University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2016

The Influence of Model Components and Misspecification Type on the Performance of the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA) in Structural Equation Modeling

Mahler, Claudia

Mahler, C. (2016). The Influence of Model Components and Misspecification Type on the Performance of the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA) in Structural Equation Modeling (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. doi:10.11575/PRISM/28214

http://hdl.handle.net/11023/3240

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

The Influence of Model Components and Misspecification Type on the Performance of the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA) in Structural Equation Modeling

by

Claudia Mahler

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

AUGUST, 2016

© Claudia Mahler 2016

Abstract

This thesis examined the performance of two popular fit indices used in structural equation modeling: the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). Of interest were the indices' sensitivities to different sources of misspecification as well as sensitivities to model components that may affect index behavior over and above misspecification. Index performances were evaluated in confirmatory factor analysis models involving one of three sources of misspecification: omitted error covariances, omitted cross-loadings, or an incorrectly modeled latent structure. In addition, model components—including model complexity, loading size, factor correlation size, and model balance—were manipulated to determine their effects on index behavior. It was revealed that CFI is more sensitive to latent misspecifications, while RMSEA is more sensitive to misspecifications due to omitted error covariances. Both indices are affected to some extent by model components, particularly model complexity and loading size.

Acknowledgements

I would like to offer my thanks to my graduate supervisors, Dr. Gemai Chen and Dr. Hua Shen, for their help and guidance throughout my time in this program. Without their assistance, recommendations, and time, this thesis would not have been possible. Gratitude also goes to Dr. Elena Braverman, Dr. Bingrui Cindy Sun, and Dr. Ying Yan for their time in serving on my examination committee.

Additional thanks goes to my wonderful husband, Nate, who has been incredibly supportive throughout this process, to my amazing mom, who has been encouraging and motivating me throughout many, many years of school, and to my dad, who has offered more support than he knows. Without their help, I would have never made it this far.

I would also like to acknowledge many of my friends, including Charles, Gowri, Jialin, Saifa, Salha, and Tayler, for their support, motivation, and encouragement, with respect to both research and classwork. Finally, I would like to thank Mrs. Yanmei Fei for all her hard work, recommendations, and kindness.

Dedication

I dedicate this thesis to my husband, Nate, who has been and continues to be a constant source of love and kindness in my life.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
RMSEA	Root Mean Square Error of Approximation
SEM	Structural Equation Modeling
p	The number of indicator variables in a model
k	The number of factors in a model
λ	A single factor loading value
$\boldsymbol{\lambda}$	A $p \times 1$ vector of factor loadings
$\boldsymbol{\Lambda}$	A $p \times k$ matrix of factor loadings
ψ	A single residual value
$\boldsymbol{\psi}$	A $p \times p$ covariance matrix of the residual
φ	A single factor correlation value
$\boldsymbol{\Phi}$	A a $k \times k$ matrix of factor correlations
$\boldsymbol{\theta}$	A $q \times 1$ vector of population parameters
$\hat{\boldsymbol{\theta}}$	A $q \times 1$ vector of population parameter estimates
$\mathbf{S}$	A sample covariance matrix
$\boldsymbol{\Sigma}^*$	A population covariance matrix
$\boldsymbol{\Sigma}(\boldsymbol{\theta})$	The structure of the population covariance matrix implied by the researcher's proposed model
$\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$	The estimated population covariance matrix based on the researcher's proposed model

Chapter One: Introduction

1.1 Structural Equation Modeling

Structural equation modeling (SEM) is a statistical modeling technique that lets researchers construct and test causal connections amongst variables. This is done by allowing the relationships between the variables to be expressed as functions of the parameters of a hypothesized model. In SEM, these relationships are expressed as the covariances (or correlations) between the variables.

SEM can be used to model relationships between latent variables (factors), between observed variables (indicators), or between latent and observed variables. These relationships can be expressed as a series of structural equations in which model parameters are estimated. The restrictions imposed by these parameters are then applied to a given sample to determine whether or not the model suggesting these parameters actually holds in the population from which the sample was taken.

Because SEM is a general method that comprises multiple different modeling techniques, the present research focuses solely on confirmatory factor analysis (CFA), one of the most commonly used types of SEM. A structural equation model typically comprises two main components: a structural model and a measurement model. In CFA, the structural model describes the relationships amongst the k factors while the measurement model represents the set of p indicators of the k factors (McDonald and Ho, 2002; Perry et al., 2015). A complete model is a combination of the structural model and the measurement model and serves as an expression of the causal connections amongst the factors as well as the causal connections between the factors and their relevant indicators. These connections are functions of the model parameters.

The model itself is a theory-based representation of how the variables within it relate to each other in reality. For example, suppose a developmental psychologist suspects that there is a relationship between mathematical intelligence and verbal intelligence (both latent factors) and that these can be measured by a child's performance in various school-related processes (reading comprehension, math test performance, etc.). The psychologist's model of the relationship amongst the two latent factors and four selected indictor variables is represented in Figure 1.1.

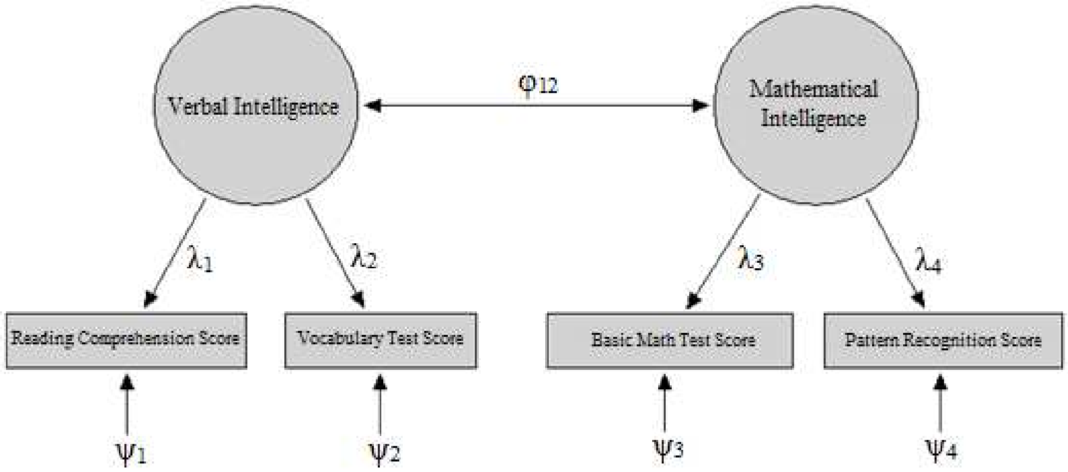

Figure 1.1: A structural equation model relating two latent factors, Verbal Intelligence and Mathematical Intelligence, to each other and to four indicators (Reading Comprehension Score, Vocabulary Test Score, Basic Math Test Score, and Pattern Recognition Score).

The process of SEM involves estimating model parameters. Often, these model parameters include loading size, error variance, and factor correlation. Each indicator variable can be said to "load onto" or measure one or more of the latent factors. In the model presented in Figure 1.1, Reading Comprehension Score and Vocabulary Test Score both load onto the Verbal Intelligence factor, while Basic Math Test Score and Pattern Recognition Score load onto the Mathematical Intelligence factor. The values λ_1, λ_2, λ_3, and λ_4 represent *loading sizes*, or the strengths of the relationships between indicators and factors. Loading sizes can be different for different pairs of indicator variables and factors.

Each indicator variable also has an *error variance* associated with it. In SEM, error variance is the portion of an indicator's variance that does not covary with the latent factor, such as measurement error. In Figure 1.1, the error variances associated with each of the four indicator variables are denoted by ψ_1, ψ_2, ψ_3, and ψ_4.

In some cases, it may be suspected that latent factors are related to some degree. Researchers can include one or more *factor correlations* in their model to represent these relationships. In Figure 1.1, the psychologist suspects there is a relationship between Verbal Intelligence and Mathematical Intelligence, and this relationship is expressed in the factor correlation, φ_{12}, between them.

Loading sizes, error variances, and factor correlations are all model parameters that represent the relationships amongst factors and indicators. Unless a researcher chooses to fix any of these parameters at certain values, the values of these parameters will be estimated in the SEM process. This estimation process is discussed in further detail in section 1.4.1.

SEM offers the ability to test these theory-driven models against empirical data, which is one of the main reasons the technique has grown in popularity amongst scientists in many different fields, including psychology, environmental science, and education (Fan et al., 1999). Another reason behind its popularity is its allowance for the modeling of latent variables. Many research problems, particularly those in the social sciences, require a way of relating measurable variables, or indicators, to related latent factors.

One of the most important components of the SEM process is assessing model fit: how well do the causal inferences contained in a hypothesized model (like the one proposed in Figure

1.1) reflect the actual relationships amongst the variables? Traditionally, the chi-square test statistic has been the sole criterion by which model fit is judged. However, notable problems arise with the test statistic's performance in both large and small samples, under different estimation methods, and in cases where the underlying distributional assumptions are violated.

In response to these problems, a multitude of "goodness-of-fit" indices have been developed to aid researchers in accurately assessing model fit in situations where the chi-square may prove inaccurate. Due to their increased inclusion in popular SEM software, it is not uncommon for researchers to report one or two fit index values alongside a chi-square test statistic when evaluating the fit of a model.

However, use of these indices does not come without its own set of problems. One particular problem is the sheer number of indices available. Programs such as SAS, EQS, and LISREL are capable of printing upwards of a dozen indices in addition to the chi-square test statistic (Fan et al., 1999; Gerbing and Anderson, 1992; Hu and Bentler, 1998). Without proper knowledge, it may be difficult for a researcher to know which indices to report (Bollen and Long, 1992).

An additional problem stems from the fact that not all indices have been developed under the same theoretical rationale. For example, some indices have been developed to penalize overly-complex models (models with a large number of indicators), while other indices have no such penalty. Because of differing theoretical rationales, indices may perform differently under certain model structures or misspecification types. These differences in performance could, in some cases, lead to conflicting conclusions about the appropriateness of a model's fit (Fan et al., 1999).

When using a fit index to determine goodness-of-fit, it is common practice to employ the use of a “cutoff value.” Not unlike a significance level for a hypothesis test, a cutoff value is used as a threshold of model acceptance/rejection. Different indices have different widely-accepted cutoff values; however, there have been multiple demonstrations in the literature (e.g., Marsh et al., 1988; Beauducel and Wittmann, 2005) indicating that these cutoff values may not be generalizable across all modeling situations. Other studies (e.g., Chau and Hocevar, 1995; Fan and Sivo, 2007; Hu and Bentler, 1999; Kenny and McCoach, 2003; Marsh et al., 1988) have revealed that model components such as sample size, model complexity, and estimation procedure all affect how index values change, over and above the effect of misspecification.

Thus, while attempts have been made to develop alternatives to the chi-square, there still exist problems with the use of fit indices in evaluating model fit, particularly when trying to do so across different model and misspecification types.

1.2 Goals of the Present Research

The aim of the present research is to investigate the performance of two popular fit indices, the comparative fit index (CFI) and the root mean squared error of approximation (RMSEA) in various model and misspecification conditions. Specifically, I will examine the effects of different sources of misspecification (e.g., omitted error covariances, omitted cross-loadings) in conjunction with different model components (e.g., model complexity, loading size, latent factor correlation) on index values.

To eliminate any variability in index performance that may be due to sample size, the population-based indices are studied. The purpose of this is to contrast the results of the current study with previous studies (most of which manipulate sample size), to determine if any general conclusions about fit index behavior, regardless of sample size, can be made. In conducting this research, the goal is to address the following four questions:

1. To what extent is fit index value affected by the source of the misspecification?
2. To what extent is the relationship between the degree of model misspecification and fit index value moderated by different model components?
3. Does the current research support the use of universal cutoff values across different model and misspecification types?
4. Can guidelines for the use of different indices under different models be developed?

In addressing the first question, the hope is to provide a clearer understanding of whether either index (RMSEA or CFI) is more sensitive to certain types of misspecification than others. RMSEA and CFI values will be evaluated in models with one of three different sources of misspecification. The first source of misspecification arises from one or more omitted error covariances. That is, a hypothesized model omits one or more error covariances that are present in the true (population) model. Using the psychology example presented above and the model in Figure 1.1, if the two indicators loading onto Verbal Intelligence were measured using the same instrument, it may be expected that their error variances are correlated. Figure 1.1, then, would be misspecified in the sense that it omits this error covariance. Because covarying errors are common in many disciplines that make use of SEM, it is important that fit indices show worse fit for hypothesized models that omit any error covariances (such as the covariance of ψ_1 and ψ_2) that are present in the population.

The second source of misspecification arises from one or more omitted cross-loadings. If latent factors are strongly correlated in the population, it is likely that an indicator that loads onto one of the factors also loads onto the other factor. Again, using the psychology example, suppose that Reading Comprehension Score actually loads onto both Verbal Intelligence and Mathematical Intelligence (perhaps a child needs good reading comprehension skills to fully understand word problems in math). A model would omit this cross-loading if it either omitted the loading between Verbal Intelligence and Reading Comprehension Score or Mathematical Intelligence and Reading Comprehension Score. If a hypothesized model omits one (or more) cross-loadings that are present in the population, this misspecification should ideally be picked up by fit indices.

The third and final source of misspecification arises from a misspecified latent structure. This occurs when a hypothesized model includes either more factors or fewer factors than there are present in the population. Suppose in the psychology example that there was only one factor representing intelligence in the population. If this were the case, Figure 1.1 would have a missepecified latent structure, as it includes more factors than there are present in the population. Misspecifications of this sort may be considered more serious than those of the previous two, since misspecification is not due to an omitted pathway (an error covariance or cross-loading) but is instead due to the failure to include the correct number of factors present in the population.

The second question involves looking at how different model components may affect index values. Model components, in the context of this study, are any aspects of the modeling procedure that may affect index value over and above the size of the actual misspecification. The model components studied here include loading size, factor correlation size (in models with two or more latent variables), model complexity, and model balance (in models with two or more latent variables). Model complexity will be judged by the total number of indicator variables, the total number of factors, and the ratio of indicators to factors in a given model. The goal is to determine which indices are sensitive to which types of model components, and to what degree.

With respect to the third question, the aim is not to suggest specific cutoff criteria; rather, the commonly applied cutoff values for CFI and RMSEA will be evaluated with the goal to determine whether the indices behave consistently enough to warrant the use of universal cutoff values across varying modeling situations.

Finally, the goal of the fourth question is to attempt to put forward a set of guidelines regarding the use of the two fit indices. Index behavior is examined here in a way that allows both the strengths and weaknesses of the indices to be revealed under different misspecification types and under the influence of different nuisance parameters.

1.3 Thesis Structure

In the next section of this chapter, I begin by providing a brief discussion of the steps of a typical SEM process. The focus is on the estimation procedure utilized when fitting a hypothesized model to sample data. Both the chi-square test statistic and the two fit indices of interest (CFI and RMSEA) are introduced. The remaining portion of the chapter is dedicated to a review of the previous literature concerning the chi-square and fit indices. I highlight the theoretical issues surrounding the use of these measures of goodness-of-fit and review studies that have illustrated how CFI and RMSEA are affected by various model components.

Chapter 2 describes the methodology used in the present study to examine CFI and RMSEA performance in models with one of three different misspecification sources. The rationale behind the use of simulations is discussed, as are the details of how the simulations are constructed and carried out.

Chapter 3 focuses on the performances of CFI and RMSEA in several specific misspecification scenarios. First, the performances of CFI and RMSEA are examined in CFA models in which misspecification is due to one or several omitted error covariances. Second, the performances of CFI and RMSEA are examined in two-factor CFA models in which misspecification is due to one or several omitted cross-loadings. Finally, I focus on the performances of CFI and RMSEA in CFA models with a misspecified latent structure.

In addition to examining the effect of misspecification type and size on index performance, the scenarios also include the manipulation of certain model components in order to determine their effects on index performance as well.

Chapter 4 presents a brief summary of the results of Chapter 3, including a comparison of the results of this study to results found in previous literature. I discuss the benefits of researchers using CFI and RMSEA in conjunction and offer several recommendations and guidelines for the use of these two indices.

In Chapter 5, practical applications of the recommendations in Chapter 4 is described. Using data from two previously published SEM-related studies, I discuss alternative models for these datasets and explain how CFI and RMSEA together can be used to possibly identify the sources of any model misspecifications. Finally, Chapter 6 consists of a brief review of this study’s findings, the limitations of the study design and simulations, and suggestions for relevant future research.

1.4 The Process of SEM

There are five general steps that characterize most applications of SEM (Bollen and Long, 1992). The first step, model specification, involves specifying a particular model to represent the relationships amongst particular variables of interest. This is often based on previous research, theory, and information related to the variables. Figure 1.1, reprinted here, represents a model that specifies certain relationships amongst the two factors and four indicator variables of interest.

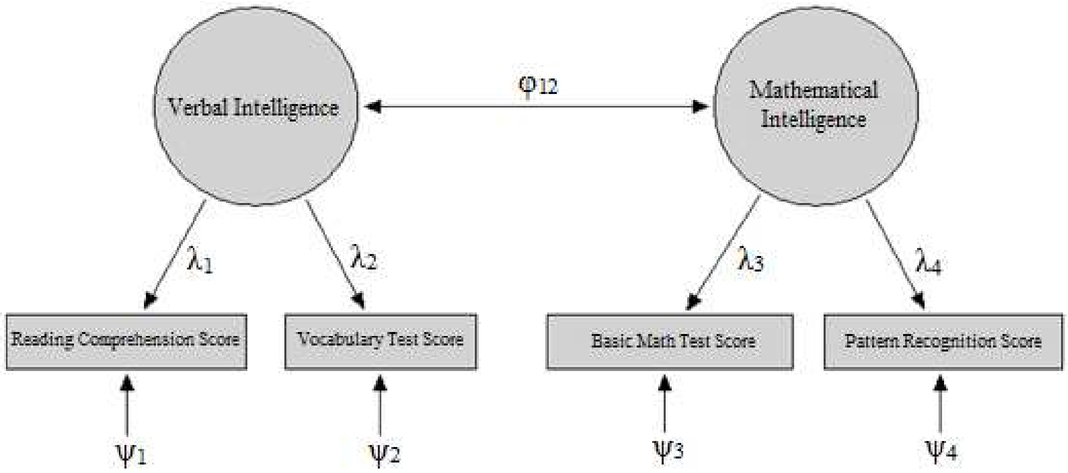

Figure 1.1: A structural equation model relating two latent factors, Verbal Intelligence and Mathematical Intelligence, to each other and to four indicators (Reading Comprehension Score, Vocabulary Test Score, Basic Math Test Score, and Pattern Recognition Score).

Once a model has been specified, the second step of the SEM process, known as *model identification*, can be performed. In SEM, the relationships amongst variables can be written as a set of structural equations based on the covariance matrix of the variables in the model. The structural equations implied by the hypothesized model are used to construct a model-implied covariance matrix (Hu and Bentler, 1999). In the example presented above, the model-implied covariance matrix would be based on the relationships amongst the variables as shown in Figure 1.1. For a 2-factor model like the one in question, the covariance structure is given by $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \boldsymbol{\psi}$, where $\mathbf{\Lambda}$ is a $p \times k$ matrix of factor loadings, $\mathbf{\Phi}$ is a $k \times k$ matrix of factor correlations, and $\boldsymbol{\psi}$ is the $p \times p$ covariance matrix of the residuals, where k represents the number of latent factors in the model and p represents the number of indicators in the model. The covariance structure implied by Figure 1.1 is given as

$$\mathbf{\Sigma} = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ 0 & \lambda_3 \\ 0 & \lambda_4 \end{bmatrix} \begin{bmatrix} \varphi_{12} & 0 \\ 0 & \varphi_{12} \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & \lambda_4 \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 \\ 0 & 0 & 0 & \psi_4 \end{bmatrix}.$$

The third step of the SEM process is *estimation*. Though SEM seeks to make conclusions about the relationships amongst variables in the population, in most cases researchers must rely on a sample from the population of interest. Thus, the goal of the estimation step is to find parameter values such that the discrepancies between the sample covariance matrix and the model-implied covariance matrix are minimized. The parameter values in SEM are often comprised of the "pathways" in a hypothesized model. These pathways are usually represented by arrows in a diagram of the model. In Figure 1.1, the parameters to be estimated include the loadings of the indicators onto their respective factors (the λ terms), the correlation between the two factors (φ_{12}), and the error variances (the ψ terms).

After a model has been estimated, the fourth step of the SEM processes involves determining how well the hypothesized model *fits* the data. In other words, how well is the theoretical model underlying the hypothesized model supported by the data? While a general measure of fit can be obtained by looking at the residuals between the sample covariance matrix and the model-implied covariance matrix (Bollen and Long, 1992), more formal measures of fit, such as the chi-square test and goodness-of-fit indices, are often used in this step to assess how well a proposed model is supported by the data.

The final step, which may be repeated multiple times, is *model respecification*. Depending on how well the original hypothesized model fits the data, restructuring the model may be necessary. If a model is restructured, the SEM procedure can begin again at step one and continue until a model with an acceptable fit is obtained (it should be emphasized, however, that any restructuring of a model should be done under the guidance of relevant theory and past research).

The focus of this thesis is on how well the fit indices CFI and RMSEA accurately reflect the degree of misspecification for a given model. In order to gain a better understanding of how these indices work within the context of the SEM process, I will discuss steps 3 and 4 (*estimation* and *determining fit*) in more detail in the following sections.

1.4.1 Estimation

Sample data based on N = n + 1 subjects and p indicator variables are summarized in a $p \times p$ sample covariance matrix $\mathbf{S}$. It is hypothesized that a population covariance matrix $\mathbf{\Sigma^*}$ exists and is generated by q true but unknown parameters. The $q \times 1$ vector of these unknown parameters, $\boldsymbol{\theta}$, corresponds to the particular structure of $\mathbf{\Sigma^*}$. If the sample size corresponding to $\mathbf{S}$ were to increase to infinity, $\mathbf{S}$ would converge to $\mathbf{\Sigma^*}$ and its structure would be known (Bentler and Bonett, 1980).

In order to test the SEM null hypothesis $\mathbf{\Sigma^*} = \mathbf{\Sigma(\theta)}$, which states that the population covariance matrix $\mathbf{\Sigma^*}$ has the structure implied by the researcher's hypothesized model, estimates of the unknown population parameters $\boldsymbol{\theta}$ as well as the matrix $\mathbf{\Sigma^*}$ must be calculated under the hypothesized model. Once a vector $\widehat{\boldsymbol{\theta}}$ of estimated model parameters has been obtained, an estimated covariance matrix $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ can be constructed as a function of the estimated model parameters (Bentler and Bonett, 1980).

While the ideal case would involve a direct test of $\mathbf{\Sigma^*} = \mathbf{\Sigma(\theta)}$, in reality, the true population covariance matrix $\mathbf{\Sigma^*}$ is never actually known. Instead, researchers must compare the hypothesized model's covariance matrix to the sample matrix **S**.

The primary goal of the estimation step is to arrive at parameter estimates $\widehat{\boldsymbol{\theta}}$ such that $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$, the hypothesized model's covariance structure based on these estimates, is as similar to the structure of **S** as possible (Moshagen, 2012). Obtaining these parameter estimates is achieved by the minimization of some discrepancy function $F(\boldsymbol{\theta})$ which, if given a set of parameters, provides an assessment of the difference between the model-implied covariance matrix $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ and the sample covariance matrix **S**, based on the residuals between these two matrices (Folnes et al., 2012).

According to Anderson and Gerbing (1984) and Moshagen (2012), the predominately used estimation procedure (and the default estimation method in nearly all major SEM packages) is the maximum likelihood (ML) procedure. The traditional maximum likelihood fit function $F_{ML}(\boldsymbol{\theta})$, hereby written as $F(\boldsymbol{\theta})$, is based on the log likelihood ratio. In the population, this value is given by

$$F(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \ln|\boldsymbol{\Sigma}^*| + \text{tr}[\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-\mathbf{1}}] - p, \tag{1.1}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ represents the structure of the covariance matrix implied by the hypothesized model, $\boldsymbol{\Sigma}^*$ represents the population covariance matrix, and p is the number of observed variables (Bollen and Long, 1992). When $\boldsymbol{\Sigma}^*$ is not known, the sample covariance matrix **S** can replace $\boldsymbol{\Sigma}^*$ and (1.1) can be expressed as

$$F(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \ln|\mathbf{S}| + \text{tr}[\mathbf{S}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-\mathbf{1}}] - p. \tag{1.2}$$

Minimizing (1.2) yields $F(\widehat{\boldsymbol{\theta}})$, hereby written as $\widehat{F}$, and the corresponding $q \times 1$ vector of parameter estimates $\widehat{\boldsymbol{\theta}}$. $\widehat{F}$ attains the value of 0 if and only if $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{S}$; otherwise, $\widehat{F}$ is positive and increases as the discrepancy between $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and **S** increases (MacCallum et al., 1996).

1.4.2 Assessing Model Fit

Once the estimation step has been performed, the result is a vector $\widehat{\boldsymbol{\theta}}$ of parameter estimates that minimize the fit function $F(\boldsymbol{\theta})$. The corresponding model-implied covariance matrix $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ can be assessed to determine how well it matches the structure of the sample covariance **S**, which in turn is an indication of how well the hypothesized covariance structure is supported by the data.

Discussed in the following sections are two broad methods of assessing model fit: the use of the chi-square test statistic to assess what is known in the literature as "exact fit," and the use of fit indices to assess "close fit."

1.4.2.1 The Chi-Square Test Statistic

Traditionally, the assessment of model fit in SEM has been accomplished via the dichotomous decision of hypothesis testing. That is, a model is either accepted or rejected based on how well it appears to fit the sample data. When assessing model fit, the null hypothesis is the claim that **Σ*** = **Σ(θ)**, or that the hypothesized model's covariance structure, **Σ(θ)**, exactly matches the population covariance matrix of the observed variables, **Σ***. Thus, evaluating fit in this way is a test of "exact fit" (Bollen and Long, 1992; Hsu et al., 2015).

Testing exact fit involves the use of the chi-square test statistic and its associated p-value and is based on the minimized fit function $\hat{\mathrm{F}}$ found in the *estimation* step (discussed in the previous section). The chi-square test statistic is found by multiplying $\hat{\mathrm{F}}$ by (N – 1) to yield:

$$\mathrm{T} = (\mathrm{N} - 1)\hat{\mathrm{F}}. \tag{1.3}$$

This T statistic is used to test the null hypothesis **Σ* = Σ(θ)** (Hu and Bentler, 1999). Under the assumptions that the model is correct and that the data are multivariate normal, T has an asymptotic χ^2_d distribution with degrees of freedom $d = \frac{p(p+1)}{2} - q$, where *p* is the number of indicator variables in the model and *q* is the number of unique parameters to be estimated (Hu and Bentler, 1998; MacCallum et al., 1996).

Larger values of $\hat{\mathrm{F}}$ correspond to greater discrepancies (greater residuals) between the model-implied covariance matrix and the sample covariance matrix (Bentler, 1990). Thus, if the residuals are larger than what would be expected due to sampling fluctuation, the T statistic will exceed a critical value of the chi-square distribution at a pre-specified α-level (Hu and Bentler, 1998). The null hypothesis $\mathbf{\Sigma^* = \Sigma(\theta)}$ will then be rejected, indicating that the hypothesized model does not exactly describe the underlying population covariance structure from which the sample was drawn. On the other hand, if $\hat{F}$ is small, the resulting T will be small as well, leading to acceptance of the model if the T value is smaller than the critical value.

Different estimation methods (generalized least squares, maximum likelihood, etc.) arrive at the minimized fit function value $\hat{F}$ in different ways. Therefore, chi-square test statistics resulting from different estimation methods have slightly different interpretations. For maximum likelihood, the estimation method used in this thesis, the resulting chi-square is a likelihood ratio test statistic (Bentler and Bonett, 1980). The likelihood of observing the sample data under the hypothesized model is compared to the likelihood of observing the data under a saturated model—a model containing as many parameter estimates as degrees of freedom. Small values of this ratio imply that the data are as likely to occur under both the saturated model and the hypothesized model. Large values, on the other hand, suggest that the structure imposed by the hypothesized model, when compared to the saturated model, is not overly restrictive to the point that it fails to adequately fit the patterns found in the data (McDonald and Marsh, 1990).

1.4.2.2 Fit Indices

As structural equation modeling gained popularity in the 1990s and early 2000s, several criticisms arose with respect to the chi-square test as a measure of model fit. One major criticism is that the test is highly sensitive to sample size (this will be discussed in a later section). Other criticisms relate to the nature of the chi-square as a test of exact fit. It is unrealistic, many argue, to assume that any given covariance structure will *exactly* match that of the population covariance structure; thus, it would be more appropriate to assess the *degree* of lack of fit rather than exact fit (Hsu et al., 2015; Marsh et al., 1988; McDonald and Marsh, 1990).

Another concern related to the chi-square's use is the lack of information it provides over and above whether the model-implied covariance matrix is equal to the sample covariance matrix. In other words, apart from revealing that a model either fits perfectly or does not fit perfectly, the chi-square cannot provide any information about the magnitude or cause of the misfit, details that would be beneficial to researchers trying to refine their model (Hsu et al., 2015).

In response to these and other problems surrounding the use of the chi-square, a number of fit indices have been developed to either replace the use of the chi-square (in cases where sample size is an issue) or to be used alongside it as a supplemental way of assessing fit.

While the chi-square test leads to a binary fit/no fit decision, most fit indices have been developed to represent goodness-of-fit along a continuum rather than to prompt the researcher to accept or reject a model based on a prespecified critical value. Thus, fit indices are to be interpreted as gauges of "close fit" rather than of exact fit.

Like the chi-square test statistic, fit indices make use of the residuals between the sample covariance matrix and the model-implied covariance matrix, $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$, to assess fit. In fact, many popular fit indices can be expressed as functions of the chi-square test statistic. Table 1.1 provides a brief summary of the two indices studied in this thesis— comparative fit index and the root mean square error of approximation—and includes their equations both at the sample level and at the population level.

Table 1.1: Names, sample definitions, and population definitions of CFI and RMSEA.

Index Name	Sample Definition	Population Definition
Comparative Fit Index (CFI)	$\frac{(\chi^2_I - df_I) - (\chi^2 - df)}{(\chi^2_I - df_I)}$	$1 - \frac{\hat{F}}{\hat{F_I}}$
Root Mean Square Error of Approximation (RMSEA)	$\sqrt{\frac{\frac{\chi^2}{df} - 1}{n - 1}}$	$\sqrt{\frac{\hat{F}}{df}}$

Note. Where n is the sample size, χ^2_I and χ^2 stand for the chi-square values for the independent (baseline) model and the hypothesized model, respectively; df_I and df are the degrees of freedom for the independent model and the hypothesized model, respectively; $\hat{F_I}$ and $\hat{F}$ stand for the minimized fit function for the independent and hypothesized models, respectively.

To derive the population equations, $\hat{F} \cdot N$ (the minimized fit function multiplied by $N = n + 1$) was used to replace the chi-square values found in the sample definitions of the indices. Then n was then allowed to tend to infinity. The value of $\hat{F}$ can be utilized in the population as a measure of model misspecification. Models that are not exactly true (i.e., that do not exactly match the structure of the population covariance matrix) will yield $\hat{F}$ values that do not equal zero, with increasingly large $\hat{F}$ values corresponding to models with increasingly poor fit (Bentler, 1990; Steiger et al., 1985).

The current study focuses on the performance of fit indices in the population, and thus the population definitions in Table 1.1 are used to calculate CFI and RMSEA values in the simulations presented later. Properties of the two indices of interest, as well as rationale behind their use, will be described in more detail in the following sections.

1.5 Literature Review

I now turn to a review of the existing literature for a discussion of the issues surrounding model fit in SEM. I begin with a brief summary of the literature involving the issues affecting the chi-square test statistic, then focus on the literature surrounding the use of the fit indices. I present a more in-depth introduction to the two indices of interest in this thesis, then discuss concerns regarding the use of these indices as measures of model fit. While the current research is concerned with index behavior at the population level, the majority of existing research has been carried out at the sample level. Therefore, existing findings on the effects of sample size on index behavior are reviewed briefly as well.

1.5.1 Chi-Square

1.5.1.1 Sample Size

As is true with any statistical test, the power of the chi-square test is a direct function of the sample size. Thus, as the sample size tends toward infinity, even the smallest difference between a proposed model and the true model will be reflected in the chi-square value. In large sample, the test statistic may be so sensitive to the size of the sample that conclusions based on the chi-square value might not be trusted.

This sensitivity is well-documented in SEM literature. As early as the 1970's, researchers such as Joreskog (1970) and Bentler and Bonett (1980) have noted that unless a model fits perfectly, an increase in sample size will inflate the chi-square value. Thus, in situations involving large enough samples, a model with a trivial misspecification may be rejected solely due to the chi-square's sensitivity to sample size rather than due to any actual severe misspecification (Hsu et al., 2015; Perry et al., 2015).

The chi-square's sensitivity to sample size limits its practical use in SEM. This is because it is often expected that a hypothesized model will not fit the data *exactly* (Bearden et al., 1982; Gerbing and Anderson 1992). However, a model can fit well enough to warrant its use as an appropriate representation of the population covariance structure. Because the chi-square can reject even a trivially misspecified model if the sample size is large enough, models that might be *practically* useful may be discarded on the basis that the chi-square test shows that they are a poor fit to the data. While one solution to this problem would be to allow the use of "appropriate" p-values as cutoff values (e.g., set a lower p-value as the cutoff value so it is more difficult to reject a particular model), most studies involving SEM do not employ this technique and instead rely on typical cutoff values instead.

It is also worth noting that in many hypothesis testing settings (such as regression or ANOVA), the goal is to find enough evidence to reject the null. In the context of SEM, however, the goal is to retain the null, as the null suggests that the researcher's model provides a good description of the relationship amongst the variables of interest in the population. Thus, a large sample size, which is desirable in nearly every application of statistics, may work against the goals of a researcher using SEM (Bentler and Bonett, 1980).

Problems with the chi-square can also arise when the sample size is small. The T statistic defined in (1.3) follows an asymptotic chi-square distribution which may not be well approximated in smaller samples (Bentler and Yuan, 1999). Bentler and Yuan (1999), Hooper et al. (2008), and Hu et al. (1992), among others, have demonstrated that in small samples, the chi-square lacks power and tends to over-reject the null. This behavior could lead to incorrect conclusions about the adequacy of a model.

1.5.1.2 Other Issues

The assumption of multivariate normality must also be taken into consideration when using the chi-square. Violations of this assumption can lead to severely inflated T statistics, resulting in the rejection of models that are properly specified (Hooper et al., 2008; Moshagen, 2012). The chi-square test statistic is also sensitive to the size of the proposed model. Moshagen, (2012) has demonstrated that as the size of the proposed model increases, T will inflate and increase rejection rates for models that may be properly specified.

1.5.2 Fit Indices

Before reviewing the literature discussing the behavior of fit indices in SEM, I present a more in-depth introduction to the two fit indices shown in Table 1.1. Despite there being over two dozen indices readily available for use by researchers using SEM, I chose to focus my research on CFI and RMSEA for several specific reasons.

