UNIVERSITY OF CALIFORNIA

Los Angeles

Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Education

by

Ching-Yun Yu

2002

The dissertation of Ching-Yun Yu is approved.

Yeow-Meng Thum

Sandra Graham

Steve Reise

Bengt Muthén, Committee Chair

University of California, Los Angeles

2002

I dedicate this dissertation to my parents, whose love and patience helped to make a dream reality. I also dedicate this work to my advisor Bengt Muthén, without whose support and guidance, none of this would have been possible.

Table of Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I cannot express enough appreciation and gratitude to my advisor and committee chair, Professor Bengt Muthén, for his unwaivering support and academic guidance. The evolution and completion of this study would not have been possible without his teaching, assistance and critique. Bengt has inspired me in so many ways, and I was very fortunate to work with him and be involved in his research group.

I will always be indebted to my committee members Professors Yeow-Meng Thum, Sandra Graham and Steve Reise for their guidance, support and thoughtful comments. I am also truly thankful to members of Bengt's research group, especially Andreas Klein, Booil Jo, Klaus Larsen, Katherine Masyn, Gitta Lubke, Christy Boscardin and Koo-Hee Kim for their warm friendship, helpful suggestions and recommendations in this study.

I would like to thank many people whose encouragement and support indirectly contributed to the completion of this dissertation. Professors Huo-yan Shyu at Academia Sinica in Taiwan and Seock-Ho Kim at University of Georgia have been mentors in my academic and personal endeavors, and their tremendous amount of support and encouragement have meant a lot to me. Thanks also go to my dear friends, Chun-Chu Lin, Tsai-wei Sun, Meng-Jie Lin, Xiaoxia Ai, Alfonso Ang, Chih-Chien Yang and Ryan Metz for their wise advice, constant praise and support. You have made my life in UCLA rich and unforgettable.

Last, but not least, I thank my family, especially my parents, for unconditionally supporting my endeavors, for always being there when I needed them, and for being proud of all my accomplishments.

VITA

October 17, 1971	Born, Taipei, Taiwan
1994	B.A., Political Science National Taiwan University Taipei, Taiwan
1994-1996	Research Assistant Sun Yat-sen Institute for Social Sciences and Philosophy (ISSP), Academia Sinica Taipei, Taiwan
1997-1998	Statistical Consultant University of Georgia Athens, Georgia
1998	M.Ed., Educational Psychology University of Georgia Athens, Georgia
1998-2001	Graduate Student Researcher University of California, Los Angeles Los Angeles, California
2000-2002	Teaching Assistant University of California, Los Angeles Los Angeles, California

PUBLICATIONS AND PRESENTATIONS

Kim, S.-H., Olejnik, S. F., & Yu, C.-Y. (1998). A Monte Carlo Study of Factors That Affect the Mantel Test and the Generalized Mantel-Haenszel Test of Differential Item Functioning. Paper presented at the 11th European Meeting of the Psychometric Society.

Yu, C. –Y. (2000). Estimating Interaction between Latent Variables Using a Mixture-Modeling Approach. A poster presentation at the annual meeting of the American Educational Research Association, April 25, 2000, New Orleans.

Yu, C. –Y., & Muthén, B. O. (2002). Evaluation of Model Fit Indices for Latent Variable Models with Categorical and Continuous Outcomes. Paper presented at the annual conference of the American Educational Research Association, April 4, 2002, New Orleans.

ABSTRACT OF THE DISSERTATION

Evaluating cutoff criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes

by

Ching-Yun Yu

Doctor of Philosophy in Education

University of California, Los Angeles, 2002

Professor Bengt Muthén, Chair

The aims of this study are to first evaluate the performance of various model fit measures under different model and data conditions, and, secondly, to examine the adequacy of cutoff criteria for some model fit measures. Model fit indices, along with some test statistics, are meant to assess model fit in latent variable models. They are frequently applied to judge whether the model of interest is a good fit to the data. Since Bentler and Bonett (1980) popularized the concept of model fit indices, numerous studies have been done to propose new fit indices or to compare various fit indices. Most of the studies, however, are limited to continuous outcomes and to measurement models, such as confirmatory factor analysis models (CFA). The present study broadens the structure of models by including the multiple causes and multiple indicators (MIMIC) and latent growth curve models. Moreover, both binary and continuous outcomes are investigated in the CFA and MIMIC models.

Weighted root-mean-square residual (WRMR), a new fit index, is empirically evaluated and compared to the Tucker-Lewis Index (TLI), the Comparative Fit Index (CFI), the root-mean-square error of approximation (RMSEA) and the standardized root-mean-square residual (SRMR). Few studies have investigated the adequacy of cutoff criteria for fit indices. This study applies the method demonstrated in Hu and Bentler (1999) to evaluate the adequacy of cutoff criteria for the fit indices. The adequacy of a conventional probability level of 0.05 for χ^2 to assess model fit is also investigated. With non-normal continuous outcomes, the Satorra-Bentler rescaled χ^2 (SB) is incorporated into the calculation of TLI, CFI and RMSEA, and these SB-based fit measures are evaluated under various cutoff values. An example of applying adequate cutoff values of overall fit indices is illustrated using the Holzinger and Swineford data. Generally speaking, the use of SRMR with binary outcomes is not recommended. A cutoff value close to 1.0 for WRMR is suitable under most conditions but it is not recommended for latent growth curve models with more time points. CFI performs relatively better than TLI and RMSEA, and a cutoff value close to 0.96 for CFI has acceptable rejection rates across models when $N \geq 250$.

CHAPTER 1

INTRODUCTION

In the development and evaluation of latent variable models and related procedures, an important and consistent theme for the last two decades has been model evaluation/ selection and model fit indices (e.g., Austin & Calderón, 1996; Bollen & Long, 1993). After a substantively based model is specified and estimates are obtained, the researcher might desire to evaluate its model fit and check whether the model is consistent with the data. Many model fit measures have been proposed to serve this purpose of assessing overall model fit of a hypothesized model.

Let $\mathbf{\Sigma}$ represent the population covariance matrix of observed variables and $\mathbf{\Sigma}(\theta)$ represent the covariance matrix written as a function of parameters θ for a hypothesized model. The overall fit measures assess whether the covariance structure hypothesis $\mathbf{\Sigma} = \mathbf{\Sigma}(\theta)$ is valid, and, if not, they measure the discrepancy between $\mathbf{\Sigma}$ and $\mathbf{\Sigma}(\theta)$ (Bollen, 1989). There are two types of the more popular overall fit measures. One type is the chi-square (χ^2) test statistic, and the other type is an array of various fit indices (Hu & Bentler, 1999). These two types of fit measures are described below.

In latent variable modeling (LVM), the asymptotic χ^2 test statistic was developed under the framework of hypothesis testing and was the first one applied widely to assess overall model fit. The probability level associated with the χ^2 value (Chi-P) serves as a criterion to evaluate whether a hypothesized model is a good fit to the data, and the decision is to either accept or reject the specified model. Currently the most widely used χ^2

test statistic is obtained from maximum likelihood (ML) estimation. However, as noted by many researchers, there are a few problems associated with model evaluation based on the ML χ^2 statistical test. One problem is that, not only model adequacy, but sample size and violation of some underlying assumptions also affect its value. For continuous outcomes, a ML χ^2 approximation makes the following assumptions (Bollen, 1989).

1. *The null hypothesis (H_0): $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta)$ holds exactly.* Because the χ^2 statistic is a test of the hypothesis $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta)$, the information it provides is whether the model fits the data exactly or not. However, in practice it might not be realistic to assume that a specific model $\boldsymbol{\Sigma}(\theta)$ exists in the population.
2. *The sample is sufficiently large.* By the asymptotic distribution theory the χ^2 statistic approximates a χ^2 distribution only in large samples. A simulation study by Anderson and Gerbing (1984) shows that the ML χ^2 test statistic in small samples tends to be large and this leads to too many rejections of H_0.
3. *The observed variables have no kurtosis.* That is, the observed variables should have a multivariate normal distribution. This implies that the ML χ^2 estimation is not accurate for non-normal or categorical data. Leptokurtic (more peaked than normal) distributions result in too many rejections of the null models (Browne, 1984).

The general χ^2 test statistic in covariance structure analysis also requires the first and second assumptions to be valid. The power of the χ^2 test is partially a function of sample size (N). The χ^2 estimator increases in direct proportion to N - 1, and its power increases as

N increases. In small samples, its power is not sufficient and the chance to commit a Type II error increases (Tanaka, 1993). The dependence on sample size suggests that the χ^2 statistic might not be comparable across samples. Moreover, since the χ^2 statistic assumes that $\mathbf{\Sigma} = \mathbf{\Sigma}(\theta)$ holds exactly, a trivially false model $\mathbf{\Sigma}(\theta)$ (that is, residuals with no practical significance) might be rejected in large samples, whereas a less appropriate model might be accepted as adequate in small samples. Fit indices have been designed to avoid some problems of the (ML) χ^2 statistic mentioned above.

A fit index is an overall summary statistic that evaluates how well a particular covariance structure model explains the sample data. Rather than test hypotheses, fit indices are meant to quantify features such as the sum of residuals or variance accounted for by the proposed model (Hu & Bentler, 1998) and "to provide information about the degree to which a model is correctly or incorrectly specified for the given data" (Fan, Thompson & Wang, 1999). Fit indices are often used to supplement the χ^2 test to evaluate the acceptability of latent variable models. Since Bentler and Bonett (1980) introduced the use of model fit indices to the analysis of covariance structures, numerous fit indices based on different rationales have been proposed and studied (e.g., Anderson & Gerbing, 1984; Marsh, et al., 1988; Bentler, 1990; Mulaik et al., 1989; Williams & Holahan, 1994). Fit indices that will be investigated in this study are the Tucker-Lewis Index (TLI), the Comparative Fit Index (CFI), the root-mean-square error of approximation (RMSEA), the standardized root-mean-square residual (SRMR), the weighted root-mean-square residual (WRMR). Detailed descriptions of these fit indices will be presented in the next section.

Among the five fit indices, WRMR was recently proposed in Muthén and Muthén

(1998-2001) and has not yet been studied. This study will thus evaluate WRMR and compare its performance to other fit indices listed above. To investigate whether the fit indices are (or which fit index is) more relevant to model adequacy than the χ^2 test statistic, it is important to evaluate the performance of fit indices under various model and data conditions. The conditions investigated in the past, for example, included sensitivity of fit indices to model misspecification, sample size, estimation methods, model complexity, and violation of distribution assumptions. My study will evaluate these fit indices under conditions such as model misspecification (true, misspecified and trivially misspecified models), type of outcome variables (normal, non-normal continuous and binary outcomes), type of model specification and various sample sizes. These conditions are explained in detail below.

Satorra (1990) mentioned that two types of assumptions, structural and distributional, are needed to justify the validity of analyses in structural equation modeling. Structural assumptions set up a model $\mathbf{\Sigma}(\theta)$ and imply a specific structure for the population covariance matrix $\mathbf{\Sigma}$. A correct structural specification means that the hypothesized model reflects the population structure, whereas a model misspecification would result in violations of structural assumptions. There are two kinds of misspecified models: over-parameterized (estimating one or more parameters when their population values are, in fact, zero) and under-parameterized misspecified model (specifying values of one or more parameters at zero when their population values are non-zero) (Hu & Bentler, 1998). Previous studies (e.g., La Du & Tanaka, 1989; Hu & Benter, 1999) have indicated that the over-parameterized misspecified models have zero population noncentrality and do not have significantly different estimates for model fit indices. Thus, only under-parameterized misspecification is considered in this study. The χ^2 test statistic has long been known to be too powerful for trivially misspecified models with large samples, thus it is interesting to compare its performance with that of fit indices in trivially misspecified models. Good fit measures should be sensitive to degree of misspecification but should not possess too much power in trivially false models. The performance of the χ^2 test statistic and fit indices in the correct specified, misspecified and trivially misspecified models will be investigated in this study.

The effects of violations of distributional assumptions on model fit measures will also be investigated. The distribution form of observed variables is often assumed to be multivariate normal and the ML method is often used. However, non-normal and categorical outcomes frequently occur in applications and, thus, the values of model fit measures based on ML estimation might be biased. The Satorra-Bentler robust χ^2 (SB) was proposed by Satorra and Bentler (1988) to yield a better approximation of a χ^2 variate under non-normality. Previous studies (e.g., Chou et al., 1991; Hu et al., 1992; Curran et al., 1996; Anderson, 1996) have shown that, with non-normal continuous outcomes, the SB χ^2 outperforms the ML χ^2 under some combinations of non-normality and model specification at adequate sample size. The SB-based McDonald Fit Index was found to perform better in correctly specified models (Anderson, 1996), and the results in Nevitt and Hancock (2000) showed that, relative to the ML-based RMSEA, the SB-based RMSEA at a cutoff of 0.05 had better control over type I errors under non-normality and smaller sample size. Thus, besides comparing the SB χ^2 and ML χ^2 test statistics, this study will also incorporate the SB χ^2 into the estimation of TLI, CFI and RMSEA and compare their performance under various cutoff values with the "unadjusted" ones. In addition to normal and non-normal continuous outcomes, performance of model fit measures with respect to binary outcomes will also be investigated. Some questions that will be discussed are: does the SB χ^2 perform better than the ML χ^2 in confirmatory factor analysis models with non-normal data? Do the model fit indices such as TLI and CFI adjusted using the SB χ^2 information also perform better with non-normal data?

Hu and Bentler (1998, 1999) have shown that some fit indices are sensitive to different types of model specification. For example, the ML-based TLI, CFI, and RMSEA are more sensitive to models with misspecified factor loadings, whereas SRMR is more sensitive to models with misspecified factor covariances. Sample size has also been shown to be a prominent factor that affects the performance of model fit indices. Thus, these two factors will be investigated in this study.

Most models investigated in the fit-index study have been limited to measurement models, such as the confirmatory factor analysis (CFA) models (e.g., Anderson & Gerbing, 1984; Marsh et al. 1988; La Du & Tanaka, 1989; Curran et al., 1996; Hu & Benter, 1998:1999; Nevitt and Hancock, 2000). However, applications of LVM have proceeded beyond these measurement models. For example, since Jöreskog introduced it in 1980s (Jöreskog, 1971; Jöreskog & Goldberger, 1975), the multiple causes and multiple indicators (MIMIC) models have been frequently fitted to data by substantive researchers. In recent years, one trend in structural equation modeling has been to incorporate and conceptualize growth curve modeling under the framework of LVM (e.g., Muthén &

Curren, 1997; Muthén & Khoo, 1998). Also a recent trend is to model longitudinal data using a combination of continuous and categorical latent variables (Muthén, 2001). Few studies, however, evaluated the performance of fit indices for these more complex models. There is a need to expand the fit index study to a variety of models, such as structural or growth curve models, so that the practitioners will have some rules of thumb to apply these fit indices more accurately.

Another important issue for the study of fit indices is the selection of adequate cutoff criteria. Typical practice in LVM has been to adopt some cutoff criteria for fit indices as the decision rules to evaluate model fit. For example, CFI values larger than 0.9 suggest that the model might be useful, and those below 0.9 indicate that there might be some inconsistency between model and data. Similarly, Steiger (1989) and Browne and Cudeck (1993) suggested that, for RMSEA, values below 0.05 would suggest a good fit, and those larger than 0.1 would suggest a poor fit. However, there is not enough evidence or rationale to support these conventional rules of thumb, and the legitimacy of these conventional cutoff criteria has been questioned (e.g., Marsh, 1995; Marsh & Hau, 1996; Carlson & Mulaik, 1993). Hu and Benter (1999) empirically evaluated the adequacy of cutoff values based on the criterion that the adequate cutoff values should result in minimum type I and type II errors. Their study, however, is limited to continuous outcomes and does not consider trivially misspecified models. We often encounter categorical outcomes in social and behavioral sciences, and it is important to know how well fit indices perform and what their adequate cut-off values for categorical outcomes might be. Thus, in addition to continuous outcomes, this study will compare these fit indices and obtain their adequate cut-off values in CFA and MIMIC models with dichotomous outcomes.

1.1 Research Questions to be Addressed

The first purpose of this study is to compare the performance of fit indices with that of Chi-P and provide researchers information on how well the fit indices work under different model and data conditions. In addition to CFA models, the performance of model fit measures in MIMIC and latent growth curve models will be evaluated. Continuous outcomes will be considered in CFA, MIMIC and latent growth curve models. Both continuous and binary outcomes will be investigated in CFA and MIMIC models, and the fit indices will be compared to Chi-P under conditions such as model misspecification, types of model specification and various sample sizes.

Some questions that will be discussed in relation to sample sizes and model misspecification, for example, are:

- Does sample size affect the performance of model fit indices?
- How large should the sample size be to allow enough power for model fit indices?
- Which fit index has more power to detect misspecified models?

With respect to non-normal outcomes, some questions that will be tackled are:

- Does the SB χ^2 outperform the ML χ^2?
- Do the SB χ^2-based model fit indices such as TLI, CFI and RMSEA outperform their ML χ^2-based counterparts?

- Are fit indices robust to non-normality?

The second purpose of this study is, based on the results obtained earlier, to suggest the adequate cutoff criteria for TLI, CFI, RMSEA, SRMR and WRMR. The suitable cutoff criteria of TLI, CFI, RMSEA, SRMR, WRMR and Chi-P will be evaluated in CFA, MIMIC and latent growth curve models with continuous outcomes. Some questions that will be discussed in relation to continuous outcomes, for example, are:

- What are the adequate cutoff criteria (rules of thumb) for the fit indices?
- Do the adequate cutoff criteria vary with different types of model specification and/ or sample sizes?

The suitable cutoff criteria for TLI, CFI, RMSEA, SRMR and WRMR will also be obtained in CFA and MIMIC models with dichotomous outcomes. Some questions that will be tackled in relation to dichotomous outcomes, for example, are:

- Which cutoff criteria for the fit indices generate less type I and type II errors?
- What are the adequate cutoff criteria for the fit indices?
- Are cutoff values and performance of fit measures with dichotomous outcomes different from those with continuous ones?

1.2 Scope and Significance of the Study

Model selection is no doubt an important part of the modeling process. After specifying a model based on substantive theory and fitting the model to the data, one needs assistance to judge whether this model is a good or useful one. In LVM, model fit indices are important alternatives to the χ^2 test statistic for model selection because, for example, the χ^2 statistic is a function of N and it tests unrealistic null hypotheses. Applied researchers often assess fit using the χ^2 test statistic with a probability level of 0.05, and/ or using model fit indices with some conventional cutoff values. However, the adequacy of these cutoff values is questionable. This study adopts a few methods to empirically evaluate the performance of the model fit measures, and thus may provide more objective guidelines for applying these fit measures.

A newly developed and yet-to-be-evaluated fit index, WRMR, is included and compared to the other fit indices in this study. This study will also shed new light on the use of model fit indices with binary outcomes, and on the use of fit indices in latent variable models where the predictors are included and/ or longitudinal designs are taken into account. A wide array of multivariate distributions and models will be studied, and the robustness of the unadjusted or adjusted model fit indices against non-normality will be investigated. In addition, the adequacy of using the conventional or scaled χ^2 test statistics to assess model fit will be investigated. This study will provide applied researchers valuable information concerning which fit index is preferable and which cutoff criteria for fit indices are more suitable.

CHAPTER 2

LITERATURE REVIEW

2.1 Measures Used for Assessing Model Fit

2.1.1 Likelihood Ratio Test/ χ^2 Test Statistic

A likelihood function, $L(\theta)$, gives the probability of observing the specific data given a set of parameters θ that specified in our model of interest. Once estimates of the specific model are obtained, the likelihood ratio test (LRT) procedure can be applied to test model fit. The LRT statistic is

$$LRT = -2\log \{\max[L(\theta_i)] / \max[L(\theta_j)]\},$$

where $L(\theta_i)$ is the likelihood for model i with parameters θ_i, and $L(\theta_j)$ is the likelihood for model j with parameters θ_j. Model j is the more restricted model, and is nested within model i. Each likelihood is evaluated at the values of θ that maximize it. The large sample distribution of LRT under H_0, with estimates that maximize likelihoods, is a χ^2 distribution. Its associated degrees of freedom are the number of difference between the freely estimated parameters of models i and j. The associated probability level of the χ^2 represents the probability of obtaining a χ^2 value larger than the value calculated, given that the H_0 is true. The higher the probability level of the calculated χ^2, the better the fit of our hypothesized model.

In regular practice of hypothesis testing in LVM, the model of interest is the more restricted model j, and it is often compared to a saturated model (H_1: $\mathbf{\Sigma} = \mathbf{S}$). **S** is the sample covariance matrix of the observed variables. For a covariance structure $\mathbf{\Sigma}(\theta)$, population parameters θ are estimated with the goal to minimize the discrepancy between the observed and estimated population covariance matrices [**S** and $\boldsymbol{\Sigma}(\hat{\theta})$]. The estimates $\hat{\theta}$ can be obtained by minimizing a fitting function $F(\theta)$, where $F(\theta)$ is a function representing the discrepancy between **S** and $\boldsymbol{\Sigma}(\theta)$. Denote the minimum of the $F(\theta)$ as $F(\hat{\theta})$ and assuming a Wishart distribution for **S**, then, under the H_0, $(N-1)F(\hat{\theta})$ has an asymptotic χ^2 distribution with degrees of freedom equal to $p(p+1)/2 - q$, where p is the number of observed variables and q is the number of parameters that are freely estimated in model $\boldsymbol{\Sigma}(\theta)$.

Currently the ML fitting function, $F_{ML}(\theta)$, is the most widely used one. The product of N-1 and $F_{ML}(\hat{\theta})$, under the H_0 and the assumption of multivariate normality, approximates a χ^2 distribution in large samples. These assumptions limit the use of the ML estimation under certain data and model conditions. The χ^2 approximation is sensitive to sample size and violation of the multivariate normality assumption. The research done by Muthén and Kaplan (1992) showed that the χ^2 approximation is also sensitive to model complexity. Moreover, a model is tentative and is only regarded as a proxy to reality, thus the reject or fail-to-reject decision obtained from testing the hypothesis $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta)$ seems not be the main research interest. The information that researchers often want to know is whether the model fit is adequate and how closely their model fits the data. Therefore, just as Jöreskog and Sörbom (1981) mentioned, the χ^2 approximation is probably better regarded as a fit measure instead of a test statistic because its assumptions are seldom fulfilled in practice.

A few approaches have been proposed to address the limitations of the ML estimation method. Among them, one is to develop alternative methods of estimation that is robust to small sample sizes and non-normal data. For example, Browne (1982, 1984) proposed an asymptotic distribution free (ADF) method and test statistic that were not based on the multivariate normality assumption. (N-1) $F_{ADF}(\hat{\theta})$ is asymptotic distributed as a χ^2 (Browne, 1984). However, some studies (e.g., Hu, Bentler & Kano, 1992; Yuan & Bentler, 1998) have shown that the ADF estimator performed poorly under sample sizes of 2500. Additionally, Muthén and Kaplan (1992) have shown that the ADF estimator performed worse with increasing model complexity and/ or at smaller sample sizes. The ADF estimator requires very large samples to behave like a χ^2 (Bentler & Yuan, 1999), and models with more than 20 variables are not feasibly estimated (Browne, 1984).

Another approach is to adjust and rescale the existing estimators. For example, Satorra and Bentler (1988) rescaled the ML χ^2 for the presence of non-zero kurtosis under non-normality, and developed the SB χ^2. Recently, Bentler and Yuan (1999) proposed a statistic based on an adjustment to the ADF χ^2, and found it performed well at small sample sizes such as 60 to 120.

2.1.2 Comparative Fit Indices

Comparative fit indices measure the improvement of fit by comparing the hypothesized model with a more restricted baseline model. The baseline model commonly used is a null or independent model where the observed variables, with variances to be estimated, are mutually uncorrelated (Bentler & Bonett, 1980). The model fit information obtained from these fit indices are very different from that obtained from the χ^2 measure where a hypothesized model is compared to a saturated model.

Tucker-Lewis Index (TLI)

Originating from Tucker and Lewis (1973), Bentler and Bonett (1980) applied TLI to covariance structure analysis and claimed that it can be used to compare a particular model across samples. TLI is calculated as

$$TLI = \frac{\chi_b^2 / df_b - \chi_{H_0}^2 / df_{H_0}}{(\chi_b^2 / df_b) - 1}, \quad (1)$$

where df_b and df_{Ho} are the degrees of freedom for the baseline and the hypothesized (under H_0) models, respectively. TLI can exceed the 0 to 1 range. Anderson and Gerbing (1984) show that TLI values tend toward 1 for a correctly specified model, but in small samples (sample size smaller than 100) its value is underestimated (that is, indicates a bad fit for an acceptable model) and has large sampling variability. Hu and Bentler (1999) recommended a cutoff value of TLI close to 0.95.

Comparative Fit Index (CFI)

To avoid TLI's problems concerning underestimation of fit and considerable sampling variability in small samples, Bentler (1995, 1990) proposed CFI in 1988. Bentler (1990) conducted a simulation study to compare TLI, CFI, the Normed Fit Index (NFI) and the Incremental Fit Index (IFI), and concluded that CFI is the best index.

Although TLI indicates a greater degree of misspecification, CFI has the advantages of having a 0-1 range and smaller sampling variability. CFI is defined as

$$CFI = 1 - \frac{\max[(\chi^2_{H_0} - df_{H_0}), 0]}{\max[(\chi^2_{H_0} - df_{H_0}), (\chi^2_b - df_b), 0]}. \quad (2)$$

Hu and Bentler (1999) recommended a cutoff value of CFI close to 0.95. Both TLI and CFI are incremental fit indices, which measure the improvement of fit by comparing a H_0 model with a more restricted baseline model.

2.1.3 Error-of-Approximation Indices

Cudeck and Henly (1991) mentioned that three types of discrepancy functions could be used as a basis for model selection. They are the sample discrepancy, the overall discrepancy and the discrepancy due to approximation. The sample discrepancy is F[$\mathbf{S}$, $\mathbf{\Sigma}(\hat{\theta})$], which represents the discrepancy between the sample covariance matrix $\mathbf{S}$ and the estimated covariance matrix $\mathbf{\Sigma}(\hat{\theta})$ for the model fitted to the *sample*. It is stochastic and depends on sample size. The χ^2 measure mentioned earlier are based on the sample discrepancy function. Let $\mathbf{\Sigma}(\tilde{\theta})$ denote the best fit of model to the population covariance matrix $\mathbf{\Sigma}$, the discrepancy due to approximation is then F[$\mathbf{\Sigma}$, $\mathbf{\Sigma}(\tilde{\theta})$]. It is a population quantity and does not depend on sample data. Based on the error of approximation, Steiger and Lind (1980) introduced fit indices, the root mean square (RMS) and RMSEA, for model selection.

Root-mean-square Error of Approximation (RMSEA)

Introduced by Steiger and Lind (1980) and Browne and Cudeck (1993), RMSEA for continuous outcomes is calculated as

$$RMSEA = \sqrt{\max[(\frac{2F(\hat{\theta})}{d} - \frac{1}{N}), 0]}\ , \quad (3)$$

where d denotes the degrees of freedom of the model, and $F(\hat{\boldsymbol{\theta}})$ is the minimum of the fitting function $F(\boldsymbol{\theta})$. N in (3) sometimes is replaced by N-1, which is motivated by assuming a Wishart distribution for **S** (e.g., Bollen, 1989; Nevitt & Hancock, 2000). Here we adopt the use of N in Muthén and Muthén (1998-2001, p. 360). With categorical outcomes, d in (3) is replaced by a function of the sample variances (Muthén & Muthén, 1998-2001). RMSEA has a known distribution and, thus, permits the calculation of confidence intervals. Browne and Cudeck (1993) suggested that RMSEA values larger than 0.1 are indicative of poor-fitting models, values in the range of 0.05 to 0.08 are indicative of fair fit, and values less than 0.05 are indicative of close fit. Hu and Bentler (1999) recommended a cutoff value of RMSEA close to 0.06.

2.1.4 Residual-Based Fit Indices

SRMR and WRMR measure the (weighted) average differences between the sample and estimated population variances and covariances.

Standardized Root-mean-square Residual (SRMR)

Introduced by Bentler (1995), it is a standardized version of the root mean square residual (RMR) which was developed by Jöreskog and Sörbom (1981). SRMR for continuous outcomes is defined as

$$SRMR = \sqrt{\sum_j \sum_{k \le j} \frac{r_{jk}^2}{e}}, \tag{4}$$

$$\text{where} \quad r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}} - \frac{\hat{\sigma}_{jk}}{\sqrt{\hat{\sigma}_{jj}}\sqrt{\hat{\sigma}_{kk}}}, \tag{5}$$

and e = (p(p + 1))/2. s_{jk} and $\hat{\sigma}_{jk}$ are the sample and model-estimated covariance between the continuous outcomes y_j and y_k, respectively. s_{jj} and s_{kk} are the sample variances for the continuous outcomes y_j and y_k. p is the number of continuous outcomes. When $s_{jj} = \hat{\sigma}_{jj}$ and $s_{kk} = \hat{\sigma}_{kk}$, this formula coincides with Hu and Bentler (1998).

SRMR has a 0-1 range. For categorical outcomes, Muthén and Muthén (1998-2001) define SRMR when all outcomes are categorical with no threshold structure or covariates. Hu and Bentler (1999) recommended a cutoff value close to 0.08 for SRMR.

Weighted Root-mean-square Residual (WRMR)

WRMR was proposed in Muthén and Muthén (1998-2001). It is defined as

$$WRMR = \sqrt{\sum_r^e \frac{(s_r - \hat{\sigma}_r)^2 / v_r}{e}}, \tag{6}$$

where e is the number of sample statistics. s_r and $\hat{\sigma}_r$ are elements of the sample statistics and model-estimated vectors, respectively. v_r is an estimate of the asymptotic variance of s_r. WRMR is suitable for models where sample statistics have widely disparate variances and when sample statistics are on different scales such as in models with mean and/or threshold structures. It is also suitable with non-normal outcomes.

WRMR, for categorical outcomes, is equivalent to:

$$WRMR = \sqrt{\frac{2\,\mathrm{NF}(\hat{\boldsymbol{\theta}})}{e}}\,, \tag{7}$$

where $F(\hat{\boldsymbol{\theta}}) = \min[(1/2)(\mathbf{s} - \hat{\boldsymbol{\sigma}})' W_D^{-1} (s - \hat{\boldsymbol{\sigma}})]$, which is the minimum of the weighted least squares (WLS) fitting function. $(W_D)_{jj}$ is the asymptotic variance of s_j (see Muthén and Muthén,1998-2001, p. 361-362).

2.2 Review of Related Monte Carlo Studies

Small sample size and departure from normality are two factors that have been found to affect the estimates of fit measures. The Monte Carlo study by Boomsma (1983) has shown that the χ^2 approximation is not accurate for N ≤ 50, and sample sizes larger than 200 might be necessary to avoid problems of non-convergence or improper solutions. Boomsma investigated four different CFA and structural equation models under various sample sizes with up to 300 replications, and the model parameters in his study ranged from twelve to seventeen. Boomsma also found that a high skewness value might lead to high χ^2 estimates and thus result in rejecting the population models too often.

Hu, Bentler and Kano (1992) conducted a large-scale simulation study to evaluate the use of the χ^2 test statistics for assessing model fit with continuous outcomes. The χ^2 test statistics estimated by various methods were evaluated in a fifteen-variable, three-factor CFA model (87 degrees of freedom) under seven data distributions at various sample sizes (150, 250, 500, 1000, 2500 and 5000). The iterations were 200 per condition. The results showed that the ML χ^2 had inflated rejection rates under non-normality, and both ML χ^2

and SB χ^2 had inflated type I errors (reject when Chi-P < 0.05) at N ≤ 250 even with multivariate normal data.

The same fifteen-variable, three-factor CFA model of Hu, Bentler and Kano (1992) was simulated in Hu and Benter (1998; 1999). In the latter study, two more model conditions were considered. They were model specification (misspecified factor covariances or factor loadings) and type of model misspecification (correctly specified or under-parameterized misspecifications). Hu and Bentler (1998) investigated the sensitivity of the ML-, generalized least squares (GLS)- and ADF-based fit indices to model misspecification. It was found that SRMR was the most sensitive to models with misspecified factor covariances, whereas TLI, CFI and RMSEA were more sensitive to models with misspecified factor loadings. For the ML method, RMSEA, TLI and CFI were found to have high correlations (the absolute values of correlations ranged from 0.96 to 1.0), and SRMR was the least similar to the other fit indices. The use of SRMR, supplemented by TLI, CFI or RMSEA, was recommended. Most of the ML-based fit indices outperformed the GLS- and ADF-based ones. Hu and Bentler in 1999 examined the adequacy of some conventional cutoff criteria for various fit indices. With the ML method, they suggested a cutoff value close to 0.95 for TLI and CFI, close to 0.06 for RMSEA and close to 0.08 for SRMR. However, using the suggested cutoff criteria, RMSEA and TLI tended to overreject properly specified models at small sample sizes.

Curran, West and Finch (1996) conducted a simulation study based on a nine-indicator, three-factor CFA model to study the robustness of the ML, SB and ADF χ^2 to non-normality. Data were generated from three levels of data distribution (normality, moderate and severe non-normality) and four different model specifications at sample sizes of 100, 200, 500 and 1000. There were 200 replications per condition. It was found that the SB χ^2 was underestimated with increasing non-normality, and the power of the SB to reject misspecified models attenuated with non-normal data. The ML χ^2 was inflated with increasing non-normality for misspecified models. Finally, in comparison with ML and ADF, the SB χ^2 was found to perform better across nearly all condition. It is notable that, as opposed to the results in Hu, Benter and Kano (1992), the SB χ^2 was not inflated at small sample size in the nine-indicator model.

Nevitt and Hancock (2000) conducted a simulation study based on the same CFA model as Curran et al. (1996). Two model specifications (properly specified and misspecified), four sample sizes and three data distributions were considered to compare the use of the ML-based RMSEA to that of the SB-based and the bootstrap adjusted RMSEA for assessing model fit. There were 200 replications per condition. For properly specified models, it was found that the RMSEA means tended to decrease with increasing sample sizes. In addition, the ML-based RMSEA increased with increased non-normality, whereas the SB-based RMSEA appeared to be more stable across different data distributions. The ML-based RMSEA in the close-fit test (reject models when RMSEA > 0.05) had type I error rates near 5% level across sample sizes under normality but overrejected true models systematically with increasing non-normality. With respect to the misspecified model (under-parameterized factor loadings), average values of the ML-based RMSEA tended to increase while those of the SB-based RMSEA tended to decrease with increasing non-normality. The rejection rates of the ML- and SB-based

RMSEA at a cutoff of 0.05 generally increased with increasing sample sizes. Finally, with the rejection rule of RMSEA > 0.05, the power of the SB-based RMSEA tended to decrease with increasing departure from non-normality, and the SB-based RMSEA maintained better type I error control than the ML-based RMSEA under non-normality.

CHAPTER 3

METHODOLOGY

3.1 Estimators

The Mplus (Muthén and Muthén, 1998-2001) program is used to assess the values of the model fit measures. The χ^2 test statistic and the five model fit indices can all be estimated with both continuous and categorical outcomes in Mplus. Depending on scales of the outcomes, three different estimators are used in this study. The first two estimators are ML based and are designed to estimate models with continuous outcomes. They are named ML and MLM in Mplus. For estimating dichotomous outcomes, we will use a robust (mean- and variance-adjusted) method of WLS; it is named WLSMV in Mplus. Below is a brief description of these three estimators (Muthén & Muthén, 1998-2001, p. 38).

ML - maximum likelihood parameter estimates with conventional standard errors and χ^2 test statistics. Estimates are obtained by minimizing the fitting function

$$F_{ML}(\theta) = tr(\mathbf{\Sigma}^{-1}\mathbf{S}) - \log|\mathbf{\Sigma}^{-1}\mathbf{S}| - p, \quad (8)$$

with respect to θ, where p is the number of the observed outcomes (Jöreskog & Sörbom, 1979). Denote the minimum of (8) as $F_{ML}(\hat{\theta})$, the ML χ^2 statistic is then given by 2N $F_{ML}(\hat{\theta})$ (Muthén & Muthén, 1998-2001, p. 359). Note that the factor N, instead of N-1, is used.

