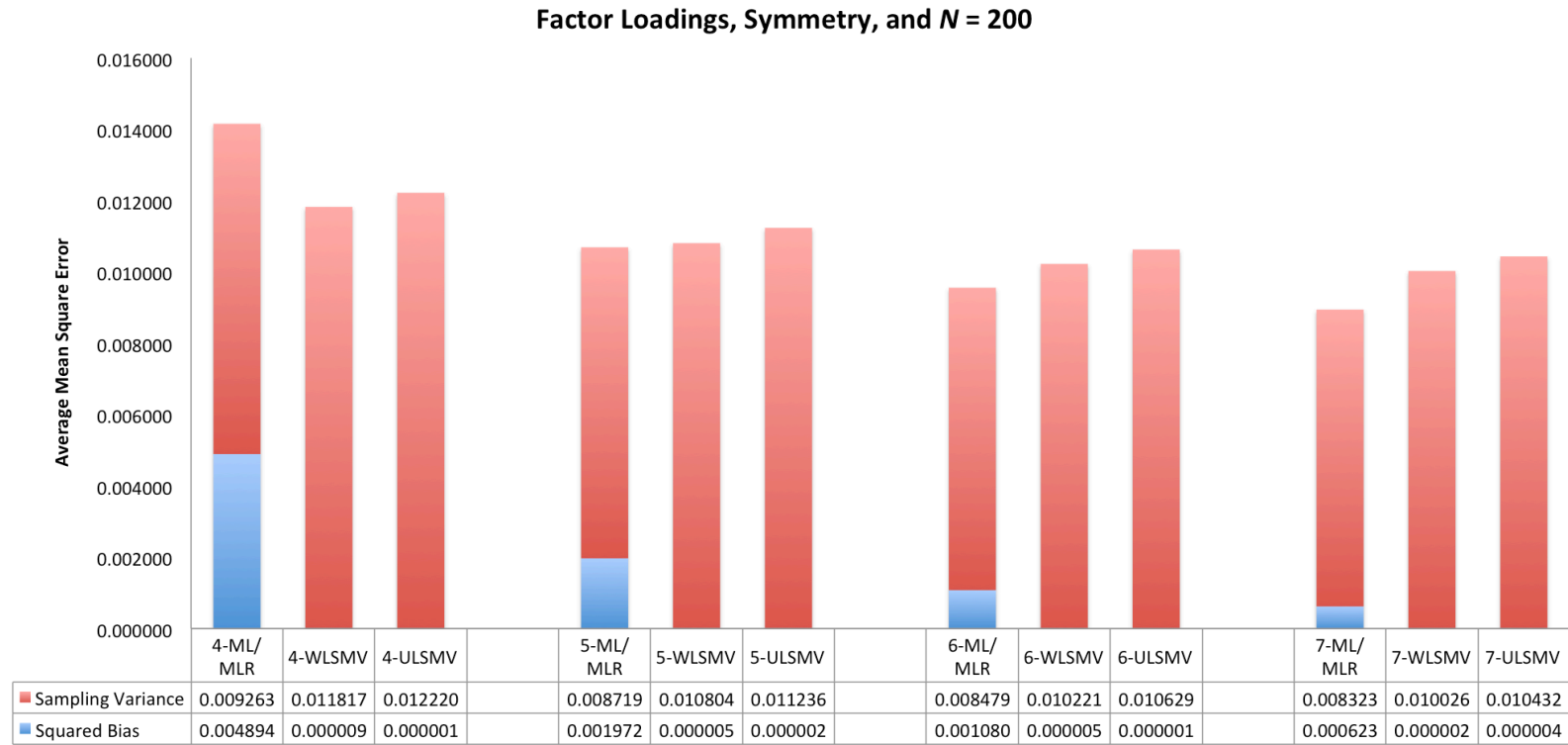


Figure 3. Average mean squared error for the factor loading estimates across the number of categories with symmetric data and the smallest sample size $N = 200$.

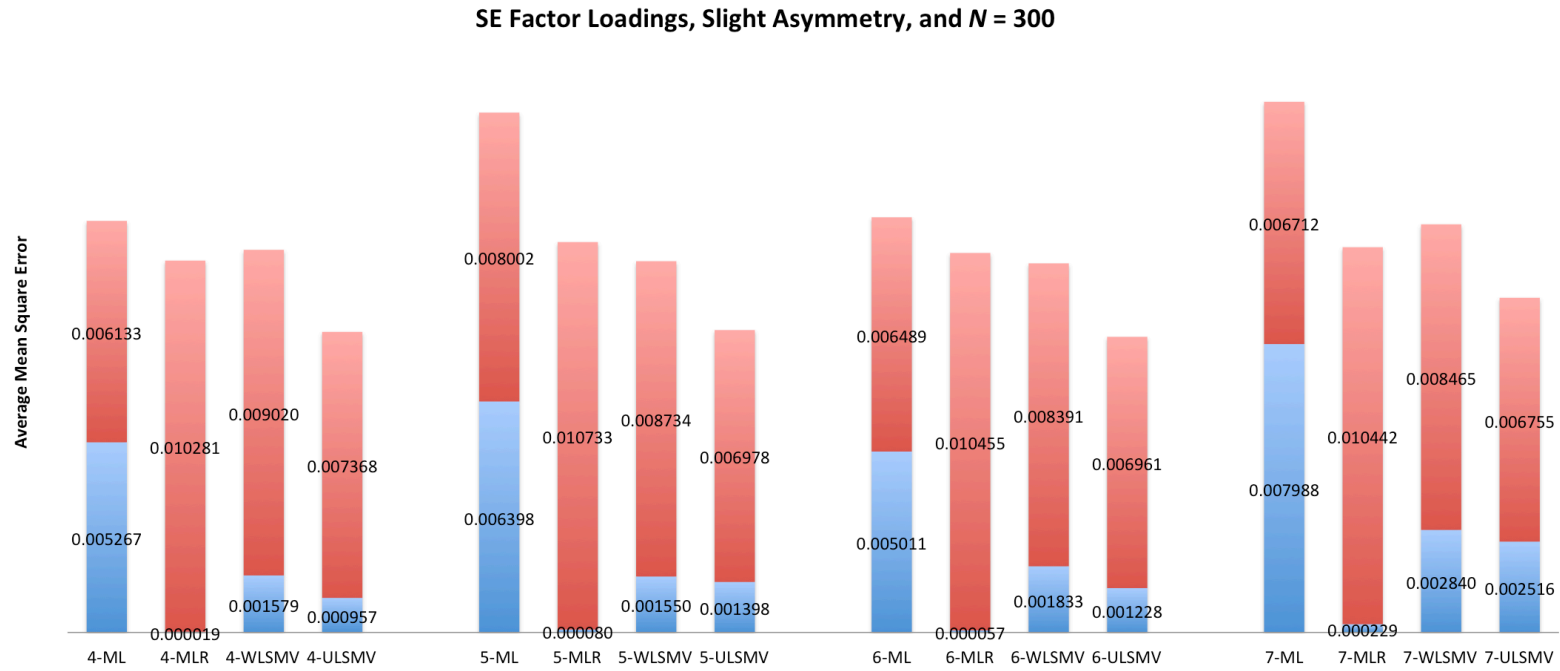


Figure 4. Average mean squared error for the standard error estimates of factor loadings across the number of categories with slightly asymmetric data and the sample size $N = 300$.

SE Factor Loadings, Slight Asymmetry, and $N = 1,000$

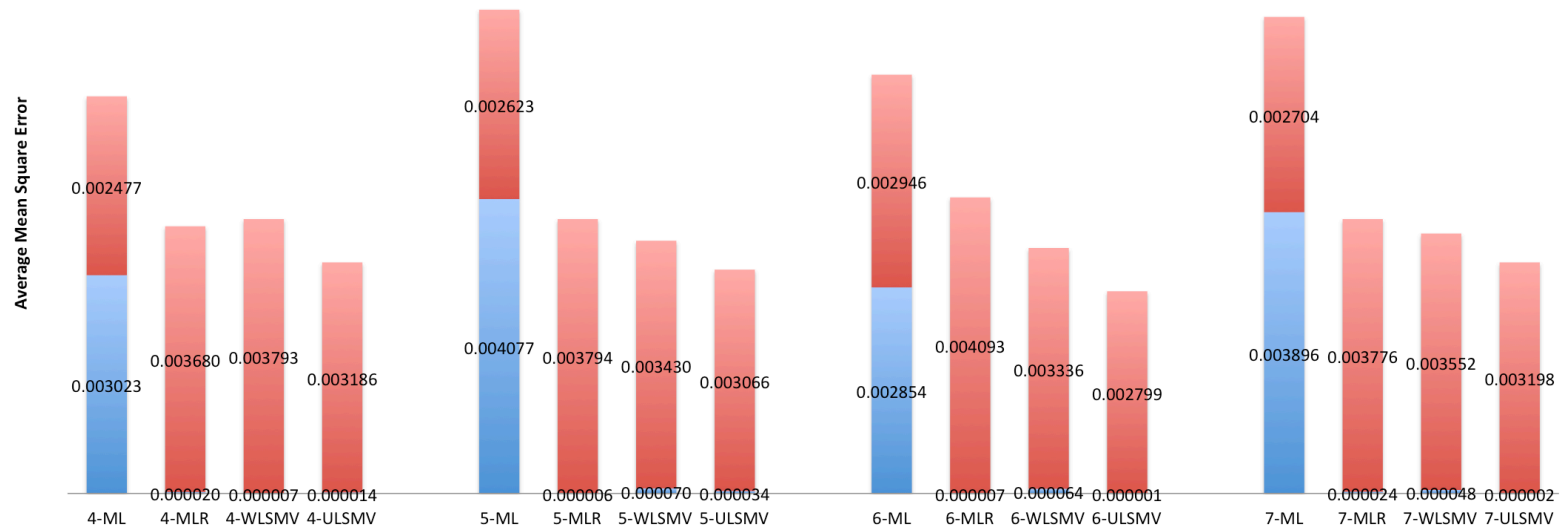


Figure 5. Average mean squared error for the standard error estimates of factor loadings across the number of categories with slightly asymmetric data and the sample size $N = 1,000$.