First, while the chi-square test statistic is almost universally reported as a measure of model fit, certain fit indices are more commonly used than others. According to McDonald and Ho (2002), the most commonly reported fit index in the psychological literature is CFI, followed by RMSEA. Rigdon (1996) claims that CFI and RMSEA are two of the most commonly reported fit indices across multiple fields. Due to the popularity of these two indices, it was decided that the research here should focus on their performances across many different modeling scenarios rather than on the performances of several different indices in only a few scenarios.

Second, the popularity of CFI and RMSEA has led to a great deal of research focusing on their behaviors in different modeling situations. This large amount of research allows for a greater number of comparisons that can be made between how the indices performed in other studies versus how they perform in the scenarios presented here.

Finally, despite the fact that both indices are based on the chi-square value, CFI and RMSEA evaluate model fit in different ways. CFI is from a family of indices known as *relative* (or *incremental*) *fit indices*. These indices evaluate the hypothesized model in relation to a more restricted "baseline" model, often a model in which all the observed variables are uncorrelated (Themessl-Huber, 2014). Relative fit indices assess how well a hypothesized model fits in comparison to the more restricted baseline model, thus requiring more than just the discrepancies between a hypothesized model's covariance matrix and the sample covariance matrix in their calculations (Rigdon, 1996).

RMSEA is from a family of indices known as *absolute fit indices*. Indices of this family evaluate how well a researcher's hypothesized model reproduces the sample data and is a function of the discrepancies between that hypothesized model's covariance matrix and the sample covariance matrix (Hooper et al., 2008; Rigdon, 1996). More will be said on the differences between absolute and relative fit indices in the following two sections.

1.5.2.1 Population Definitions

1.5.2.1.1 Comparative Fit Index (CFI)

In the population, CFI can be defined as:

$$\mathrm{CFI} = 1 - \frac{\hat{\mathrm{F}}}{\hat{\mathrm{F}}_{\mathrm{I}}}, \tag{1.4}$$

where $\hat{F}$ is the minimized fit function of the hypothesized model and $\widehat{F_I}$ is the minimized fit function of a baseline model. Commonly, this baseline model is defined as the model in which all variables are modeled as mutually uncorrelated, and thus is also known as the “independent model” (though the baseline can be any other model selected by the researcher) (Bentler and Bonett, 1980; Hooper et al., 2008).

Defined as a relative fit index, CFI uses the baseline model as an additional reference point for the evaluation of the covariance structure of the hypothesized model. It is expected that the value of $\widehat{F_I}$ is large, indicating poor fit. In addition, it is hoped that the value of $\hat{F}$, corresponding to the proposed model, is small, indicating good fit. If the hypothesized model fits perfectly, the ratio of $\hat{F}$ to $\widehat{F_I}$ will be zero (because $\hat{F}$ will equal zero) and CFI will equal one. The smaller the CFI value is, the more it suggests that the hypothesized model is no more an adequate model for the data than the baseline model (Bentler and Bonett, 1980).

CFI is bound by 0 and 1, with larger values indicating better fit. The generally accepted cutoff value for the CFI in the literature (e.g., Beauducel and Wittmann, 2005; Hooper et al., 2008; Hu and Bentler, 1999) is .95, with values greater than .95 indicating good model fit.

1.5.2.1.2 Root Mean Square Error of Approximation (RMSEA)

First developed by Steiger and Lind (1980), cited in Steiger (1990), then expanded upon and popularized by Browne and Cudeck (1992), RMSEA is an absolute fit index based on the minimized fit function $\hat{F}$. A population expression of RMSEA can be written as

$$\mathrm{RMSEA} = \sqrt{\frac{\hat{F}}{\mathrm{df}}}, \tag{1.5}$$

where *df* is the degrees of freedom for the hypothesized model. The index is a measure of the average discrepancy between the sample covariance matrix and the model-implied covariance matrix per degrees of freedom (Hsu et al., 2015). Because the degrees of freedom are a function of the number of estimated parameters in a model, RMSEA is said to have a parsimony correction, as it penalizes models that are "too large" (i.e., too many parameters to estimate) (Hooper et al., 2008; Themessl-Huber, 2014). Often, this means that models with a large number of indicators will be penalized more than models with a smaller number of indicators, as more indicators implies more loading parameters that need to be estimated.

When $\boldsymbol{\Sigma^*} = \boldsymbol{\Sigma(\theta)}$, $\hat{\mathrm{F}} = 0$; thus, an RMSEA value of zero is achieved when the model fit is perfect (RMSEA values cannot be negative, as $\hat{\mathrm{F}}$ cannot be negative). Because of this, RMSEA is often thought of as a measure of "badness of fit," as its size increases as model fit grows worse, though values rarely exceed one (Hu and Bentler, 1999). The most generally accepted cutoff value in the literature is .06, as proposed initially by Hu and Bentler (1999). That is, if RMSEA of a given model is less than .06, the model is said to be a "close fitting" model and accurately describes the relationships amongst the variables in the model.

1.5.2.1.3 Omitted Indices

This study omits several commonly used fit indices in favor of being able to focus more on the behaviors of two of the most popular indices reported in the literature. In this section, I briefly make note of other indices that are often used to assess the goodness of fit of models in SEM. For sample and population definitions of these and other commonly used fit indices, please see Appendix A.

In addition to the RMSEA, other absolute fit indices include the goodness of fit index (GFI) proposed by Jöreskog and Sörbom (1981) and the adjusted goodness of fit index (AGFI), a modified version of the GFI and developed by Jöreskog and Sörbom (1981) when studies revealed that the GFI shows an improvement in model fit as additional parameters were included in the model (e.g., Cudeck and Browne, 1983; MacCallum and Hong, 1997). These indices calculate the proportion of variance accounted for in the sample covariance matrix by the covariance matrix derived from the hypothesized model.

Another commonly used absolute fit index is the standardized root mean square residual (SRMR), also developed by Jöreskog and Sörbom (1981). The SRMR assesses the average size of the residuals between the sample and hypothesized covariance matrices and is similar to the RMSEA in the sense that it is a measure of "badness of fit"—the index value increases as the model fit gets worse (Chen et al., 2008).

Other relative fit indices frequently reported in the literature include the normed fit index (NFI) proposed by Bentler and Bonett (1980), the Bentler-Bonnet Index (BBI) (Bentler and Bonett, 1980), and Bollen's Incremental Fit Index (IFI) (Bollen, 1989). Though the sample definitions of these indices differ slightly, at the population level they are all defined with the same equation as the CFI. Each of these indices is influenced not only by the goodness of fit of the proposed model but also by the choice and fit of the baseline model as well.

Several other indices, such as the non-normed fit index (NNFI) proposed by Bentler and Bonett (1980) and the Tucker-Lewis Index (TLI) proposed by Tucker and Lewis (1973), are slightly modified relative fit indices and include the ratio of the degrees of freedom of the proposed and baseline models as well. Indices such as these have also gained popularity in the literature.

1.5.3 Concerns Regarding the Use of Fit Indices

I now turn to a summary of the concerns raised with respect to using fit indices as measures of goodness of fit in SEM. In general, fit indices have been developed to compensate for some of the shortcomings of the chi-square test when assessing the fit of a model. Specifically, many have been designed to be less sensitive to both sample size and model size and to be unaffected by estimation method.

However, studies addressing index performance across different modeling situations have revealed that the indices are not without their own set of problems. Some problems stem from how indices are selected and interpreted by researchers. Other problems arise due to how the indices perform given different sample sizes, estimation methods, model sizes and types, and different misspecification types.

I will begin by summarizing the theoretical and methodological concerns pertaining to the use of fit indices, focusing mainly on the issue of the cutoff values commonly used to determine whether a model's fit is acceptable. I then turn to the application-related concerns that have been addressed in the literature. I summarize relevant studies that examine the behavior of fit indices with respect to different aspects of the modeling procedure and different model components, including estimation method, sample size, model size, parameter size, and misspecification type. I conclude with a summary of the pertinent results and discuss how the present research aims to address some of the remaining concerns.

1.5.3.1 Theoretical and Methodological Concerns

1.5.3.1.1 Selecting an Index

In SEM, it is considered standard practice to report the chi-square test statistic and its associated p-value for a hypothesized model. However, there exist no generally accepted guidelines for which index (or indices) to report alongside the chi-square for those who wish to do so. As previously mentioned, popular SEM programs are capable of printing more than one dozen fit indices by default. Because of this, the issue can arise where printed indices conflict—some show the model fit is adequate, while others do not. These conflicting results often arise because different indices have been designed to assess different criteria of fit (Gerbing and Anderson, 1992). Thus, researchers can arrive at different conclusions about a particular model depending on which indices they choose to examine (Hu and Bentler, 1998).

There have been attempts by researchers in SEM to define guidelines that outline which indices to report in which situations; however, there does not appear to be any agreement over which indices should always be reported alongside the chi-square. Crowley and Fan (1997) suggest that it is necessary to report a variety of indices due to the fact that different indices reflect different aspects of model fit. Several index pairings are recommended by Hu and Bentler (1999), including pairing the SRMR with the TLI or the RMSEA with the CFI. Others (e.g., Hooper et al., 2008; Kline, 2005) suggest different sets of indices to report. The lack of more concrete and universal guidelines with respect to index reporting may make it difficult for researchers to determine which indices to focus on when assessing model fit.

A final concern relates to the use of fit indices by those researchers who are not highly familiar with SEM but still utilize the procedure. A researcher who needs to use SEM in one or two studies, for example, may not be as familiar with the different fit indices as a researcher who uses SEM more frequently. If presented with a dozen different fit indices for a fitted model, a less knowledgeable researcher may be unware of the difference in how model fit is calculated for each index and therefore might simply choose to report the indices they may have seen in previous papers or, possibly, even just the indices that support their model. Therefore, knowing indices' strengths and weaknesses and making these known to the general academic public is a necessity.

1.5.3.1.2 Cutoff Values

Another important concern regarding the use fit indices is the application of index cutoff values to judge the adequacy of a model. Similar to a significance level in hypothesis testing, a cutoff value is a set value for a given fit index that can be used to decide whether or not a model adequately fits the data (Marsh et al., 2004). For example, it is common practice to use a cutoff value of .95 for most incremental fit indices (such as the CFI), where index values greater than or equal to .95 suggest adequate model fit and index values less than .95 suggest a less than optimal fit.

The use of cutoff values has grown in popularity as fit indices have become more widely used in SEM research. Part of the appeal of cutoff values is that they allow researchers to make a "yes" or "no" conclusion when determining whether a model exhibits good fit. When reporting the value of a fit index, researchers have the cutoff value to compare it to in order to justify their conclusion regarding model fit.

However, unlike the significance level use in hypothesis testing, there is little to no rationale behind why any specific index value is used as the cutoff criterion. For example, while there may be agreement among researchers that RMSEA values greater than .06 suggest poor model fit, there are no strong arguments in the literature as to why .06 should be chosen as the cutoff value rather than .05 or .07. While a model with an RMSEA value of .062, for example, would be rejected using the strict cutoff of .06, there is no evidence to suggest that this model has substantially worse fit than a model with an RMSEA value of .058 (which would be accepted according to the cutoff criterion).

In addition to these concerns, there have been many studies (e.g., Beauducel and Wittmann, 2005; Hsu et al., 2015; Marsh et al., 2004; Sivo et al., 2006) demonstrating that a single cutoff value cannot be used reliably under all measurement and data conditions. Sivo et al. (2006) criticize the use of "universal" cutoff values, noting that sample size and model type both affect index behavior. Thus, even within the same index, there is a general lack of comparability across different model types and sample sizes, making a universal cutoff value impractical. Marsh et al. (1988), Marsh et al. (2004), and Fan et al. (1999) all support this view.

Other criticisms of cutoff values stem from the reason fit indices were developed in the first place. Initially, many indices were developed to gauge model fit along a continuum from no fit to perfect fit (Bentler and Bonett, 1980; Marsh et al., 2004). This provided additional information regarding the adequacy of a model's fit over and above the fit/no fit conclusion based on the chi-square test. However, if researchers are using index cutoff criteria in a fashion similar to a significance level in hypothesis testing, some argue that fit indices provide no more information about model fit over the information provided by the chi-square test (Hu and Bentler, 1998).

Even Hu and Bentler (1998, 1999), whose work had focused on developing more appropriate cutoff criteria, argued against using them as "golden rules." They and others (e.g., Fan et al., 1999; Marsh et al., 2004; Perry et al., 2015) emphasize that the purpose of a fit index is to allow researchers to examine the degree of misspecification rather than to come to a strict conclusion about accepting or rejecting a model based on its fit. Indices are meant to complement the "fit vs. no fit" conclusion of the chi-square rather than to be used as another method to come to that conclusion. Thus, while fit indices are still of use in assessing model fit, researchers must exercise caution when interpreting index values with respect to commonly used cutoff criteria.

1.5.3.2 Application-Related Concerns

In addition to the more theoretical and methodological issues surrounding the use of fit indices, consideration must also be given to how indices perform given varying aspects of the modeling procedure. Previous research has shown that index behavior is affected by such things as sample size, estimation method, and various model components such as model size, model parameter size, and misspecification type.

While it is expected that indices will perform differently under different conditions, it is important that researchers are aware of how changes in their modeling procedure (e.g., using generalized least squares instead of maximum likelihood as the estimation method) may affect index behavior. In the following sections, I summarize several important studies in SEM literature that focus on how popular indices are affected by various aspects of the modeling procedure.

While the effect of sample size is commonly studied with respect to its influence on index behavior, the present research addresses the behavior of indices in the population and thus is not concerned with the effect of sample size. As such, only a brief overview of the effects of sample size will be summarized here. For the remainder of the literature review, the effects of sample size will only be discussed when they are presented in conjunction with other influences (such as model size, for example) that remain of interest at the population level.

1.5.3.2.1 Sample Size

One of the primary concerns with the chi-square test is its sensitivity to sample size. Many fit indices were developed in response to this concern, the goal being to create a test of fit that was less sensitive to the effects of sample size. However, research has shown that the majority of fit indices are not immune to the effects of sample size, with some indices being just as sensitive as the chi-square.

In a large study by Marsh et al. (1988), the performances of 29 different fit indices were examined to determine which, if any, were relatively independent of sample size. Index performance was examined in seven sample size conditions (ranging from n = 25 to n = 1,600) when fitting a three-factor model with nine indicator variables (three for each factor) to four different sets of data. The variation of an index's values across the different sample size conditions was used to determine to what degree that index was independent of the influence of sample size.

Of the 29 indices studied, the authors found 24 of them to have values significantly affected by sample size. This was true despite the fact that the degree of misspecification in the models remained the same. The five indices significantly unaffected by sample size were the TLI and the four indices in the study that were based on the TLI.

A similar study by Fan et al. (1999) examined the effects of sample size, estimation procedure, and model misspecification on the performances of nine popular fit indices. Five levels of sample size (from n = 50 to n = 1,000), three levels of misspecification (true model, slightly misspecified model, moderately misspecified model), and two estimation methods (maximum likelihood and generalized least squares) were incorporated into the $5 \times 3 \times 2$ design carried out by Monte Carlo simulation. Misspecification was achieved by deleting true paths and adding false paths (the slightly misspecified model involved deleted paths only; the moderately misspecified model involved both deleted paths and false paths).

The model structure included four factors, each of which had three or four indicators. The authors commented that this model structure replicates what is commonly found in SEM in practice (specifically, models with two to six factors and three or four indicator variables per each factor).

The authors found that regardless of estimation method and model misspecification, GFI and AGFI were most strongly affected by sample size, showing an increase in fit as the sample size increased. While still affected by sample size, RMSEA, CFI, and NNFI were less sensitive to its effects than GFI and AGFI. RMSEA, CFI, and NNFI were also most sensitive to the size of the misspecification, a desirable result.

Another similar study was performed by Sharma et al. (2005) in which the effects of sample size, model size, and factor correlations between factors were examined for RNI, TLI, RMSEA. The authors used simulations to empirically assess the effects of these factors on fit indices as well as on the use of prespecified cutoff values.

The models included in the study were two specifications of two-, four-, six-, and eight-factor CFA models with four indicators per factor. This resulted in a range of relatively small models (two factors with eight indicators and one factor correlation) and relatively large models (eight factors with 32 indicators and 28 factor correlations). Each model was either correctly specified (the hypothesized model was identical with the population model) or misspecified (the hypothesized model was not the same as the population model).

The authors found that in the misspecified cases, the results suggested that as the models grew larger, a larger sample size was needed before these indices became insensitive to the effects of the sample size. However, GFI was most sensitive to sample size regardless of the size of the model, while RMSEA was least affected by the interaction between sample size and model size.

Other studies have been carried out examining fit indices in different modeling scenarios (e.g, Anderson and Gerbing, 1984; Bearden et al., 1982; Kenny et al., 2015; La Du and Tanaka, 1989). In all of these studies, the effects of sample size have been well-documented, suggesting that most indices are significantly sensitive to sample size, despite their having been developed to be less sensitive than the chi-square to sample size effects.

1.5.3.2.2 Estimation Method

In the estimation step of the SEM process (discussed in detail in section 1.4.1 above), the minimization of some discrepancy function $F(\theta)$ leads to parameter estimates $\hat{\theta}$ such that the hypothesized model's covariance structure, when based on these estimates, is as similar to the structure of the sample covariance matrix S as possible. There exist different estimation methods, such as maximum likelihood (ML) and generalized least squares (GLS), that are used to achieve the minimization of $F(\theta)$. It has been shown that certain indices perform differently depending on the estimation method used. This section will provide only a brief summary of studies examining this effect, as the current research focuses on index behavior only under ML estimation.

Sugawara and MacCallum (1993) found that absolute indices, such as RMSEA and SRMR, have a tendency to behave more consistently across different estimation methods than relative fit indices (such as CFI and NNFI). These findings were supported by Fan et al. (1999) in the same study described in the previous section. Recall that this study included a true model with four latent variables with three or four indicators each, and two misspecified models. Misspecification was achieved by deleting true paths and adding false paths; the slightly misspecified model involved deleted paths only and the moderately misspecified model involved both deleted paths and false paths. Index values were examined at five different sample sizes using both ML and GLS estimation procedures.

While Fan et al. (1999) found that estimation method had no effect on index value in the true model case, when the models were misspecified to any degree, large differences in index value were found for NFI, CFI, and NNFI under different estimation procedures, while relatively small differences were found for indices not defined as relative fit indices. In general, all indices showed better fit under GLS estimation than ML estimation.

1.5.3.2.3 Model Size

Apart from sample size, perhaps the most researched influence on fit index value is model size. Model size can be defined both by the number of factors (k), which is usually determined by theory, and by the number of indicator variables (p) in a given model. The issue of index sensitivity to model size has raised concern amongst researchers since model size is an aspect of a researcher's model that can be changed relatively easily (in comparison to something like the sample size or the size of the parameter estimates).

As described in section 1.4.2.1, the traditional method of assessing model fit in SEM is by using the T statistic, which has a χ^2_d distribution asymptotically. In addition to this statistic's sensitivity to sample size, it has been documented that it is sensitive to model size as well (e.g., Fornell, 1983; Moshagen, 2012). In finite samples, the size of a model has been shown to affect the goodness of approximation of the T statistic to the asymptotic chi-square distribution. Specifically, the T statistic tends to become "inflated" as more variables are added to the model and thus may lead to the rejection of a correct model simply because the model is very large.

Moshagen (2012) notes that many of the popular fit indices, including RMSEA and CFI, are based on the T statistic and might also be misleading when it comes to representing fit for larger models. The concern expressed by Moshagen (2012), Kenny and McCoach (2003), Chau and Hocevar (1995) and others is that if models with more variables exhibit worse fit, researchers may be tempted to adopt a variety of strategies that would reduce the number of variables in the model in order to improve fit. For instance, items may be collapsed to form parcels, larger models could be broken down into submodels that contain only a subset of the variables, or variables could simply be trimmed from the model (Kenny and McCoach, 2003).

On the other hand, if indices show an improvement in fit when the number of variables increase, it could possibly lead to researchers including variables that should not theoretically belong in the model but are added solely to improve fit. Because of these concerns, the effect of model size on index behavior has been studied from several different aspects.

Many studies have been conducted examining the effect of the number of indicators on index behavior. A study by Kenny and McCoach (2003) involved simulating perfectly specified and misspecified models while varying the number of indicators. The study focused on the performances of TLI, CFI, and RMSEA. For three different sample size conditions (100, 200, 1,000), a hypothesized 1-factor model had either 4, 6, 10, 12, 14, 20, or 25 indicators and was either perfectly specified or misspecified in one of three ways. The first misspecification involved the omission of the loadings of a second minor population factor. The second misspecification involved the case where the hypothesized 1-factor model omitted the correlation between the two factors present in the population. The third misspecification involved omitted error covariances amongst the indicators in the fitted 1-factor model.

The authors found that for perfectly specified models, CFI and TLI tended to show worse fit as model size increased (especially in small sample size conditions), while RMSEA tended to show an improvement of fit as the model size increased, but only in the larger sample size condition (N = 1,000). As model size increased, both CFI and NFI showed a decrease in fit in both misspecification cases where the misspecification involved the latent structure of the model (the omission of a second minor factor and the omission of a factor correlation), but showed an increase in fit when the misspecification was due to omitted error covariances.

RMSEA showed an improvement in fit as model sized increased regardless of the type of misspecification. This result is consistent with the findings of other studies (e.g, Browne, 1987), supporting the claim that RMSEA shows better fit as models grow larger. Kenny and McCoach (2003) suggest that the decline in RMSEA value, indicating improved fit, is due to the decline in the ratio of the model chi-square to its degrees of freedom, since adding more observable variables to a model increases the degrees of freedom faster than it increases the chi-square value.

The effect of the number of indicators on index behavior was also examined in a study by Chau and Hocevar (1995). The authors examined five popular fit indices (GFI, AGFI, NFI, CFI, and TLI) to determine which were least susceptible to the effect of model size on index value. Their initial CFA model consisted of 28 indicators loading on to seven intercorrelated factors (resulting in four indicators per factor). In two subsequent manipulations, the same CFA model was maintained but the number of observable variables per factor was reduced by random deletion down to three per factor (21 indicators total) and then down to two per factor (14 indicators total).

The authors found that while all indices in the study showed worse fit for models with more indicators, CFI, NFI, and TLI were more stable than the others, meaning that as model size increased, the values of these indices showed worse fit but did not differ substantially from their values for smaller models. This suggests that while all these indices are affected by model size to some degree, the three relative fit indices included (and possibly other relative fit indices as well) may be less sensitive to model size than absolute fit indices are.

I point out that in this study by Chau and Hocevar (1995) and in the study by Kenny and McCoach (2003), no effort was made to control for changes in the ratio of indicators (*p*) to factors (*k*). That is, as *p* decreased, the ratio of *p:k* decreased as well, from 4:1 to 3:1 to 2:1. In addition to examining how the number of indicators in a model might affect index behavior, the current study will also examine how the ratio of *p:k* affects the behavior of RMSEA and CFI. A comparison of my results and the results found by Chau and Hocevar (1995) will allow insight as to whether these indices are sensitive to the ratio of observable variables to indicator variables over and above changes in *p* (or *k*) alone.

Other studies have manipulated both the number of indicators as well as the number of factors. In a simulation study by Sharma et al. (2005), the authors constructed four CFA models with different numbers of factors (2, 4, 6, and 8) with four indicators each (leading to a total of 8, 16, 24, and 32 indicators variables for each model, respectively). Two specifications for each model were created. In the first specification, the model was correctly specified and matched the population structure perfectly; in the second specification, the model was not correctly specified (due to omitted factor correlations). The authors also manipulated factor loading sizes (.3, .5, .7), factor correlation sizes (.3, .5, .7), and sample sizes (100, 200, 400, 800). The performances of RNI, TLI, GFI, and RMSEA were assessed for each condition.

Sharma et al. (2005) found that GFI was significantly affected by both sample size and model size. Specifically, GFI showed worse fit as the sample and model size increased for the misspecified model case. RMSEA was significantly affected by the size of the model; it showed an improvement in fit for the misspecified model as the model size increased.

In contrast to the studies done by Chau and Hocevar (1995) and Kenny and McCoach (2003), I note here that the ratio of *p:k* was controlled in this study. That is, the number of indicators per factor is held constant as the number of factors increases, meaning that the *p:k* ratio stays the same as the model increases in size. The present study will examine the effect of this ratio in further detail and compare the results to the results found by Chau and Hocevar (1995), Kenny and McCoach (2003), and Sharma et al. (2005).

1.5.3.2.4 Parameter Values

Other model components that may have an effect on fit index behavior are the values of the model parameters, such as loading size or factor covariance size. These values can be specified (or "fixed") in a model based on theory and on previous research. For example, in personality psychology, the covariance between factors Extraversion and Agreeableness in a given model may be based on what previous research suggests is an appropriately strong relationship between the two factors. Parameters can also be estimated during the modeling procedure. Most of the studies addressing the effect of model parameters on index behavior focus on the influence of fixed parameters rather than on the influence of estimating parameters.

A recent study by Themessl-Huber (2014) examined the effect of loading size on the behavior of SRMR, RMSEA, and CFI. The study included three CFA models, each with 24 indicators and three factors (each factor had 8 indicators loading onto it). The first model was a correctly specified model with uncorrelated factors and no cross-loadings. The second model was a misspecified version of the first model and included correlated factors but no cross-loadings (the factor correlations were .3, .4, or .5). The third model was also a misspecified version of the first model and included cross-loadings but no correlated factors (the primary loadings were all between .3 and .9, while the cross-loadings were no greater than .2).

Index performance in these scenarios was assessed using the type I and type II error rates when model rejection/acceptance was based on the popular cutoff values for the indices (models were accepted when RMSEA $< .06$, SRMR $< .08$, and CFI $> .95$). The author found that when factor loadings were low or medium (less than .6), CFI had trouble accepting correctly specified models. RMSEA and SRMR had better rates of acceptance for low loadings. However, CFI did the best of the three indices when it came to the misspecified models, especially when the loadings were high (.8, .9). That is, even when loadings were high in the misspecified models, CFI still rejected these models due to their poor fit.

In general, Themessl-Huber (2014) found that all indices had trouble detecting misspecified models when the factor loadings were low. This finding is cause for concern, as in certain disciplines where it is common to have low loadings (such as psychology), a model could be accepted as having good fit simply because the loadings are small rather than because the model actually fits the data well.

Another study by Miles and Shevlin (2007) also showed that index performance can change due to loading size alone. In the second half of a two-part study, the authors fit a one-factor model to two-factor data, fixing the loadings to .8 and the factor correlation to .5. Of the indices included in their study (RMSEA, CFI, NNFI, and RMR), all indices showed the model fitting the data poorly. However, when they fixed the loadings to .5 instead of .8, RMSEA indicated a well-fitting model. Both CFI and NNFI still showed a poor fit.

These results suggest that when the loadings are small, RMSEA may not be sensitive enough to detect when a model omits a factor, which can be considered to be a rather large misspecification. The authors argue that this finding supports the use of comparative fit indices such as the CFI alongside the chi-square to gain a better understanding of the source of model misspecification.

The modeling scenario presented by Miles and Shevlin (2007) is similar to the scenarios carried out in the present study to explore how sensitive fit indices are to detecting cases of latent structure misspecification. Given the results of their study, I may expect to find that the RMSEA's ability to detect latent structure misspecification is in part a function of the size of the loadings.

1.5.3.2.5 Other Model Components

Before I summarize the literature discussing how fit index behavior is affected by the type of model misspecification, I wish to briefly discuss other model components that may play a role in how fit indices behave.

As discussed in the previous section, research has shown that fit index behavior is affected by model size. Many popular indices, including the RMSEA, tend to show a better fit for large models over comparable small models. However, Moshagen (2012) notes that the notion of a “large” model is vague. To some researchers, the definition of “large” is based upon the number of indicators (e.g, Anderson and Gerbing, 1984; Kenny and McCoach, 2003; Marsh et al., 1998), while to others, the definition is based on the number of parameters needed to be estimated in the model (e.g., Boomsma, 2000; Curran et al., 2002). Some consider the degrees of freedom to be indicative of model size, as it is based on both the number of unique elements of the covariance matrix and the number of free parameters (recall that the degrees of freedom $d = \frac{p(p+1)}{2} - q$, where p is the number of observed variables in the model and q is the number of unique parameters to be estimated).

In CFA, both the number of free parameters and the degrees of freedom increase as the number of indicators increases, which may explain why there is little concern as to what is actually meant by a “large” model in the literature. However, it should be noted that models with the same number of indicators can lead to different degrees of freedom. Moshagen (2012) offers the example of model A, which has three correlated factors and 10 observable variables each (here, p = 30, q = 63 and df = 402), compared to model B, which has 10 correlated factors and three observable variables each (p = 30, q = 105, and df = 360).

A limited number of studies have made this distinction between the number of indicators and the size of the degrees of freedom when assessing how model size affects index behavior. A study by Kenny et al. (2015) examined the behavior of RMSEA in models with small degrees of freedom. The study included seven different degrees of freedom conditions (1, 2, 3, 5, 10, 20, or 50) and six different sample size conditions (50, 100, 200, 400, 600, or 1,000). All models were correctly specified.

The authors found that when both the degrees of freedom and the sample size were large, RMSEA never rejected the model (that is, it showed that the model had good fit). This replicated similar findings by Chen et al. (2008). However, for both very small sample sizes (100 or below) and small degrees of freedom, RMSEA rejected a substantial proportion of correctly specified models. Even with large sample sizes combined with small degrees of freedom, RMSEA suggested poor model fit.

It is noted that while Kenny et al. (2015) focused on the effect of the degrees of freedom rather than on the effect of the number of indicators, their conclusions regarding the behavior of RMSEA are the same as in the study by Chen et al. (2008). That is, the larger the degrees of freedom, the greater the tendency of RMSEA to show good model fit. Other studies (summarized above) have shown that the larger the number of indicators, the greater the tendency of the RMSEA to show good model fit. This perhaps suggests that since the degrees of freedom are in part a function of the number of indicators that RMSEA is in fact sensitive to model size in terms of the number of indicators rather than the degrees of freedom.

Another model component that may affect fit index behavior is the balance of the indicators across the different factors in the model. In all of the simulated models in the studies summarized above, there is an equal number of indicators loading onto each factor. For example, if a model has three factors and 12 indicators, each factor has four indicators loading onto it.

However, there is nothing to suggest that there needs to be an equal number of indicators per factor in real life models. In an application of SEM, for example, there may be a three-factor model where one factor has four indicators loading onto it, the second factor has two indicators loading onto it, and the third factor has six indicators loading onto it.

Despite the fact that such “unbalanced models” are completely reasonable model designs, there has been no research done on the effect of balance on index behavior. The present study aims to offer some insight as to how CFI and RMSEA behave when the number of indicators is not equally spread across all factors in a given model.

1.5.3.2.6 Misspecification Type and Severity

While fit indices are designed to measure the degree of fit of a model, the previous sections reveal that indices are sensitive to other model components that do not directly involve the degree of misspecification. Thus, sensitivity to components such as sample size and estimation method may be considered a weakness of fit indices.

However, it is desirable for fit indices to be sensitive to the nature and degree of the misfit between a hypothesized model and the data. As previously stated, many fit indices were designed to supplement the chi-square. The chi-square results in a binary fit/no fit decision. It is not designed to provide any information about the source of a possible misspecification or how severe it may be.

If fit indices are sensitive to the source misspecification (as well as the severity of the misspecification), they become more useful as a supplement to the chi-square, as they can provide information about where a possible misspecification may exist (e.g., if it is due to an omitted error covariance or due to a misspecified latent structure).

Indices become even more useful if different indices are sensitive to different sources of misspecification. If, for example, it is known that CFI is sensitive to a certain type of misspecification and RMSEA is sensitive to another type, then researchers can examine both indices (in addition to the chi-square) to possibly determine where a misspecification may be occurring in their model.

There have been several studies that have examined index sensitivity to misspecification type and misspecification severity. The earliest comprehensive study was done by Hu and Bentler (1998), who examined the behavior of 15 indices with respect to the severity of model misspecification. The goal was to determine which indices, if any, accurately reflected the degree of misspecification. The authors examined index performance in CFA models involving three factors and 15 indicators. For what they called the "simple" misspecification scenario, there were three different models: one that was correctly specified, one that involved one omitted factor covariance, and one that omitted two factor covariances. For the "complex" misspecification scenario, there were also three different models: one that was correctly specified, on that involved one omitted cross-loading, and one that omitted two cross-loadings.

In addition to the differing degrees of misspecification, the authors also included three estimation method conditions (ML, GLS, and asymptotic distribution free (ADF)) and six sample size conditions (n = 50 to n = 5,000). They measured sensitivity to misspecification by using an ANOVA. The larger the amount of variance accounted for by the model misspecification (and the smaller the amount of variance accounted for by the sample size and estimation method), the better and index was said to be.

Results from the study showed that in the case of the "simple" misspecification (omitted factor covariances), SRMR, TLI, BL89, RNI, CFI, Mc, gamma hat, and RMSEA all performed well in terms of large proportions of their variances being accounted for by the size of the misspecification. Other indices, including GFI, AGFI, NFI, CAK, and CK, were highly affected by sample size.

In the case of the "complex" misspecification, the authors found that large proportions of the variances of TLI BL89, CFI, and RMSEA were accounted for by misspecification. Overall, the authors found that all indices apart from SRMR were more sensitive to misspecifications due to omitted cross-loadings than those due to omitted factor covariances. Based on this, they recommended a two-index presentation in results reporting, coupling SRMR with one other index.

One criticism of Hu and Bentler's (1998) study design is that the severity of model misspecification was neither defined nor controlled. Therefore, the simple and complex misspecificaitons in the study may not have comparable degrees of misspecification. Fan and Sivo (2005) sought to expand upon Hu and Bentler's (1998) study by quantifying the degree of misspecification in each model so as to keep it consistent across the two misspecification types (simple and complex).

To quantify the degree of misspecification, Fan and Sivo (2005) treated the chi-square values for the models as noncentrality parameters. That is, these chi-square values (and their associated degrees of freedom) describe the amount of shift from the central chi-square distribution to the non-central chi-square distributions due to model misspecification. Since these values are blind to the type of misspecification, the authors argued that they would be a good way to compare the severity of misspecification of the simple and complex cases presented by Hu and Bentler (1998).

By comparing the ($\chi^2 - \mathrm{df}$) from the different models, they found that the misspecification was less severe in the "simple" case than in the "complex" case. Thus, the results in Hu and Bentler's (1998) study may have been partly due to these differences in misspecification severity. To determine if this was the case, Fan and Sivo (2005) repeated Hu and Bentler's (1998) original study while controlling for the severity of model misspecification. To do so, they adjusted the population model parameters in the original designs so that the models would have comparable degrees of misspecification.