MLM - maximum likelihood parameter estimates with robust standard errors and a mean-adjusted χ^2 test statistic. Estimates are obtained by minimizing the fitting function

$F_{MLM}(\theta) = tr(\mathbf{\Sigma}^{-1}\,\mathbf{T}) - \log|\mathbf{\Sigma}^{-1}\,\mathbf{S}| - p,$ (9)

where $\mathbf{T} = \mathbf{S} + (\bar{\boldsymbol{v}} - \boldsymbol{u})(\bar{\boldsymbol{v}} - \boldsymbol{u})'$, (10)

$\bar{\boldsymbol{v}}$ is the sample mean vector and $\boldsymbol{u}$ is the population mean vector (e.g., Jöreskog & Sörbom, 1979; Jöreskog & Sörbom, 1996; Muthén & Muthén, 1998-2001). This mean-adjusted χ^2 statistic is defined as

$G_M = 2N\,F(\hat{\theta})/c,$ (11)

where $c = tr(\hat{\boldsymbol{U}}\,\hat{\boldsymbol{V}}_{ss})/d.$ (12)

c is a scaling correction factor. $\hat{\boldsymbol{V}}_{ss}$ is obtained from the sample covariances and fourth-order multivariate product moments, $\hat{\boldsymbol{U}}$ is the residual weight matrix and weight matrix under the model, and d denotes the degrees of freedom of the model (see pages 212 and 218 of Bentler, 1995; or page 357 of Muthén & Muthén, 1998-2001). This mean-adjusted χ^2 is commonly called the SB χ^2. When the multivariate normality assumption underlying the ML estimation is violated, the ML χ^2 test statistic is inflated and does not follow the expected χ^2 distribution. By modifying the model χ^2 with the correlation factor c, the SB χ^2 yields a better approximation of a χ^2 variate under non-normality. With normal data, the SB χ^2 simplifies to the ML χ^2.

WLSMV – weighted least square parameter estimates using a diagonal weight matrix with robust standard errors and mean- and variance-adjusted χ^2 test statistic. As mentioned earlier, WLS estimates are obtained by minimizing the fitting function

$F_{WLS}(\theta) = (\boldsymbol{s} - \boldsymbol{\sigma})'\boldsymbol{W}_D^{-1}(\boldsymbol{s} - \boldsymbol{\sigma}),$ (13)

where $(W_D)_{jj}$ is the asymptotic sample variance of the outcome y_j. A mean- and variance-adjusted χ^2 statistic has similar form as (9) but with different definitions of c and d.

$$c = \mathrm{tr}(\hat{\boldsymbol{U}}\hat{\boldsymbol{V}}_{ss})/d' \quad (14)$$

$$d^* = [\mathrm{tr}(\hat{\boldsymbol{U}}\hat{\boldsymbol{V}}_{ss})]^2/\mathrm{tr}[(\hat{\boldsymbol{U}}\hat{\boldsymbol{V}}_{ss})^2], \quad (15)$$

where d* is computed as the closest integer to d′ (see pages 357 and 358 of Muthén & Muthén, 1998-2001).

3.2 Assessment of Fit Indices

3.2.1 Sensitivity to Model Misspecification

Misspecifications of loadings larger than 0.7 or covariances larger than 0.3 would yield practically as well as statistically significant effects on the parameter estimates and model fit, thus model fit measures should possess strong power in this case. Moreover, the more the misspecified coefficients, the larger power the fit indices should have. The χ^2 test statistic has long been known to be too powerful for trivially false models, thus it is important to know whether the alternative model fit indices have the same drawback. Adequate model fit measures should not have too much power in trivially false models.

3.2.2 Cutoff Criteria of Fit Measures

For each cutoff value of the Chi-P and fit indices, the rejection rates (percentage of the 500 replications that reject the proposed model) were calculated. Adequate cut-off values of fit indices should exert higher probabilities of accepting correct models and rejecting misspecified models (power); that is, they should have small type I and type II errors.

From our design, type I errors can be evaluated as the rejection rates for the true models, whereas type II errors can be evaluated as the acceptance rates for the misspecified models. Power corresponds to the rejection rates for the misspecified models, and its compliment is the probability of a type II error. By balancing and minimizing type I and type II errors, we can then evaluate and obtain adequate cutoff values of the fit indices.

CHAPTER 4

MONTE CARLO STUDY 1: CFA

This paper applies a Monte Carlo method to evaluate the performance and to examine the adequate cutoff values for fit indices. By means of Monte Carlo simulation, samples are generated from an "assumed" true model in the population and, thus, the extent to which the model fit indices identify the true models are known. The class of latent variable models is broad, and it is plausible that the performance and suitable cutoff criteria of model fit measures vary with different types of models. In this dissertation, the performance and adequacy of cutoff criteria for model fit measures are evaluated in CFA, MIMIC and latent growth curve models with the aim to expand the generalizability and usefulness of the fit-index investigation. The designs of the three main Monte Carlo studies are summarized in Table 4.1. For each model, the design and methods used to investigate the adequacy of cutoff criteria are described first, followed by discussions on results.

A CFA model is often applied to "confirm" the hypothesized relationship between a set of observed variables and a set of latent variables. Researchers can impose substantively based constraints on factor loadings, factor correlations or error variances/ covariances in a CFA model. The CFA models in Hu and Bentler (1998) are used to evaluate model fit measures in this study. Model fit measures are evaluated under conditions such as model misspecification (true, misspecified and trivially misspecified models), type of outcome variables (normal, non-normal continuous and binary outcomes), type of model specification and various sample sizes (100, 250, 500 and 1000). Detailed designs and results for the CFA models are presented next.

4.1 Design of Simulation

Model specification

The confirmatory factor models of Hu and Bentler (1998) were used to generate data in this study. There were in total fifteen observed variables and three factors. The model can be expressed as $y = \Lambda \xi + \varepsilon$, where y, ξ and ε are, respectively, observed-variable, common-factor and error vectors; Λ is a factor loadings matrix.

There were two types of under-parameterized misspecified models designed in this study; one had misspecified factor covariances (Hu and Bentler called it the "simple" model) and the other had misspecified factor loadings ("complex" model). The model structure is presented in Figure 4.1 and values of the population parameters are listed in Table 4.2.

Simple model. In the simple models there were no cross loadings. Three specifications of the simple model were considered. The first model was correctly specified such that the estimated parameters in the sample exactly corresponded to their population structure (which will be referred to as the True model). The other two models excluded factor covariances from the sample that existed in the population. The population covariance between factor 1 and factor 2 (COV(F1, F2)) was 0.5 and covariance between factor 1 and factor 3 (COV(F1, F3)) was 0.4. The first misspecified model (referred to as Miss1) fixed COV(F1, F2) at zero in estimation (that is, path a of Figure 4.1 was set to 0);

the second one (referred to as Miss2) fixed both COV(F1, F2) and COV(F1, F3) at zero in estimation (paths a and b of Figure 4.1 were set to 0). Although we estimated three factor covariance parameters in the simple true models, we only estimated two (or one) factor covariance parameters in the Miss1 (or Miss2) models.

Complex model. In addition to the fifteen factor loadings and three factor covariances of the sample models, the complex models added three cross-factor loadings with values 0.7 (paths d, e and f in Figure 4.1). Besides the True model, two misspecified models were also designed for the complex models. The first one fixed the loading of the 1st observed variable on F3 (path d in Figure 4.1) which existed in the population as zero, and the second one, in addition to path d, fixed the loading of the 4th variable on F2 (path e) as zero in estimation. For the purpose of identification, the factor loading of the last indicator on each factor estimated in the sample was fixed at 0.8 for both simple and complex models.

Simple model with trivial covariances (simple trivial model). The design for the simple trivial model was similar to the simple model, except that the covariances (path a, b and c in Figure 4.1; also see Table 1) were all set to 0.05 in the population structure. The value of 0.05 was chosen because, first, a covariance (equal to correlation here) value of 0.05 indicated a small and ignorable relationship. Secondly, ignoring one or two covariances of 0.05 was found not to affect much the parameter estimates and the residuals between the observed and estimated covariances. The relative bias of the parameter estimates (the obtained estimate minus the population value divided by the population value) was below the 10% level. Three simple trivial models were specified for estimation. They were the True (correctly specified), Miss1 (misspecified path a as zero) and Miss2 (misspecified paths a and b as zeros), respectively.

Complex model with trivial loadings (complex trivial model). The three cross loadings (paths d, e and f in Figure 4.1) were specified to be 0.1 in the population structure. Misspecified one or two loadings of 0.1 to be 0 was found not affect much the residual covariances, and the relative bias of the parameter estimates was below the 10% level. There were also three complex trivial models specified for estimation: the True (correctly specified), Miss1 (misspecified path d as zero) and Miss2 (misspecified paths d and e as zeros) models, respectively.

<u>Type of outcome variables</u>

Continuous outcomes. Three population distributions were considered for each of the model specifications with continuous outcomes. The first distribution was multivariate normal with zero skewness and zero kurtosis. The second distribution was moderately non-normal with univariate skewness of 2.0 and kurtosis of 7.0. The third distribution represented severely non-normal with univariate skewness of 3.0 and kurtosis of 21.0. According to Curran et al. (1996), these levels of non-normality reflect the real data distributions that they found in community-based mental health and substance abuse research.

Binary outcomes. Variables were generated from multivariate normal factors/ errors and then were dichotomized according to different choices of thresholds. Two population distributions were considered. Distribution 1 (referred to as an Equal case) dichotomized observed variables at a threshold of 0.5, and the proportion of these two categories for all the fifteen observed variables was 0.31 to 0.69. Distribution 2 (referred to as an Unequal case) dichotomized the fifteen variables unequally such that the proportion of the two categories for the 1st, 2nd, 6th, 7th, 11th and 12th observed variables was 0.4 to 0.6. The proportion of the two categories for the 3rd, 4th, 8th, 9th, 13th and 14th observed variables was 0.25 to 0.75, and that for the 5th, 10th and 15th was 0.1 and 0.9. The uneven proportion of observed variables was chosen to reflect the real data sets that we often encounter in applications. Note that the variances of the fifteen outcomes were scaled to be one in the complex models with binary outcomes.

Sample size

Since there were thirty-three and thirty-six parameters respectively in simple and complex true models, a sample size of one hundred allowed for approximately three cases per estimated parameter and seemed minimal to estimate the models properly. Thus, one hundred was chosen as the minimum sample size investigated. Three other sample sizes of 250, 500 and 1000 were also considered.

Replications

Five hundred replications were obtained for each condition.

Procedure

Continuous outcomes. SAS (SAS Institute, 1988) was used to generate data based on three different distributional specification levels (normal, moderately and severely non-normal), two model specification types (simple and complex), and four different sample sizes (N=100, 250, 500 or 1000). The simple and complex trivial models are generated under normal distributions and four different sample sizes. Vale and Maurelli's (1983) procedures were used to generate multivariate non-normal distributions. All fifteen observed variables were generated from the model implied covariance matrix $\mathbf{\Sigma}(\theta)$ with the same degree of non-normality. The non-normal distributions were generated in the observed level rather than the latent level to better manipulate the degree of non-normality observed in applications. The same procedures were also adopted in some previous studies such as Chou et al. (1991), Curran et al. (1996) and Nevitt et al. (2000).

Correctly and incorrectly specified structural models (True, Miss1 and Miss2) were fitted to each simulated sample using the ML or MLM estimators in Mplus. The maximum number of iterations to convergence was set to 1000 in Mplus by default. With multivariate normal outcomes, the MLM estimator would generate similar results as the ML estimator except for sample size of less or equal to 100. Thus, with normal outcomes and at the sample size of 100, MLM results for model-fit measures are also provided. With moderately and severely non-normal outcomes, both ML and MLM estimation results are provided for Chi-P, TLI, CFI and RMSEA. Currently WRMR is only available with the MLM estimator in Mplus. Trivial models were fitted to samples with normal outcomes, and ML estimation results are provided.

Binary outcomes. Data was generated based on three conditions: distributional specification levels (2 levels), model specification (2 types) and sample sizes (4 levels). In each of the sixteen conditions, five hundred samples were drawn. Then, correctly and incorrectly specified structural models were fitted to each simulated sample using the WLSMV estimator in Mplus.

A new added utility RUNALL, a set of batch files in Mplus version 2, was used in this simulation study. The RUNALL utility is designed for Monte Carlo studies with data sets generated outside the Mplus. It fits one model (specified in one input file) to each of the data sets in Mplus and automatically saves the results into one file so that values of fit measures can be easily accessed and summarized. This utility can be downloaded from the Mplus webpage at http://www.statmodel.com/runutil.html.

4.2 Results

This section is organized as follows. The performance and suitable cutoff criteria of the Chi-P and fit indices are discussed for continuous outcomes, binary outcomes and trivially misspecified models, respectively. Three different continuous outcomes (normal, moderately and severely non-normal) and two binary outcomes (Equal and Unequal cases) are examined. The performance and a conventional cutoff value of 0.05 for Chi-P are evaluated. For each of the fit indices, the cutoff criteria suggested by Hu and Bentler (1999) are investigated first with continuous outcomes, followed by discussions on whether these cutoff values are applicable to other model and data conditions, such as trivial models and binary outcomes.

4.2.1 Continuous Outcomes

Chi-p. Table 4.3.1 presents the results of rejection rates for Chi-P under multivariate normal distributions. When N = 100 under normality, the ML-based Chi-P with a conventional cutoff value of 0.05 (that is, applying the rejection rule of Chi-P < 0.05) rejected 14.8% and 13.8% of the simple and complex true-population models; the type I error rates were slightly high. With this cutoff value, the power of the χ^2 test statistic ranged from 0.58 to 1 for the simple and complex misspecified models across sample sizes. The rejection rates of the SB-based Chi-P were slightly higher than those of the ML-based Chi-P at N = 100. With N ≥ 250, the conventional cutoff Chi-P value of 0.05 yielded reasonable rejection rates for both true and misspecified models. Tables 4.3.2 and 4.3.3 present the results of rejection rates for Chi-P under moderately and severely non-normal distributions, respectively. Consistent with previous studies, the ML χ^2 under non-normal distributions was inflated tremendously and thus resulted in high rejection rates (for example, rejected 78.4% to 100% of the samples with a cutoff value of 0.05) in both true and misspecified models. Moreover, increasing non-normality was accompanied by an increase in type I errors of the ML χ^2. In comparison with the ML χ^2, the SB χ^2 exhibited a better control over type I errors under non-normality with increasing sample sizes. However, with a conventional cutoff value of 0.05, the SB-based Chi-P still overrejected true models at N ≤ 250 under moderate non-normality (type I error rates ranged from 16.8% to 42.4%) and across all four sample sizes under severe non-normality (type I error rates ranged from 14.0% to 74.2%).

TLI. Rejection rate summaries of the ML- and MLM-based TLI for multivariate normal, moderately and severely non-normal distributions are presented in Tables 4.4.1, 4.4.2 and 4.4.3, respectively. It was shown in Table 4.4.1 that a cutoff value around 0.95

for the ML-based TLI has type I error rates close to the nominal rate of 5% if the multivariate normality assumption can sustain. With moderately and severely non-normal distributions, TLI with a cutoff value of 0.95 tended to overreject true-population models at samples less than 250, just as Hu and Benter (1999) observed. Table 4.4.3 shows that, under severe non-normality, the ML- (SB-) based TLI at a cutoff value of 0.95 rejected 57.2% to 93% (10.6% to 73.2%) of the true models when N ≤ 250. At a cutoff of 0.95, the SB-based TLI was preferable at N ≤ 250 under severely non-normal distributions, although its type I errors at N = 100 were still high (73.2%). Note that, in general, the power of TLI deteriorated in simple models but remained strong in complex (especially for Miss2) models with increasing sample sizes. The results are consistent with those in Hu and Bentler (1999).

CFI. Rejection rate summaries of the ML- and SB-based CFI for multivariate normal, moderately and severely non-normal distributions are presented in Tables 4.5.1, 4.5.2 and 4.5.3, respectively. With cutoff values less than 0.94 for the ML-based CFI under normality, the simple misspecified models appeared to be underrejected (rejection rates ranged from 0% to 33.6%). Therefore, the ideal cutoff value of CFI should be 0.95 or larger. In comparison with TLI, rejection rates of CFI were generally lower in both true and misspecified models with continuous outcomes. That is, CFI, with the same cutoff value, had lower type I errors as well as lower power than TLI did. With normal distributions, the low power associated with CFI at cutoff values of equal and less than 0.95 in simple misspecified models can be improved by increasing the cutoff value to 0.96. In doing so, CFI then had acceptable power in simple Miss2 models (rejection rates ranged from 59.8% to 65.2%) along with reasonable type I error rates.

With increasing non-normality, the ML-based CFI tended to reject a high percentage of true models and result in inflated type I errors at small sample sizes. The results for the non-normal continuous outcomes showed that the SB-based CFI had lower type I errors and provided better rejection rates in true models at $N \leq 250$. With $N \geq 500$, the power of the SB-based CFI to detect simple misspecified models deteriorated. One might wish to raise the cutoff value of CFI to 0.96 to maintain sufficient power, however, the type I error rates of CFI at this cutoff were large at $N \leq 250$ under severely non-normal distributions (ranged from 12% to 75%). Under the combination of a small sample size (N = 250 or less) and severe non-normality, the SB-based CFI was preferable in order to control the inflated type I errors. Similar to TLI, CFI had better power to detect complex misspecified models than to detect simple misspecified models with increasing sample size.

RMSEA. Rejection rate summaries of the ML- and SB-based RMSEA for multivariate normal, moderately and severely non-normal distributions are presented in Tables 4.6.1, 4.6.2 and 4.6.3, respectively. Under normal distributions (Table 4.6.1), the type I error summary of the ML-based RMSEA at cutoff values less than 0.06 was not satisfactory at N = 100 (rejection rates ranged from 11.6% to 23.2% for simple true models and 11.0% to 23.0% for complex true models). With a cutoff value of 0.06, RMSEA had acceptable type I errors; its power ranged from 0.8 to 1 to detect complex misspecified models and ranged from 0 to .54 to detect simple misspecified models.

With increasing non-normality, rejection rates of the ML-based RMSEA increased and resulted in higher type I errors at $N \leq 250$. For example, under severely non-normal distributions, the ML-based RMSEA at a cutoff value of 0.06 rejected 89.6% and 95.6% of the simple and complex true models at N = 100, and it also rejected 55.2% and 70.4% of the simple and complex true models at N = 250. Type I errors of the ML-based RMSEA were too high under the combination of non-normality and small sample sizes, and the SB-based RMSEA were able to control some of the inflated type I errors. For example, with N = 250 under severe non-normality, the SB-based RMSEA only rejected up to 0.4% of the simple and complex true models, which exhibited a good control of type I errors.

Tables 4.6.1, 4.6.2 and 4.6.3 show that the power of the ML-based RMSEA increased with increasing non-normality. But the seeming advantage of power for the ML-based RMSEA should be interpreted cautiously. Nevitt et al. (2000) argued that the high power of the ML-based RMSEA was only an artifact caused by the inflated ML χ^2 test statistic under non-normality. Their results show that with increasing non-normality, the average ML-based RMSEA values increased monotonously whereas the average SB-based RMSEA values might decrease. Because the average ML-based RMSEA values increased with increasing non-normality, rejection rates of the ML-based RMSEA also increased.

This study also found that, the ML- and SB-based RMSEA (especially the SB-based RMSEA) with a cutoff around 0.06 tended to underreject the simple misspecified models at N ≥ 250. Reducing the cutoff value of RMSEA to 0.05, although with inflated type I errors at N = 100, rejected a better percentage of the misspecified models.

SRMR. Tables 4.7.1, 4.7.2 and 4.7.3 present the rejection rates of SRMR under multivariate normal, moderately and severely non-normal distributions, respectively. Normal distribution (Table 4.7.1) is discussed first. With a cutoff value of SRMR smaller or equal to 0.06 at N = 100, true-population models were overrejected (rejection rates ranged from 42.4% to 97.8% for simple true models and 15.6% to 90.6% for complex true models). Thus, a cutoff value higher than 0.06 should be adopted. At a cutoff value of 0.08, power of SRMR to detect simple misspecified models (ranging from 99.8% to 100%) was much higher than its power to detect complex misspecified models (ranging from 0% to 75.4%). Moreover, with increasing sample size, the power of SRMR stayed large to detect simple misspecified models but deteriorated to detect complex ones. In comparison with the cutoff value of 0.08, SRMR at a cutoff value around 0.07 demonstrated a similar pattern but had better power in complex misspecified models. Thus, with normal outcomes SRMR ≤ 0.07 seems to be a better criterion than SRMR ≤ 0.08 to indicate good models.

The power of SRMR showed a similar pattern under non-normality (Tables 4.7.2 and 4.7.3). The power of SRMR in simple models was larger than that in complex models, and an increasing sample size was associated with a decrease of power in complex models. Moreover, the rejection rates of SRMR increased with increasing non-normality. To maintain reasonable power in complex models at larger sample sizes, a cutoff value of 0.07 might be better than 0.08 for SRMR when N ≥ 250.

WRMR. Tables 4.8.1, 4.8.2 and 4.8.3 present the rejection rates of WRMR under multivariate normal, moderately and severely non-normal distributions, respectively. With cutoff values equal to 0.6, 0.7 or 0.8 under normal distributions, the true models were overrejected (rejection rates ranged from 36.8% to 96.2% for simple models and ranged from 13.8% to 86.4% for complex models). With cutoff values of 0.95 or 1.0, WRMR had moderate or strong power to detect simple and complex models (ranging from 47.6% to

100%) with reasonable type I errors (ranging from 0 to 3.8%). Overall, 0.95 or 1.0 seemed to be appropriate cutoff values for WRMR under normality.

Increasing non-normality was accompanied by an increase of type I errors for WRMR in both simple and complex models. Under moderate non-normality, a cutoff value of 0.95 for WRMR exhibited small type I error rates except for the simple model at N = 100 (rejection rate was 18.4%). A cutoff value close to 1.0 for WRMR might be preferable in order to control type I errors at small samples. Under severe non-normality, WRMR with a cutoff value of 0.95 should be applied with caution to the simple models at N ≤ 250 (type I error rates ranged from 22.2% to 43.2%). WRMR ≤ 1.0 can be used to identify good complex models when N ≥ 250 and good simple models when N ≥ 500. Different from the previous four fit indices (TLI, CFI, RMSEA and SRMR), the power of WRMR increased in *both* simple and complex models with increasing sample sizes.

To evaluate the similarities between the performance of fit measures, Figure 4.2 and Figure 4.3 present the pairwise scatterplots for fit measures under the true and Miss1 models based on 500 replications. Each of these replications had a sample size of 500 with normal outcomes. Under true models, it appeared that there were two clusters of correlated fit measures. SRMR and WRMR clustered with high correlation (0.99), and the other cluster of high intercorrelations included Chi-P, TLI, CFI and RMSEA (the absolute values of correlations ranged from 0.84 to 0.98). Under Miss1 models, Chi-P had values very close to 0 for all 500 replications. These seemed to be still two clusters. TLI and CFI appeared to be perfectly correlated, and RMSEA still correlated more highly with TLI and CFI than with WRMR and SRMR. The general trend and relationships between pairs of the fit indices seemed to be similar in the true and Miss1 models. Since some fit indices correlated highly (such as the cluster of TLI and CFI or the cluster of WRMR and SRMR), it might be sufficient to use or report just one fit index from each of the clusters.

4.2.2 Dichotomous Outcomes

Chi-P. Tables 4.9.1 and 4.9.2 present the rejection rates of Chi-P for the Equal and Unequal cases, respectively. Table 4.9.1 (Equal case) shows that except for slightly low power in complex misspecified models at N = 100, Chi-P with a cutoff value of 0.05 had satisfactory power (ranging from 0.85 to 1) to detect both simple and complex misspecified models with acceptable type I error rates (2.4% to 5%). Just like its performance with normal outcomes, the conventional cutoff value of 0.05 for Chi-P tended to overreject true models at N = 100 in the Unequal case (Table 4.9.2; rejection rates were 14.0% and 11.0% for the simple and complex true models). The rejection rule of Chi-P < 0.05 might not be suitable for small samples and, to reduce the inflated type I errors, an alternative way might be lowering the cutoff value to, e.g., 0.01.

TLI. Tables 4.10.1 and 4.10.2 present the rejection rates of TLI for the Equal and Unequal cases, respectively. Although in the Unequal case TLI tended to overreject the true models at N = 100 (rejection rates were 17.6% and 9.4% for the simple and complex true models), a cutoff value of 0.95 seemed to still be a reasonable cutoff value for TLI with binary outcomes at N ≥ 250. In contrast against the continuous outcomes, the power of TLI increased in simple models but decreased in complex models with increasing sample sizes for binary outcomes. This resulted in better performance of TLI in models with misspecified factor covariances rather than models with misspecified factor loadings for binary outcomes.

CFI. With the same cutoff value of 0.95, the type I errors and power of CFI were larger than those of TLI in both Equal and Unequal (Table 4.11.1 and 4.11.2) models. The rejection rule of CFI < 0.95 seemed to overreject the simple true models when N = 100 (rejecting 10.2% and 19.6% of the samples in the Equal and Unequal cases, respectively). To maintain reasonable power in complex misspecified models with binary outcomes, the rejection rule CFI < 0.95 seemed to be better than TLI < 0.95. At N ≥ 250, a cutoff value of 0.96 for CFI had better type I and type II error rates than the cutoff value of 0.95.

RMSEA. Tables 4.12.1 and 4.12.2 present the rejection rates of RMSEA for the Equal and Unequal cases, respectively. With a cutoff value around 0.06, RMSEA tended to overreject the true-population unequal models at N = 100, especially with the Unequal case. For example, Table 4.12.2 shows that RMSEA rejected 21.6% and 18.0% of the simple and complex true models. Thus, a cutoff value around 0.06 for RMSEA was less preferable at small sample sizes. To control type I errors, it might be preferable to increase the cutoff value of RMSEA to 0.07 or 0.08. However, in doing so, RMSEA underrejected the complex misspecified models with increasing sample sizes. Overall, with N ≥ 250, a cutoff value around 0.05 was better than 0.06 for RMSEA in that the former had less type II errors. Similar to CFI and TLI, the power of RMSEA was larger to detect simple misspecified models than to detect complex misspecified models with binary outcomes.

SRMR. Tables 4.13.1 and 4.13.2 show that, with a cutoff value close to 0.08 for SRMR, its rejection rates in true models were excessively large at N = 100 (ranging from 98.4% to

100.0%). For the Unequal case (Table 4.13.2), the type I error rates of SRMR with the cutoff of 0.08 and at N = 250 were also large (67.2 and 37.8% for simple and complex models, respectively). Moreover, with increasing sample sizes, the power of SRMR decreased substantially in complex models (ranging from 0 to 1). With binary unequal outcomes, there seemed to be no adequate cutoff value for SRMR (either the type I errors were overly inflated at N ≤ 250 or the power were too low to reject complex misspecified models at N ≥ 500).

WRMR. Tables 4.14.1 and 4.14.2 show the rejection rates of WRMR for the Equal and Unequal cases, respectively. In the Unequal case, WRMR at the cutoff value of 0.95 overrejected the simple true model at N = 100 (rejection rate was 14.4%). In comparison with a cutoff value of 0.95, WRMR at a cutoff value of 1.0 had acceptable and lower type I error rates in both simple and complex models. However, WRMR at a cutoff value of 1.0 tended to underreject the complex misspecified models at small sample sizes (e.g., when N = 100 in the complex Miss1 models, the rejections rates were 4.8% and 14.5% for the Equal and Unequal cases, respectively). Overall, 1.0 seemed to be an acceptable cutoff value of WRMR for both continuous and dichotomous outcomes.

Note that, when we simply lowered cutoff criterion of the ML or SB-based Chi-P to 0.01, the rejection rates of Chi-P were then close to those of WRMR with a cutoff value of 1.0. It implied that WRMR might not have much advantage over the χ^2 test statistic if we adopted a more conservative probability level for χ^2 with binary outcomes. In addition, the power of WRMR and χ^2 was very large (often 1.0) when sample size was large, and this posed a potential concern that WRMR, similar to χ^2, might overreject (trivially) misspecified models. The χ^2 test statistic has long been known to be too powerful for trivially false models, and it is important to know whether WRMR has the same pattern. The performance of the fit measures in trivially misspecified models is discussed below.

4.2.3 Trivially Misspecified Models

Tables 4.15 to 4.20 present the rejection rates of the Chi-P and fit indices under various cutoff values for normal distributions in trivially misspecified models. Generally speaking, there were two patterns shown in the trivially misspecified models among the six model fit measures. TLI, CFI, RMSEA and SRMR at their suggested cutoff criteria rarely rejected the trivially false models, especially at large sample sizes. On the contrary, WRMR and Chi-P gained more power as N increased. When sample sizes were large, the conventional rejection rule of Chi-P < 0.05 exhibited stronger power to reject trivially misspecified complex models than the other fit indices at previously suggested cutoff values. The rejection rates of WRMR at a cutoff value close to 1.0 (rejection rates ranged from 0.6% to 18.0%) in complex trivial models were similar to those of Chi-P at a cutoff value close to 0.01 (rejection rates ranged from 2.6% to 21.2%). The rejection rates of WRMR were very different from the Chi-P and other fit indices, however, in simple trivial models. Even with just one correlation of 0.05 misspecified (simple Miss1 model), WRMR at a cutoff value of 1.0 rejected models over 50% of the time when N = 1000.

4.3 Summary and Conclusion

This chapter evaluated the adequacy of cutoff values for fit measures under different sample size, model specification, type of model misspecification and type of outcomes variables for CFA models. The results were summarized with respect to the types of outcomes. When the outcomes were normal, a cutoff value of 0.05 for Chi-P overrejected the true models at N = 100. Relatively speaking, for all four sample sizes, the better choices of cutoff values for the ML-based TLI, CFI, RMSEA and SRMR were close to 0.95, 0.96, 0.05 and 0.07 respectively. A cutoff value close to 0.05 or 0.045 for RMSEA was necessary to maintain reasonable power at N ≥ 250, although it tended to overreject the true models at N = 100. WRMR ≤ 0.95 or 1.0 to indicate good fit had moderate to large power to detect *both* simple and complex misspecified models with small type I errors. In some cases, a few cutoff values for the same fit index may be acceptable, such as cutoff values of 0.95 or 0.96 for CFI and TLI, values of 0.07 or 0.08 for SRMR, values of 0.05 or 0.06 for RMSEA, and values close to 0.95 or 1.0 for WRMR. Thus, the suggested cutoff values should not be seen as fixed rules. Comparatively speaking, the power of TLI, CFI and RMSEA to detect misspecified complex models was larger than that to detect misspecified simple models; the power of SRMR was larger to detect misspecified simple models; WRMR and Chi-P had strong power to detect both types of misspecified models.

From the previous cutoff criteria and under moderate non-normality, only WRMR with a cutoff value of 1.0 sustained reasonable power (ranged from 0.42 to 1) to detect misspecified models with acceptable type I errors for *both* simple and complex models across all four sample sizes. The ML-based TLI ≥ 0.95, the ML-based CFI ≥ 0.95, SRMR ≤ 0.07 and the SB-based Chi-p ≥ 0.01 seemed to be acceptable indications of good simple and complex models at N ≥ 250. Note that the SB-based Chi-P, TLI, CFI and RMSEA at the suggested cutoff values still overrejected true models at N = 100 under non-normality, although their type I error rates were greatly reduced comparing to their ML-based counterparts. A cutoff value close to 0.05 for the ML-based RMSEA had reasonable rejection rate results at N ≥ 500.

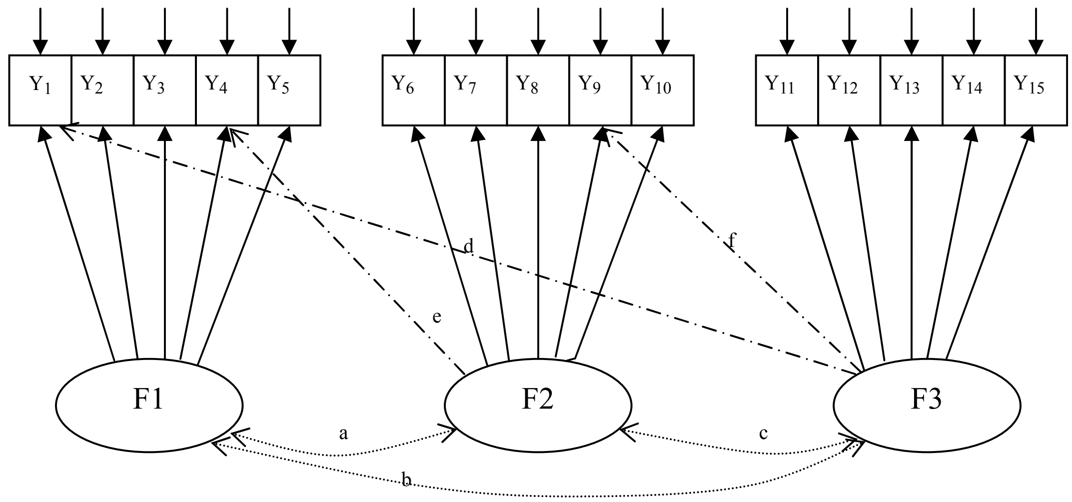
Under severe non-normality, none of the model fit indices at the suggested cutoff values had acceptable type I errors at N = 100. SRMR at a cutoff value close to 0.7, the ML-based TLI and CFI at a cutoff value close to 0.95 and WRMR at a cutoff value close to 1.0 were still applicable when sample size was equal or larger than 500. At a small sample size (N = 250), a cutoff value of 0.95 for the SB-based CFI and a cutoff value of 0.07 for SRMR had acceptable type I and type II errors.

With binary outcomes, the results suggested that SRMR was not a good fit index. In addition, none of the fit indices had ideal type I errors and at the same time maintained strong power for both simple and complex models across all samples. WRMR with a cutoff value of 1.0, except for its low power in complex models at small sample sizes, had acceptable type I and type II errors for *both* simple and complex models. In comparison with TLI ≥ 0.95, the performance of CFI ≥ 0.95 to indicate good models was better because its power to detect complex misspecified models stayed strong across samples. With binary outcomes, Chi-P ≥ 0.05, CFI ≥ 0.95 (or 0.96), RMSEA ≤ 0.05 and WRMR ≤ 1.0 can be indications of good models with binary outcomes at N ≥ 250.

From the above summary, some cutoff values of the fit indices seemed to be applicable to both continuous and binary outcomes across most sample sizes. A cutoff value close to 0.95 for CFI (using the SB-based CFI under the combination of severe non-normality and N = 250) was applicable to both continuous and binary outcomes when N ≥ 250, but one needed to be cautious of its low power in simple models with normal and moderately non-normal outcomes, especially when a sample size was large. Applying WRMR ≤ 1.0 to indicate good models, except for its potential low power at small samples with binary outcomes, can work well for normal continuous, moderately non-normal continuous and binary outcomes as well as for both simple and complex models across samples. However, it appeared that WRMR might be too powerful for trivial misspecification of factor covariances. A cutoff value of 0.05 for Chi-P resulted in inflated type I errors at N = 100 with all but binary equal outcomes. Thus, except for binary equal outcomes, Chi-P ≥ 0.05 was not an appropriate good-model indicator at N = 100. We would like to avoid incorrectly rejecting the model if it is true, thus the aim was to select cutoff values and fit measures that have minimum type I errors with acceptable type II error rates. Note that the rules of thumb suggested above had type I error rates lower than 5% with liberal type II error rates in some cases around or up to 70%.

Similar to previous studies, it was found that the SB χ^2 and SB-based TLI, CFI and RMSEA outperformed the ML χ^2 and ML-based TLI, CFI and RMSEA under some combinations of non-normality and model specification at certain sample sizes. The SB χ^2 had much lower type I errors than the ML χ^2 under non-normality, but it still overrejected the true models at N ≤ 250 under moderate non-normality and across all four sample sizes under severe non-normality. Generally speaking, with severely non-normal outcomes and small sample size, the SB-based TLI, CFI and RMSEA were recommended to reduce type I errors. However, the true models were still overrejected at N = 100 under non-normality using the cutoff values close to 0.95 (or 0.96), 0.95 (or 0.96) and 0.05 (or 0.06) for the SB-based TLI, CFI and RMSEA.