SE Structural Coefficients, Slight Asymmetry, and $N = 300$

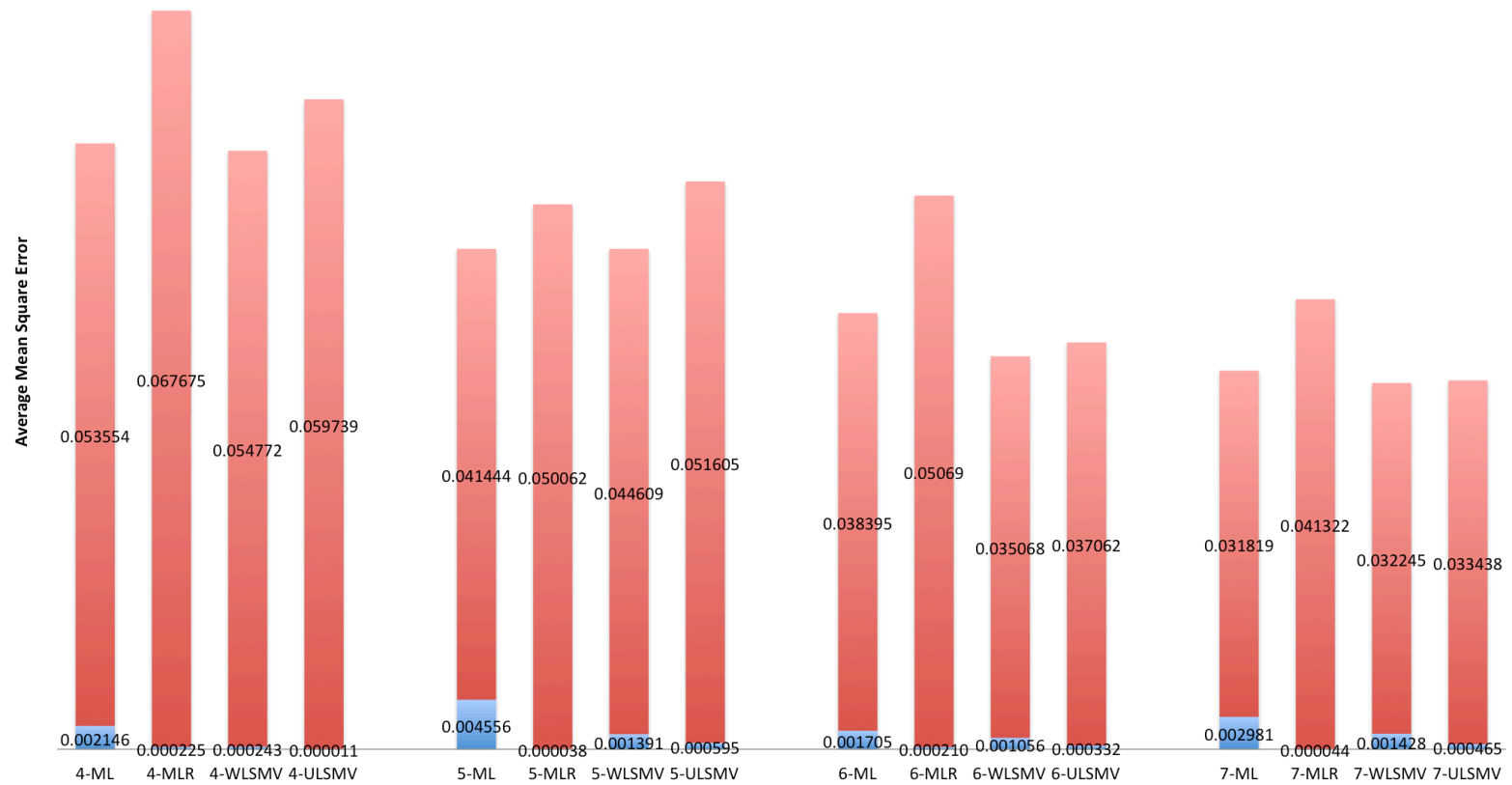
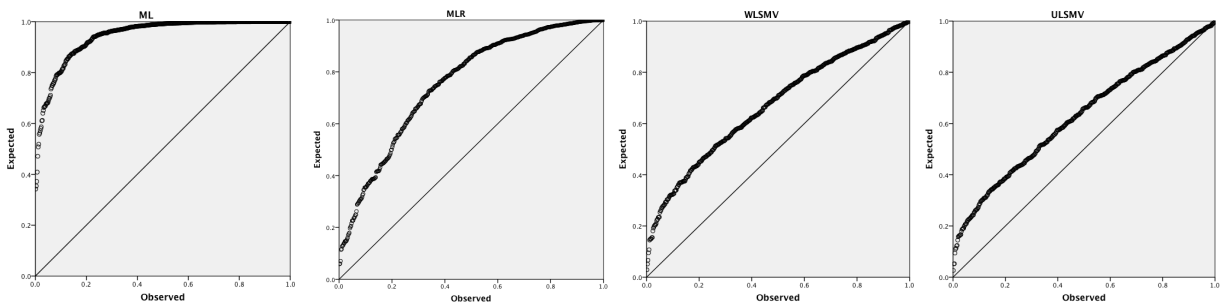
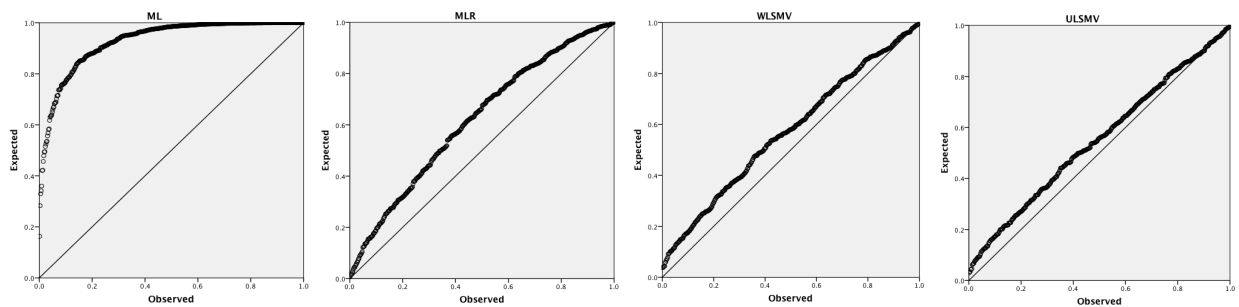


Figure 6. Average mean squared error for the standard error estimates of structural coefficients across the number of categories with slightly asymmetric data and the sample size $N = 300$.

Sample size $N = 200$



Sample size $N = 500$



Sample size $N = 1,000$

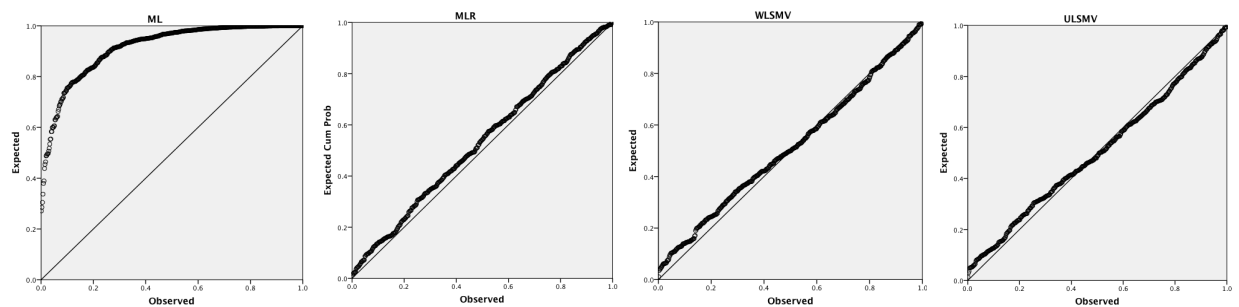
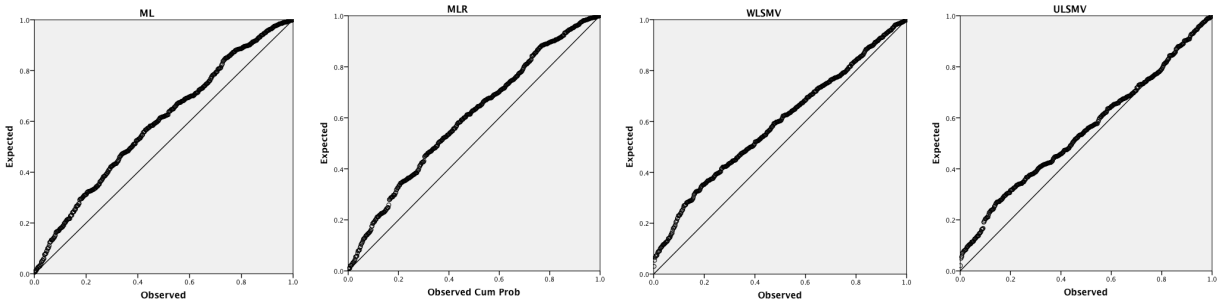
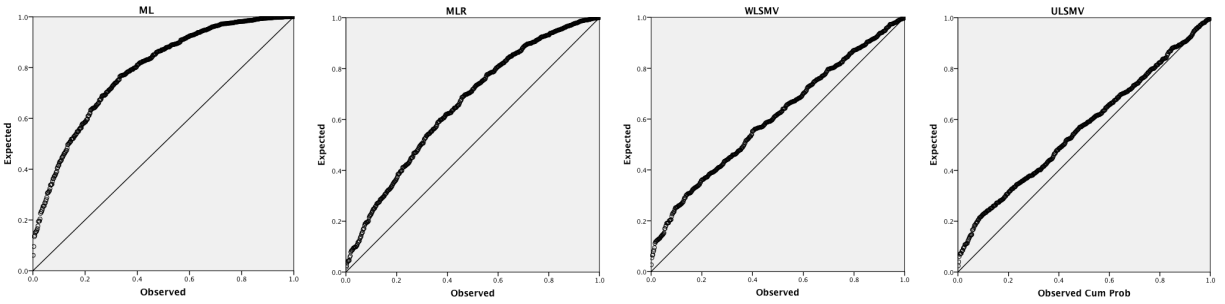


Figure 7. P-P plots for T_{ML} , T_{MLR} , T_{WLSMV} , and T_{ULSMV} (Moderate Asymmetry and 7-category)

