Title
Model Close Match as a Criterion for Structured Model Comparison and Its Robust
Statistical Tests

Permalink
https://escholarship.org/uc/item/2432d2x7

Authors
Li, Libo
Bentler, Peter M

Publication Date
2006-08-21

Robust Statistical Tests for Evaluating the Hypothesis of Close Fit of Misspecified Mean
and Covariance Structural Models

Libo Li and Peter M. Bentler*

University of California, Los Angeles

June 11, 2006

Robust Statistical Tests for Evaluating the Hypothesis of Close Fit of Misspecified Mean
and Covariance Structural Models

Abstract

Model close fit is one key issue in the mean and covariance structure analysis. In this article, we utilize the latest results on the general distribution of likelihood ratio statistic in this methodology and propose several distribution free root mean square error of approximation (RMSEA) tests for evaluating the hypothesis of close fit of misspecified models. Simulation studies show that three of these tests have robust and desirable performance in spite of severe nonnormality across the examples when sample size is as large as 300. A new two-stage procedure which combines model exact fit tests and the proposed RMSEA tests for model close fit is further proposed for overall model fit evaluation.

Keywords: Likelihood ratio statistic, RMSEA, model close fit, asymptotics, noncentral chi-square distribution, model misspecification.

1. Introduction

In the last thirty years, since the publication of Goldberger and Duncan's (1973) seminal volume on structural equation modeling with latent variables, this methodology has become a well-established research methodology in many disciplines such as biometrics, education, marketing, psychology, and sociology. Illustrative overviews of applications are given by Bollen and Curran (2006), Byrne (2006), Hays, Revicki, and Coyne (2005), Holbert and Stephenson (2002), MacCallum and Austin (2000), Martens (2005), Peek (2000), Penny, Stephan, Mechelli, and Friston (2004), and Smith and Langfield-Smith (2004). A short technical summary is given by Yuan and Bentler (in press), and a complete review of statistical issues is given in the 18 chapters of S.-Y. Lee's (in press) *Handbook of Structural Equation Models.* See also Skrondal and Rabe-Hesketh (2004).

Among the variety of technical developments, methods for evaluating overall model fit are critical for both statistical inference and practical applications. After all, without a validated model, there is not much point to worrying about any kind of specific parameter estimate within a model, such as a path (either within- or between-group paths, or within- or across-level paths), or a variance or covariance. One way of model validation is to look whether the model is correctly specified. Among the variety of model testing procedures, in addition to the classical normal theory based likelihood ratio (NTLR) test, several extensions of Browne's (1984) asymptotically distribution free (ADF) test statistics for mean and covariance structures, such as the Satorra-Bentler scaled test (Satorra & Bentler, 1988, 1994) or residual-based tests (Yuan & Bentler, 1997, 1998, 1999), were developed for this purpose. Simulation studies have shown that these methods can provide valuable information for evaluating model exact fit even when the data is nonnormal with a reasonably small sample size and some missing values (Yuan & Bentler, 1997, 1998, 1999, 2000).

Although model correctness is important, this way of model validation may not be realistic or complete in practice. It is likely that any model that we use in applications of mean and covariance structure analysis in social sciences is nothing more than an approximation to reality (e.g., Bentler & Bonett, 1980; Browne & Cudeck, 1993; de Leeuw, 1988). In discussing models generally, MacCallum (2003, p. 113) notes "All of these models, in their

attempt to provide a parsimonious representation of psychological phenomena, are wrong to some degree and are thus implausible if taken literally." In fact, no specific model may be assumed to exist in population. Except for the saturated model, the population mean or covariance matrix may not be reproducible precisely by any specific group of parameters as assumed in exact fit model evaluation. A true population mean or covariance matrix with some specific structure may only exist in simulation studies. On the other hand, a perfectly fitted model may not be interpretable or parsimonious.

As a consequence, measures of model close fit, namely, so-called fit indices, were developed years ago to assess the degree of fit or misfit of a model, and are often recommended for practical use (e.g., Akaike, 1987; Bentler, 1990; Bentler & Bonett, 1980; Bollen, 1986, 1989; James, Mulaik, & Brett, 1982; Joreskog & Sorbom, 1981; Marsh, Balla & McDonald, 1988; McDonald, 1989; McDonald & Marsh, 1990; Steiger & Lind, 1980; Tanaka, 1987; Tanaka & Huba, 1985; Tucker & Lewis, 1973). Today, model close fit indices are so popular that they are extensively studied by many people, are provided as standard output of most software packages, and are reported in most application articles. Only a few people are very critical of their use (Yuan, 2005).

The popularity of model close fit indices reflects the fact that a model with some misspecification is typical in practice. It also reflects the lack of reliable tests related to a misspecified model. The main barrier to development of such tests is that the general or asymptotic distribution of existing exact fit test statistics under misspecification may be not suitable for the testing of model misspecification or is not well understood. In the literature of mean and covariance structure analysis, no distribution theory exists for some test statistics like the Satorra-Bentler scaled statistic or the Yuan-Bentler residual based ADF test statistic (Yuan & Bentler, 1998). While the ADF test statistic and Browne's residual based ADF test statistic (Browne, 1984) are asymptotically noncentral chi-square distributed, and the Yuan-Bentler F statistic is asymptotically noncentral F distributed under misspecification (e.g., Browne, 1984; Shapiro, 1983; Yuan & Bentler, 1999), the noncentrality parameters of these distributions contain an asymptotic covariance matrix which is based on the distribution of the data and varies with its nonnormality. Such sample dependent noncentrality parameters

clearly makes these distributions not ideal for testing model misspecification. Although the NTLR statistic and the generalized least squares (GLS, Browne, 1974) test statistic do not have this problem, derivations of their noncentral chi-square approximation (e.g., Satorra & Saris, 1985; Steiger, Shapiro, & Browne, 1985) requires the following assumptions

1. the sample size is not too small

2. the discrepancy function is correctly specified for the distribution of the data

3. the population drift assumption (Wald, 1943). That is, the population value of the mean and covariance matrix are regarded as being a function of sample size n and converges at a rate of $O(1/\sqrt{n})$ to a point where the model is satisfied.

The first assumption is common. However, the second one is not consistent with the NTLR statistic or the GLS test statistic when the data is not normally distributed. The third one is also problematic. As a working assumption, it allows mathematical derivations to work literally. But in practice, it is hard to imagine that the population value is sample size dependent. It also hard to implement, e.g., in simulation studies. Notice that standard deviations of sample moments are typically of order $O(1/\sqrt{n})$. An explanation of the population drift assumption is that model misspecification (nonstochastic error) is of the same magnitude as the sampling error (stochastic error). A more understandable explanation that is widely used in the literature is that the model misspecification is not too large. However, there is no theory to tell us what size of model misspecification would allow the noncentral chi-square approximation to be applicable in a specific case. More importantly, such an explanation itself implies that the noncentral chi-square approximation is not a good enough approximation in general.

Although the assumptions of normality and population drift are critical as well as hard to satisfy or verify in practice, this has not prevented the noncentral chi-square distribution to be used for the NTLR statistic and the GLS test statistic in some important issues in mean and covariance structure analysis, such as development of confidence intervals for the root mean square error of approximation (RMSEA) (Browne & Cudeck, 1993), or power analysis (Satorra & Saris, 1985; MacCallum, Browne & Sugawara, 1996; Kano, 2001; Hancock, 2001;

Kim, 2003). However, recently a few simulation studies with normally distributed data have questioned the adequacy of the noncentral chi-square approximation for the NTLR statistic and the GLS test statistic under model misspecification. For example, Curran et al. (2003) found that when the misspecification is large, the confidence interval estimates of RMSEA based on the noncentral chi-square distribution of the NTLR statistic do not perform well. Olsson, Foss and Breivik (2004) further found that NTLR statistic more closely follows a normal distribution than a noncentral chi-square distribution for a large model in spite of the degree of misspecification. On the other hand, when the model is small, the NTLR statistic does follow a noncentral chi-square distribution even with severe misspecifications.

Given the inadequacy of the noncentral chi-square approximation, Yuan, Hayashi and Bentler (2005) applied the theory of Vuong (1989) to mean and covariance structure analysis and derived an ADF-type normal approximation to the NTLR statistic under model misspecification. This approximation holds under some standard regularity conditions and the assumptions of normality and population drift are not needed. Clearly, such an ADF-type approximation should be useful in the general case and may provide a new foundation for studying model misspecification. More importantly, the principle of the likelihood ratio statistic is very general. So the results may be extendable to many other situations such as multigroup data, multilevel data and so on. Although this approximation seems very promising, Yuan, Hayashi and Bentler (2005) did not study the performance of the new statistic in much detail. In their study, only 6 variables were used in the simulation and only normal data were studied. Clearly, such limited evaluation can not give strong support for the practical use of this approximation.