After controlling for the severity of misspecification, Fan and Sivo (2005) found results similar to those in the original study. TLI, BL89, RNI, CFI, Mc, and RMSEA all appeared more sensitive to misspecifications due to omitted cross-loadings than misspecifications due to omitted factor covariances. This suggests that certain indices are indeed more sensitive to certain types of misspecification, even when the degree of misspecification is controlled across the types.

In a follow-up study, Fan and Sivo (2007) sought to further examine fit index behavior with respect to the degree of model misspecification. They argued that if universal index cutoff values are to be of any practical use in SEM, it is important that fit indices be sensitive to the severity of model misspecification (regardless of the source) but not be sensitive to different types of models that have the same degree of misspecification.

To determine which indices (if any) were sensitive to model type if the degree of misspecification was the same across the different models, the authors constructed two different CFA models. CFA-a contained misspecifications due to omitted cross-loadings, and CFA-b contained misspecifications due to omitted factor covariances. Both models had three different levels of misspecification: no misspecification, a single instance of misspecification (e.g., one omitted cross-loading), and two instances of misspecification (e.g., two omitted cross-loadings). The degree of misspecification was held constant across model types for each of the three levels of misspecification.

Of the twelve indices included in the study, NFI, RHO1, and SRMR were shown to be most sensitive to model type, with 20% or more of their variation attributable to this component.

Gamma, Mc, and RMSEA were shown to be the least sensitive to model type but most sensitive to the degree of model misspecification. CFI, while sensitive to the degree of misspecification, was also slightly sensitive to the type of model as well.

Fan and Sivo (2007) also included a condition to the study which compared index behavior in the original models to index behavior in models that were smaller but otherwise similar to the originals. They found that though the severity of misspecification was the same in the two smaller models as it was in the original models, RMSEA values were dramatically higher for the smaller models, suggesting that these models had a higher degree of misspecification than the larger models. These results agree with research focused on model size and suggest, as other studies do, that RMSEA is highly sensitive to model size.

A more recent study by Heene et al. (2012) also explored the sensitivity of commonly used indices under different sources and degrees of model misspecification. The study focused on the behaviors of RMSEA, SRMR, and CFI for two different CFA models with 24 indicators equally distributed across two correlated factors. One population model, Model A, contained three correlated errors between items loading onto the two different factors. The other population model, Model B, contained six correlated errors.

The study contained five sample size conditions (150, 250, 500, 1,000, and 2,500) and two factor loadings conditions (one with loadings ranging from .3 to .6, the other with slightly higher loadings ranging from .5 to .8). To create the misspecification, the fitted model (fitted to simulated data from both Model A and Model B) assumed entirely uncorrelated errors.

The authors found that regardless of loading size, sample size, or degree of misspecification, both SRMR and RMSEA values were always below their commonly used cutoff values, demonstrating that these indices failed to reject models that had a large degree of misspecification (i.e., six omitted error covariances). CFI values, on the other hand, were almost always below the suggested cutoff value except with the weakest misspecification and the larger sample sizes (greater than 500), suggesting that the index is sensitive to misspecification due to omitted error covariances. CFI values were also higher (indicating better fit) when loadings were higher, suggesting that this index is also affected by loading size.

In the present study, one of the focuses will be on index behavior with respect to misspecification type. Also of interest will be how CFI and RMSEA will perform when the severity of misspecification is altered. The results from the current study will be compared to the results from previous studies to determine if the same conclusions are reached.

1.5.3.3 Summary

The above sections discuss the prominent literature focusing on fit index behavior and the aspects of the modeling procedure that affect different popular indices. In this section, I briefly summarize the results pertaining to the two indices of interest in the current study and discuss how I expect these indices to behave based on what has been demonstrated in the literature.

Previous research points to RMSEA as being very sensitive to model size. More specifically, as the size of a model increases (whether that increase is measured by the number of indicators included in the model or by the degrees of freedom of the model), RMSEA has a tendency to show an improvement in fit, regardless of the type of misspecification. This trend is evident regardless of whether the ratio of indicators to factors (*p:k*) is held constant or allowed to change as the model size increases. (Browne, 1987; Kenny and McCoach, 2003; Sharma et al., 2005).

CFI shows a decrease in model fit as model size increases, but only when the misspecification is due to an incorrect latent structure. When the misspecification is due to an omitted error covariance, CFI shows an improvement in fit as model size increases (Kenny and McCoach, 2003). However, other studies that have duplicated this study find that CFI values are much more stable compared to RMSEA values as model size increases (Chau and Hocevar, 1995). Based on this previous research, I expect that RMSEA will show an improvement in fit as model size increases, regardless of the type of misspecification. However, I suspect that CFI's behavior as model size increases will be affected by the type of misspecification.

Previous research also shows that both indices are sensitive to loading size. While low loadings affect CFI's ability to detect correctly specified models (meaning that low loadings lead CFI to reject models that fit the data perfectly), RMSEA is slightly less affected by loadings (Themessl-Huber, 2014). When loadings are high, CFI is appropriately sensitive, meaning that the index still rejects models that have poor fit. RMSEA, on the other hand, has a tendency to show models as having better fit based solely on the size of the loadings (better fit when loadings are larger), and thus may accept a poorly-fitting model if the loadings are high enough (Miles and Shevlin, 2007).

Based on what has been found in the literature, I suspect that loading size will affect index behavior regardless of the type of misspecification. Specifically, I anticipate that RMSEA will show improved fit (at any degree of misspecification) as loading sizes increase, while CFI will show significantly worse fit as loading sizes decrease. I also suspect that loading size will influence behavior regardless of other modeling factors (e.g., model size).

Finally, previous research shows that while both indices appear to be sensitive to misspecifications due to omitted factor covariances, RMSEA is less sensitive to misspecifications due to omitted error covariances than CFI is (Fan and Sivo, 2005; Heene, et al., 2012; Hu and Bentler, 1998). While the current study does not examine omitted factor covariances, it does examine misspecified latent structure, which involves fitting a model with a certain number of factors to data that comes from a population with a greater (or fewer) number of factors than the fitted model. Based on index behavior with respect to omitted factor covariances, I anticipate that both indices will be sensitive to misspecified latent structures, while CFI will be more sensitive to omitted error covariances than RMSEA.

Chapter Two: Methods

In this chapter, I discuss the methodology used to carry out the simulations presented in Chapter 3. I begin with a brief overview of the *estimation* and *fit assessment* steps of SEM, which are discussed in more detail in Chapter 1. I then describe the simulation methods employed here, focusing on the construction of the “population” and “hypothesized” covariance matrices, the minimization of the discrepancy function, and the plotting of the results. Finally, I re-state the goals of the present study and describe how I will attempt to achieve them.

2.1 Estimation and Assessing Model Fit

As discussed in Chapter 1, the null hypothesis in SEM states that the population covariance matrix $\boldsymbol{\Sigma}$* has the structure implied by the researcher’s hypothesized model. That is, the null claims that $\boldsymbol{\Sigma^*} = \boldsymbol{\Sigma(\theta)}$, with $\boldsymbol{\theta}$ being a vector of unknown population parameters. In order to test this null hypothesis, both $\boldsymbol{\Sigma^*}$ and $\boldsymbol{\theta}$ must be calculated under the researcher’s hypothesized model.

However, since the population covariance matrix $\boldsymbol{\Sigma^*}$ is not known, we cannot directly test the null hypothesis that $\boldsymbol{\Sigma^*} = \boldsymbol{\Sigma(\theta)}$. Instead, it is assumed that the covariance matrix **S** of a sample drawn from the population of interest is a good representation of $\boldsymbol{\Sigma^*}$. The primary goal of the estimation step, then, is to find the vector of parameter estimates $\boldsymbol{\hat{\theta}}$ such that the difference between $\boldsymbol{\Sigma}(\boldsymbol{\hat{\theta}})$, the hypothesized models’ covariance structure based on these estimates, and **S**, the sample covariance matrix, is as small as possible.

These parameter estimates are obtained by minimizing some discrepancy function F($\boldsymbol{\theta}$). The most commonly used estimation procedure in SEM is the maximum likelihood (ML) procedure. Recall equation (1.1) from the previous chapter:

$$\mathrm{F}(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \ln|\boldsymbol{\Sigma}^*| + \mathrm{tr}[\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}] - p, \tag{1.1}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ represents the structure of the covariance matrix implied by the hypothesized model, $\boldsymbol{\Sigma}^*$ represents the population covariance matrix, and p is the number of indicator variables. Since $\boldsymbol{\Sigma}^*$ is not known, the sample covariance matrix **S** replaces $\boldsymbol{\Sigma}^*$ and (1.1) is expressed as

$$\mathrm{F}(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \ln|\mathbf{S}| + \mathrm{tr}[\mathbf{S}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}] - p. \tag{1.2}$$

When F(θ) is minimized, we obtain $\hat{\mathrm{F}}$ and the corresponding vector of parameter estimates $\hat{\boldsymbol{\theta}}$. $\hat{\mathrm{F}}$ attains the value of zero if and only if $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{S}$. Otherwise, $\hat{\mathrm{F}}$ is positive and increases as the discrepancy between $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and **S** increases.

Once $\hat{\mathrm{F}}$ and $\hat{\boldsymbol{\theta}}$ have been found, the next step is to use them to assess how well the hypothesized model fits the data. Specifically, we want to compare $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ to **S** in order to determine how similar they are. Fewer discrepancies between the two matrices suggests that the proposed model is a good fit to the underlying population covariance structure from which the sample was drawn. The traditional method of assessing goodness-of-fit involves using the chi-square test statistic, which is based on the minimized fit value $\hat{\mathrm{F}}$. Recall from Chapter 1 that the chi-square test statistic T can be expressed as

$$\mathrm{T} = (\mathrm{N} - 1)\hat{\mathrm{F}}. \tag{1.3}$$

Most fit indices that are commonly used in SEM rely on $\hat{\mathrm{F}}$ as well, expressed as a chi-square value.

2.1.1 Sample versus Population Fit Indices

Note that in order for $\hat{F}$ to be expressed as a chi-square value, it must be multiplied by (N – 1) and thus depends on the sample size (which in turn implies that any fit index based on $\hat{F}$ is also influenced by sample size). Many previous studies (e.g., Fan et al., 1999; Kenny et al., 2015; Marsh, et al., 1988) have shown that fit index behavior is, in most cases, heavily influenced by sample size. In the present study, the goal is to examine index behavior independently of sample size.

In order to do so, the equations for the two fit indices of interest (CFI and RMSEA) must be re-written so as to not depend on the sample size n. To derive these population equations, $\hat{F} \cdot N$, where $N = n + 1$, was used to replace the chi-square values found in the sample definitions of the indices. The sample size n was then allowed to tend to infinity.

The value of $\hat{F}$, as stated above, is obtained by minimizing the function $F(\boldsymbol{\theta})$, which relies partially on the sample covariance matrix **S**. Because **S** is affected by n, $\hat{F}$ is still affected by sample size. Rather than relying on sample covariance matrices (**S**), I instead define population covariance matrices ($\boldsymbol{\Sigma}$*****) whose structures are explicitly defined to reflect the relationships amongst variables in the population. These $\boldsymbol{\Sigma}$***** matrices therefore do not rely on n and, subsequently, $\hat{F}$ is independent of sample size in this study. In other words, I compute $\hat{F}$ based on equation (1.1) rather than (1.2). Please refer again to Appendix A for the population values of commonly used fit indices. The population equations for CFI and RMSEA were used to calculate the fit index values for the model simulations presented in this study.

2.2 Steps of the Simulation Procedure

The simulations presented in the following chapter illustrate how these indices behave in specific modeling scenarios. In order to control the source and severity of misspecification (as well as different model components, such as loading size, model size, factor correlation, etc.), the simulations involve first creating a "population" model and then fitting deliberately misspecified models to that population model.

All simulations and calculations were carried out using the statistical software R, and several randomly chosen points for each scenario were verified using EQS 6.3. In cases where convergence problems were present, the outlier values were replaced by the average of the surrounding values. The steps involved in the simulation procedure are described here.

Step 1: Definition of the Model Components and Loops

The first step in each simulation is to define the general components of the model that were fixed (i.e., not manipulated at all during the simulation). Depending on the scenario, these components include any number of the following: the number of indicators (p), the number of factors (k), loading sizes (λ), and factor correlation (ϕ). For example, Figure 3.1 represents a scenario in which a hypothesized 1-factor model with 8 indicators is misspecified in the sense that it omits an error covariance that is present in the population. In this scenario, neither the number of indicators in the model nor the number of factors in the model are changed, so these values ($p = 8$ and $k = 1$) are fixed at the beginning of the simulation.

In most simulations, one or two model components are manipulated so that the simulation produces results for different values of said model components. In such cases, a loop in the code is created so that the simulation procedure—specifically, the minimization procedure and calculation of fit index values—occurs for each value of the manipulated component. Using Figure 3.1 again as an example, it is of interest to see how different loading sizes affected model fit. Thus, a vector of loading sizes is defined (λ = .4, .5, .6, .7, .8, or .9) and a loop is created so that the fit function is minimized and CFI and RMSEA values are calculated for each value of λ.

Step 2: Construction of the Population Covariance Matrix

As previously stated, the goal of the present research is to examine the behavior of CFI and RMSEA independent of the influence of sample size. Thus, there is no need to simulate data in order to carry out the calculations of the population CFI and RMSEA values. Instead, simulations and calculations are carried out strictly based on covariance matrices constructed based on an assumed "true" population model and a "hypothesized" model.

Before the construction of these models is discussed, it should be noted that the calculations of the population fit indices can rely entirely upon covariance matrices because the fit function being minimized only requires three components: the covariance matrix implied by the hypothesized model, the population covariance matrix, and the number of indicator variables (see equation (1.1)). Thus, as long as the two covariance matrices can be constructed in some way, there is no need to simulate data in order to produce them.

The first covariance matrix constructed in the simulations is $\mathbf{\Sigma^*}$. This matrix represents the "true" or "population" covariance matrix. To obtain $\mathbf{\Sigma^*}$, it is implied that certain relationships exist amongst variables in the population and that these relationships can be described by a structural equation model. For the simulations in this study, the focus is on 1- and 2-factor CFA models. Therefore, $\mathbf{\Sigma^*}$ is calculated either as

$$\mathbf{\Sigma}^{*} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\psi} \tag{2.1}$$

for a 1-factor model, where $\boldsymbol{\lambda}$ is a $p \times 1$ vector of population factor loadings and $\boldsymbol{\psi}$ is the $p \times p$ covariance matrix of the residuals, where p represents the number of indicator variables in the model, or as

$$\mathbf{\Sigma}^{*} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \boldsymbol{\psi} \tag{2.2}$$

for a 2-factor model, where $\mathbf{\Lambda}$ is a $p \times k$ matrix of population factor loadings, $\mathbf{\Phi}$ is a $k \times k$ matrix of population factor correlations, and $\boldsymbol{\psi}$ is the $p \times p$ covariance matrix of the residuals, where k represents the number of latent factors in the model.

In most scenarios presented in the following chapter, the discrepancy between the population covariance matrix **Σ*** and the hypothesized model is due to the hypothesized model failing to include a component of the model underlying the population. Thus, an additional loop is created within the simulations in order to vary the size of this component of interest in **Σ*** to observe how its omission from the hypothesized model affects model fit.

For example, in Figure 3.1, there exists a single error covariance in the population model, which is in turn reflected in **Σ*** (specifically, $\boldsymbol{\psi}$ is not diagonal). The hypothesized model omits this error covariance ($\boldsymbol{\psi}$ is diagonal in the hypothesized model), which is reflected in the estimated covariance matrix ($\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$, discussed in Step 4). A visual representation of the population model and hypothesized model is given in Figure 2.1.

In the code for Figure 3.1, a loop is inserted such that model fit is assessed for population error covariances ranging from 0 to .84, in .01 increments. That is, for a given loading size, the fit function minimization is performed and recorded 85 separate times for the 85 different error covariance sizes in the population. The description of the specific fit function used in this simulation is given in the next step.

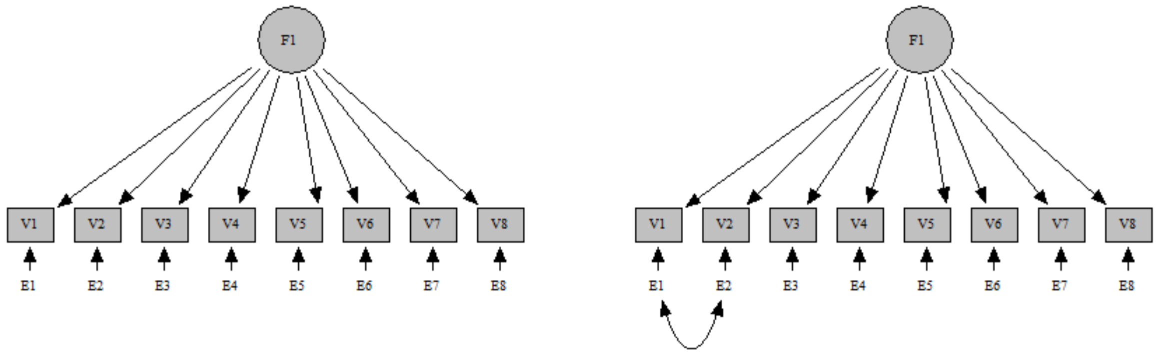

Figure 2.1: The hypothesized model for Figure 3.1 (left), omitting the error covariance that exists in the population model for Figure 3.1 (right).

Step 3: Define the Fit Functions to be Minimized

Within the loop described in step 2, the fit function must be minimized in order to obtain $\hat{\mathrm{F}}$ and the and the corresponding $q \times 1$ vector of parameter estimates $\hat{\boldsymbol{\theta}}$, used in the estimated covariance matrix. Note, however, that there are actually two fit functions to be minimized. Recall from Chapter 1 that CFI requires information not only from the hypothesized model but from a "baseline" or "independent" model as well. Since the model most often used for this baseline model is one in which all observed variables are uncorrelated, this is the model used in the present study. Thus, fit functions for both the hypothesized model (used for both CFI and RMSEA calculations) and for the baseline model (used for CFI) are constructed.

For the hypothesized model's fit function, a series of starting values are defined to be used in the minimization procedure. These starting values are used to construct $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, the matrix based on the hypothesized model's covariance matrix structure. Following this, the fit function $\mathrm{F}(\boldsymbol{\theta})$ is defined as in equation (1.1):

$$\mathrm{F}(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \ln|\boldsymbol{\Sigma}^{*}| + \mathrm{tr}[\boldsymbol{\Sigma}^{*}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-\mathbf{1}}] - p \tag{1.1}$$

and is minimized using the R function `nlm`, which carries out the minimization of the function using a Newton-type algorithm. An additional function, `try`, is used to handle errors from the `nlm` function. Specifically, if an error is produced by `nlm`, the `try` function allows for an additional loop to be created in which random starting values are employed until a useable result from `nlm` is produced.

For the baseline model's fit function, the above procedures are the same, except **$\boldsymbol{\Sigma}(\boldsymbol{\theta})$** is no longer based on the structure of the hypothesized model but is instead constructed assuming that the variables are all uncorrelated (a diagonal covariance matrix).

Step 4: Obtain and Store the CFI and RMSEA Values

Minimization of the functions in the above step produces $\widehat{\mathrm{F}}$ and $\widehat{\mathrm{F}_{\mathrm{I}}}$, the minimized fit functions corresponding to the hypothesized model and the baseline model, respectively, as well as $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}$, the $q \times 1$ vectors of parameter estimates responsible for minimizing the fit functions for the hypothesized and baseline models, respectively. Using the estimates $\widehat{\boldsymbol{\theta}}$ and the form of the hypothesized covariance matrix yields $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$, the hypothesized model's covariance matrix that minimizes the differences between it and the population covariance matrix **$\boldsymbol{\Sigma}$***. Also computed are the CFI and RMSEA values, calculated based on the minimized fit values obtained in Step 3.

In order to store all the CFI and RMSEA values produced by the loops in these simulations, appropriately sized matrices are created, initially empty, and are subsequently filled with the CFI and RMSEA values resulting from each loop. For example, in Figure 3.1, six different factor loading sizes are considered, as are 85 different population error covariance values. In order to store the CFI and RMSEA values for each combination of loading size and error covariance size, two 85×6 empty matrices, one for the CFI values and one for the RMSEA values, are constructed. These matrices are then filled with the index values as the code cycles through the loops.

2.3 Plotting the Results

A unique feature employed in the current study is the use of continuous curves to examine and display index performance. Most previous research has focused on assessing index performance at a few select values. Here, I plot curves showing index value as a continuous function of the modeling components of interest or the size of the misspecification (e.g, the size of the model or the size of an omitted error covariance). This method of presentation was originally utilized by Savalei (2010) and Mahler (2011) and allows for a clearer representation of index performance under the model aspects and misspecifications of interest.

In addition to the curves, horizontal lines indicating the most commonly agreed upon cutoff value for each index are plotted as well. This is done so that index behavior can be examined against these cutoff values as model components and misspecifications change. The most commonly agreed upon cutoff for CFI is .95 (models with CFI values greater than .95 are said to fit the data well). The most commonly agreed upon cutoff for RMSEA is .06 (models with RMSEA values less than .06 are said to fit the data well).

In order to increase the ease of comparison between the plots of CFI values and the plots of RMESA values, the (1 – RMSEA) values are plotted instead so that they can be interpreted in a similar way to the CFI values. That is, these values should be interpreted as showing better fit as values increase (and perfect fit when (1 – RMSEA) = 1). RMSEA's commonly accepted cutoff value has also been plotted as (1 – .06) = .94, and the corresponding interpretation should be that models with (1 – RMSEA) values greater than .94 show an acceptable degree of fit.

2.4 Research Questions

The goal of the current research is to investigate the performance of RMSEA and CFI in various model and misspecification conditions. To limit the scope of the study, only CFA models are considered. Listed here again are the four questions of interest:

1. To what extent is fit index value affected by the source of the misspecification?
2. To what extent is the relationship between the degree of model misspecification and fit index value moderated by model components?
3. Does the current research support the use of uniform cutoff values across different model and misspecification types?
4. Can guidelines for the use of different indices under different models be developed?

To address the first question, RMSEA and CFI values will be evaluated in models with one of three different sources of misspecification. The three sources of misspecification covered in this study are one or more omitted error covariances, one or more omitted cross-loadings, and a misspecified latent structure. The behaviors of the indices will be compared across the different misspecification sources to determine if either index appears to be more sensitive to certain misspecifications than others.

To address the second question, various model components are manipulated within the three different misspecification types. The different model components addressed in this study include loading size, factor correlation size (in multiple-factor models), model size (as measured by the number of indicators, the total number of factors, and the ratio of indicators to factors), and model balance (as measured by how equally distributed indicators are amongst the factors in a multiple-factor model).

The effect of the size or degree of model misspecification is also of concern here and will be manipulated as well. When the source of the misspecification is due to omitted error covariances or omitted cross-loadings, the size of the omitted error covariance (or cross-loading) or the total number of omitted covariances (or cross-loadings) are used to define the size of misspecification. When the source of misspecification is due to an incorrect latent structure, the degree of misspecification is defined by the difference in the number of factors in the fitted model versus the number of factors in the population.

The third question will be addressed by examining whether models would be accepted or rejected in the scenarios presented in this study based on the most commonly used cutoff values in the research. This study is not offering to redefine appropriate cutoff values; rather, the aim is to determine if there is evidence to dissuade the use of the common cutoff values in all modeling situations.

Finally, the fourth question will be addressed by comparing the behaviors of RMSEA and CFI and determining in what cases it might be beneficial for researchers to examine both of these indices when assessing model fit. It has already been documented in previous research that these two indices behave differently under certain modeling conditions. It is one of the goals of this study to determine if any information can be gathered regarding the source of misspecification in cases where the two indices disagree about model fit.

Chapter Three: Results

In this chapter, the results of the simulation studies are presented and summarized. The chapter is broken into three sections corresponding to the three sources of misspecification examined in this study. The first section focuses on misspecification involving one or more omitted error covariances, the second focuses on misspecification involving one or more omitted cross-loadings, and the third focuses on misspecified latent structures.

In each section, different modeling scenarios are presented. These are used to examine the effects of misspecification type on index behavior with respect to various model components. These components include model size, loading size, factor correlation (when applicable), and model balance (when applicable). Plots for both CFI and RMSEA are presented for each scenario.

3.1 Misspecification Source: One or More Omitted Error Covariances

The scenarios presented in this section include both one- and two-factor CFA models. The covariance structure of a 1-factor model is given by $\boldsymbol{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\psi}$, where $\boldsymbol{\lambda}$ is a $p \times 1$ vector of factor loadings and $\boldsymbol{\psi}$ is the $p \times p$ covariance matrix of the residuals, where p represents the number of indicator variables in the model. The covariance structure of a 2-factor model is given by $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\psi}$, where $\boldsymbol{\Lambda}$ is a $p \times k$ matrix of factor loadings, $\boldsymbol{\Phi}$ is a $k \times k$ matrix of factor correlations, and $\boldsymbol{\psi}$ is the $p \times p$ covariance matrix of the residuals, where k represents the number of latent factors in the model. In each model, the number of unique parameters to be estimated, q, includes all factor loadings (λ values), all residuals (ψ values), and, in 2-factor models, the factor correlation φ.

This section focuses on misspecifications due to one or more omitted error covariances. This misspecification occurs when a researcher's proposed model fails to include the covariance of one (or more) pairs of error terms that covary in the true (population) model. In the scenarios presented here, the covariance matrices corresponding to the true model are constructed to include one or more pairs of covarying error terms (that is, $\boldsymbol{\psi}$ is not diagonal). However, the hypothesized models in this section omit these error covariances, suggesting that all residuals are uncorrelated ($\boldsymbol{\psi}$ is diagonal in all hypothesized models).

In addition to the size and number of misspecifications due to omitted error covariances, the influence of loading size, model size, factor correlation (in 2-factor models), and model balance (in 2-factor models) on index behavior are examined as well.

3.1.1 Effects of Misspecification Size

The first two figures examine the effect of the size of a single omitted error covariance on index values. Figure 3.1 corresponds to a scenario involving a hypothesized 1-factor model with one omitted error covariance. That is, the true model's covariance matrix in this scenario is constructed to represent data from a population with one error covariance. The hypothesized model omits this error covariance.

The plots of Figure 3.1 show the relationships between index value (plotted on the y-axes) and the size of the error covariance omitted from the model (plotted on the x-axes). The six colored curves correspond to six loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. All indicators in the model have the same loading size. Solid lines represent a model with eight indicator variables (p = 8); dashed lines represent a model with 16 indicator variables (p = 16). For each plot, a horizontal

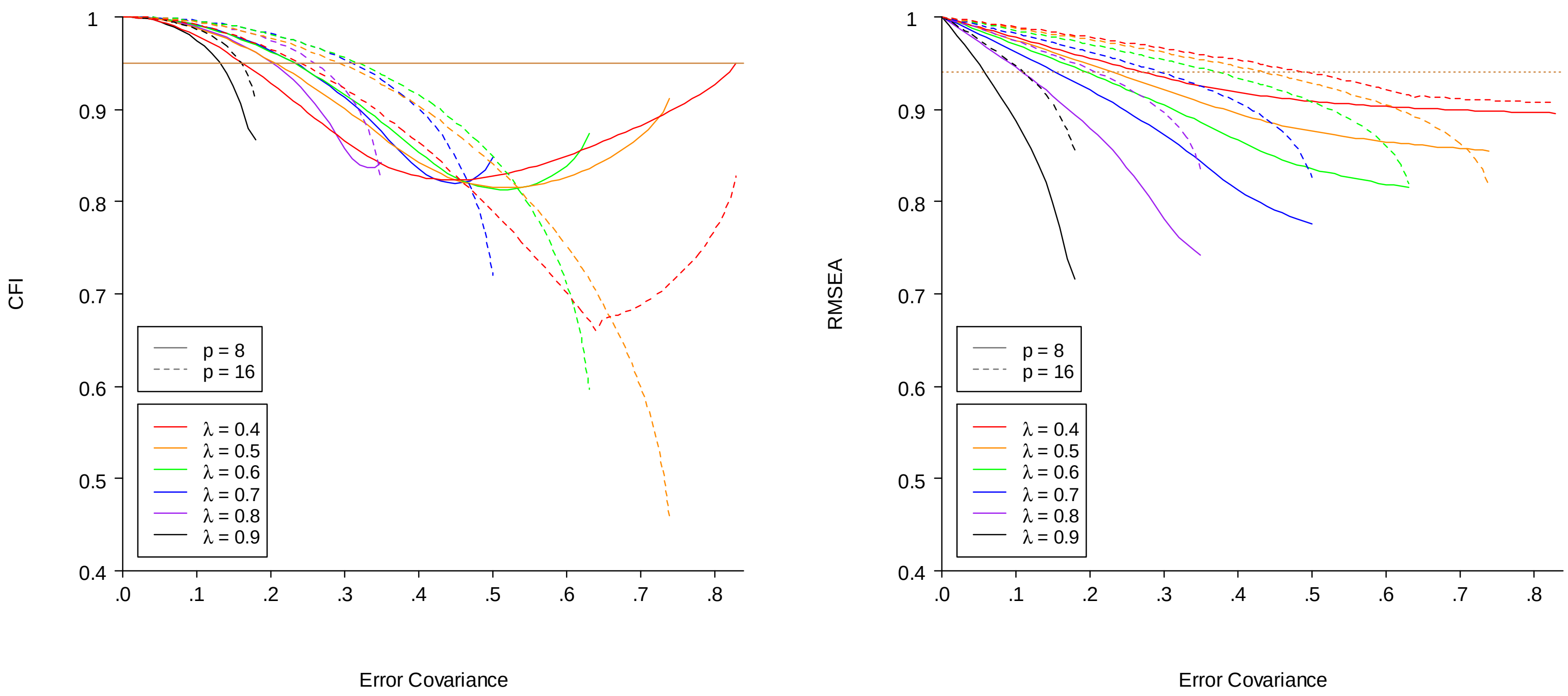

Figure 3.1: Plots of population fit index values vs. a single omitted error covariance for a 1-factor model with 8 indicators (solid lines) or 16 indicators (dashed lines). The colored curves correspond to different loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively.

brown line has been drawn to represent the commonly used cutoff value for that index (with .96 used for (1 – RMSEA)). Note that the colored curves end at difference sizes of the omitted error covariance. This is because the maximum value of the omitted error covariance is a function of the loading size. Specifically, error covariance values exceeding $1 - \lambda^2$ result in a nonpositive definite $\boldsymbol{\psi}$ matrix corresponding to the true model.

Ideally, fit indices should show worse fit as the size of the omitted error covariance increases, regardless of the size of the loadings. However, when $p = 8$, CFI shows a non-monotone relationship between index value and the size of the omitted error covariances. The index shows worse fit when the size of the omitted error covariance is moderate (between .4 to .6, depending on the loading size) and an improvement in fit for either small or large omitted error covariances. This is especially the case for lower loadings (.4, .5).

Figure 3.1 shows that RMSEA, regardless of loading size or number of indicators, behaves properly in that it shows worse fit as the size of the omitted error covariance increases. RMSEA does, however, appear to be quite sensitive to loading sizes in this situation. When loadings are higher (.8 or .9), RMSEA shows poor fit when the size of the omitted error covariance is as small as about .08. However, for loadings of .4, the omitted error covariance has to be larger than about .28 for the model to be rejected based on the traditional cutoff value.

Figure 3.2 extends the scenario presented in Figure 3.1 to a 2-factor model. The addition of a second factor allows for an additional form of misspecification. That is, an error covariance can exist between two variables loading onto the same factor or between two variables loading onto different factors. Figure 3.2 presents the same scenario as in Figure 3.1 but for a 2-factor model with eight indicators (four indicators per factor). The correlation between the two factors is held at .4. As in Figure 3.1, the six colored curves correspond to six loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. Here, the type of line represents where the omitted error covariance is located. Solid lines correspond to a model that omits an error covariance between variables loading onto the same factor. Dashed lines correspond to a model that omits an error covariance between variables loading onto different factors.

For both CFI and RMSEA, the curves in Figure 3.2 are similar in shape to the curves in Figure 3.1, suggesting that the addition of a second factor does not greatly affect the pattern of the relationships between index values and the size of the omitted error covariance. However, both indices generally show better fit in the 2-factor model case than in the 1-factor model case.

It is apparent that CFI is more sensitive to an omitted error covariance of any size if the covariance occurs between variables loading onto different factors (dashed lines). The index appears far less sensitive to this type of misspecification when the covariance occurs between variables loading onto the same factor (solid lines). For example, in the 8-indicator 1-factor case (solid lines in Figure 3.1), CFI values are below .95 for omitted error covariances larger than .3, regardless of the loading size. However, for the 8-indicator 2-factor model case, CFI values do not fall below .95 when the omitted error covariance occurs between variables loading onto the same factor (solid lines in Figure 2). While RMSEA, too, appears to be more sensitive to an omitted error covariance between variables loading onto different factors than one between variables loading onto the same factor, the difference in sensitivity is not as extreme as it is for CFI.

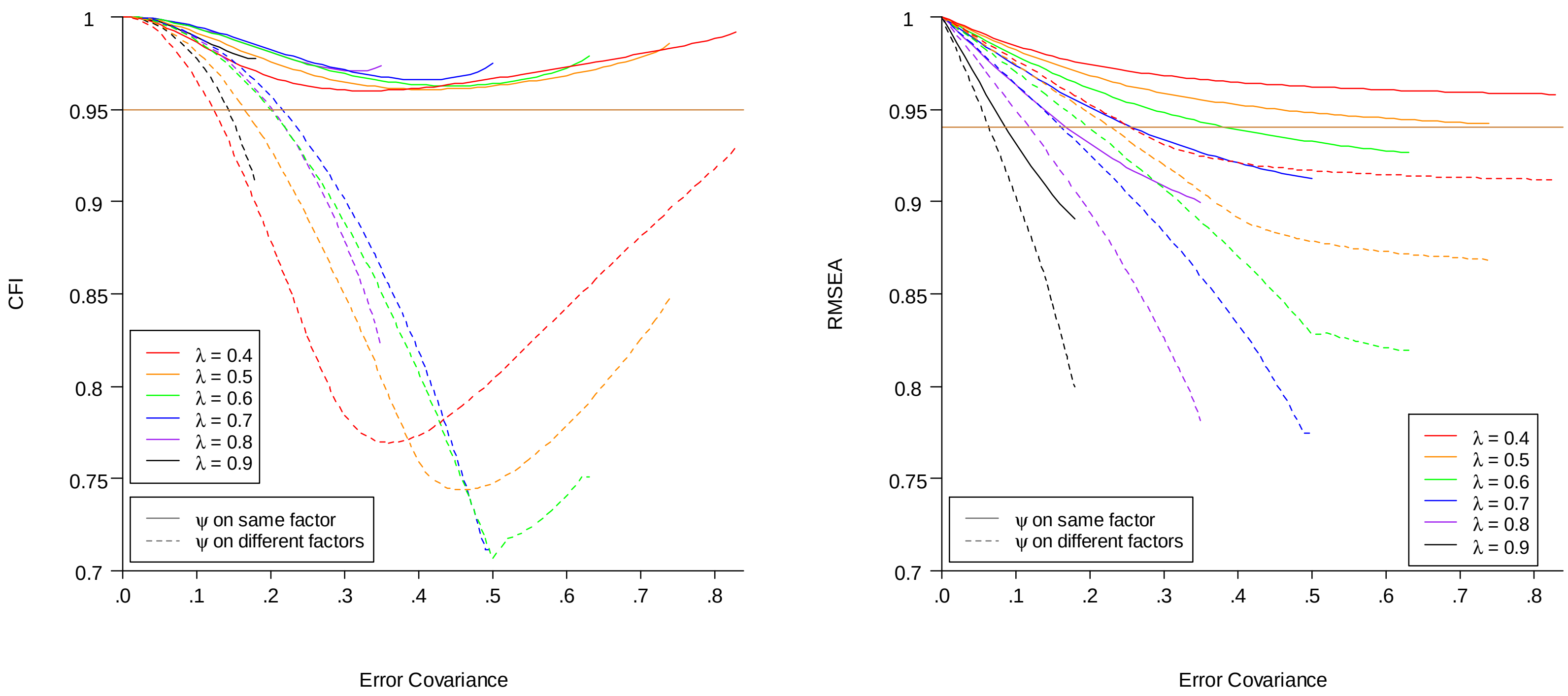

Figure 3.2: Plots of population fit index values vs. a single omitted error covariance for a 2-factor model with 8 indicators (4 per factor). The omitted error covariance occurs either between indicators of the same factor (solid lines) or indicators of different factors (dashed lines). The colored curves correspond to different loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively.