Symmetry



Slight Asymmetry



Moderate Asymmetry

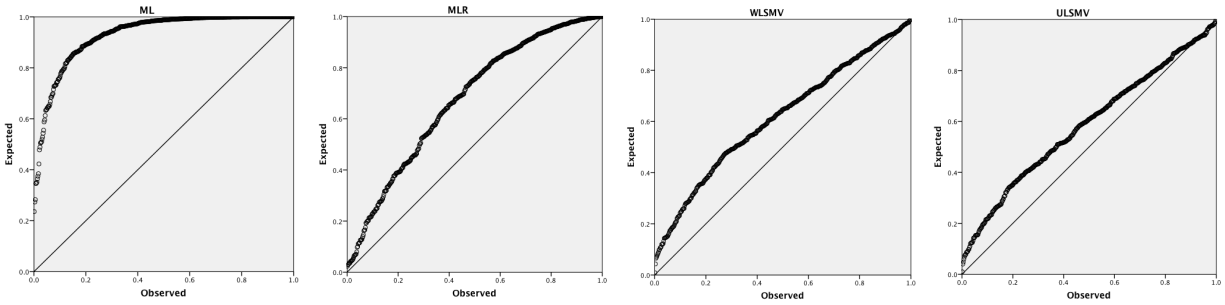


Figure 8. P-P plots for T_{ML} , T_{MLR} , T_{WLSMV} , and T_{ULSMV} ($N = 300$ and 7-category)

Appendix C

Technical Details

1. Robust correction to the chi-square statistic for WLSM

The mean-adjusted chi-square statistic can also be implemented in the D-WLS estimator (Muthén & Muthén, 2010):

$$T_{D-WLSM} = \frac{df}{\text{trace}(\tilde{\mathbf{U}}\tilde{\mathbf{V}})} T_{WLS}, \quad df = s - t, \quad (\text{A.1})$$

where $T_{WLS} = (N - 1) F_{WLS}(\hat{\underline{\theta}}, \underline{s})$, $\hat{\mathbf{V}}$ is the estimated asymptotic covariance matrix of \underline{s} , $\tilde{\mathbf{U}} = \mathbf{W}_D^{-1} - \mathbf{W}_D^{-1} \hat{\mathbf{A}} (\hat{\mathbf{A}}' \mathbf{W}_D^{-1} \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}' \mathbf{W}_D^{-1}$, s = the number of unique elements in \underline{s} , and t = the number of independent model parameters.

2. Robust corrections to the standard error for MLM or MLMV

A consistent estimator of the asymptotic covariance matrix of the parameter estimates $\hat{\underline{\theta}}$ for MLM or MLMV can be expressed as (Muthén & Muthén, 2010; Satorra & Bentler, 1994):

$$\widehat{\text{aCov}}(\hat{\underline{\theta}})_{\text{MLM or MLMV}} = N^{-1} (\tilde{\mathbf{A}}' \mathbf{W}_{NT} \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}' \mathbf{W}_{NT} \tilde{\mathbf{V}} \mathbf{W}_{NT} \tilde{\mathbf{A}} (\tilde{\mathbf{A}}' \mathbf{W}_{NT} \tilde{\mathbf{A}})^{-1}, \quad (\text{A.2})$$

and

$$\mathbf{W}_{NT} = \frac{1}{2} N \{ \mathbf{D}' [\boldsymbol{\Sigma}^{-1}(\hat{\underline{\theta}}) \otimes \boldsymbol{\Sigma}^{-1}(\hat{\underline{\theta}})] \mathbf{D} \}, \quad (\text{A.3})$$

where $\tilde{\mathbf{A}} = \frac{\partial \underline{g}(\hat{\underline{\theta}})}{\partial \underline{\theta}}$ is the matrix of model first derivatives evaluated at the parameter estimates $\hat{\underline{\theta}}$, \mathbf{W}_{NT} is the normal-theory weight matrix (see Browne, 1974), $\tilde{\mathbf{V}}$ is the estimated asymptotic covariance matrix of \mathbf{S} , \mathbf{D} is the “duplication” matrix (see Magnus & Neudecker, 1986) and \otimes denotes a Kronecker product.

3. Robust corrections to the chi-square statistic for MLM and MLMV

The mean-adjusted chi-square statistic is available in the robust ML estimator (also known as the Satorra-Bentler scaled chi-square statistic: Satorra & Bentler, 1994; Muthén, 1993):

$$T_{MLM} = T_{SB} = \frac{df}{\text{trace}(\tilde{\mathbf{U}}\tilde{\mathbf{V}})} T_{ML}, \quad df = s - t, \quad (\text{A.4})$$

where $T_{ML} = (N - 1) F_{ML}(\hat{\boldsymbol{\theta}}, \mathbf{S})$, $\tilde{\mathbf{V}}$ is the estimated asymptotic covariance matrix of \mathbf{S} , $\tilde{\mathbf{U}} = \mathbf{W}_{NT} - \mathbf{W}_{NT}\tilde{\mathbf{A}}(\tilde{\mathbf{A}}'\mathbf{W}_{NT}\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}'\mathbf{W}_{NT}$, s = the number of unique elements in \mathbf{S} , and t = the number of total model parameters.

Alternatively, the mean- and variance-adjusted chi-square statistic can also be implemented in the robust ML estimator (Asparouhov & Muthén, 2010):

$$T_{MLMV} = \sqrt{\frac{df}{\text{trace}(\tilde{\mathbf{U}}\tilde{\mathbf{V}}\tilde{\mathbf{U}}\tilde{\mathbf{V}})}} T_{ML} + df - \sqrt{\frac{df [\text{trace}(\tilde{\mathbf{U}}\tilde{\mathbf{V}})]^2}{\text{trace}(\tilde{\mathbf{U}}\tilde{\mathbf{V}}\tilde{\mathbf{U}}\tilde{\mathbf{V}})}}, \quad df = s - t, \quad (\text{A.5})$$

where $T_{ML} = (N - 1) F_{ML}(\hat{\boldsymbol{\theta}}, \mathbf{S})$, $\tilde{\mathbf{V}}$ is the estimated asymptotic covariance matrix of \mathbf{S} , $\tilde{\mathbf{U}} = \mathbf{W}_{NT} - \mathbf{W}_{NT}\tilde{\mathbf{A}}(\tilde{\mathbf{A}}'\mathbf{W}_{NT}\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}'\mathbf{W}_{NT}$, s = the number of unique elements in \mathbf{S} , and t = the number of total model parameters.

Appendix D
Mplus Code for Data Generation and Analysis

1. *Mplus* code for data generation

TITLE: Data generation in an SR model with symmetry data, 4 categories, and $N = 200$

MONTECARLO:

```
NAMES = y1-y20;  
NOBSERVATIONS = 200; ! sample size  $N = 200$   
NREPS = 500; ! number of replications = 500  
SEED = 4533;  
REPSAVE = ALL;  
SAVE = ex1_rep*.dat;  
! The SAVE option is used to name the files to which the 500 datasets were written.  
! The asterisk * was replaced by the replication number. A file, ex1_replist.dat, was also  
! produced. The file contains the file names of the 500 generated datasets.  
GENERATE = y1-y20 (3); ! number of thresholds = 3  
CATEGORICAL = y1-y20;
```

MODEL POPULATION:

```
F1 BY y1*.8 y2*.7 y3*.6 y4*.5; ! standardized factor loadings  
F1@1; ! latent variance
```

```
[y1$1*-1.282 y1$2*0 y1$3*1.282]; ! pre-specified thresholds  
[y2$1*-1.282 y2$2*0 y2$3*1.282];  
[y3$1*-1.282 y3$2*0 y3$3*1.282];  
[y4$1*-1.282 y4$2*0 y4$3*1.282];
```

```
y1*.36 y2*.51 y3*.64 y4*.75; ! residual variances
```

```
F2 BY y5*.8 y6*.7 y7*.6 y8*.5;  
F2@1; ! latent variance
```

```
[y5$1*-1.282 y5$2*0 y5$3*1.282];  
[y6$1*-1.282 y6$2*0 y6$3*1.282];  
[y7$1*-1.282 y7$2*0 y7$3*1.282];  
[y8$1*-1.282 y8$2*0 y8$3*1.282];
```

```
y5*.36 y6*.51 y7*.64 y8*.75;
```

```
F3 BY y9*.8 y10*.7 y11*.6 y12*.5;  
F3@.336; ! residual variance of latent variable
```

```
[y9$1*-1.282 y9$2*0 y9$3*1.282];  
[y10$1*-1.282 y10$2*0 y10$3*1.282];  
[y11$1*-1.282 y11$2*0 y11$3*1.282];  
[y12$1*-1.282 y12$2*0 y12$3*1.282];
```

y9*.36 y10*.51 y11*.64 y12*.75;

F4 BY y13*.8 y14*.7 y15*.6 y16*.5;
F4@.4364; ! residual variance of latent variable

[y13\$1*-1.282 y13\$2*0 y13\$3*1.282];
[y14\$1*-1.282 y14\$2*0 y14\$3*1.282];
[y15\$1*-1.282 y15\$2*0 y15\$3*1.282];
[y16\$1*-1.282 y16\$2*0 y16\$3*1.282];

y13*.36 y14*.51 y15*.64 y16*.75;

F5 BY y17*.8 y18*.7 y19*.6 y20*.5;
F5@.3798; ! residual variance of latent variable

[y17\$1*-1.282 y17\$2*0 y17\$3*1.282];
[y18\$1*-1.282 y18\$2*0 y18\$3*1.282];
[y19\$1*-1.282 y19\$2*0 y19\$3*1.282];
[y20\$1*-1.282 y20\$2*0 y20\$3*1.282];

y17*.36 y18*.51 y19*.64 y20*.75;

F1 WITH F2*.3; ! inter-factor correlation

F3 ON F1*.4 F2*.6; ! gamma coefficients
F4 ON F1*.4 F2*.2;
F5 ON F1*.1 F2*.1;

F4 ON F3*.3; ! beta coefficients
F5 ON F3*.2 F4*.5;

MODEL:

F1 BY y1*.8 y2*.7 y3*.6 y4*.5;
F1@1;

[y1\$1*-1.282 y1\$2*0 y1\$3*1.282];
[y2\$1*-1.282 y2\$2*0 y2\$3*1.282];
[y3\$1*-1.282 y3\$2*0 y3\$3*1.282];
[y4\$1*-1.282 y4\$2*0 y4\$3*1.282];