In this article we will first review the relevant statistical theories, especially Vuong's theory, and their application to the NTLR statistic. Then some new ADF-like test statistics for evaluating model misspecification will be proposed. The several statistics are then compared in simulation studies. The results from this comparison will be presented and some suggestions for future studies will be proposed.

2. Theoretical background

In classical single population mean and covariance structure analysis, the simultaneous relationships among p-observed variables in a $p \times 1$ random vector $X = (\mathbf{x}_1, \ldots, \mathbf{x}_p)'$ and m-unobserved factors are hypothesized to depend on q unknown structural parameters included in a $q \times 1$ parameter vector θ. The hypothesized model leads to the model-implied mean $\mu(\theta)$ and covariance matrix $\mathbf{\Sigma}(\theta)$. Now let $\mu = E(X)$, $\mathbf{\Sigma} = \text{cov}(X)$, $\bar{X}$ and $\mathbf{S}$ be the corresponding mean and unbiased sample estimator and $\mathbf{S}^* \equiv (n-1) \cdot \mathbf{S}/n$ be the MLE estimator of $\mathbf{\Sigma}$, where n is the sample size. Let β denote the parameter vector of the saturated model, then in this case $\beta = (\mu', \text{vech}(\mathbf{\Sigma})')'$, where $\text{vech}(\cdot)$ is an operator which transforms a symmetric matrix into a vector by stacking the nonduplicated elements of the matrix. Further, $\hat{\beta}^* \equiv (\bar{X}', \text{vech}(\mathbf{S}^*)')'$ and $\hat{\beta} \equiv (\bar{X}', \text{vech}(\mathbf{S})')'$ will be the MLE and unbiased estimator of β separately. Although there is some difference between these two estimators, such difference will become very slight when the sample size n is large (e.g., Anderson, 1984).

Suppose that the data $X_i = (x_{i1}, \ldots, x_{ip}), i = 1, \ldots, n = N + 1$ are identically and independently drawn from X, then the normal theory based log likelihood function of observations is given by

$$l_n(\beta) = \sum_{i=1}^{n} \log f(X_i; \mu, \mathbf{\Sigma}) = \text{constant} - \frac{n}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)'\mathbf{\Sigma}^{-1}(X_i - \mu)$$

where $f(X_i; \mu, \mathbf{\Sigma})$ is the density function of the multivariate normal distribution for individual observation X_i. Obviously, $\hat{\beta}^*$ is the maximizer of $l_n(\beta)$.

Let μ_0, $\mathbf{\Sigma}_0$ denote the population counterpart to μ, $\mathbf{\Sigma}$ and $\beta_0 \equiv (\mu_0', \text{vech}(\mathbf{\Sigma}_0)')$. Let $\mathbf{\Gamma}$ be the asymptotic covariance matrix of $\hat{\beta}$ and thus $\hat{\beta}^*$, then under some standard regularity conditions (e.g., Kano, 1986; Shapiro, 1984), $\hat{\beta}^*$ and thus $\hat{\beta}$, will be strongly consistent and asymptotically normally distributed, that is,

$$\sqrt{n}(\hat{\beta}^* - \beta_0) \overset{a}{=} \sqrt{n}(\hat{\beta} - \beta_0) \overset{L}{\longrightarrow} N(0, \mathbf{\Gamma}) \tag{1}$$

where $\overset{a}{=}$ refers to asymptotic equality (i.e., the difference between both sides of the equality tends to zero in probability as $n \to \infty$). Further, $\mathbf{\Gamma}$ can be shown to be equal to $A^{-1}(\beta_0)B(\beta_0)A^{-1}(\beta_0)$ (e.g., Vuong, 1989; Yuan & Jennrich, 1998) and

$$A(\beta_0) = -E\left[\frac{\partial^2 l_i(\beta_0)}{\partial\beta_0\partial\beta_0'}\right] \qquad B(\beta_0) = E\left[\frac{\partial l_i(\beta_0)}{\partial\beta_0}\frac{\partial l_i(\beta_0)}{\partial\beta_0'}\right]$$

where $E(\cdot)$ denotes the expectation with respect to the true distribution of X.

When μ and Σ are parameterized as $\mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$, then the log likelihood function becomes

$$l_n(\theta) = \sum_{i=1}^{n} \log f(X_i; \theta) = \text{constant} - \frac{n}{2}\log|\Sigma(\theta)| - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu(\theta))'\Sigma^{-1}(\theta)(X_i - \mu(\theta))$$

An estimator of θ can be obtained by minimizing the well-known normal theory maximum likelihood discrepancy function (Browne & Arminger, 1995)

$$\begin{aligned} F_{ML}(\bar{X}, \mathbf{S}^*; \mu(\theta), \Sigma(\theta)) &= (\bar{X} - \mu(\theta))'\Sigma^{-1}(\theta)(\bar{X} - \mu(\theta)) + \log|\Sigma(\theta)| + \text{tr}(\mathbf{S}^*\Sigma^{-1}(\theta)) \\ &\quad -\log|\mathbf{S}^*| - p \end{aligned}$$

The minimizer, $\hat{\theta}_{NML}$, is the maximum likelihood estimator and the NTLR test statistic T_{NML} is given by

$$T_{NML} = nF_{ML}(\bar{X}, \mathbf{S}^*; \mu(\hat{\theta}_{NML}), \Sigma(\hat{\theta}_{NML})) = 2\sum_{i=1}^{n} \log\left[\frac{f(X_i; \hat{\beta}^*)}{f(X_i; \hat{\theta}_{NML})}\right] \equiv 2LR_n(\hat{\beta}^*, \hat{\theta}_{NML}) \quad (2)$$

In practice, the discrepancy function $F_{ML}(\bar{X}, \mathbf{S}; \mu(\theta), \Sigma(\theta))$ is used in most literature instead of $F_{ML}(\bar{X}, \mathbf{S}^*; \mu(\theta), \Sigma(\theta))$. Its minimizer, $\hat{\theta}_{ML}$, and the corresponding NTLR statistic $T_{ML} = NF_{ML}(\bar{X}, \mathbf{S}; \mu(\hat{\theta}_{ML}), \Sigma(\hat{\theta}_{ML}))$ are given in the standard output of typical software packages (e.g., EQS, Bentler 2006; Mplus, Muthen & Muthen 2003). Although there is some difference between $\hat{\theta}_{NML}$ and $\hat{\theta}_{ML}$, and between T_{NML} and T_{ML}, such differences will become very slight when the sample size n increases to large (e.g., Bentler, 2006; Browne & Arminger, 1995).

Now suppose that the model is correctly specified, or in other words, the null hypothesis of model exact fit $H_0^E : \mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$ holds, then there exists a θ_0 by which $\mu_0 = \mu(\theta_0)$, $\Sigma_0 = \Sigma(\theta_0)$ and $F_{ML}(\mu_0, \Sigma_0; \mu(\theta_0), \Sigma(\theta_0)) = 0$. Further, under the null hypothesis of model exact fit H_0^E and some standard regularity conditions (e.g., Kano, 1986; Shapiro, 1984), $\hat{\theta}_{NML}$, thus $\hat{\theta}_{ML}$, will be strongly consistent and asymptotically normally distributed (Vuong, 1989; Yuan & Jennrich, 1998), and

$$\sqrt{n}(\hat{\theta}_{ML} - \theta_0) \stackrel{a}{=} \sqrt{n}(\hat{\theta}_{NML} - \theta_0) \stackrel{L}{\longrightarrow} N(0, \Omega_{\theta_0})$$

where $\Omega_{\theta_0} = A^{-1}(\theta_0)B(\theta_0)A^{-1}(\theta_0)$. When normality is assumed, then $\Omega_{\theta_0} = A^{-1}(\theta_0)$ and the NTLR statistics T_{NML} and T_{ML} are chi-square distributed with $df = p^* - q$, where $p^* = p(p+3)/2$.

3. Model misspecification and likelihood ratio statistic

There are two kinds of model misspecifications: overparameterized model misspecification and underparameterized model misspecification. In an overparameterized model, more structural parameters are used than necessary. The problem of such misspecification is parameter redundancy. The NTLR test can be used for this problem when the data is normally distributed. The Wald test and a scaled difference chi-square test by Satorra and Bentler (2001) can handle this problem in more general situations. Since this type of overparameterization is not a "real" misspecification in some sense, in the text below "misspecification" only refers to the second misspecification: an underparameterized model misspecification. In this case, the null hypothesis of model exact fit H_0^E does not hold any more and there are two possibilities. In one possibility, β_0 is attainable and can be reproduced by a group of parameters θ_0. However, the model parameterized by such θ_0 may be hard to find or has no interpretability or substantive usefulness. In another possibility, β_0 is unattainable and can not be reproduced by any group of parameters, or in other words, $F_{ML}(\mu_0, \mathbf{\Sigma}_0; \mu(\theta), \mathbf{\Sigma}(\theta)) > 0$ for any possible value of θ except in a saturated model where θ_0 is equal to, or is some transformation of, β_0.