3.1.2 Effects of the Number of Misspecifications

In a model with a large number of indicator variables, it is possible that there may be more than one error covariance present. Thus, in addition to examining how a single omitted error covariance affects CFI and RMSEA, it was also of interest to examine index performance when the number of omitted error covariances was manipulated. The next two figures show the effect of an increasing number of omitted error covariances on CFI and RMSEA values.

Figure 3.3 plots fit index values against the number of omitted error covariances (1 to 10) in a 1-factor model with 20 indicator variables. Error covariances occur between variables that do not share error covariances with other variables. That is, variables 1 and 2 share an error covariance, variables 3 and 4 share an error covariance, etc. While the number of omitted error covariances is measured as a categorical variable, the neighboring points in the figure have been connected for readability. The five colored curves correspond to five loading sizes, with red, orange, green, blue, and violet corresponding to loadings of .4, .5, .6, .7, and .8, respectively. The different types of lines correspond to different sizes of the omitted error covariances. Solid, dashed, and dotted lines represent the case where all omitted error covariances are set to .05, .2, and .35, respectively.

Both CFI and RMSEA appropriately show a decrease in fit as the number of omitted error covariances—and thus, the degree of misspecification—increases, both when the omitted error covariances are of size .2 and .35. Both indices also appear to be more sensitive to loading sizes as the sizes of the omitted error covariances increase, but not as the number of omitted error covariances increase. When the omitted error covariances are all .05, neither index would reject the model as being a poor fit for the data, even when there are 10 error covariances present in the population but omitted from the model.

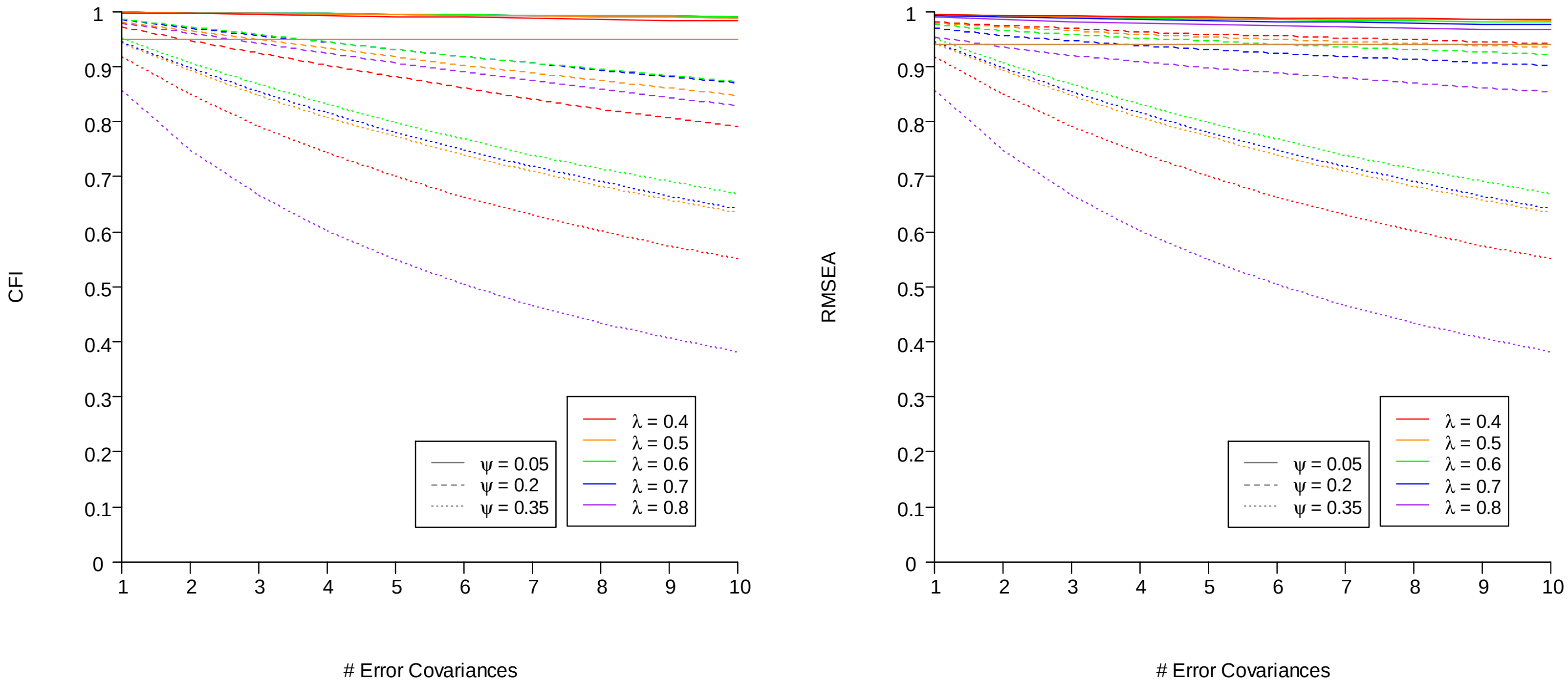

Figure 3.3: Plots of population fit index values vs. the number of omitted error covariances (1 to 10) for a 1-factor model with 20 indicators. The size of the omitted error covariances is set to .05 (solid lines), .2 (dashed lines), or .35 (dotted lines). The colored curves correspond to different loading sizes, with red, orange, green, blue, and purple corresponding to loadings of .4, .5, .6, .7, and .8, respectively. Neighboring points are connected for readability.

Figure 3.4 presents the same scenario as Figure 3.3, except for a model with an additional latent factor. That is, Figure 3.4 plots index values against an increasing number of omitted error covariances (1 to 10) for a 2-factor model with 20 indicator variables (10 per factor). As in Figure 3.3, the five colored curves here correspond to five loading sizes, with red, orange, green, blue, and violet corresponding to loadings of .4, .5, .6, .7, and .8, respectively. The size of the omitted error covariance has been fixed to .2, as this value can be considered large enough to be an omission that may be of concern in a model. The different line types in Figure 3.4 represent different factor correlation sizes, with solid, dashed, dotted lines representing factor correlations of .1, .4, and .7, respectively. Neighboring points have again been connected for readability.

From Figure 3.4, it appears that regardless of the size of the factor correlation and loading size, CFI shows poor model fit as soon as the number of omitted error covariances increases above three. Comparing CFI's performance in the 2-factor model case (Figure 3.4) with the 1-factor model case when the size of the omitted error covariances was fixed at .2 (Figure 3.3, dashed lines), the addition of one more latent factor to the model does not greatly affect CFI's ability to detect misspecifications due to an increasing number of omitted error covariances.

RMSEA, too, behaves similarly in the 2-factor case as it did in the 1-factor case where the omitted error covariances were fixed at .2. Specifically, RMSEA shows worsening fit as the number of omitted error covariances increases. In addition, Figure 3.4 also shows that RMSEA values are not affected by the size of the factor correlation for this type of misspecification, at least when the model is relatively large (10 indicators per factor).

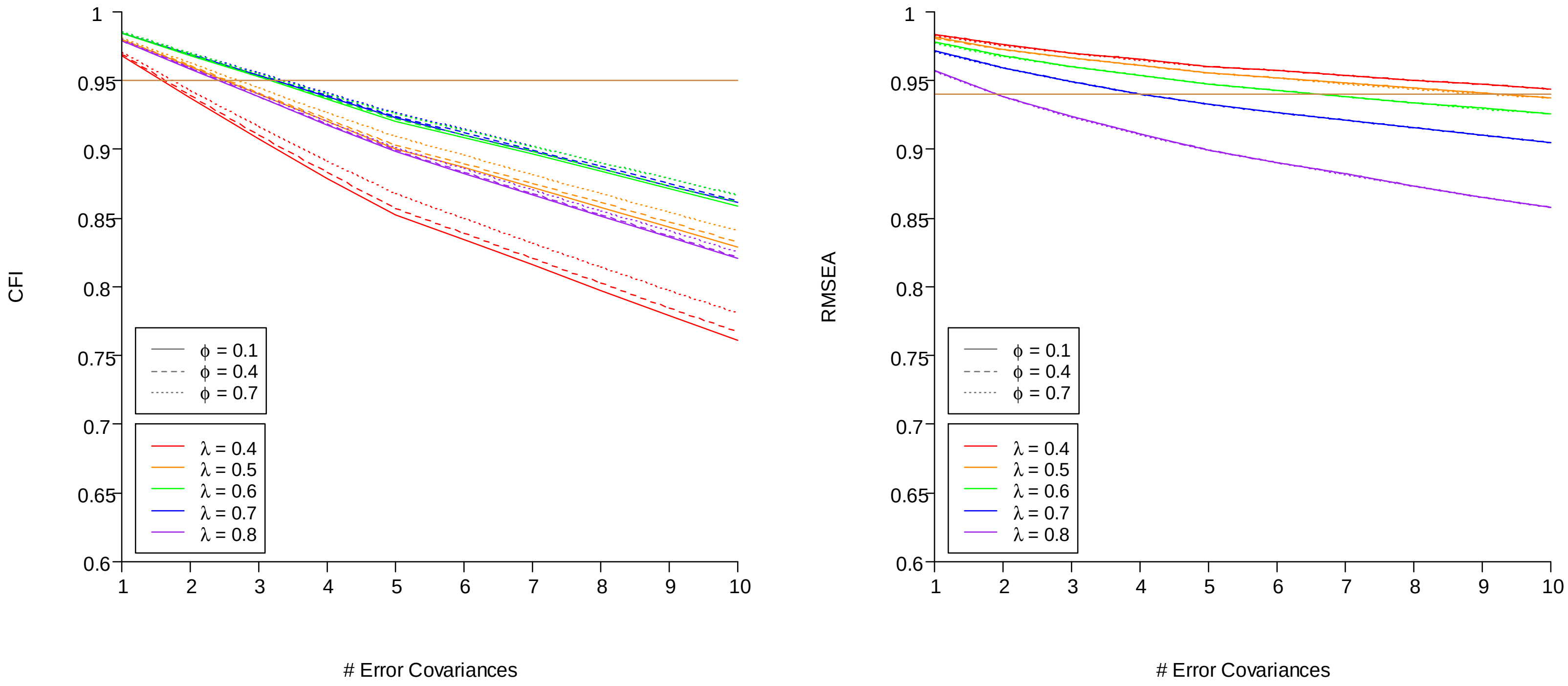

Figure 3.4: Plots of population fit index values vs. the number of omitted error covariances (1 to 10) for a 2-factor model with 20 indicators (10 per factor). The size of the omitted error covariances is fixed at .2. Factor correlation is set to .1 (solid lines), .4 (dashed lines), or .7 (dotted lines). The colored curves correspond to different loading sizes, with red, orange, green, blue, and purple corresponding to loadings of .4, .5, .6, .7, and .8, respectively. Neighboring points are connected for readability.

3.1.3 Effects of Model Size

In the previous four figures, the effects of the size of one omitted error covariance (Figures 3.1 and 3.2) and an increasing number of omitted error covariances (Figures 3.3 and 3.4) were examined both in 1-factor models and 2-factor models. Thus, these figures show how one aspect of model size, the number of latent factors, can influence index behavior when combined with misspecifications due to one or more omitted error covariances. I now wish to examine the influence of another aspect of model size, the number of observable variables, on index behavior when the misspecification is due to an omitted error covariance.

Figure 3.5 plots index values against an increasing number of indicator variables (p = 4, 6, 8, 10, 12, 14, 16, 18, 20) for a 1-factor model. The five colored curves correspond to five different sizes of omitted error covariance, with red, orange, green, blue, and purple corresponding to omitted error covariance sizes of .1, .2, .3, .4, and .5, respectively. Solid lines correspond to loadings of .4; dashed lines correspond to loadings of .7. Note that while the number of indicator variables (represented on the x-axis) is measured as a categorical variable, neighboring points are connected for readability.

Figure 3.5 shows that both CFI and RMSEA are appropriately sensitive to the size of the omitted error covariances. That is, regardless of the number of indicator variables, the indices show the worst fit for the largest omitted error covariance size (ψ = .5) and the best fit for the smallest omitted error covariance size (ψ = .1). Except when the omitted error covariance is .5, CFI appears to be more sensitive to this type of misspecification when the loadings are lower (.4, solid lines) than when they are higher (.7, dashed lines). Regardless of the size of the omitted error covariance, RMSEA is more sensitive to this type of misspecification when the loadings

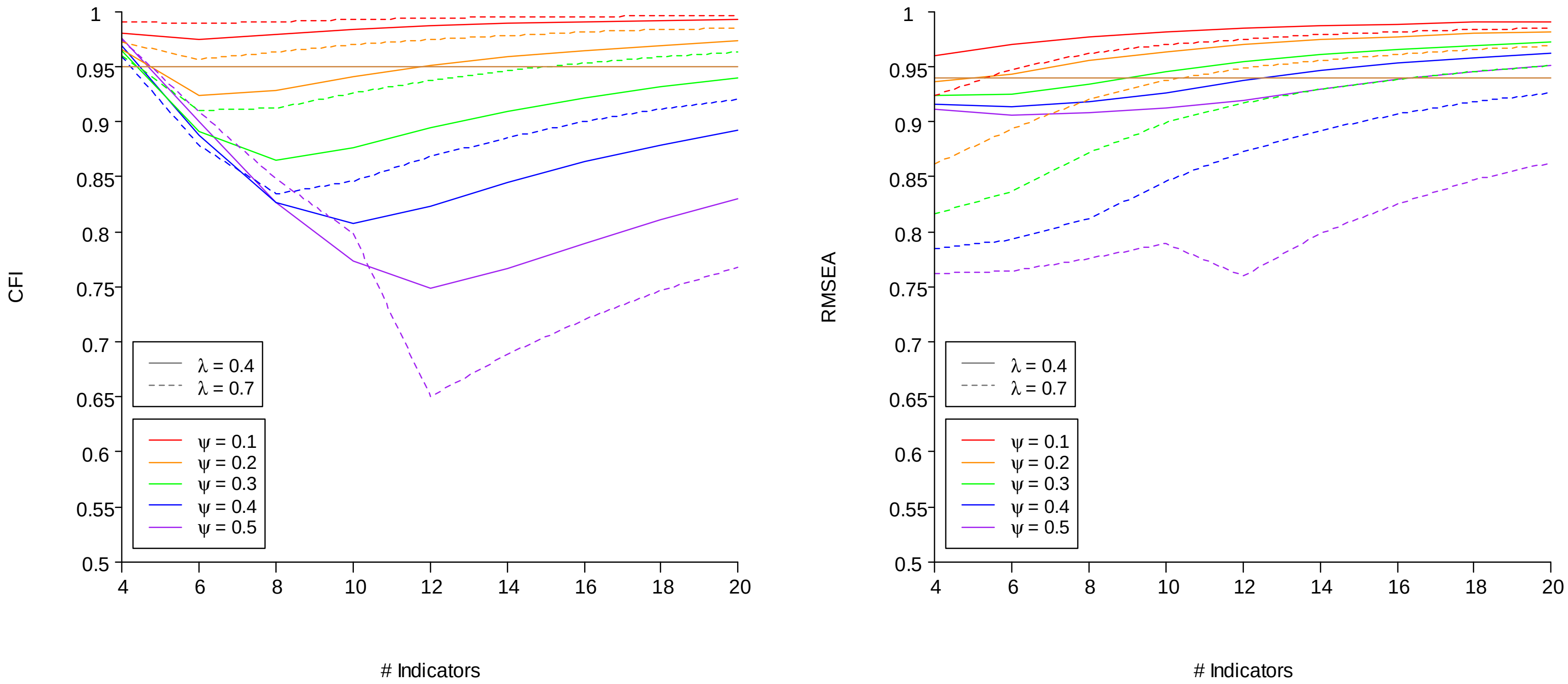

Figure 3.5: Plots of population fit index values vs. an increasing number of indicators (*p* = 4, 6, 8, 10, 12, 14, 16, 18, 20) for a 1-factor model. Loadings are .4 (solid lines) or .7 (dashed lines). The colored curves correspond to different sizes of the single omitted error covariance, with red, orange, green, blue, and purple corresponding to omitted error covariances of .1, .2, .3, .4, and .5, respectively. Neighboring points are connected for readability.

are higher (.7) than lower (.4), and again appears more sensitive to loading size overall than CFI. This is consistent with RMSEA's behavior in Figure 3.1.

With respect to the number of indicator variables, CFI appears to show a decrease in model fit as the number of indicators increase up to a certain point, then appears to show an improvement in fit as the number of indicators continues to increase. The point at which CFI changes from showing a decrease to an increase in fit is dependent upon both the size of the omitted error covariance and the size of the loadings, with smaller omitted error covariances and smaller loadings leading to an improvement in model fit for a smaller number of indicators For example, when loadings are .4, CFI starts showing an improvement of fit for $\psi = .2$ at $p = .6$; for $\psi = .5$, the improvement in fit doesn't start until $p = 12$. In contrast to the non-linear relationship between CFI and the number of indicators included in the model, RMSEA shows an improvement in fit as p increases, regardless of loading size or the size of the omitted error covariance.

Figure 3.6 extends the scenario presented in Figure 3.5 to a 2-factor model. Figure 3.6 plots index values against an increasing number of indicators (p = 4, 6, 8, 10, 12, 14, 16, 18, 20) for a 2-factor model. For the different model sizes, there are an equal number of indicators loading onto each factor (e.g., for $p = 10$, each factor has five indicators loading onto it). The five colored curves correspond to five different sizes of omitted error covariance, with red, orange, green, blue, and purple corresponding to omitted error covariance sizes of .1, .2, .3, .4, and .5, respectively. Solid lines correspond to loadings of .4; dashed lines correspond to loadings of .7. Neighboring points are connected for readability.

Compared to the 1-factor model scenario examined in Figure 3.5, there is higher model complexity in Figure 3.6 due to the addition of a second factor in the model. This added complexity affects index behavior when compared to Figure 3.5. While there still exists a non-linear relationship between CFI values and the number of indicator variables for a 2-factor model, the relationship is much less dramatic. The inflection points at which CFI values cease to decrease and begin to increase appear at larger values of p in Figure 3.6 than they do in Figure 3.5, and the non-linear relationship actually disappears for $\psi = .5$.

The addition of a second latent factor also appears to reduce but not eliminate RMSEA's tendency to show an improvement of fit as the number of indicator variables increases. RMSEA shows good fit regardless of the size of the model, except for cases where a higher loading size (.7) is combined with a larger omitted error covariance size (.3 to .5).

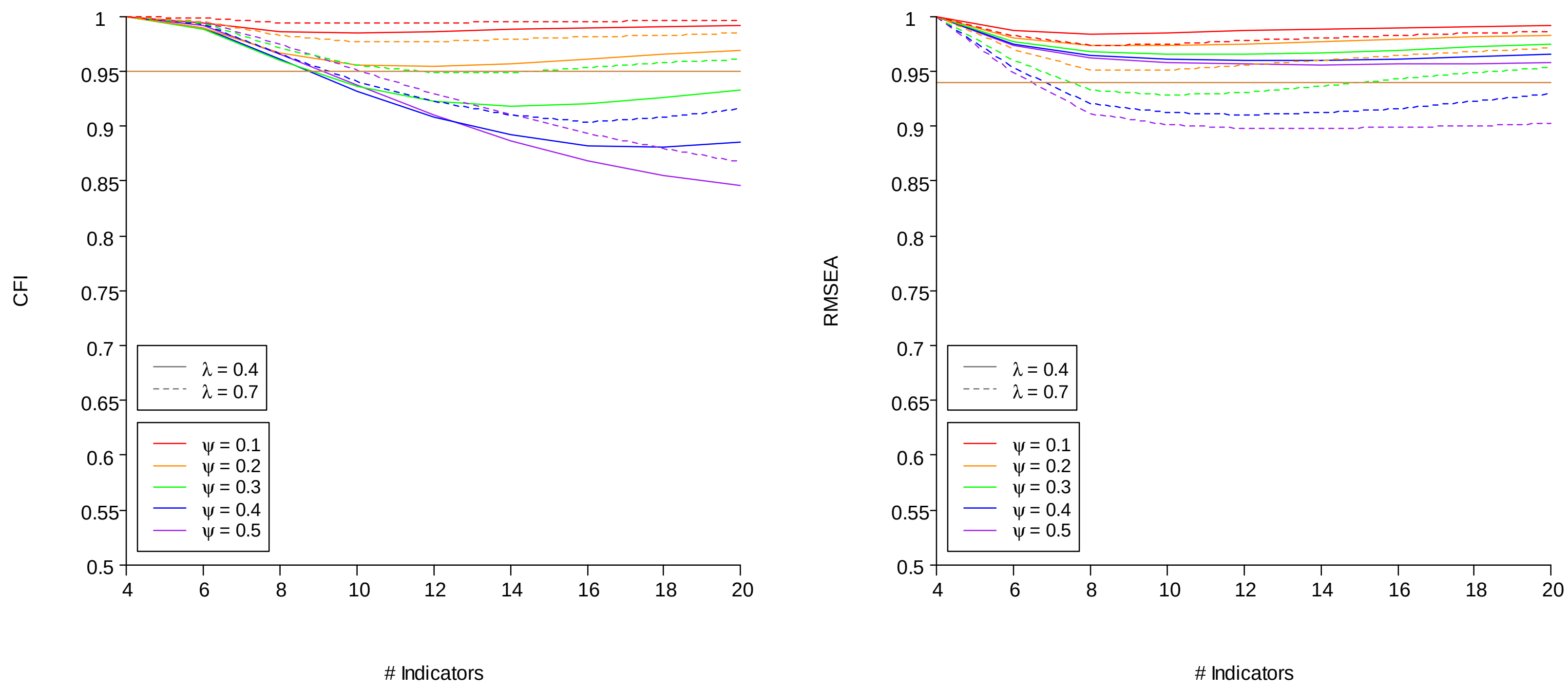

Figure 3.6: Plots of population fit index values vs. an increasing number of indicators (*p* = 4, 6, 8, 10, 12, 14, 16, 18, 20) for a 2-factor model. Loadings are .4 (solid lines) or .7 (dashed lines). The colored curves correspond to different sizes of the single omitted error covariance, with red, orange, green, blue, and purple corresponding to omitted error covariances of .1, .2, .3, .4, and .5, respectively. Neighboring points are connected for readability.

3.1.4 Effects of Factor Correlation

In the case of a model with two or more latent factors, the size of the factor correlation(s) may also affect fit index behavior in certain misspecification scenarios. Figure 3.7 plots index values against an increasing factor correlation size (0 to 1) for a 2-factor model with eight indicator variables (four per factor). The five colored curves correspond to five different sizes for the single omitted error covariance, with red, orange, green, blue, and purple corresponding to omitted error covariance sizes of .1, .2, .3, .4, and .5. Solid lines correspond to loadings of .4, dashed lines correspond to loadings of .7.

As is seen in Figure 3.7, both CFI and RMSEA show good model fit when the factor correlation is near zero, but then gradually show a decrease in fit as the factor correlation increases to one. This suggests, in addition to the results from Figures 3.1 and 3.2, that both indices are more sensitive to misspecification due to an omitted error covariance in a 1-factor model versus a 2-factor model. As seen in previous plots (e.g., Figure 3.5), CFI again appears to be more sensitive to the misspecification when loadings are low (.4) rather than high, while RMSEA appears more sensitive to the misspecification when loadings are high rather than low. RMSEA again seems more sensitive to loading size in general than CFI, as the differences between the $\lambda = .4$ case (solid lines) and the $\lambda = .7$ case (dashed lines) appear more dramatic for RMSEA than they do for CFI.

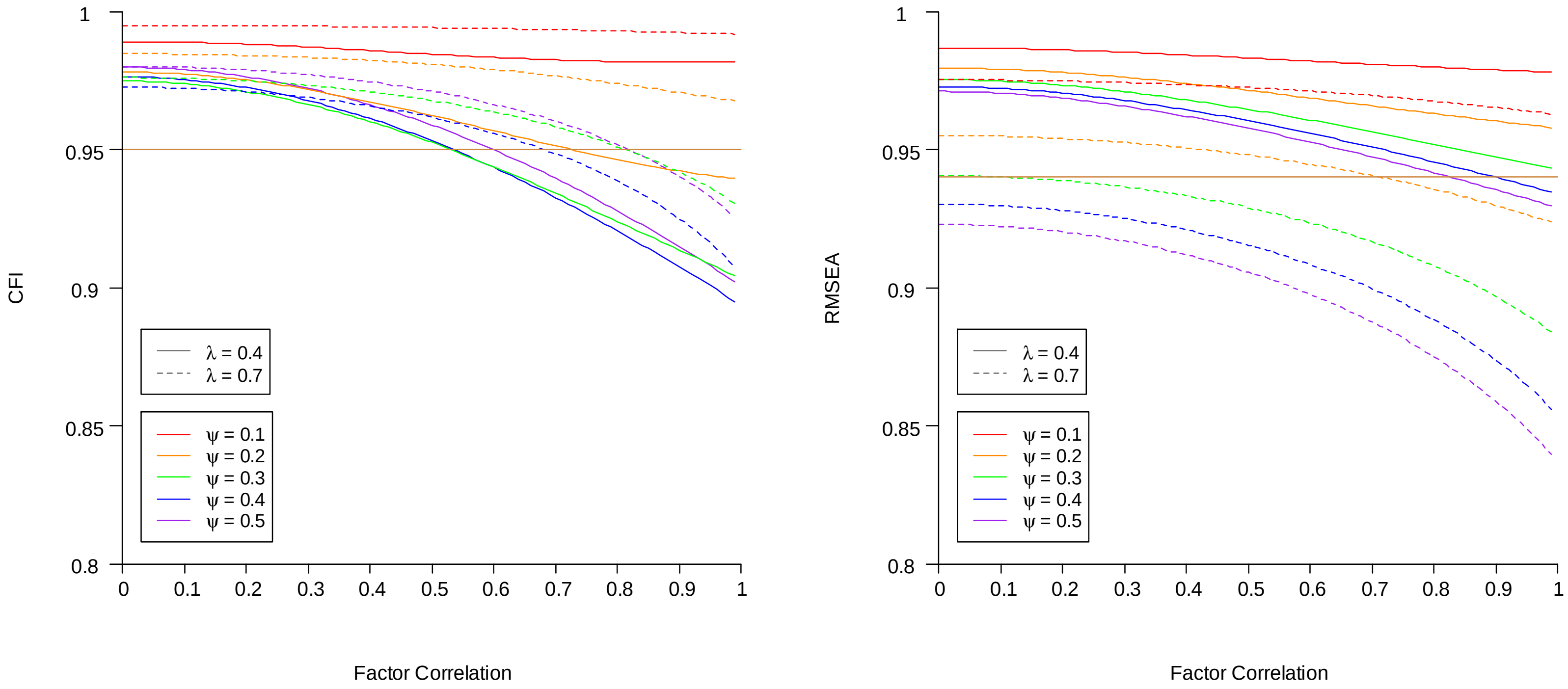

Figure 3.7: Plots of population fit index values vs. an increasingly large factor correlation (0 to 1) for a 2-factor model with 8 indicators. Loadings are .4 (solid lines) or .7 (dashed lines). The colored curves correspond to different sizes of the single omitted error covariance, with red, orange, green, blue, and purple corresponding to omitted error covariances of .1, .2, .3, .4, and .5, respectively.

3.1.5 Effects of Model Imbalance

An additional feature unique to multi-factor models is that of model balance, or how the indicator variables are distributed amongst the latent variables. While it is often the case that hypothesized models are “balanced,” meaning that there are an equal number of indicator variables loading onto the different factors (e.g, a 2-factor model with six indicators total would be balanced if each factor had three indicator variables loading exclusively onto it), there exist hypothesized models in which certain factors have more indictors loading onto them than others. However, previous studies in the literature that examine index behavior involve balanced models almost exclusively. In rare cases when an imbalanced model is used, it is usually the case that only one additional indicator loads onto one factor while the second factor has only one less indicator. In such cases, the imbalance is noted, but its effects are never discussed.

Here, I seek to explore the effect of model balance/imbalance more thoroughly. Figure 3.8 plots index values against the size of a single omitted error covariance for a 2-factor model with 24 indicator variables. Factor correlation is set to .4 and loadings are set to .4. The six colored curves correspond to six different degrees of model imbalance. Red corresponds to a balanced model (12 indicators per factor). Orange corresponds to a model with 11 and 13 indicators per factor, green corresponds to a model with 10 and 14 indicators per factor, blue corresponds to a model with 9 and 15 indicators per factor, purple corresponds to a model with 8 and 16 indicators per factor, and black corresponds to a model with 7 and 17 indicators per factor. Solid lines correspond to case where the omitted error covariance occurs between variables loading onto the “larger” factor; dashed lines correspond to the case where the omitted error covariance occurs between variables loading onto the “smaller” factor. It is worth clarifying that the imbalance is modeled correctly. That is, the population (true) models for this figure have an unequal number of indicators per factor, but the hypothesized models accurately represent this imbalance. Thus, the only source of misspecification is the omitted error covariance.

Figure 3.8 reveals that CFI values are highly affected by model balance, particularly when the size of the omitted error covariance is moderate to large. For all levels of balance, CFI shows good model fit until the omitted error covariance increases above about .25. However, once the size of the misspecification increases, CFI appears much more sensitive to its effects when the misspecification occurs between variables loading onto the “larger” factor (solid lines). For example, in the most severely imbalanced case, where one factor has seven indicators loading onto it and the other has 17, when the omitted error covariance is .7, CFI is about .66 when the misspecification is on the “larger” factor and about .94 when the misspecification is on the “smaller” factor. It is also worth noting that we again see the non-linear relationship between CFI values and the size of the omitted error covariance (as seen in Figure 3.1), regardless of the balance/imbalance of the model. That is, CFI shows better fit for smaller and larger omitted error covariance sizes and worse fit for moderate omitted error covariance sizes. However, unless the model is severely imbalanced and the misspecification occurs on the “small” factor (black dashed line), CFI values do not increase enough to lead a researcher to accept a model based on the commonly used cutoff criterion.

While RMSEA shows a linear decrease in fit for all levels of model imbalance as the size of the omitted error covariance increases, the index only shows poor fit for the most severely imbalanced model when the misspecification occurs on the “larger” factor (solid black line). This suggests that RMSEA is not as affected by model imbalance as CFI. However, as was seen in Figures 3.5 and 3.6, RMSEA shows an improvement in fit as the number of indicators in a model increases. Thus, since the model is quite large in this scenario (p = 24), it was suspected that

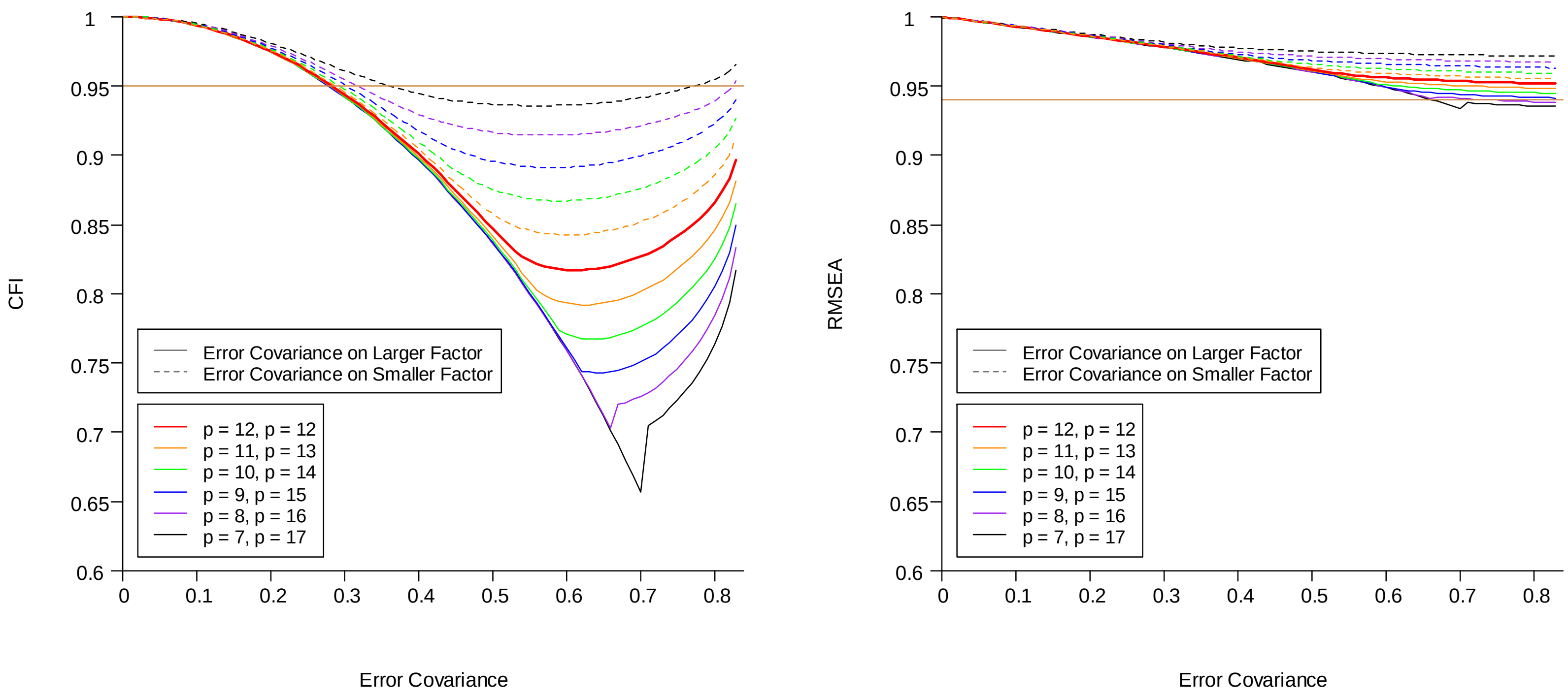

Figure 3.8: Plots of population fit index values vs. a single omitted error covariance for a 2-factor model with 24 indicators. Factor correlation is .4 and loadings are .4. The colored curves correspond to differently balanced models, with red corresponding to the balanced model and black corresponding to the most imbalanced model. Solid lines correspond to the case where the error covariance is omitted within the larger factor; dashed lines correspond to the case where the error covariance is omitted within the smaller factor.