Figure 4.1. Structures of true-population and misspecified models for the CFA models



Source:

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure analysis: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424-453.

Figure 4.2. Scatterplot of the relationship between pairs of fit measures in the CFA true models

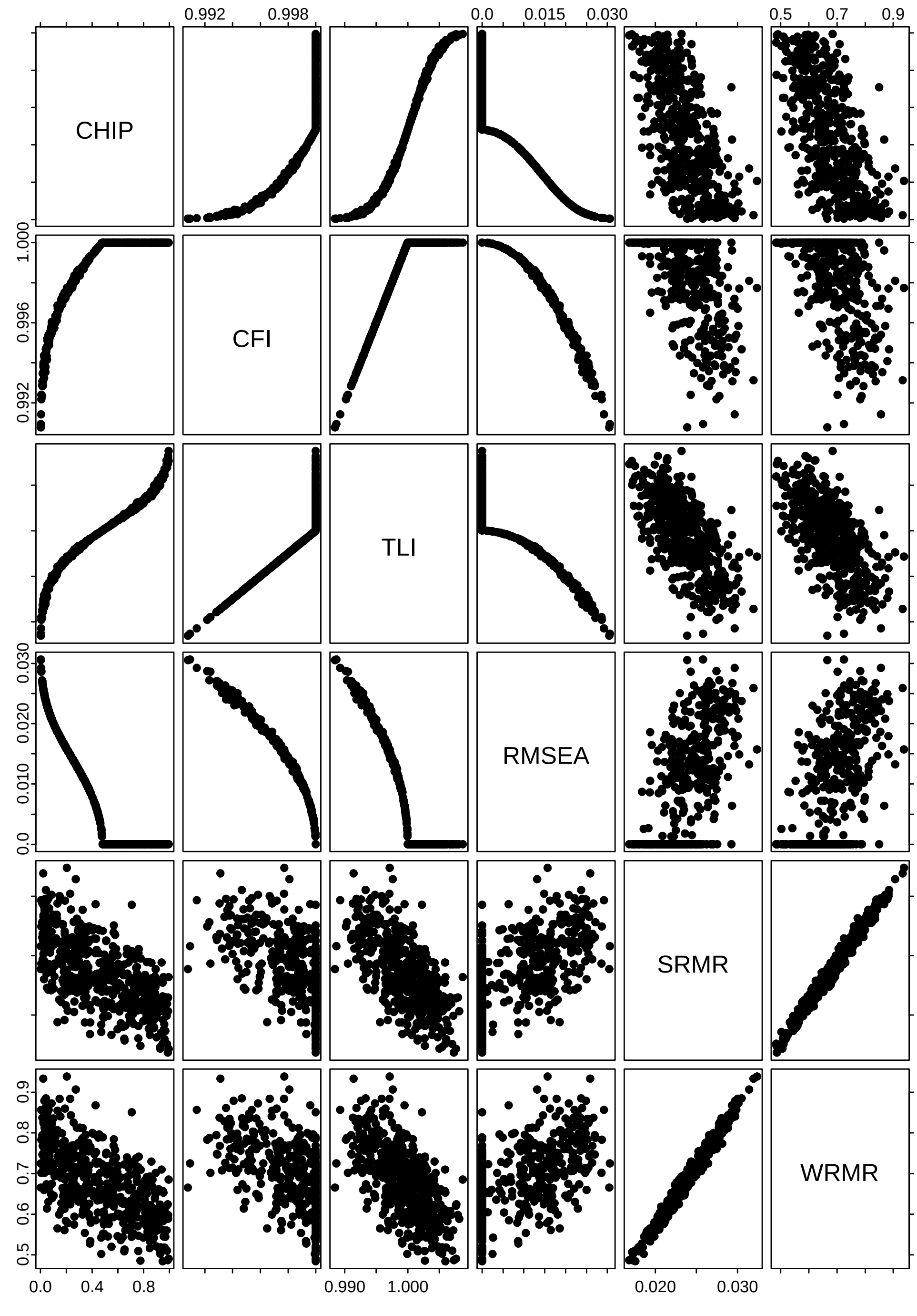


Figure 4.3. Scatterplot of the relationship between pairs of fit measures in the CFA Miss1 models

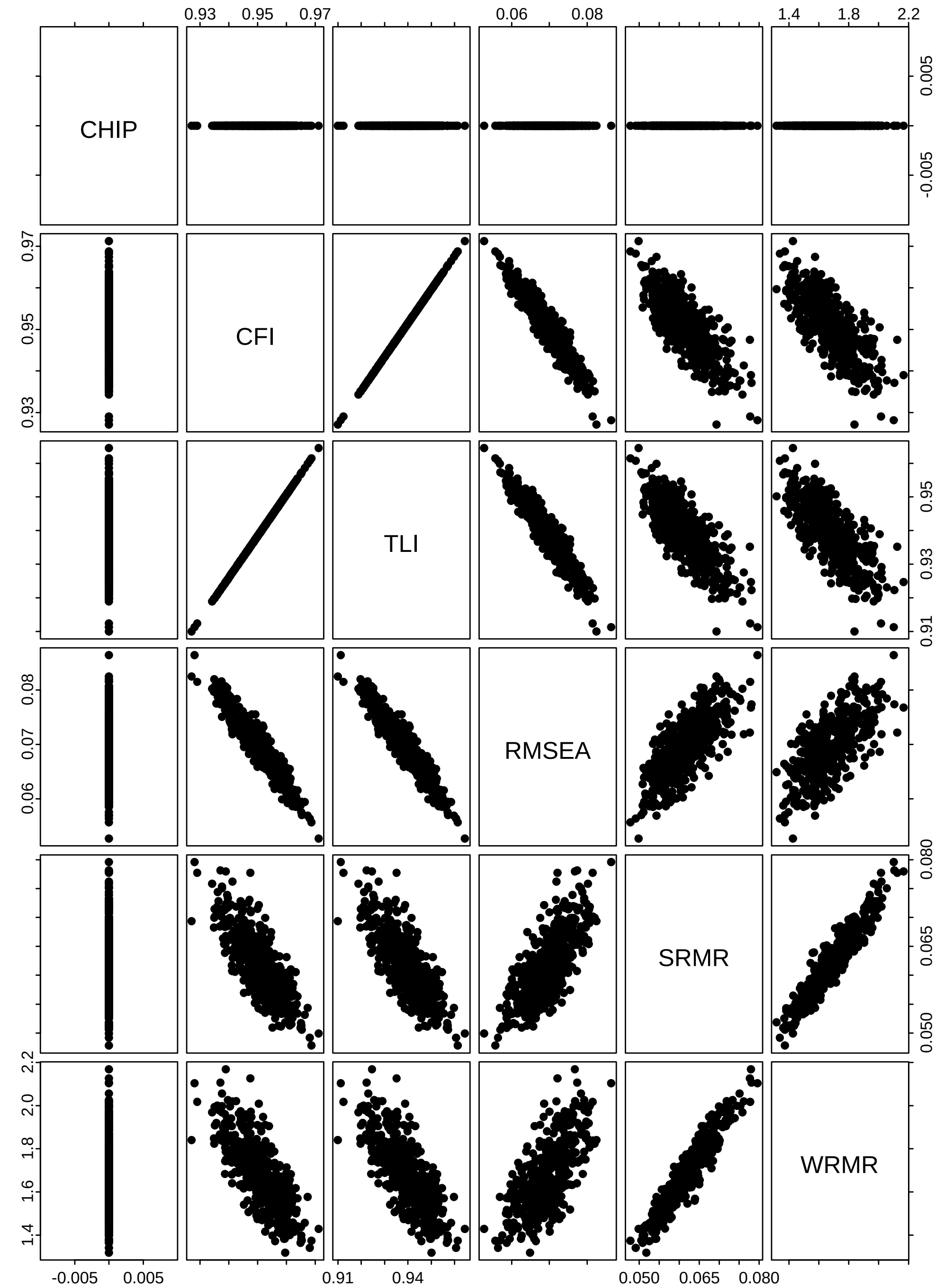


Table 4.1. Model and data conditions evaluated in the simulation studies

Conditions	Study 1:CFA	Study 2: MIMIC	Study3: Latent growth curve model
Model misspecification	True, misspecified, trivially misspecified	True, misspecified	True, misspecified
Type of outcomes	Normal, non-normal continuous, binary	Normal, non-normal continuous, binary	Normal
Sample sizes	100, 250, 500, 1000	100, 250, 500, 1000	100, 250, 500, 1000
Model specification	Misspecified factor covariances/ loadings	Misspecified factor loadings	Misspecified growth trend (five and eight time points)

Table 4.2. Parameter values for the CFA simple and complex models

Variables	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}
Factor Loadings for the simple model															
F1	.7	.7	.75	.8	.8	0	0	0	0	0	0	0	0	0	0
F2	0	0	0	0	0	.7	.7	.75	.8	.8	0	0	0	0	0
F3	0	0	0	0	0	0	0	0	0	0	.7	.7	.75	.8	.8
Error variances for the simple model															
	.51	.51	.44	.36	.36	.51	.51	.44	.36	.36	.51	.51	.44	.36	.36
Factor Loadings for the complex model															
F1	.7	.7	.75	.8	.8	0	0	0	0	0	0	0	0	0	0
F2	0	0	0	**.7**	0	.7	.7	.75	.8	.8	0	0	0	0	0
F3	**.7**	0	0	0	0	0	0	0	**.7**	0	.7	.7	.75	.8	.8
Factor variances and covariances															
	F1	F2	F3												
F1	1														
F2	.5	1													
F3	.4	.3	1												

Table 4.3.1. CFA model rejection rates of Chi-P at various cutoff values under normality

Cutoff Value	Simple Model					Complex Model				
	Sample Size					Sample Size				
	100		250	500	1000	100		250	500	1000
	ML	SB				ML	SB			
0.01										
True	4.6	6.4	2.0	1.2	1.4	3.8	7.8	1.6	1.0	1.0
Miss1	36.2	40.4	86.8	100.0	100.0	75.2	78.8	100.0	100.0	100.0
Miss2	50.6	54.6	99.2	100.0	100.0	98.8	99.0	100.0	100.0	100.0
-										
0.03										
True	11.6	13.6	4.8	4.0	3.4	10.8	12.4	4.2	4.6	2.4
Miss1	48.6	55.0	94.8	100.0	100.0	83.6	86.6	100.0	100.0	100.0
Miss2	64.0	71.0	100.0	100.0	100.0	99.6	99.8	100.0	100.0	100.0
0.04										
True	13.4	15.2	6.2	4.8	3.6	12.2	14.8	6.0	6.6	3.8
Miss1	53.8	61.0	95.8	100.0	100.0	85.6	88.4	100.0	100.0	100.0
Miss2	70.2	75.6	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0
0.05										
True	14.8	17.0	8.2	6.2	4.6	13.8	17.2	7.6	8.2	4.6
Miss1	58.0	64.4	96.8	100.0	100.0	87.8	89.4	100.0	100.0	100.0
Miss2	74.8	78.2	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0
0.06										
True	15.2	19.8	9.0	8.0	5.4	15.6	20.2	8.2	10.0	5.2
Miss1	62.6	65.6	98.2	100.0	100.0	88.8	90.6	100.0	100.0	100.0
Miss2	78.2	81.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.3.2. CFA model rejection rates of Chi-P at various cutoff values under moderate non-normality

Cutoff Value	Simple Model								Complex Model							
	Sample Size 100		250		500		1000		Sample Size 100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.01																
True	58.6	16.0	65.6	5.2	62.6	2.4	60.0	2.0	71.2	20.4	70.2	6.4	72.2	3.8	77.6	1.6
Miss1	86.0	47.4	99.0	79.2	100.0	99.6	100.0	100.0	97.8	81.0	100.0	98.4	100.0	100.0	100.0	100.0
Miss2	91.4	59.0	99.8	94.8	100.0	100.0	100.0	100.0	100.0	97.6	100.0	100.0	100.0	100.0	100.0	100.0
0.03																
True	71.8	27.0	77.6	10.8	76.2	5.0	76.6	4.0	81.6	34.0	79.6	12.8	80.8	7.6	86.0	4.4
Miss1	91.8	65.6	99.6	90.4	100.0	99.8	100.0	100.0	98.6	90.4	100.0	99.8	100.0	100.0	100.0	100.0
Miss2	96.0	75.6	100.0	97.8	100.0	100.0	100.0	100.0	100.0	99.2	100.0	100.0	100.0	100.0	100.0	100.0
0.04																
True	76.2	30.4	81.6	14.8	79.6	6.6	79.6	4.8	84.0	38.8	83.6	15.6	84.2	10.6	88.2	5.0
Miss1	94.2	71.0	99.6	91.4	100.0	99.8	100.0	100.0	99.2	91.8	100.0	99.8	100.0	100.0	100.0	100.0
Miss2	97.2	80.2	100.0	98.8	100.0	100.0	100.0	100.0	100.0	99.4	100.0	100.0	100.0	100.0	100.0	100.0
0.05																
True	78.4	33.6	83.8	16.8	82.8	8.6	82.0	6.4	86.0	42.4	85.6	17.6	85.4	12.2	89.4	6.4
Miss1	95.0	74.0	99.8	92.8	100.0	99.8	100.0	100.0	99.2	93.6	100.0	99.8	100.0	100.0	100.0	100.0
Miss2	97.8	83.0	100.0	99.2	100.0	100.0	100.0	100.0	100.0	99.4	100.0	100.0	100.0	100.0	100.0	100.0
0.06																
True	80.0	36.2	85.8	18.6	85.0	9.6	84.6	7.0	87.4	46.6	87.0	19.8	87.0	13.2	90.0	7.6
Miss1	95.8	76.6	99.8	93.4	100.0	99.8	100.0	100.0	99.2	94.0	100.0	100.0	100.0	100.0	100.0	100.0
Miss2	98.8	85.4	100.0	99.2	100.0	100.0	100.0	100.0	100.0	99.4	100.0	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.3.3. CFA model rejection rates of Chi-P at various cutoff values under severe non-normality

Cutoff Value	Simple Model								Complex Model							
	Sample Size								Sample Size							
	100		250		500		1000		100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.01																
True	87.8	44.2	92.8	16.0	95.2	7.6	98.2	3.4	94.4	50.6	96.2	21.8	99.2	7.2	99.6	6.0
Miss1	95.0	72.2	99.8	76.0	100.0	94.8	100.0	100.0	99.8	92.2	100.0	96.2	100.0	99.8	100.0	100.0
Miss2	97.2	80.2	100.0	90.8	100.0	99.8	100.0	100.0	100.0	99.4	100.0	100.0	100.0	100.0	100.0	100.0
0.03																
True	92.0	60.2	95.4	26.0	97.4	15.4	99.0	8.0	96.6	67.2	98.8	35.6	99.8	14.0	99.8	11.4
Miss1	97.8	82.4	99.8	87.0	100.0	98.2	100.0	100.0	100.0	96.2	100.0	98.2	100.0	99.8	100.0	100.0
Miss2	98.6	89.2	100.0	95.8	100.0	100.0	100.0	100.0	100.0	99.6	100.0	100.0	100.0	100.0	100.0	100.0
0.04																
True	93.4	65.2	96.4	30.8	98.0	19.8	99.0	11.6	97.2	72.0	99.4	38.8	100.0	16.6	99.8	13.8
Miss1	98.0	85.8	100.0	88.6	100.0	99.0	100.0	100.0	100.0	96.8	100.0	99.0	100.0	99.8	100.0	100.0
Miss2	98.8	90.6	100.0	97.2	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0
0.05																
True	94.2	68.4	97.4	33.4	98.4	22.8	99.4	14.0	97.6	74.2	99.4	42.6	100.0	18.6	100.0	16.6
Miss1	98.2	87.6	100.0	89.8	100.0	99.4	100.0	100.0	100.0	97.2	100.0	99.8	100.0	99.8	100.0	100.0
Miss2	99.2	91.8	100.0	97.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.06																
True	95.0	71.0	97.8	37.8	99.0	24.6	99.4	16.2	98.0	77.6	99.4	44.4	100.0	19.8	100.0	18.6
Miss1	98.6	88.0	100.0	90.6	100.0	99.4	100.0	100.0	100.0	98.0	100.0	99.8	100.0	100.0	100.0	100.0
Miss2	99.4	93.0	100.0	98.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.4.1. CFA model rejection rates of TLI at various cutoff values under normality

Cutoff Value	Simple Model					Complex Model				
	Sample Size					Sample Size				
	100		250	500	1000	100		250	500	1000
	ML	SB				ML	SB			
0.90										
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	3.0	3.8	0.0	0.0	0.0	13.6	17.0	0.2	0.0	0.0
Miss2	8.4	9.8	0.2	0.0	0.0	69.0	70.4	74.2	80.8	89.8
0.93										
True	0.4	0.6	0.0	0.0	0.0	0.4	0.4	0.0	0.0	0.0
Miss1	21.2	24.8	1.2	0.0	0.0	49.4	51.2	28.4	14.2	6.0
Miss2	35.0	38.0	7.6	0.4	0.0	93.2	94.4	100.0	100.0	100.0
0.94										
True	2.4	3.6	0.0	0.0	0.0	1.2	2.2	0.0	0.0	0.0
Miss1	33.6	36.6	4.6	0.0	0.0	61.2	66.4	53.4	51.0	50.4
Miss2	46.2	50.0	24.4	9.8	2.0	96.0	96.8	100.0	100.0	100.0
0.95										
True	7.8	9.6	0.0	0.0	0.0	3.6	4.2	0.0	0.0	0.0
Miss1	44.4	48.8	17.6	4.2	0.2	74.6	76.8	77.8	87.6	95.6
Miss2	59.6	64.8	51.4	50.2	48.0	98.6	98.6	100.0	100.0	100.0
0.96										
True	14.0	16.6	0.0	0.0	0.0	9.8	10.4	0.0	0.0	0.0
Miss1	55.8	61.2	43.6	34.4	26.6	83.4	85.6	93.8	99.4	99.8
Miss2	75.0	77.6	79.4	88.4	95.6	99.6	99.6	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.4.2. CFA model rejection rates of TLI at various cutoff values under moderate non-normality

Cutoff Value	Simple Model								Complex Model							
	Sample Size 100		250		500		1000		Sample Size 100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.90																
True	13.4	3.8	0.0	0.0	0.0	0.0	0.0	0.0	11.0	2.2	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	44.6	23.0	2.8	0.4	0.0	0.0	0.0	0.0	68.4	48.6	20.2	7.4	1.2	0.2	0.0	0.0
Miss2	56.4	32.0	9.6	1.4	0.2	0.0	0.0	0.0	95.6	85.4	95.8	84.6	94.0	89.2	96.4	97.8
0.93																
True	40.8	17.0	0.2	0.0	0.0	0.0	0.0	0.0	36.6	14.6	0.2	0.0	0.0	0.0	0.0	0.0
Miss1	79.2	52.2	36.6	12.2	4.8	0.6	0.0	0.0	92.2	75.4	78.4	53.0	59.6	38.0	32.8	34.0
Miss2	85.4	64.4	63.4	28.0	24.8	7.6	2.4	1.4	99.4	97.0	99.8	99.2	100.0	100.0	100.0	100.0
0.94																
True	52.8	26.0	2.0	0.2	0.0	0.0	0.0	0.0	52.2	22.4	1.0	0.2	0.0	0.0	0.0	0.0
Miss1	84.6	66.0	58.8	25.6	17.0	4.8	1.0	0.2	95.4	84.4	93.4	70.8	86.2	72.4	83.4	78.8
Miss2	90.8	75.2	83.6	50.0	56.4	30.2	25.0	16.8	100.0	98.6	100.0	99.8	100.0	100.0	100.0	100.0
0.95																
True	66.0	34.2	7.0	0.2	0.0	0.0	0.0	0.0	67.4	31.6	4.6	0.4	0.0	0.0	0.0	0.0
Miss1	91.2	74.4	79.0	42.2	49.0	20.4	15.0	8.8	97.4	91.2	98.2	86.2	98.8	92.4	98.8	97.8
Miss2	94.6	82.4	94.0	73.8	86.0	66.2	79.0	70.2	100.0	99.4	100.0	100.0	100.0	100.0	100.0	100.0
0.96																
True	77.2	45.0	18.0	2.8	0.2	0.0	0.0	0.0	79.2	47.2	15.0	1.4	0.0	0.0	0.4	0.0
Miss1	95.6	82.6	92.8	68.8	83.2	55.8	66.6	50.8	99.0	94.4	99.8	95.8	100.0	99.0	100.0	100.0
Miss2	98.6	89.6	99.0	90.8	98.6	91.6	99.6	98.6	100.0	99.6	100.0	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.4.3. CFA model rejection rates of TLI at various cutoff values under severe non-normality

Cutoff Value	Simple Model								Complex Model							
	Sample Size 100		250		500		1000		Sample Size 100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.90																
True	55.8	29.6	3.0	0.4	0.0	0.0	0.0	0.0	52.6	28.2	2.0	0.2	0.0	0.0	0.0	0.0
Miss1	77.6	57.8	32.8	6.4	2.0	0.2	0.0	0.0	91.4	76.2	63.0	34.2	21.2	7.0	3.2	2.8
Miss2	82.0	66.0	50.6	13.2	11.0	0.8	0.0	0.0	99.4	95.4	98.2	91.6	98.0	93.2	97.4	97.0
0.93																
True	79.0	55.2	22.0	4.2	0.4	0.0	0.0	0.0	82.6	52.6	23.4	3.4	0.4	0.0	0.0	0.0
Miss1	91.2	79.6	75.0	36.2	37.8	8.8	3.0	0.4	98.2	92.8	94.8	74.2	88.0	59.0	71.8	61.8
Miss2	94.2	86.2	86.2	58.4	68.4	29.4	23.8	9.4	100.0	99.6	100.0	99.4	100.0	99.6	100.0	100.0
0.94																
True	84.8	64.8	37.8	5.6	2.2	0.0	0.0	0.0	88.8	62.8	36.8	6.4	1.0	0.2	0.0	0.0
Miss1	93.4	85.4	85.4	52.6	64.4	23.6	17.2	4.4	99.2	96.0	98.8	82.8	97.8	80.4	94.8	88.2
Miss2	96.4	89.8	92.6	72.0	87.8	56.8	63.6	41.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	100.0
0.95																
True	89.8	73.2	58.4	10.6	8.4	0.6	0.0	0.0	93.0	74.2	57.2	13.4	4.6	0.4	0.0	0.0
Miss1	97.2	90.2	93.0	67.6	86.0	47.0	54.6	26.6	99.8	97.0	100.0	91.4	99.4	94.0	100.0	97.4
Miss2	98.4	93.8	97.8	83.8	98.0	80.0	93.4	83.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.96																
True	93.2	81.0	72.2	19.4	22.2	1.4	0.0	0.0	96.2	83.2	76.2	22.6	14.2	1.4	0.4	0.0
Miss1	98.6	94.0	97.2	80.6	97.4	74.2	92.0	70.2	100.0	99.0	100.0	96.6	100.0	98.8	100.0	100.0
Miss2	99.0	97.4	99.8	94.0	99.8	95.4	99.8	98.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.5.1. CFA model rejection rates of CFI at various cutoff values under normality

Cutoff Value	Simple Model					Complex Model				
	Sample Size					Sample Size				
	100		250	500	1000	100		250	500	1000
	ML	SB				ML	SB			
0.90										
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	0.2	0.6	0.0	0.0	0.0	4.4	4.4	0.0	0.0	0.0
Miss2	1.0	1.8	0.0	0.0	0.0	41.2	44.2	27.6	14.0	7.6
0.93										
True	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	10.0	12.2	0.0	0.0	0.0	27.4	30.2	3.6	0.6	0.0
Miss2	22.0	25.0	0.6	0.0	0.0	82.0	84.6	95.4	98.8	99.4
0.94										
True	0.4	0.4	0.0	0.0	0.0	0.4	0.2	0.0	0.0	0.0
Miss1	19.2	22.6	0.6	0.0	0.0	44.4	47.2	18.4	7.8	1.4
Miss2	33.6	37.0	6.6	0.0	0.0	92.2	93.0	99.8	100.0	100.0
0.95										
True	2.4	3.4	0.0	0.0	0.0	1.0	1.8	0.0	0.0	0.0
Miss1	34.0	36.6	4.6	0.2	0.0	58.0	63.0	48.4	42.8	39.6
Miss2	47.2	51.2	26.8	11.6	3.0	95.6	96.4	100.0	100.0	100.0
0.96										
True	8.6	10.2	0.0	0.0	0.0	3.6	4.2	0.0	0.0	0.0
Miss1	48.6	52.2	23.2	8.2	1.6	75.4	77.4	79.2	89.0	96.2
Miss2	63.6	68.8	59.8	63.2	65.2	98.6	99.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.5.2. CFA model rejection rates of CFI at various cutoff values under moderate non-normality

Cutoff Value	Simple Model								Complex Model							
	Sample Size 100		250		500		1000		Sample Size 100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.90																
True	4.6	1.8	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	28.4	9.4	0.2	0.0	0.0	0.0	0.0	0.0	43.4	21.0	1.8	0.6	0.0	0.0	0.0	0.0
Miss2	39.4	16.6	0.8	0.0	0.0	0.0	0.0	0.0	86.0	70.4	71.6	54.0	48.2	40.2	25.8	35.6
0.93																
True	25.8	7.8	0.0	0.0	0.0	0.0	0.0	0.0	19.0	5.4	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	64.0	38.0	13.8	3.0	0.4	0.0	0.0	0.0	81.4	60.4	45.4	24.0	13.0	6.4	1.6	1.6
Miss2	74.6	50.4	37.0	11.4	4.2	0.8	0.0	0.0	98.6	92.6	99.6	96.4	99.8	99.6	99.8	100.0
0.94																
True	37.2	15.8	0.2	0.0	0.0	0.0	0.0	0.0	30.2	11.4	0.2	0.0	0.0	0.0	0.0	0.0
Miss1	77.0	50.2	33.0	10.6	3.4	0.6	0.0	0.0	89.8	72.6	70.6	44.0	43.6	27.4	17.2	17.0
Miss2	85.4	63.6	61.6	26.2	22.4	7.0	2.0	0.6	99.4	96.4	99.8	99.2	100.0	100.0	100.0	100.0
0.95																
True	52.8	25.6	2.0	0.2	0.0	0.0	0.0	0.0	47.6	20.8	0.8	0.2	0.0	0.0	0.0	0.0
Miss1	84.8	66.0	60.0	26.0	18.2	4.8	1.0	0.4	94.8	83.2	90.6	67.2	82.0	67.2	75.6	72.2
Miss2	91.0	76.4	85.4	53.6	60.8	33.4	30.6	20.8	100.0	98.6	100.0	99.8	100.0	100.0	100.0	100.0
0.96																
True	68.0	35.2	7.8	0.4	0.0	0.0	0.0	0.0	67.4	31.6	4.6	0.4	0.0	0.0	0.0	0.0
Miss1	91.6	76.0	85.0	49.6	57.4	27.0	21.6	14.0	97.4	91.4	98.2	87.4	99.2	93.2	99.2	97.8
Miss2	95.6	85.6	96.2	80.2	93.0	75.0	90.6	83.8	100.0	99.4	100.0	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.5.3. CFA model rejection rates of CFI at various cutoff values under severe non-normality

Cutoff Value	Simple Model								Complex Model							
	Sample Size 100		250		500		1000		Sample Size 100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.90																
True	36.4	17.6	0.4	0.2	0.0	0.0	0.0	0.0	32.0	13.2	0.6	0.0	0.0	0.0	0.0	0.0
Miss1	63.8	39.8	14.6	2.0	0.4	0.0	0.0	0.0	77.6	54.2	30.0	11.0	1.4	0.6	0.0	0.0
Miss2	73.0	50.2	25.0	4.0	1.2	0.2	0.0	0.0	97.2	88.8	90.2	72.4	79.4	61.2	60.4	62.6
0.93																
True	72.0	42.4	9.6	1.6	0.0	0.0	0.0	0.0	65.6	37.0	7.2	0.8	0.0	0.0	0.0	0.0
Miss1	85.2	72.8	56.4	19.0	13.6	1.4	0.0	0.0	96.4	85.6	83.0	51.8	48.8	24.4	18.6	13.0
Miss2	90.0	79.4	74.4	34.2	36.4	10.0	2.6	0.8	100.0	98.6	99.8	96.8	99.6	98.4	99.8	100.0
0.94																
True	77.4	53.6	19.2	3.8	0.2	0.0	0.0	0.0	78.0	48.0	16.2	2.0	0.2	0.0	0.0	0.0
Miss1	90.6	78.6	73.4	34.4	33.6	7.4	2.2	0.4	97.6	92.0	92.6	69.4	79.8	49.0	60.4	47.6
Miss2	94.2	86.0	85.6	57.8	67.6	28.6	22.0	8.8	100.0	99.4	100.0	99.2	100.0	99.6	100.0	100.0
0.95																
True	84.4	64.8	37.4	5.6	2.0	0.0	0.0	0.0	87.4	60.8	33.4	5.2	0.8	0.2	0.0	0.0
Miss1	93.6	85.4	85.6	53.6	64.8	24.2	17.2	4.6	99.0	95.6	98.6	82.4	96.2	77.4	92.8	82.8
Miss2	97.2	90.2	93.4	72.8	89.4	58.6	68.4	45.4	100.0	99.8	100.0	99.8	100.0	99.8	100.0	100.0
0.96																
True	90.0	74.6	61.8	11.6	11.2	0.0	0.0	0.0	93.0	74.2	57.2	13.4	4.6	0.4	0.0	0.0
Miss1	97.4	91.6	93.6	70.8	89.4	54.6	64.4	33.8	99.8	97.0	100.0	91.6	99.4	94.6	100.0	97.6
Miss2	98.8	94.8	98.6	86.8	98.6	86.6	96.4	90.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.6.1. CFA model rejection rates of RMSEA at various cutoff values under normality

Cutoff Value	Simple Model					Complex Model				
	Sample Size 100 ML	100 SB	250	500	1000	Sample Size 100 ML	100 SB	250	500	1000
0.045										
True	23.2	28.0	0.4	0.0	0.0	23.0	29.0	0.6	0.0	0.0
Miss1	69.8	74.0	72.6	79.0	87.8	92.4	94.0	100.0	100.0	100.0
Miss2	84.4	87.4	93.2	98.4	99.8	100.0	100.0	100.0	100.0	100.0
0.05										
True	15.0	18.4	0.0	0.0	0.0	15.8	20.2	0.2	0.0	0.0
Miss1	61.2	65.0	51.4	43.6	37.6	88.8	90.4	99.2	100.0	100.0
Miss2	76.2	79.6	81.8	90.2	97.4	100.0	100.0	100.0	100.0	100.0
0.055										
True	11.6	14.0	0.0	0.0	0.0	11.0	13.6	0.0	0.0	0.0
Miss1	48.6	55.0	27.4	11.8	3.2	84.4	87.0	96.4	99.8	100.0
Miss2	63.6	70.4	61.4	66.4	70.4	99.6	99.8	100.0	100.0	100.0
0.06										
True	6.0	8.8	0.0	0.0	0.0	7.0	9.2	0.0	0.0	0.0
Miss1	40.2	43.4	10.8	1.4	0.0	79.6	81.8	84.8	96.2	99.4
Miss2	54.2	59.6	36.8	27.0	16.0	99.4	99.4	100.0	100.0	100.0
0.07										
True	0.8	1.4	0.0	0.0	0.0	1.2	1.4	0.0	0.0	0.0
Miss1	18.6	21.8	0.2	0.0	0.0	60.2	63.8	49.8	47.0	43.8
Miss2	30.2	34.8	4.4	0.2	0.0	95.2	97.0	100.0	100.0	100.0
0.08										
True	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.0
Miss1	6.6	7.6	0.0	0.0	0.0	34.4	41.4	9.4	2.4	0.0
Miss2	12.6	16.0	0.2	0.0	0.0	87.8	89.6	97.4	99.8	100.0
0.09										
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	0.0	1.2	0.0	0.0	0.0	14.6	19.0	0.2	0.0	0.0
Miss2	2.0	3.6	0.0	0.0	0.0	66.2	70.8	71.2	77.6	87.6
0.10										
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	0.0	0.0	0.0	0.0	0.0	3.4	4.4	0.0	0.0	0.0
Miss2	0.0	0.0	0.0	0.0	0.0	38.8	45.0	22.8	11.2	3.6

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.6.2. CFA model rejection rates of RMSEA at various cutoff values under moderate non-normality

Cutoff Value	Simple Model								Complex Model							
	Sample Size 100		250		500		1000		Sample Size 100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.045																
True	85.4	47.4	44.6	1.6	2.4	0.0	0.0	0.0	90.4	57.8	51.2	2.6	6.4	0.0	0.0	0.0
Miss1	97.4	82.8	96.8	55.8	96.6	27.4	97.8	11.4	100.0	96.2	100.0	95.0	100.0	97.8	100.0	100.0
Miss2	99.4	89.0	99.4	83.8	99.8	71.0	100.0	71.2	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0
0.05																
True	79.8	35.6	24.4	0.2	0.2	0.0	0.0	0.0	87.4	46.6	35.0	0.6	0.8	0.0	0.0	0.0
Miss1	95.4	76.0	94.4	33.2	86.6	6.6	75.2	0.2	99.2	93.8	100.0	87.0	100.0	90.0	100.0	92.4
Miss2	98.4	85.0	98.2	61.6	99.0	32.8	99.2	16.4	100.0	99.4	100.0	100.0	100.0	100.0	100.0	100.0
0.055																
True	71.8	27.2	12.4	0.0	0.0	0.0	0.0	0.0	82.0	35.2	21.8	0.4	0.0	0.0	0.0	0.0
Miss1	91.8	65.6	86.8	17.6	64.2	1.0	31.6	0.0	98.8	91.0	99.8	73.6	100.0	64.8	100.0	56.8
Miss2	95.8	75.2	95.4	38.0	91.8	7.0	91.6	0.4	100.0	99.2	100.0	99.8	100.0	100.0	100.0	100.0
0.06																
True	63.0	19.2	5.8	0.0	0.0	0.0	0.0	0.0	76.0	25.0	10.6	0.0	0.0	0.0	0.0	0.0
Miss1	88.6	52.6	72.8	4.2	35.2	0.2	7.6	0.0	98.4	85.6	99.2	54.4	100.0	30.0	100.0	14.8
Miss2	91.8	61.8	90.8	14.2	73.2	1.0	52.8	0.0	100.0	98.2	100.0	99.2	100.0	100.0	100.0	100.0
0.07																
True	41.0	7.0	0.4	0.0	0.0	0.0	0.0	0.0	55.2	9.8	2.0	0.0	0.0	0.0	0.0	0.0
Miss1	75.8	30.2	33.0	0.2	3.2	0.0	0.0	0.0	96.0	67.0	92.0	15.4	86.4	1.0	79.4	0.0
Miss2	83.2	38.6	56.6	0.4	17.0	0.2	1.0	0.0	99.6	94.8	100.0	88.2	100.0	86.2	100.0	86.4
0.08																
True	22.2	1.8	0.0	0.0	0.0	0.0	0.0	0.0	32.4	2.4	0.2	0.0	0.0	0.0	0.0	0.0
Miss1	55.4	10.6	6.8	0.0	0.0	0.0	0.0	0.0	87.4	42.8	63.0	1.2	29.8	0.0	7.6	0.0
Miss2	62.4	16.0	17.6	0.0	1.4	0.0	0.0	0.0	98.6	82.8	99.8	54.0	100.0	25.6	100.0	9.8
0.09																
True	8.4	0.6	0.0	0.0	0.0	0.0	0.0	0.0	15.6	0.8	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	32.6	2.8	0.4	0.0	0.0	0.0	0.0	0.0	74.4	19.0	25.4	0.0	2.8	0.0	0.0	0.0
Miss2	41.6	5.0	2.6	0.0	0.0	0.0	0.0	0.0	95.2	60.4	95.2	12.0	92.6	0.6	96.0	0.0
0.10																
True	2.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	15.4	0.6	0.0	0.0	0.0	0.0	0.0	0.0	49.4	5.8	5.4	0.0	0.0	0.0	0.0	0.0
Miss2	22.2	0.8	0.0	0.0	0.0	0.0	0.0	0.0	85.6	34.0	72.6	1.0	47.2	0.0	22.8	0.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.6.3. CFA model rejection rates of RMSEA at various cutoff values under severe non-normality