No matter of what kind of possibility it is, suppose now we have a model parameterized by a $q \times 1$ structural parameter vector θ as we assumed before. Due to the misspecification (parameter underrepresentation), the minimizer θ_* of $F_{ML}(\mu_0, \mathbf{\Sigma}_0; \mu(\theta), \mathbf{\Sigma}(\theta))$ differs from θ_0 in value and maybe in dimension too (see Yuan & Bentler, 2004; Yuan, Marshall & Bentler, 2003 for details). It also has been shown that under the same regularity conditions, the $\hat{\theta}_{NML}$, thus $\hat{\theta}_{ML}$, will be strongly consistent for θ_* and be asymptotically normally distributed (Arminger & Schoenberg, 1989; Vuong, 1989; Yuan & Jennrich, 1998), and

$$\sqrt{n}(\hat{\theta}_{ML} - \theta_*) \stackrel{a}{=} \sqrt{n}(\hat{\theta}_{NML} - \theta_*) \stackrel{L}{\longrightarrow} N(0, \Omega_{\theta_*}) \tag{3}$$

where $\Omega_{\theta_*} = A^{-1}(\theta_*)B(\theta_*)A^{-1}(\theta_*)$. When normality is also assumed, then $\Omega_{\theta_*} = A^{-1}(\theta_*)$

Now let $F_0 = F_{ML}(\mu_0, \boldsymbol{\Sigma}_0; \mu(\theta_*), \boldsymbol{\Sigma}(\theta_*))$. When the model is misspecified, it is well-known that the NTLR statistic $T_{ML} \xrightarrow{L} \chi^2_{df}(NF_0)$ under the assumption of normality and the assumption of population drift, that is,

$$\mu_0 - \mu(\theta_*) = O(1/\sqrt{n}) \text{ and } \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}(\theta_*) = O(1/\sqrt{n}) \tag{4}$$

Let $\sigma_* = (\mu(\theta_*)', \text{vech}(\boldsymbol{\Sigma}(\theta_*))')'$ and $\dot{\sigma}_* = \partial \sigma_* / \partial \theta'_*$. Let $\mathbf{D}_p$ be the duplication matrix as defined by Magnus and Neudecker (1988), $\mathbf{W} = \text{diag}\left[\boldsymbol{\Sigma}_0^{-1}, 2^{-1}\mathbf{D}'_p(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\mathbf{D}_p\right]$, $\mathbf{W}_* = \text{diag}\left[\boldsymbol{\Sigma}^{-1}(\theta_*), 2^{-1}\mathbf{D}'_p(\boldsymbol{\Sigma}^{-1}(\theta_*) \otimes \boldsymbol{\Sigma}^{-1}(\theta_*))\mathbf{D}_p\right]$ and $\mathbf{U} = \mathbf{W}_* - \mathbf{W}_* \dot{\sigma}_* (\dot{\sigma}_*' \mathbf{W}_* \dot{\sigma}_*)^{-1} \dot{\sigma}_*' \mathbf{W}_*$. Then it has been shown that under the population drift assumption (4) (Yuan & Marshall, 2004),

$$\mathbf{AE}(T_{ML}) = NF_0 + \text{tr}(\mathbf{U}\boldsymbol{\Gamma}) \tag{5}$$

where $\mathbf{AE}$ represents the asymptotic expectation with respect to the true distribution of X. When normality is assumed, (5) reduces to

$$\mathbf{AE}(T_{ML}) = NF_0 + df \tag{6}$$

Theorem 1. Under the Assumptions A1-A5 of Vuong (1989),

$$\mathbf{AE}(T_{NML}) = nF_0 + \text{tr}(A^{-1}(\beta_0)B(\beta_0) - A^{-1}(\theta_*)B(\theta_*))$$

where $\text{tr}(A^{-1}(\beta_0)B(\beta_0) - A^{-1}(\theta_*)B(\theta_*))$ reduces to $\text{tr}(\mathbf{U}\boldsymbol{\Gamma})$ if the population drift assumption (4) is assumed or a correct model is specified and reduces to df if normality is assumed.
Proof. see the Appendix.

Yuan, Hayashi and Bentler (2005) applied the theory of Vuong (1989) to mean and covariance structure analysis and derived the asymptotic distribution of T_{ML} under the alternative hypothesis of model exact fit without the assumptions of normality and population drift. Given the unfamiliarity of Vuong's theory in the literature of mean and covariance structure analysis, we give a brief explanation of Vuong's theory and its application first.

The theory of Vuong (1989) focuses on T_{NML} instead of T_{ML}. In (2), we notice that

$$\frac{1}{2n}T_{NML} = \frac{1}{n}\sum_{i=1}^{n} \log\left[\frac{f(X_i; \hat{\beta}^*)}{f(X_i; \hat{\theta}_{NML})}\right] \equiv LR_n(\hat{\beta}^*, \hat{\theta}_{NML}) \tag{7}$$

8

As mentioned above, $\hat{\theta}_{NML}$ is strongly consistent to θ_* and $\hat{\theta}_{NML} - \theta_* = O_p(1/\sqrt{n})$ under the misspecification and some standard regularity conditions. Another fact is that by (1) $\hat{\beta}^*$ is also strongly consistent to β_0 and $\hat{\beta}^* - \beta_0 = O_p(1/\sqrt{n})$ in spite of the misspecification. Given these properties of $\hat{\beta}^*$ and $\hat{\theta}_{NML}$, by using a Taylor expansion of $LR_n(\beta_0, \theta_*)$ around $(\hat{\beta}^{*\prime}, \hat{\theta}'_{NML})'$, we can get

$$
\begin{aligned}
\frac{1}{n} & LR_n(\beta_0, \theta_*) \\
& \equiv \frac{1}{n} \sum_{i=1}^{n} \log \left[\frac{f(X_i; \beta_0)}{f(X_i; \theta_*)} \right] \quad\quad\quad\quad\quad\quad\quad\quad\quad (8) \\
& = \frac{1}{n} \sum_{i=1}^{n} \log \left[\frac{f(X_i; \hat{\beta}^*)}{f(X_i; \hat{\theta}_{NML})} \right] + \frac{1}{n} \cdot V \cdot \left[(\beta_0', \theta_*')' - (\hat{\beta}^{*\prime}, \hat{\theta}'_{NML})' \right] + O_p(1/n)
\end{aligned}
$$

where

$$
V = \partial \sum_{i=1}^{n} \log \left[\frac{f(X_i; \hat{\beta}^*)}{f(X_i; \hat{\theta}_{NML})} \right] \Bigg/ \partial(\hat{\beta}^{*\prime}, \hat{\theta}'_{NML})
$$

Since $\hat{\beta}^*$ and $\hat{\theta}_{NML}$ are MLE estimators, $V = 0$. Then by some algebra, we get

$$
\frac{1}{n} \sum_{i=1}^{n} \log \left[\frac{f(X_i; \hat{\beta}^*)}{f(X_i; \hat{\theta}_{NML})} \right] = \frac{1}{n} \sum_{i=1}^{n} \log \left[\frac{f(X_i; \beta_0)}{f(X_i; \theta_*)} \right] + O_p(1/n) \quad\quad (9)
$$

Now assume that X_i is i.i.d. sampled from X, then $\log \left[f(X_i; \beta_0) / f(X_i; \theta_*) \right]$ is also i.i.d. sampled from some unknown distribution H. By the Law of Large Numbers,

$$
\frac{1}{n} \sum_{i=1}^{n} \log \left[\frac{f(X_i; \beta_0)}{f(X_i; \theta_*)} \right] \xrightarrow{a.s.} E \left[\log \left[\frac{f(X_i; \beta_0)}{f(X_i; \theta_*)} \right] \right]
$$

The term on the right side of the equation is the Kullback-Leibler (1951) Information Criterion in statistical theory. Suppose $E \left[\log \left[f(X_i; \beta_0) / f(X_i; \theta_*) \right] \right]^2$ is finite, then the second central moment of this unknown distribution H is

$$
\omega^2 = E \left[\log \left[\frac{f(X_i; \beta_0)}{f(X_i; \theta_*)} \right] \right]^2 - \left[E \left[\log \left[\frac{f(X_i; \beta_0)}{f(X_i; \theta_*)} \right] \right] \right]^2
$$