RMSEA's behavior may be due in part to the fact that the index shows better fit for larger models in general. Thus, an additional model (not presented here), identical to Figure 3.8 except having only eight indicators instead of 24, was created to determine whether RMSEA's behavior was due solely to the size of the model in Figure 3.8. In the smaller model, RMSEA also appeared to be unaffected by model balance, which suggests that the index's behavior in Figure 3.8 is due to a general insensitivity to model balance than to the size of the model itself.

3.2 Misspecification Source: One or More Omitted Cross-Loadings

This section focuses on misspecifications due to one or more omitted cross-loadings. Recall that an indicator variable can, in some situations, load onto more than one factor in a given model. This is especially likely if the two factors are highly correlated. The misspecifications examined in this section occur when a researcher's proposed model fails to include one (or more) non-zero cross-loadings that exist in the true (population) model.

The scenarios presented here focus on 2-factor CFA models, as at least two factors are required for a non-zero cross-loading to exist. As described in the first section of this chapter, the covariance structure of a two-factor model is given by $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\psi}$, where $\boldsymbol{\Lambda}$ is a $p \times k$ matrix of factor loadings, $\boldsymbol{\Phi}$ is a $k \times k$ matrix of factor correlations, and $\boldsymbol{\psi}$ is the $p \times p$ covariance matrix of the residuals, where p represents the number of indicator variables and k represents the number of latent factors.

For these misspecifications, the covariance matrices corresponding to the true models are constructed to include one or more cross-loadings. This is reflected in the structure of $\boldsymbol{\Lambda}$. For example, suppose a 2-factor model with eight indicator variables has four indicators loading exclusively onto the first factor, three indicators loading exclusively onto the second factor, and one indicator loading both on the first and second factor. The structure of the loadings matrix $\mathbf{\Lambda}$ may be as follows:

$$\mathbf{\Lambda} = \begin{bmatrix} * & 0 \\ * & 0 \\ * & 0 \\ * & 0 \\ 0 & * \\ 0 & * \\ 0 & * \\ * & * \end{bmatrix}$$

where * represents a non-zero value of the loading for that indicator onto its corresponding factor. The hypothesized models in this section omit the cross-loadings that exist in the population. For example, if the $\mathbf{\Lambda}$ above represented the structure of the loadings matrix in the population, the hypothesized $\mathbf{\Lambda}$ might look as follows:

$$\mathbf{\Lambda} = \begin{bmatrix} * & 0 \\ * & 0 \\ * & 0 \\ * & 0 \\ 0 & * \\ 0 & * \\ 0 & * \\ 0 & * \end{bmatrix}$$

Note that the hypothesized $\mathbf{\Lambda}$ has the last indicator variable loading only onto the second factor, whereas in the population, it loads onto both. In addition to the size and number of misspecifications due to omitted cross-loadings, the influence of loading size, model size, factor correlation, and model balance on index behavior are examined as well.

3.2.1 Effects of Misspecification Size

As in the first section of this chapter, I begin examining the effect of a single misspecification—in this case, a single omitted error covariance—on index behavior. Figure 3.9 plots index values (y-axes) against an increasingly large omitted cross-loading (x-axes) for a 2-factor model. The four colored curves correspond to four loading sizes (apart from the size of the cross-loading). Red, orange, green, and blue corresponding to loadings of .4, .5, .6, and .7. Solid lines correspond to a 2-factor model with eight indicator variables (four per factor), and dashed lines correspond to a 2-factor model with 16 indicator variables (eight per factor). The factor correlation is set to .1.

As in the case with a single increasingly large omitted error covariance, we again see a non-monotone relationship between CFI values and the size of the omitted cross-loading. When $p = 8$, CFI values indicate worse fit as the size of the omitted cross-loading increases. However, at a certain size of the omitted cross-loading (depending on the size of the other loadings), CFI starts showing an improvement in fit. It should be noted, though, that while CFI values begin to increase again as the size of the omitted cross-loading grows larger, CFI never increases above the commonly used cutoff value of .95 for any combination of model size or loading size. Thus, there would be no danger in this scenario of a researcher accepting a model as having good fit (based on the cutoff criterion) once the size of the omitted cross-loading was greater than about .42.

RMSEA shows a steady decrease in model fit as the size of the omitted cross-loading increases. This is true for all loading sizes and both model sizes. However, when loadings are low ($\lambda = .4$), RMSEA values fail to indicate poor fit even when the omitted cross-loading is as big as .7, nearly double the size of the rest of the loadings. Both indices are less sensitive to the

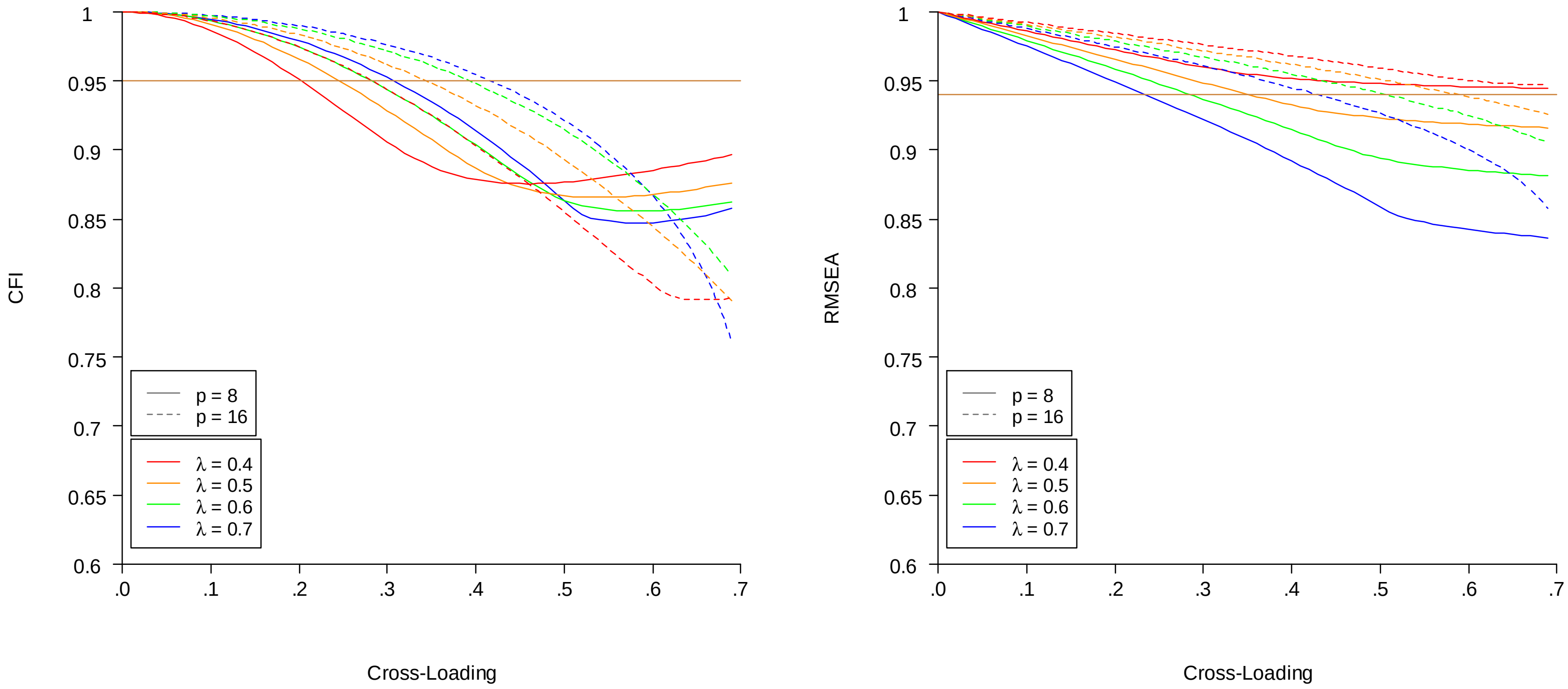

Figure 3.9: Plots of population fit index values vs. a single omitted cross-loading for a 2-factor model with 8 indicators (solid lines) or 16 indicators (dashed lines). The colored curves correspond to different loading sizes, with red, orange, green, and blue corresponding to loadings of .4, .5, .6, and .7, respectively.

effects of an omitted cross-loading when the number of indicators is increased from p = 8 to p = 16. This result makes sense, as the effect of a single omitted cross-loading is likely “diluted” by the larger number of loadings in the larger model.

3.2.2 Effects of the Number of Misspecifications

Now I examine the effects of an increasingly large number of cross-loadings omitted by the hypothesized model. Figure 3.10 plots index values against the number of omitted cross-loadings (1 to 10) in a 2-factor model with 20 indicator variables (10 per factor). The four colored curves correspond to four different loading sizes, with red, orange, green, and blue corresponding to loadings of .3, .4, .5, and .6. All omitted cross-loadings are set to λ = .3. Solid lines correspond to a factor correlation of .1; dashed lines correspond to a factor correlation of .3.

For both CFI and RMSEA, we actually see an increase in fit as the number of omitted cross-loadings increases. CFI shows poorest fit when the number of omitted cross-loadings is two, three, or four (depending on the loading size), but then shows better fit and finally nearly perfect fit as the number of omitted cross-loadings increases to 10.

CFI is more sensitive to the number of omitted cross-loadings when the factor correlation is .3 than when it is .1. The same is true for RMSEA, but the effect of the factor correlation is less present for RMSEA than for CFI.

RMSEA appears even less sensitive to omitted cross-loadings than CFI. RMSEA values are smallest when there are four omitted cross-loadings and show nearly perfect fit when the number of omitted cross-loadings is 10. However, at no point are RMSEA values smaller than the cutoff value of .94. Thus, regardless of the number of omitted cross-loadings in the hypothesized model, RMSEA shows this model as having good fit when judged by the cutoff

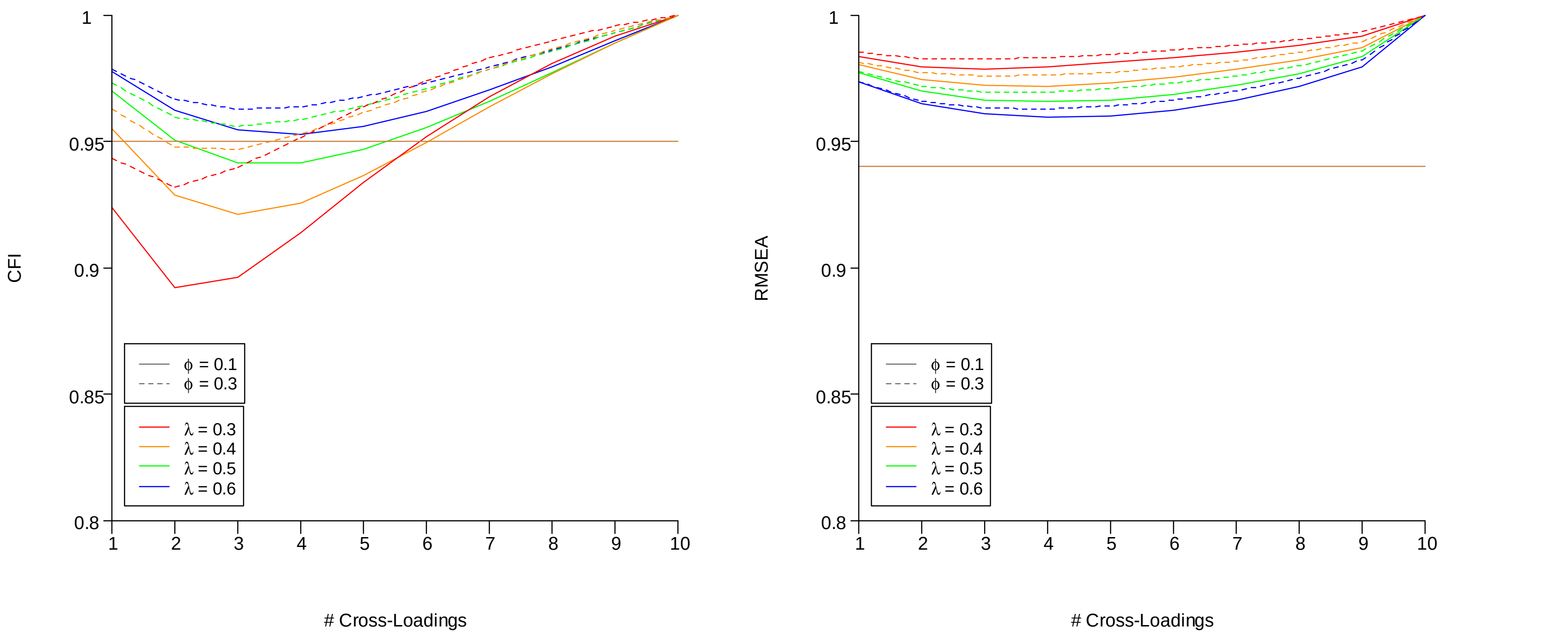

Figure 3.10: Plots of population fit index values vs. the number of omitted cross-loadings (1 to 10) for a 2-factor model with 20 indicators (10 per factor). The size of the omitted cross-loadings is fixed at .3. Factor correlation is set to .1 (solid lines) or .3 (dashed lines). The colored curves correspond to different loading sizes (except for the omitted cross-loading), with red, orange, green, and blue corresponding to loadings of 0.3, .4, .5, and .6, respectively. Neighboring points are connected for readability.

criterion. While these results may suggest that RMSEA is insensitive to misspecifications due to one or multiple omitted cross-loadings, recall that the model in this scenario has 20 indicator variables ($p = 20$). As we have seen for larger models (e.g., in Figures 3.5 and 3.6), the sensitivity of RMSEA to any degree of misspecification has been substantially less than CFI.

3.2.3 Effects of Model Size

After studying the effects of the size and number of omitted cross-loadings on index value, I next examine how the number of indicator variables in a model might affect indices' behavior when misspecification is due to an omitted cross-loading. In Figure 3.11, index values are plotted against an increasing number of indicator variables (p = 4, 6, 8, 10, 12, 14, 16, 18, and 20) for a 2-factor model. In the hypothesized model, each factor has an equal number of indicator variables exclusively loading onto it (e.g., for $p = 8$, each factor has exactly four indicators loading onto it). In the true (population) model, however, one of the indicators loads onto both factors. The size of this cross-loading, omitted in the hypothesized model, is represented by the six colored curves. Red, orange, green, blue, purple, and black correspond to the size of the omitted cross-loading being set to .1, .2, .3, .4, .5, or .6, respectively. The size of the other loadings is set to either .3 (represented by the solid lines) or .7 (represented by the dashed lines). The factor correlation is .1. Neighboring points are connected for readability.

Both CFI and RMSEA appropriately show worse fit for larger omitted cross-loadings, regardless of the number of indicators or the size of the other loadings (.3 or .7). For all cross-loading sizes, CFI shows an initial decrease in model fit as the number of indicators increases, but then begins to show an increase in fit as the number of indicators continues to grow. The inflection point at which CFI changes from decreasing to increasing appears to depend on both

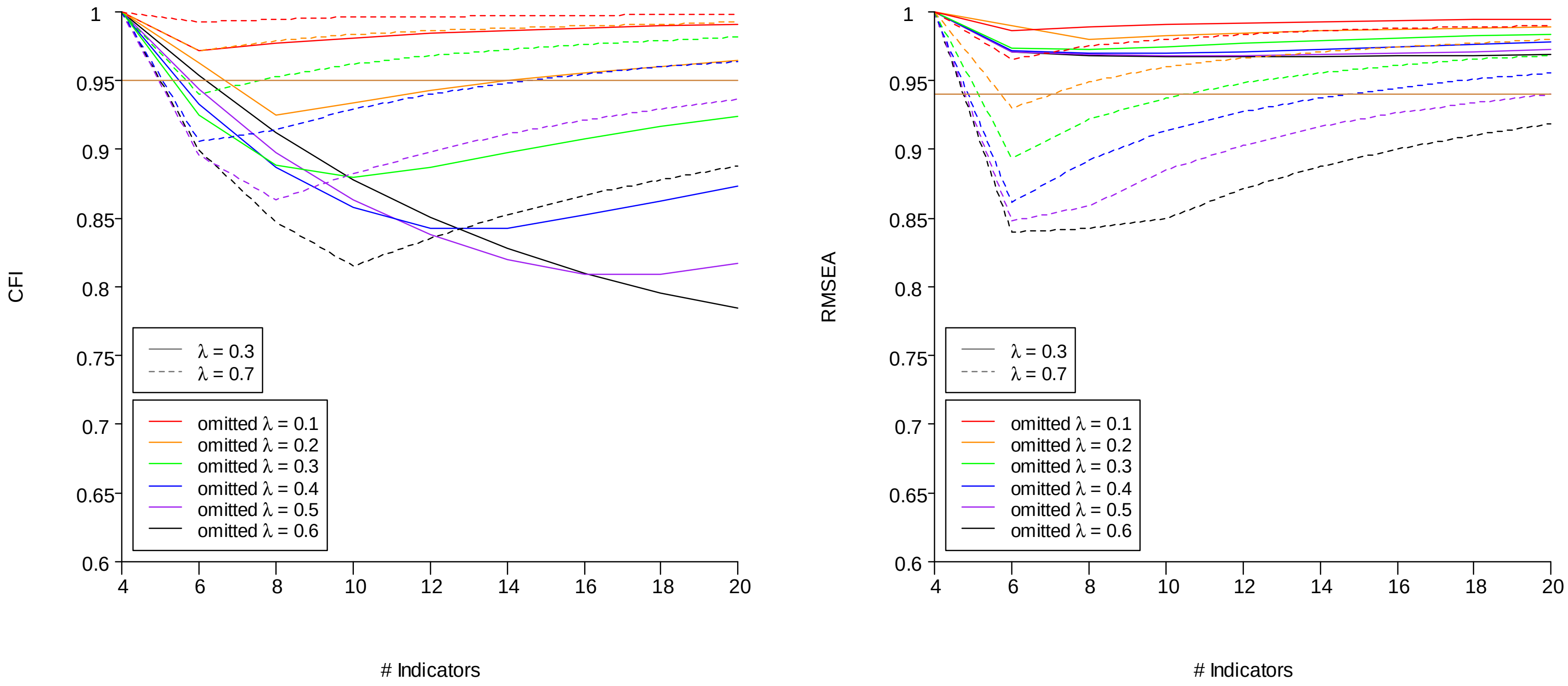

Figure 3.11: Plots of population fit index values vs. an increasing number of indicators (*p* = 4, 6, 8, 10, 12, 14, 16, 18, 20) for a 2-factor model. Loadings are .3 (solid lines) or .7 (dashed lines). The colored curves correspond to different sizes of the single omitted cross-loading, with red, orange, green, blue, purple, and black corresponding to omitted cross-loadings of .1, .2, .3, .4, .5, and .6, respectively. Neighboring points are connected for readability.

the size of the omitted cross-loading and on the size of the other loadings in the model. This pattern is less evident when the other loadings are high (.7) and the omitted cross-loading is high as well (.5, .6). In general, this non-monotone relationship between CFI and model size would not affect a researcher's decision about a model's fit based on the commonly used cutoff value (e.g., a model that omits a cross-loading of size $\lambda = .6$ will always be rejected for poor fit, except in the smallest models). However, in a few cases in this scenario, a model of moderate size with a certain omitted cross-loading size might be rejected as having poor fit, while a model with the same size omitted cross-loading but with a larger number of indicators might be accepted as having good fit based on the cutoff criterion.

This same pattern is evident for RMSEA, though much less so when the loadings in the model are set at .3 rather than .7. When the loadings are .3 (solid lines), RMSEA values fail to fall below .94 and thus show good fit regardless of the number of indicators or the size of the omitted cross-loading. When the loadings are .7 (dashed lines), RMSEA shows worse fit when $p = 6$ (for all omitted cross-loading sizes), but then steadily shows an improvement in fit as the number of indicator variables increases. When the size of the omitted cross-loading is .5 or .6, RMSEA shows poor fit for all model sizes, apart from when $p = 4$.

3.2.4 Effects of Factor Correlation

It is worth acknowledging the role the size of the correlation between factors may play in determining the severity of a misspecification due to an omitted cross-loading. An omitted cross-loading between highly correlated factors might be considered more of a severe misspecification than an omitted cross-loading between slightly correlated factors, due to the fact that if an indicator variable loads highly onto one factor, it is likely it would also load highly onto another highly correlated factor. However, the less correlated two factors are, the more "obvious" an omitted cross-loading might be in a model. This may be because an indicator variable that legitimately loads onto two quite different (i.e., slightly correlated) factors may be an important and unique feature of a population. If this feature is not included in a hypothesized model, may render that hypothesized model a poor fit to the unique structure underlying the population.

Figure 3.12 plots index values against an increasing factor correlation (0 to 1) for a 2-factor model. The six colored curves correspond to six different sizes of the omitted cross-loading, with red, orange, green, blue, purple, and black corresponding to omitted cross-loading sizes of .1, .2, .3, .4, .5, and .6, respectively. The remaining loadings are set at .4. Solid lines correspond to a model with eight indicators; dashed lines correspond to a model with 16 indicators.

From Figure 3.12, it is clear that both CFI and RMSEA show an improvement in fit as the factor correlation increases from 0 to 1. CFI appears more sensitive to larger omitted cross-loadings (blue, purple, and black dashed lines) when the model is larger but more sensitive to smaller cross-loadings (red and orange solid lines) when the model is smaller. However, this switch in sensitivity is slight and is likely of no practical concern. For RMSEA, the index is consistently more sensitive to omitted cross-loadings when the model is smaller (solid lines) than larger (dashed lines), but this difference is also very slight.

Again, we see with RMSEA that index values never fall below the cutoff value. However, this pattern occurs both when $p = 8$ (solid lines) and when $p = 16$ (dashed lines). Since a 2-factor model with eight indicator variables might be considered small to moderately-sized, this suggests that RMSEA's lack of sensitivity toward omitted cross-loadings may not be due to

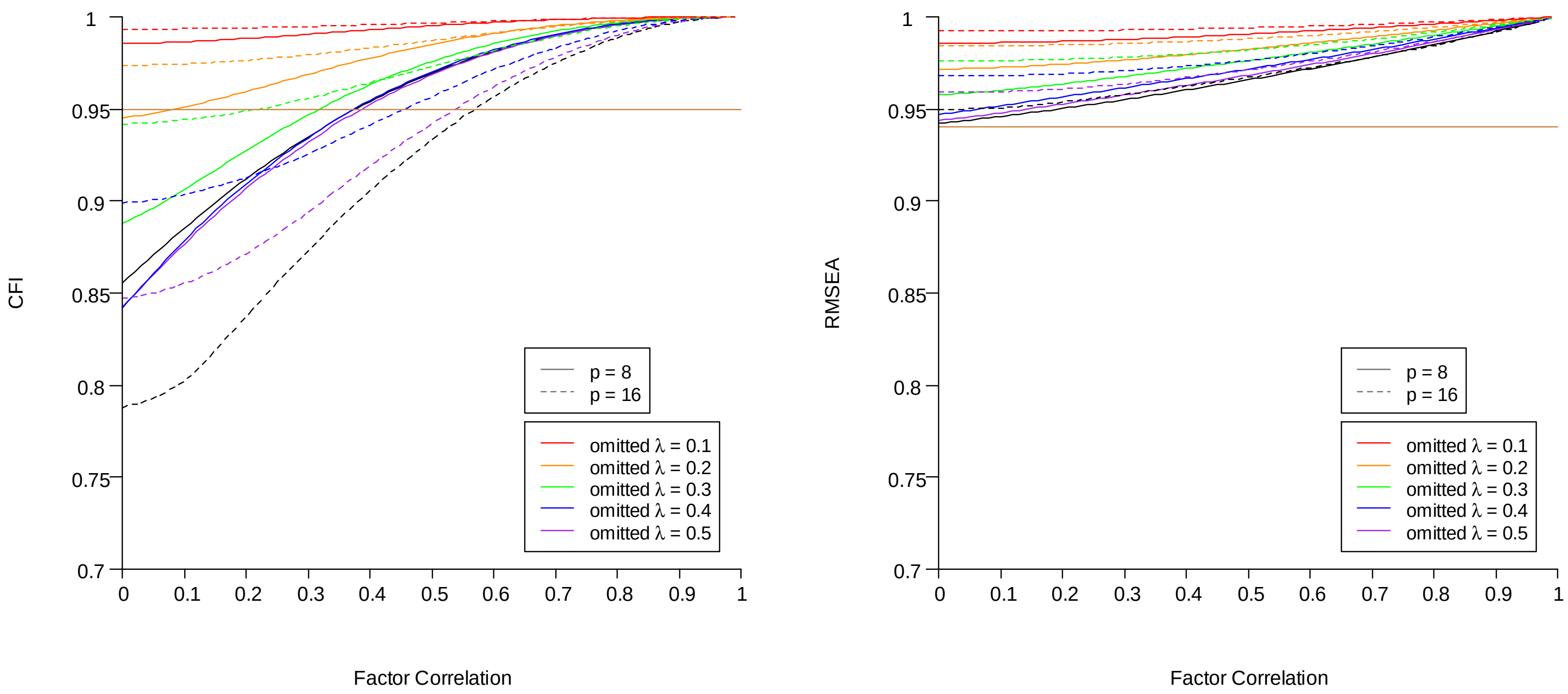

Figure 3.12: Plots of population fit index values vs. an increasingly large factor correlation (0 to 1) for a 2-factor model with 8 indicators (solid lines) or 16 indicators (dashed lines). Loadings are .4. The colored curves correspond to different sizes of the single omitted cross-loading, with red, orange, green, blue, and purple corresponding to omitted cross-loadings of .1, .2, .3, .4, and .5, respectively.

an interaction with model size, but instead might simply indicate that RMSEA is not very sensitive to this type of misspecification.

3.2.5 Effects of Model Imbalance

Finally, I examine the effects of model balance on index behavior when the misspecification is due to an omitted cross-loading. Figure 3.13 plots index values against an increasingly large omitted cross-loading for a 2-factor model with a total of 24 indicator variables. Loadings (apart from the cross-loading) are set to .4 and the factor correlation is set to .1. The six colored curves correspond to six different degrees of model imbalance. Red corresponds to a balanced model (12 indicators per factor). Orange corresponds to a model with 11 and 13 indicators per factor, green corresponds to a model with 10 and 14 indicators per factor, blue corresponds to a model with 9 and 15 indicators per factor, purple corresponds to a model with 8 and 16 indicators per factor, and black corresponds to a model with 7 and 17 indicators per factor.

In the population, one of the 24 indicator variables loads onto both factors (regardless of the balance/imbalance of the model). There are two different ways that the hypothesized model can omit a cross-loading in this scenario. The hypothesized model can either fail to include the cross-loading onto the smaller factor or can fail to include the cross-loading onto the larger factor. For example, in the most severely imbalance model scenario, the larger factor has 17 indicators and the smaller factor has seven. The black curves in Figure 3.13 represent these cases, with solid lines corresponding to the cross-loading being omitted from the larger factor; dashed lines correspond to the cross-loading being omitted from the smaller factor.

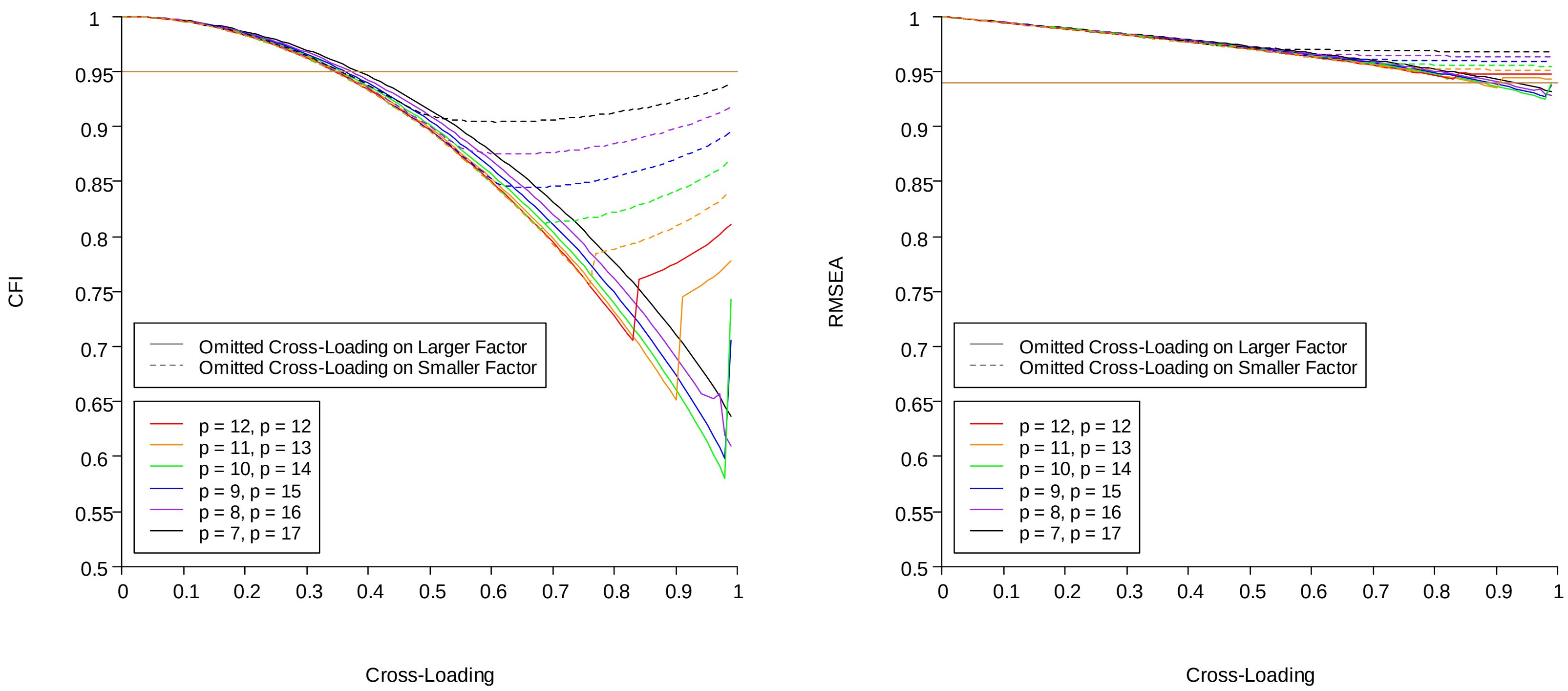

Figure 3.13: Plots of population fit index values vs. a single omitted cross-loading for a 2-factor model with 24 indicators. Factor correlation is .1 and loadings are .4. The colored curves correspond to differently balanced models, with red corresponding to the balanced model and black corresponding to the most imbalanced model. Solid lines correspond to the case where the cross-loading is omitted from the larger factor; dashed lines correspond to the case where the cross-loading is omitted from the smaller factor.

As was the case for Figure 3.8, it is worth clarifying here that the imbalance is modeled correctly. The population (true) models for this figure have an unequal number of indicators per factor, but the hypothesized models accurately model this imbalance. Thus, the only source of misspecification is the omitted cross-loading.

Appropriately, both CFI and RMSEA show a decrease in fit as the size of the omitted cross-loading increases, regardless of the balance of the model or where the omitted cross-loading is located. For CFI, both model balance and cross-loading location have little effect on index value until the size of the omitted cross-loading increases above about .35. Once the omitted cross-loading is larger than .35, the balanced model (red line) shows the poorest fit, while the best fit is shown for the most imbalanced model when the cross-loading is omitted from the smaller factor (black dashed line). However, it is worth noting that once the omitted cross-loading increases above .35, CFI values do not increase above the commonly used cutoff value of .95.

In contrast, RMSEA only shows poor model fit once the size of the omitted cross-loading increases above about .85, and even then only for the most imbalanced models and when the cross-loading is omitted from the larger factor. Once again, we see RMSEA exhibiting insensitivity to misspecifications due to an omitted cross-loading. RMSEA, in comparison to CFI, also appears to be less sensitive to model balance in general.

3.3 Misspecification Source: Misspecified Latent Structure

This section focuses on the third and final source of misspecification examined in this thesis: misspecification of the latent structure. Recall that the factors of a CFA model are considered latent variables. Latent variables are those that cannot be directly observed or measured, but are inferred through indicator variables that can be observed and directly measured.

Because the factors in a model cannot be directly observed, situations may arise in which a researcher's proposed model does not accurately reflect the number of factors in the model underlying the population. For example, a proposed model may contain only one factor when the actual population model contains two. The misspecifications examined in this section occur when a researcher's proposed model fails to include the same number of factors that exist in the true (population) model.

As in the previous two sections, we again focus on CFA models. In the scenarios presented here, the covariance matrices corresponding to the true model are constructed to include a specific number of factors and the hypothesized models are constructed so as to either underestimate or overestimate the number of factors. This misspecification is reflected in the covariance structure of the hypothesized model.

To use the example presented above, if there is only *one* factor in the population, the population model has a covariance structure given by $\boldsymbol{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\psi}$, where $\boldsymbol{\lambda}$ is a $p \times 1$ vector of factor loadings and $\boldsymbol{\psi}$ is the $p \times p$ covariance matrix of the residuals, where p represents the number of indicator variables in the model.

If the hypothesized model suggests that there are *two* factors underlying the data, it is overestimating the number of factors in the population. The hypothesized model then has a covariance structure given by $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\psi}$, where $\mathbf{\Lambda}$ is a $p \times 2$ matrix of factor loadings, $\mathbf{\Phi}$ is a 2×2 matrix of factor correlations, and $\mathbf{\psi}$ is the $p \times p$ covariance matrix of the residuals. Note that $\mathbf{\Phi}$ is absent from the structure of the true model in this example, since there is only one factor present and thus no factor correlations to be measured.

In addition to the size of the misspecification (as measured by how different the latent structure of the hypothesized model is when compared to the true model), the influence of loading size, model size, and factor correlation on index behavior are examined as well.

3.3.1 Effects of Misspecification Size

For scenarios involving a misspecified latent structure, the size of the misspecification can be defined by how much a hypothesized model overestimates or underestimates the number of factors in the true (population) model. While the actual number of factors in the population can never be truly known, it is worth exploring how sensitive CFI and RMSEA are to misspecifications that are strictly due to an "incorrect" number of factors included in a hypothesized model rather than any other source (e.g., an omitted error covariance or an omitted cross-loading).