F2 BY y5*.8 y6*.7 y7*.6 y8*.5;
F2@1;

[y5\$1*-1.282 y5\$2*0 y5\$3*1.282];
[y6\$1*-1.282 y6\$2*0 y6\$3*1.282];
[y7\$1*-1.282 y7\$2*0 y7\$3*1.282];
[y8\$1*-1.282 y8\$2*0 y8\$3*1.282];

```
F3 BY y9*.8 y10*.7 y11*.6 y12*.5;  
F3@1;
```

```
[y9$1*-1.282 y9$2*0 y9$3*1.282];  
[y10$1*-1.282 y10$2*0 y10$3*1.282];  
[y11$1*-1.282 y11$2*0 y11$3*1.282];  
[y12$1*-1.282 y12$2*0 y12$3*1.282];
```

```
F4 BY y13*.8 y14*.7 y15*.6 y16*.5;  
F4@1;
```

```
[y13$1*-1.282 y13$2*0 y13$3*1.282];  
[y14$1*-1.282 y14$2*0 y14$3*1.282];  
[y15$1*-1.282 y15$2*0 y15$3*1.282];  
[y16$1*-1.282 y16$2*0 y16$3*1.282];
```

```
F5 BY y17*.8 y18*.7 y19*.6 y20*.5;  
F5@1;
```

```
[y17$1*-1.282 y17$2*0 y17$3*1.282];  
[y18$1*-1.282 y18$2*0 y18$3*1.282];  
[y19$1*-1.282 y19$2*0 y19$3*1.282];  
[y20$1*-1.282 y20$2*0 y20$3*1.282];
```

```
F1 WITH F2*.3;
```

```
F3 ON F1*.4 F2*.6;  
F4 ON F1*.4 F2*.2;  
F5 ON F1*.1 F2*.1;
```

```
F4 ON F3*.3;  
F5 ON F3*.2 F4*.5;
```

```
OUTPUT: TECH9;
```

! The TECH9 option is used to request error messages related to convergence for each
! replication.

Notes: (1) This is an example *Mplus* code for ordinal indicators that have symmetric distributions and four categories in a sample size of $N = 200$. The number of thresholds, the pre-specified values of thresholds, and sample size (i.e., the NOBSERVATIONS option) can be correspondingly modified to target different experimental conditions. (2) See Chapter 12: Monte Carlo simulation studies of the *Mplus* User's Guide for further details about other commands and options. (3) The exclamation mark ! is used to make notes and comments but not read by *Mplus*.

2. *Mplus* code for data analysis using ML and MLR

TITLE: Data analysis in an SR model using ML

DATA: FILE=ex1_replist.dat;
! The FILE option is used to carry out data analysis for each replication.
! "ex1_replist.dat" contains the file names of the 500 generated datasets.
TYPE = MONTECARLO;

VARIABLE:
NAMES= y1-y20;

ANALYSIS:
ESTIMATOR = ML;
! One can replace ML by MLR to obtain robust maximum likelihood estimation.

MODEL:
F1 BY y1* y2-y4;
F1@1;

F2 BY y5* y6-y8;
F2@1;

F3 BY y9* y10-y12;
F3@1;

F4 BY y13* y14-y16;
F4@1;

F5 BY y17* y18-y20;
F5@1;

F1 WITH F2;

F3 ON F1 F2;
F4 ON F1 F2;
F5 ON F1 F2;

F4 ON F3;
F5 ON F3 F4;

OUTPUT: STDYX; ! The STDYX option is used to request standardized solutions.

SAVEDATA: RESULTS ARE <Name of Results File>; ! The SAVEDATA command is used to save estimation results obtained from the 500 replications.

3. *Mplus* code for data analysis using ULSMV and WLSMV

TITLE: Data analysis in an SR model using ULSMV

DATA: FILE=ex1_replist.dat;
 TYPE = MONTECARLO;

VARIABLE:
NAMES= y1-y20;
CATEGORICAL= y1-y20;

ANALYSIS:
ESTIMATOR = ULSMV;
! One can replace ULSMV by WLSMV to obtain robust weighted least squares estimation.

MODEL:
 F1 BY y1* y2-y4;
 F1@1;

 F2 BY y5* y6-y8;
 F2@1;

 F3 BY y9* y10-y12;
 F3@1;

 F4 BY y13* y14-y16;
 F4@1;

 F5 BY y17* y18-y20;
 F5@1;

 F1 WITH F2;

 F3 ON F1 F2;
 F4 ON F1 F2;
 F5 ON F1 F2;

 F4 ON F3;
 F5 ON F3 F4;

OUTPUT: STDYX;

SAVEDATA: RESULTS ARE <Name of Results File>;

Appendix E

Results for sample sizes of $N = 400$, 750, and 1,500 are presented below:

1. Tables E1–E3 display average relative bias (RB_A) and average mean squared error (MSE_A) of factor loadings and structural coefficients by number of categories and observed distributions for all three robust estimators.

Table E1. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 400$)

		ML/MLR				WLSMV				ULSMV			
		RB_A		MSE_A		RB_A		MSE_A		RB_A		MSE_A	
Dis.	Cat.	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC
sym	4	-6.97	-1.08	0.0094	0.3552	0.11	-0.76	0.0058	0.3705	-0.09	-0.93	0.0061	0.3751
	5	-4.49	-0.02	0.0063	0.3315	0.00	0.21	0.0053	0.3494	-0.18	0.04	0.0055	0.3577
	6	-3.22	-0.37	0.0053	0.2930	0.03	-0.24	0.0051	0.3095	-0.14	-0.39	0.0053	0.3136
	7	-2.48	-0.59	0.0047	0.2963	0.01	-0.37	0.0049	0.3112	-0.16	-0.50	0.0051	0.3172
slight	4	-9.95	-1.04	0.0155	0.4576	0.08	-0.48	0.0066	0.4513	-1.14	-0.58	0.0068	0.4452
	5	-6.88	-0.71	0.0100	0.4010	-0.01	0.12	0.0058	0.3897	-0.20	-0.02	0.0061	0.3904
	6	-5.92	-1.12	0.0086	0.3694	0.03	-0.52	0.0055	0.3361	-0.15	-0.70	0.0057	0.3396
	7	-5.37	-0.65	0.0080	0.3950	0.00	-0.01	0.0052	0.3479	-0.17	-0.17	0.0054	0.3509
mod	4	-11.67	-2.93	0.0212	0.5442	0.02	-0.07	0.0081	0.5969	-0.28	-0.29	0.0085	0.6034
	5	-9.03	-2.60	0.0149	0.4678	0.10	-0.54	0.0067	0.4472	-0.14	-0.81	0.0070	0.4511
	6	-8.62	-2.47	0.0142	0.4849	0.06	-0.21	0.0062	0.4317	-0.15	-0.40	0.0065	0.4308
	7	-8.60	-2.57	0.0143	0.4901	0.04	-0.17	0.0060	0.4147	-0.17	-0.36	0.0063	0.4139

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML/MLR = maximum likelihood/robust maximum likelihood, WLSMV =

robust weighted least squares, ULSMV = robust unweight least squares. FL represents factor loadings and SC is structural coefficients.

Table E2. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 750$)

		ML/MLR				WLSMV				ULSMV			
		RB_A		MSE_A		RB_A		MSE_A		RB_A		MSE_A	
Dis.	Cat.	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC
sym	4	-6.85	-0.30	0.0071	0.1879	0.17	-0.21	0.0031	0.1979	0.06	-0.26	0.0032	0.2011
	5	-4.31	-0.05	0.0041	0.1606	0.15	-0.01	0.0027	0.1732	0.04	-0.04	0.0028	0.1762
	6	-3.09	-0.39	0.0031	0.1572	0.12	-0.33	0.0026	0.1698	0.03	-0.35	0.0027	0.1732
	7	-2.34	-0.29	0.0027	0.1490	0.12	-0.23	0.0025	0.1618	0.03	-0.27	0.0026	0.1654
slight	4	-9.80	-1.34	0.0125	0.2267	0.18	-0.38	0.0034	0.2357	0.05	-0.44	0.0035	0.2406
	5	-6.71	-1.25	0.0072	0.1936	0.14	-0.30	0.0030	0.1940	0.03	-0.29	0.0031	0.1975
	6	-5.81	-1.14	0.0060	0.1832	0.13	-0.28	0.0028	0.1812	0.02	-0.33	0.0029	0.1860
	7	-5.23	-0.94	0.0053	0.1753	0.11	-0.21	0.0026	0.1689	0.01	-0.25	0.0028	0.1720
mod	4	-11.54	-3.17	0.0173	0.2488	0.11	-0.31	0.0042	0.2875	-0.05	-0.39	0.0044	0.2917
	5	-8.97	-2.61	0.0115	0.2277	0.10	-0.26	0.0035	0.2377	-0.02	-0.32	0.0036	0.2427
	6	-8.52	-2.70	0.0108	0.2186	0.10	-0.23	0.0033	0.2086	-0.02	-0.31	0.0034	0.2137
	7	-8.48	-2.52	0.0108	0.2177	0.12	-0.01	0.0031	0.2079	0.00	-0.06	0.0033	0.2115

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML/MLR = maximum likelihood/robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. FL represents factor loadings and SC is structural coefficients.