By the Central Limit Theorem and (8),

$$
\sqrt{n} \left\{ \frac{1}{n} LR_n(\beta_0, \theta_*) - E \left[\log \left[\frac{f(X_i; \beta_0)}{f(X_i; \theta_*)} \right] \right] \right\} \xrightarrow{L} N(0, \omega^2) \quad\quad (10)
$$

Combining (7), (9) and (10), we obtain

$$
\sqrt{n} \left\{ \frac{1}{2n} T_{NML} - E \left[\log \left[\frac{f(X_i; \beta_0)}{f(X_i; \theta_*)} \right] \right] \right\} \xrightarrow{L} N(0, \omega^2) \quad\quad (11)
$$

9

One point which should be mentioned is that this asymptotic approximation holds only when $\omega^2 \neq 0$. Vuong (1989) showed that $\omega^2 = 0$ and $f(X_i; \beta_0) = f(X_i; \theta_*)$ in (11) are equivalent in general (see Lemma 4.1 by Vuong). For nested models, like the saturated model and a nested structural model, Vuong (1989) further showed that $f(X_i; \beta_0) = f(X_i; \theta_*)$ and $\beta_0 = \beta(\theta_*)$ are equivalent to each other under standard regularity conditions (see Lemma 7.1 by Vuong). So the rejection of model exact fit: $\beta_0 = \beta(\theta_*)$ which is equivalent to $f(X_i; \beta_0) = f(X_i; \theta_*)$ is the way to establish $\omega^2 \neq 0$ and model misspecification is necessary for the use of (11).

Lemma 1. The following identity holds

$$E\left[\log\left[\frac{f(X_i; \beta_0)}{f(X_i; \theta_*)}\right]\right] = \frac{1}{2}F_0$$

Proof.

$$
\begin{aligned}
&E\left[\log\left[\frac{f(X_i; \beta_0)}{f(X_i; \theta_*)}\right]\right] \\
&= E\left[\log f(X_t; \beta_0)\right] - E\left[\log f(X_i; \theta_*)\right] \\
&= \frac{1}{2}E\left[\log|\Sigma(\theta_*)| + (X_t - \mu(\theta_*))'\Sigma^{-1}(\theta_*)(X_t - \mu(\theta_*))\right] \\
&\quad - \frac{1}{2}E\left[\log|\Sigma_0| + (X_t - \mu_0)'\Sigma_0^{-1}(X_t - \mu_0)\right] \\
&= \frac{1}{2}\left[\log|\Sigma(\theta_*)| + \mathrm{tr}(\Sigma_0\Sigma^{-1}(\theta_*)) + (\mu_0 - \mu(\theta_*))'\Sigma^{-1}(\theta_*)(\mu_0 - \mu(\theta_*))\right. \\
&\quad \left. -\log|\Sigma_0| - p\right] \\
&= \frac{1}{2}F_0
\end{aligned}
$$

Let $\hat{F}_{ML} \equiv \frac{1}{N}T_{ML} = F_{ML}(\bar{X}, \mathbf{S}; \mu(\hat{\theta}_{ML}), \Sigma(\hat{\theta}_{ML}))$. Then by (5), (6), (11), Lemma 1 and the asymptotic equivalence between T_{NML} and T_{ML}, we obtain the following corollary (Yuan, Hayashi & Bentler, 2005, Corollary 2 and 3)

Corollary 1. Under standard regularity conditions as in Yuan and Bentler (1997) and model misspecification,

$$\sqrt{n}\left(\hat{F}_{ML} - F_0 - \frac{\mathrm{tr}(\mathbf{U}\boldsymbol{\Gamma})}{n}\right) \xrightarrow{L} N(0, 4\omega^2)$$

when normality is assumed, this reduces to

$$\sqrt{n}\left(\hat{F}_{ML} - F_0 - \frac{df}{n}\right) \xrightarrow{L} N(0, 4\omega^2)$$

Notice that Corollary 1 has no conflict with (11) because the extra term $\text{tr}(\mathbf{U\Gamma})/n$ and df/n in Corollary 1 approach zero as n goes to infinity. Vuong (1989) further gave a consistent estimator of ω^2, that is,

$$\hat{\omega}^2_{\text{Vuong}} = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left[\frac{f(X_i;\hat{\beta}^*)}{f(X_i;\hat{\theta}_{NML})}\right]\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{f(X_i;\hat{\beta}^*)}{f(X_i;\hat{\theta}_{NML})}\right]\right]^2$$

Given the asymptotic equivalence between $\hat{\theta}_{NML}$ and $\hat{\theta}_{ML}$, we obtain the following estimator of ω^2, that is,

$$\hat{\omega}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left[\frac{f(X_i;\hat{\beta})}{f(X_i;\hat{\theta}_{ML})}\right]\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{f(X_i;\hat{\beta})}{f(X_i;\hat{\theta}_{ML})}\right]\right]^2 \tag{12}$$

Clearly, $\hat{\omega}^2$ is also a consistent estimator of ω^2.

Yuan, Hayashi and Bentler (2005) further derived the explicit form of ω^2 under various conditions and gave the corresponding estimators. Although their work is valuable, the preliminary results from a simulation study of normal data show that there is no big difference in performance between their estimators and $\hat{\omega}^2$ in (12). More importantly, their estimators are limited to single group mean and covariance structure analysis and are not as general as $\hat{\omega}^2$. So in this article, we only use $\hat{\omega}^2$ for the tests that follow.

4. RMSEA and tests of close fit

As mentioned before, in the literature of mean and covariance structure analysis, many fit indices have been proposed as measures of model misspecification. Despite their popularity, most of them remain at the level of descriptive statistics of a sample. It does not seem possible to use them for inference about misspecification in the population.

There is one well-known exception: RMSEA. The population RMSEA was proposed and defined as (Steiger & Lind, 1980; Browne & Cudeck, 1993) as

$$\text{RMSEA}_0 = \sqrt{\frac{F_0}{df}}$$

where RMSEA_0 denotes its true value. Clearly, RMSEA_0 is greater than or equal to zero and will be zero only if the model is correctly specified. On account of the asymptotic bias

of T_{ML} as the estimator of NF_0 in (6), the sample RMSEA is defined as (Steiger & Lind, 1980; Browne & Cudeck, 1993)

$$\widehat{\text{RMSEA}} = \sqrt{\max(\hat{F}_{ML}/df - 1/n, 0)}$$

Notice that in the parenthesis we use $1/n$ instead of $1/N$, used by most people. However, the difference between $1/n$ and $1/N$ will be tiny as n goes to infinity, so this is not a big problem. As mentioned before in (5), the asymptotic bias of T_{ML} differs from df in nonnormal conditions. So it is not hard to define a robust sample RMSEA as follows

$$\widetilde{\text{RMSEA}} = \sqrt{\max\left(\frac{\hat{F}_{ML}}{df} - \frac{\text{tr}(\hat{\mathbf{U}}\hat{\mathbf{\Gamma}})}{n \cdot df}, 0\right)}$$

where $\hat{\mathbf{\Gamma}}$ is the consistent estimator of $\mathbf{\Gamma}$ (e.g., Bentler, 2006) and $\hat{\mathbf{U}}$ is a consistent estimator of $\mathbf{U}$ obtained by replacing θ_* by $\hat{\theta}_{ML}$.

As stated before, misspecifications are inevitable in practice. So it is unrealistic to ask RMSEA_0, the measure of misspecification, to be zero. Instead, we set a small positive value a for RMSEA_0. If RMSEA_0 is less than or equal to such a small value, then we can say the model is closely fitted and still acceptable, even with some minor misspecification. Otherwise, the model will be rejected. As a "rule of thumb", the value of .05 was suggested as such a cutoff value of close fit and is widely used in the literature.

As a population value, RMSEA_0 is unknown in practice. So we need a statistic to test the null hypothesis $H_0 : \text{RMSEA}_0 \leq a$ against its alternative one $H_1 : \text{RMSEA}_0 > a$ and decide on rejection or acceptance of the model. The null and alternative hypotheses above can also be written as $H_0 : F_0 \leq df \cdot a^2$ against $H_1 : F_0 > df \cdot a^2$. Let $\chi^2_{df,.95}(N \times df \times a^2)$ be the 95 percent quantile of the noncentral chi-square distribution with NCP equal to $N \times df \times a^2$ and the degree of freedom df. Suppose now $H_0 : F_0 \leq df \cdot a^2$ is true and $T_{ML} \xrightarrow{L} \chi^2_{df}(NF_0)$, then $\Pr[T_{ML} > \chi^2_{df,.95}(N \times df \times a^2)] \xrightarrow{L} .05$. So a test of close fit can be proposed. The null hypothesis will be rejected in favor of the alternative if T_{ML} is greater than $\chi^2_{df,.95}(N \times df \times a^2)$. Otherwise, the null hypothesis can not be rejected.