Figure 3.14 plots fit index value against the number of factors in the true model (2 to 8) when the hypothesized model includes only one factor (i.e., the hypothesized model is *underestimating* the number of factors in the true population). The number of indicator variables is held at p = 24, and the indicators are equally distributed across the factors in the population (e.g., for the 2-factor population model, each factor has 12 indicators loading onto it; for the 6-factor population model, each factor has four indicators loading onto it). The size of the loadings (both in the true and hypothesized model) are represented by the six colored curves. Red, orange,

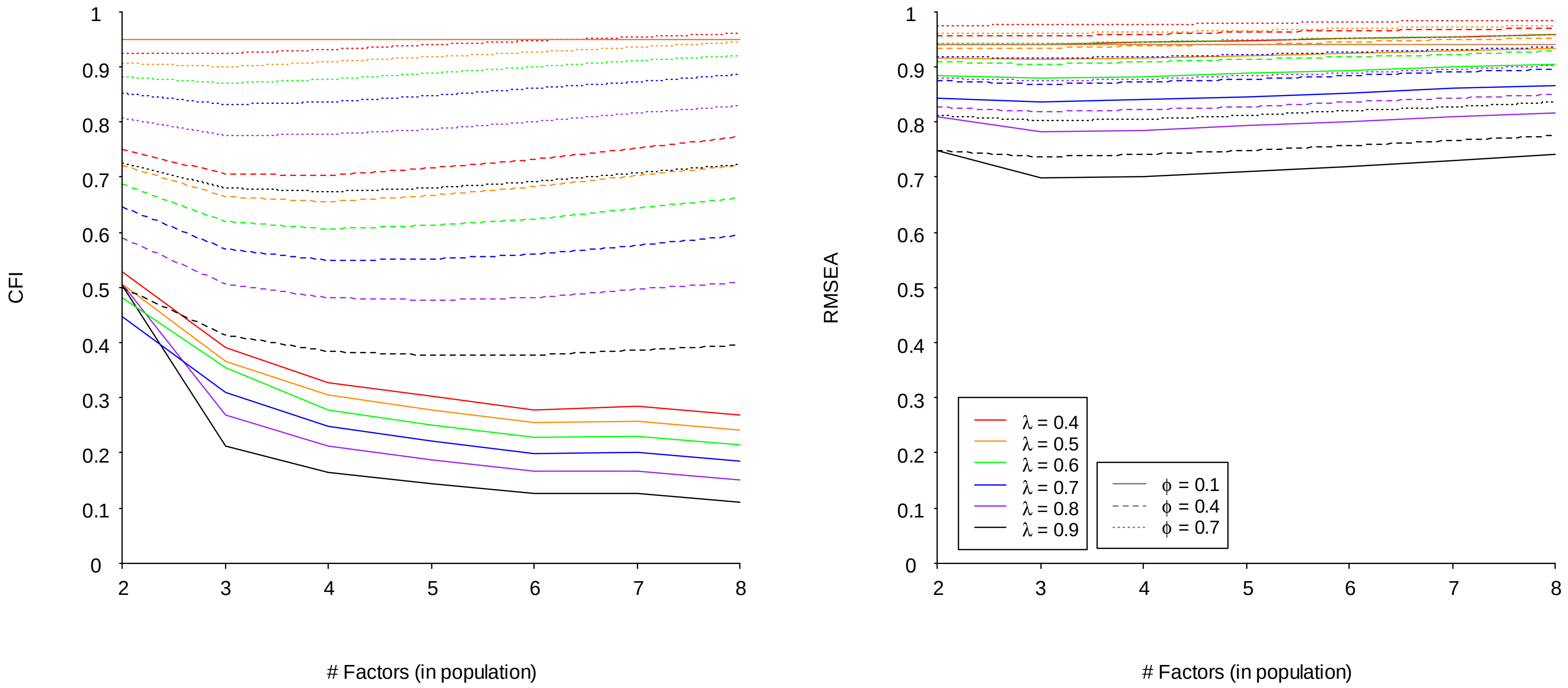

Figure 3.14: Plots of population fit index values vs. the number of latent factors in the population model (2 to 8) when the hypothesized model is a 1-factor model. The number of indicators is held at 24. Correlations amongst latent factors are set at .1 (solid lines), .4 (dashed lines), or .7 (dotted lines). The colored curves correspond to different loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively.

green, blue, purple, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. The correlations amongst factors in the population are set to .1 (solid lines), .4 (dashed lines), or .7 (dotted lines). Note that the legend for CFI was omitted due to lack of space, but that the legend for the RMSEA plot also applies to the CFI plot.

While it is unlikely that a researcher would have such an extremely misspecified latent structure as would result from fitting a 1-factor model to data from a population that actually has eight factors, the hope is that both CFI and RMSEA would show a decrease in fit as the 1-factor model is fit to data from a population with an increasing number of factors.

Figure 3.14 shows that CFI generally shows a poor fit when misspecification is due to an underestimation of the number of factors in the hypothesized model. CFI is very sensitive to factor correlations in this scenario. When the factor correlations are set to .1 in the population, CFI shows the worst fit for the 1-factor hypothesized model, with the largest CFI value being about .53. This is an appropriate and desired behavior, as it suggests that CFI is very sensitive to a model that fails to incorporate the correct number of mildly correlated factors in the population. CFI values are higher, in general, as the factor correlations increase. It is worth reiterating here that the factor correlations are those amongst the factors in the true (population) model, and thus would be unknown to the researcher.

When the factor correlations are either .4 or .7, CFI shows an improvement in fit as the size of the misspecification grows (that is, as the number of population factors increases). However, this improvement is very slight, and CFI values only increase above the commonly used cutoff value of .95 for the case when $\lambda = .4$ and the factor correlations are .7. As mentioned above, a misspecification this severe is highly unlikely in real research settings.

RMSEA, on the other hand, shows a slight improvement in fit as the size of the misspecification increases, regardless of the factor correlations in the population. For all factor correlation sizes and loading sizes, RMSEA appears to show worst fit when the actual number of factors in the population is three. For the smallest loading included in this scenario ($\lambda = .4$), RMSEA would show good model fit regardless of the size of the misspecification. RMSEA appears much less sensitive to the factor correlation sizes than CFI. Both fit indices show an improvement in fit, to some degree, as the size of the misspecification grows (though CFI appears more sensitive than RMSEA to misspecified latent structures).

A notable concern with the setup of the scenario in Figure 3.14 is the fact that the *p:k* (indicators to factors) ratio is not constant. Recall that the number of indicators in Figure 3.14 is held constant at $p = 24$ across all values of *k*. This means, for example, that when $k = 3$, the ratio of indicators to factors is 8:1, while when $k = 6$, the ratio is 4:1. Thus, the results observed from this figure may be confounded by the changing *p:k* ratio.

Figure 3.15 presents the same scenario as Figure 3.14, except holding *p:k* constant. As in Figure 3.14, Figure 3.15 plots index values against an increasing number of factors in the true model (2 to 8) when the hypothesized model includes only one factor. Now, however, instead of *p* being held constant at 24, the number of indicators appropriately increases as *k* increases, keeping the *p:k* ratio constant. This is done by allowing either three, five, or seven indicator variables to load onto each factor in the true model. The colored curves correspond to different loading sizes in both the hypothesized and true models, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. The scenario involving three indicators per factor is represented by the solid lines, the scenario involving five

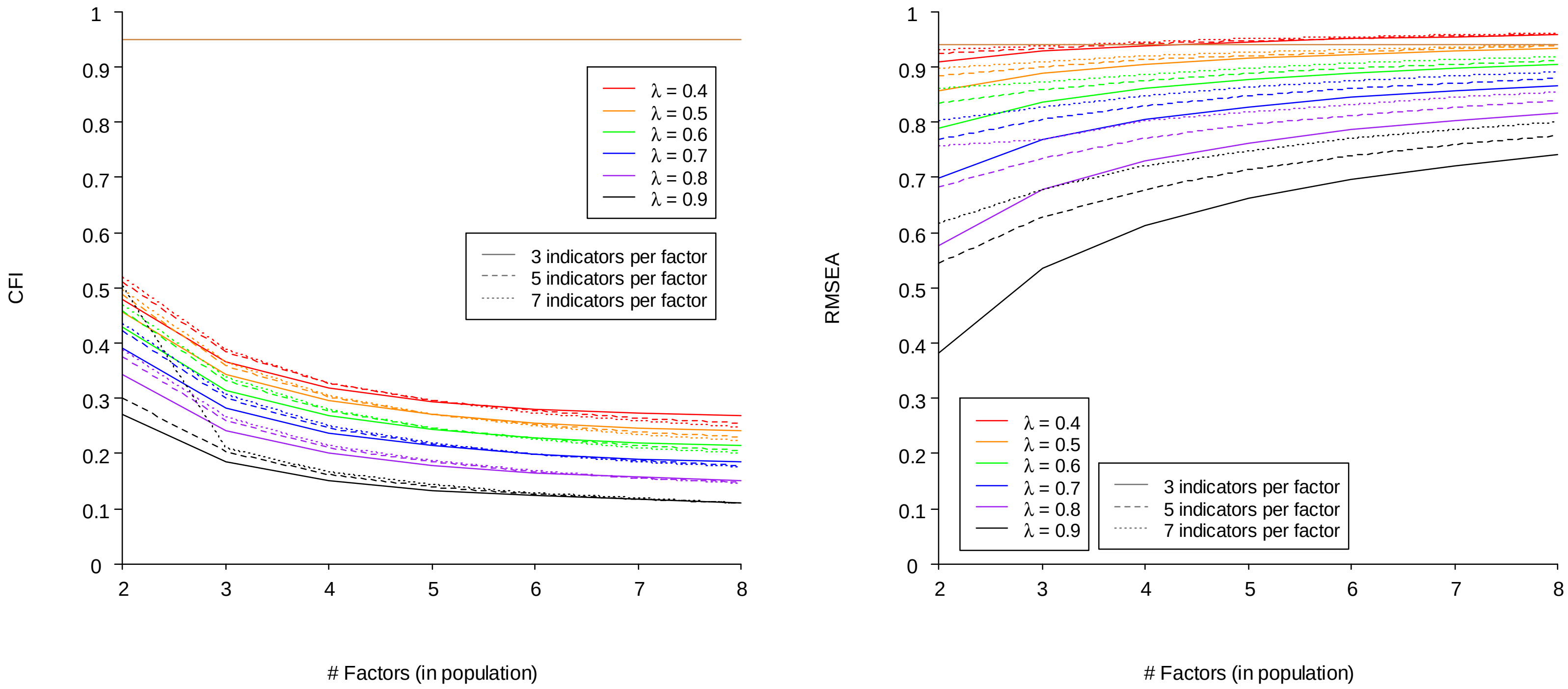

Figure 3.15: Plots of population fit index values vs. the number of latent factors in the population model (2 to 8) when the hypothesized model is a 1-factor model. The ratio of indicators to factors is held constant, with either 3, 5, or 7 indicators per factor (solid, dashed, and dotted lines, respectively). The colored curves correspond to different loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively.

indicators per factor is represented by the dashed lines, and the scenario involving seven indicators per factor is represented by the dotted lines.

Comparing Figure 3.15 with Figure 3.14, it is clear that maintaining the *p:k* ratio as the number of factors increases (Figure 3.15) changes the behavior of CFI from the case where the *p*:*k* ratio decreased as the number of factors increased (Figure 3.14). In Figure 3.15, a more desirable behavior is seen for CFI. Specifically, regardless of the loading size and the number of indicators per factor, CFI shows a decrease in fit as the size of the misspecification (the number of factors in the population) increases. In fact, the largest CFI value in this scenario is about .52, which suggests that CFI is highly sensitive to the misspecified latent structure in this scenario.

For RMSEA, we again see an improvement in fit as the size of the misspecification increases (as we saw in Figure 3.14). This may be due to the fact that RMSEA is sensitive to the number of indicator variables in a model and shows an improvement in fit as the number of indicators increases. (the same phenomenon that was seen in Figures 3.5 and 3.11, for example). In Figure 3.15, in, order to keep the *p:k* ratio constant as k increases, p must get quite large. For example, when $k = 8$ and the number of indicators per factor is held at seven (dotted lines), $p = 56$. Thus, the behavior of RMSEA in Figure 3.15 may be due more to the increasing model size rather than the increasingly misspecified latent structure. However, considering that RMSEA also showed an improvement in fit in Figure 3.14, when the size of the hypothesized model was held constant, these two results combined may suggest that RMSEA is simply not as sensitive as CFI is to incorrectly modeled latent structures.

3.3.2 Effects of Model Size

As was seen in Figure 3.15, it appears that the size of the model (in terms of the number of indicators) may affect the behaviors of both CFI and RMSEA in the case of a misspecified latent structure. In this section, the effect of the number of indicator variables is examined in further detail.

Figure 3.16 plots index values against an increasing number of indicators (p = 4, 6, 8, 10, 12, 14, 16, 18, and 20) for the case when a hypothesized 1-factor model is fit to 2-factor data. The six colored curves correspond to six different loading sizes (both in the hypothesized and true model), with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9. The population (true) model has a factor correlation of .1 (solid lines), .4 (dashed lines), or .7 (dotted lines). In the population model, there are an equal number of indicators per factor (e.g., when p = 10, each factor has five indicators loading onto it).

The behavior of CFI is highly influenced by the size of the factor correlation in the population, a trend that has been seen before (e.g., Figure 3.14). When the factor correlation is low (.1), CFI shows the worst fit, with values below about .55 regardless of the number of indicators. This is desirable behavior, as it suggests that CFI is sensitive to the case where a model omits a second factor that is quite different (only slightly correlated) from the first factor. As the factor correlation increases in size, so do the CFI values. However, the values generally do not increase above the commonly used cutoff of .95. These results suggest that regardless of model size, CFI appears to retain its sensitivity to misspecified latent structures and lead to the decision that the 1-factor model is a poor fit for 2-factor data.

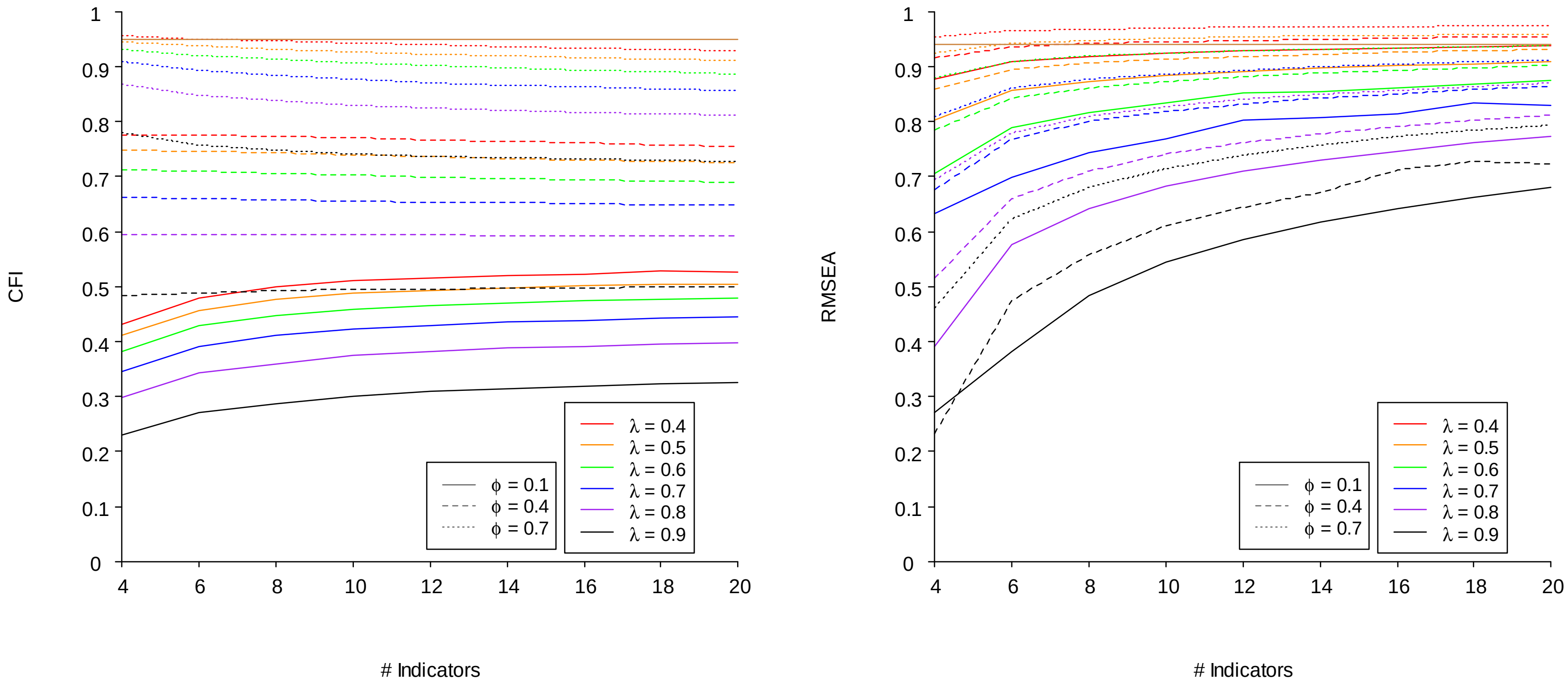

Figure 3.16: Plots of population fit index values vs. the number of indicators (*p* = 4, 6, 8, 10, 12, 14, 16, 18, 20) when a 1-factor model is fit to 2-factor data. Factor correlation is .1 (solid lines), .4 (dashed lines), or .7 (dotted lines). The colored curves correspond to different loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. Neighboring points are connected for readability.

For RMSEA, an improvement in fit is seen as the number of indicators increases, a behavior that has been noted in other scenarios as well. However, except for the case where the factor correlation is .4 or .7 and the loadings are low (.4, .5), RMSEA values remain below the commonly used cutoff value, suggesting the model is a poor fit. This indicates that while RMSEA shows an improvement in fit as the number of indicators increases, this improvement in fit would, in most cases, not lead a researcher to accept a model with a larger number of indicators while rejecting a model with fewer indicators, holding all other things constant.

In all of the misspecified latent structure scenarios presented thus far, the hypothesized model has underestimated the number of factors in the population. It is also of interest to examine index behavior when the hypothesized model *overestimates* the number of factors in the population instead. Figure 3.17 is similar to Figure 3.16 in that it plots index values against an increasing number of indicators (p = 4, 6, 8, 10, 12, 14, 16, 18, and 20). Now, however, instead of a hypothesized 1-factor model being fit to 2-factor data, a hypothesized 2-factor model is fit to 1-factor data (only one factor underlying the population). As in Figure 3.16, the six colored curves in Figure 3.17 correspond to six different loading sizes (both in the hypothesized and true model), with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9. The hypothesized model has a factor correlation of .1 (solid lines), .4 (dashed lines), or .7 (dotted lines). In the hypothesized model, there are an equal number of indicators per factor (e.g., when p = 10, each factor has five indicators loading onto it).

Comparing Figure 3.16 to Figure 3.17, the behaviors of CFI and RMSEA are fairly similar across both scenarios. When the hypothesized model overestimates the number of factors in the population (Figure 3.17), CFI values generally do not increase above the commonly used cutoff, similar to when the hypothesized model underestimates the number of factors

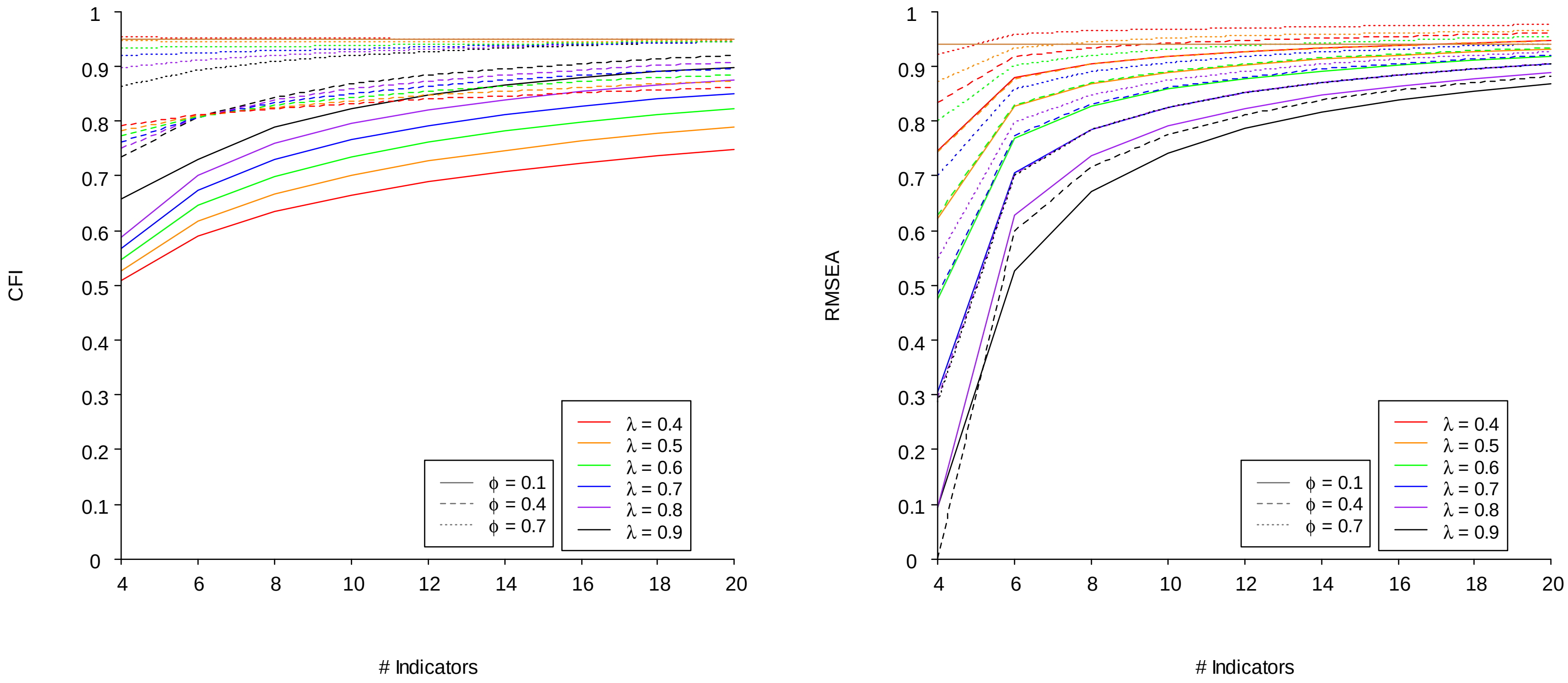

Figure 3.17: Plots of population fit index values vs. the number of indicators (p = 4, 6, 8, 10, 12, 14, 16, 18, 20) when a 2-factor model is fit to 1-factor data. Factor correlation is .1 (solid lines), .4 (dashed lines), or .7 (dotted lines). The colored curves correspond to different loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. Neighboring points are connected for readability.

(Figure 3.16). While CFI appears to show an improvement in fit in Figure 3.17 as the number of indicators increases, the only case for which CFI values are above the cutoff value are when $\phi = .7$ and $\lambda = .4$. Thus, just as in Figure 3.16, CFI appears to retain its sensitivity to misspecified latent structures when the hypothesized model overestimates the number of parameters in the population. RMSEA's behavior in Figure 3.17 is almost identical to its behavior in Figure 3.16, suggesting that the fit index is not affected by how the latent structure is misspecified (overestimating or underestimating the number of factors in the population).

3.3.3 Effects of Factor Correlation

Figure 3.18 examines the effects of factor correlation in scenarios involving a misspecified latent structure. Specifically, a 1-factor model is fit to 2-factor data. Since the hypothesized model contains only one factor, the correlation being manipulated in this scenario is the correlation between the two factors in the population. While this value is, in practice, unknown, it is of interest to see how the population factor correlation affects how both CFI and RMSEA perform for this type of misspecified latent structure.

Figure 3.18 plots index values against an increasingly large (population) factor correlation when a hypothesized 1-factor model is fit to 2-factor data. The six colored curves correspond to six different loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9. The number of indicator variables is either 12 (solid lines) or 24 (dashed lines). In the 2-factor population model, there are an equal number of indicators loading onto both factors.

Both CFI and RMSEA show desirable behavior in this scenario. That is, they show worst fit when the factor correlation is low, and perfect fit when the factor correlation is 1. While both indices follow the general trend of showing better fit as the factor correlation size increases, CFI values begin much lower than RMSEA values, and no model would be accepted as having good fit according to CFI until the factor correlation in the population was above about .7. RMSEA, on the other hand, shows good model fit when the factor correlation is as low as about .15 in the case where $p = 24$ and $\lambda = .4$. This contrast between CFI and RMSEA gives further evidence to suggest that CFI is, on average, more sensitive than RMSEA to misspecified latent structures. CFI also appears to be less affected by the model size (solid and dashed lines) in this scenario than RMSEA.

In the previous figure, the hypothesized model underestimates the number of factors in the population. The last scenario I examine here is the same setup as the scenario in Figure 3.18, except now a 2-factor model is fit to 1-factor data, representing the case when the hypothesized model overestimates the number of factors in the population. Figure 3.19 plots index values against an increasingly large factor correlation when a hypothesized 2-factor model is fit to 1-factor data. The factor correlation being manipulated here is the factor correlation in the hypothesized model, not the population, since the population has only one factor. The six colored curves correspond to six different loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9. The number of indicator variables is either 12 (solid lines) or 24 (dashed lines). In the 2-factor hypothesized model, there are an equal number of indicators loading onto both factors.

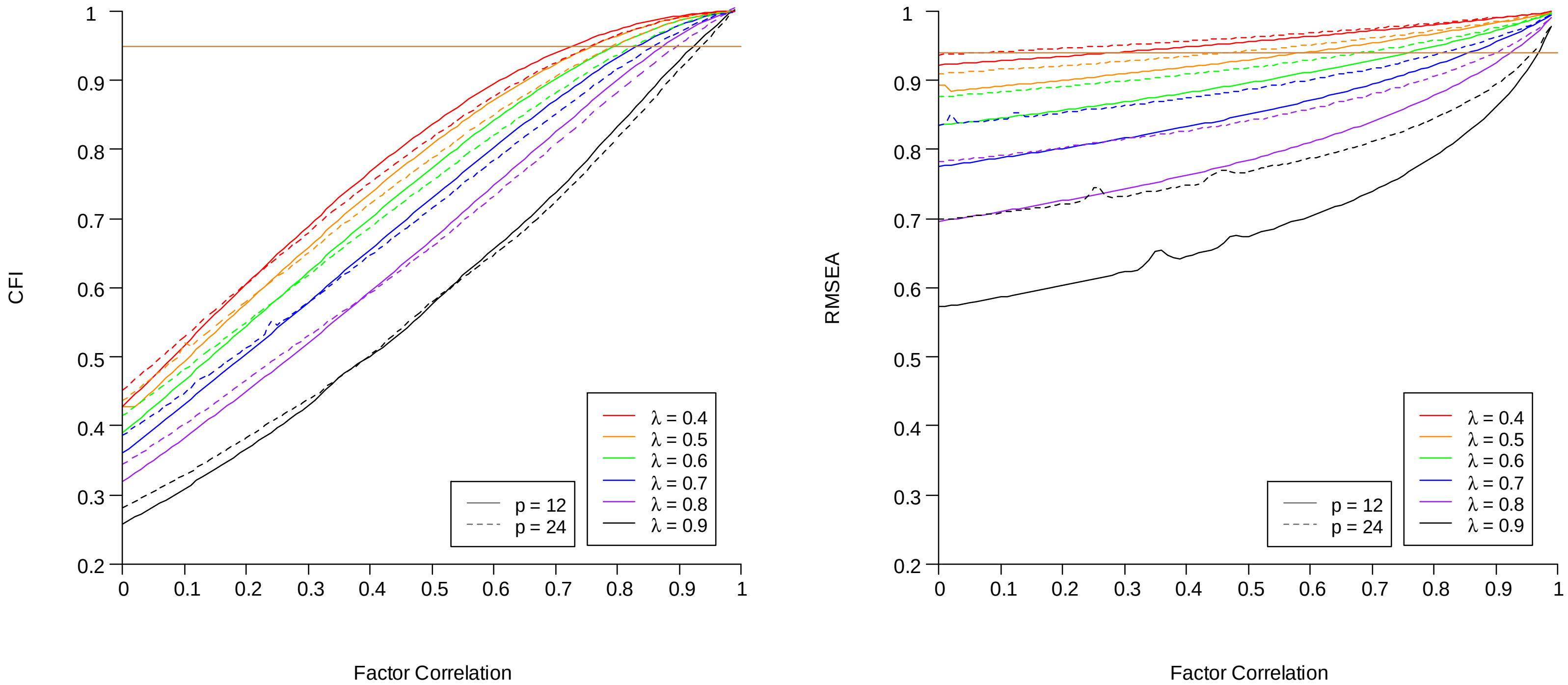

Figure 3.18: Plots of population fit index values vs. factor correlation (0 – 1) when a 1-factor model is fit to 2-factor data. The number of indicators is *p* = 12 (solid lines) or *p* = 24 (dashed lines). The colored curves correspond to different loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively.

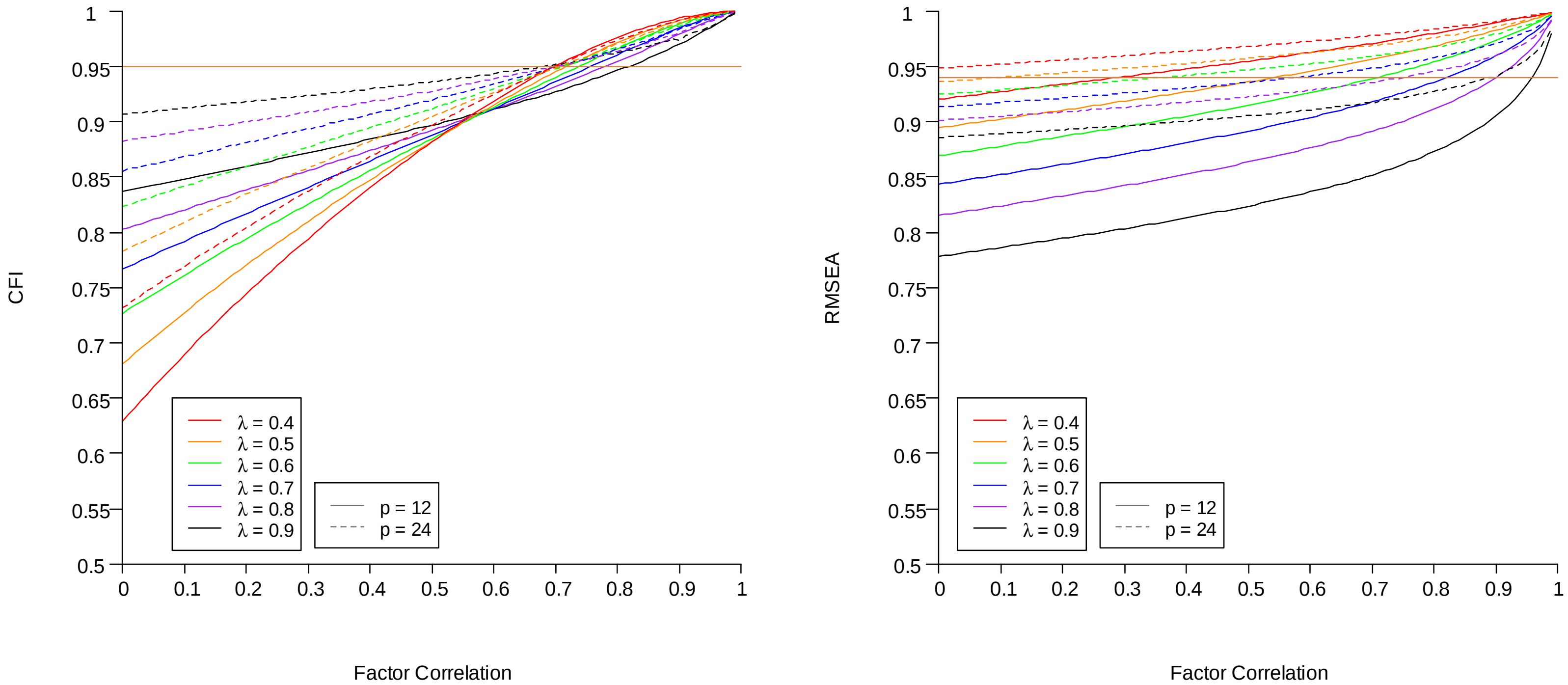

Figure 3.19: Plots of population fit index values vs. factor correlation (0 – 1) when a 2-factor model is fit to 1-factor data. The number of indicators is *p* = 12 (solid lines) or *p* = 24 (dashed lines). The colored curves correspond to different loading sizes, with red, orange, green, blue, purple, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively.

Comparing Figures 3.18 and 3.19, it is clear that both CFI and RMSEA behave similarly across both scenarios, suggesting that the type of latent misspecification (overestimating or underestimating the number of factors in the population) does not dramatically affect index behavior. As in Figure 3.18, CFI shows poor fit in Figure 3.19 until the factor correlation in the hypothesized model is above about .7. While CFI values are not as low in Figure 3.19 as they are in Figure 3.18, if a researcher is assessing fit based on the commonly used cutoff value, their conclusions about fit would remain the same whether the hypothesized model is overestimating or underestimating the number of factors in the population. RMSEA behaves almost identically in both scenarios, showing good model fit in the case where $p = 24$ and $\lambda = .4$, but poor fit in all other cases.

Chapter Four: Discussion

In this chapter, a summary of the results of Chapter 3 is provided. I first focus on the behavior of CFI, broken down by the source of misspecification (omitted error covariance, omitted cross-loading, and misspecified latent structure). I provide possible explanations for certain index behaviors, mentioning results that have been seen in the literature as well as any interesting behaviors that have not been discussed in previous studies. I then present a review of the behavior of RMSEA in the same fashion.

Following these summaries, I discuss the possibility of combining the use of CFI and RMSEA when assessing model fit. Combining the use of these two indices, I claim, will provide researchers with more information about possible model misspecifications over and above any information given by relying upon either index on its own.

4.1 Summarized Results for CFI

4.1.1 Misspecification Source: One or More Omitted Error Covariances

I begin by discussing CFI's behavior in situations where misspecification is due to one or more omitted error covariances. Perhaps the most interesting result for CFI involves the non-monotone relationship between index value and the size of a single omitted error covariance, a result that has been seen both in Savalei (2010) and Mahler (2011). As seen in Figures 3.1 and 3.2, CFI appears to show worse fit for a "moderately" sized omitted error covariance, while showing better fit both when the omitted error covariance is small and when it is close to .9. This suggests that relying solely on CFI in a situation where misspecification may be due to an omitted error covariance may lead a researcher to accept a model with a larger misspecification (e.g., $\psi = .8$) while rejecting a model with a moderate misspecification (e.g., $\psi = .4$).