Table E3. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients ($N = 1,500$)

Dis.	Cat.	ML/MLR				WLSMV				ULSMV			
		RB_A		MSE_A		RB_A		MSE_A		RB_A		MSE_A	
		FL	SC	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC
sym	4	-6.88	-0.01	0.0059	0.0805	0.08	0.12	0.0015	0.0840	0.03	0.12	0.0016	0.0855
	5	-4.33	-0.18	0.0030	0.0781	0.07	0.03	0.0013	0.0820	0.03	0.03	0.0014	0.0839
	6	-3.08	-0.14	0.0020	0.0715	0.08	0.02	0.0013	0.0748	0.03	0.01	0.0014	0.0764
	7	-2.36	-0.28	0.0016	0.0692	0.06	-0.13	0.0012	0.0720	0.02	-0.12	0.0013	0.0737
slight	4	-9.80	-0.93	0.0111	0.1058	0.10	-0.05	0.0017	0.1042	0.04	-0.01	0.0017	0.1051
	5	-6.72	-0.78	0.0059	0.0918	0.08	0.07	0.0015	0.0902	0.03	0.08	0.0015	0.0924
	6	-5.81	-0.65	0.0047	0.0855	0.07	-0.03	0.0014	0.0800	0.03	-0.03	0.0015	0.0814
	7	-5.22	-0.70	0.0040	0.0869	0.08	0.05	0.0013	0.0790	0.03	0.07	0.0014	0.0806
mod	4	-11.57	-2.92	0.0155	0.1237	0.04	-0.19	0.0021	0.1309	-0.04	-0.20	0.0022	0.1333
	5	-8.98	-2.40	0.0099	0.1053	0.07	-0.12	0.0017	0.1012	0.00	-0.13	0.0018	0.1026
	6	-8.51	-2.53	0.0091	0.1075	0.06	-0.18	0.0016	0.0934	0.01	-0.16	0.0017	0.0944
	7	-8.48	-2.49	0.0091	0.1064	0.07	-0.10	0.0015	0.0881	0.02	-0.09	0.0016	0.0897

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML/MLR = maximum likelihood/robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. FL represents factor loadings and SC is structural coefficients.

2. The RB_A and MSE_A for standard errors of factor loadings and structural coefficients are presented in Tables E4–E6.

Table E4. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 400$)

		ML				MLR			
		RB_A		MSE_A		RB_A		MSE_A	
Dis.	Cat.	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}
sym	4	0.18	-0.06	0.0048	0.0201	0.46	1.03	0.0071	0.0227
	5	-0.21	-1.44	0.0052	0.0162	0.08	-0.31	0.0075	0.0189
	6	-0.95	-0.18	0.0054	0.0152	-0.55	0.80	0.0077	0.0177
	7	-0.80	-0.93	0.0052	0.0141	-0.40	0.02	0.0075	0.0164
slight	4	-5.45	-7.23	0.0077	0.0289	1.16	-1.61	0.0082	0.0290
	5	-6.39	-7.05	0.0094	0.0241	0.16	-1.25	0.0088	0.0243
	6	-6.32	-6.10	0.0089	0.0220	-0.37	-0.86	0.0081	0.0233
	7	-6.93	-8.47	0.0098	0.0254	0.27	-2.75	0.0081	0.0232
mod	4	-16.07	-14.70	0.0316	0.0514	-0.22	-1.44	0.0087	0.0385
	5	-14.28	-13.28	0.0264	0.0405	-0.17	-1.40	0.0086	0.0292
	6	-14.58	-14.59	0.0259	0.0429	0.20	-2.51	0.0085	0.0301
	7	-15.33	-14.83	0.0282	0.0428	0.10	-2.24	0.0084	0.0287

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table E4 (cont'd)

		WLSMV				ULSMV			
		RB _A		MSE _A		RB _A		MSE _A	
Dis.	Cat.	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}
sym	4	-2.21	-1.28	0.0072	0.0223	-1.41	-0.55	0.0060	0.0230
	5	-2.49	-2.50	0.0072	0.0198	-2.12	-2.27	0.0078	0.0220
	6	-3.66	-2.61	0.0079	0.0183	-3.52	-1.68	0.0067	0.0186
	7	-3.66	-3.36	0.0077	0.0175	-3.33	-2.57	0.0064	0.0176
slight	4	-2.19	-3.00	0.0078	0.0293	-1.86	-1.53	0.0064	0.0290
	5	-2.97	-3.17	0.0080	0.0227	-2.85	-2.29	0.0068	0.0228
	6	-3.63	-1.89	0.0079	0.0211	-3.14	-1.04	0.0066	0.0219
	7	-3.78	-4.78	0.0079	0.0206	-3.43	-3.82	0.0065	0.0203
mod	4	-3.76	-2.96	0.0098	0.0406	-3.75	-3.11	0.0082	0.0429
	5	-3.33	-2.05	0.0084	0.0301	-3.57	-1.97	0.0072	0.0295
	6	-3.15	-5.10	0.0085	0.0279	-3.72	-4.11	0.0072	0.0271
	7	-3.71	-5.04	0.0086	0.0243	-4.00	-4.12	0.0075	0.0239

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table E5. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 750$)

Dis.	Cat.	ML				MLR			
		RB_A		MSE_A		RB_A		MSE_A	
		SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}
sym	4	-0.03	-2.20	0.0027	0.0108	-0.19	-1.54	0.0039	0.0120
	5	1.23	-1.28	0.0036	0.0090	1.12	-0.60	0.0047	0.0102
	6	0.63	-1.63	0.0032	0.0086	0.69	-1.06	0.0044	0.0096
	7	0.59	-1.47	0.0034	0.0077	0.66	-0.88	0.0045	0.0087
slight	4	-5.11	-6.27	0.0053	0.0167	1.00	-0.85	0.0045	0.0161
	5	-5.21	-5.84	0.0061	0.0130	1.10	-0.14	0.0055	0.0124
	6	-4.38	-5.70	0.0048	0.0133	1.38	-0.58	0.0048	0.0128
	7	-5.32	-5.88	0.0060	0.0134	1.54	-0.24	0.0053	0.0126
mod	4	-14.02	-11.18	0.0222	0.0267	1.29	2.03	0.0049	0.0198
	5	-11.69	-11.75	0.0162	0.0250	2.02	-0.15	0.0051	0.0156
	6	-13.00	-11.59	0.0195	0.0256	1.40	0.69	0.0051	0.0170
	7	-13.91	-12.25	0.0220	0.0258	1.06	0.40	0.0047	0.0152

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table E5 (cont'd)

		WLSMV				ULSMV			
		RB _A		MSE _A		RB _A		MSE _A	
Dis.	Cat.	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}
sym	4	-1.07	-3.16	0.0044	0.0124	-0.53	-2.76	0.0041	0.0123
	5	0.63	-2.39	0.0045	0.0108	0.44	-1.96	0.0039	0.0108
	6	-0.35	-3.42	0.0045	0.0106	-0.26	-2.99	0.0039	0.0103
	7	-0.17	-3.37	0.0043	0.0098	-0.22	-2.97	0.0035	0.0098
slight	4	0.04	-3.53	0.0044	0.0153	0.43	-3.02	0.0039	0.0156
	5	-0.30	-2.49	0.0044	0.0117	0.02	-2.09	0.0037	0.0119
	6	-0.71	-3.36	0.0041	0.0115	-0.46	-3.07	0.0036	0.0117
	7	0.07	-3.39	0.0047	0.0109	-0.19	-2.72	0.0038	0.0108
mod	4	-1.01	-0.76	0.0050	0.0196	-0.78	-0.33	0.0041	0.0202
	5	-0.57	-2.68	0.0045	0.0149	-0.57	-2.38	0.0041	0.0151
	6	-1.13	-2.50	0.0049	0.0138	-1.33	-2.14	0.0043	0.0141
	7	-1.31	-3.18	0.0045	0.0127	-1.28	-2.73	0.0042	0.0128

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table E6. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients ($N = 1,500$)

Dis.	Cat.	ML				MLR			
		RB_A		MSE_A		RB_A		MSE_A	
		SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}
sym	4	1.10	1.69	0.0023	0.0048	0.68	2.18	0.0027	0.0057
	5	1.19	-0.41	0.0023	0.0041	0.88	0.15	0.0027	0.0048
	6	0.42	0.78	0.0017	0.0039	0.32	1.20	0.0023	0.0046
	7	0.94	0.52	0.0021	0.0038	0.86	1.00	0.0026	0.0047
slight	4	-5.12	-5.92	0.0046	0.0090	0.66	-0.56	0.0028	0.0069
	5	-4.71	-5.24	0.0045	0.0078	1.39	0.40	0.0035	0.0065
	6	-5.07	-4.50	0.0045	0.0062	0.42	0.59	0.0029	0.0054
	7	-5.19	-5.87	0.0049	0.0079	1.43	-0.33	0.0032	0.0055
mod	4	-14.52	-13.61	0.0230	0.0246	0.25	-0.92	0.0030	0.0083
	5	-12.40	-11.33	0.0173	0.0183	0.83	0.17	0.0032	0.0071
	6	-12.98	-12.37	0.0188	0.0207	1.02	-0.40	0.0031	0.0072
	7	-13.69	-12.78	0.0208	0.0218	0.92	-0.41	0.0033	0.0072

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table E6 (cont'd)

		WLSMV				ULSMV			
		RB _A		MSE _A		RB _A		MSE _A	
Dis.	Cat.	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}
sym	4	0.31	1.36	0.0026	0.0056	-0.09	1.54	0.0021	0.0058
	5	0.67	-0.60	0.0026	0.0051	0.60	-0.48	0.0025	0.0052
	6	-0.18	0.14	0.0024	0.0046	-0.68	0.40	0.0020	0.0048
	7	0.32	0.17	0.0025	0.0045	0.05	0.41	0.0023	0.0047
slight	4	0.65	-1.10	0.0030	0.0065	1.07	-0.62	0.0031	0.0065
	5	0.50	-0.35	0.0027	0.0064	0.30	-0.32	0.0024	0.0065
	6	-0.57	0.29	0.0024	0.0047	-0.81	0.63	0.0020	0.0049
	7	0.53	-0.90	0.0029	0.0049	0.52	-0.65	0.0028	0.0050
mod	4	-0.53	-1.07	0.0034	0.0083	-0.74	-0.99	0.0031	0.0084
	5	-0.76	0.21	0.0026	0.0070	-0.98	0.56	0.0023	0.0072
	6	-0.49	0.58	0.0026	0.0064	-0.32	0.93	0.0023	0.0066
	7	-0.26	1.12	0.0025	0.0061	-0.37	1.31	0.0023	0.0062

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

3. Tables E7–E9 present findings for chi-square test statistics and RMSEA with MLR, ULSMV, and WLSMV estimation.

Table E7. Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 400$)

		ML				MLR			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	2.55	8.00	0.008	0.00	2.78	8.40	0.008	0.00
	5	3.63	6.80	0.009	0.00	3.79	7.20	0.009	0.00
	6	2.74	6.80	0.009	0.00	2.91	7.40	0.009	0.00
	7	2.91	7.00	0.009	0.00	3.06	7.40	0.009	0.00
slight	4	12.08	30.60	0.016	0.00	5.55	14.40	0.011	0.00
	5	11.80	28.40	0.015	0.00	4.72	10.80	0.010	0.00
	6	11.39	27.00	0.015	0.00	4.82	11.40	0.010	0.00
	7	13.17	31.60	0.016	0.00	5.55	12.60	0.011	0.00
mod	4	25.15	63.53	0.024	0.00	5.37	14.23	0.010	0.00
	5	23.67	62.73	0.023	0.00	5.98	11.62	0.011	0.00
	6	25.48	64.80	0.024	0.00	6.32	13.00	0.011	0.00
	7	26.49	68.80	0.025	0.00	6.20	13.80	0.011	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. M = mean.