The close fit test proposed here actually is the same as one proposed by Browne and Cudeck (1993). But they use a different terminology. The obvious drawback of this close fit

test is that it is based on a noncentral chi-square distribution. It works only in normal data with a misspecification satisfying the population drift assumption.

In the last section, the asymptotic distribution of T_{ML} under model misspecification is derived in Corollary 1 and a consistent estimator of ω^2 is given in (12). Based on these results, the following two test statistics are proposed: Vuong's test statistic (T_1) which is:

$$T_1 = \frac{\sqrt{n}\left(\hat{F}_{ML} - df \cdot a^2 - df/n\right)}{2\hat{\omega}_1}$$

and the robust Vuong's test statistic (T_2) which is

$$T_2 = \frac{\sqrt{n}\left(\hat{F}_{ML} - df \cdot a^2 - \mathrm{tr}(\hat{\mathbf{U}}\hat{\mathbf{\Gamma}})/n\right)}{2\hat{\omega}_1}$$

Corollary 2. Under some standard regularity conditions as in Yuan and Bentler (1997) and model misspecification ($\mathrm{RMSEA}_0 \neq 0$),

$$T_1 \overset{a}{=} T_2 \overset{L}{\longrightarrow} N(\sqrt{n}\delta_1, 1) \qquad \text{and} \qquad \delta_1 = \frac{df}{\omega} \cdot \left[\frac{a_0^2 - a^2}{2}\right]$$

where a_0 is the value of RMSEA_0. Further,

1. When $\mathrm{RMSEA}_0 = a$, then $\delta_1 = 0$ and $T_1 \overset{a}{=} T_2 \overset{L}{\longrightarrow} N(0,1)$.

2. When $\mathrm{RMSEA}_0 > a$, then $\delta_1 > 0$ and $T_1 \overset{a}{=} T_2 \longrightarrow +\infty$ as $n \longrightarrow +\infty$.

3. When $0 < \mathrm{RMSEA}_0 < a$, then $\delta_1 < 0$ and $T_1 \overset{a}{=} T_2 \longrightarrow -\infty$ as $n \longrightarrow +\infty$.

Although the Corollary above states that T_1 and T_2 are asymptotically equivalent, they differ from each other in nonnormal data especially when n is small. Let $\lambda_{.95}$ be 95 percent quantile of the standard normal distribution. Then under model misspecification and the null hypothesis $H_0 : \mathrm{RMSEA}_0 \leq a$, $\Pr[T_1 \text{ or } T_2 > \lambda_{.95}] \overset{L}{\longrightarrow} .05$. Clearly, T_1 and T_2 can be used to test the hypothesis of close fit for a misspecified model. For each, the null hypothesis $H_0 : \mathrm{RMSEA}_0 \leq a$ will be rejected if it is greater than $\lambda_{.95}$. Otherwise, it can not be rejected.

Although the theory of Vuong (1989) has been in the literature for a long time and has been used in several areas (see Golden, 2003), and it has a nice distribution free property theoretically, it is unknown in psychometrics and to practical users. One explanation for

the unpopularity of Vuong's method is that it is an asymptotic result. As noticed by Clarke (2003, 2005), the convergence rate of Vuong's test is very slow. A very large sample is needed for satisfactory performance. Clearly, such slow convergence inevitably will prevent it from practical use in psychology, in which small or medium sized samples are common. Although Yuan, Hayashi, and Bentler (2005) reported a desirable performance of Vuong's tests with a small sample size, the examples they used have only 6 observed variables and were only studied in normal data. Clearly, such limited simulation is inadequate to establishing Vuong's test for practical use.

In mainstream statistics, power transformations have a long history as methods for improving the normality of a statistic. For example, Wilson and Hilferty (1931) applied the cube root transformation to χ^2 random variables for improvement towards normality. Recently Chen and Deo (2004) developed a method of determining an appropriate power transformation to improve a normal approximation in small samples. However, they assumed that population parameters such as mean and variance are known. So their methods still need improvement before practical application.

Given the unavailability of a general method to determine the optimal power transformation for a specific statistic, and the popularity of RMSEA in mean and covariance structure analysis, we make use of a square root transformation to hopefully improve the performance of the above tests. This approach makes sense since RMSEA_0 is a square root transformation of the standard NCP F_0 divided by a constant $\sqrt{df}$. Now by Corollary 1 and the Delta method, we obtain the following approximation

Corollary 3. Under some standard regularity conditions as in Yuan and Bentler (1997) and model misspecification ($\text{RMSEA}_0 \neq 0$),

$$\sqrt{n}\left(\widehat{\text{RMSEA}} - \text{RMSEA}_0\right) \xrightarrow{L} N(0, \frac{\omega^2}{df \times F_0})$$

or for the general case,

$$\sqrt{n}\left(\widetilde{\text{RMSEA}} - \text{RMSEA}_0\right) \xrightarrow{L} N(0, \frac{\omega^2}{df \times F_0})$$

14

Proof.

$$\sqrt{n}\left(\widehat{\mathrm{RMSEA}} - \mathrm{RMSEA}_0\right) = \sqrt{n}\left(\sqrt{\max(\hat{F}_{ML}/df - 1/n, 0)} - \sqrt{F_0/df}\right)$$
$$\overset{a}{=} \sqrt{n}\left(\sqrt{\hat{F}_{ML}/df} - \sqrt{F_0/df}\right)$$
$$\overset{L}{\longrightarrow} N(0, \frac{\omega^2}{df \times F_0}) \qquad \text{(Delta method)}$$

The property of $\widetilde{\mathrm{RMSEA}}$ can be proved in the same way.

Clearly, by Corollary 3, two new test statistics can be defined as follows: the RMSEA test statistic T_3 which is

$$T_3 = \frac{\sqrt{n}\left(\widehat{\mathrm{RMSEA}} - a\right)}{\hat{\omega}\left/\sqrt{df \times (\hat{F}_{ML} - df/n)}\right.}$$

and the robust RMSEA test statistic T_4 which is

$$T_4 = \frac{\sqrt{n}\left(\widetilde{\mathrm{RMSEA}} - a\right)}{\hat{\omega}\left/\sqrt{df \times (\hat{F}_{ML} - df/n)}\right.}$$

Now let $\hat{c} = (\mathrm{tr}(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}}) - df)/n$, then we further define another two RMSEA test statistics T_5 and T_6 as

$$T_5 = \frac{\sqrt{n}\left(\widetilde{\mathrm{RMSEA}} - a\right)}{\sqrt{\hat{\omega}^2 - \hat{c}}\left/\sqrt{df \times (\hat{F}_{ML} - df/n + \hat{c})}\right.}$$

and

$$T_6 = \frac{\sqrt{n}\left(\widetilde{\mathrm{RMSEA}} - a\right)}{\sqrt{\hat{\omega}^2 - 2.5 \cdot \hat{c}}\left/\sqrt{df \times (\hat{F}_{ML} - df/n + \hat{c})}\right.}$$

Clearly, $\hat{c}$ is an estimator of $c_0 \equiv (\mathrm{tr}(\mathbf{U}\boldsymbol{\Gamma}) - df)/n$ and converges to c_0 in the order of $O_p(n^{-3/2})$. When the data is normally distributed, c_0 is equal to zero and $\hat{c}$ will converge to zero in the order of $O_p(n^{-3/2})$. So in this condition, T_3, T_4, T_5 and T_6 should have similar performance. When the data is nonnormal, c_0 and thus $\hat{c}$ carry the information on nonnormality of the data. So compared to T_3, T_4 has a correction in numerator and T_5 and T_6 have a correction both in numerator and denominator. Even though such corrections should not matter asymptotically, they may make a difference in performance with small samples.

15

Another point which should be mentioned here is that when one or several quantities among $\hat{F}_{ML} - df/n$, $\hat{F}_{ML} - df/n + \hat{c}$, $\hat{\omega}^2 - \hat{c}$ and $\hat{\omega}^2 - 2.5 \cdot \hat{c}$ are less than or equal to zero due to conditions such as a small sample size, then the corresponding test statistics T_3, T_4, T_5 or T_6 will be undefined respectively. So during the simulations below, replications with such problems will be discarded and the number of these replications will be reported in Tables.