A possible cause for the non-monotone pattern of CFI in these figures has to do with how CFI is calculated. Recall the population definition of CFI:

$$\mathrm{CFI} = 1 - \frac{\hat{\mathrm{F}}}{\hat{\mathrm{F}}_{\mathrm{I}}}. \tag{2.2}$$

In the population, CFI is a function of both the minimized fit function of the hypothesized model ($\hat{F}$) and the minimized fit function of the baseline model ($\hat{F_I}$). As previously mentioned, the current study used a baseline model in which all indicator variables are uncorrelated.

It was suspected that the non-monotone relationship between CFI values and the size of an omitted error covariance was due to different rates of change for $\hat{F}$ and $\hat{F_I}$. Thus, I plotted the values of $\hat{F}$ and $\hat{F_I}$ separately against an increasing omitted error covariance in a 1-factor model with 8 indicators (the same model used in Figure 3.1). These plots revealed that $\hat{F}$ and $\hat{F_I}$ indeed have different rates of change that depend on the size of the omitted error covariance. This suggests that there is a relationship between the size of the misspecification and its influence on the minimized fit functions for both the hypothesized and the baseline model.

Based on the results of Figure 3.1 and Figure 3.2, it might be appealing to discourage the use of CFI in cases where a researcher may suspect the possibility of omitted error covariances in their model. However, there are three things to note regarding CFI's behavior in this situation. First, while there is a curvilinear relationship between CFI values and misspecification size, if a researcher were to employ the commonly used cutoff criterion for CFI, the smaller model in this scenario (p = 8) would be regarded as fitting poorly regardless of the loading sizes once the omitted error covariance rises above about .2. Thus, the non-monotone relationship does not affect the cutoff-based decision about the model once the omitted error covariance is large enough.

Second, when the number of indicators is increased from $p = 8$ to $p = 16$, the curvilinear relationship between CFI values and omitted error covariance size disappears for all but the smallest loadings case, and the index appropriately shows a decrease in fit as the size of the omitted error covariance increases. This suggests that the non-monotone relationship between index size and omitted error covariance size might not even be present in models with a larger number of indicator variables.

Finally, it should be noted that CFI does behave appropriately when the number of omitted error covariances increases (Figure 3.3, Figure 3.4). That is, a model with a greater number of omitted error covariances will be shown to have worse fit than a model with a fewer number of omitted error covariances. This result agrees with findings by Heene et al. (2012). In addition, CFI did not appear to be overly sensitive to multiple omitted error covariances if the covariances were quite small ($\psi = .05$). This can be seen as a positive trait of the index; it suggests that the omission of a few very small error covariances, which is likely not to be considered a severe misspecification, will not affect CFI's value enough to cause the researcher to reject the model. CFI's sensitivity to more misspecifications is also not greatly affected by the number of latent variables ($k = 1$ or $k = 2$).

The influence of the number of indicators in CFI behavior, however, appears to be confounded with the number of latent variables. CFI initially shows a decrease in fit as the number of indicator variables increases, a trend that has been observed in previous literature (e.g., Chau and Hocevar, 1995; Kenny and McCoach, 2003; Moshagen, 2012) when the misspecification was due to omitted error covariances. However, in the 1-factor model case

(Figure 3.3), this trend reverses once the number of indicator variables reaches a certain amount, and CFI shows an improvement in fit. In the 2-factor model case (Figure 3.4), CFI shows a decrease in fit regardless of the number of indicators. While this difference in behavior between the 1- and 2-factor models is worth noting, it is of little practical concern. In the scenario presented in Figure 3.5, the change from a decrease in fit to an improvement in fit does not affect whether a researcher would reject or accept a model based on the commonly used cutoff value except in two cases where the omitted error covariances were quite small (ψ = .2, .3). In such cases, the misspecification is small enough that it may not even warrant rejecting the model.

A final result worth noting involves CFI's behavior in an imbalanced model scenario. As seen in Figure 3.8, while model imbalance appears to affect CFI values for larger omitted error covariances, once the size of the omitted error covariance increases above about .3, CFI shows poor model fit by the commonly used cutoff value, regardless of the degree of imbalance. This suggests that researchers need not worry about even severely imbalanced models (e.g., one factor with 7 indicators and another factor with 17 indicators) greatly affecting CFI values.

4.1.2 Misspecification Source: One or More Omitted Cross-Loadings

Next, I summarize the performance of CFI in situations where misspecification is due to one or more omitted cross-loadings. As was the case when a single error covariance was omitted from the model, we see another non-linear relationship between CFI values and the size of a single omitted cross-loading (Figure 3.9). Specifically for the p = 8 scenario, CFI values decrease until the size of the omitted cross-loading is approximately .4, then begin to increase again.

It should be noted, however, that CFI values never increase above the common cutoff value of .95. Thus, while this non-linear behavior may appear troubling at first, it is of little practical concern. Even if a researcher were to employ a more conservative cutoff value (say, .90), the slight increase in CFI values for very large cross-loadings would not lead the researcher to accept the model as having adequate fit in a scenario like the one presented in Figure 3.9.

CFI exhibits interesting behavior when the number of omitted cross-loadings increases in a given model. As the number of omitted cross-loadings increases, CFI actually shows an improvement in fit, to the point where the index shows perfect fit in the case where there exist 10 cross-loadings in the true population model but all 10 are omitted from the researcher's model. This is concerning, as it implies that a researcher could, in theory, accept a model that omits as many as 10 cross-loadings based on the CFI value. Such findings are not present in previous literature (e.g., Fan and Sivo, 2007; Hu and Bentler, 1998), which has demonstrated that CFI is appropriately sensitive to an increasing number of omitted cross-loadings.

CFI does exhibit a useful behavior, however, with respect to the relationship between factor correlation and cross-loadings. As was seen in Figure 3.12, as the correlation between factors of a 2-factor model increases from 0 to 1, CFI shows an improvement from poor fit to perfect fit, regardless of the number of indicators in the model or the size of the omitted cross-loading. This suggests that CFI appropriately reflects the "severity" of the misspecification with respect to how correlated the two factors are. Specifically, when $\phi = 1$, the two factors can essentially be interpreted as being the same factor and the model becomes a 1-factor model. Thus, there is really no cross-loading present at all, and CFI shows perfect fit because nothing is being omitted. However, as the factors become more and more distinct (ϕ decreases), the omitted cross-loading becomes more obvious, as there is an indicator loading onto two distinct factors in the population, but this is not being reflected in the researcher's model. Thus, model fit decreases as the factor correlation decreases, reflecting that the cross-loading is an important relationship in the population that is being excluded from the model.

The effect of the number of indicators on CFI value is similar to the effects seen when the misspecification is due to an omitted cross-loading. That is, while CFI initially shows a decrease in fit as the number of indicators increases, after a certain point, it shows fit improving as the number of indicators continues to increase (Figure 3.11). This effect is likely due to CFI's sensitivity to model size, which has been documented in previous studies (e.g., Chau and Hocevar, 1995; Moshagen, 2012). Thus, while this result may not indicate anything specific about CFI's sensitivity to omitted cross-loadings, it does suggest that researchers should be aware that model size influences CFI values, regardless of the source of any misspecification their model might have.

In all scenarios examined here, CFI is appropriately sensitive to the size of the omitted cross-loading. The larger the omitted cross-loading is, the worse fit CFI shows. This is true regardless of the size of the other loadings. In addition, CFI appears to be more sensitive to misspecifications due to omitted cross-loadings when the other loadings in the model are higher versus when they are lower. This result agrees with several previous studies (e.g., Miles and Shevlin, 2007; Themessl-Huber, 2014), which showed that CFI is less sensitive to misspecification when loadings are, on average, fairly low.

Finally, I note the result of model imbalance on CFI behavior. As Figure 3.13 shows, CFI values are affected by the location of the omitted cross-loading. Specifically, when the cross-loading is omitted from the factor with fewer indicator variables, CFI appears less sensitive to the misspecification than when the cross-loading is omitted from the factor with more indicator variables. However, similar to the imbalanced model case when the misspecification was due to an omitted error covariance, CFI shows poor model fit regardless of either the location of the misspecification or the degree of imbalance once the omitted cross-loading is larger than about .4. Thus, this suggests that while model imbalance and the location of the omitted cross-loading do affect CFI values, they do not do so in a way that would practically affect the selection or rejection of a model.

4.1.3 Misspecification Source: Misspecified Latent Structure

As was mentioned in the previous chapter, while the actual number of factors in the population can never be truly known, it is worth exploring how sensitive CFI and RMSEA are to cases where a hypothesized model either overestimates or underestimates the true number of factors in a population to see if either index is particularly useful in detecting this sort of misspecification.

In Figure 3.14, a 1-factor hypothesized model is fit to data arising from a population with an increasing number of factors, from two to eight. Regardless of loading size or factor correlation size, CFI appears to be very sensitive to misspecified latent structure, even when the discrepancy between the number of factors in the hypothesized model and the number of factors underlying the population is "small" (e.g., a 1-factor model fit to 2-factor data). In the scenario presented in Figure 3.14, there is no combination of factor correlation and loading size for which CFI shows good model fit when the number of population factors is anywhere from two to six. This indicates that CFI will reflect a misspecified latent structure even if that misspecification is "small".

CFI does appear to show slight improvement in fit as the misspecification increases for moderate or high factor correlation ($\phi = .4$ or $\phi = .7$). However, this slight increase in CFI values may be due to the fact that in Figure 3.14, the number of indicator variables was held constant at 24. Thus, for the $k = 2$ model, each factor had 12 indicators loading onto it, while in the $k = 8$ model, each factor only had three indicators loading onto it. That is, the fact that the *p:k* ratio is not constant may affect CFI's behavior in the case of a misspecified latent structure.

To determine if this is true, Figure 3.15 involved the same scenario as Figure 3.14, except the *p:k* ratio was held constant. The results indicated that CFI is even more sensitive to misspecified latent structure when, overall, a model is larger (more indicator variables when there are a larger number of factors). This is consistent with findings by Chau and Hocevar (1995) and Sharma et al. (2005). CFI shows a desirable decrease in fit as the size of the misspecification increases, regardless of loading size and regardless of the number of indicators loading onto each factor.

The effect of the number of indicator variables alone was also examined, and it was found that CFI tends to be rather insensitive to changes in the number of indicator variables when latent misspecification exists. This behavior was noted in a study by Chau and Hocevar (1995). As Figures 3.16 and 3.17 show, CFI still shows poor fit (values below the common cutoff value) when a 1-factor model is fit to 2-factor data and when a 2-factor model is fit to 1-factor data. The CFI values are generally unchanged regardless of the number of indicator variables included in the hypothesized model in either case. This is a desirable result, as it suggests that for both small and large models, CFI appears to be highly sensitive to a misspecified latent structure, regardless of whether the misspecification is due to the hypothesized model overestimating the number of population factors or underestimating the number of population factors.

Finally, Figures 3.18 and 3.19 show that when a 1-factor model is fit to 2-factor data or when a 2-factor model is fit to 1-factor data, CFI shows very poor fit when the two factors are slightly correlated, but shows perfect fit when the factors have a perfect correlation (when, essentially, there is only one factor). This is a desirable result. If the two factors in the population are only slightly correlated, fitting a 1-factor model to data from this population (Figure 3.18) can be considered a larger misspecification than fitting a 1-factor model to data from a population with moderately- or highly-correlated factors. Similarly, if the two factors in a hypothesized model are only slightly correlated, fitting this model to data with 1-factor (Figure 3.19) can be considered a larger misspecification than fitting a 2-factor model with highly correlated factors to 1-factor data.

4.1.4 Summary

From the results of the various scenarios presented in this study, a general summary of CFI's behavior in different misspecification circumstances can be obtained. When model misspecification is due to either one or more omitted error covariances or one or more omitted cross-loadings, the relationship between CFI and the severity of the misspecification appears to be non-linear. Specifically, CFI shows good model fit both when the misspecification is small (an omitted error covariance of .2, for example) and when the misspecification can be considered large (an omitted error covariance of .9, for example). The index shows worst fit when the misspecification is moderate.

For the scenarios presented here, this non-linear behavior would not affect a researcher's decision to reject or fail to reject a model based on the commonly used cutoff of .95. That is, even though CFI showed an improvement in fit as the severity of the misspecification increased, CFI values did not increase above .95 in situations where an omitted error covariance or an omitted cross-loading was large. However, it should be noted that such non-linear behavior may affect a researcher's decision for models that differ from the specific ones presented here. For example, in Figure 3.1, only models of size $p = 8$ and $p = 16$ were examined. If a researcher were to have a model with fewer indicator variables, there is the chance that the non-linear behavior of CFI may cause the researcher to accept a model containing a severe misspecification (an omitted error covariance of .9, for example).

In addition to this, the results in this study showed that while CFI did show worse fit as the number of omitted error covariances increased in a model, it showed better fit as the number of omitted cross-loadings increased in a model. These results, combined with the non-linear behavior, may suggest that CFI may not accurately reflect how well a model fits (or fails to fit) if misspecification is due to omitted error covariances or to omitted cross-loadings.

In contrast, however, the behavior of CFI in cases where models have a misspecified latent structure suggests that this index is highly sensitive to misspecifications of this type. Regardless of model size, loading size, and the correlations between factors in the population, CFI appears to show poor model fit when a 1-factor model is fit to data from a population with more than one factor and when a 2-factor model is fit to data from a population with only one factor. Considering that a misspecified latent structure can be considered a more severe form of misspecification than one due to omitted error covariances or cross-loadings, these findings indicate that CFI can be a useful tool in detecting misspecifications of this type.

The influence of other modeling components (model size, factor loadings, etc.) can also be summarized here. In general, higher factor loadings increase CFI's sensitivity to misspecifications arising from any source. That is, the higher the factor loadings, the lower CFI values will be for a misspecified model. While CFI generally shows better fit as more indicators are added to a model, this trend appears to be less apparent in models with more factors and is not as severe as for RMSEA (discussed below). Finally, it appears that CFI is not highly affected by model balance.

4.2 Summarized Results for RMSEA

4.2.1 Misspecification Source: One or More Omitted Error Covariances

In contrast to the non-linear behavior presented by CFI, RMSEA behaves more appropriately fit in situations where misspecification is due to an omitted error covariance. RMSEA shows a decrease in fit as the size of the omitted error covariance increases. This is the case for 1-factor and 2-factor models (Figures 3.1 and 3.2). In addition, Figures 3.3 and 3.4 showed that RMSEA also shows worse fit as the number of omitted error covariances increases. These results are consistent with findings by Hu and Bentler (1998) and Sharma et al. (2005). Like CFI, RMSEA does not appear to be overly sensitive to multiple omitted error covariances if the covariances are quite small ($\psi = .05$).

It is important, however, to note RMSEA's sensitivity to loading sizes in these scenarios. RMSEA appears to be more sensitive to loadings than CFI is, at least when misspecification is due to one or more omitted error covariances. For the same severity of misspecification, a model with low factor loadings (.4) may be accepted as having good fit according to the commonly used cutoff, while a model with higher factor loadings (.8 or .9) would be rejected as having poor fit. This specific sensitivity to loading size has been observed in previous studies as well (e.g., Miles and Shevlin, 2007; Themessl-Huber, 2014) and may suggest that a universal cutoff value may not be appropriate for RMSEA across different loading sizes.

Another important result involves RMSEA's sensitivity to model size (as measured by the number of indicator variables). When misspecification is due to an omitted error covariance, regardless of loading size or the severity of the misspecification, RMSEA shows an improvement in fit as more indicator variables are included in a model, regardless if the number of factors was one (Figure 3.5) or two (Figure 3.6), though the effect was lessened by the addition of another factor. This behavior is expected, as previous research (e.g., Browne, 1987; Kenny and McCoach, 2003; Sharma et al., 2005), has demonstrated that RMSEA tends to show better fit as model size increases. This improvement in fit as p increases, both for RMSEA and CFI, might be due to the fact that as the model gets larger, the addition of more indicators "dilutes" the misspecification and thus masks its effects.

Finally, as was the case for CFI, the effect of model balance does not highly influence RMSEA values when misspecification is due to an omitted error covariance (Figure 3.8). This indicates that even in models as imbalanced as the ones included in this study (e.g., one factor having seven indicators loading onto it and the other factor having 17 indicators loading onto it), RMSEA values will not be significantly influenced by the imbalance.

4.2.2 Misspecification Source: One or More Omitted Cross-Loadings

When misspecification is due to a single omitted cross-loading, RMSEA shows an appropriate decrease in fit as the size of the omitted cross-loading increases (Figure 3.9), which is consistent with findings by Hu and Bentler (1998) and Fan and Sivo (2007). However, similar to CFI, when the number of omitted cross-loadings increases, RMSEA shows an improvement in fit.

In fact, in the scenario presented in Figure 3.10, a researcher would fail to reject a model as having poor fit no matter how many cross-loadings have been omitted from it. This is true regardless of factor correlation size and loading size. However, it should be noted that there are a total of 20 indicators (p = 20) in the model in Figure 3.10. Thus, RMSEA's insensitivity to the number of omitted cross-loadings may be canceled out by the fact that RMSEA tends to show better fit in larger models, regardless of the misspecification type or size.

In Figure 3.11, the effect of the number of indicators was explicitly examined in the omitted cross-loading misspecification. As was the case with CFI, as the number of indicator variables increased, RMSEA began to show an improvement in fit once the number of indicators rose above six. This was true regardless of the size of the omitted cross-loading.

However, the sizes of the other loadings appeared to have a large effect on RMSEA values. When loadings (apart from the cross-loading) were low, RMSEA failed to show poor fit, regardless of the size of the model and the size of the omitted cross-loading. But when the other loadings were high (.7), RMSEA appeared to be sensitive to the omitted cross-loading in smaller models, but then still show an improvement in model fit as the number of indicator variables increased. This may suggest that RMSEA's sensitivity to omitted cross-loadings may be influenced not only by the number of indicators in a model but also by the size of the other loadings. Previous work by Themessl-Huber (2014) and Miles and Shevlin (2007) also show that RMSEA appears to be less sensitive to misspecifications in general when the loadings are, on average, low.

Figure 3.12 shows the effect of factor correlation (in the population) on index' ability to detect misspecification due to an omitted cross-loading. While RMSEA exhibits the same general pattern as CFI—showing better fit as the factor correlation increases—it should be noted that RMSEA values failed to suggest good model fit according to the commonly used cutoff value, regardless of the factor correlation or the size of the omitted cross-loading.

Finally, as was the case when misspecification was due to an omitted error covariance, model balance does not appear to affect RMSEA values to any significant degree (Figure 3.13), which suggests that any imbalance in the number of indicators per factor will not highly affect RMSEA values.

4.2.3 Misspecification Source: Misspecified Latent Structure

Finally, I discuss the behavior of RMSEA in scenarios involving a misspecified latent structure. Figure 3.14 shows a 1-factor model being fit to data from a population with more than one factor. In this figure, the lines of the RMSEA values appear to be flat (or increasing only slightly) as the discrepancy between the number of factors in the hypothesized model and the number of factors underlying the population increases. In addition, for smaller loading sizes ($\lambda = .4$), RMSEA indicates good model fit regardless of how many factors are in the population model and the degree to which they are correlated. These results are similar to those achieved by Miles and Shevlin (2007) and may suggest that RMSEA is, in general, not highly sensitive to latent structure misspecifications.

However, as was noted in the summary of CFI's behavior, the $p{:}k$ ratio in Figure 3.14 is not held constant; thus, RMSEA's behavior may be partially due to the inconsistent $p{:}k$ ratio as k increases and p remains the same. In Figure 3.15, this ratio was held constant by increasing p as k increased, so as to have three, five, or seen indicators per factor, regardless of the number of factors.

In Figure 3.15, RMSEA actually shows an improvement in fit as the discrepancy between the number of factors in the hypothesized model and the number of factors in the true population increases. That is, regardless of loading size or factor correlation size, a 1-factor model fit to 2-factor data fits worse than a 1-factor model fit to 8-factor data, according to RMSEA. A possible explanation for this behavior again goes back to the well-documented fact that RMSEA is highly sensitive to model size, showing better fit in for models with more indicators, all other things held constant. In Figure 3.15, p must be substantially increased with each incremental increase of k in order to keep the $p{:}k$ ratio constant. For example, when $k = 2$ and there are seven indicators per factor, $p = 14$. However, if $k = 8$ and there are seven indicators per factor, then $p = 56$. There is a large difference in the size of these models, which may explain why RMSEA shows a better fit for larger k. In fact, the results of Figure 3.15 are consistent with results of a similar setup by Sharma et al. (2005).

The effect of model size was also examined in Figures 3.16 and 3.17, and the same trend is observed. When a 1-factor model is fit to 2-factor data and when a 2-factor model is fit to 1-factor data, RMSEA shows better fit for larger models, regardless of loading size and factor correlation size. However, it should be noted that except when loadings are small ($\lambda = .4$), a researcher would still reject these models as having poor fit.

Finally, similar to CFI, RMSEA behaves appropriately with respect to the relationship between the degree of misspecification and the size of the population factor correlations. In Figures 3.18 and 3.19, a 1-factor model is fit to 2-factor data (Figure 3.18) and a 2-factor model is fit to 1-factor data (Figure 3.19). RMSEA shows poor fit when the factor correlation is small, but increases to show perfect fit when the factor correlation is one. This is useful behavior, as it suggests that RMSEA would be more sensitive to a more “severe” misspecification (modeling one factor when the population involves two highly different factors, or modeling two highly different factors when the population involves only one factor) than a less severe misspecification (modeling one factor when the population involves two very correlated factors, or modeling two very correlated factors when the population involves only one factor).

4.2.4 Summary

When misspecification is due to one or more omitted error covariances, RMSEA shows worse fit as the size of a single omitted error covariance increases as well as when more error covariances are omitted from a hypothesized model. This indicates that RMSEA appropriately reflects a lack of fit when a hypothesized model fails to include the correct number of error covariances that exist in the population.

While RMSEA shows a decrease in fit as the size of an omitted cross-loading increases, it actually shows an improvement in fit as more cross-loadings are omitted from a hypothesized model. In addition, RMSEA values tend to suggest good model fit according to the common cutoff value when misspecification is due to an omitted cross-loading. This suggests that regardless of other modeling components (e.g., factor correlation size, model size), RMSEA is not very sensitive to omitted cross-loadings. Thus, if a researcher were to rely solely on RMSEA when assessing model fit, a model may be accepted as having good fit even if there is a rather large misspecification due to one or more omitted cross-loadings.

RMSEA also does not appear to reflect the severity of misspecification when a model fails to include an accurate number of latent factors. Specifically, in the scenario presented here, RMSEA values are similar for a 1-factor model fit to 2-factor data and a 1-factor model fit to

8-factor data. The fact that it is not sensitive to the difference in severity of these misspecifications suggests that RMSEA should not be used in situations where competing models with different latent structures exist and a researcher is trying to determine which has better fit.

Despite RMSEA's insensitivity to the severity of a misspecified latent structure, the index in general shows poor fit when this type of misspecification is present, which suggests that it can be a useful tool when detecting misspecifications of this type.

In addition to the influence of different misspecification types on RMSEA's behavior, it is important to note the effects of different modeling components, such as loading size, model size, and model balance. In comparison to CFI, RMSEA appears to be more heavily influenced by loading size. Both indices tend to be more sensitive to misspecifications when loadings are higher (.8, .9) versus when they are lower (,4, .5). However, for RMSEA, the differences between index values for different loading sizes is more dramatic.

This sensitivity, combined with the use of a cutoff value to determine if a model fits or not, might lead a researcher to accept a particular model if that model had small enough loadings, but would lead them to reject the same model if that model included higher loadings. While it is unlikely that a given researcher would have two competing models that only differ in loading sizes, the fact that RMSEA is so sensitive to this modeling component may suggest that a universal cutoff value is not appropriate when assessing model fit. Instead, consideration should perhaps be given to the average loading sizes present in a model before a cutoff-like value is selected.

RMSEA is also highly sensitive to model size, as measured by the number of indicator variables in a model. As has been shown in several studies prior to this one, RMSEA shows an improvement in fit as the number of indicators increases. This is true regardless of the source of misspecification and regardless of the ratio of the number of indicators to the number of factors, and suggests that researchers should not rely solely on RMSEA to accurately assess fit in larger models and should be aware that adding more indicator variables will likely artificially improve model fit according to RMSEA.

4.3 Recommendations

Based on the summaries presented above, it is clear that CFI and RMSEA perform differently in certain modeling scenarios that include different modeling components and sources of misspecification. This study backs up previous research that has shown that neither index performs universally "better" than the other. However, the fact that CFI's and RMSEA's behaviors tend to complement each other in several different cases suggests that researchers may gain a better understanding of a hypothesized model if they were to use both CFI and RMSEA in conjunction.

While the idea of combining the use of fit indices is not a new one, and has in fact been mentioned multiple times in the literature, little has been said about combining CFI and RMSEA, and to my knowledge, nothing has been published on how best to interpret their combined values in order to assess what might be the cause of a model's misspecification.

In this section, I briefly reiterate the previous literature on index combining, as first discussed in Chapter 1, then present ways of using CFI and RMSEA in conjunction in order to best understand the possible sources of misspecification in a hypothesized model.

4.3.1 A Brief Recap of Previous Research on Combining Indices

The fact that different indices appear to be sensitive to different sources of misspecification (as well as to different modeling components) has been used in previous studies to justify the claim that more than one index should be reported when stating the fit of a hypothesized model.

Hu and Bentler (1999) make one of the earliest claims that a presentation of two fit indices rather than one might better reflect the true fit of a given model. They recommend pairing either CFI and SRMR or RMSEA and SRMR, noting that SRMR was the only index studied whose sensitivity to certain misspecification types was significantly different than other indices' sensitivities. Hu and Bentler (1999) also suggest using stricter cutoff values, with .96 being used for CFI and .05 for RMSEA. It should be noted, though, that the main sources of misspecification in Hu and Bentler's (1999) study were misspecified factor covariances and misspecified factor loadings, which are different than the misspecifications included in this study.

Other authors recommend combining indices based on index type. Hooper et al. (2008), Kline, 2005, and Boomsma (2000) suggest reporting CFI, RMSEA, and SRMR in addition to the chi-square. Since each of these three indices were developed under a different philosophy of measuring model fit, these authors argue that each index should be sensitive to different types of misspecification. Combining their results should allow researchers to better assess how well a model fits as well as what might be the source behind any degree of misfit.

While these authors recommend reporting multiple fit indices, little to nothing is said about what might be the reasons behind possible discrepancies amongst the indices (e.g., if RMESA shows good fit but SRMR does not) or how to interpret such discrepancies if they are witnessed. More importantly, nothing is said about how researchers could actually benefit from observing discrepancies amongst fit indices in order to possibly improve their model to better reflect what is going on in the population.

Based on the current study, I suggest the use of CFI and RMSEA in conjunction when assessing model fit, and attempt to provide a series of general guidelines for researchers to interpret what certain discrepancies between these indices might mean and how they can use this information to possibly improve the fit of their model.

4.3.2 Combining CFI and RMSEA

From the results of the present study, CFI and RMSEA appear to be sensitive to different sources of misspecification as well as to different modeling components, such as model size, loading size, and factor correlation (in models with two or more factors). These results suggest that combining CFI and RMSEA when assessing model fit may not only help a researcher back up their claim of having a well-fitting model when the indices both show good fit, but can also help them better understand the possible source(s) of misspecification if one or both indices show poor fit. Here, I discuss some general guidelines and suggestions based on the results of the simulations presented above.

4.3.2.1 CFI Shows Good Fit, RMSEA Shows Poor Fit

It may be the case for a given model that CFI shows good fit (> .95) while RMSEA shows poor fit (> .06). If this is the situation, it may be worth examining possible sources of misspecification. If it is suspected that the latent structure of the hypothesized model is correct, a possible source of the discrepancy between CFI and RMSEA may be the omission of one or more error covariances. As the results in the present study show, CFI has a tendency to suggest good fit (values just slightly greater than .95) even if an error covariance of .9 is omitted from the hypothesized model. RMSEA, on the other hand, shows poor fit when a large error covariance is omitted. Thus, it is a possibility that misspecification is due to a large omitted error covariance, but this misspecification is simply not being picked up by CFI. This may be especially likely in smaller models (models with fewer than eight indicators), as the nonlinear relationship between CFI and the size of an omitted error covariance appears to be more exaggerated the fewer indicator variables there are in a given model.

Another possible cause of this type of discrepancy may simply be the loading sizes. As previously stated, while both CFI and RMSEA become more sensitive to any type of misspecification if the loadings in a model are higher (.7, .8, .9), RMSEA appears much more influenced by loading sizes than CFI. If a researcher observes a discrepancy between CFI and RMSEA where CFI is showing good fit but RMSEA is showing poor fit, it might be worth taking note of the size of the loadings in the model. If the loadings are generally high, then the cause of the discrepancy may simply be RMSEA's heightened sensitivity to misspecifications in models with higher loadings.

4.3.2.2 CFI Shows Poor Fit, RMSEA Shows Good Fit

For some models, it may be the case that CFI shows poor fit (< .95) while RMSEA shows good fit (< .06). The present research has shown that there may be several causes for this type of discrepancy. First, while neither index appeared to be highly sensitive to misspecification due to one or more omitted cross-loadings, CFI was shown to be more sensitive to this type of misspecification than RMSEA. For example, when a 2-factor model with 10 indicators per factor omitted three cross-loadings that exist in the population, CFI showed the model as having poor fit when the loadings and omitted cross-loadings were all .3. However, RMSEA still showed the model as having good fit. CFI also appears more sensitive to omitted cross-loadings regardless of loading size or model size. If a researcher suspects a possible omitted cross-loading as a source of misspecification and observes poor fit with CFI and good fit with RMSEA, this may suggest that a cross-loading is, in fact, not present in the hypothesized model when it should be.

As was shown in the previous chapter, CFI appears to be more sensitive than RMSEA to misspecified latent structures. While both indices showed poor fit when a 1-factor model was fit to 2-factor data (Figure 3.18) and when a 2-factor model was fit to 1-factor data (Figure 3.19), the values of CFI were very low (as low as .27), indicating much worse fit than the RMSEA values, which never decreased below .55. If a researcher were in a situation where both CFI and RMSEA showed poor fit, but CFI showed much worse fit than RMSEA, it might indicate that the poor fit is due to a misspecified latent structure. Looking for this pattern with CFI and RMSEA can be especially helpful if it is suspected that the latent structure may not be accurate.

Finally, an additional cause of this type of discrepancy may be the size of the model in question. The present research confirms what has previously been shown in the literature: regardless of the source of misspecification, RMSEA tends to show better fit for models with a large amount of indicator variables. This is especially true when loadings are high (.7, .8, .9). While CFI also shows a similar trend, it is much less pronounced and actually disappears with larger loadings. Thus, if a researcher has a relatively large amount of indicator variables in a specific model and observes that CFI is showing poor fit while RMSEA is showing good fit, the discrepancy may simply be due to the fact that the large number of indicators is artificially improving fit for RMSEA while not affecting CFI to the same extent. If the researcher has no cause to suspect any major misspecification, noting RMSEA's decreased sensitivity to misspecification in large models can help the researcher explain the discrepancy between CFI's poor fit and RMSEA's good fit.

4.3.2.3 Both CFI and RMSEA Show Poor Fit

In the present study, there was not a scenario in which both CFI and RMSEA showed poor fit without there being a rather serious misspecification. Thus, if a researcher observes both CFI and RMSEA indicating poor fit, it is likely that there is something legitimately wrong with the hypothesized model in question.

However, when both show poor fit, it may be more difficult to determine what is causing the misspecification in the first place based solely on the index values themselves. Thus, the best course of action in such a situation may be for the researcher to either refer to other SEM studies conducted on the topic of interest or to re-examine the theory behind the topic and determine if any reasonable adjustments to the model can be made.

4.3.2.4 Both CFI and RMSEA Show Good Fit

If both CFI and RMSEA show good fit, it is likely the case that the model in question is a good representation of the covariance structure underlying the population. However, I make note of two model components that may, in certain cases, influence both CFI and RMSEA to the extent that they indicate good fit when, in fact, there is at least a moderate degree of misspecification.

The first model component that researchers should always take note of is the size of the loadings. As shown in this study, both CFI and RMSEA values are influenced by loading size. The lower the average loadings are in a model, the less sensitive these indices are to any type of misspecification. Thus, if a particular model has very low loadings on average (e.g., most loadings are around .2), it may be the case that both CFI and RMSEA are unable to detect even moderately-sized misspecifications.

A second model component is the size of the model, as measured by the number of indicator variables. As discussed in previous sections, RMSEA shows better fit for larger models, regardless of the source or size of misspecification. While this behavior is much less pronounced for CFI (and in fact is hardly present when the source of misspecification is a misspecified latent structure), there may be cases where models consist of enough indicators to cause both CFI and RMSEA to show good fit when, in fact, a misspecification is present.

Chapter Five: Application

Discussed in the previous chapter were various suggestions of how to interpret CFI and RMSEA values in order to determine the source of model misspecification. In this chapter, the goal is to provide some applications of these suggestions to real-life data. I will present data from two different studies and discuss competing models for each set of data. I wish to show that examining both the CFI and RMSEA values for specific models may help guide a researcher to the source or sources of possible misspecification.

For each of the two sets of data, I first give a brief overview of the purpose and importance of the original studies, as well as some background on the data involved. I then discuss different proposed models, computing the CFI and RMSEA values and using them to attempt to explain any possible sources of model misspecification.

5.1 The Causal Dimension Scale (McAuley, et. al.)

The first data come from a paper by McAuley et al. (1992), titled *Measuring Causal Attributions: The Revised Causal Dimension Scale (CDSII)*. In psychology, Attribution Theory seeks to explain why people behave the way they do. The theory reduces the causes of behavior to three dimensions: *locus of causality* (is the cause internal or external to the person?), *stability* (is the cause constant over time or changeable?), and *control* (can the cause be controlled?) (Weiner, 1985).

The Causal Dimension Scale (Russell, 1982) was developed to measure how individuals perceived causes in terms of these three dimensions. Since the development of this scale, however, several researchers (e.g., McAuley and Gross, 1983; Russell et al., 1987) have expressed concerns regarding the scale's structure. Specifically, they claim that the *control* dimension tends to correlate highly with the *locus of causality* dimension and that the *control* dimension lacks internal consistency. In addition, concerns have been raised over whether the difference between personal control versus external control should be addressed within the control dimension.

In response to these concerns, McAuley et al. (1992) sought to examine the *control* dimension in further detail by creating a new variation of the Causal Dimension Scale (which they denoted as CDSII). In addition to three items assessing locus of causality and three items assessing stability, the authors included six items to assess control. Three of these items were specifically created to measure personal control, and three were specifically created to measure external control. After examining four different factor structures of the 12-item CDSII, the authors concluded that a model consisting of four factors—*locus of causality, stability, external control*, and *internal control*—was the best fit to the data.

5.1.1 Model 1

In a paper overviewing the use of factor analysis in the *Personality and Social Psychology Bulletin*, Russell (2002) takes a closer look at the relationship between the *locus of causality* and the *stability* subscales of the CDSII developed by McAuley et al. (1992). Russell claims that in a typical multifactor CFA model, items typically have non-zero loadings onto their respective factors and zero loadings onto any other factors present in the model. He uses the items and factors of the *locus of causality* and *stability* subscales as defined in McAuley et al. (1992) to demonstrate such a model. Note that the factors are modeled as being orthogonal (uncorrelated).