Table E7 (cont'd)

		WLSMV				ULSMV			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	1.30	5.20	0.007	0.00	0.12	3.20	0.006	0.00
	5	2.28	3.80	0.008	0.00	1.15	2.81	0.007	0.00
	6	2.63	5.40	0.008	0.00	1.42	4.80	0.007	0.00
	7	2.97	5.80	0.008	0.00	1.84	4.80	0.008	0.00
slight	4	2.91	6.20	0.008	0.00	1.69	4.20	0.007	0.00
	5	2.50	5.40	0.008	0.00	1.31	3.80	0.007	0.00
	6	2.67	5.60	0.008	0.00	1.52	4.20	0.007	0.00
	7	3.86	5.00	0.009	0.00	2.58	3.40	0.008	0.00
mod	4	2.42	4.02	0.008	0.00	1.43	3.81	0.007	0.00
	5	3.13	6.43	0.008	0.00	2.13	5.00	0.008	0.00
	6	3.52	7.80	0.009	0.00	2.54	6.40	0.008	0.00
	7	4.20	6.80	0.009	0.00	3.25	5.40	0.009	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. M = mean.

Table E8. Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 750$)

		ML				MLR			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	2.18	9.60	0.006	0.00	2.12	9.60	0.006	0.00
	5	1.95	6.60	0.006	0.00	1.80	6.40	0.006	0.00
	6	2.11	9.80	0.006	0.00	2.04	9.80	0.006	0.00
	7	1.85	8.20	0.006	0.00	1.75	8.00	0.006	0.00
slight	4	9.37	22.20	0.010	0.00	1.87	5.80	0.006	0.00
	5	10.98	25.20	0.011	0.00	2.81	8.80	0.006	0.00
	6	10.55	24.00	0.010	0.00	2.86	8.80	0.006	0.00
	7	10.71	26.80	0.011	0.00	2.11	6.80	0.006	0.00
mod	4	24.44	63.60	0.017	0.00	3.14	9.20	0.007	0.00
	5	22.23	57.20	0.016	0.00	3.11	8.60	0.006	0.00
	6	23.67	63.40	0.017	0.00	3.07	8.60	0.007	0.00
	7	24.38	66.60	0.017	0.00	2.77	8.00	0.006	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. M = mean.

Table E8 (cont'd)

		WLSMV				ULSMV			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	1.11	5.40	0.005	0.00	0.48	4.80	0.005	0.00
	5	0.89	4.60	0.005	0.00	0.27	3.41	0.005	0.00
	6	1.60	5.60	0.005	0.00	0.99	5.20	0.005	0.00
	7	1.66	6.40	0.006	0.00	1.02	5.40	0.005	0.00
slight	4	1.07	5.20	0.005	0.00	0.34	3.80	0.005	0.00
	5	1.75	3.80	0.006	0.00	1.13	3.80	0.005	0.00
	6	1.85	5.80	0.006	0.00	1.23	5.00	0.005	0.00
	7	1.41	5.40	0.005	0.00	0.72	4.20	0.005	0.00
mod	4	1.32	4.60	0.005	0.00	0.72	3.80	0.005	0.00
	5	2.00	6.80	0.006	0.00	1.45	6.00	0.005	0.00
	6	1.94	6.40	0.006	0.00	1.39	5.80	0.005	0.00
	7	2.04	6.20	0.006	0.00	1.48	5.20	0.005	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. M = mean.

Table E9. Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA ($N = 1,500$)

		ML				MLR			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	0.90	6.20	0.004	0.00	0.69	6.20	0.004	0.00
	5	1.02	5.80	0.004	0.00	0.71	5.40	0.004	0.00
	6	1.15	6.80	0.004	0.00	0.93	6.60	0.004	0.00
	7	1.11	6.60	0.004	0.00	0.84	6.00	0.004	0.00
slight	4	9.85	21.80	0.007	0.00	1.66	6.60	0.004	0.00
	5	10.34	21.20	0.007	0.00	1.53	8.00	0.004	0.00
	6	9.74	21.60	0.007	0.00	1.44	7.40	0.004	0.00
	7	10.91	23.20	0.008	0.00	1.68	8.00	0.004	0.00
mod	4	24.03	63.60	0.012	0.00	1.87	7.40	0.004	0.00
	5	21.89	59.00	0.011	0.00	1.95	7.40	0.004	0.00
	6	23.41	59.60	0.012	0.00	1.98	7.20	0.004	0.00
	7	24.69	62.80	0.012	0.00	2.20	7.60	0.004	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. ML = maximum likelihood, MLR = robust maximum likelihood. M = mean.

Table E9 (cont'd)

		WLSMV				ULSMV			
		Chi-square		RMSEA		Chi-square		RMSEA	
Dis.	Cat.	bias	%	M	%	bias	%	M	%
sym	4	0.57	5.00	0.004	0.00	0.36	4.21	0.004	0.00
	5	0.40	5.40	0.004	0.00	0.19	4.20	0.004	0.00
	6	0.98	6.40	0.004	0.00	0.79	6.40	0.004	0.00
	7	0.86	6.20	0.004	0.00	0.61	5.20	0.004	0.00
slight	4	1.04	6.00	0.004	0.00	0.75	5.84	0.003	0.00
	5	0.86	5.60	0.004	0.00	0.67	5.60	0.004	0.00
	6	0.96	5.80	0.004	0.00	0.76	5.40	0.003	0.00
	7	0.81	6.20	0.004	0.00	0.57	5.80	0.004	0.00
mod	4	0.67	5.20	0.004	0.00	0.49	5.42	0.004	0.00
	5	1.22	6.00	0.004	0.00	1.08	5.00	0.004	0.00
	6	1.05	7.60	0.004	0.00	0.83	7.00	0.004	0.00
	7	1.33	7.60	0.004	0.00	1.16	7.60	0.004	0.00

Note. Dis. = distribution type and Cat. = number of categories. sym = symmetric distribution, slight = slightly asymmetric distribution, mod = moderately asymmetric distribution. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. M = mean.

Appendix F

Results for bipolarization data are presented below:

1. Table F1 displays average relative bias (RB_A) and average mean squared error (MSE_A) of factor loadings and structural coefficients by number of categories and sample sizes for all three robust estimators.

Table F1. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Factor Loadings and Structural Coefficients with Bipolarization Distribution

		ML/MLR				WLSMV				ULSMV			
		RB_A		MSE_A		RB_A		MSE_A		RB_A		MSE_A	
	Cat	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC	FL	SC
$N = 200$	4	-10.36	-2.09	0.0219	1.2664	0.20	-0.67	0.0140	1.3447	-0.28	-0.80	0.0145	1.2803
	5	-9.89	-2.36	0.0208	1.2004	0.17	-0.58	0.0131	1.2148	-0.29	-0.79	0.0136	1.1765
	6	-9.04	-2.15	0.0189	1.1454	0.14	-0.57	0.0127	1.0955	-0.29	-0.77	0.0132	1.0687
	7	-7.74	-1.55	0.0163	0.9636	0.15	-0.10	0.0117	0.9089	-0.26	-0.26	0.0121	0.8856
$N = 300$	4	-10.36	-2.08	0.0180	0.6648	0.08	-0.29	0.0089	0.6808	-0.03	0.18	0.0079	0.4693
	5	-9.82	-2.18	0.0168	0.6092	0.08	-0.22	0.0085	0.6339	-0.21	-0.42	0.0089	0.6274
	6	-8.96	-1.93	0.0150	0.5489	0.10	-0.02	0.0081	0.5497	-0.18	-0.14	0.0084	0.5469
	7	-7.69	-1.60	0.0126	0.5139	0.08	0.19	0.0075	0.5025	-0.18	0.09	0.0078	0.4991
$N = 400$	4	-10.33	-1.40	0.0163	0.4485	0.08	-0.06	0.0069	0.4910	-0.16	-0.19	0.0072	0.4941
	5	-9.81	-1.52	0.0152	0.4281	0.06	-0.28	0.0066	0.4641	-0.17	-0.40	0.0069	0.4656
	6	-8.98	-1.67	0.0135	0.3999	0.02	-0.42	0.0063	0.4311	-0.19	-0.57	0.0066	0.4320
	7	-7.70	-1.50	0.0111	0.3795	0.04	-0.35	0.0059	0.3845	-0.16	-0.50	0.0061	0.3863

Table F1 (cont'd)

<i>N</i> = 500	4	-10.34	-2.20	0.0151	0.3299	0.04	-0.96	0.0055	0.3649	-0.16	-1.06	0.0057	0.3698
	5	-9.83	-2.33	0.0141	0.3161	0.03	-0.93	0.0052	0.3445	-0.15	-1.01	0.0054	0.3503
	6	-8.97	-2.17	0.0123	0.3082	0.04	-0.77	0.0049	0.3198	-0.13	-0.84	0.0051	0.3249
	7	-7.69	-2.03	0.0100	0.2864	0.06	-0.71	0.0045	0.2898	-0.11	-0.80	0.0047	0.2955
<i>N</i> = 750	4	-10.21	-1.57	0.0133	0.2309	0.16	-0.12	0.0036	0.2558	0.03	-0.10	0.0038	0.2588
	5	-9.69	-1.45	0.0122	0.2207	0.15	-0.02	0.0034	0.2395	0.03	-0.03	0.0035	0.2430
	6	-8.84	-1.44	0.0106	0.2117	0.14	-0.12	0.0032	0.2252	0.02	-0.14	0.0034	0.2298
	7	-7.58	-1.17	0.0084	0.1992	0.11	-0.04	0.0030	0.2051	0.00	-0.05	0.0031	0.2097
<i>N</i> = 1,000	4	-10.21	-1.51	0.0126	0.1629	0.12	-0.26	0.0027	0.1752	0.02	-0.32	0.0028	0.1758
	5	-9.70	-1.35	0.0116	0.1622	0.12	-0.19	0.0025	0.1689	0.03	-0.27	0.0026	0.1697
	6	-8.84	-1.42	0.0099	0.1521	0.13	-0.22	0.0024	0.1566	0.04	-0.27	0.0025	0.1579
	7	-7.59	-1.21	0.0078	0.1475	0.09	-0.16	0.0022	0.1438	0.01	-0.22	0.0023	0.1456
<i>N</i> = 1,500	4	-10.26	-1.39	0.0120	0.0955	0.07	-0.09	0.0018	0.1033	0.00	-0.12	0.0019	0.1052
	5	-9.74	-1.40	0.0110	0.0939	0.07	-0.03	0.0017	0.0977	0.02	-0.06	0.0018	0.0996
	6	-8.90	-1.34	0.0094	0.0912	0.06	-0.05	0.0016	0.0939	0.01	-0.07	0.0017	0.0956
	7	-7.62	-1.22	0.0072	0.0876	0.06	-0.03	0.0015	0.0885	0.01	-0.04	0.0016	0.0902

Note. Cat. = number of categories. ML = maximum likelihood, MLR = robust maximum likelihood, WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. FL represents factor loadings and SC is structural coefficients.