Corollary 4. Under some standard regularity conditions as in Yuan and Bentler (1997) and model misspecification (RMSEA$_0 \neq 0$),

$$T_3 \stackrel{a}{=} T_4 \stackrel{a}{=} T_5 \stackrel{a}{=} T_6 \stackrel{L}{\longrightarrow} N(\sqrt{n}\delta_2, 1) \qquad \text{and} \qquad \delta_2 = \frac{df}{\omega} \cdot \left(a_0^2 - a \cdot a_0\right)$$

Further,

1. When RMSEA$_0 = a$, then $\delta_2 = 0$ and $T_3 \stackrel{a}{=} T_4 \stackrel{a}{=} T_5 \stackrel{a}{=} T_6 \stackrel{L}{\longrightarrow} N(0, 1)$.

2. When RMSEA$_0 > a$, then $\delta_2 > 0$ and $T_3 \stackrel{a}{=} T_4 \stackrel{a}{=} T_5 \stackrel{a}{=} T_6 \longrightarrow +\infty$ as $n \longrightarrow +\infty$.

3. When $0 < \text{RMSEA}_0 < a$, then $\delta_2 < 0$ and $T_3 \stackrel{a}{=} T_4 \stackrel{a}{=} T_5 \stackrel{a}{=} T_6 \longrightarrow -\infty$ as $n \longrightarrow +\infty$.

Clearly, like T_1 and T_2 discussed before, T_3, T_4, T_5 or T_6 can be used to test the hypothesis of close fit for a misspecified model too. For each of them, the null hypothesis $H_0 : \text{RMSEA}_0 \leq a$ will be rejected if it is greater than $\lambda_{.95}$. Otherwise, it can not be rejected.

Corollary 5. Under $H_1 : \text{RMSEA}_0 > a$ and some standard regularity conditions as in Yuan and Bentler (1997), then T_3, T_4, T_5 and T_6 have more asymptotic power than T_1 and T_2 to detect the overmisspecification.

Proof. By Corollary 2 and 4,

$$T_1 \stackrel{a}{=} T_2 \stackrel{L}{\longrightarrow} N(\sqrt{n}\delta_1, 1) \qquad\qquad T_3 \stackrel{a}{=} T_4 \stackrel{a}{=} T_5 \stackrel{a}{=} T_6 \stackrel{L}{\longrightarrow} N(\sqrt{n}\delta_2, 1)$$

where

$$\delta_1 = \frac{df}{\omega} \cdot \left[\frac{a_0^2 - a^2}{2}\right] \qquad\qquad \delta_2 = \frac{df}{\omega} \cdot \left(a_0^2 - a \cdot a_0\right)$$

Clearly, $\delta_2 > \delta_1$ when $a_0 > a > 0$.

5. Simulation studies

In the sections above, we discussed seven test statistics and their corresponding critical values for testing. They are T_{ML}, T_1, T_2, T_3, T_4, T_5 and T_6. In order to establish these statistics as reliable tools for testing $H_0 : \text{RMSEA}_0 \leq a$, we first need to look at the asymptotic approximation and one-sided type I errors of these statistics when $\text{RMSEA}_0 = a$. It is hard to manipulate the level of RMSEA_0 to a specific value a such as the widely accepted value $a = .05$. Instead, we set a equal to the value of RMSEA_0 for all statistics since RMSEA_0 is known in a simulation study. Thus, for each statistic, if it has a desirable approximation to the corresponding theoretical distribution and its exceedance probability over the 95 percent quantile of that distribution is close to .05 across conditions, then it can be suggested as a reliable test of the hypothesis of close fit. Otherwise, it can not be used.

Since the statistics we proposed are asymptotically distribution free, we generated data under three distribution conditions for each of two examples below. They are: normal, mild nonnormal and severe nonnormal. In the mild nonnormal condition, the skewness and kurtosis of each observed variable is set to 1.0 and 3.0 during the data generation. In the severe nonnormal condition, they are set to 2.0 and 7.0. For our two examples, the sample size levels are set to 150, 300, 400, 500, 1000 and 2500. So there are $3 \times 6 = 18$ data conditions for each example. The number of replications is set to 2000 under each data condition.

The whole data generation and analysis were conducted by using EQS 6.1 (Bentler, 2006). In addition, we specified SE=OBS during the analysis. Thus, the term $(\hat{\sigma}_*'\hat{\mathbf{W}}_*\hat{\sigma}_*)$, the Fisher information estimator, in $\hat{\mathbf{U}}$ is replaced by the estimator of the Hessian or observed information matrix.

Example 1. This example is a factor model used by Hu and Bentler (1992, 1998, 1999). This is a classic example in model close fit research. The data are generated with the covariance structure: $\boldsymbol{\Sigma}_0 = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$ where

$$\boldsymbol{\Lambda} = \begin{bmatrix} .70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 \\ .00 & .00 & .00 & .00 & .00 & .70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 \\ .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .70 & .70 & .75 & .80 & .80 \end{bmatrix}'$$

$$\mathbf{\Phi} \;=\; \begin{bmatrix} 1.0 & .50 & .40 \\ .50 & 1.0 & .30 \\ .40 & .30 & 1.0 \end{bmatrix}$$

and $\mathbf{\Psi} =$ diag$(.51,.51,.44,.36,.36,.51,.51,.44,.36,.36,.51,.51,.44,.36,.36)$

In the fitted three-factor model, ϕ_{21} and ϕ_{12} are set to zero for misspecification and the factor loadings of the 5th, 10th and 15th variables on their corresponding factors are set to .80 for identification. All other nonzero parameters in the true model are set free during the estimation. For this misspecified model, $df = 88$ and RMSEA$_0 = 0.0489$.

The rejection rates of T_{ML}, T_1 to T_6 are presented in Table 1 (normal condition), Table 2 (mild nonnormal condition) and Table 3 (severe nonnormal condition). From Table 1, we see that, except for a slight overrejection when $n = 150$, T_{ML} follows $\chi^2_{88}(N \times 88 \times 0.0489^2)$ perfectly in normal data across sample size levels. However, as shown in Tables 2 and 3, when the data is nonnormal, T_{ML} always overrejects the model.

In Table 1, T_1 and T_2 always overaccept the model except when $n = 2500$. When the data is nonnormal, T_1 has good rejection rates across sample size levels in Table 2 (mild nonnormal condition) but it overrejects model substantially too much in Table 3 (severe nonnormal condition). As to T_2, it overaccepts the model too much in both tables.

Under normality, T_3, T_4, T_5 and T_6 have very good rejection rates in Table 1 when $n \geq 300$ even though they slightly overreject the model when $n = 150$. When the data is nonnormal, T_3 always overrejects the model especially in Table 3 while T_4 always overaccepts the model on the other hand (see Table 2 and 3). As to T_5, it performs well when $n \geq 300$ in Table 2 and 3. T_6 also performs well in most cases in two tables when $n \geq 300$ although it overrejects a little bit when $n = 300$. So combining Tables 1, 2 and 3, we may conclude that T_5 and T_6 perform desirably across distribution conditions in this example when $n \geq 300$.

In the discussion above, we set $a = $ RMSEA$_0 = 0.0489$ to examine the one-side type I error of the statistics. In the next step, we set a back to .05 and look at the acceptance performance of these statistics when $a > $ RMSEA$_0 = 0.0489$. Let $T_{ML,.05}$, $T_{1,.05}$, $T_{2,.05}$, $T_{3,.05}$, $T_{4,.05}$, $T_{5,.05}$, and $T_{6,.05}$ denote the corresponding test statistics when a is set to .05. The rejection rates of these statistics are presented in Table 4 (normal condition), Table 5 (mild nonnormal condition) and Table 6 (severe nonnormal condition).

Since in this example RMSEA$_0$ is equal to 0.0489, which is slightly less than the cutoff value .05, δ_1 and δ_2 by Corollary 2 and 4 should not be very large. Although there is more model acceptance across these three tables, seven statistics do not reach complete acceptance as expected according to Corollary 2 and 4 even when $n = 2500$.