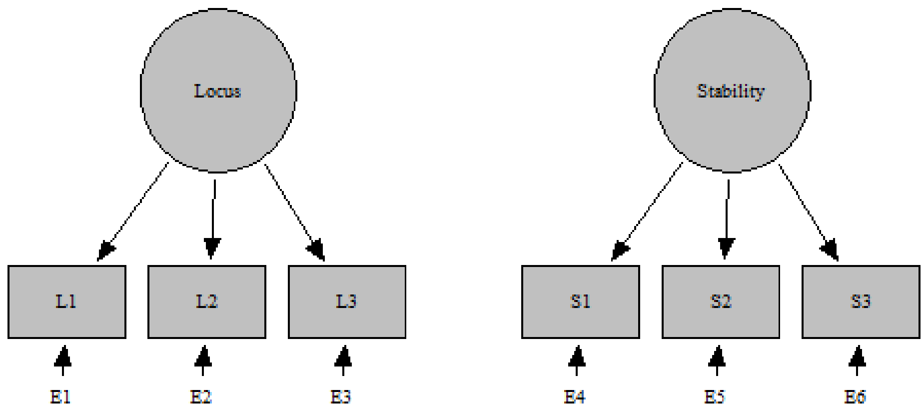

Figure 5.1: A visual representation of Russell's (2002) hypothesized relationship between the factors *locus of causality* and *stability*, determined to be orthogonal, and six subscale items from McAuley et al. (1992).

Russell (2002) then tests this model using the data originally obtained by McAuley et al. (1992) (n = 380 college students). The results indicate that the model fits the data, with $\chi^2(9) = 14.44$, $p = .11$. These values were obtained using LISREL 8.3. I confirmed these values using EQS 6.3 and also obtained the values of CFI = .985 and RMSEA = .039. Both fit indices also suggest good model fit, as CFI is larger than the commonly used cutoff (.95) and RMSEA is smaller than the commonly used cutoff (.06).

Based on the suggestions offered at the end of Chapter 4 of this thesis, it is likely that this model is indeed a good representation of the covariance structure underlying the population. Recall that both loading sizes and model size may influence both indices to show good fit when there is actually a substantial misspecification. Specifically, models with low loadings as well as models with a large number of indicator variables may show good fit regardless of the size of any misspecifications. However, in this particular case, loading sizes estimated to be average to high (all greater than .53) and the model is small in terms of the number of indicators (p = 6). Thus, this model is likely a good representation of the relationship amongst these variables and factors in the population.

5.1.2 Model 2

Russell (2002) offers an additional model of the relationship of the *locus of causality* and *stability* subsets. This second model is identical to the one presented above, except the two factors are allowed to correlate. The author's justification for this change is to see if allowing the correlation will significantly improve the fit of the model.

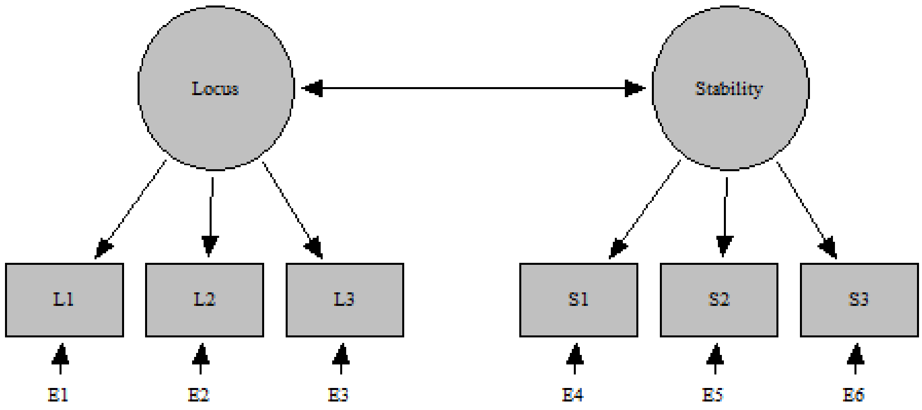

Figure 5.2: A visual representation of Russell's (2002) hypothesized relationship between the factors *locus of causality* and *stability*, with the factors allowed to correlate, and six subscale items from McAuley et al. (1992).

For this model, the resulting chi-square indicates that it also fits the data, with $\chi^2(8) = 14.39$, $p = .07$. This result was confirmed using EQS 6.3 and the index values of CFI = .982 and RMSEA = .045 were also computed. A difference in chi-square values between this model and the previous was also given by Russell (2002) as $\chi^2(1) = .05$, $p = .82$, suggesting that this second model does not lead to a significant improvement in fit over the first model.

5.1.3 Model 3

In the model in which the *locus of causality* and *stability* factors were permitted to correlate, the author allowed for LISREL 8.3 to estimate the value of the correlation. It was found that this correlation, estimated at $r = .02$, was nonsignificant. This, combined with the fact that the model with two uncorrelated factors appeared to fit the data well, suggests that it is likely that two distinct factors do exist in the population.

As mentioned in the beginning of this section, there has been debate surrounding the structure of the Causal Dimension Scale, particularly regarding the relationships amongst the three proposed dimensions. Suppose, as in the original study by McAuley et al. (1992), that a researcher wishes to test other possible models for this scale. Specifically, suppose they suspected that there was one underlying factor for all six items (indicators) presented in the previous models. That is, instead of three items loading on to the *locus of causality* factor and three items loading onto the *stability* factor, all six items loaded onto one common factor.

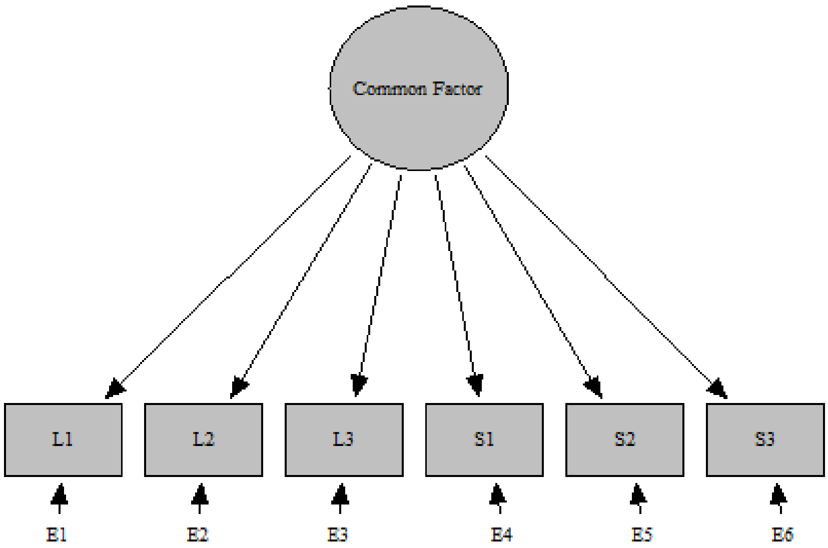

Figure 5.3: A visual representation of an alternative 1-factor model with six subscale items from McAuley et al. (1992).

Testing this model using EQS 6.3, I obtain $\chi^2(9) = 186.599$, $p < .001$, with CFI = .502 and RMSEA = .228. Judging by the value of the chi-square test statistic alone, this model does not fit the data well. However, the information given to us by the CFI and RMSEA values can be used to try and determine what the source or sources of misspecification might be.

In this case, both fit indices show poor fit (CFI < .95, RMSEA > .06). As was mentioned at the end of the previous chapter of this thesis, when both indices show poor fit, it is likely that there is a serious misspecification in the model. Thus, it can be useful to take a closer look at how both indices are behaving in order to determine if we can figure out a possible source of misspecification.

Notice in this scenario that while both indices show poor fit, the value of CFI, .502, is quite low when compared to the commonly used cutoff of .95. The simulations carried out in this thesis (specifically, 3.16, 3.17, 3.18, 3.19) showed that CFI is highly sensitive to latent structure misspecifications, to the point that the index's values will show very poor fit when a model is omitting a latent factor that exists in the population.

Given what we already know about the fit of the two-factor model in this scenario, the high value of RMSEA combined with the very low value of CFI might suggest to us that this is a case where the latent structure of the proposed model does not match the latent structure underlying the population. Specifically, the proposed model is not including the correct number of latent factors.

5.2 Reliability and Stability in Panel Models (Wheaton, et. al.)

The second set of data come from *Assessing Reliability and Stability in Panel Models* by Wheaton, et al. (1977). When individuals are measured on certain constructs of interest (or factors) over time, the result is referred to as panel data (or longitudinal data). It is often of interest to study the relationships amongst variables across time points. These relationships amongst indicator variables and the underlying factors are often represented by a measurement model.

In many cases, the factors have multiple indicators per time point. For example, suppose the factor of interest is general intelligence and how it changes throughout childhood and adolescence. A researcher who is interested in how this factor changes over time may measure a number of indicator variables assumed to load onto general intelligence (such as the scores on a vocabulary test, scores on a math test, and verbal ability) at different ages to determine if there is a pattern of change.

Of concern to those working with panel data is the specification of this measurement model to best explain the relationships between the factors and indicator variables in such a situation where measurements are taken across multiple time points. Wheaton, et al. (1977) address the issue of representing not only the relationships amongst indicator variables but relationships amongst factors as well in the context of panel data.

A component of their original article involves data from a longitudinal study of the effects of industrial development on individuals in a rural part of Illinois. The authors wished to determine if certain social attitudes were stable over time or if they were prone to change with changes in the environment (the change, in this case, being the industrial development). Six attitude scales, along with a measure of education and a measure of socioeconomic standing, were administered to n = 932 individuals at three different time points: 1966, 1967, and 1971.

While Wheaton, et al. (1977) developed and tested several different models to represent the relationships amongst the variables of interest, these models were ultimately more complex than the simulated models presented in this study and as a consequence contained modeling components that were not examined in this thesis. However, the data used in the original study have been used by various other researchers in attempt to model the relationships amongst the variables of interest. A portion of the original data and variables are even used in the Lisrel (EQS 6) Program Model (Bentler, 2006) as a demonstration.

Thus, the two models I discuss in the following sections are those developed by other authors for subsets of the original sets of variables.

5.2.1 Model 1

The first model of interest comes from Gonzalez and Griffin (2001). An adaptation of the data presented in the Lisrel Program Manual (Bentler, 2006), originally from Wheaton, et al. (1977), here the authors focus only on four variables of interest: the scores on the Anomie scale as measured in 1967 (labeled “X1”), the scores on the Powerlessness scale as measured in 1967 (“X2”), the scores on the Anomie scale as measured in 1971 (“X3”), and the scores on the Powerlessness scale as measured in 1971 (“X4”).

In the model presented by Gonzalez and Griffin (2001), two latent factors are included, one of which with the two Anomie scale variables loading onto it and the other with the two Powerlessness scale variables loading onto it. The authors allow the latent factors to correlate and, based on the fact that the data are panel data, allow the error variances of the two Anomie scale variables to covary with each other and the error variances of the two Powerlessness scale variables to covary with each other as well. Note that the loadings and error variances for each of the two scale measurements are set to be equal.

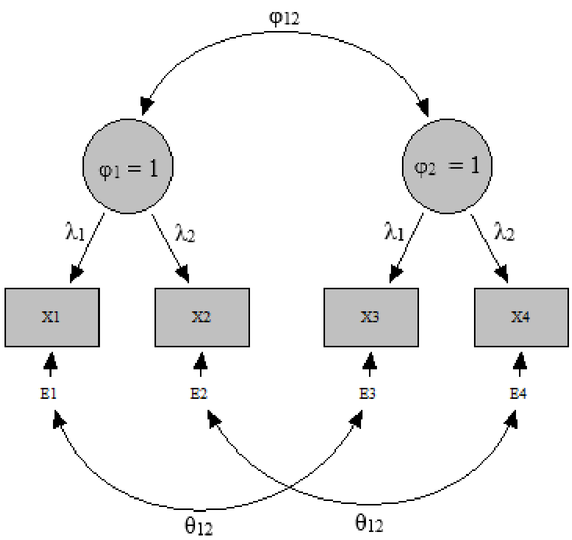

Figure 5.4: Figure from Gonzalez and Griffin (2001), showing two latent factors, with two indicator variables each. The model includes two error covariance terms.

Fitting this model to the data, the authors obtained a non-statistically significant chi-square test statistic value ($\chi^2(4) = 2.969$, $p = .563$), suggesting that the model is a good fit for the data. I verified this result using EQS 6.3, and also obtained CFI = 1 and RMSEA = 0. Given that both fit indices are at the extreme "good fit" ends of their respective scales, there is good evidence to suggest that this two-factor model with correlated error variances is an accurate representation of the relationships amongst these variables in the population.

5.2.2 Model 2

Given the setup of the panel data used by Gonzalez and Griffin (2001), it makes sense for the authors to include the two error covariances in their proposed model. Since X1 and X3 are the same variable measured at different times (same with X2 and X4), it is likely that there exists some covariance between their error terms.

However, the case may arise in which an individual, working with this same set of data, may not recognize the need to include covarying error terms in their model. Suppose that Gonzalez and Griffin's (2001) original model is re-created by another researcher, the only difference being that the two error covariances are omitted.

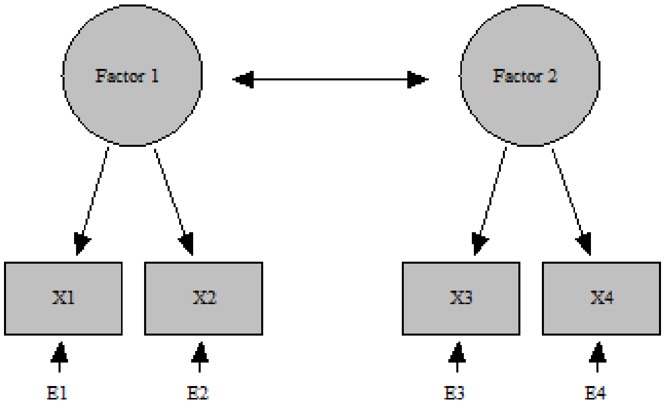

Figure 5.5: The same model as in Figure 5.4, except with the two error covariance terms omitted.

Testing this model using EQS 6.3, I obtained a statistically significant chi-square test statistic ($\chi^2(5) = 63.868$, $p < .001$), suggesting that this model is a poor fit for the data. In addition, I obtained CFI = .962 and RMSEA = .112. It is interesting to note that in this case, both the chi-square test statistic and RMSEA both show the model as having poor fit, while CFI shows good fit using the commonly used cutoff of .95. While this model would be rejected based on the chi-square value, it is of interest to see what might be causing the discrepancy between the CFI and RMSEA values to possibly determine the source of the misspecification.

In the previous chapters, it was found that RMSEA, more so than CFI, is sensitive to misspecifications due to omitted error covariances, especially in models with a small number of indicator variables. Given that both the chi-square test statistic and RMSEA suggest that this model is a poor fit to the data, it may be the case that this model fails to include covariance terms that exist in the population.

Another possible source for the discrepancy between CFI and RMSEA values may be the average loading sizes in the model. As was shown in previous chapters, RMSEA appears to be affected by loading sizes, in that if loadings sizes are large on average in a given model, RMSEA is prone to being more sensitive to any source of misspecification. Again, however, it is useful to note that RMSEA agrees with the chi-square test statistic in this case. Given what we know about the variables in this particular situation, the discrepancy between CFI and RMSEA likely points to the omission of one or more error covariances that should, in fact, be included in the model.

Chapter Six: Conclusion

In the final chapter of this thesis, I focus on two main topics. First, I revisit the questions of interest in this study, as first presented in Chapter 1, and discuss briefly how these questions were addressed. Second, I discuss the limitations of the current study and offer suggestions for possible future research.

6.1 Revisiting the Questions of the Study

In Chapter 1, four questions of interest were listed with the goal of the present study being to address each of these questions. In doing so, the hope was that more information could be gathered regarding how both CFI and RMSEA behave with respect to different modeling conditions as well as different sources of misspecification.

Question 1: To what extent is fit index value affected by the source of the misspecification?

As has been shown in the previous chapters, CFI and RMSEA appear to be sensitive to different sources of misspecification. Specifically, RMSEA seems to be more sensitive to misspecifications due to omitted error covariances, while CFI is more sensitive to misspecifications of the latent structure of a model. This difference in sensitivity suggests that combining the results of CFI and RMSEA when assessing model fit can help a researcher determine the possible source of any misspecification in a given model. Guidelines for interpreting combinations of CFI and RMSEA values were given in Chapter 4, section 3.2.

Question 2: To what extent is the relationship between the degree of model misspecification and fit index value moderated by different model components?

A key component to the present research was including a wide variety of models in terms of different model components. Recall that model components, as defined within this study, are any aspects of the modeling procedure that may affect index value over and above any actual misspecification. The model components of interest in this study included loading size, factor correlation size (in models with two or more latent variables), model complexity, and model balance (in models with two or more latent variables).

As was demonstrated in the previous chapters, both CFI and RMSEA are sensitive to changes in certain model components, regardless of the size or type of model misspecification. While such sensitivities cannot practically be avoided in many research settings (e.g., it may not be ethical to increase the number of indicators in a model in order for RMSEA to show a better fit), it is useful for researchers to know that things such as model complexity and loading sizes may have an effect on the behavior of CFI or RMSEA over and above any effects due to model misspecification.

Question 3: Does the current research support the use of universal cutoff values across different model and misspecification types?

It is clear from the results of this research that specific, universally applied cutoff values for either CFI or RMSEA may not be the most appropriate method of distinguishing between a well-fitting model and a poor-fitting one. While this study did not examine the appropriateness of the most commonly used cutoff values (.95 for CFI, .06 for RMSEA), the wide range of values attained by the fit indices across the various different models suggests that any universally applied cutoff value would, in some cases, either lead a researcher to reject a model with only a slight misspecification or to retain a model with a severe misspecification.

If any guidance can be given with respect to the use of cutoff values, it relates to the influence of loading size on fit index value. In nearly all modeling scenarios presented in this study, loading size affected the sensitivity of both CFI and RMSEA to model misspecification. In particular, lower loadings tended to correspond to less sensitivity to misspecifications of any type, while higher loadings tended to correspond to greater sensitivity to misspecificaitons of any type. Thus, as far as a broad guideline for the use of cutoffs, I recommend the use of stricter cutoffs for models that contain, on average, lower loadings, and more relaxed cutoffs for models that contain, on average, higher loadings.

Question 4: Can guidelines for the use of different indices under different models be developed?

This study revealed that CFI and RMSEA differ in sensitivity to different combinations of misspecification type and model components. Thus, in Chapter 4, a loose set of guidelines was presented on how to interpret the combination of CFI and RMSEA values. In particular, when the two indices disagree (e.g., CFI shows poor fit but RMSEA shows good fit), the guidelines suggest a possible underlying source of misspecification.

While it is not possible in a real-life research setting to determine the actual cause of model misspecification, such guidelines as presented in Chapter 4 allow for a “starting point” as to what a possible source of misspecification might be. This is an improvement over simply relying on the chi-square test statistic’s binary “fit” or “no fit” decision, as it allows for researchers to locate possible points of model readjustment if it is deemed appropriate.

6.2 Limitations of the Current Study and Possible Future Research

The current study is limited in the sense that its focus was solely on confirmatory factor analysis (CFA) models. The methodology of structural equation modeling encompasses a wide range of model types, including path analysis models, structural regression models, and latent change models, among others. Since these other models were not included in the present study, it cannot be said whether the results found here can be generalized beyond CFA models, or even to more complex CFA models (for example, models in which latent factors load onto other latent factors).

While most other research that examines fit index behavior has done so using CFA models, as this study has done, there have been fewer studies that focus on how these indices perform in different types of models, and to my knowledge, no studies that examine performance in as many modeling scenarios as presented here. A possible direction for future research, therefore, would be to examine the performance of CFI and RMSEA in these other types of models, particularly focusing on similar model components (such as model complexity or loading size) and their effects on index behavior.

Another limitation of the present study is the focus on only three sources of model misspecification, and each one in isolation. That is, each source of misspecification—omitted error covariance, omitted cross-loading, or misspecified latent structure—occurred in a model in which there was no other source of misspecification. Combinations of misspecifications, such as a model containing an omitted error covariance and an omitted cross-loading, were not examined here. In reality, it is possible that a model may contain multiple sources of misspecification.

This leads to another direction of possible future research: examining the effects of multiple sources of misspecification on index behavior. Similar models to the ones in this study could be examined, but containing not just one source of misspecification but perhaps two or three. Examining multiple, simultaneous different sources of misspecification is not something that has been examined in much detail in the current literature; thus, studying the effects of multiple sources of misspecification on index behavior is somewhat of a new research direction.

A final limitation of this study worth mentioning is a limitation common to SEM simulations such as those presented here. Because of the way the "hypothesized" and "true" models were constructed in this study, I was able to control the source and severity of the misspecification in every model. Thus, for any given hypothesized model, I knew the exact cause of the misspecification as well as the degree of the misspecification's severity, and was able to generate guidelines on how to best combine CFI and RMSEA to determine possible sources of model misspecification based on the simulation results.

In reality, however, the "true" model is never actually known, and thus a possible limitation of this study is that it may be difficult to generalize my suggested guidelines to real life modeling situations in which the source or sources of misspecification may be more complex.

An attempt to overcome this limitation was made by including models based on actual theories and data in Chapter 5. In these applied cases, the "true" model can never actually be known. However, in Chapter 5, I first examined research-based models that appeared to accurately represent the relationships amongst variables in the population, and then compared these models' fit indices to other models that were, based on theory, a "worse fit" than the well-fitting models. The well-fitting models thus acted as good approximations to the "true" models, and allowed me to compare more misspecified models to them to determine if my suggested guidelines were useful in determining possible sources of misspecification.

Further research could be carried out in a similar fashion: find models in the literature that are well-established and considered to be accurate representations of the population, obtain other possible models that differ from the well-fitting model in certain ways (e.g., a different number of factors, different loadings, etc.), and compare the CFI and RMSEA values of these models to assess how well the guidelines given in this thesis can help aid researchers in assessing the cause of misspecification in hypothesized models.

References

Anderson, J. C. and D. W. Gerbing (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.

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

Beauducel, A. and W. W. Wittmann (2005), "Simulation study on fit indexes in CFA based on data with slightly distorted simple structure." *Structural Equation Modeling: A Multidisciplinary Journal*, 12, 41–75.

Bentler, P. M. *EQS 6 Structural Equations Program Manual.* Encino, CA" Multivariate Software, Inc., 2006. Print.

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

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

Bentler, P. M. and K. H. Yuan (1999), “Structural equation modeling with small samples: Test statistics.” *Multivariate Behavioral Research*, 34, 181–197.

Bollen, K. A. (1989), “A new incremental fit index for general structural equation models.” *Sociological Methods & Research*, 17, 303–316.

Bollen, K. A. and J. S. Long (1992), “Test for structural equation models.” *Sociological Methods & Research*, 21, 123–131.

Boomsma, A. (2000), “Reporting analyses of covariance structures.” *Structural Equation Modeling: A Multidisciplinary Journal*, 7, 461–483.

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

Browne, M. W. and R. Cudeck (1992), “Alternative ways of assessing model fit.” *Sociological Methods & Research*, 21, 230–258.

Chau, H. and D. Hocevar. “The effects of number of measured variables on goodness-of-fit in confirmatory factor analysis.” *Proceedings of the American Educational Research Association, San Francisco, CA, April 1995.*

Chen, F., P. J. Curran, K. A. Bollen, J. Kirby, and P. Paxton (2008), "An empirical evaluatin of the use of fixed cutoff points in RMSEA test statistic in structural equation models." *Sociological Methods & Research*, 36, 462–494.

Crowley, S. L. and X. Fan (1997), "Structural equation modeling: Basic concepts and applications in personality assessment research." *Journal of Personality Assessment*, 68, 508–531.

Cudeck R. and M. W. Browne (1983), "Cross-validation of covariance structures." *Multivariate Behavioral Research*, 18, 147–167.

Curran, P. J., K. A. Bollen, P. Paxton, J. Kirby, and F. Chen (2002), "The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a Monte Carlo simulation." *Multivariate Behavioral Research*, 37, 1–36.

Fan, X. and S. A. Sivo (2005), "Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revised." *Structural Equation Modeling: A Multidisciplinary Journal*, 12, 343–367.

Fan, X. and S. A. Sivo (2007), "Sensitivity of fit indices to model misspecification and model types." *Multivariate Behavioral Research*, 42, 509–529.

Fan, X., B. Thompson, and L. Wang (1999), “Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes.” *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 56–83.

Folnes, N., T. Foss, and U. H. Olsson (2012), “Residuals and the residual-based statistic for testing goodness of fit in structural equation models.” *Journal of Educational and Behavioral Statistics*, 37, 367–386.

Fornell, C. (1983), “Issues in the application of covariance structure analysis: A comment.” *Journal of Consumer Research*, 2, 443–450.

Gerbing, D. W. and J. C. Anderson (1992), “Monte Carlo evaluations of goodness of fit indices for structural equation models.” *Sociological Methods & Research*, 21, 132–160.

Gonzalez, R. and D. Griffin (2001), “Testing parameters in structural equation modeling: Every “one” matters.” *Psychological Methods*, 6, 258–269.

Hayduk, L. (2014), “Seeing perfectly fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse.” *Educational and Psychological Measurement*, 74, 905–926.

Heene, M., S. Hilbert, H. H. Freudenthaler, and M. Bühner (2012), "Sensitivity of SEM fit indices with respect to violations of uncorrelated errors." *Structural Equation Modeling: A Multidisciplinary Journal*, 19, 36–50.

Hooper, D., J. Coughlan, and M. Mullen (2008), "Structural equation modelling: Guidelines for determining model fit." *Electronic Journal of Business Research Methods*, 6, 53–60.

Hsu, H. Y., O. Kwok, J. H. Lin, and S. Acosta (2015), "Detecting misspecified multilevel structural equation models with common fit indices: A Monte Carlo study." *Multivariate Behavioral Research*, 50, 197–215.

Hu, L. and P. M. Bentler (1998), "Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification." *Psychological Methods*, 3, 424–453.

Hu, L., and P. M. Bentler (1999), "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives." *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55.

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

Joreskog, K. G. (1970), "A general method for analysis of covariance structures." *Biometrika*, 57, 239–251.

Jöreskog, K. G., and Dag Sörbom. *LISREL V: Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Methods.* Chicago, IL: National Educational Resources, 1981. Print.

Kenny, D. A., B. Kaniskan, and D. B. McCoach (2015), "The Performance of RMSEA in models with small degrees of freedom." *Sociological Methods & Research*, 44, 486–507.

Kenny, D. A. and D. B. McCoach (2003), "Effect of the number of variables on measures of fit in structural equation modeling." *Structural Equation Modeling: A Multidisciplinary Journal*, 10, 333–351.

Kline, Rex B. *Principles and Practice of Structural Equation Modeling*. 2nd ed. New York: Guilford, 2005. Print.

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

La Du, T. J. and J. S. Tanaka (1995), "Incremental fit index changes for nested structural equation models." *Multivariate Behavioral Research*, 30, 289–316.

MacCallum, R. C., M. W. Browne, and H. M. Sugawara (1996), "Power analysis and determination of sample size for covariance structure modeling." *Psychological Methods*, 1, 130–149.

MacCallum, R. C. and S. Hong (1997), "Power analysis in covariance structure modeling using GFI and AGFI." *Multivariate Behavioral Research*, 32, 193–210.

Mahler, C. "The effects of misspecification type and nuisance variables on the behaviors of population fit indices used in structural equation modeling." MA thesis. University of British Columbia, 2011. Print.

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

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

Marsh, H. W., K. T. Hau, and Z. Wen (2004), "In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings." *Structural Equation Modeling: A Multidisciplinary Journal*, 11, 320–341.

McAuley, E., T. E. Duncan, and D. W. Russell (1992), "Measuring causal attributions: The revised causal dimension scale (CDSII)." *Personality and Social Psychology Bulletin*, 18, 566–573.

McAuley, E., and J. B. Gross (1983), "Perceptions of causality in sport: An application of the Causal Dimension Scale." *Journal of Sport Psychology*, 5, 72–76.

McDonald, R. P. and Ho, M. R. (2002), "Principles and practice in reporting structural equation analyses." *Psychological Methods,* 7, 64–82.

McDonald, R. P. and Marsh, H. W. (1990), "Choosing a multivariate model: Noncentrality and goodness of fit." *Psychological Bulletin*, 107, 247–255.

Miles, J. and M. Shevlin (2007). "A time and place for incremental fit indices." *Personality and Individual Differences*, 42, 869–874.

Moshagen, M. (2012), "The model size effect in SEM: Inflated goodness-of-fit statistics are due to the size of the covariance matrix." *Structural Equation Modeling: A Multidisciplinary Journal*, 19, 86–98.

Perry, J. L., A. R. Nicholls, P. J. Clough, and L. Crust (2015), “Assessing model fit: Caveats and Recommendations for confirmatory factor analysis and exploratory structural equation modeling.” *Measurement in Physical Education and Exercise Science*, 19, 12–21.

Rigdon, E. E. (1996), “CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling.” *Structural Equation Modeling: A Multidisciplinary Journal*, 3, 369–379.

Russell, D. (1982), “The Causal Dimension Scale: A measure of how individuals perceive causes.” *Journal of Personality and Social Psychology*, 42, 1137–1145.

Russell, D. (2002), “In search of underlying dimensions: The use (and abuse) of factor analysis in *Personality and Social Psychology Bulletin.*” *Personality and Social Psychology Bulletin*, 28, 1629–1646.

Russell, D., E. McAuley, and V. Tarico (1987), “Measuring causal attributions for success and failure: A comparison of methodologies for assessing causal dimensions.” *Journal of Personality and Social Psychology*, 52, 1248–1257.

Savalei, V. (2010), “The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models.” *Educational and Psychological Measurement*, 72, 910–932.

Sharma, S., S. Mukherjee, A. Kumar, and W. R. Dillon (2005), "A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research*, 58, 935–943.

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

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

Steiger, J. H. (2007), "Understanding the limitations of global fit assessment in structural equation modeling." *Personality and Individual Differences*, 42, 893–898.

Steiger, J. H. and J. C. Lind. "Statistically-based tests for the number of common factors." *Proceedings of the Annual Meeting of the Psychometric Society, Iowa City, IA, June 1980*. Handout for a talk.

Steiger, J. H., A. Shapiro, and M. W. Browne (1985), "On the multivariate asymptotic distribution of sequential chi-square tests. *Psychometrika*, 50, 253–264.

Sugawara, H. M. and R. C. MacCallum (1993), "Effect of estimation method on incremental fit indexes for covariance structure models." *Applied Psychological Measurement*, 17, 365–377.

Themessl-Huber, M. (2014), "Evaluation of the χ^2-statistic and different fit-indices under misspecified number of factors in confirmatory factor analysis." *Psychological test and Assessment Modeling*, 56, 219–236.

Tucker, L. R. and C. Lewis (1973), "The reliability coefficient for maximum likelihood factor analysis." *Psychometrika*, 38, 1–10.

Weiner, B. (1985), "An attributional theory of achievement motivation and emotion." *Psychological Review*, 92, 548–573.

Wheaton, B., B. Muthen, D. F. Alwin, and G. F. Summers (1977), "Assessing reliability and stability in panel models." *Sociological Methodology*, 8, 84–136.

Yuan, K. H. (2005), "Fit indices versus test statistics." *Multivariate Behavioral Research*, 40, 115–148.

APPENDIX A: NAMES, SAMPLE DEFINITIONS, AND POPULATION DEFINITIONS OF COMMONLY USED FIT INDICES

Index Name(s)	Sample Definition	Population Definition
Comparative Fit Index (CFI)	$\frac{(\chi_I^2 - df_I) - (\chi^2 - df)}{(\chi_I^2 - df_I)}$	$1 - \frac{\hat{F}}{\hat{F}_I}$
Normed Fit Index (NFI) *Bentler-Bonnet Index (BBI)* *BL86* *Bollen's Fit Index* *Δ₁*	$\frac{\chi_I^2 - \chi^2}{\chi_I^2}$	$1 - \frac{\hat{F}}{\hat{F}_I}$
Bollen's Incremental Fit Index *(IFI)* *BL89* *Normed Fit Index 2 (NFI2)* *Δ₂*	$\frac{\chi_I^2 - \chi^2}{\chi_I^2 - df}$	$1 - \frac{\hat{F}_I}{\hat{F}}$
Non-Normed Fit Index (NNFI) *Tucker-Lewis Index (TLI)* *Bentler-Bonnet Non-Normed Fit Index (BBNFI)*	$\frac{\frac{\chi_I^2}{df_I} - \frac{\chi^2}{df}}{\frac{\chi_I^2}{df_I} - 1}$	$1 - \frac{\hat{F}_I}{\hat{F}} \times \frac{df_I}{df}$
Bollen86 *Relative Fit Index (RFI)*	$\frac{\frac{\chi_I^2}{df_I} - \frac{\chi^2}{df}}{\frac{\chi_I^2}{df_I}}$	$1 - \frac{\hat{F}_I}{\hat{F}} \times \frac{df_I}{df}$

Appendix A (Continued)

Index Name(s)	*Sample Definition*	*Population Definition*
Root Mean Square Error of Approximation (RMSEA)	$\sqrt{\frac{\frac{\chi^2}{df} - 1}{n-1}}$	$\sqrt{\frac{\hat{F}}{df}}$
Goodness of Fit Index (GFI)	$1 - \frac{\chi^2}{\chi_I^2}$	$1 - \frac{Tr[\left(\Sigma(\hat{\theta})^{-1}(\Sigma^* - \Sigma(\hat{\theta}))\right)^2]}{Tr((\Sigma(\hat{\theta})^{-1}\Sigma^*)^2)}$
Adjusted Goodness of Fit Index (AGFI)	$1 - GFI \times \left(\frac{df}{df_I}\right)$	$1 - \frac{p(p+1)}{2df} \times (1 - GFI)$
Standardized Root Mean Square Residual (SRMR)	$\sqrt{\frac{2}{n(n+1)} \sum_i^n \sum_j^i (s_{ij} - c_{ij})}$	$\sqrt{\frac{\sum(R^* - R(\hat{\theta}))^2}{p(p+1)}}$
Gamma	$\frac{p}{p + 2(\frac{\chi^2 - df}{n-1})}$	$\frac{p}{p + 2df(RMSEA^2)}$

Note. Where *p* is the number of indicators, *n* is the sample size, χ_I^2 and χ^2 stand for the chi-square values for the independent (baseline) model and the hypothesized model, respectively; df_I and *df* are the degrees of freedom for the independent model and the hypothesized model, respectively; $\hat{F}$ and $\hat{F}_I$ stand for the minimized fit function for the hypothesized and independent models, respectively; R* is the population correlation matrix, and $R(\hat{\theta})$ is the model-predicted correlation matrix.