Cutoff Value	Simple Model								Complex Model							
	Sample Size 100		250		500		1000		Sample Size 100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.045																
True	96.6	78.0	86.2	7.0	52.0	0.0	2.0	0.0	98.6	84.6	92.6	11.4	64.8	0.0	12.6	0.0
Miss1	99.2	92.8	99.4	54.8	99.4	10.6	99.8	0.0	100.0	98.8	100.0	89.8	100.0	77.4	100.0	64.2
Miss2	99.8	95.8	100.0	74.4	100.0	36.4	100.0	5.4	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0
0.05																
True	94.6	70.4	77.6	2.4	31.0	0.0	0.2	0.0	98.0	77.6	87.2	5.4	45.2	0.0	2.6	0.0
Miss1	98.6	87.8	98.0	32.0	98.4	1.2	94.4	0.0	100.0	98.0	100.0	80.8	100.0	46.6	100.0	18.0
Miss2	99.4	92.6	99.6	51.6	100.0	8.6	100.0	0.2	100.0	100.0	100.0	99.4	100.0	99.4	100.0	98.8
0.055																
True	92.0	60.8	67.8	0.6	17.4	0.0	0.0	0.0	96.6	68.2	81.2	1.2	24.8	0.0	1.0	0.0
Miss1	97.8	82.4	94.6	15.4	92.8	0.0	74.2	0.0	100.0	96.4	100.0	64.4	100.0	19.2	100.0	2.0
Miss2	98.4	89.2	98.0	27.6	99.0	0.6	96.6	0.0	100.0	99.8	100.0	97.6	100.0	96.8	100.0	91.2
0.06																
True	89.6	49.6	55.2	0.4	8.6	0.0	0.0	0.0	95.6	56.6	70.4	0.0	13.2	0.0	0.2	0.0
Miss1	96.8	75.6	89.8	6.6	80.0	0.0	36.8	0.0	99.8	94.0	100.0	42.0	99.6	3.0	100.0	0.2
Miss2	97.6	82.4	95.8	12.6	95.8	0.0	84.2	0.0	100.0	99.6	100.0	93.4	100.0	79.8	100.0	59.8
0.07																
True	81.0	30.6	25.2	0.0	0.8	0.0	0.0	0.0	89.2	35.4	44.0	0.0	2.6	0.0	0.0	0.0
Miss1	91.2	52.8	75.2	0.4	37.2	0.0	1.6	0.0	99.2	82.6	98.8	11.2	98.2	0.0	94.0	0.0
Miss2	93.4	62.4	84.8	0.6	65.0	0.0	15.4	0.0	100.0	97.0	100.0	66.2	100.0	20.0	100.0	2.6
0.08																
True	69.0	14.0	8.2	0.0	0.0	0.0	0.0	0.0	80.6	17.6	20.4	0.0	0.6	0.0	0.0	0.0
Miss1	83.4	31.8	46.0	0.0	7.8	0.0	0.0	0.0	97.2	61.2	90.6	0.4	70.8	0.0	45.0	0.0
Miss2	86.4	37.2	60.6	0.0	21.6	0.0	0.8	0.0	99.8	89.2	100.0	22.4	99.8	0.8	100.0	0.0
0.09																
True	50.2	4.0	2.6	0.0	0.0	0.0	0.0	0.0	66.4	8.6	7.2	0.0	0.0	0.0	0.0	0.0
Miss1	70.2	13.6	20.6	0.0	0.8	0.0	0.0	0.0	93.4	37.0	70.0	0.0	29.0	0.0	5.8	0.0
Miss2	75.8	17.8	32.4	0.0	2.4	0.0	0.0	0.0	99.0	73.2	98.8	2.8	98.0	0.0	97.0	0.0
0.10																
True	34.0	1.6	0.6	0.0	0.0	0.0	0.0	0.0	48.0	2.8	1.4	0.0	0.0	0.0	0.0	0.0
Miss1	55.4	3.4	6.4	0.0	0.2	0.0	0.0	0.0	85.6	16.4	42.6	0.0	5.4	0.0	0.4	0.0
Miss2	61.6	4.2	10.6	0.0	0.2	0.0	0.0	0.0	97.8	50.0	89.4	0.0	79.6	0.0	56.6	0.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 4.7.1. CFA model rejection rates of SRMR at various cutoff values under normality

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.045								
True	97.8	2.8	0.0	0.0	90.6	0.2	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.05								
True	89.0	0.2	0.0	0.0	67.0	0.0	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	99.8	99.0	99.4	99.2
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.055								
True	68.6	0.0	0.0	0.0	38.6	0.0	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	99.0	93.4	88.8	84.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.06								
True	42.4	0.0	0.0	0.0	15.6	0.0	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	96.8	79.6	60.2	42.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.07								
True	5.0	0.0	0.0	0.0	2.2	0.0	0.0	0.0
Miss1	99.8	100.0	100.0	100.0	76.6	31.8	7.4	0.8
Miss2	100.0	100.0	100.0	100.0	96.8	87.0	78.6	68.4
0.08								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	99.6	100.0	100.0	100.0	42.2	2.8	0.0	0.0
Miss2	100.0	100.0	100.0	100.0	75.4	37.0	10.6	0.8
0.09								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	99.4	100.0	100.0	100.0	16.6	0.2	0.0	0.0
Miss2	100.0	100.0	100.0	100.0	41.0	2.4	0.0	0.0
0.10								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	98.0	99.4	99.6	100.0	3.8	0.0	0.0	0.0
Miss2	99.6	100.0	100.0	100.0	14.6	0.4	0.0	0.0
0.11								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	94.4	97.4	98.8	99.8	1.2	0.0	0.0	0.0
Miss2	99.0	100.0	100.0	100.0	3.0	0.0	0.0	0.0

Table 4.7.2. CFA model rejection rates of SRMR at various cutoff values under moderate non-normality

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.045								
True	100.0	34.4	0.0	0.0	99.2	12.4	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.05								
True	99.0	7.2	0.0	0.0	92.6	3.0	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	99.4	98.0	98.4
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.055								
True	92.6	0.6	0.0	0.0	80.8	0.6	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	97.2	91.6	86.2
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.06								
True	76.4	0.2	0.0	0.0	56.0	0.0	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	98.6	88.0	69.0	55.8
Miss2	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0
0.07								
True	36.2	0.0	0.0	0.0	14.0	0.0	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	90.8	48.4	19.8	3.6
Miss2	100.0	100.0	100.0	100.0	99.2	92.2	81.4	75.2
0.08								
True	5.6	0.0	0.0	0.0	2.2	0.0	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	60.8	11.6	0.6	0.0
Miss2	100.0	100.0	100.0	100.0	88.0	57.0	23.0	6.0
0.09								
True	1.2	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Miss1	99.6	99.4	100.0	100.0	31.0	2.6	0.0	0.0
Miss2	100.0	100.0	100.0	100.0	58.2	13.8	0.6	0.2
0.10								
True	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	97.4	99.0	99.4	100.0	12.8	0.2	0.0	0.0
Miss2	100.0	100.0	100.0	100.0	32.2	1.8	0.0	0.0
0.11								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	95.4	97.2	97.4	99.4	3.6	0.0	0.0	0.0
Miss2	99.2	100.0	100.0	100.0	11.4	0.0	0.0	0.0

Table 4.7.3. CFA model rejection rates of SRMR at various cutoff values under severe non-normality

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.045								
True	100.0	72.0	3.2	0.0	100.0	52.0	0.8	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	99.8	99.8	99.8
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.05								
True	99.8	42.6	0.6	0.0	97.8	21.8	0.2	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	98.8	98.8	98.2
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.055								
True	98.4	17.6	0.0	0.0	92.2	7.2	0.2	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	96.6	92.2	88.8
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.06								
True	93.4	7.0	0.0	0.0	79.0	2.0	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	99.4	90.0	71.0	61.0
Miss2	100.0	100.0	100.0	100.0	100.0	99.6	99.6	99.4
0.07								
True	65.0	0.4	0.0	0.0	39.0	0.4	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	93.6	60.8	26.8	12.6
Miss2	100.0	100.0	100.0	100.0	99.2	94.0	84.8	76.8
0.08								
True	29.8	0.0	0.0	0.0	12.0	0.0	0.0	0.0
Miss1	99.8	100.0	100.0	100.0	72.8	27.2	5.2	0.8
Miss2	100.0	100.0	100.0	100.0	92.0	64.4	33.0	15.2
0.09								
True	7.6	0.0	0.0	0.0	3.6	0.0	0.0	0.0
Miss1	99.4	99.4	100.0	100.0	44.4	6.6	0.8	0.0
Miss2	100.0	100.0	100.0	100.0	74.8	29.4	5.8	0.6
0.10								
True	2.6	0.0	0.0	0.0	1.0	0.0	0.0	0.0
Miss1	98.4	99.2	100.0	100.0	22.2	1.2	0.0	0.0
Miss2	99.8	100.0	100.0	100.0	48.0	6.8	0.4	0.0
0.11								
True	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0
Miss1	94.4	95.6	98.8	99.2	8.4	0.0	0.0	0.0
Miss2	99.6	100.0	100.0	100.0	24.0	1.2	0.0	0.0

Table 4.8.1. CFA model rejection rates of WRMR at various cutoff values under normality

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.60								
True	96.2	98.2	97.8	98.0	86.4	84.2	79.6	79.8
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.70								
True	77.8	73.8	73.2	70.6	47.8	44.8	39.6	35.2
Miss1	100.0	100.0	100.0	100.0	99.2	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.80								
True	36.8	31.6	28.8	23.6	13.8	11.6	8.4	6.0
Miss1	100.0	100.0	100.0	100.0	93.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	99.0	100.0	100.0	100.0
0.90								
True	9.2	6.0	5.0	5.2	3.6	0.6	0.6	0.2
Miss1	100.0	100.0	100.0	100.0	74.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	94.8	100.0	100.0	100.0
0.95								
True	3.6	2.2	1.8	2.8	1.4	0.2	0.0	0.0
Miss1	99.8	100.0	100.0	100.0	60.2	99.6	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	89.2	100.0	100.0	100.0
1.00								
True	1.4	0.8	0.4	1.6	0.2	0.0	0.0	0.0
Miss1	99.8	100.0	100.0	100.0	47.6	97.8	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	77.6	100.0	100.0	100.0
1.10								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	99.6	100.0	100.0	100.0	22.4	88.6	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	52.4	100.0	100.0	100.0

Table 4.8.2. CFA model rejection rates of WRMR at various cutoff values under moderate non-normality

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.60								
True	97.8	97.2	96.4	97.2	91.6	83.8	78.6	75.4
Miss1	100.0	100.0	100.0	100.0	99.2	99.8	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0
0.70								
True	86.8	83.0	72.6	66.8	65.4	52.4	43.4	33.2
Miss1	100.0	100.0	100.0	100.0	95.6	99.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	99.2	100.0	100.0	100.0
0.80								
True	59.2	45.6	35.6	27.8	34.6	19.4	10.0	7.4
Miss1	100.0	100.0	100.0	100.0	87.2	97.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	95.2	99.4	100.0	100.0
0.90								
True	31.0	17.6	8.8	6.0	12.8	3.8	1.6	0.6
Miss1	100.0	100.0	100.0	100.0	66.0	89.4	99.4	100.0
Miss2	100.0	100.0	100.0	100.0	84.6	98.0	100.0	100.0
0.95								
True	18.4	6.8	3.2	1.8	6.2	1.4	0.2	0.0
Miss1	100.0	100.0	100.0	100.0	54.8	84.8	98.2	100.0
Miss2	100.0	100.0	100.0	100.0	77.8	97.2	100.0	100.0
1.00								
True	9.0	2.4	0.8	0.6	2.6	1.0	0.2	0.0
Miss1	100.0	100.0	100.0	100.0	42.2	76.0	95.8	100.0
Miss2	100.0	100.0	100.0	100.0	67.6	95.2	100.0	100.0
1.10								
True	2.4	0.0	0.2	0.0	0.4	0.4	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	22.0	55.0	88.0	100.0
Miss2	100.0	100.0	100.0	100.0	45.4	86.6	99.4	100.0

Table 4.8.3. CFA model rejection rates of WRMR at various cutoff values under severe non-normality

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.60								
True	98.0	98.4	97.0	95.0	94.4	90.2	83.4	78.4
Miss1	100.0	100.0	100.0	100.0	98.2	99.8	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	99.4	100.0	100.0	100.0
0.70								
True	95.0	90.0	83.0	75.8	77.6	68.4	48.0	37.6
Miss1	100.0	100.0	100.0	100.0	94.6	98.4	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	97.4	99.6	100.0	100.0
0.80								
True	81.4	64.4	49.6	43.0	52.6	41.4	19.8	13.0
Miss1	100.0	100.0	100.0	100.0	87.2	93.6	99.6	100.0
Miss2	100.0	100.0	100.0	100.0	92.4	99.0	100.0	100.0
0.90								
True	57.0	34.8	23.8	14.0	29.4	16.8	5.8	2.6
Miss1	100.0	100.0	100.0	100.0	70.6	85.8	94.8	99.6
Miss2	100.0	100.0	100.0	100.0	83.6	96.0	100.0	100.0
0.95								
True	43.2	26.0	13.6	7.8	22.2	8.2	2.0	0.8
Miss1	100.0	100.0	100.0	100.0	61.0	79.4	91.0	99.2
Miss2	100.0	100.0	100.0	100.0	79.0	93.8	99.8	100.0
1.00								
True	29.6	15.2	7.8	2.2	14.0	4.2	1.0	0.4
Miss1	99.8	100.0	100.0	100.0	52.6	69.8	87.0	98.0
Miss2	100.0	100.0	100.0	100.0	73.2	90.0	98.4	100.0
1.10								
True	12.8	4.4	1.8	0.0	6.2	0.2	0.2	0.0
Miss1	99.2	100.0	100.0	100.0	32.2	46.4	68.8	94.6
Miss2	100.0	100.0	100.0	100.0	56.0	78.2	93.8	100.0

Table 4.9.1. CFA model rejection rates of Chi-P at various cutoff values for binary equal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.01								
True	0.8	1.6	0.6	1.0	1.0	0.8	0.4	1.6
Miss1	75.4	99.8	100.0	100.0	11.0	64.6	98.4	100.0
Miss2	88.2	100.0	100.0	100.0	27.2	95.4	100.0	100.0
0.03								
True	3.2	3.4	1.4	2.4	2.8	1.6	2.2	3.2
Miss1	84.4	100.0	100.0	100.0	24.0	77.8	99.2	100.0
Miss2	93.4	100.0	100.0	100.0	45.4	98.8	100.0	100.0
0.04								
True	4.0	3.8	2.2	2.8	4.2	2.8	3.4	4.4
Miss1	85.4	100.0	100.0	100.0	29.2	80.6	99.2	100.0
Miss2	94.4	100.0	100.0	100.0	51.4	99.4	100.0	100.0
0.05								
True	5.0	5.0	2.4	3.6	5.0	3.0	4.4	4.8
Miss1	86.6	100.0	100.0	100.0	32.2	84.6	99.6	100.0
Miss2	95.0	100.0	100.0	100.0	57.0	99.8	100.0	100.0
0.06								
True	6.6	6.0	4.0	5.4	6.0	3.8	4.4	6.0
Miss1	88.0	99.8	100.0	100.0	35.6	85.8	99.6	100.0
Miss2	95.4	100.0	100.0	100.0	61.0	100.0	100.0	100.0

Table 4.9.2. CFA model rejection rates of Chi-P at various cutoff values for binary unequal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100 [a]	250	500	1000
0.01								
True	3.6	1.2	1.0	0.8	4.6	1.6	0.8	1.2
Miss1	74.6	98.8	100.0	100.0	15.5	47.2	96.0	100.0
Miss2	88.2	99.6	100.0	100.0	25.4	87.2	100.0	100.0
0.03								
True	9.4	3.0	2.8	3.8	7.4	3.2	1.6	3.0
Miss1	81.6	99.4	100.0	100.0	27.4	63.4	98.2	100.0
Miss2	94.2	99.6	100.0	100.0	43.7	93.0	100.0	100.0
0.04								
True	11.6	3.8	3.8	4.2	9.0	4.2	2.2	3.2
Miss1	84.6	99.6	100.0	100.0	31.6	66.0	98.2	100.0
Miss2	94.8	99.6	100.0	100.0	48.1	95.0	100.0	100.0
0.05								
True	14.0	4.4	5.0	5.0	11.0	5.4	3.0	4.4
Miss1	85.4	99.6	100.0	100.0	35.0	69.2	98.4	100.0
Miss2	95.8	99.8	100.0	100.0	51.9	96.6	100.0	100.0
0.06								
True	16.2	5.6	6.8	6.0	12.8	5.8	4.2	5.0
Miss1	87.4	99.6	100.0	100.0	39.0	71.6	98.8	100.0
Miss2	96.6	99.8	100.0	100.0	55.5	97.6	100.0	100.0

Note. [a] For the True models, 499 out of 500 data sets have converged results; 497 out of 500 data sets have converged results for the Miss1 and Miss2 models.

Table 4.10.1. CFA model rejection rates of TLI at various cutoff values for binary equal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.90								
True	0.4	0.0	0.0	0.0	0.2	0.0	0.0	0.0
Miss1	75.4	87.4	95.4	98.2	2.2	0.0	0.0	0.0
Miss2	90.6	99.2	100.0	100.0	6.2	0.6	0.0	0.0
0.93								
True	2.8	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Miss1	85.8	96.4	99.8	100.0	7.2	0.6	0.0	0.0
Miss2	95.4	100.0	100.0	100.0	18.4	10.2	3.8	0.8
0.94								
True	5.2	0.0	0.0	0.0	0.8	0.0	0.0	0.0
Miss1	89.4	97.8	100.0	100.0	11.6	2.8	0.0	0.0
Miss2	96.4	100.0	100.0	100.0	27.2	21.4	13.2	9.0
0.95								
True	7.6	0.0	0.0	0.0	2.0	0.0	0.0	0.0
Miss1	92.2	99.0	100.0	100.0	20.8	7.4	2.4	0.2
Miss2	97.2	100.0	100.0	100.0	38.6	36.2	34.8	35.4
0.96								
True	11.8	0.4	0.0	0.0	5.2	0.0	0.0	0.0
Miss1	94.0	100.0	100.0	100.0	30.8	18.6	11.8	6.4
Miss2	98.0	100.0	100.0	100.0	52.8	59.2	67.0	77.2

Table 4.10.2. CFA model rejection rates of TLI at various cutoff values for binary unequal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100 [a]	250	500	1000
0.90								
True	3.8	0.0	0.0	0.0	1.6	0.0	0.0	0.0
Miss1	71.2	83.2	93.4	97.6	6.2	0.0	0.0	0.0
Miss2	89.4	98.2	99.8	100.0	11.3	1.2	0.0	0.0
0.93								
True	10.2	0.4	0.0	0.0	3.8	0.0	0.0	0.0
Miss1	83.0	95.6	99.8	100.0	15.3	2.6	0.2	0.0
Miss2	95.2	99.2	100.0	100.0	23.7	12.8	3.6	1.6
0.94								
True	13.4	0.4	0.0	0.0	7.0	0.0	0.0	0.0
Miss1	86.6	97.8	99.8	100.0	20.5	4.2	0.8	0.0
Miss2	96.2	99.2	100.0	100.0	33.0	21.6	11.2	9.0
0.95								
True	17.6	0.6	0.0	0.0	9.4	0.2	0.0	0.0
Miss1	92.2	98.2	100.0	100.0	27.2	11.6	1.8	0.4
Miss2	97.8	99.4	100.0	100.0	41.0	35.8	35.0	35.0
0.96								
True	22.2	1.4	0.0	0.0	13.4	0.4	0.0	0.0
Miss1	94.8	98.8	100.0	100.0	36.4	19.0	10.6	6.0
Miss2	98.8	99.8	100.0	100.0	51.9	58.4	68.2	75.4

Note. [a] For the True models, 499 out of 500 data sets have converged results; 497 out of 500 data sets have converged results for the Miss1 and Miss2 models.

Table 4.11.1. CFA model rejection rates of CFI at various cutoff values for binary equal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size				Sample Size			
	100	250	500	1000	100	250	500	1000
0.90								
True	0.6	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Miss1	76.0	88.6	96.6	99.2	3.4	0.2	0.0	0.0
Miss2	89.0	98.2	99.8	100.0	10.4	8.0	4.8	2.6
0.93								
True	4.2	0.0	0.0	0.0	0.6	0.0	0.0	0.0
Miss1	86.2	96.4	99.8	100.0	13.2	6.4	2.2	1.6
Miss2	95.2	99.6	100.0	100.0	30.4	35.8	46.2	62.8
0.94								
True	6.4	0.2	0.0	0.0	2.6	0.0	0.0	0.0
Miss1	89.2	98.2	100.0	100.0	21.0	13.0	9.2	6.8
Miss2	95.4	100.0	100.0	100.0	40.6	54.6	71.0	87.6
0.95								
True	10.2	0.2	0.0	0.0	4.6	0.0	0.0	0.0
Miss1	91.8	98.8	100.0	100.0	30.2	25.6	25.6	26.2
Miss2	96.8	100.0	100.0	100.0	51.6	71.6	88.6	97.8
0.96								
True	16.2	1.4	0.0	0.0	8.0	0.2	0.0	0.0
Miss1	94.0	100.0	100.0	100.0	39.6	42.0	46.2	59.8
Miss2	97.6	100.0	100.0	100.0	64.0	87.8	97.8	100.0

Table 4.11.2. CFA model rejection rates of CFI at various cutoff values for binary unequal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100 [a]	250	500	1000
0.90								
True	4.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0
Miss1	72.0	84.0	94.0	98.0	7.8	0.6	0.0	0.0
Miss2	87.8	96.8	100.0	100.0	15.1	5.6	2.2	0.6
0.93								
True	11.8	0.4	0.0	0.0	4.8	0.0	0.0	0.0
Miss1	83.2	95.4	99.8	100.0	19.3	6.8	1.8	0.4
Miss2	94.0	99.2	100.0	100.0	30.2	30.0	30.8	40.6
0.94								
True	15.4	0.4	0.0	0.0	7.6	0.2	0.0	0.0
Miss1	86.6	97.8	99.8	100.0	25.8	12.4	4.6	2.4
Miss2	95.8	99.2	100.0	100.0	37.8	43.4	57.8	72.4
0.95								
True	19.6	1.0	0.0	0.0	11.0	0.4	0.0	0.0
Miss1	92.0	98.2	100.0	100.0	32.4	18.8	16.2	14.0
Miss2	97.0	99.2	100.0	100.0	49.1	62.0	83.4	93.2
0.96								
True	25.6	3.0	0.0	0.0	16.2	1.0	0.0	0.0
Miss1	95.0	98.8	100.0	100.0	42.9	35.2	37.0	42.2
Miss2	98.2	99.8	100.0	100.0	58.8	78.8	94.8	99.6

Note. [a] For the True models, 499 out of 500 data sets have converged results; 497 out of 500 data sets have converged results for the Miss1 and Miss2 models.

Table 4.12.1. CFA model rejection rates of RMSEA at various cutoff values for binary equal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.045								
True	26.6	2.2	0.0	0.0	21.6	1.2	0.0	0.0
Miss1	96.0	100.0	100.0	100.0	68.4	71.2	71.2	87.8
Miss2	98.4	100.0	100.0	100.0	84.8	96.8	99.8	100.0
0.05								
True	20.6	1.2	0.0	0.0	17.0	0.6	0.0	0.0
Miss1	94.2	99.8	100.0	100.0	60.4	55.0	53.2	59.6
Miss2	97.8	100.0	100.0	100.0	81.0	92.0	98.2	100.0
0.055								
True	14.4	0.2	0.0	0.0	12.6	0.0	0.0	0.0
Miss1	93.2	99.2	100.0	100.0	51.4	40.4	32.6	30.6
Miss2	97.2	100.0	100.0	100.0	75.4	85.0	90.4	99.0
0.06								
True	10.6	0.2	0.0	0.0	8.4	0.0	0.0	0.0
Miss1	90.2	98.2	100.0	100.0	42.8	27.8	16.4	9.8
Miss2	96.4	100.0	100.0	100.0	68.2	74.0	77.6	88.6
0.07								
True	3.8	0.0	0.0	0.0	4.6	0.0	0.0	0.0
Miss1	85.8	94.8	98.8	100.0	29.0	7.8	1.2	0.2
Miss2	94.8	99.6	100.0	100.0	51.0	42.4	34.2	32.8
0.08								
True	1.0	0.0	0.0	0.0	1.2	0.0	0.0	0.0
Miss1	80.0	90.0	96.2	98.6	15.6	2.0	0.0	0.0
Miss2	91.2	98.8	99.8	100.0	32.6	14.8	4.6	1.2
0.09								
True	0.2	0.0	0.0	0.0	0.6	0.0	0.0	0.0
Miss1	71.4	81.0	86.2	93.8	6.6	0.2	0.0	0.0
Miss2	86.2	96.2	99.0	100.0	15.8	3.2	0.0	0.0
0.10								
True	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Miss1	62.2	67.4	66.4	74.2	4.0	0.0	0.0	0.0
Miss2	80.4	91.8	96.8	99.0	8.8	0.4	0.0	0.0
0.11								
True	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Miss1	53.4	48.0	40.8	37.6	1.4	0.0	0.0	0.0
Miss2	73.8	82.8	90.0	94.8	4.2	0.2	0.0	0.0

Table 4.12.2. CFA model rejection rates of RMSEA at various cutoff values for binary unequal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100 [a]	250	500	1000
0.045								
True	34.8	2.6	0.0	0.0	31.7	2.6	0.0	0.0
Miss1	96.0	99.2	100.0	100.0	67.2	55.8	56.2	56.2
Miss2	99.4	99.6	100.0	100.0	82.5	90.4	97.4	100.0
0.05								
True	31.8	1.2	0.0	0.0	26.7	1.2	0.0	0.0
Miss1	95.0	98.8	100.0	100.0	61.4	41.0	30.6	25.4
Miss2	98.8	99.6	100.0	100.0	77.5	83.8	92.0	95.6
0.055								
True	26.0	0.4	0.0	0.0	22.4	0.6	0.0	0.0
Miss1	93.4	97.8	99.8	100.0	55.9	30.0	14.6	7.6
Miss2	98.0	99.6	100.0	100.0	72.6	72.8	78.4	82.6
0.06								
True	21.6	0.4	0.0	0.0	18.0	0.2	0.0	0.0
Miss1	91.6	96.0	99.6	100.0	47.5	16.2	5.6	0.4
Miss2	97.6	99.6	100.0	100.0	65.2	56.4	51.8	53.2
0.07								
True	12.2	0.2	0.0	0.0	10.6	0.0	0.0	0.0
Miss1	86.0	90.2	95.8	98.8	34.8	5.6	0.8	0.0
Miss2	96.6	98.8	100.0	100.0	50.5	24.8	9.6	6.0
0.08								
True	6.2	0.0	0.0	0.0	6.4	0.0	0.0	0.0
Miss1	78.6	77.8	85.0	91.8	20.1	1.0	0.0	0.0
Miss2	93.0	95.6	98.8	100.0	33.8	7.4	0.6	0.0
0.09								
True	2.8	0.0	0.0	0.0	3.2	0.0	0.0	0.0
Miss1	70.8	62.2	65.4	66.6	12.1	0.0	0.0	0.0
Miss2	87.0	89.2	95.4	98.2	20.1	1.2	0.0	0.0
0.10								
True	1.4	0.0	0.0	0.0	1.8	0.0	0.0	0.0
Miss1	59.2	41.0	36.4	32.0	5.8	0.0	0.0	0.0
Miss2	80.4	76.8	83.8	89.4	12.3	0.2	0.0	0.0
0.11								
True	0.2	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Miss1	47.8	26.2	14.6	9.2	3.0	0.0	0.0	0.0
Miss2	72.0	60.2	63.0	66.4	6.6	0.0	0.0	0.0

Note. [a] For the True models, 499 out of 500 data sets have converged results; 497 out of 500 data sets have converged results for the Miss1 and Miss2 models.

Table 4.13.1. CFA model rejection rates of SRMR at various cutoff values for binary equal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.045								
True	100.0	100.0	86.4	0.2	100.0	100.0	51.2	0.2
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.05								
True	100.0	100.0	45.4	0.0	100.0	99.4	14.8	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	99.6	96.2
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.055								
True	100.0	99.4	12.4	0.0	100.0	92.0	1.2	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	97.2	77.4
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.06								
True	100.0	94.6	0.8	0.0	100.0	72.4	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	99.6	81.0	40.6
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.8
0.07								
True	100.0	53.4	0.0	0.0	100.0	15.8	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	91.2	30.4	3.2
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	87.6	61.6
0.08								
True	100.0	9.6	0.0	0.0	98.4	1.6	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	99.6	52.4	3.2	0.0
Miss2	100.0	100.0	100.0	100.0	100.0	90.4	41.4	7.0
0.09								
True	99.0	1.0	0.0	0.0	85.6	0.0	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	97.6	17.6	0.0	0.0
Miss2	100.0	100.0	100.0	100.0	99.0	54.8	5.6	0.0
0.10								
True	88.8	0.0	0.0	0.0	58.6	0.0	0.0	0.0
Miss1	100.0	99.8	99.4	99.8	85.8	2.2	0.0	0.0
Miss2	100.0	100.0	100.0	100.0	94.6	18.0	0.0	0.0
0.11								
True	61.2	0.0	0.0	0.0	24.4	0.0	0.0	0.0
Miss1	99.8	97.2	98.4	98.0	62.4	0.6	0.0	0.0
Miss2	100.0	100.0	100.0	100.0	79.4	3.6	0.0	0.0

Table 4.13.2. CFA model rejection rates of SRMR at various cutoff values for binary unequal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100 [a]	250	500	1000
0.045								
True	100.0	100.0	99.4	16.2	100.0	100.0	96.4	2.2
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.05								
True	100.0	100.0	93.8	1.8	100.0	100.0	76.4	0.2
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.4
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.055								
True	100.0	100.0	68.8	0.4	100.0	100.0	41.6	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	99.6	91.8
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.06								
True	100.0	100.0	37.8	0.0	100.0	99.2	16.4	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	97.2	68.2
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.6
0.07								
True	100.0	95.8	3.8	0.0	100.0	78.8	0.4	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	99.8	74.2	12.8
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	97.2	82.8
0.08								
True	100.0	67.2	0.0	0.0	100.0	37.8	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	89.8	22.6	0.0
Miss2	100.0	100.0	100.0	100.0	100.0	99.6	78.2	21.0
0.09								
True	99.8	24.0	0.0	0.0	99.4	8.2	0.0	0.0
Miss1	100.0	100.0	100.0	100.0	100.0	57.8	2.8	0.0
Miss2	100.0	100.0	100.0	100.0	100.0	87.6	27.4	0.8
0.10								
True	98.6	4.2	0.0	0.0	96.0	0.6	0.0	0.0
Miss1	100.0	99.8	99.8	99.8	99.0	26.4	0.2	0.0
Miss2	100.0	100.0	100.0	100.0	99.8	58.8	2.4	0.0
0.11								
True	95.2	0.6	0.0	0.0	80.8	0.0	0.0	0.0
Miss1	100.0	98.8	98.4	98.2	94.6	6.4	0.0	0.0
Miss2	100.0	100.0	100.0	100.0	97.6	25.2	0.2	0.0

Note. [a] For the True models, 499 out of 500 data sets have converged results; 497 out of 500 data sets have converged results for the Miss1 and Miss2 models.

Table 4.14.1. CFA model rejection rates of WRMR at various cutoff values for binary equal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.60								
True	99.8	99.6	100.0	99.2	96.4	93.4	93.6	95.6
Miss1	100.0	100.0	100.0	100.0	99.6	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.70								
True	86.6	82.2	80.4	81.8	60.6	52.8	53.0	54.6
Miss1	100.0	100.0	100.0	100.0	91.6	99.6	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	98.0	100.0	100.0	100.0
0.80								
True	40.2	29.6	23.6	26.2	16.0	8.6	7.2	8.4
Miss1	98.8	100.0	100.0	100.0	61.2	91.4	99.8	100.0
Miss2	99.8	100.0	100.0	100.0	80.0	100.0	100.0	100.0
0.90								
True	8.8	4.0	1.2	2.2	1.6	0.2	0.2	0.8
Miss1	95.2	100.0	100.0	100.0	23.2	62.0	96.6	100.0
Miss2	99.0	100.0	100.0	100.0	46.8	94.4	100.0	100.0
0.95								
True	2.6	1.6	0.4	0.4	0.4	0.0	0.0	0.4
Miss1	90.8	100.0	100.0	100.0	9.0	41.4	89.8	100.0
Miss2	97.8	100.0	100.0	100.0	29.4	87.4	100.0	100.0
1.00								
True	0.6	0.4	0.2	0.0	0.4	0.0	0.0	0.0
Miss1	86.0	100.0	100.0	100.0	4.8	25.0	76.0	100.0
Miss2	96.2	100.0	100.0	100.0	14.2	72.6	100.0	100.0
1.10								
True	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0
Miss1	71.6	98.4	100.0	100.0	1.2	5.6	43.8	99.6
Miss2	91.0	100.0	100.0	100.0	4.2	36.8	97.6	100.0

Table 4.14.2. CFA model rejection rates of WRMR at various cutoff values for binary unequal outcomes

Cutoff Value	Simple Model				Complex Model			
	Sample Size 100	250	500	1000	Sample Size 100 [a]	250	500	1000
0.60								
True	99.8	100.0	99.4	99.8	99.2	98.8	98.2	98.4
Miss1	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.70								
True	92.4	90.8	86.0	87.0	79.4	71.2	66.6	65.6
Miss1	100.0	100.0	100.0	100.0	95.6	99.8	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	98.6	100.0	100.0	100.0
0.80								
True	59.2	43.8	39.0	36.6	33.1	20.4	14.8	12.4
Miss1	99.4	100.0	100.0	100.0	70.8	92.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	86.1	99.8	100.0	100.0
0.90								
True	23.0	6.0	5.4	4.8	11.6	2.6	0.4	0.8
Miss1	96.4	99.8	100.0	100.0	36.4	56.6	95.8	100.0
Miss2	100.0	100.0	100.0	100.0	55.1	92.4	100.0	100.0
0.95								
True	14.4	2.8	0.6	0.8	6.6	1.0	0.0	0.0
Miss1	92.6	99.8	100.0	100.0	23.1	38.4	88.8	99.8
Miss2	99.4	100.0	100.0	100.0	40.6	81.6	99.8	100.0
1.00								
True	7.2	1.0	0.0	0.2	3.4	0.4	0.0	0.0
Miss1	87.8	99.2	100.0	100.0	14.5	22.2	76.4	99.8
Miss2	98.0	99.8	100.0	100.0	24.5	67.4	99.2	100.0
1.10								
True	1.8	0.4	0.0	0.0	1.0	0.0	0.0	0.0
Miss1	73.0	97.8	100.0	100.0	4.0	6.4	34.8	96.2
Miss2	93.4	99.6	100.0	100.0	9.7	30.0	93.4	100.0

Note. [a] For the True models, 499 out of 500 data sets have converged results; 497 out of 500 data sets have converged results for the Miss1 and Miss2 models.