Example 2. This example is a MIMIC model whose misspecification is due to the omission of some paths from the covariates to factors. This example was used by Curran, Bollen, Chen, Paxton & Kirby (2003) in their RMSEA study (see Misspecification 3 for their population model 3). The data are generated by

$$\mathbf{y} = \mathbf{\Pi}\eta + \epsilon$$

$$\eta = \mathbf{B}\eta + \mathbf{\Lambda}\mathbf{x} + \zeta$$

where $\mathbf{x}$, ϵ and ζ are independent to each other with $E(\epsilon) = 0$, Cov$(\epsilon)=\mathbf{\Psi}$, $E(\zeta) = 0$, Cov$(\zeta)=\mathbf{\Xi}$, $E(\mathbf{x}) = 0$, Cov$(\mathbf{x})=\mathbf{\Phi}$. Moreover, $\mathbf{\Psi} = \text{diag}(.11, .11, .11, .11, .22, .17, .36, .75, .75)$, $\mathbf{\Xi} = \text{diag}(.05, .11, .36)$,

$$\mathbf{\Pi} = \begin{bmatrix} 1.0 & 1.0 & 1.0 & .39 & .00 & .00 & .00 & .00 & .00 \\ .00 & .00 & .00 & .91 & 1.0 & 1.2 & .53 & .00 & .00 \\ .00 & .00 & .00 & .00 & .00 & .195 & .95 & 1.0 & 1.0 \end{bmatrix}', \quad \mathbf{B} = \begin{bmatrix} .00 & .00 & .00 \\ 1.0 & .00 & .00 \\ .00 & 1.0 & .00 \end{bmatrix},$$

$$\mathbf{\Lambda} = \begin{bmatrix} .40 & -.50 & -.60 \\ -.40 & .00 & .00 \\ -.20 & .25 & .35 \\ .20 & .00 & .00 \end{bmatrix}', \quad \text{and} \quad \mathbf{\Phi} = \begin{bmatrix} .25 & .11 & .30 & .15 \\ .11 & .20 & .18 & .18 \\ .30 & .18 & 1.0 & .50 \\ .15 & .18 & .50 & 1.0 \end{bmatrix}$$

In addition, the values of ϕ_{ii} here may be somewhat different from what Curran et al. (2003) used. This is because they did not clearly provide these values.

In the fitted MIMIC model, λ_{21}, λ_{23}, λ_{31} and λ_{33} are set to zero for misspecification. π_{21}, π_{52} and π_{83} are set to 1.0 for identification. All other nonzero parameters in the population model are set free during the estimation. For this misspecified model, $df = 54$ and RMSEA$_0 = 0.08345$.

Tables 7–9 summarize the rejection rates of T_{ML}, T_1 to T_6 in normal, mild nonnormal and severe nonnormal conditions, respectively. Again, T_{ML} shows good rejection rates across sample size levels in the normal condition (see Table 7). However, for severe nonnormal data, T_{ML} performs poorly (see Table 9).

When the data is normal, T_3, T_4, T_5 and T_6 have good rejection rates across sample size levels while T_1 and T_2 always overaccept the model except when $n = 2500$ (see Table 7). When the data is nonnormal, T_1 again has good rejection rates under mild nonnormality across the sample size levels (see Table 8), but somewhat overrejects the model under severe nonnormality (see Table 9). In Tables 8 and 9, when the data is (mild and severe) nonnormal, T_2 overaccepts the model most of the time, except in large samples. Unlike in Example 1, T_3 has good rejection rates under mild nonnormality (see Table 8), but it overrejects the model under severe nonnormality (see Table 9). In Table 8 and 9, only T_4, T_5 and T_6 have good rejection performance generally when $n \geq 300$.

In this example, RMSEA_0 is equal to 0.08345 which is between .08 and .10. So by the rule of thumb, this model contains some serious misspecifications. Tables 10, 11 and 12 show that $T_{ML,.05}$, $T_{1,.05}$ to $T_{6,.05}$ receive very close to complete rejection with a sample size of 300 across the distribution conditions. In addition, when $n = 150$, it is clear that $T_{3,.05}$, $T_{4,.05}$, $T_{5,.05}$ and $T_{6,.05}$ have more rejection power than $T_{1,.05}$ and $T_{2,.05}$. This result agrees with Corollary 5. Finally, from these three tables, $T_{ML,.05}$ has a similar rejection rate as $T_{3,.05}$, $T_{4,.05}$, $T_{5,.05}$, and $T_{6,.05}$. This may suggest that they have a similar asymptotic rejection power in evaluating misspecification.

6. Discussion

In this article, we apply the theory of Vuong (1989) and the results of Yuan, Hayashi and Bentler (2005) to RMSEA in mean and covariance structure analysis and propose four ADF-like RMSEA test statistics for evaluating a misspecified model without using the normality and population drift assumptions. In our simulation studies, two of these statistics (T_5 and T_6), in terms of type I error across the distribution conditions, can appropriately accept the models when $n \geq 300$. Another statistic (T_4) also performs desirably in convergence and model acceptance across distribution conditions when $n \geq 300$, although compared to T_5 and T_6, it is generally more conservative in rejection especially under nonnormality. In contrast to these three statistics, T_3 is not reliable across the conditions and examples that we studied. Furthermore, the two test statistics T_1 and T_2, which are directly implied

by Vuong (1989) and Yuan, Hayashi and Bentler (2005), also perform poorly. As to the noncentral chi-square based T_{ML}, it always performs very well in normal data but poorly in nonnormal ones. Since in this article the misspecifications in the examples vary from minor to serious, the contrasting performance between normal and nonnormal data across the examples suggests that the adequacy of the noncentral chi-square approximation is more sensitive to the violation of normality than it is to the population drift assumption.

As mentioned before, some standard cutoff values have been established for the population RMSEA. Given the desirable acceptance performance of T_4, T_5 and T_6 demonstrated in our simulation studies, we set the a in these statistics to be equal to some cutoff values, for example $a = .05$ for close fit, and obtain $T_{4,.05}$, $T_{5,.05}$ and $T_{6,.05}$. Clearly, $T_{4,.05}$, $T_{5,.05}$ and $T_{6,.05}$ will be very useful in evaluating the null hypothesis of model close fit $H_0 : \text{RMSEA}_0 \leq .05$ for misspecified models. When RMSEA_0 is exactly equal to .05, then $T_{4,.05}$, $T_{5,.05}$ and $T_{6,.05}$ will accept the misspecified model with some desirable alpha level of type I errors as long as the sample size is reasonably large (e.g. $n \geq 300$ according to our examples). When RMSEA_0 lies between zero and the cutoff value .05, then $T_{4,.05}$, $T_{5,.05}$ and $T_{6,.05}$ will tend to completely accept the misspecified model as n increases as they did in Examples 1. In contrast, when RMSEA_0 is over the cutoff value .05, $T_{4,.05}$, $T_{5,.05}$ and $T_{6,.05}$ will incline to completely reject the misspecified model as n increases as they did in Examples 2. In Corollaries 2 and 4, we give the general distribution of $T_{1,.05}$, $T_{2,.05}$, $T_{3,.05}$, $T_{4,.05}$, $T_{5,.05}$ and $T_{6,.05}$. Clearly, an analysis of the power of rejecting misspecified models can be further studied for these test statistics based on these general distributions.

One important point we want to emphasize again is that the RMSEA tests we propose only work for misspecified models. These statistics do not apply to correctly specified models, at least according to the theory. Since in practice nobody would know exactly if one model is correctly specified or not, model exact fit should probably be evaluated first and RMSEA tests would be applicable only after the rejection of model exact fit.

Fortunately, in mean and covariance analysis, many asymptotic distribution free tests such as the Satorra-Bentler scaled test (Satorra & Bentler, 1988, 1994) or residual-based tests (Yuan & Bentler, 1997, 1998, 1999) have been demonstrated to reliably evaluate model

exact fit even when the data is nonnormal with a reasonably small sample size and some missing values. As a result, these exact fit tests can be combined with the proposed RMSEA tests for misspecified models. Thus, a sequential two-stage procedure is proposed for overall model evaluation: accept the model if it satisfies the exact fit tests such as the Satorra-Bentler scaled test or residual-based tests; or, accept the model if it is rejected by the exact fit tests but still satisfies the RMSEA tests such as $T_{4,.05}$, $T_{5,.05}$ and $T_{6,.05}$.

One potential problem of the two-stage procedure above is its significance level during overall model evaluation. Notice that $H_0 : \text{RMSEA}_0 \leq a$ is a composite of H_0^E and $H_0 - H_0^E$. Let T_E denote some reliable exact fit test statistic such as the Satorra-Bentler scaled test or a residual-based test, and let T_C denote some reliable close fit test statistic such as $T_{4,.05}$, $T_{5,.05}$ and $T_{6,.05}$. Further, let $A \equiv \{ T_E > \chi^2_{df,\alpha}\}$ and $B \equiv \{ T_C > \lambda_\alpha \}$. Then $\Pr[\text{reject } H_0 | H_0] = \Pr[A \cap B | H_0] \leq \max\{ \Pr(A \cap B | H_0^E), \Pr(A \cap B | H_0 - H_0^E)\} \leq \max\{ \Pr(A | H_0^E), \Pr(B | H_0 - H_0^E)\}$. Let α_E and α_C be the asymptotic significance levels of T_E and T_C respectively, then $\Pr(A | H_0^E) \rightarrow \alpha_E$ and $\Pr(B | H_0 - H_0^E) \rightarrow \alpha_C$. So the significance level of the two-stage strategy is asymptotically bounded above by the maximum of the asymptotic significance levels α_E and α_C.