2. The RB_A and MSE_A for standard errors of factor loadings and structural coefficients are presented in Table F2.

Table F2. The Average Relative Bias (RB_A) and Average Root Mean Squared Error (MSE_A) for Standard Errors (SE) of Factor Loadings and Structural Coefficients with Bipolarization Distribution

		ML				MLR			
		RB_A		MSE_A		RB_A		MSE_A	
	Cat.	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}	SE_{FL}	SE_{SC}
$N = 200$	4	-5.94	-9.17	0.0123	0.1581	0.53	-2.57	0.0124	0.1957
	5	-5.80	-8.76	0.0122	0.2144	0.85	-1.99	0.0124	0.2638
	6	-5.88	-9.01	0.0123	0.1749	0.54	-2.59	0.0124	0.2026
	7	-5.76	-8.33	0.0124	0.0712	0.36	-2.24	0.0125	0.0817
$N = 300$	4	-4.55	-6.27	0.0087	0.0695	0.94	-0.65	0.0082	0.0735
	5	-4.72	-5.51	0.0088	0.0544	0.94	0.28	0.0081	0.0592
	6	-4.71	-4.16	0.0085	0.0449	0.76	1.56	0.0077	0.0501
	7	-4.62	-4.50	0.0088	0.0392	0.62	0.96	0.0082	0.0446
$N = 400$	4	-4.99	-5.21	0.0070	0.0282	-0.17	0.17	0.0057	0.0293
	5	-5.13	-5.18	0.0072	0.0271	-0.13	0.23	0.0057	0.0280
	6	-5.08	-4.39	0.0070	0.0243	-0.20	0.97	0.0055	0.0261
	7	-4.98	-4.37	0.0071	0.0216	-0.32	0.66	0.0058	0.0227

Table F2 (cont'd)

<i>N</i> = 500	4	-4.83	-4.86	0.0061	0.0198	-0.22	0.30	0.0047	0.0204
	5	-5.10	-4.49	0.0063	0.0188	-0.33	0.77	0.0045	0.0198
	6	-4.69	-4.84	0.0060	0.0179	-0.01	0.23	0.0047	0.0182
	7	-4.21	-4.10	0.0056	0.0157	0.33	0.73	0.0048	0.0166
<i>N</i> = 750	4	-3.99	-5.13	0.0044	0.0151	0.21	-0.43	0.0033	0.0147
	5	-3.72	-5.33	0.0044	0.0146	0.66	-0.57	0.0036	0.0140
	6	-3.58	-5.26	0.0042	0.0138	0.72	-0.64	0.0035	0.0131
	7	-3.37	-4.56	0.0040	0.0129	0.78	-0.17	0.0034	0.0128
<i>N</i> = 1,000	4	-3.50	-5.96	0.0033	0.0109	0.46	-1.44	0.0025	0.0086
	5	-3.72	-6.50	0.0038	0.0116	0.41	-1.89	0.0029	0.0088
	6	-3.14	-5.77	0.0033	0.0102	0.94	-1.25	0.0028	0.0081
	7	-3.29	-5.89	0.0034	0.0099	0.62	-1.62	0.0028	0.0076
<i>N</i> = 1,500	4	-4.56	-2.29	0.0042	0.0060	-0.83	2.23	0.0025	0.0068
	5	-4.60	-2.41	0.0044	0.0060	-0.71	2.21	0.0026	0.0068
	6	-4.28	-2.73	0.0040	0.0056	-0.45	1.75	0.0024	0.0059
	7	-4.33	-2.96	0.0039	0.0060	-0.63	1.28	0.0023	0.0061

Note. Cat. = number of categories. ML = maximum likelihood, MLR = robust maximum likelihood. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

Table F2 (cont'd)

		WLSMV				ULSMV			
		RB _A		MSE _A		RB _A		MSE _A	
	Cat.	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}	SE _{FL}	SE _{SC}
<i>N</i> = 200	4	-3.90	-5.10	0.0138	0.1609	-2.72	-3.71	0.0106	0.1311
	5	-3.31	-5.41	0.0138	0.1372	-2.30	-4.39	0.0108	0.1120
	6	-4.05	-5.02	0.0141	0.1114	-2.90	-3.91	0.0112	0.0987
	7	-4.10	-4.81	0.0139	0.0715	-3.10	-3.36	0.0111	0.0692
<i>N</i> = 300	4	-0.99	-0.97	0.0095	0.0609	-1.97	1.41	0.0080	0.0456
	5	-1.43	-0.86	0.0093	0.0576	-1.51	0.54	0.0074	0.0601
	6	-1.47	0.43	0.0089	0.0435	-1.22	1.69	0.0071	0.0459
	7	-1.83	-0.70	0.0092	0.0390	-1.50	0.53	0.0073	0.0419
<i>N</i> = 400	4	-1.90	-1.07	0.0068	0.0285	-1.67	-0.34	0.0054	0.0284
	5	-1.92	-2.32	0.0068	0.0273	-1.62	-1.52	0.0056	0.0274
	6	-2.13	-2.19	0.0069	0.0267	-1.87	-1.27	0.0057	0.0265
	7	-2.40	-1.57	0.0069	0.0207	-2.38	-0.71	0.0057	0.0206

Table F2 (cont'd)

<i>N</i> = 500	4	-0.93	-0.86	0.0063	0.0227	-0.55	-0.27	0.0052	0.0237
	5	-1.14	-0.98	0.0059	0.0208	-0.77	-0.39	0.0049	0.0223
	6	-0.88	-0.70	0.0059	0.0192	-0.66	-0.04	0.0049	0.0202
	7	-0.55	-0.56	0.0061	0.0163	-0.36	0.01	0.0051	0.0174
<i>N</i> = 750	4	-0.66	-1.95	0.0040	0.0155	-0.31	-1.67	0.0036	0.0157
	5	-0.16	-2.76	0.0041	0.0146	0.32	-2.54	0.0037	0.0146
	6	-0.02	-2.93	0.0040	0.0134	0.24	-2.79	0.0036	0.0137
	7	0.25	-2.44	0.0039	0.0130	0.35	-2.18	0.0035	0.0132
<i>N</i> = 1,000	4	-0.14	-1.79	0.0030	0.0092	-0.16	-1.21	0.0026	0.0093
	5	0.04	-1.89	0.0033	0.0090	0.15	-1.33	0.0027	0.0091
	6	0.32	-1.87	0.0031	0.0081	0.12	-1.43	0.0026	0.0083
	7	0.40	-1.60	0.0031	0.0073	-0.05	-1.20	0.0027	0.0075
<i>N</i> = 1,500	4	-0.80	2.15	0.0026	0.0072	-1.04	2.19	0.0024	0.0074
	5	-0.42	1.92	0.0025	0.0067	-0.40	1.99	0.0024	0.0069
	6	0.08	1.63	0.0026	0.0063	-0.19	1.79	0.0024	0.0066
	7	0.06	0.38	0.0025	0.0061	-0.01	0.55	0.0023	0.0064

Note. Cat. = number of categories. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. SE_{FL} represents standard errors of factor loadings and SE_{SC} is standard errors of structural coefficients.

3. Table F3 presents findings for chi-square test statistics and RMSEA with ML, MLR, ULSMV, and WLSMV estimation.

Table F3. Bias and Rejection Rates of Chi-Square Statistics as well as Means and Rejection Rates of RMSEA with Bipolarization Distribution

		ML				MLR			
		Chi-square		RMSEA		Chi-square		RMSEA	
	Cat.	bias	%	M	%	bias	%	bias	%
<i>N</i> = 200	4	11.86	29.20	0.022	0.00	7.02	14.60	0.017	0.00
	5	11.92	27.80	0.022	0.00	6.93	15.40	0.017	0.00
	6	12.06	28.60	0.022	0.00	7.15	15.20	0.017	0.00
	7	11.77	26.65	0.022	0.00	7.13	15.63	0.017	0.00
<i>N</i> = 300	4	8.53	19.80	0.015	0.00	3.46	9.00	0.011	0.00
	5	8.82	22.80	0.015	0.00	3.55	8.80	0.010	0.00
	6	8.74	18.80	0.015	0.00	3.61	10.20	0.010	0.00
	7	8.46	17.60	0.015	0.00	3.55	10.20	0.010	0.00
<i>N</i> = 400	4	8.25	17.00	0.013	0.00	2.93	8.60	0.009	0.00
	5	8.23	17.20	0.013	0.00	2.74	8.60	0.009	0.00
	6	7.80	16.80	0.012	0.00	2.47	7.80	0.008	0.00
	7	7.57	15.00	0.012	0.00	2.50	6.60	0.008	0.00

Table F3 (cont'd)

<i>N</i> = 500	4	7.74	18.40	0.011	0.00	2.30	10.20	0.008	0.00
	5	7.74	18.20	0.011	0.00	2.13	9.00	0.007	0.00
	6	7.72	18.40	0.011	0.00	2.22	9.20	0.007	0.00
	7	7.62	19.60	0.011	0.00	2.39	10.60	0.007	0.00
<i>N</i> = 750	4	7.79	17.00	0.009	0.00	2.22	7.80	0.006	0.00
	5	7.94	16.20	0.009	0.00	2.19	9.60	0.006	0.00
	6	7.78	18.20	0.009	0.00	2.17	9.20	0.006	0.00
	7	7.43	17.00	0.009	0.00	2.08	8.20	0.006	0.00
<i>N</i> = 1,000	4	6.85	15.00	0.007	0.00	1.18	7.00	0.005	0.00
	5	7.10	16.60	0.007	0.00	1.26	6.80	0.005	0.00
	6	6.89	16.40	0.007	0.00	1.19	7.40	0.005	0.00
	7	6.71	14.40	0.007	0.00	1.26	5.40	0.005	0.00
<i>N</i> = 1,500	4	7.01	16.40	0.006	0.00	1.28	8.60	0.004	0.00
	5	7.25	15.80	0.006	0.00	1.33	6.00	0.004	0.00
	6	7.34	15.00	0.006	0.00	1.56	8.00	0.004	0.00
	7	6.99	16.80	0.006	0.00	1.47	7.40	0.004	0.00

Note. Cat. = number of categories. ML = maximum likelihood, MLR = robust maximum likelihood. M = mean.