Table 4.15. CFA model rejection rates of Chi-P at various cutoff values in trivial models

Cutoff Value	Simple trivial model (covariances=.05)				Complex trivial model (loadings=.1)			
	Sample Size				Sample Size			
	100	250	500	1000	100	250	500	1000
0.01								
True	4.4	1.8	0.4	2.0	5.2	2.0	1.4	1.6
Miss1	4.6	1.6	0.6	2.6	6.0	2.6	4.8	7.6
Miss2	4.6	2.0	1.0	3.0	7.0	3.8	11.0	21.2
0.03								
True	9.8	5.6	3.0	3.8	10.8	4.6	4.6	2.8
Miss1	10.4	6.4	3.2	4.6	11.8	6.8	12.2	13.8
Miss2	11.0	6.8	4.0	6.4	12.6	10.6	20.4	38.2
0.04								
True	13.6	7.6	3.8	4.4	12.6	6.0	6.2	3.2
Miss1	13.4	8.0	5.0	5.8	13.6	9.4	13.6	17.0
Miss2	13.6	8.4	5.6	7.2	14.8	12.6	23.8	43.0
0.05								
True	15.8	9.4	5.4	4.4	14.8	6.8	8.4	4.2
Miss1	16.4	9.2	6.6	7.2	15.4	11.2	16.6	20.0
Miss2	17.2	10.0	7.2	8.6	17.0	14.6	27.0	48.4
0.06								
True	18.0	11.0	6.8	6.2	17.0	7.8	9.6	5.4
Miss1	18.8	11.4	8.0	8.2	17.2	12.4	18.0	22.0
Miss2	19.4	11.0	8.6	11.2	19.6	15.6	30.4	51.2

Table 4.16. CFA model rejection rates of TLI at various cutoff values in trivial models

Cutoff Value	Simple trivial model (covariances=.05)				Complex trivial model (loadings=.1)			
	Sample Size				Sample Size			
	100	250	500	1000	100	250	500	1000
0.90								
True	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss2	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0
0.93								
True	2.4	0.0	0.0	0.0	1.4	0.0	0.0	0.0
Miss1	1.8	0.0	0.0	0.0	2.0	0.0	0.0	0.0
Miss2	2.0	0.0	0.0	0.0	2.8	0.0	0.0	0.0
0.94								
True	4.6	0.0	0.0	0.0	4.2	0.0	0.0	0.0
Miss1	4.8	0.0	0.0	0.0	4.2	0.0	0.0	0.0
Miss2	4.4	0.0	0.0	0.0	4.4	0.0	0.0	0.0
0.95								
True	9.4	0.0	0.0	0.0	7.8	0.0	0.0	0.0
Miss1	9.8	0.0	0.0	0.0	8.8	0.0	0.0	0.0
Miss2	9.8	0.0	0.0	0.0	9.0	0.0	0.0	0.0
0.96								
True	17.4	0.0	0.0	0.0	12.8	0.0	0.0	0.0
Miss1	16.8	0.2	0.0	0.0	13.8	0.0	0.0	0.0
Miss2	17.0	0.2	0.0	0.0	15.0	0.0	0.0	0.0

Table 4.17. CFA model rejection rates of CFI at various cutoff values in trivial models

Cutoff Value	Simple trivial model (covariances=.05)				Complex trivial model (loadings=.1)			
	Sample Size				Sample Size			
	100	250	500	1000	100	250	500	1000
0.90								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.93								
True	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	1.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Miss2	1.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0
0.94								
True	1.8	0.0	0.0	0.0	1.2	0.0	0.0	0.0
Miss1	1.4	0.0	0.0	0.0	1.8	0.0	0.0	0.0
Miss2	1.4	0.0	0.0	0.0	2.6	0.0	0.0	0.0
0.95								
True	4.6	0.0	0.0	0.0	3.8	0.0	0.0	0.0
Miss1	4.8	0.0	0.0	0.0	4.0	0.0	0.0	0.0
Miss2	5.6	0.0	0.0	0.0	4.2	0.0	0.0	0.0
0.96								
True	11.0	0.0	0.0	0.0	7.8	0.0	0.0	0.0
Miss1	11.6	0.0	0.0	0.0	9.4	0.0	0.0	0.0
Miss2	11.8	0.0	0.0	0.0	9.6	0.0	0.0	0.0

Table 4.18. CFA model rejection rates of RMSEA at various cutoff values in trivial models

Cutoff Value	Simple trivial model (covariances=.05)				Complex trivial model (loadings=.1)			
	Sample Size				Sample Size			
	100	250	500	1000	100	250	500	1000
0.045								
True	24.0	0.2	0.0	0.0	23.8	0.4	0.0	0.0
Miss1	24.2	0.2	0.0	0.0	26.2	0.6	0.0	0.0
Miss2	24.8	0.2	0.0	0.0	27.2	1.2	0.0	0.0
0.05								
True	18.0	0.0	0.0	0.0	17.0	0.0	0.0	0.0
Miss1	17.8	0.0	0.0	0.0	17.2	0.0	0.0	0.0
Miss2	18.4	0.0	0.0	0.0	19.2	0.2	0.0	0.0
0.055								
True	10.0	0.0	0.0	0.0	11.2	0.0	0.0	0.0
Miss1	10.4	0.0	0.0	0.0	12.2	0.0	0.0	0.0
Miss2	10.8	0.0	0.0	0.0	12.6	0.0	0.0	0.0
0.06								
True	5.2	0.0	0.0	0.0	7.4	0.0	0.0	0.0
Miss1	5.2	0.0	0.0	0.0	8.0	0.0	0.0	0.0
Miss2	5.8	0.0	0.0	0.0	8.4	0.0	0.0	0.0
0.07								
True	1.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0
Miss1	1.0	0.0	0.0	0.0	1.8	0.0	0.0	0.0
Miss2	1.0	0.0	0.0	0.0	2.6	0.0	0.0	0.0
0.08								
True	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss2	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0
0.09								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.10								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.11								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 4.19. CFA model rejection rates of SRMR at various cutoff values in trivial models

Cutoff Value	Simple trivial model (covariances=.05)				Complex trivial model (loadings=.1)			
	Sample Size				Sample Size			
	100	250	500	1000	100	250	500	1000
0.045								
True	100.0	10.8	0.0	0.0	94.4	0.8	0.0	0.0
Miss1	100.0	43.2	5.4	0.0	97.6	3.0	0.0	0.0
Miss2	100.0	67.2	18.6	1.8	98.6	6.2	0.0	0.0
0.05								
True	96.6	2.2	0.0	0.0	76.4	0.0	0.0	0.0
Miss1	99.4	24.0	2.8	0.0	84.0	0.2	0.0	0.0
Miss2	99.6	42.6	6.8	0.4	89.4	0.8	0.0	0.0
0.055								
True	84.2	0.2	0.0	0.0	49.0	0.0	0.0	0.0
Miss1	95.2	11.6	1.2	0.0	61.2	0.0	0.0	0.0
Miss2	98.8	25.4	3.4	0.0	69.6	0.0	0.0	0.0
0.06								
True	61.0	0.0	0.0	0.0	22.8	0.0	0.0	0.0
Miss1	82.6	4.2	0.4	0.0	33.2	0.0	0.0	0.0
Miss2	92.2	14.2	1.4	0.0	41.2	0.0	0.0	0.0
0.07								
True	11.0	0.0	0.0	0.0	3.2	0.0	0.0	0.0
Miss1	41.0	0.2	0.0	0.0	5.0	0.0	0.0	0.0
Miss2	64.0	3.4	0.0	0.0	6.4	0.0	0.0	0.0
0.08								
True	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	15.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0
Miss2	31.4	1.4	0.0	0.0	0.6	0.0	0.0	0.0
0.09								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss2	14.4	0.2	0.0	0.0	0.2	0.0	0.0	0.0
0.10								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	1.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss2	4.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.11								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss2	2.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 4.20. CFA model rejection rates of WRMR at various cutoff values in trivial models

Cutoff Value	Simple trivial model (covariances=.05)				Complex trivial model (loadings=.1)			
	Sample Size				Sample Size			
	100	250	500	1000	100	250	500	1000
0.60								
True	99.6	99.6	99.4	99.2	91.2	89.6	90.0	92.2
Miss1	100.0	99.8	99.8	99.8	96.0	95.8	98.8	99.6
Miss2	100.0	100.0	100.0	100.0	97.6	98.8	99.8	100.0
0.70								
True	93.6	91.4	90.6	89.8	58.0	54.0	48.8	49.8
Miss1	98.4	96.8	97.4	98.0	67.4	68.8	81.6	92.0
Miss2	99.2	98.4	99.8	99.6	73.4	78.0	93.0	99.0
0.80								
True	63.2	60.8	57.8	59.0	20.2	14.2	11.2	12.4
Miss1	83.0	84.4	85.8	89.8	29.4	28.0	40.4	66.0
Miss2	92.8	92.4	95.4	97.6	36.8	39.4	57.4	87.4
0.90								
True	22.2	23.4	17.4	19.0	5.4	2.2	1.0	1.8
Miss1	53.6	55.4	64.0	73.0	8.4	6.0	11.4	25.8
Miss2	74.4	77.0	84.6	92.6	10.4	9.6	23.0	54.0
0.95								
True	9.6	10.0	7.6	8.4	1.2	0.6	0.6	0.4
Miss1	38.4	43.6	51.6	65.8	3.0	2.8	4.6	13.0
Miss2	61.0	66.8	76.6	86.8	4.0	4.6	11.2	34.8
1.00								
True	3.0	5.6	3.2	3.8	0.4	0.0	0.4	0.0
Miss1	25.8	32.2	40.4	55.8	1.0	0.6	1.6	5.2
Miss2	48.4	54.4	65.8	82.8	1.2	1.2	4.8	18.0
1.10								
True	0.2	1.0	0.0	0.2	0.0	0.0	0.0	0.0
Miss1	11.6	18.6	23.0	39.8	0.0	0.0	0.4	0.4
Miss2	27.2	34.6	48.2	71.0	0.2	0.0	0.4	2.0

CHAPTER 5

MONTE CARLO STUDY 2: MIMIC

MIMIC models were introduced by Jöreskog & Goldberger (1975). Compared to CFA models, a MIMIC model has background variables, and it allows the regression of latent variables on the background variables. The information gained from adding a set of relevant background variables allows us to validate constructs, study the relative importance of the predictors, detect population heterogeneity and detect measurement noninvariance (Muthén et al., 1993). A typical MIMIC model has both measurement and structural components. The measurement component refers to the hypothesized relationship between a latent variable and its indicators. The structural component consists of the coefficients of regressing latent variables on the background variables, and it reflects the relative importance of the background variables on the latent variables.

5.1 Design of Simulation

The simulation design was similar to the first study except for the addition of a covariate. A covariate (referred to as x), which was normally distributed with a variance of one and a mean of zero, was added to the previous CFA model to predict all the three factors. The model is expressed as

$$y = \Lambda \xi + \varepsilon, \quad (16)$$

$$\xi = \Gamma x + \zeta. \quad (17)$$

The first equation is a regular CFA model. Γ is a matrix that captures the regression coefficients of factors on covariates, and it reflects the relationship between the covariate(s) and factors. ζ is a vector that captures the residual variances of factors after taking into account covariates.

The regression coefficients of the three factors on x were set to be 0.7, 0.5 and 0.4, respectively. In doing so, the R^2 values for the three factors accounted by x ranged from 0.16 to 0.5, which are often seen in real data. Models with misspecified factor loadings (the complex model) were further investigated in this study, and the parameter values for the measurement component of the MIMIC model were the same as the complex model in Table 4.2. Three specifications of the complex model were investigated. The first model was properly specified, and the other two models excluded factor loadings from the sample that existed in the population. The number of free parameters were thirty-nine in the properly specified model. The setup of the misspecified models stayed the same in that one cross loading of 0.7 (path d in Figure 4.1) was misspecified in the Miss1 model and two loadings of 0.7 (path d and e in Figure 4.1) were misspecified in the Miss2 model.

For continuous outcomes, Chi-P, TLI, CFI, RMSEA, SRMR and WRMR were evaluated under three distributional levels (normal, moderately and severely non-normal), three levels of model misspecification (one properly specified and two under-parameterized misspecified models) and four sample sizes (100, 250, 500, 1000). For binary outcomes, Chi-P, TLI, CFI, RMSEA and WRMR were evaluated under two levels of distributional specification (equal and unequal cases), three levels of model misspecification and four sample sizes. The simulation of non-normal data for continuous outcomes and choices of cut points for binary outcomes were the same as the previous CFA study. Note that SRMR was not included because it has not yet been defined for categorical outcomes with threshold structure or covariates.

Data was generated in SAS and analyzed by Mplus 2.1. The convergence rates ranged from 99% to 100%, and runs that failed to converge within 1000 iterations were discarded.

5.2 Results

It was found in the previous CFA study that a few cutoff values for the same fit index may be acceptable, such as cutoff values close to 0.95 or 0.96 for CFI and TLI, close to 0.07 or 0.08 for SRMR, close to 0.05 or 0.06 for RMSEA, and close to 0.95 or 1.0 for WRMR. In this section these cutoff values suggested in the previous study are evaluated first to see whether they can be applicable also to the MIMIC models. In addition to the investigation of cutoff criteria, the performance of fit indices under different sample sizes and model misspecifications is also discussed.

5.2.1 Continuous Outcomes

Tables 5.1.1 to 5.6.2 present the rejection rates of Chi-P, TLI, CFI, RMSEA, SRMR and WRMR under various cutoff values with continuous outcomes. To generally evaluate the performance of fit indices under different sample sizes and model misspecifications for the CFA and MIMIC models, the values of means and standard deviations (SDs) for the six fit measures under normality are provided in Table 5.12.

Chi-P. Table 5.12 shows that for both CFA and MIMIC True models with multivariate normal data, the ML-based Chi-P values on average tended to increase with increasing sample sizes. In both CFA and MIMIC Miss1 models, Chi-P tended to have larger average values (0.023) and spread (SDs ranged from 0.06 to 0.07) at N = 100, whereas at N ≥ 250 its values were consistently close to 0. Inspection of the rejection rates in Table 4.3.1 and Table 5.1.1 shows that similar to the CFA models, rejection rates of the SB χ^2 were almost always higher than those of the ML χ^2 at N = 100 in the MIMIC model. The discrepancy of rejection rates between the ML-based and SB-based Chi-P in the MIMIC model appeared to be larger than the CFA model. Moreover, similar to the CFA models, a probability level of 0.05 for Chi-P overrejected true models at N = 100 in the MIMIC models under normality (rejection rates were 16.2% for the ML-based Chi-P and 23.0% for the SB-based Chi-P).

The ML χ^2 and SB χ^2 showed similar patterns and rejection rates between the MIMIC models (Table 5.1.2) and the CFA models (Tables 4.3.2 – 4.3.3) with non-normal data. The SB χ^2 had much lower type I error rates than the ML χ^2 under non-normality, but it still tended to overreject the properly specified models at smaller sample sizes. A cutoff value of 0.05 for the SB Chi-P overrejected true models at N ≤ 500 (rejection rates ranged from 14.2% to 52.6%) under moderate non-normality and across all four sample sizes under severely non-normality (rejection rates ranged from 18.4% to 81.6%).

TLI. It was indicated in Table 5.2.1 and Table 5.2.2 that a cutoff value of 0.95 for the ML-based TLI was still applicable across all four sample sizes under normality, at N ≥ 250 under moderate non-normality, and at N ≥ 500 under severe non-normality in the MIMIC models. The rejection rates of the ML- and SB-based TLI in the MIMIC misspecified models were lower than those in the CFA misspecified models. Table 5.12 gives the means and standard deviations for the ML-based TLI under normality, and it shows that the TLI values on average were very similar for both the CFA and MIMIC True models. In addition, the TLI means in the MIMIC Miss1 model were slightly higher than the CFA Miss1 model. Similar patterns occurred to CFI. The CFI means tended to be larger and thus result in lower rejection rates in MIMIC misspecified models under certain cutoff values.

CFI. Similar to CFA models, a cutoff value of 0.96 for the ML-based CFI exhibited satisfactory power (ranged from 0.7 to 0.89) with acceptable type I error rates (ranged from 0% to 4.2%) under normality (Table 5.3.1). A cutoff value of 0.95 appeared to lack power in the Miss1 models at larger sample sizes. Under non-normality (Table 5.3.2), a cutoff value of 0.95 or 0.96 for the ML-based CFI had acceptable type I and type II error rates at $N \geq 250$ under moderate non-normality and at $N \geq 500$ under severe non-normality. The cutoff value of 0.96 for TLI exhibited strong power but also had higher type I error rates at smaller sample size. Under severe non-normality with $N = 250$, the rejection rates in the True models was 32.8% for the ML-based CFI comparing to 7.6% for the SB-based CFI at a cutoff of 0.95, thus the use of the SB-based CFI was more adequate.

RMSEA. For both CFA and MIMIC complex models, a cutoff value of 0.06 for the ML-based RMSEA was suitable across all four sample sizes with normal data (Table 5.4.1 and Table 4.6.1). With this cutoff value, however, the type II error rates were too high at large sample size under non-normality (rejected 2.4% of the Miss1 models under moderate non-normality and 0% of the Miss1 models under severe non-normality at $N = 1000$). The cutoff value close to 0.05 was still suitable for the ML-based RMSEA at $N \geq 250$ under normality and at N ≥ 500 under non-normality. At N = 250, the SB-based RMSEA at a cutoff of 0.05 had acceptable type I and II error rates under non-normality (rejection rates ranged from 77.8% to 100% for misspecified models and 0% to 4% for true models in Table 5.4.2). Finally, Table 5.12 shows that the average RMSEA values tend to be higher in the CFA models and, similar to CFI and TLI, the rejection rates of RMSEA in the MIMIC misspecified models were lower than those in the CFA misspecified models.

SRMR. Table 5.5.1 shows that SRMR does not have a suitable cutoff value with acceptable type I error and type II error rates across *all* four samples under normality. At a cutoff value of 0.07, SRMR only rejected 21% of the Miss2 models and 0% of the Miss1 models at N = 1000. One might want to decrease the cutoff value to 0.055 or 0.05, but the type I errors inflated tremendously at N = 100 under these smaller cutoff values. A cutoff value close to 0.07 under normality and close to 0.08 under non-normality for SRMR had suitable type I and type II error rates at N = 100. However a cutoff value close to 0.06 or 0.055 was necessary to have reasonable power when sample size was larger (N ≥ 250).

WRMR. Generally speaking, WRMR at a cutoff value of 1.0 had acceptable type I and type II error rates with normal and non-normal continuous outcomes (Table 5.6.1 and Table 5.6.2). A cutoff of 0.95 for WRMR provided better rejection rates in the Miss1 and Miss2 models especially when N ≤ 250, but it also resulted in an inflated type I error rate (12.8%) at N = 100 under severe non-normality. Similar to TLI, CFI, RMSEA and SRMR, The rejection rates of WRMR in the MIMIC misspecified models were smaller than those in the CFA misspecified models. Table 5.12 shows that the RMSEA, SRMR and WRMR values on average were lower in the MIMIC Miss1 than in the CFA Miss1 models, whereas the

TLI and CFI mean values in general were higher in the MIMIC Miss1 model. Except for Chi-P, all other fit indices tended to indicate better fit for the MIMIC misspecified models.

5.2.2 Dichotomous Outcomes

Table 5.13 provides the means and SDs for Chi-P, TLI, CFI, RMSEA and WRMR with binary unequal outcomes, and Tables 5.7 to 5.11 present the rejection rates of these fit measures under various cutoff values with binary outcomes. A conventional cutoff value of 0.05 for Chi-P yielded reasonable type I error rates (rejection rates ranged from 4.2% to 7.2%). Its rejection rates in misspecified models ranged from 78.2 % to 100% at $N \geq 250$, and ranged from 33.9% to 67.1% at $N = 100$.

When $N \geq 250$, using a cutoff value of 0.95 for TLI rejected 0 to 0.2% of the true models, 49% to 52.8% of the Miss 1 models, and 88.6% to 99.6% of the Miss2 models. To increase power for Miss1 models, one might want to increase the cutoff value of TLI to 0.96 (rejected 65.6% to 85.2% of the Miss1 models). At the same cutoff value of 0.95 $N \geq 250$, CFI exhibited similar patterns as TLI. It rejected 0% to 0.4% of the true models, 53.2% to 72.4% of the Miss1 models, and 91% to 100% of the Miss2 models at $N \geq 250$. To have higher power for the Miss1 models, one might also want to increase the cutoff of CFI to 0.96. A cutoff value of 0.95 for TLI and CFI exerted higher power across samples but also inflated type I errors at $N = 100$ in MIMIC models. Generally speaking, a cutoff value of 0.95 was still suitable for TLI and CFI with binary outcomes at $N \geq 250$, and, with the same cutoff value, CFI exhibited higher power than TLI.

In comparison with CFA models, it was indicated in Table 5.13 that the average CFI and TLI values tended to be lower and the variability tended to be larger in the MIMIC true and misspecified models.

RMSEA with a cutoff value close to 0.05 or 0.06 was not preferable at N = 100 because it overrejected true models (rejection rates ranged from 11.8% to 27.4%). At N ≥ 250, the cutoff value of 0.06 for RMSEA rejected only 0.8% to 30.4% of the Miss1 models and 49.8% to 86.8% of the Miss2 models. Thus, a cutoff value of close to 0.05 or 0.045 might be more desirable in order to have reasonable power at larger sample sizes. The performance of RMSEA was similar for both CFA and MIMIC complex models with binary outcomes. According to Table 5.13, the means and SDs of RMSEA were very similar for both the CFA and MIMIC models. In addition, comparing Table 5.10 with Table 4.12.1 and Table4.12.2, RMSEA seemed to have similar rejection rate summaries.

WRMR at cutoff values close to 0.95 or 1.0 had acceptable type I error rates (ranged from 0.8% to 7.6%) across all four sample sizes. However, WRMR rejected only around 20% of the Miss1 models and rejected 43.6% to 45.1% of the Miss2 models with the cutoff value of 1.0 at N = 100. With a cutoff close to 0.95, WRMR rejected 34.5% to 37.1% of the Miss1 models and 62% to 67% of the Miss2 models at N = 100. In comparison with the cutoff of 1.0, WRMR at the cutoff of 0.95 also had much higher power (rejected 70% to 74% of the Miss1 models and 95% to 98% of the Miss2 models) at N = 250.

Table 5.13 shows that the average WRMR values in the MIMIC true and misspecified models were higher across samples and were more homogenous (especially at smaller sample sizes) than those in the CFA models. Table 5.11 presents the rejection rates of WRMR under various cutoff values with binary outcomes. It was found that the rejection rates of WRMR in MIMIC models (Table 5.11) were generally larger than those in CFA models (Table 4.14.1 and Table 4.14.2) thus resulted in better power with WRMR in MIMIC models.

5.3 Summary and Conclusion

This chapter evaluated the adequacy of cutoff values for Chi-P, TLI, CFI, RMSEA, SRMR and WRMR in MIMIC models under data and model conditions such as different sample sizes, various types of model misspecification and outcome variables. In comparison with CFA models under normality, the average RMSEA, SRMR and WRMR values were lower and the average TLI and CFI values were higher in the MIMIC misspecified models. TLI, CFI, RMSEA, SRMR and WRMR on average tended to reject fewer MIMIC misspecified models.

Continuous Outcomes. With normal data, the results indicated that a cutoff value close to 0.01 for Chi-P, 0.95 for TLI, 0.96 for CFI, and 0.06 for RMSEA had power around or above 0.7 with type I error rates lower than 5% across all four sample sizes. A cutoff value of 0.95 (or 1.0) for WRMR was roughly applicable to all samples, but note that it only rejected about 45% (34.4%) of the Miss1 models at N = 100. Similar to the previous CFA study, a cutoff value of 0.05 for the ML or SB-based Chi-P overrejected the MIMIC True models at small sample sizes under various data distributions. Moreover, the SB χ^2 with the same cutoff value exerted much lower type I errors than the ML χ^2 under non-normality, but it still overrejected true models at smaller sample sizes.

Under moderate non-normality, cutoff values of 0.95 for the ML-based TLI and 0.96

for the ML-based CFI were still applicable at N ≥ 250. A cutoff value of 0.05 for the ML-based RMSEA can be applicable at N ≥ 500, and the use of the SB-based RMSEA was preferable at N = 250. The suitable cutoff value for SRMR varied with sample size. The commonly used cutoff value of close to 0.07 was only suitable at small sample sizes. A cutoff value close to 0.95 for WRMR exhibited acceptable type I and type II error rates at N ≥ 250.

To reduce type II errors a cutoff value of 0.95 for WRMR might be preferable, but with severely non-normal data, a cutoff value of 1.0 might be preferred to maintain type I error control at N = 100. At N ≥ 250, WRMR at a cutoff close to 0.95 performed well. The ML-based CFI at a cutoff value of 0.95 was applicable at N ≥ 500, and at N = 250 the use of the SB-based CFI was preferable. With moderate non-normality, a cutoff value of 0.05 for the ML-based RMSEA was applicable at N ≥ 500, and the use of the SB-based RMSEA was preferable at N = 250.

Binary Outcomes. In comparison with CFA models, the average Chi-P, TLI and CFI values were lower whereas the average WRMR values were higher in MIMIC models with binary outcomes. The rejection rates of these fit indices were higher in the MIMIC models. RMSEA on average had similar values in the CFA and MIMIC models.

With binary outcomes, none of the fit measures under previously investigated cutoff values had power larger than 0.7 with type I errors lower than 5% at N = 100. Relatively speaking, at small sample sizes Chi-P at a probability level of 0.05 and WRMR at a cutoff value of 0.95 had better type I and type II error rates than the other fit indices. Chi-P at a cutoff of 0.05 rejected 34% to 67% of the misspecified models, whereas WRMR at a cutoff of 0.95 rejected 35% to 67% of the misspecified models at N = 100. TLI at a cutoff close to 0.95, CFI at close to 0.95 (0.96) and RMSEA at close to 0.05 (0.06) all tended to overreject true models at N = 100.

Generally speaking, when only models with misspecified factor loadings (complex models) were taken into account, the performance and rejection rates of these fit measures were similar for both the CFA and MIMIC models. The suitable cutoff criteria discussed in the previous CFA models also seemed to be applicable to the MIMIC models. Table 5.14 summarizes the suitable cutoff criteria for the CFA and MIMIC complex models. It is shown in Table 5.14 that Chi-P ≥ .01, TLI ≥ .95, CFI ≥ .96, RMSEA ≤ .06 can be indications of good models across samples under normality with type I and type II error rates lower than 5% and 30%. A cutoff value of 0.9 for WRMR also had type I and type II error rates lower than 5% and 30% at N ≥ 250. Note that with different acceptable levels of type I and type II errors, different cutoff values may be chosen as a result.

Under non-normality there were few suitable cutoff criteria that can maintain type I error rates lower than 5% and type II error rates lower than 30% across sample sizes. The ML-based Chi-P at a cutoff of 0.01, CFI at 0.95 and RMSEA at 0.05 appeared to be robust to non-normality at N ≥ 500, and their SB counterparts were preferred at N = 250 under non-normality. For SRMR and RMSEA, they tended to lack power at larger sample sizes under some commonly used cutoff values. Because only complex models were investigated, relatively SRMR could not perform well comparing to other fit indices. Most fit indices performed well with larger sample sizes. However, with small sample sizes, these fit indices at commonly suggested cutoff criteria tended to have inflated type I error rates and/or low power.

Table 5.1.1. MIMIC model rejection rates of Chi-P at various cutoff values under normality

Cutoff Value	Complex Model				
	Sample Size				
	100		250	500	1000
	ML	SB			
0.01					
True	**5.0**	**10.2**	**2.6**	**1.6**	**2.0**
Miss1	**72.8**	**80.2**	**99.8**	**100.0**	**100.0**
Miss2	**97.4**	**98.8**	**100.0**	**100.0**	**100.0**
0.03					
True	10.8	17.8	5.6	5.0	5.2
Miss1	81.8	88.0	100.0	100.0	100.0
Miss2	99.2	99.6	100.0	100.0	100.0
0.04					
True	14.2	20.4	7.2	6.0	6.6
Miss1	84.6	90.6	100.0	100.0	100.0
Miss2	99.6	99.8	100.0	100.0	100.0
0.05					
True	**16.2**	**23.0**	**8.6**	**7.2**	**7.6**
Miss1	**86.2**	**91.8**	**100.0**	**100.0**	**100.0**
Miss2	**99.6**	**100.0**	**100.0**	**100.0**	**100.0**
0.06					
True	18.4	26.0	11.2	7.6	8.6
Miss1	87.8	93.2	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 5.1.2. MIMIC model rejection rates of Chi-P at various cutoff values under non-normality

Cutoff Value	Moderate Non-normality								Severe Non-normality							
	Sample Size 100		250		500		1000		Sample Size 100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.01																
True	**68.0**	**32.0**	**70.0**	**9.2**	**70.6**	**4.4**	**75.4**	**1.6**	**93.4**	**66.6**	**98.2**	**31.6**	**99.4**	**13.0**	**99.8**	**6.4**
Miss1	**96.8**	**85.8**	**100.0**	**99.0**	**100.0**	**100.0**	**100.0**	**100.0**	**99.8**	**95.2**	**100.0**	**99.0**	**100.0**	**100.0**	**100.0**	**100.0**
Miss2	**100.0**	**98.8**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**99.8**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
0.03																
True	78.8	45.6	79.0	18.8	83.2	8.0	85.6	4.8	97.4	78.2	99.0	44.2	99.8	23.2	99.8	12.8
Miss1	99.2	93.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	98.6	100.0	99.8	100.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.04																
True	81.6	50.2	82.0	21.8	84.6	11.2	88.6	7.4	97.4	80.2	99.0	48.4	99.8	28.2	99.8	17.2
Miss1	99.2	94.2	100.0	99.8	100.0	100.0	100.0	100.0	100.0	98.8	100.0	99.8	100.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.05																
True	**83.6**	**52.6**	**84.8**	**24.6**	**86.4**	**14.2**	**90.8**	**8.6**	**97.4**	**81.6**	**99.4**	**51.8**	**99.8**	**31.0**	**99.8**	**18.4**
Miss1	**99.4**	**94.6**	**100.0**	**99.8**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**99.2**	**100.0**	**99.8**	**100.0**	**100.0**	**100.0**	**100.0**
Miss2	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
0.06																
True	86.0	55.2	86.4	27.4	88.0	15.8	92.2	10.2	97.8	83.4	99.4	54.2	99.8	34.6	100.0	19.4
Miss1	99.8	95.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	99.4	100.0	99.8	100.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 5.2.1. MIMIC model rejection rates of TLI at various cutoff values under normality

Cutoff Value	Complex Model				
	Sample Size				
	100		250	500	1000
	ML	SB			
0.90					
True	0.0	0.0	0.0	0.0	0.0
Miss1	13.6	12.2	0.2	0.0	0.0
Miss2	62.6	67.0	55.0	58.6	62.8
0.93					
True	0.4	1.2	0.0	0.0	0.0
Miss1	39.0	47.4	28.4	14.2	6.0
Miss2	90.6	92.8	98.4	100.0	100.0
0.94					
True	1.4	2.4	0.0	0.0	0.0
Miss1	55.8	61.8	41.2	35.0	23.6
Miss2	94.6	96.2	99.8	100.0	100.0
0.95					
True	**3.6**	**7.2**	**0.0**	**0.0**	**0.0**
Miss1	**68.8**	**75.8**	**71.8**	**78.4**	**87.0**
Miss2	**97.0**	**98.2**	**100.0**	**100.0**	**100.0**
0.96					
True	9.6	13.8	0.0	0.0	0.0
Miss1	80.6	85.6	90.4	97.4	100.0
Miss2	99.2	99.6	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 5.2.2. MIMIC model rejection rates of TLI at various cutoff values under non-normality

Cutoff Value	Moderate Non-normality								Severe Non-normality							
	Sample Size 100		250		500		1000		Sample Size 100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.90																
True	10.6	3.8	0.0	0.0	0.0	0.0	0.0	0.0	52.6	31.2	3.0	1.0	0.0	0.0	0.0	0.0
Miss1	65.6	46.0	14.4	4.8	0.2	0.2	0.0	0.0	90.4	79.2	61.4	30.8	16.2	4.8	1.4	0.6
Miss2	95.2	88.2	90.8	77.8	88.0	80.0	80.6	82.0	99.4	96.4	98.4	89.2	96.2	86.6	93.2	89.2
0.93																
True	36.2	16.8	0.0	0.0	0.0	0.0	0.0	0.0	81.0	61.8	20.2	5.0	0.2	0.0	0.0	0.0
Miss1	90.0	77.8	72.6	47.2	43.4	27.2	17.0	15.2	98.6	93.0	95.6	73.8	84.4	54.2	56.8	37.6
Miss2	99.4	98.2	100.0	99.4	100.0	99.8	100.0	100.0	100.0	100.0	100.0	99.6	100.0	100.0	100.0	100.0
0.94																
True	48.8	26.8	1.6	0.0	0.0	0.0	0.0	0.0	86.2	71.2	38.2	8.4	1.4	0.0	0.0	0.0
Miss1	94.6	85.4	89.4	68.2	80.4	60.0	61.4	52.6	99.6	96.6	98.4	85.8	95.8	77.0	88.6	70.4
Miss2	100.0	99.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.95																
True	**61.6**	**39.0**	**3.6**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**91.6**	**79.2**	**57.8**	**14.6**	**7.2**	**0.4**	**0.0**	**0.0**
Miss1	**96.8**	**91.4**	**97.2**	**85.6**	**97.2**	**88.4**	**95.6**	**93.0**	**99.8**	**98.8**	**99.0**	**93.2**	**99.6**	**91.4**	**98.6**	**93.8**
Miss2	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
0.96																
True	74.4	52.0	13.6	2.2	0.0	0.0	0.0	0.0	96.2	85.4	78.8	25.2	22.8	1.2	0.2	0.0
Miss1	99.0	94.8	99.2	95.6	99.4	98.4	99.8	99.6	100.0	99.6	100.0	97.6	100.0	98.0	100.0	99.4
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 5.3.1. MIMIC model rejection rates of CFI at various cutoff values under normality

Cutoff Value	Complex Model				
	Sample Size				
	100		250	500	1000
	ML	SB			
0.90					
True	0.0	0.0	0.0	0.0	0.0
Miss1	2.4	2.6	0.0	0.0	0.0
Miss2	33.4	39.0	11.4	4.4	0.6
0.93					
True	0.0	0.0	0.0	0.0	0.0
Miss1	19.4	25.6	1.6	0.0	0.0
Miss2	77.2	82.8	84.8	94.0	98.8
0.94					
True	0.2	0.6	0.0	0.0	0.0
Miss1	34.6	40.4	11.8	1.0	0.4
Miss2	88.6	91.0	97.8	100.0	100.0
0.95					
True	1.4	1.4	0.0	0.0	0.0
Miss1	53.4	59.4	36.4	27.8	15.6
Miss2	94.6	95.8	99.8	100.0	100.0
0.96					
True	**3.6**	**7.2**	**0.0**	**0.0**	**0.0**
Miss1	**70.0**	**76.2**	**73.4**	**79.6**	**89.4**
Miss2	**97.4**	**98.4**	**100.0**	**100.0**	**100.0**

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 5.3.2. MIMIC model rejection rates of CFI at various cutoff values under non-normality

Cutoff Value	Moderate Non-normality								Severe Non-normality							
	Sample Size								Sample Size							
	100		250		500		1000		100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.90																
True	2.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	30.6	16.0	0.6	0.2	0.0	0.0	0.0	0.0
Miss1	36.4	20.2	1.4	0.4	0.0	0.0	0.0	0.0	77.6	59.0	26.4	9.6	2.0	0.8	0.0	0.0
Miss2	82.0	68.8	56.4	39.0	28.6	21.8	9.4	10.6	97.0	88.8	87.8	65.2	67.0	44.8	35.2	35.0
0.93																
True	17.6	7.6	0.0	0.0	0.0	0.0	0.0	0.0	66.8	43.8	8.4	1.8	0.0	0.0	0.0	0.0
Miss1	79.2	61.8	35.8	17.0	6.4	2.2	0.2	0.2	94.4	87.8	80.6	48.2	48.4	18.6	8.8	4.6
Miss2	98.6	93.8	99.0	93.4	98.2	96.8	99.8	99.6	100.0	98.6	99.8	98.0	99.8	97.4	100.0	99.4
0.94																
True	28.8	13.6	0.0	0.0	0.0	0.0	0.0	0.0	78.2	57.6	14.6	4.2	0.0	0.0	0.0	0.0
Miss1	87.4	74.4	63.6	39.2	32.2	17.2	6.4	6.2	97.8	91.6	93.2	67.0	77.4	44.0	40.8	25.2
Miss2	99.2	96.8	100.0	99.4	99.8	99.6	100.0	100.0	100.0	99.8	100.0	99.4	100.0	99.4	100.0	100.0
0.95																
True	**45.8**	**24.6**	**1.2**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**84.8**	**67.8**	**32.8**	**7.6**	**0.4**	**0.0**	**0.0**	**0.0**
Miss1	**93.2**	**84.4**	**87.2**	**62.6**	**73.2**	**54.2**	**51.8**	**45.2**	**99.6**	**95.8**	**98.2**	**83.8**	**94.8**	**73.2**	**83.4**	**64.2**
Miss2	**100.0**	**99.2**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
0.96																
True	**61.6**	**39.0**	**3.6**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**91.6**	**79.2**	**57.8**	**14.6**	**7.2**	**0.4**	**0.0**	**0.0**
Miss1	**97.2**	**91.4**	**97.4**	**86.4**	**97.4**	**88.8**	**95.8**	**94.0**	**99.8**	**98.8**	**99.2**	**93.6**	**99.6**	**91.8**	**98.6**	**95.0**
Miss2	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 5.4.1. MIMIC model rejection rates of RMSEA at various cutoff values under normality

Cutoff Value	Complex Model				
	Sample Size				
	100		250	500	1000
	ML	SB			
0.045					
True	22.0	30.0	0.2	0.0	0.0
Miss1	90.8	95.0	98.6	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0
0.05					
True	**15.6**	**23.0**	**0.0**	**0.0**	**0.0**
Miss1	**86.2**	**91.8**	**96.0**	**100.0**	**100.0**
Miss2	**99.6**	**100.0**	**100.0**	**100.0**	**100.0**
0.055					
True	9.4	16.0	0.0	0.0	0.0
Miss1	80.4	86.2	89.6	97.0	100.0
Miss2	98.8	99.4	100.0	100.0	100.0
0.06					
True	**5.4**	**10.8**	**0.0**	**0.0**	**0.0**
Miss1	**73.6**	**80.2**	**75.2**	**84.6**	**95.6**
Miss2	**97.6**	**98.8**	**100.0**	**100.0**	**100.0**
0.07					
True	1.4	3.0	0.0	0.0	0.0
Miss1	48.2	60.4	30.4	19.2	7.4
Miss2	93.0	95.4	99.6	100.0	100.0
0.08					
True	0.0	0.2	0.0	0.0	0.0
Miss1	23.0	34.6	2.2	0.0	0.0
Miss2	80.6	86.2	86.0	94.6	99.6
0.09					
True	0.0	0.0	0.0	0.0	0.0
Miss1	6.2	12.4	0.0	0.0	0.0
Miss2	53.0	66.0	41.2	33.8	25.6
0.10					
True	0.0	0.0	0.0	0.0	0.0
Miss1	1.2	3.0	0.0	0.0	0.0
Miss2	23.8	35.2	6.4	0.8	0.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 5.4.2. MIMIC model rejection rates of RMSEA at various cutoff values under non-normality

Cutoff Value	Moderate Non-normality								Severe Non-normality							
	Sample Size 100		250		500		1000		Sample Size 100		250		500		1000	
	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB	ML	SB
0.045																
True	87.6	61.4	47.4	3.2	2.6	0.0	0.0	0.0	98.0	88.2	95.2	14.8	69.8	0.0	7.0	0.0
Miss1	100.0	96.0	100.0	95.8	100.0	96.2	100.0	99.0	100.0	99.6	100.0	93.2	100.0	76.2	100.0	49.8
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.05																
True	**83.4**	**52.2**	**31.2**	**0.6**	**0.4**	**0.0**	**0.0**	**0.0**	**97.4**	**81.6**	**90.0**	**7.4**	**48.8**	**0.0**	**2.0**	**0.0**
Miss1	**99.4**	**94.6**	**99.8**	**88.2**	**100.0**	**83.4**	**100.0**	**82.4**	**100.0**	**99.0**	**100.0**	**82.6**	**100.0**	**42.2**	**100.0**	**9.2**
Miss2	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**99.2**	**100.0**	**98.2**
0.055																
True	77.8	43.2	18.2	0.0	0.0	0.0	0.0	0.0	96.8	75.8	81.2	3.0	27.6	0.0	0.8	0.0
Miss1	98.8	92.0	99.2	74.4	99.4	53.4	100.0	34.0	100.0	97.6	100.0	66.6	100.0	15.4	100.0	0.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.4	100.0	92.8	100.0	84.6
0.06																
True	69.4	33.0	7.0	0.0	0.0	0.0	0.0	0.0	94.0	67.0	70.8	1.6	13.4	0.0	0.0	0.0
Miss1	97.0	86.0	98.0	50.8	98.0	20.8	98.2	4.4	99.8	95.4	99.6	45.0	100.0	3.2	99.0	0.0
Miss2	100.0	99.0	100.0	99.4	100.0	99.0	100.0	100.0	100.0	100.0	100.0	96.0	100.0	75.2	100.0	44.8
0.07																
True	49.0	13.6	1.2	0.0	0.0	0.0	0.0	0.0	88.0	46.2	37.4	0.2	1.0	0.0	0.0	0.0
Miss1	92.8	70.8	82.6	11.2	65.0	0.0	40.8	0.0	99.0	86.2	97.8	10.2	92.4	0.0	76.0	0.0
Miss2	100.0	96.2	100.0	85.6	100.0	75.6	100.0	66.6	100.0	98.0	100.0	66.6	100.0	12.6	100.0	0.2
0.08																
True	28.6	5.4	0.0	0.0	0.0	0.0	0.0	0.0	76.0	26.0	14.4	0.0	0.2	0.0	0.0	0.0
Miss1	82.6	46.4	47.2	0.6	11.6	0.0	0.8	0.0	96.6	68.8	87.0	1.0	59.2	0.0	17.2	0.0
Miss2	98.2	83.0	99.0	42.8	98.8	10.6	99.8	1.2	100.0	91.0	100.0	21.2	99.6	0.2	99.6	0.0
0.09																
True	11.6	1.4	0.0	0.0	0.0	0.0	0.0	0.0	60.8	12.2	5.8	0.0	0.0	0.0	0.0	0.0
Miss1	63.2	20.2	13.2	0.0	2.8	0.0	0.0	0.0	89.8	44.6	62.2	0.2	17.2	0.0	1.0	0.0
Miss2	93.2	63.2	86.0	6.8	74.6	0.0	59.4	0.0	99.4	74.8	98.4	2.2	92.4	0.0	83.8	0.0
0.10																
True	4.2	0.4	0.0	0.0	0.0	0.0	0.0	0.0	42.2	3.2	1.8	0.0	0.0	0.0	0.0	0.0
Miss1	41.6	5.8	1.2	0.0	0.0	0.0	0.0	0.0	79.6	25.0	31.0	0.0	2.6	0.0	0.0	0.0
Miss2	78.4	37.6	44.4	0.0	13.2	0.0	2.4	0.0	96.0	49.0	85.4	0.0	57.2	0.0	18.8	0.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 5.5.1. MIMIC model rejection rates of SRMR at various cutoff values under normality

Cutoff Value	Complex Model			
	Sample Size 100	250	500	1000
0.045				
True	86.8	0.8	0.0	0.0
Miss1	100.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0
0.05				
True	60.6	0.0	0.0	0.0
Miss1	100.0	96.6	95.0	96.4
Miss2	100.0	100.0	100.0	100.0
0.055				
True	33.8	0.0	0.0	0.0
Miss1	98.8	88.4	74.8	60.8
Miss2	100.0	100.0	100.0	100.0
0.06				
True	13.6	0.0	0.0	0.0
Miss1	94.6	65.6	39.0	13.4
Miss2	99.8	98.2	97.0	98.8
0.07				
True	**0.6**	**0.0**	**0.0**	**0.0**
Miss1	**66.4**	**13.8**	**1.6**	**0.0**
Miss2	**93.8**	**68.2**	**46.0**	**21.4**
0.08				
True	0.0	0.0	0.0	0.0
Miss1	31.6	1.6	0.0	0.0
Miss2	62.6	14.0	1.2	0.0
0.09				
True	0.0	0.0	0.0	0.0
Miss1	6.2	0.0	0.0	0.0
Miss2	28.2	0.6	0.0	0.0
0.10				
True	0.0	0.0	0.0	0.0
Miss1	1.2	0.0	0.0	0.0
Miss2	5.0	0.0	0.0	0.0
0.11				
True	0.0	0.0	0.0	0.0
Miss1	0.2	0.0	0.0	0.0
Miss2	0.4	0.0	0.0	0.0

Table 5.5.2. MIMIC model rejection rates of SRMR at various cutoff values under non-normality

Cutoff Value	Moderate Non-normality				Severe Non-normality			
	Sample Size				Sample Size			
	100	250	500	1000	100	250	500	1000
0.045								
True	99.0	8.2	0.0	0.0	99.6	46.4	1.4	0.0
Miss1	100.0	100.0	100.0	99.8	100.0	100.0	99.8	99.2
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.05								
True	92.0	1.2	0.0	0.0	97.2	19.2	0.0	0.0
Miss1	100.0	99.4	96.2	94.4	100.0	100.0	97.0	91.8
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.055								
True	74.0	0.2	0.0	0.0	88.4	5.4	0.0	0.0
Miss1	99.6	93.4	83.0	67.6	100.0	96.4	85.0	70.2
Miss2	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0
0.06								
True	47.4	0.0	0.0	0.0	72.6	2.2	0.0	0.0
Miss1	98.2	77.8	54.0	27.4	99.4	85.0	62.6	39.8
Miss2	100.0	99.4	98.4	98.4	100.0	99.6	97.8	97.2
0.07								
True	**9.8**	**0.0**	**0.0**	**0.0**	**32.8**	**0.2**	**0.0**	**0.0**
Miss1	**83.6**	**33.4**	**6.8**	**0.4**	**93.4**	**45.8**	**16.8**	**4.2**
Miss2	**97.4**	**80.6**	**64.6**	**33.2**	**99.4**	**87.0**	**69.4**	**46.6**
0.08								
True	0.4	0.0	0.0	0.0	7.4	0.0	0.0	0.0
Miss1	53.0	6.4	0.2	0.0	64.4	16.8	3.0	0.6
Miss2	80.8	32.2	5.2	0.2	89.6	49.4	18.4	3.0
0.09								
True	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0
Miss1	20.0	1.2	0.0	0.0	35.4	4.2	0.8	0.0
Miss2	46.4	4.4	0.0	0.0	61.0	13.0	2.0	0.2
0.10								
True	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Miss1	4.6	0.0	0.0	0.0	18.2	1.0	0.2	0.0
Miss2	15.2	0.6	0.0	0.0	30.4	3.4	0.4	0.0
0.11								
True	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	1.0	0.0	0.0	0.0	6.8	0.0	0.0	0.0
Miss2	3.4	0.0	0.0	0.0	13.2	0.2	0.0	0.0

Table 5.6.1. MIMIC model rejection rates of WRMR at various cutoff values under normality

Cutoff Value	Complex Model			
	Sample Size 100	250	500	1000
0.60				
True	79.2	80.2	80.6	79.6
Miss1	100.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0
0.70				
True	40.4	36.2	33.2	32.8
Miss1	98.6	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0
0.80				
True	12.6	6.4	7.4	5.8
Miss1	87.2	100.0	100.0	100.0
Miss2	98.6	100.0	100.0	100.0
0.90				
True	**1.2**	**1.2**	**1.0**	**0.4**
Miss1	**63.2**	**99.8**	**100.0**	**100.0**
Miss2	**90.0**	**100.0**	**100.0**	**100.0**
0.95				
True	**0.2**	**0.4**	**0.0**	**0.0**
Miss1	**45.2**	**98.8**	**100.0**	**100.0**
Miss2	**79.2**	**100.0**	**100.0**	**100.0**
1.00				
True	**0.0**	**0.4**	**0.0**	**0.0**
Miss1	**34.4**	**94.8**	**100.0**	**100.0**
Miss2	**64.6**	**100.0**	**100.0**	**100.0**
1.10				
True	0.0	0.0	0.0	0.0
Miss1	15.0	82.8	100.0	100.0
Miss2	40.2	99.8	100.0	100.0

Table 5.6.2. MIMIC model rejection rates of WRMR at various cutoff values under non-normality

Cutoff Value	Moderate Non-normality				Severe Non-normality			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.60								
True	89.8	81.8	75.6	72.8	93.6	89.2	86.0	76.0
Miss1	98.6	100.0	100.0	100.0	99.4	99.8	99.4	100.0
Miss2	100.0	100.0	100.0	100.0	99.6	99.8	100.0	100.0
0.70								
True	59.8	42.8	33.8	30.4	77.8	66.0	50.4	33.4
Miss1	94.4	99.0	99.8	100.0	95.6	97.4	99.0	100.0
Miss2	98.2	100.0	100.0	100.0	98.2	99.2	99.8	100.0
0.80								
True	25.4	13.6	9.2	5.4	49.2	33.0	19.2	9.6
Miss1	82.0	95.4	99.6	100.0	86.0	93.2	98.0	99.6
Miss2	93.0	99.4	100.0	100.0	93.0	97.2	99.6	100.0
0.90								
True	**7.8**	**2.8**	**0.8**	**0.8**	**23.0**	**12.0**	**4.8**	**1.4**
Miss1	**58.2**	**84.6**	**99.0**	**100.0**	**70.8**	**80.4**	**94.0**	**99.0**
Miss2	**81.4**	**97.0**	**99.8**	**100.0**	**83.4**	**94.6**	**98.4**	**100.0**
0.95								
True	**4.4**	**0.6**	**0.2**	**0.2**	**12.8**	**6.2**	**3.0**	**0.6**
Miss1	**43.8**	**77.2**	**96.8**	**100.0**	**58.2**	**70.6**	**89.8**	**98.0**
Miss2	**71.8**	**94.8**	**99.8**	**100.0**	**76.6**	**91.0**	**98.2**	**100.0**
1.00								
True	**1.4**	**0.2**	**0.0**	**0.0**	**8.4**	**3.8**	**1.4**	**0.0**
Miss1	**29.4**	**63.8**	**93.8**	**100.0**	**46.2**	**56.4**	**82.2**	**97.4**
Miss2	**59.2**	**92.0**	**99.6**	**100.0**	**66.6**	**84.6**	**97.2**	**99.8**
1.10								
True	0.0	0.0	0.0	0.0	3.0	1.0	0.4	0.0
Miss1	12.4	38.6	83.4	99.6	22.4	37.4	60.8	91.4
Miss2	31.2	80.2	98.8	100.0	46.6	70.2	94.0	98.6

Table 5.7. MIMIC model rejection rates of Chi-P at various cutoff values for binary outcomes

Cutoff Value	Equal Case				Unequal Case			
	Sample Size 100 [a]	250	500	1000	Sample Size 100 [b]	250	500	1000
0.01								
True	0.8	0.4	0.2	0.8	2.0	0.4	0.8	2.0
Miss1	16.6	71.6	98.4	100.0	13.0	55.4	94.2	100.0
Miss2	40.5	96.4	100.0	100.0	28.7	91.2	100.0	100.0
0.03								
True	5.2	2.6	1.6	2.8	3.6	2.6	2.4	3.6
Miss1	30.9	82.8	99.6	100.0	25.2	71.2	96.6	100.0
Miss2	57.1	98.4	100.0	100.0	45.1	95.2	100.0	100.0
0.04								
True	6.4	4.4	3.2	4.0	4.2	3.2	3.2	4.6
Miss1	36.7	85.8	99.6	100.0	29.0	75.2	96.8	100.0
Miss2	62.9	98.8	100.0	100.0	51.7	96.0	100.0	100.0
0.05								
True	**7.2**	**5.4**	**4.2**	**5.6**	**6.0**	**4.2**	**4.0**	**5.2**
Miss1	**41.3**	**86.8**	**99.8**	**100.0**	**33.9**	**78.2**	**97.2**	**100.0**
Miss2	**67.1**	**99.0**	**100.0**	**100.0**	**57.4**	**96.8**	**100.0**	**100.0**
0.06								
True	9.2	7.4	5.4	6.2	8.9	5.8	4.8	6.6
Miss1	43.5	87.8	100.0	100.0	37.9	79.4	97.2	100.0
Miss2	70.7	99.4	100.0	100.0	61.6	97.0	100.0	100.0

Note. [a] For the True models, 498 out of 500 data sets have converged results; 499 out of 500 data sets have converged results for the Miss1 and Miss2 models. [b] There are 497, 493 and 495 converged results for the true, Miss1 and Miss2 models, respectively.

Table 5.8. MIMIC model rejection rates of TLI at various cutoff values for binary outcomes

Cutoff Value	Equal Case				Unequal Case			
	Sample Size 100 [a]	250	500	1000	Sample Size 100 [b]	250	500	1000
0.90								
True	2.4	0.0	0.0	0.0	4.6	0.0	0.0	0.0
Miss1	20.6	4.4	0.6	0.0	23.7	5.8	0.8	0.2
Miss2	47.1	30.0	23.0	15.6	46.3	32.2	20.6	17.6
0.93								
True	8.2	0.0	0.0	0.0	13.5	0.0	0.0	0.0
Miss1	41.9	24.8	11.6	6.4	43.2	25.6	13.0	5.8
Miss2	67.7	69.8	76.4	89.4	68.3	69.6	74.0	84.8
0.94								
True	12.2	0.0	0.0	0.0	17.5	0.0	0.0	0.0
Miss1	50.5	37.2	28.8	21.2	52.3	36.0	28.4	19.4
Miss2	75.4	83.4	89.6	97.6	74.5	79.0	87.8	96.4
0.95								
True	**16.3**	**0.0**	**0.0**	**0.0**	**24.7**	**0.2**	**0.0**	**0.0**
Miss1	**60.1**	**52.8**	**49.6**	**51.2**	**60.9**	**51.2**	**47.8**	**49.0**
Miss2	**82.6**	**92.4**	**98.0**	**99.6**	**79.0**	**88.6**	**94.6**	**99.4**
0.96								
True	24.3	0.6	0.0	0.0	33.2	2.0	0.0	0.0
Miss1	68.3	70.2	72.8	85.2	68.2	65.6	70.8	80.2
Miss2	86.4	96.0	99.8	100.0	85.5	94.6	99.0	100.0

Note. [a] For the True models, 498 out of 500 data sets have converged results; 499 out of 500 data sets have converged results for the Miss1 and Miss2 models. [b] There are 497, 493 and 495 converged results for the true, Miss1 and Miss2 models, respectively.

Table 5.9. MIMIC model rejection rates of CFI at various cutoff values for binary outcomes

Cutoff Value	Equal Case				Unequal Case			
	Sample Size 100 [a]	250	500	1000	Sample Size 100 [b]	250	500	1000
0.90								
True	2.2	0.0	0.0	0.0	4.8	0.0	0.0	0.0
Miss1	21.2	7.4	1.2	0.4	22.7	5.8	1.0	0.2
Miss2	47.7	38.6	40.2	50.0	45.7	34.8	30.8	33.8
0.93								
True	8.6	0.0	0.0	0.0	11.9	0.0	0.0	0.0
Miss1	41.9	30.6	23.2	18.2	41.0	27.8	18.2	9.4
Miss2	68.5	78.2	89.4	98.0	68.7	73.0	81.2	94.0
0.94								
True	12.0	0.0	0.0	0.0	17.5	0.0	0.0	0.0
Miss1	51.1	44.2	40.2	44.0	52.1	38.8	32.6	28.6
Miss2	77.0	89.0	96.8	99.4	75.6	82.8	91.0	98.4
0.95								
True	16.1	0.0	0.0	0.0	24.1	0.4	0.0	0.0
Miss1	61.3	58.6	62.2	72.4	59.2	53.6	53.2	61.0
Miss2	83.4	94.6	99.2	100.0	79.4	91.0	97.2	99.8
0.96								
True	**25.7**	**1.4**	**0.0**	**0.0**	**32.6**	**1.8**	**0.0**	**0.0**
Miss1	**69.5**	**75.2**	**82.4**	**93.2**	**67.1**	**68.6**	**75.8**	**87.2**
Miss2	**88.2**	**97.4**	**99.8**	**100.0**	**85.1**	**95.4**	**99.4**	**100.0**

Note. [a] For the True models, 498 out of 500 data sets have converged results; 499 out of 500 data sets have converged results for the Miss1 and Miss2 models. [b] There are 497, 493 and 495 converged results for the true, Miss1 and Miss2 models, respectively.

Table 5.10. MIMIC model rejection rates of RMSEA at various cutoff values for binary outcomes

Cutoff Value	Equal Case				Unequal Case			
	Sample Size 100 [a]	250	500	1000	Sample Size 100 [b]	250	500	1000
0.045								
True	31.1	0.6	0.0	0.0	34.6	0.4	0.0	0.0
Miss1	73.3	72.8	72.4	83.4	70.6	58.0	55.4	57.0
Miss2	90.6	96.6	99.8	100.0	87.1	91.8	96.6	99.4
0.05								
True	**22.5**	**0.2**	**0.0**	**0.0**	**27.4**	**0.0**	**0.0**	**0.0**
Miss1	**68.3**	**58.4**	**53.6**	**57.0**	**63.5**	**44.6**	**33.4**	**23.6**
Miss2	**86.6**	**93.4**	**99.0**	**99.6**	**82.8**	**84.8**	**89.6**	**97.4**
0.055								
True	15.9	0.0	0.0	0.0	19.7	0.0	0.0	0.0
Miss1	59.9	43.6	32.4	26.0	54.4	29.4	15.2	4.8
Miss2	82.8	85.8	92.2	98.0	78.6	73.0	74.4	81.6
0.06								
True	**11.8**	**0.0**	**0.0**	**0.0**	**13.5**	**0.0**	**0.0**	**0.0**
Miss1	**50.7**	**30.4**	**13.6**	**8.0**	**47.1**	**17.4**	**5.0**	**0.8**
Miss2	**77.6**	**72.6**	**75.0**	**86.8**	**71.9**	**58.4**	**51.6**	**49.8**
0.07								
True	6.0	0.0	0.0	0.0	4.4	0.0	0.0	0.0
Miss1	34.7	8.4	1.0	0.2	28.6	3.8	0.2	0.0
Miss2	59.5	45.0	31.2	27.4	49.7	24.8	10.6	4.2
0.08								
True	1.0	0.0	0.0	0.0	2.4	0.0	0.0	0.0
Miss1	18.8	1.0	0.2	0.0	15.2	0.6	0.0	0.0
Miss2	42.1	17.4	4.6	0.8	33.5	6.4	0.2	0.0
0.09								
True	0.6	0.0	0.0	0.0	0.8	0.0	0.0	0.0
Miss1	8.2	0.2	0.0	0.0	6.1	0.0	0.0	0.0
Miss2	24.6	3.8	0.2	0.0	15.8	1.0	0.0	0.0
0.10								
True	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0
Miss1	2.8	0.0	0.0	0.0	2.6	0.0	0.0	0.0
Miss2	11.8	0.4	0.0	0.0	7.5	0.4	0.0	0.0

Note. [a] For the True models, 498 out of 500 data sets have converged results; 499 out of 500 data sets have converged results for the Miss1 and Miss2 models. [b] There are 497, 493 and 495 converged results for the true, Miss1 and Miss2 models, respectively.

Table 5.11. MIMIC model rejection rates of WRMR at various cutoff values for binary outcomes

Cutoff Value	Equal Case				Unequal Case			
	Sample Size 100 [a]	250	500	1000	Sample Size 100 [b]	250	500	1000
0.60								
True	100.0	99.8	100.0	100.0	100.0	99.6	100.0	100.0
Miss1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.70								
True	92.0	88.8	87.0	87.6	95.2	93.8	92.6	93.0
Miss1	99.6	100.0	100.0	100.0	99.8	100.0	100.0	100.0
Miss2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.80								
True	51.0	37.0	32.4	31.0	64.0	54.6	49.0	44.0
Miss1	88.4	98.4	100.0	100.0	91.7	98.2	100.0	100.0
Miss2	96.8	100.0	100.0	100.0	97.2	100.0	100.0	100.0
0.90								
True	**11.2**	**4.6**	**2.6**	**2.2**	**20.1**	**7.6**	**4.4**	**5.4**
Miss1	**51.7**	**88.2**	**99.6**	**100.0**	**57.8**	**83.8**	**97.2**	**100.0**
Miss2	**80.2**	**99.4**	**100.0**	**100.0**	**80.8**	**98.6**	**100.0**	**100.0**
0.95								
True	**4.6**	**1.0**	**0.2**	**0.6**	**7.6**	**1.2**	**0.8**	**2.4**
Miss1	**34.5**	**73.6**	**98.0**	**100.0**	**37.1**	**69.6**	**95.4**	**100.0**
Miss2	**61.9**	**98.0**	**100.0**	**100.0**	**66.7**	**95.4**	**100.0**	**100.0**
1.00								
True	**0.8**	**0.0**	**0.0**	**0.0**	**2.4**	**0.0**	**0.2**	**0.0**
Miss1	**20.0**	**56.2**	**95.4**	**100.0**	**20.7**	**48.6**	**91.6**	**100.0**
Miss2	**45.1**	**93.0**	**100.0**	**100.0**	**43.6**	**89.4**	**100.0**	**100.0**
1.10								
True	0.2	0.0	0.0	0.0	0.4	0.0	0.0	0.0
Miss1	2.6	21.8	75.2	100.0	3.7	16.2	65.8	99.8
Miss2	15.0	68.2	99.8	100.0	14.5	59.8	98.6	100.0

Note. [a] For the True models, 498 out of 500 data sets have converged results; 499 out of 500 data sets have converged results for the Miss1 and Miss2 models. [b] There are 497, 493 and 495 converged results for the true, Miss1 and Miss2 models, respectively.

Table 5.12. Means (SDs) for fit measures in the CFA and MIMIC models (normal outcomes)

Cutoff Value	CFA Model				MIMIC Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
Chi-P								
True	**0.367**	**0.422**	**0.456**	**0.498**	**0.368**	**0.429**	**0.461**	**0.475**
	(0.287)	(0.273)	(0.295)	(0.290)	(0.288)	(0.286)	(0.286)	(0.292)
Miss1	**0.023**	**0.000**	**0.000**	**0.000**	**0.023**	**0.000**	**0.000**	**0.000**
	(0.066)	(0.000)	(0.000)	(0.000)	(0.056)	(0.001)	(0.000)	(0.000)
TLI								
True	**0.990**	**0.998**	**0.999**	**1.000**	**0.990**	**0.998**	**0.999**	**1.000**
	(0.022)	(0.008)	(0.004)	(0.002)	(0.021)	(0.008)	(0.004)	(0.002)
Miss1	**0.931**	**0.939**	**0.939**	**0.940**	**0.936**	**0.943**	**0.943**	**0.944**
	(0.029)	(0.014)	(0.009)	(0.006)	(0.027)	(0.013)	(0.009)	(0.006)
CFI								
True	**0.989**	**0.997**	**0.998**	**0.999**	**0.989**	**0.997**	**0.999**	**0.999**
	().013)	(0.004)	(0.002)	(0.001)	(0.013)	(0.004)	(0.002)	(0.001)
Miss1	**0.944**	**0.950**	**0.951**	**0.951**	**0.948**	**0.954**	**0.954**	**0.955**
	(0.024)	(0.011)	(0.007)	(0.005)	(0.022)	(0.011)	(0.007)	(0.004)
RMSEA								
True	**0.025**	**0.013**	**0.008**	**0.005**	**0.025**	**0.013**	**0.008**	**0.005**
	(0.022)	(0.013)	(0.009)	(0.006)	(0.022)	(0.013)	(0.009)	(0.006)
Miss1	**0.072**	**0.070**	**0.069**	**0.069**	**0.068**	**0.066**	**0.065**	**0.065**
	(0.018)	(0.008)	(0.005)	(0.004)	(0.016)	(0.008)	(0.005)	(0.003)
SRMR								
True	**0.053**	**0.034**	**0.024**	**0.017**	**0.052**	**0.033**	**0.023**	**0.016**
	(0.007)	(0.004)	(0.003)	(0.002)	(0.007)	(0.004)	(0.003)	(0.002)
Miss1	**0.079**	**0.066**	**0.062**	**0.059**	**0.075**	**0.063**	**0.059**	**0.056**
	(0.011)	(0.007)	(0.006)	(0.004)	(0.010)	(0.007)	(0.005)	(0.004)
WRMR								
True	**0.700**	**0.690**	**0.677**	**0.669**	**0.683**	**0.675**	**0.671**	**0.668**
	(0.097)	(0.090)	(0.087)	(0.083)	(0.095)	(0.083)	(0.081)	(0.078)
Miss1	**0.998**	**1.284**	**1.676**	**2.263**	**0.950**	**1.230**	**1.598**	**2.150**
	(0.145)	(0.143)	(0.153)	(0.161)	(0.134)	(0.137)	(0.147)	(0.146)

Table 5.13. Means (SDs) for fit measures in the CFA and MIMIC models (binary unequal outcomes)

Cutoff Value	CFA Model				MIMIC Model			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
Chi-P								
True	**0.344**	**0.423**	**0.454**	**0.495**	**0.341**	**0.410**	**0.441**	**0.466**
	(0.242)	(0.258)	(0.270)	(0.276)	(0.231)	(0.251)	(0.261)	(0.279)
Miss1	**0.171**	**0.055**	**0.004**	**0.000**	**0.160**	**0.045**	**0.005**	**0.000**
	(0.190)	(0.011)	(0.025)	(0.000)	(0.170)	(0.097)	(0.031)	(0.000)
TLI								
True	**0.985**	**0.997**	**0.999**	**1.000**	**0.976**	**0.995**	**0.999**	**1.000**
	(0.029)	(0.011)	(0.005)	(0.002)	(0.042)	(0.015)	(0.007)	(0.004)
Miss1	**0.961**	**0.971**	**0.973**	**0.973**	**0.932**	**0.947**	**0.950**	**0.950**
	(0.038)	(0.018)	(0.010)	(0.007)	(0.051)	(0.027)	(0.018)	(0.012)
CFI								
True	**0.980**	**0.993**	**0.997**	**0.999**	**0.970**	**0.991**	**0.996**	**0.998**
	(0.027)	(0.011)	(0.004)	(0.002)	(0.033)	(0.011)	(0.005)	(0.003)
Miss1	**0.956**	**0.963**	**0.963**	**0.961**	**0.932**	**0.946**	**0.947**	**0.946**
	(0.039)	(0.022)	(0.013)	(0.010)	(0.048)	(0.027)	(0.019)	(0.012)
RMSEA								
True	**0.033**	**0.015**	**0.009**	**0.005**	**0.032**	**0.015**	**0.009**	**0.006**
	(0.029)	(0.016)	(0.010)	(0.006)	(0.026)	(0.014)	(0.009)	(0.007)
Miss1	**0.057**	**0.048**	**0.046**	**0.046**	**0.055**	**0.047**	**0.046**	**0.046**
	(0.030)	(0.016)	(0.009)	(0.006)	(0.025)	(0.014)	(0.009)	(0.006)
WRMR								
True	**0.784**	**0.749**	**0.736**	**0.726**	**0.833**	**0.804**	**0.796**	**0.791**
	(0.101)	(0.082)	(0.067)	(0.064)	(0.084)	(0.067)	(0.063)	(0.065)
Miss1	**0.877**	**0.939**	**1.062**	**1.307**	**0.925**	**1.002**	**1.144**	**1.402**
	(0.118)	(0.111)	(0.103)	(0.119)	(0.096)	(0.101)	(0.113)	(0.116)

Table 5.14. A summary of suitable cutoff criteria under various model and data conditions (MIMIC complex models[1])

Type of outcomes	Sample sizes	Cutoff values[2] to indicate good models at certain N	Criteria across all four sample size
Normal	N = 100 N = 250 N ≥ 500	 WRMR (.9[4]) WRMR (.9) *{SRMR (.07 or .08) lacks power}*	Chi-P[3] (.01), TLI (.95), CFI (.96), RMSEA (.06).
Moderately non-normal continuous	N = 100 N = 250 N ≥ 500	 **SB Chi-P[5] (.01)**, TLI (.95), CFI (.96), **SB RMSEA (.05)**, WRMR (.9). SB Chi-P (.01), TLI (.95), CFI (.96), RMSEA (.05), WRMR (.9). *{RMSEA (.06) and SRMR (.07 or .08) lack power}*	
Severely non-normal Continuous	N = 100 N = 250 N ≥ 500	 **SB CFI (.95)**, SRMR (.07), **SB RMSEA (.05),** WRMR (.95). **SB Chi-P (.01)**, TLI (.95), CFI (.95), WRMR (.95). *{SRMR (.07) lacks power}*	
binary equal	N = 100 N ≥ 250 N ≥ 500	 Chi-P (.05), CFI (.95), RMSEA (.045), WRMR (.9).	
binary unequal	N = 100 N ≥ 250 N ≥ 500	 Chi-P (.05), CFI (.95), WRMR (.9).	

[1] Applied to CFA complex models as well except for binary outcomes; [2] These suggested cutoff criteria have type I and type II error rates close to or lower than 5% and 30%, respectively. [3] The fit measures not denoted specifically are ML-based. [4] Values in parentheses are the suggested cutoff values. [5] The Satorra-Bentler based fit indices.

CHAPTER 6

MONTE CARLO STUDY 3: LATENT GROWTH CURVE MODEL

One major method for analyzing longitudinal data and studying change is growth curve modeling. It allows us to model individual differences in development over time and to explain individual differences using background variables. Individual differences in growth are captured by random coefficients in growth curve models. Latent growth curve modeling incorporates growth curve modeling into the framework of LVM, and random coefficients are conceptualized as latent variables (see, e.g., Muthén & Khoo, 1998). Different from the first and second studies where the mean values of latent variables and the intercepts of measurement equations were ignored, latent growth modeling (LGM) takes into account mean structures. Therefore, it is valuable to study the performance of the fit indices in this new situation.

6.1 Design of Simulation

Two quadratic LGMs with normal continuous outcomes were generated: one had five and the other had eight time points. The quadratic LGMs can be expressed as

$$y = \Lambda \xi + \varepsilon, \tag{18}$$

where y is a T x 1 vector of observed variables for the T time points, Λ is a T x 3 matrix of factor loadings, ξ is a 3 x 1 vector of growth factors, and ε is a T x 1 vector of errors. For the quadratic LGM with five time points,

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} .$$

The coefficients in the second column of Λ were for modeling linear growth whereas those in the third column were for modeling quadratic growth. The population means for the intercept, linear and quadratic growth factors were set to be 0, 0.5 and –0.1, respectively. The variances of the three growth factors were 1, 0.3 and 0.1, which are common variance ratios among growth factors. The intercept and linear factors had a correlation of 0.25, and the intercepts of the outcome variables were fixed to zero. The five outcomes all had a constant R^2 of 0.5. Hence, the residual variances increased over time due to increasing growth curve variance. The parameter values for the LGM with eight time points stayed the same.

Sample sizes of 100, 250, 500 and 1000 were considered, and 500 replications were obtained for each condition. Two specifications of the LGMs were fitted to sample mean and covariance matrices generated from each replication. The first model was properly specified (True model) such that the estimated parameters in the sample exactly corresponded to their population structure (quadratic growth curve). The LGM true models with five and eight time points had twelve and fifteen parameters, respectively. The second model was a misspecified model in which a linear growth curve was estimated (Miss1 model). An adequate cutoff value of fit indices should have a high probability of not rejecting true models and a high probability of rejecting the misspecified linear growth models. Data was generated in SAS and analyzed by Mplus 2.02. The maximum number of iterations to convergence was set to 1000 in Mplus by default, and all the runs were converged. At N = 100, both ML and SB estimation results were provided for Chi-P, TLI, CFI and RMSEA.

6.2 Results

Tables 6.1 to 6.6 present the rejection rates of the Chi-P, TLI, CFI, RMSEA, SRMR and WRMR under various cutoff values, and Table 6.7 provides the means and SDs of the six fit measures for the LGMs with five and eight time points. The rejection rates of Chi-P at a conventional probability level of 0.05 in true models varied around the expected value of 5% (ranged from 5.6% to 7%). Chi-P at a cutoff of 0.05 rejected 59% to 100% of the LGM Miss1 models with five time points and rejected all the LGM Miss1 models with eight time points. As shown in Table 6.7, Chi-P had higher values and larger spread (tended to accept Miss1 models) under the combination of five time points and N =100. Moreover, the results show that the power of Chi-P to reject linear growth models increased with larger sample sizes and more time points.