Table F3 (cont'd)

		WLSMV				ULSMV			
		Chi-square		RMSEA		Chi-square		RMSEA	
	Cat.	bias	%	bias	%	M	%	bias	%
<i>N</i> = 200	4	2.65	3.82	0.011	0.00	1.46	1.41	0.010	0.00
	5	2.54	4.41	0.011	0.00	1.40	3.21	0.009	0.00
	6	2.89	3.82	0.011	0.00	1.74	3.01	0.010	0.00
	7	3.23	5.01	0.012	0.00	1.98	4.22	0.010	0.00
<i>N</i> = 300	4	1.19	3.40	0.008	0.00	0.23	3.40	0.007	0.00
	5	1.04	3.40	0.008	0.00	0.22	2.80	0.007	0.00
	6	1.27	3.20	0.008	0.00	0.46	2.60	0.007	0.00
	7	1.42	4.80	0.008	0.00	0.61	4.20	0.008	0.00
<i>N</i> = 400	4	1.66	4.41	0.007	0.00	1.01	3.81	0.007	0.00
	5	1.63	4.40	0.007	0.00	1.01	4.00	0.007	0.00
	6	1.62	4.20	0.008	0.00	0.99	2.80	0.007	0.00
	7	1.82	3.60	0.008	0.00	1.11	3.40	0.007	0.00

Table F3 (cont'd)

<i>N</i> = 500	4	0.87	4.40	0.006	0.00	0.29	4.00	0.006	0.00
	5	0.61	4.60	0.006	0.00	0.04	4.00	0.006	0.00
	6	0.60	5.00	0.006	0.00	0.04	4.00	0.006	0.00
	7	0.69	4.00	0.006	0.00	0.04	3.40	0.005	0.00
<i>N</i> = 750	4	1.15	5.80	0.005	0.00	0.76	6.20	0.005	0.00
	5	1.18	5.60	0.005	0.00	0.79	5.60	0.005	0.00
	6	1.21	6.00	0.005	0.00	0.80	5.40	0.005	0.00
	7	0.78	4.60	0.005	0.00	0.34	5.00	0.005	0.00
<i>N</i> = 1,000	4	0.15	4.80	0.004	0.00	-0.11	4.40	0.004	0.00
	5	0.21	5.40	0.004	0.00	-0.04	5.20	0.004	0.00
	6	0.16	4.80	0.004	0.00	-0.11	3.80	0.004	0.00
	7	0.28	3.80	0.004	0.00	-0.01	4.20	0.004	0.00
<i>N</i> = 1,500	4	0.42	5.40	0.003	0.00	0.33	5.40	0.003	0.00
	5	0.57	5.00	0.004	0.00	0.49	5.40	0.004	0.00
	6	0.95	6.20	0.004	0.00	0.86	6.20	0.004	0.00
	7	0.86	5.60	0.004	0.00	0.75	5.20	0.004	0.00

Note. Cat. = number of categories. WLSMV = robust weighted least squares, ULSMV = robust unweight least squares. M = mean.

REFERENCES

REFERENCES

- Anderson, R. D. (1996). An Evaluation of the Satorra-Bentler distribution misspecification correction applied to the McDonald fit index. *Structural Equation Modeling*, 3, 203-227.
- Asparouhov, T., & Muthén, B. O. (2005). Multivariate statistical modeling with survey data. Retrieved from: http://www.fcs.m.gov/05papers/Asparouhov_Muthen_IIA.pdf
- Asparouhov, T., & Muthén, B. O. (2010). *Simple second order chi-square correction*. Retrieved from: http://www.statmodel.com/download/WLSMV_new_chi21.pdf
- Bandalos, D. L. (2006). The use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 385-426). Greenwich, CT: Information Age.
- Beauducel, A., & 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(2), 186-203.
- Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural equation modeling. *Sociological Methods and Research*, 16, 78-117.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bolt, D. M. (2005). Limited- and full-information estimation of item response theory models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics* (pp. 27-71). Mahwah, NJ: Lawrence Erlbaum Associates.
- Boomsma, A. (2013). Reporting Monte Carlo studies in structural equation modeling. *Structural Equation Modeling*, 20, 518-540.
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 58, 430-450.
- Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, 107, 260-273.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematics and Statistical Psychology*, 37, 62-83.
- Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, 8, 1-24. Reprinted in 1977 in D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socioeconomic models* (pp. 205-226). Amsterdam: North Holland.

- Brown, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research*, 29, 468-508.
- Christoffersson, A. (1977). Two-step weighted least squares factor analysis of dichotomized variables. *Psychometrika*, 40, 433-438.
- Coenders, G., Satorra, A., & Saris, W. E. (1997). Alternative approaches to structural modeling of ordinal data: A Monte Carlo study. *Structural Equation Modeling*, 4, 261- 282.
- Cook, C., Heath, F., & Thompson, R. L. (2001). Score reliability in web- or internet-based surveys: unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement*, 61, 697-706.
- Curran, P. J., Bollen, K. A., Paxton, P., Kirby, J., & Chen, F. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a Monte Carlo simulation. *Multivariate Behavioral Research*, 37, 1-36.
- Ding, L., Velicer, W. F., Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling*, 2, 119-144.
- DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment*, 23, 225-241.
- Ethington, C. A., (1987). The robustness of LISREL estimates in structural equation models with categorical variables. *The Journal of Experimental Education*, 55, 80-88.
- Flora, D. B., & Curran P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466-491.
- Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*, 14, 275-299.
- Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicator: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling*, 16, 625-641.
- Gagné, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, 41, 65-83.

- Gartside, P. (2001). Letters to the editor. *The American Statistician*, 55, 171-174.
- 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, 225-271.
- Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling*, 16, 1-27.
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and meta-analysis. *Sociological Methods & Research*, 26, 329-367.
- Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112, 351-362.
- Jackson, D. L., Gillaspay, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14, 6-23.
- Johnson D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, 48(3), 398-407.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.
- Jöreskog, K. G. (2005). *Structural equation modeling with ordinal variables using LISREL*. Retrieved from: <http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf>.
- Jöreskog, K. G., & Sörbom, D. (1986). *PRELIS: A program for multivariate data screening and data summarization. A pre-processor for LISREL*. Mooresville, IN: Scientific Software.
- Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.
- Jöreskog, K. G., Sörbom, D., du Toit, S., & du Toit, M. (2000). *LISREL 8: New Statistical Features*. Chicago: Scientific Software International.
- Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions*. (2nd ed.). Thousand Oaks, CA: Sage.
- Lei, P. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality and Quantity*, 43, 495-507.
- Lietz, P. (2010). Research into questionnaire design: A summary of the literature. *International Journal of Market Research*, 52, 249-272.

- Magnus, J. R., & Neudecker, H. (1986). Symmetry, 0-1 matrices and Jacobians: A review. *Econometric Theory*, 2, 157-190.
- Marsh, H. W., Hau, K., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33, 181-220.
- Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika*, 71, 57-77.
- Medsker, G. M., Williams, L. J., & Holohan, P. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of Management*, 20, 439-464.
- Micceri, T. (1989). The unicorn, the normal curve, than other improbable creatures. *Psychological Bulletin*, 105, 156-166.
- Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115-132.
- Muthén, B. O. (1993). Goodness of fit with categorical and nonnormal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205-234). Newbury Park, CA: Sage.
- Muthén, B. O. (2002). *Using Mplus Monte Carlo simulation in practice: A note on assessing estimation quality and power in latent variable models*. Retrieved from: <https://www.statmodel.com/download/webnotes/mc1.pdf>.
- Muthén, B. O., & 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.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Retrieved from: http://gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf.
- Nester, M. (1996). An applied statistician's creed. *Applied Statistics*, 45, 401-410.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44, 443-460.
- Oranje, A. (2003, April). *Comparison of estimation methods in factor analysis with categorical variables: Applications to NAEP data*. Paper presented at the annual meeting of the

American Education Research Association (AERA), Chicago, IL.

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.

Raykov, T. (2012). Scale construction and development using structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 472-492). New York: The Guildford Press.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variable be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*.

Rigdon, E. E. (1998). Structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 251-294). Mahwah, NJ: Lawrence Erlbaum Associates.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54*, 131-151.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika, 6*, 309-316.

Savalei, V. (2010). Expected versus observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods, 15*, 352-367.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3*, 231-251.

Weijter, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological Methods, 15*, 96-110.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*, 58-79.

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling, 17*, 392-423.

Yuan, K. H., & Bentler, P. M. (1997). Improving parameter tests in covariance structure analysis. *Computational Statistics & Data Analysis, 26*, 177-198.

- Yuan, K. H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, 51, 289-309.
- Yuan, K. H., & Hayashi, K. (2006). Standard errors in covariance structure models: Asymptotics versus bootstrap. *British Journal of Mathematical and Statistical Psychology*, 59, 397-417.
- Yuan, K. H., & Schuster, C. (2013). Overview of statistical estimation methods. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods: Volume 1* (pp. 361-387). New York: Oxford University Press.
- Yuan, K. H., Bentler, P. M., & Zhang, W. (2005). The effect of skewness and kurtosis on mean and covariance structure analysis: The univariate case and its multivariate implication. *Sociological Methods and Research*, 34, 249-258.