At N = 100, a cutoff value of 0.95 for TLI rejected around 16% of the True models and 74% of the Miss1 models in the five-point LGM. With more time points, the type I error rates of TLI at a cutoff of 0.95 decreased and its power increased. For example, Table 6.2 shows that in the eight-point LGM, a cutoff value of 0.95 for TLI rejected only about 8% of the True models and 100% of the misspecified models. CFI had very similar means, SDs and rejection rates as TLI in the five-point LGM Miss1 models. CFI at a cutoff value of 0.95 rejected 11.8% of the five-point true models, which was more desirable than the 15.6% of TLI. A cutoff value of 0.96 for CFI rejected even more true models and was less suitable for LGMs under the combination of fewer time points and N = 100. A cutoff value of 0.94 for CFI appeared to have type I error rates around or lower than 5% and type II error rates around or lower than 30% for both five- and eight-point LGMs.

A cutoff value of 0.06 for RMSEA rejected 24.8% of the five-point LGM True models and 12.6% of the eight-point LGM True models at N = 100. With a cutoff value of 0.05, RMSEA had inflated type I error rates not only at N = 100 but also had inflated type I errors in the five-point LGM (rejected 12.4% of the True models) at N = 250. Similar to TLI and CFI under the suggested cutoff values, the type I error rates of RMSEA decreased and the power of RMSEA increased with a greater number of time points. With fewer time points and at small sample sizes, one might want to increase the cutoff value to 0.09 or 0.10 for RMSEA to control type I error rates. In doing so, however, the power of RMSEA to detect Miss1 models decreased to 0.5 or 0.6. Similar to previous CFA and MIMIC studies, in LGMs the SB-based fit measures tended to have slightly higher rejection rates than their ML counterparts with normal data. Overall, the SB χ^2 and the SB–based TLI/ CFI/ RMSEA had very similar rejection rates as their ML counterparts in LGMs.

A cutoff value of 0.08 for SRMR rejected 4.4% of the True models and 73% of the misspecified models at N = 100 in the five-point LGM. However, with this cutoff value, the power of SRMR decreased with increasing sample sizes and it only rejected 31% of the Miss1 models at N = 1000. A cutoff value of 0.07 for SRMR might be more suitable in terms of power at larger sample sizes. At N ≥ 250, SRMR at a cutoff of 0.07 rejected 77% to 100% of the Miss1 models with very small type I errors.

For the LGM with five time points, WRMR at a cutoff value of 1.0 rejected 3% to 6%

of the true models and 76% to 100% of the Miss1 models. For the LGM with eight time points, however, the type I error rates of WRMR inflated tremendously. From Table 6.7 we can see that, similar to SRMR, means of WRMR increased with increasing number of time points in both True and Miss1 models. The means of WRMR shifted very close to 0.90 and 0.95 in the eight-point LGM thus resulting in large rejection rates of the True and Miss1 models. This indicated that a cutoff value of 0.95 or 1.0 for WRMR was not suitable for latent growth curve models with more time points.

The scatterplots in Figure 6.1 show that the relationship between CFI and WRMR at N = 500 are similar under the two LGM true models (the correlation coefficient was –0.75 in the five-point LGM, and –0.63 in the eight-point LGM). However, while the CFI values ranged similarly (from 0.97 to 1.0) for *both* LGM true models, the WRMR values tended to be higher for the eight-point LGM true model (WRMR ranged from 0.2 to 1.2 for the five-point LGM, and 0.5 to 1.6 for the eight-point LGM). Similarly, Figure 6.2 shows that the relationship between SRMR and WRMR stays very strong under both LGM models (the correlation coefficients were 0.99 and 0.98 for the five-point and eight-point LGMs, respectively). SRMR ranged from 0.01 to 0.04 for the five-point LGM and 0.02 to 0.05 for the eight-point LGM, thus a cutoff value of 0.07 for SRMR rejected none of the true models. On the other hand, a cutoff value of 1.0 for WRMR rejected 3.2% of the five-point and 26.6% of the eight-point LGM true models. Because the WRMR values tended to increase significantly in LGMs with more time points, a certain cutoff value for WRMR tended to reject more true models with more time points.

6.3 Summary and Conclusion

This chapter evaluated the performance and obtained adequate cutoff criteria for Chi-P, TLI, CFI, RMSEA, SRMR and WRMR in latent growth curve models with normal data. For the LGMs with five time points and at $N \geq 250$, Chi-P ≥ 0.05, TLI ≥ 0.95, CFI ≥ 0.95, RMSEA ≤ 0.06 and SRMR ≤ 0.07 were suitable criteria to indicate good models. In each of the case, the type I error rate was reasonable and the power was higher than 0.8. At $N = 100$ and fewer time points, TLI, CFI and RMSEA under the suggested cutoff values tended to overreject true models. Chi-P at the suggested cutoff value of 0.05 had an acceptable type I error rate (5.6%) with moderate power around 0.6 at $N = 100$. CFI at the cutoff of 0.95 had inflated type I error rate (11.8%) with power of around 0.74. A cutoff value of 0.94 for CFI rejected 0 to 7% of the True models and 69% to 97% of the Miss1 models, and it seemed to be applicable across all four sample sizes. WRMR at a cutoff of 1.0 had acceptable type I and type II error rates across samples in the five-point LGM, but it overrejected true models in the eight-point LGM.

For the LGM with eight time points, Chi-P ≥ 0.05, TLI ≥ 0.95 and CFI ≥ 0.95 (0.94) were suitable indications of good models across all four sample sizes. Different cutoff values seemed to be necessary for WRMR in LGMs with different time points, which might not be desirable. Finally, considering both LGMs, the results show that Chi-P at a cutoff of 0.05 performed well across all sample sizes (with moderate power under the combination of small sample size and fewer time points). CFI at a cutoff of 0.94 had type I error rates around 5% and power higher than 0.7, and it was suitable for latent growth curve models across samples.

Generally speaking, the residual-based fit indices SRMR and WRMR performed differently from the other fit indices in latent growth curve models. With more time points, the WRMR and SRMR values on average tended to increase considerably in both true and misspecified models. Therefore, WRMR and SRMR tended to overreject true models, and the adequate cutoff value for WRMR and SRMR might vary with the number of time points. In contrast, TLI, CFI and RMSEA tended to have fewer type I errors in the LGM with more time points. Chi-P appeared to perform similarly for latent growth curve models with varying time points.

Figure 6.1. Scatterplot of the relationship between WRMR and CFI in true models based on 500 observations.

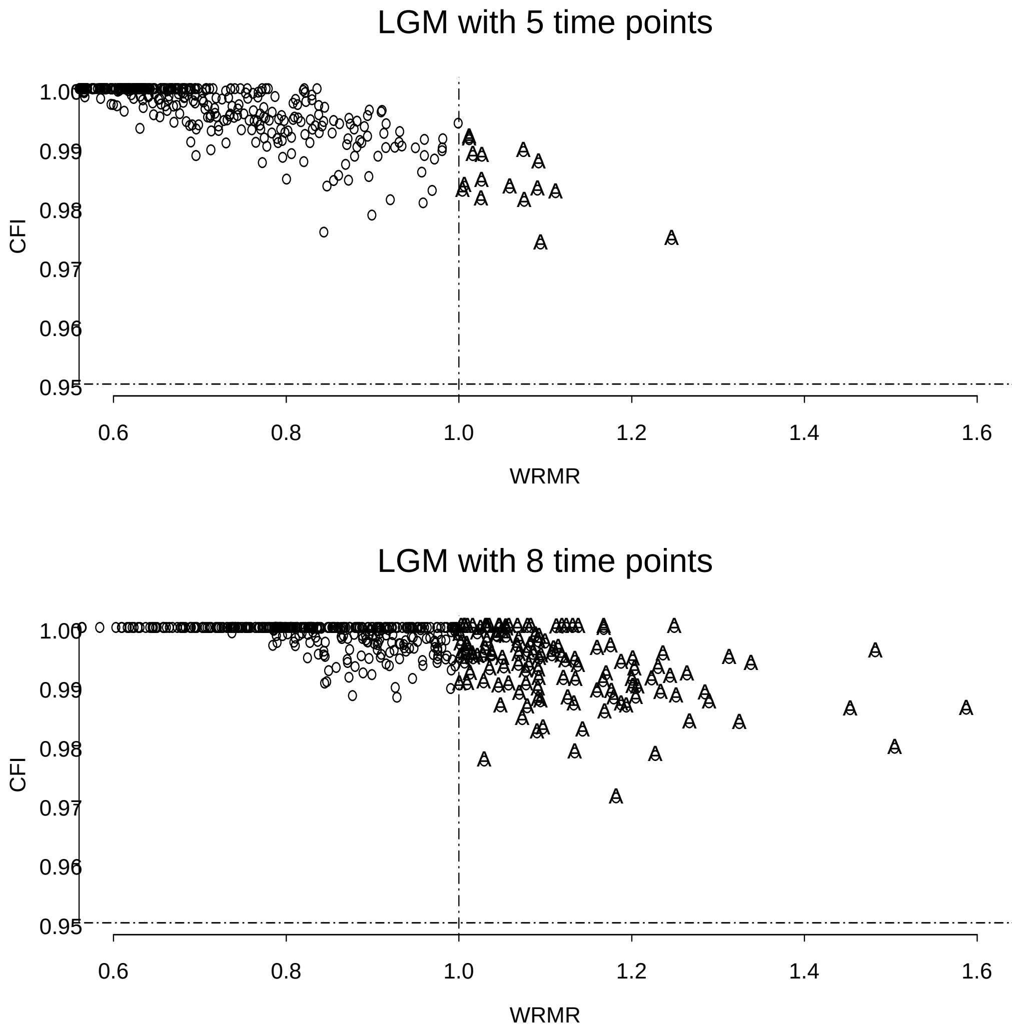


Figure 6.2. Scatterplot of the relationship between WRMR and SRMR in true models based on 500 observations.

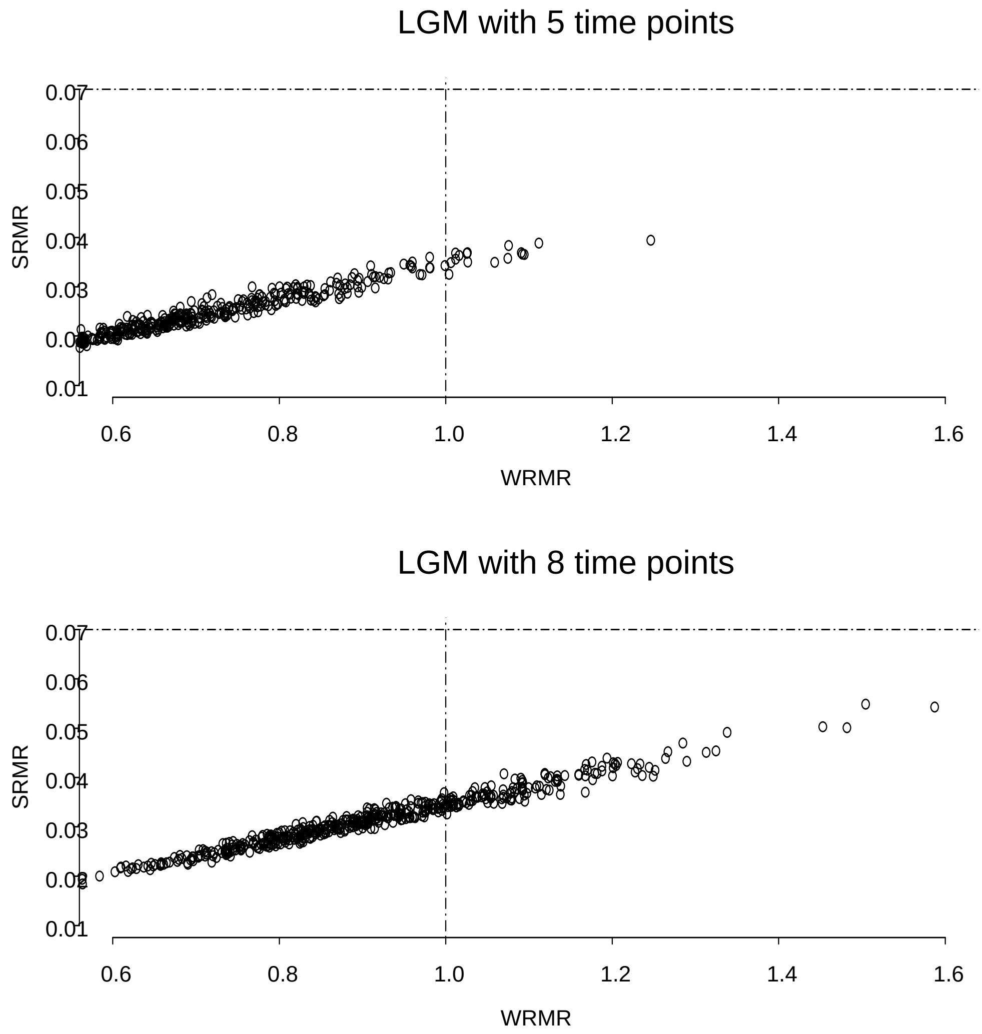


Table 6.1. LGM model rejection rates of Chi-P at various cutoff values

Cutoff Value	Five Time Points					Eight Time Points				
	Sample Size 100		250	500	1000	Sample Size 100		250	500	1000
	ML	SB				ML	SB			
0.01										
True	1.0	0.8	1.6	0.6	0.4	1.0	1.4	1.0	1.6	1.2
Miss1	33.0	34.0	85.2	100.0	100.0	99.4	99.8	100.0	100.0	100.0
0.03										
True	2.4	3.4	4.4	2.8	3.2	4.0	4.6	3.6	3.4	4.0
Miss1	50.0	52.0	92.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.04										
True	4.6	4.6	6.0	4.0	4.8	5.6	6.6	4.4	5.6	5.2
Miss1	55.0	56.8	94.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.05										
True	**5.6**	**6.6**	**7.0**	**4.8**	**6.4**	**7.0**	**7.8**	**5.4**	**7.2**	**6.0**
Miss1	**59.4**	**60.6**	**95.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
0.06										
True	6.6	7.4	7.2	5.4	7.8	7.8	10.0	6.2	8.2	7.4
Miss1	62.6	63.6	95.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 6.2. LGM model rejection rates of TLI at various cutoff values

Cutoff Value	Five Time Points					Eight Time Points				
	Sample Size 100		250	500	1000	Sample Size 100		250	500	1000
	ML	SB				ML	SB			
0.90										
True	2.8	3.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	42.0	40.2	39.0	36.6	29.6	98.4	98.2	100.0	100.0	100.0
0.93										
True	9.4	8.8	0.6	0.0	0.0	2.2	2.2	0.0	0.0	0.0
Miss1	61.8	63.2	69.2	80.6	89.2	100.0	100.0	100.0	100.0	100.0
0.94										
True	12.6	13.0	1.2	0.0	0.0	4.2	4.4	0.0	0.0	0.0
Miss1	69.0	68.8	76.0	91.0	97.2	100.0	100.0	100.0	100.0	100.0
0.95										
True	**15.6**	**16.0**	**1.2**	**0.0**	**0.0**	**8.2**	**9.2**	**0.0**	**0.0**	**0.0**
Miss1	**73.8**	**74.6**	**83.2**	**95.2**	**99.8**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
0.96										
True	20.0	20.0	4.6	0.0	0.0	13.2	14.4	0.0	0.0	0.0
Miss1	77.4	78.0	90.4	98.8	99.8	100.0	100.0	100.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 6.3. LGM model rejection rates of CFI at various cutoff values

Cutoff Value	Five Time Points					Eight Time Points				
	Sample Size					Sample Size				
	100		250	500	1000	100		250	500	1000
	ML	SB				ML	SB			
0.90										
True	1.6	1.4	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0
Miss1	42.0	40.2	39.0	36.6	29.6	99.4	99.6	100.0	100.0	100.0
0.93										
True	4.6	5.2	0.2	0.0	0.0	2.2	2.4	0.0	0.0	0.0
Miss1	61.8	63.2	69.2	80.6	89.2	100.0	100.0	100.0	100.0	100.0
0.94										
True	7.0	7.0	0.6	0.0	0.0	5.2	5.4	0.0	0.0	0.0
Miss1	69.0	68.8	76.0	91.0	97.2	100.0	100.0	100.0	100.0	100.0
0.95										
True	**11.8**	**12.4**	**1.0**	**0.0**	**0.0**	**8.6**	**9.8**	**0.2**	**0.0**	**0.0**
Miss1	**73.8**	**74.6**	**83.2**	**95.2**	**99.8**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
0.96										
True	**15.6**	**16.0**	**2.0**	**0.0**	**0.0**	**14.2**	**15.2**	**0.4**	**0.0**	**0.0**
Miss1	**77.4**	**78.0**	**90.4**	**98.8**	**99.8**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 6.4. LGM model rejection rates of RMSEA at various cutoff values

Cutoff Value	Five Time Points					Eight Time Points				
	Sample Size					Sample Size				
	100		250	500	1000	100		250	500	1000
	ML	SB				ML	SB			
0.045										
True	34.0	35.2	16.4	4.0	0.2	26.8	29.4	4.4	0.2	0.0
Miss1	86.2	87.4	98.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.05										
True	30.6	32.2	12.4	2.6	0.0	21.8	24.2	1.8	0.0	0.0
Miss1	84.4	85.4	97.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.055										
True	28.0	29.2	9.0	0.6	0.0	16.6	20.0	0.6	0.0	0.0
Miss1	83.0	83.4	95.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.06										
True	**24.8**	**25.8**	**7.0**	**0.0**	**0.0**	**12.6**	**14.2**	**0.2**	**0.0**	**0.0**
Miss1	**81.2**	**80.8**	**94.4**	**99.6**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
0.07										
True	18.2	19.6	2.8	0.0	0.0	6.4	7.0	0.0	0.0	0.0
Miss1	75.2	76.8	88.2	97.6	99.8	100.0	100.0	100.0	100.0	100.0
0.08										
True	13.2	14.0	1.0	0.0	0.0	1.8	2.2	0.0	0.0	0.0
Miss1	69.0	69.6	77.6	91.4	97.4	99.8	99.8	100.0	100.0	100.0
0.09										
True	7.8	8.6	0.6	0.0	0.0	0.4	0.8	0.0	0.0	0.0
Miss1	60.0	62.2	65.6	75.8	84.2	98.2	98.6	100.0	100.0	100.0
0.10										
True	4.8	5.6	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	49.6	51.0	48.2	53.0	50.4	95.0	94.8	99.2	100.0	100.0
0.11										
True	1.6	2.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss1	38.6	40.6	30.2	27.2	18.4	87.2	89.6	97.0	100.0	100.0

Note. ML = maximum likelihood; SB = Satorra-Bentler.

Table 6.5. LGM model rejection rates of SRMR at various cutoff values

Cutoff Value	Five Time Points				Eight Time Points			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.045								
True	67.6	8.6	0.0	0.0	99.2	46.8	1.8	0.0
Miss1	99.4	99.6	100.0	100.0	100.0	100.0	100.0	100.0
0.05								
True	52.2	2.8	0.0	0.0	97.2	23.8	0.8	0.0
Miss1	98.6	98.8	99.6	100.0	100.0	100.0	100.0	100.0
0.055								
True	39.4	1.4	0.0	0.0	91.6	10.8	0.0	0.0
Miss1	97.2	96.2	98.4	99.6	100.0	100.0	100.0	100.0
0.06								
True	27.0	1.0	0.0	0.0	81.8	3.6	0.0	0.0
Miss1	92.8	92.6	95.6	98.6	100.0	100.0	100.0	100.0
0.07								
True	**11.0**	**0.2**	**0.0**	**0.0**	**53.4**	**0.4**	**0.0**	**0.0**
Miss1	**85.2**	**77.0**	**77.8**	**77.6**	**100.0**	**100.0**	**100.0**	**100.0**
0.08								
True	**4.4**	**0.0**	**0.0**	**0.0**	**24.8**	**0.0**	**0.0**	**0.0**
Miss1	**73.4**	**53.2**	**45.2**	**31.2**	**100.0**	**100.0**	**100.0**	**100.0**
0.09								
True	1.4	0.0	0.0	0.0	9.0	0.0	0.0	0.0
Miss1	57.2	30.2	18.8	5.0	100.0	100.0	100.0	100.0
0.10								
True	0.2	0.0	0.0	0.0	3.4	0.0	0.0	0.0
Miss1	38.6	11.2	4.4	0.4	100.0	100.0	100.0	100.0
0.11								
True	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0
Miss1	24.0	2.8	0.8	0.2	100.0	100.0	100.0	100.0

Table 6.6. LGM model rejection rates of WRMR at various cutoff values

Cutoff Value	Five Time Points				Eight Time Points			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
0.60								
True	66.0	59.2	59.8	59.8	98.8	98.4	98.6	98.8
Miss1	99.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.70								
True	45.0	40.8	34.2	34.4	93.4	92.8	91.8	90.8
Miss1	97.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.80								
True	27.0	21.2	18.6	18.6	81.4	75.8	73.6	73.2
Miss1	92.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.90								
True	13.0	9.8	7.6	8.8	60.0	55.4	48.8	49.0
Miss1	86.0	99.4	100.0	100.0	100.0	100.0	100.0	100.0
0.95								
True	**8.4**	**6.6**	**5.2**	**6.6**	**47.6**	**40.6**	**36.4**	**35.4**
Miss1	**82.0**	**99.2**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
1.00								
True	**6.0**	**3.8**	**3.2**	**4.2**	**37.0**	**30.0**	**26.6**	**27.2**
Miss1	**76.2**	**99.2**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
1.10								
True	2.8	1.4	0.4	1.4	17.2	13.0	11.4	11.2
Miss1	64.0	96.4	100.0	100.0	100.0	100.0	100.0	100.0

Table 6.7. Means (SDs) for fit measures in LGMs

Cutoff Value	Five Time Points				Eight Time Points			
	Sample Size 100	250	500	1000	Sample Size 100	250	500	1000
Chi-P								
True	**0.470**	**0.488**	**0.490**	**0.500**	**0.443**	**0.493**	**0.494**	**0.477**
	(0.290)	(0.291)	(0.284)	(0.285)	(0.287)	(0.277)	(0.293)	(0.292)
Miss1	**0.114**	**0.010**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	(0.185)	(0.043)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)
TLI								
True	**0.997**	**0.999**	**1.000**	**1.000**	**0.995**	**1.000**	**1.000**	**1.000**
	(0.047)	(0.020)	(0.009)	(0.004)	(0.030)	(0.011)	(0.006)	(0.002)
Miss1	**0.907**	**0.909**	**0.907**	**0.908**	**0.784**	**0.793**	**0.794**	**0.795**
	(0.071)	(0.040)	(0.027)	(0.018)	(0.060)	(0.035)	(0.024)	(0.017)
CFI								
True	**0.984**	**0.994**	**0.997**	**0.999**	**0.985**	**0.995**	**0.997**	**0.999**
	(0.026)	(0.011)	(0.005)	(0.002)	(0.021)	(0.008)	(0.004)	(0.001)
Miss1	**0.905**	**0.909**	**0.907**	**0.908**	**0.761**	**0.771**	**0.772**	**0.773**
	(0.067)	(0.040)	(0.027)	(0.018)	(0.066)	(0.039)	(0.026)	(0.019)
RMSEA								
True	**0.030**	**0.018**	**0.012**	**0.008**	**0.024**	**0.012**	**0.009**	**0.007**
	(0.037)	(0.024)	(0.016)	(0.011)	(0.026)	(0.016)	(0.012)	(0.008)
Miss1	**0.094**	**0.097**	**0.101**	**0.101**	**0.137**	**0.136**	**0.135**	**0.135**
	(0.043)	(0.023)	(0.015)	(0.010)	(0.023)	(0.013)	(0.009)	(0.007)
SRMR								
True	**0.052**	**0.032**	**0.022**	**0.016**	**0.072**	**0.045**	**0.031**	**0.022**
	(0.015)	(0.009)	(0.006)	(0.005)	(0.013)	(0.008)	(0.006)	(0.004)
Miss1	**0.095**	**0.082**	**0.079**	**0.077**	**0.166**	**0.155**	**0.151**	**0.15**
	(0.024)	(0.015)	(0.012)	(0.008)	(0.023)	(0.015)	(0.010)	(0.008)
WRMR								
True	**0.688**	**0.662**	**0.648**	**0.652**	**0.952**	**0.919**	**0.909**	**0.908**
	(0.191)	(0.184)	(0.170)	(0.179)	(0.174)	(0.161)	(0.163)	(0.161)
Miss1	**1.211**	**1.616**	**2.194**	**2.984**	**2.123**	**3.084**	**4.217**	**5.878**
	(0.288)	(0.288)	(0.308)	(0.311)	(0.278)	(0.289)	(0.282)	(0.302)

CHAPTER 7

REAL DATA ILLUSTRATION

To decide whether a model is a well-fitting one is not hard if all the fit indices lead to similar conclusions. However, in practice researchers often encounter inconsistent results for fit indices. Different cutoff criteria for fit indices in many cases also affect the decision. This chapter uses an actual database to illustrate the use of cutoff criteria for TLI, CFI, RMSEA, SRMR and WRMR in model selection. The initial data were gathered by Holzinger and Swineford (1939) and a theory-based bi-factor model was fitted to the data. This data set is a benchmark in factor analysis for comparing methods. The aim of this illustrative example is not to demonstrate how to derive a best-fitting model from the data, but to illustrate how to consult the previous simulation results and apply suitable cutoff criteria of the model fit measures to evaluate whether a model is reasonably consistent with the data and worthy of further investigation.

7.1 Holzinger and Swineford data and the bi-factor model

This set of data consists of a battery of 24 mental ability tests administrated to 145 seventh- and eighth-grade children from the Grant-White Elementary School in Forest Park, Illinois. The basic statistics for the 24 mental ability tests are presented in Table 7.1 and, to check the non-normality of the data, the univariate skewness and kurtosis are also provided.

The univariate skewness of these outcomes ranged from –0.77 to 1.18 and kurtosis ranged from –0.68 to 2.28. The majority of variables seemed to be normally distributed. The multivariate normality tests (based on Mardia's multivariate skewness and kurtosis measures) indicated that the value of kurtosis observed in the data was not significantly different from zero ($p = 0.08$) but the skewness was significantly different from zero ($p = .0003$). These variables were believed to measure mental abilities such as spatial relations (tests 1-4), verbal intelligence (tests 5-9), perceptual speed (tests 10-13), memory (tests 14-19) and deduction (tests 20-24).

Based on Spearman's two-factor theory in which intellectual activity was believed to consist of a general function and a specific function for each element (Spearman, 1904. cf. Harman, 1976), Holzinger and Swineford (1939) proposed the bi-factor solution and fitted the bi-factor solution to these twenty-four psychological tests. Their studies aimed to investigate whether the unique and simple bi-factor theory can be applicable to mental ability tests and psychological measures. Bi-factor refers to two different types of factors. One is a general factor which describes general mental abilities, and the other is an array of group factors which are measured by the grouping of variables. Harman (1976) analyzed the data using the bi-factor method, and a re-analysis of his final solution using Mplus is presented here. The parameter estimates and bi-factor pattern (one general factor and five group factors) are provided in Table 7.2[1].

Based on the bi-factor theory, the general factor accounts for the intercorrelations among group factors, thus these six factors are not correlated to each other. The residuals

---

[1] The bi-factor model was fitted to data by Mplus 2.02, and the input program can be found on the Mplus website at http://www.statmodel.com/mplus/examples/continuous/cont14.html.

of the variables “addition” (Test 10) and “arithmetic problems” (Test 24) were allowed to be correlated in estimation. After specifying a model based on substantive theory and fitting the model to the data, the χ^2 and fit indices can be helpful to assess the overall fit of the model. The χ^2 test of model fit for the baseline model with 276 degrees of freedom was 1639.347, $p < .0001$. The observed variables were uncorrelated in the baseline model, and the test result indicated that further modeling should be considered. The χ^2 test of model fit for the bi-factor model was 308.64 with 231 degrees of freedom ($p = .0005$). TLI = 0.932, CFI = 0.943, RMSEA = 0.048, SRMR = 0.06 and WRMR = 0.892. Given the results of these overall fit indices, does the bi-factor model represent an acceptable fit? What are the adequate cutoff criteria for the fit indices in the bi-factor model? This study attempts to tackle these questions.

7.2 Performance and suitable cutoff criteria of the fit measures

Because the factor-covariance structure of the bi-factor model is fixed (factors are not correlated to each other), the misspecification of factor loadings is a more important concern than the misspecification of factor covariances. According to the results of the earlier CFA study, when sample size is around 100 under normality for the complex models, the ML-based Chi-P at a cutoff of 0.01 rejected 4% of the true models and 75% to 99% of the misspecified models; TLI at 0.95 rejected 4% of the true models and 75% to 99% of the misspecified models; CFI at 0.96 rejected 4% of the true models and 74% to 99% of the misspecified models. RMSEA at a cutoff of 0.06, SRMR at 0.07 and WRMR at 0.9 all had type I error rates lower than 5% and type I error rates lower than 25% (power higher than

075). Note that a cutoff value of 0.9 was better than 0.95 for WRMR in terms of power in the CFA complex models. Inspection of the results in the CFA trivial models shows that these suggested cutoff criteria did not tend to overreject trivially misspecified models (rejection rates were less than 10%). However, according to these cutoff values these fit indices appear to disagree in this example. The ML-based Chi-P, TLI and CFI indicate a lack of fit, whereas RMSEA, SRMR and WRMR indicate an acceptable fit for the bi-factor model.

In this example, the bi-factor model has sixty-nine parameters and the data has a sample size of 145. Past research has indicated that with small sample sizes, the test statistic might not be χ^2 distributed and TLI might be underestimated. The earlier simulation results have shown that with sample sizes of around 100, some conventional cutoff values for fit measures tend to overreject properly specified models. Moreover, sample size, model complexity and the presence of non-normality appear to affect the estimates of the fit measures and the decisions of cut points. The number of parameters in the bi-factor model is much higher than in Boomsma's study and the previous CFA study of this dissertation (thirty-six parameters in the CFA complex true model). Greater sample sizes might be needed with increasing model complexity and/or a greater number of free parameters in a model. A sample size of 145 might not provide enough information for an accurate χ^2 approximation. To investigate whether the fit measures under the suggested cutoff values have strong power with minimum type I errors for the bi-factor model with N = 145, a small Monte Carlo study similar to the earlier studies was conducted. This study, however, is based on the estimated parameter values of the real-data example.

7.2.1 Design of A Small Monte Carlo Study

Five hundred data matrices with N = 145 were drawn from the bi-factor model structure with the parameter values in Table 7.2. The outcome variables had values of skewness and kurtosis close to those provided in Table 7.1. Two specifications of the bi-factor models were fitted to sample mean and covariance matrices generated from each of the 500 replications. The first model was properly specified such that the estimated parameters in the sample exactly corresponded to the bi-factor structure (True model). An evaluation of the properly specified models is important to gain insight into the behavior of the fit indices. The other was a misspecified model in which an important loading (Test 9 on the factor "verbal intelligence") was not estimated (Miss model). Runs that failed to converge within 3000 iterations or had improper solutions (negative estimates of variances, also called "Haywood cases") were discarded. As mentioned in Boomsma (1983, p. 30), at least three strategies can be used to deal with replications with improper solutions. The "e*xclusion"* and *"inclusion"* of improper solutions are two commonly used strategies in literature. Neither strategy is perfect. In this example a decision was made to *exclude* the replications with improper solutions. The differences between these two strategies in terms of rejection rates were compared, and the results did not vary significantly whether or not the replications of improper solutions were excluded. We would like to avoid incorrectly rejecting the bi-factor model if it is true, thus an important aim is to select cutoff values and fit measures that have type I error rates around or lower than 5%.

7.2.2 Results and Discussion

In this section the cutoff values suggested in the previous studies are evaluated, and the suitable cutoff criteria for the bi-factor model are provided. The agreements between pairs of the fit indices under certain cutoff values are investigated. Also discussed are the similarities in the performance of fit indices under different model misspecifications.

Among the five hundred samples drawn from the population bi-factor model, seventy-six of them were not converged and 101 of them had improper solutions. One cause of improper solutions is model misspecification, but it cannot be the case under properly specified models in a Monte Carlo study. Another plausible cause of improper solutions is small sample size (e.g., Mattson, Olsson & Rosén, 1966; Boomsma, 1983). The high percentage of non-convergence and improper solutions implied that the sample size of 145 might not be sufficient to estimate the bi-factor model and/or that the model specification may be improved. The rejection rates of the fit measures from the proper replications are presented in Table 7.3.

Suitable Cutoff Criteria

The earlier suggested cutoff values of 0.95 for TLI, 0.96 for CFI, 0.06 for RMSEA, 0.7 for SRMR, 0.9 for WRMR and 0.01 for the ML-based Chi-P all had acceptable type I error rates. However, they reacted differently to the misspecified model. The ML-based Chi-P, TLI and CFI at the suggested cutoff values had strong power (around 0.9) to reject the bi-factor model with misspecified loadings. The ML-based Chi-P, TLI and CFI in the bi-factor model at N = 145 appeared to have similar pattern and rejection rates as the CFA complex model with normal data at N = 100. On the other hand, the RMSEA, SRMR and WRMR values of this real-data example tended to be lower than the CFA complex model in the Miss model. As a result, the power of RMSEA, SRMR and WRMR at the suggested cutoff values was low. RMSEA at a cutoff of 0.06 and SRMR at a cutoff of 0.07 rejected only 15% and 1% of the misspecified model, respectively. At the suggested cutoff values, SRMR and RMSEA were unable to detect models with an important loading incorrectly excluded. The results seemed to explain why the RMSEA, SRMR and WRMR at the suggested cutoff values did not reject the bi-factor model in the real-data findings.

A high power value associated with the fit indices implied that a lack-of-fit decision is likely to reject the *real* misspecification of important parameters in the model. In order to have strong power (> 0.8) to detect the misspecification of loadings, a cutoff value of 0.045 for RMSEA, a cutoff value lower than 0.06 for SRMR and a cutoff value lower than 0.9 for WRMR were needed. Using the adjusted cutoff values for RMSEA, SRMR and WRMR from the small simulation study, these fit indices were consistent in indicating that further modeling or re-modeling should be considered.

As was demonstrated above, the suitable cutoff values suggested from the previous CFA study with N = 100 were applied and compared to the bi-factor model with N = 145, and it was found that the same cutoff criteria might not be applicable for RMSEA, SRMR and WRMR in *both* the CFA and bi-factor models. The CFA and bi-factor models investigated earlier differed not only in model complexity but also in sample size. In addition, the magnitude of the misspecified factor loadings was different (0.7 in the CFA model versus 0.6 in the bi-factor model). To evaluate whether the different performance of

RMSEA, SRMR and WRMR between the CFA and bi-factor models was due to sample size, the simulation results for the CFA model with N = 150 are presented in Table 7.4. It was found that, except for SRMR, the suitable cutoff criteria for Chi-P, TLI, CFI, RMSEA and WRMR were similar in the CFA models with N = 100 and N = 150. In the CFA models with N = 150, a cutoff value close to 0.055 or 0.06 for SRMR were needed to have power higher than 0.7. SRMR appeared to be sensitive to sample size. Comparing Table 7.3 with Table 7.4, fit indices (except for RMSEA) generally had lower type I errors in the CFA models than the bi-factor models. Furthermore, CFI and TLI tended to have higher type II error rates whereas RMSEA, SRMR and WRMR tended to have lower type II error rates in the CFA models. Again, it showed that the performance of RMSEA, SRMR and WRMR was different from that of CFI and TLI under misspecified models, and, regardless of sample size, RMSEA/SRMR/WRMR performed very differently between the 15-variable CFA and the 24-variable bi-factor models. The results also implied that the best cutoff criteria for fit indices depended on models.

Agreements and Similarities in Performance of Fit Indices

To understand the agreements on the reject or not-reject decisions for fit indices under certain cutoff values, the frequency and probability of consistent decisions for pairs of fit indices at certain cutoff values were investigated. The estimated probability of a consistent decision is calculated by $\hat{p} = \hat{p}_{11} + \hat{p}_{00}$, where $\hat{p}_{11}$ is the estimated probability when both fit indices reject the models and $\hat{p}_{00}$ is the estimated probability when both indices accept the models. Table 7.5 presents the estimated probability and frequency of consistent classifications for pairs of fit indices at certain cutoff values for the true models. It shows that a cutoff value of 0.95 for TLI, 0.96 for CFI, 0.06 for RMSEA, 0.7 for SRMR, 0.9 for WRMR and 0.01 for the ML-based Chi-P classify over 90% of the true models consistently.

Tables 7.6 and 7.7 present the estimated probability and frequency of consistent classifications for pairs of fit indices at certain cutoff values under the misspecified models. The only difference between Table 7.6 and Table 7.7 is the cutoff values for SRMR and RMSEA. Similar to the true models, in the Miss models Chi-P, TLI and CFI at the suggested cutoff values still made consistent reject or not-reject decisions (the pairwise agreement rates still ranged from 0.97 to 0.99). However, as opposed to Table 7.5, in the Miss models RMSEA at a cutoff of 0.06 and SRMR at 0.07 in Table 7.6 made far less consistent decisions with other fit indices. The pairwise agreement rates of RMSEA at a cutoff value of 0.06 ranged from 0.22 to 0.49, and those of SRMR at 0.07 only ranged from 0.07 to 0.36. Table 7.7 shows that reducing a cutoff value of RMSEA to 0.05 and that of SRMR to 0.06, the agreement rates of RMSEA/ SRMR with other fit indices increased tremendously. With a cutoff value of 0.05, RMSEA classified around 90% of the Miss models consistently with Chi-P at a cutoff of 0.01, TLI at 0.95 and CFI at 0.95, and around 80% of the Miss models consistently with WRMR at 0.9. SRMR at a cutoff value of 0.06 classified around 50% of the Miss models consistently with Chi-P at 0.01, TLI at 0.95 and CFI at 0.95, and about 80% of the Miss models consistently with WRMR at 0.9. Reducing cutoff values of RMSEA to 0.05 and SRMR to 0.06 not only increased their power but also increased their agreement rates with other fit indices.

To explore the similarities in the performance of fit indices in the real-data setting, the relationships between pairs of the ML-based Chi-P, CFI, TLI, RMSEA, SRMR and WRMR were investigated. The pairwise scatterplots for fit measures under the true and Miss models were shown in Figure 7.1 and Figure 7.2, respectively. Under the true models Chi-P, TLI, CFI and RMSEA seemed to correlate highly (the absolute value of correlations ranged from 0.85 to 0.96), and SRMR and WRMR correlated highly (correlation value was 0.96). Under the Miss models the majority of Chi-P values were close to 0 with a few outliers, thus it did not correlate highly with the other fit indices as in the true model. Similar to true models, RMSEA appeared to correlate more highly with TLI and CFI than with WRMR and SRMR under the Miss models. The rejection rates and agreement rates between pairs of the fit indices under various cutoff values can also be eye-approximated in these scatterplots. One of the characteristics that set CFI and TLI apart from RMSEA, SRMR and WRMR was that the TLI and CFI values were clearly lower and they had larger spread under the Miss models. Thus, a cutoff value of 0.95 for TLI and CFI was sufficient to reject many misspecified models.

Generally speaking, except for Chi-P, the general trend and relationships between pairs of the fit indices were similar for the bi-factor true and Miss models. The results were very similar to Figure 4.2 and Figure 4.3 of the CFA models. RMSEA correlated highly with TLI and CFI, whereas WRMR correlated highly with SRMR. It seemed that the relationships between pairs of fit indices were not affected much by model complexity. In terms of cutoff criteria, the suggested cutoff values for TLI, CFI and Chi-P in the earlier CFA study were applicable to the real-data example. In addition, TLI at a cutoff of 0.95, CFI at 0.95 and Chi-P at 0.01 had very high agreement rates for both true and misspecified models. On the contrary, the suggested cutoff values of RMSEA, SRMR and WRMR in the CFA study needed to be adjusted in order to have high power and better agreements with other fit indices. An inspection of Figure 4.3 and Figure 7.2 revealed that the RMSEA, SRMR and WRMR values were much lower in the bi-factor Miss models than in the CFA Miss1 models. For instance, RMSEA ranged from 0.05 to 0.09 in the CFA Miss1 models but only from 0.02 to 0.07 in the bi-factor Miss models. Therefore, a cutoff value of 0.06 for RMSEA failed to reject many bi-factor Miss models. RMSEA, SRMR and WRMR had better agreements with other fit indices under true models than under misspecified models.

In LVM analyses, researchers should always choose models based more on psychological theory and substantive expertise. The choice of cutoffs can vary in different substantive fields and can be affected by the standards set by prior work, as Bollen (1989, p274) mentioned. For example, if previous research usually reported a TLI value of .90 and RMSEA of 0.6, then a TLI value of 0.94 and RMSEA of 0.5 can indicate an important improvement of the modeling. In the illustrative example, the estimates of the fit indices were very close to the cutoff values, and substantive expertise can play an important role in the decision. If moderate power (0.5 to 0.7) is acceptable, TLI $\geq$ 0.93, CFI $\geq$ 0.94, RMSEA $\leq$ 0.05 and WRMR $\leq$ 0.9 can also be suitable criteria to indicate good models.

Figure 7.1. Scatterplot of the relationship between pairs of fit measures in the bi-factor true models

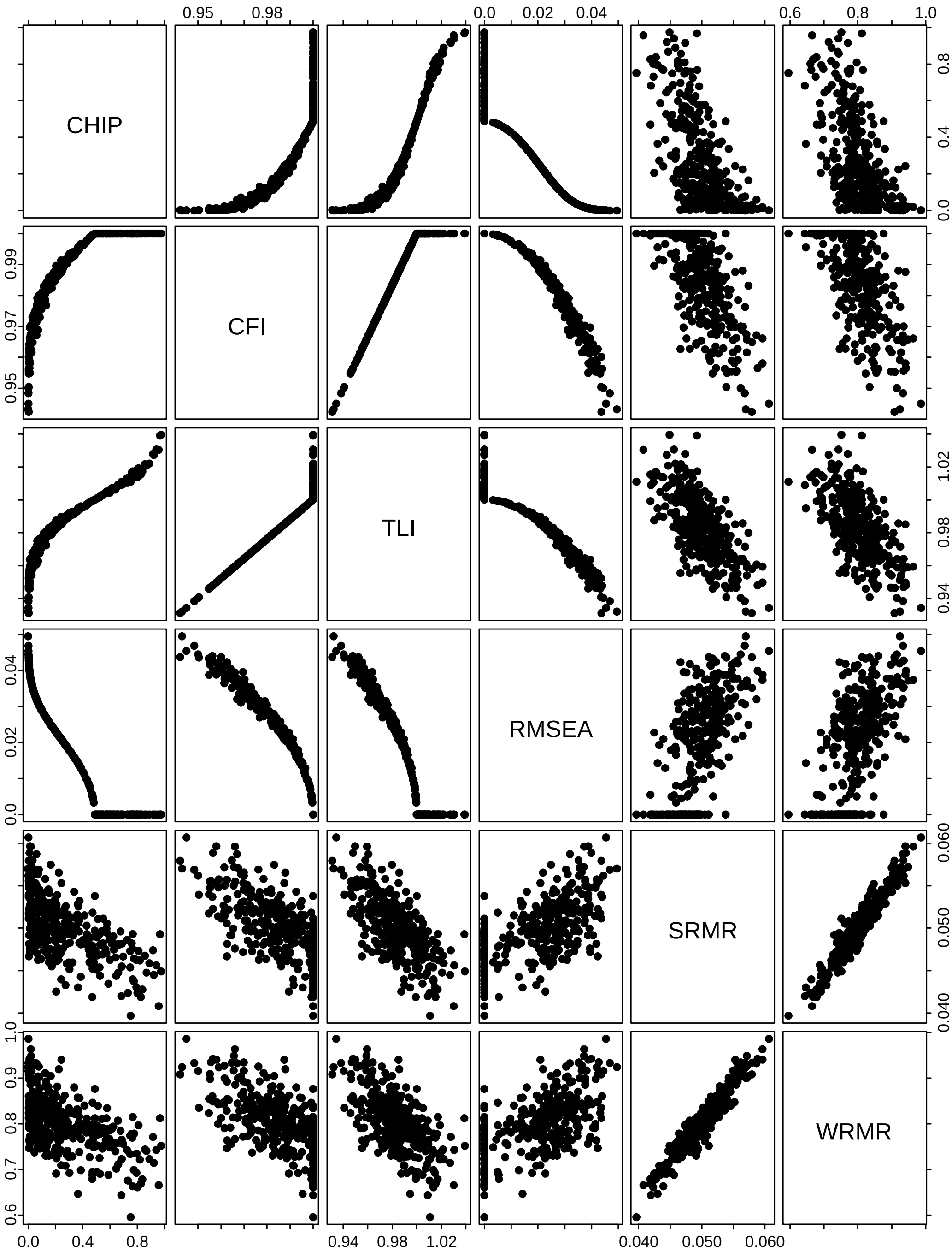


Figure 7.2. Scatterplot of the relationship between pairs of fit measures in the bi-factor Miss1 models

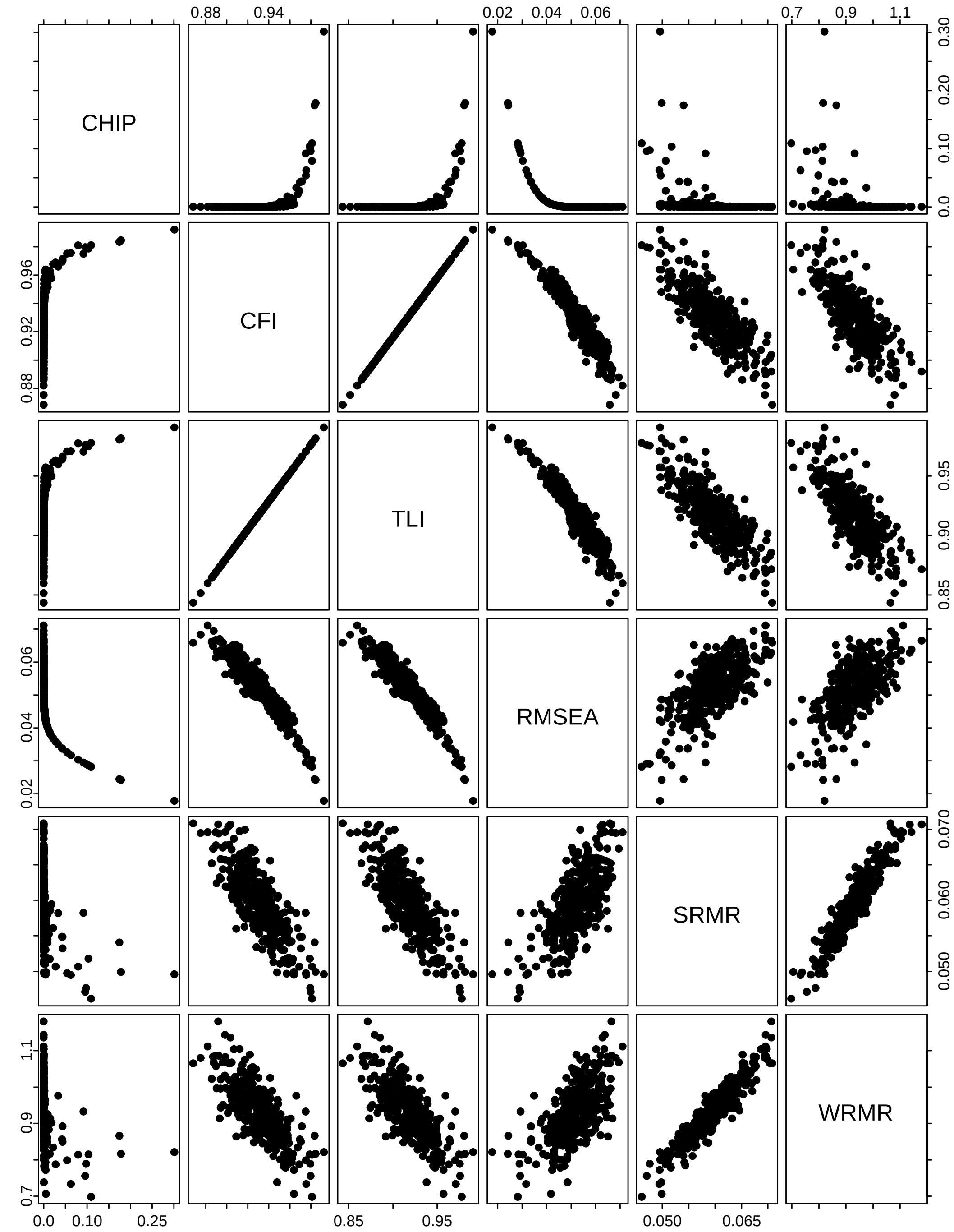


Table 7.1. Basic statistics for the 24 mental ability tests

Variable	Mean	Variance	Skewness	Kurtosis
1. Visual perception	29.58	47.80	-0.12	-0.05
2. Cubes	24.80	19.76	0.24	0.87
3. Paper form board	14.30	7.96	0.13	0.50
4. Flags	15.97	69.17	0.62	-0.45
5. General information	44.85	135.91	0.19	0.56
6. Paragraph comprehension	9.95	11.39	0.40	0.25
7. Sentence Completion	18.85	21.62	-0.55	0.22
8. Word Classification	28.20	28.70	-0.40	1.52
9. Word meaning	17.28	63.16	0.73	0.23
10. Addition	90.18	565.59	0.16	-0.36
11. Code	68.59	277.67	0.30	0.68
12. Counting dots	109.77	440.79	0.70	2.28
13. Straight-curved capitals	191.78	1371.62	0.20	0.52
14. Word recognition	176.28	117.26	-0.66	1.66
15. Number recognition	89.37	56.89	0.21	0.35
16. Figure recognition	103.40	46.26	0.10	-0.18
17. Object-number	7.21	20.87	1.18	2.82
18. Number-figure	9.44	20.29	0.38	-0.54
19. Figure-word	15.24	12.93	-0.77	0.17
20. Deduction	30.34	391.26	0.30	0.16
21. Numerical puzzles	14.46	23.35	-0.32	-0.08
22. Problem reasoning	27.73	96.52	0.14	-0.36
23. Series completion	18.75	87.62	0.04	-0.68
24. Arithmetic problems	25.83	22.21	-0.56	-0.01

Table 7.2. Bi-factor solution for the 24 mental ability tests (standardized loadings)

Test	Factors						Residual variances
	General	Spatial	Verbal	Speed	Recognition	Memory	
1	0.620[a]	0.206[a]					0.573
2	0.408*	0.322					0.730
3	0.428*	0.223					0.767
4	0.563*	0.553					0.378
5	0.587*		0.544[a]				0.360
6	0.570*		0.596*				0.320
7	0.563*		0.628*				0.288
8	0.608*		0.361*				0.500
9	0.594*		0.605*				0.282
10	0.378*			0.586[a]			0.514
11	0.504*			0.398*			0.588
12	0.445*			0.615*			0.424
13	0.579*			0.423*			0.486
14	0.385*				0.593[a]		0.500
15	0.363*				0.422*		0.690
16	0.527*				0.295*		0.636
17	0.411*				0.360*	0.484	0.467
18	0.521*					0.459[a]	0.518
19	0.442*					0.278	0.728
20	0.643*						0.586
21	0.639*						0.592
22	0.648*						0.579
23	0.722*						0.479
24	0.613*						0.624

Note. [a] Fixed parameter; * Significant at 0.05 level.

Table 7.3. Rejection rates of fit measures for the bi-factor model (N = 145)

Cutoff	ML Chi-P	SB Chi-P	Cutoff	TLI	CFI	Cutoff	RMSEA	SRMR	Cutoff	WRMR
0.01			0.9			0.045			0.7	
True	**6.8**	10.8		0.0	0.0		0.9	92.9		93.8
Miss	**94.0**	94.5		23.0	6.0		81.6	100.0		99.7
0.03			0.93			0.05			0.8	
True	14.9	18.6		0.0	0.0		0.0	47.1		49.8
Miss	95.7	96.3		69.8	45.7		62.1	96.8		95.7
0.04			0.94			0.055			0.9	
True	18.9	21.7		1.2	0.0		0.0	9.3		**7.7**
Miss	96.0	97.7		83.0	68.7		33.9	83.3		**65.2**
0.05			0.95			0.06			0.95	
True	20.7	24.5		**5.0**	1.2		**0.0**	0.3		0.6
Miss	96.8	98.0		**92.2**	83.0		**15.8**	44.0		38.5
0.06			0.96			0.07			1.0	
True	22.6	28.8		12.4	**5.9**		0.0	**0.0**		0.0
Miss	97.1	98.3		95.4	**93.4**		0.3	**1.1**		16.1

Note. There are 323 proper replications for the True model and 348 proper replications for the Miss model.

Table 7.4. Rejection rates of fit measures for the CFA model (N = 150)

Cutoff	ML Chi-P	SB Chi-P	Cutoff	TLI	CFI	Cutoff	RMSEA	SRMR	Cutoff	WRMR
0.01			0.9			0.045			0.7	
True	**2.8**	4.2		0.0	0.0		7.8	36.4		44.4
Miss	**95.6**	96.4		4.4	0.4		98.2	100.0		99.8
0.03			0.93			0.05			0.8	
True	7.4	8.4		0.0	0.0		3.8	12.0		11.2
Miss	98.2	98.8		36.2	12.6		96.2	99.8		99.2
0.04			0.94			0.055			0.9	
True	8.6	9.8		0.0	0.0		1.6	2.6		**1.8**
Miss	99.0	98.8		56.8	28.4		90.4	98.0		**92.2**
0.05			0.95			0.06			0.95	
True	9.4	10.8		**0.2**	0.0		**0.8**	0.6		0.6
Miss	99.2	99.4		**73.4**	52.0		**81.8**	90.2		85.8
0.06			0.96			0.07			1.0	
True	10.8	12.4		1.4	**0.2**		0.0	**0.0**		0.4
Miss	99.4	99.4		89.8	**74.6**		53.2	**58.6**		76.2

Table 7.5. Probability and frequency of agreements between pairs of fit measures at certain cutoff values under true models (323 replications)

Fit index	Chi-P (.01[1])	TLI (.95)	CFI (.95)	RMSEA (.06)	SRMR (.07)	WRMR (.9)
Chi-P	__	313	316	301	301	294
TLI	0.97	__	320	307	307	300
CFI	0.98	0.99	__	304	304	299
RMSEA	0.93	0.95	0.94	__	323	298
SRMR	0.93	0.95	0.94	1.00	__	298
WRMR	0.91	0.93	0.93	0.92	0.92	__

[1] Values in parentheses are the adopted cutoff value for fit indices. The underscored values in the low triangle are the estimated probability values. The frequency of consistent decisions is in the up-right triangle.

Table 7.6. Probability and frequency of agreements between pairs of fit measures at certain cutoff values under misspecified models (RMSEA at 0.06 and SRMR at 0.07; 348 replications)

Fit index	Chi-P (.01[1])	TLI (.95)	CFI (.95)	RMSEA (.06)	SRMR (.07)	WRMR (.9)
Chi-P	__	338	342	76	25	240
TLI	0.97	__	344	82	31	248
CFI	0.98	0.99	__	78	27	244
RMSEA	0.22	0.24	0.22	__	297	170
SRMR	0.07	0.09	0.08	0.85	__	125
WRMR	0.69	0.71	0.70	0.49	0.36	__

[1] Values in parentheses are the adopted cutoff value for fit indices. The underscored values in the low triangle are the estimated probability values. The frequency of consistent decisions is in the up right triangle.

Table 7.7. Probability and frequency of agreements between pairs of fit measures at certain cutoff values under misspecified models (RMSEA at 0.05 and SRMR at 0.06)

	Chi-P (.01)	TLI (.95)	CFI (.95)	RMSEA (.05)	SRMR (.06)	WRMR (.9)
Chi-P	__	338	342	305	174	240
TLI	0.97	__	344	311	180	248
CFI	0.98	0.99	__	307	176	244
RMSEA	0.88	0.89	0.88	__	213	259
SRMR	0.50	0.52	0.51	0.61	__	274
WRMR	0.69	0.71	0.70	0.74	0.79	__

[1] Values in parentheses are the adopted cutoff value for fit indices. The underscored values in the low triangle are the estimated probability values. The frequency of consistent decisions is in the up right triangle.

CHAPTER 8

SUMMARY AND DISCUSSION

The aims of this study were to first evaluate and compare the performance of various model fit measures under different model and data conditions, and, secondly, to examine the adequacy of the cutoff criteria for these model fit measures. This study applied the method demonstrated in Hu and Bentler (1999) to evaluate the adequacy of the cutoff criteria for TLI, CFI, RMSEA, SRMR and WRMR. The performance of Chi-P under various cutoff values was also evaluated and compared to fit indices. Three major simulation studies based on CFA, MIMIC and latent growth curve models were conducted to evaluate the performance and to obtain suitable cutoff criteria for model fit measures. An application of the simulation results was illustrated in Chapter 7 using the Holzinger and Swineford data.

8.1 Summary of the Results

In the CFA simulation study, fit indices were evaluated under conditions such as model specification, type of outcome variables (normal, non-normal continuous and binary outcomes), type of model misspecification (true, misspecified and trivially misspecified models) and various sample sizes. It was found that with normal outcomes, acceptable cutoff values for the maximum likelihood χ^2 (ML)-based TLI, CFI, RMSEA and SRMR were close to 0.95, 0.96, 0.05 and 0.07 respectively. A cutoff value close to 0.05 for RMSEA tended to have high type I errors and was less preferable at N = 100. This study investigated a new model fit index WRMR, and WRMR ≤ 1.0 can be an indication of good models with normal and moderately non-normal continuous outcomes across all four sample sizes. It had acceptable type I error rates across samples and distributions and its power ranged from 0.42 to 1. With severely non-normal continuous outcomes and N ≥ 250, the SB-based CFI at a cutoff value close to 0.95 and SRMR at close to 0.07 had acceptable type I and type II error rates. SRMR was found to work poorly with binary outcomes. CFI ≥ 0.96 and RMSEA ≤ 0.05 can be indications of good models with binary outcomes at N ≥ 250. WRMR with a cutoff value close to 1.0 had acceptable type I error rates with binary outcomes at N ≥ 250, but one might want to consider a cutoff value of 0.95 for WRMR to have higher power at N ≥ 250.

The SB χ^2 exhibited much lower type I error rates than the ML χ^2 under non-normality, but it still tended to overreject properly specified models at smaller sample sizes. When N ≤ 250 under severe non-normality, the use of the SB-based CFI, TLI and RMSEA was recommended to maintain type I error control. However, it appeared that the SB-based TLI, CFI and RMSEA at their suggested cutoff values still overrejected properly specified models at N = 100. When sample size was equal to or larger than 500, the ML-based TLI and CFI at the suggested values seemed to still be applicable with non-normal data. Note that the ML-based RMSEA at a cutoff value of 0.05 still had inflated type I errors at N = 500 under severe non-normality, and it required sample sizes of 1000 or larger to be robust to non-normality.

The ideal cutoff values of fit measures might vary under different conditions. In some cases, a few cutoff values for the same fit index may be considered, such as cutoff values of

0.95 or 0.96 for CFI and TLI, values of 0.07 or 0.08 for SRMR, values of 0.05 or 0.06 for RMSEA, and values close to 0.95 or 1.0 for WRMR.

In Chapter 5 the performance and adequacy of cutoff values for Chi-P, TLI, CFI, RMSEA, SRMR and WRMR in MIMIC models were evaluated under conditions such as different sample sizes, various types of model misspecification and outcome variables (binary, normal, moderately and severely non-normal outcomes). The model with misspecified factor loadings was considered and compared to the previous CFA results. It was found that rejection rates of these fit measures in the CFA and MIMIC models was similar, and the suitable cutoff criteria obtained earlier from the CFA study seemed to be applicable to the MIMIC models also.

CFI ≥ 0.96, TLI ≥ 0.95, SRMR ≤ 0.07, RMSEA ≤ 0.06, WRMR ≤ 1.0 (or 0.95) and Chi-P ≥ 0.01 were acceptable indications of good MIMIC models across all four sample sizes with normal outcome. Under moderate non-normality, a cutoff value of 0.95 for the ML-based TLI, 0.96 for the ML-based CFI, and 0.95 for WRMR were still applicable at N ≥ 250. A cutoff value of 0.05 for the ML-based RMSEA can be applicable at N ≥ 500, and the use of the SB-based RMSEA was preferable at N = 250. The cutoff values suggested above all had power higher than 0.7 with satisfactory type I error rates at N ≥ 250. Relatively speaking, at N ≥ 250, WRMR a cutoff value of 0.95 was preferable to 1.0 in both CFA and MIMIC models with binary, normal and moderately non-normal data in terms of power. On the other hand, a cutoff value of 1.0 for WRMR had the advantage of a better type I error control at N = 100. With binary outcomes, the power of WRMR at a cutoff value of 0.95, CFI at 0.96 and Chi-P at 0.05 tended to increase and their type I errors tended to decrease with increasing sample size. Therefore, they performed well at larger sample sizes (e.g. $N \geq 250$). At $N = 100$, cutoff values of 0.95 for WRMR and 0.05 for Chi-P had acceptable type I error rates with power ranging from 0.3 to 0.7.

Note that controlling type I errors under or around the nominal level of 5% was considered as a more important criterion to suggest suitable cutoff values for fit indices in this dissertation. We did not want to incorrectly reject true-population models. On the other hand, type II errors were allowed to be liberal. Rejection rates over 80% under misspecified models were deemed satisfactory, while rates of 50% to 70% were deemed moderate and still considered acceptable.

Suitable cutoff criteria for fit measures under the CFA and MIMIC models with liberal type II error rates can be summarized as below:

Normal Outcomes

- Acceptable cutoff values for the ML-based TLI, CFI, RMSEA, SRMR and WRMR were close to 0.95, 0.96, 0.05, 0.07 and 1.0, respectively.
- A cutoff value close to 0.05 for RMSEA tended to have high type I errors and is less preferable at $N = 100$.

Non-Normal Outcomes

- Under moderate non-normality, only WRMR with a cutoff value of 1.0 had moderate to strong power to detect misspecified models with acceptable type I errors across all four sample sizes.
- Under severe non-normality, none of the model fit indices at the suggested cutoff values had acceptable type I errors at $N = 100$; at $N = 250$, the cutoff value of 0.95

for the SB-based CFI and a cutoff value of 0.07 for SRMR had acceptable type I and type II errors.

Binary Outcomes

- SRMR was found to work poorly.
- CFI ≥ 0.96, RMSEA ≤ 0.05 and WRMR ≤ 1.0 can be indications of good models with binary outcomes at N ≥ 250.

Generally speaking, the power of TLI, CFI and RMSEA to detect models with misspecified loadings was higher than their power to detect models with misspecified covariances. The power of SRMR was larger to detect models with misspecified covariances. The power of WRMR and Chi-P were similar in both types of misspecified models.

In terms of the latent curve models with five and eight time points, the results show that Chi-P at a cutoff value of 0.05 performed well across all sample sizes. Except for overrejecting the five-point LGM true models at N = 100, CFI at a cutoff of 0.95 was suitable for latent growth curve models across samples. A cutoff value of 0.95 for CFI had the advantage of lower type I errors, whereas a cutoff value of 0.96 for CFI provided higher power to reject misspecified models. The choice of adequate cutoff values for the residual-based fit indices WRMR and SRMR seemed to vary with time points in the latent growth curve model.

The example illustrated in Chapter 7 represented the model complexity and sample size usually seen in practice. Practitioners often find the similar type of inconsistency among fit indices. Applying the suitable cutoff values obtained from the CFA complex models produced inconsistent conclusions among the fit indices. Chi-P at a cutoff value of 0.01, TLI at 0.95 and CFI at 0.96 suggested a lack of fit for the theoretical bi-factor model, whereas RMSEA at 0.06, SRMR at 0.07 and WRMR at 0.9 suggested goodness of fit. A small simulation study was conducted to investigate the adequacy of these suggested cutoff values in the bi-factor model with N = 145. It was found that cutoff values of close to 0.01 for Chi-P, close to 0.95 for TLI, and close to 0.96 for CFI suggested by the previous CFA study still had satisfactory power to detect the bi-factor misspecified models. In contrast, the performance of RMSEA, SRMR and WRMR differed between the bi-factor models and the three-factor CFA models. Comparing to the CFA models, the RMSEA, SRMR and WRMR values tended to be lower and thus result in lower rejection rates in the bi-factor misspecified model.

The results show that suitable cutoff criteria for some fit indices are strongly dependent on models. For example, the best cutoff values for RMSEA and SRMR in the CFA and bi-factor models at N = 150 are very different. This suggests that to find suitable cutoff criteria and to understand how fit indices perform under a certain type of models with a certain sample size, substantive researchers might want to conduct their own simulation studies like the small Monte Carlo study illustrated here.

8.2 Overall Discussion and Recommendation

The simulation results have shown that sample size, model complexity and the presence of non-normality affect the estimates of the fit measures and the decisions of cut points, thus a cut point to satisfy all conditions is hard to find. In LVM analyses, the choice of cutoffs can also vary in different substantive fields and can be affected by the standards set by prior work.

None of the fit measures with a specific cutoff value has minimum type I and type II error rates simultaneously across all sample sizes, multivariate distributions, and model specifications. The cost of type I and type II errors might vary in different research fields, and different cutoff criteria might be chosen as a result. For example, the choice of 0.95 or 0.96 for CFI depends on the relative cost of the type I and type II errors. If the type I errors are more costly, one might want to use a cutoff value of CFI close to 0.95. Moreover, with a different preferred range of power, different cutoff values might be chosen. For example, if a power value of 0.4 or 0.5 is acceptable, cutoff values of 0.95 for WRMR and 0.05 for Chi-P could be deemed suitable at N = 100 as well. In spite of the complexity in studies of the cutoff criteria for fit indices, some general conclusions and recommendations may still be drawn from the four empirical studies of the CFA, MIMIC, latent growth curve and bi-factor models.

1. The ML χ^2 with a probability level of 0.05 results in inflated type I error rates with almost all outcomes at N = 100 in the CFA and MIMIC models. The SB χ^2 still tends to overreject properly specified models at smaller sample sizes. Thus, the use of the ML and SB χ^2 (with a probability level of 0.05) are not recommended at small sample sizes and/or with non-normal data, except for latent growth models.

2. CFI at a cutoff value of 0.96 seems to be applicable to binary, normal and moderately non-normal continuous outcomes at $N \geq 250$. When $N \leq 250$ under severe non-normality, the use of the SB-based CFI is recommended to maintain type I error

control. The high agreements and similarities between CFI and TLI imply that it might be sufficient to just report one of them. Relatively speaking, CFI at the suggested cutoff value has better type I and type II error rates than TLI at a cutoff value of 0.95 and RMSEA at 0.05. RMSEA at a cutoff value of 0.05 tends to overreject properly specified models at N = 100 and, in some cases, loses power at N ≥ 500.

3. SRMR is sensitive to sample sizes, and a specific cutoff value for SRMR across samples is difficult to find. The use of SRMR with binary outcomes is not recommended.

4. WRMR at a cutoff value of 0.95 or 1.0 has acceptable type I error rates with small to moderate type II error rates in the CFA and MIMIC models but rejects too many properly specified latent growth curve models with more time points. It might also tend to reject too many models with trivial misspecification of factor covariances.

There are a few limitations in this study that should be acknowledged. Because an empirical approach is applied to explore the characteristics of these fit indices with unknown distributions, strong theoretical explanations for some phenomena observed in this study are not provided. Several issues may be investigated further. For example, it is interesting to investigate why WRMR performs less well in latent growth models with more time points, and why RMSEA and SRMR perform differently in CFA models with different number of variables and factors. Furthermore, because the Monte Carlo method is used and not all possible models are studied, caution should be used in generalizing the results and conclusions too far beyond the models investigated. More research with a wider class of models and conditions are needed.

There are a few more potentially interesting topics. Some other adjusted or rescaled test statistics, such as the Yuan-Bentler (YB) test statistic, may be incorporated into the calculation of fit indices. The SB χ^2 and SB-based TLI, CFI and RMSEA at the suggested rules of thumb tend to have inflated type I errors at small sample sizes (e.g., N = 100). The YB statistic was found to perform well in models with small sample sizes such as 60 to 120 (Bentler and Yuan, 1999), thus the YB-based TLI, CFI and RMSEA might be useful in model selection with small sample sizes. Furthermore, Hu and Bentler (1999) recommended the use of SRMR in conjunction with fit indices such as TLI, CFI and RMSEA. Since WRMR performs relatively better than SRMR especially with binary outcomes, the use of WRMR in conjunction with CFI, TLI and RMSEA may be preferable and worthy of further investigation.

References

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

Anderson, R. D. (1996). An evaluation of the Satorra-Bentler distributional misspecification correction applied to the McDonald fit index. *Structural Equation Modeling, 3*, 203-277.

Austin, J. T., & Calderón, R. F. (1996). Theoretical and technical contributions to structural equation modeling: an updated annotated bibliography. *Structural Equation Modeling, 3,* 105-175.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structure. *Psychological Bulletin, 88,* 588-606.

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

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

Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models.* Newbury Park, CA: Sage.

Boomsma, A. (1983). *On the robustness of LISREL (maximum Likelihood Estimation)* *against small sample size and non-normality.* Doctoral dissertation, University of Groningen.

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

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

Browne, 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.

Carlson, M., & Mulaik, S. (1993). Trait ratings from descriptions of behavior as mediated by components of meaning. *Multivariate Behavioral Research, 28,* 111-159.

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

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the “problem” of sample size: a clarification. *Psychological bulletin, 109,* 512-519.

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

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

Equation Modeling, 6, 56-83.

Harman, H. H. (1976). *Modern Factor Analysis*. Third Edition. Chicago: The University of Chicago Press

Holzinger, K. J., & Swineford, F. A. (1939). *A study in factor analysis: The stability of a bi-factor solution.* Supplementary Educational Monographs, No. 48, Chicago, Ill.: The University of Chicago.

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

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure analysis: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424-453.

Hu, L., & Bentler, P. M. (1999). Cutoff criterion for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409-426.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70,* 631-639.

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models.* Cambridge, MA: Abt Books.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by the method of maximum likelihood.* Mooresville, IN: Scientific

Software.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide.* Chicago, IL: Scientific Software International.

La Du, T. J., & Tanaka, J. S. (1989). Influence of sample size, estimation method, and model specification on goodness of fit assessments in structural equation models. *Journal of Applied Psychology, 74,* 625-635.

Marsh, H. W. (1995). Δ2 and χ^2I2 fit indices for structural equation models: A brief note for clarification. *Structural Equation Modeling, 2,* 246-254.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: the effect of sample size. *Psychological Bulletin, 103,* 391-410.

Marsh, H. W., & Hau, K. –H. (1996). Assessing goodness of fit: Is parsimony always desirable? *The Journal of Experimental Education, 64,* 364-390.

Mattson, A., Olsson, U., & Rosén, M. (1966). The maximum likelihood method in factor analysis with special consideration to the problem of improper solutions. *Research Report.* University of Uppsala, Department of Statistics.

Mulaik, S., James, L., Van Alstine, J., Bonnett, N., Lind, S., & Stilwell, C. (1989). Evaluation of goodness of fit indices for structural equation models. *Psychological Bulletin, 105,* 430-445.

Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of* *change, (pp. 291-322)*.

Muthén, B., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods. 2(4),* 371-402.

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171-189.

Muthén, B., & Khoo, S.T. (1998). Longitudinal studies of achievement growth using latent variable modeling. Learning and Individual Differences, Special issue: *Latent growth curve analysis, 10,* 73-101.

Muthén, L. K., & Muthén, B. O. (1998-2001). Mplus User's Guide. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Tam, T., Muthén, L., Stolzenberg, R. M., & Hollis, M. (1993). Latent variable modeling in the LISCOMP framework: Measurement of attitudes toward career choice. In D. Krebs, & P. Schmidt (Eds.), *New Directions in Attitude Measurement.* Festschrift for Karl Schuessler (pp. 277-290). Berlin: Walter de Gruyter.

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

SAS Institute Inc. (1988). SAS/ETS user's guide. Cary, NC: Author.

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

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for statistics in covariance structure analysis. Los Angeles: UCLA Statistics Series #2.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15,* 201-93.

Steiger, H. H., & Lind, J. M. (1980, June). Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 11-39). Newbury Park, CA: Sage.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10.

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

Williams, L. J., & Holahan, P. J. (1994). Parsimony-based model fit indices for multiple-indicator models: Do they work? *Structural Equation Modeling, 1.* 161-189.

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.