The theory of Vuong (1989) is based on likelihood ratio principles. In Theorem 1, we further demonstrate that $\text{tr}(\mathbf{U\Gamma})$, which is widely used in the Satorra and Bentler procedure and our RMSEA test statistics, is a special case of the more general term $\text{tr}(A^{-1}(\beta_0)B(\beta_0) - A^{-1}(\theta_*)B(\theta_*))$ based on the likelihood ratio principle. Given the generality of the likelihood ratio, and thus the results in Theorem 1, it seems that the Satorra-Bentler procedure, our RMSEA test statistics, and hence their combined two-stage procedure of model evaluation may be extendable to a wide variety of situations where the likelihood ratio principle applies. Clearly, this will tremendously increase the scope of application of the proposed methodology. Research on such extensions to a wide variety of model types or data types (e.g., multilevel models, Liang & Bentler, 2004) should be very interesting.

Appendix

Theorem 1. Under the Assumptions A1-A5 of Vuong (1989),

$$\mathbf{AE}(T_{NML}) = nF_0 + \text{tr}(A^{-1}(\beta_0)B(\beta_0) - A^{-1}(\theta_*)B(\theta_*))$$

where $\text{tr}(A^{-1}(\beta_0)B(\beta_0) - A^{-1}(\theta_*)B(\theta_*))$ reduces to $\text{tr}(\mathbf{U\Gamma})$ if the population drift assumption (4) is assumed or a correct model is specified and reduces to df if normality is assumed.

Proof. From a Taylor expansion of $l_n(\beta_0)$ around $\hat{\beta}^*$, we can get

$$l_n(\beta_0) = l_n(\hat{\beta}^*) + \frac{n}{2}(\hat{\beta}^* - \beta_0)' \left[\frac{1}{n} \cdot \sum_{i=0}^{n} \frac{\partial^2 l_i(\beta_0)}{\partial \beta_0 \partial \beta_0'}\right](\hat{\beta}^* - \beta_0) + o_p(1)$$

Similarly, we have

$$l_n(\theta_*) = l_n(\hat{\theta}_{NML}) + \frac{n}{2}(\hat{\theta}_{NML} - \theta_*)' \left[\frac{1}{n} \cdot \sum_{i=0}^{n} \frac{\partial^2 l_i(\theta_*)}{\partial \theta_* \partial \theta_*'}\right](\hat{\theta}_{NML} - \theta_*) + o_p(1)$$

Since $LR_n(\beta_0, \theta_*) = l_n(\beta_0) - l_n(\theta_*)$ and $LR_n(\hat{\beta}^*, \hat{\theta}_{NML}) = l_n(\hat{\beta}^*) - l_n(\hat{\theta}_{NML})$, we obtain

$$\begin{aligned}
2LR_n(\hat{\beta}^*, \hat{\theta}_{NML}) &= 2LR_n(\beta_0, \theta_*) - n(\hat{\beta}^* - \beta_0)' \left[\frac{1}{n} \cdot \sum_{i=0}^{n} \frac{\partial^2 l_i(\beta_0)}{\partial \beta_0 \partial \beta_0'}\right](\hat{\beta}^* - \beta_0) \\
&\quad + n(\hat{\theta}_{NML} - \theta_*)' \left[\frac{1}{n} \cdot \sum_{i=0}^{n} \frac{\partial^2 l_i(\theta_*)}{\partial \theta_* \partial \theta_*'}\right](\hat{\theta}_{NML} - \theta_*) + o_p(1) \quad \text{(A-1)}
\end{aligned}$$

It is clear that

$$\left[\frac{1}{n} \cdot \sum_{i=0}^{n} \frac{\partial^2 l_i(\beta_0)}{\partial \beta_0 \partial \beta_0'}\right] \xrightarrow{L} -A(\beta_0) \qquad \left[\frac{1}{n} \cdot \sum_{i=0}^{n} \frac{\partial^2 l_i(\theta_*)}{\partial \theta_* \partial \theta_*'}\right] \xrightarrow{L} -A(\theta_*) \qquad \text{(A-2)}$$

In addition, by (10) and Lemma 1, we obtains $2LR_n(\beta_0, \theta_*) \to nF_0$. Combining this with (1), (2), (3), (A-1) and (A-2), we obtain

$$\mathbf{AE}(T_{NML}) = nF_0 + \text{tr}(A^{-1}(\beta_0)B(\beta_0) - A^{-1}(\theta_*)B(\theta_*)) \qquad \text{(A-3)}$$

By (1), we know that $\mathbf{\Gamma} = A^{-1}(\beta_0)B(\beta_0)A^{-1}(\beta_0)$. Magnus and Neudecker (1988, p.318) also showed that $A(\beta_0) = \mathbf{W}$. Thus we obtain

$$A^{-1}(\beta_0)B(\beta_0) = A(\beta_0)A^{-1}(\beta_0)B(\beta_0)A^{-1}(\beta_0) = \mathbf{W\Gamma} \qquad \text{(A-4)}$$

. As the asymptotic covariance matrix of $\hat{\beta} \equiv (\bar{X}', \text{vech}(\mathbf{S})')'$, $\mathbf{\Gamma}$ can be partitioned as

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{\Gamma}_{\bar{X}, \text{vech}(\mathbf{S})'} \\ \mathbf{\Gamma}'_{\bar{X}, \text{vech}(\mathbf{S})'} & \mathbf{\Gamma}_{\text{vech}(\mathbf{S}), \text{vech}(\mathbf{S})'} \end{bmatrix}$$

(see Browne & Arminger, 1995). In addition, it is also clear that

$$\frac{\partial l_i(\theta_*)}{\partial \theta_*} = \frac{\partial \sigma_*'}{\partial \theta_*} \cdot \frac{\partial l_i(\sigma_*)}{\partial \sigma_*}$$

$$\frac{\partial^2 l_i(\theta_*)}{\partial \theta_* \partial \theta_*'} = \frac{\partial \sigma_*'}{\partial \theta_*} \cdot \frac{\partial^2 l_i(\sigma_*)}{\partial \sigma_* \partial \sigma_*'} \cdot \frac{\partial \sigma_*}{\partial \theta_*'} + \frac{\partial \sigma_*}{\partial \theta_* \partial \theta_*'} \cdot \frac{\partial l_i(\sigma_*)}{\partial \sigma_*}$$

$$\frac{\partial l_i(\theta_*)}{\partial \theta_*} \frac{\partial l_i(\theta_*)}{\partial \theta_*'} = \frac{\partial \sigma_*'}{\partial \theta_*} \cdot \frac{\partial l_i(\sigma_*)}{\partial \sigma_*} \cdot \frac{\partial l_i(\sigma_*)}{\partial \sigma_*'} \cdot \frac{\partial \sigma_*}{\partial \theta_*'}$$

When the population drift assumption (4) is assumed, $\beta_0 - \sigma_* = O(1/\sqrt{n})$,

$$\frac{\partial^2 l_i(\theta_*)}{\partial \theta_* \partial \theta_*'} = \frac{\partial \sigma_*'}{\partial \theta_*} \cdot \frac{\partial^2 l_i(\beta_0)}{\partial \beta_0 \partial \beta_0'} \cdot \frac{\partial \sigma_*}{\partial \theta_*'} + \frac{\partial \sigma_*}{\partial \theta_* \partial \theta_*'} \cdot \frac{\partial l_i(\beta_0)}{\partial \beta_0} + O(1/\sqrt{n}) \tag{A-5}$$

$$\frac{\partial l_i(\theta_*)}{\partial \theta_*} \frac{\partial l_i(\theta_*)}{\partial \theta_*'} = \frac{\partial \sigma_*'}{\partial \theta_*} \cdot \frac{\partial l_i(\beta_0)}{\partial \beta_0} \cdot \frac{\partial l_i(\beta_0)}{\partial \beta_0'} \cdot \frac{\partial \sigma_*}{\partial \theta_*'} + O(1/\sqrt{n}) \tag{A-6}$$

Clearly,

$$E\left[\frac{\partial^2 l_i(\beta_0)}{\partial \beta_0 \partial \beta_0'}\right] = -A(\beta_0) = -\mathbf{W} \tag{A-7}$$

Since

$$\frac{\partial l_i(\beta_0)}{\partial \mu_0} = \mathbf{\Sigma}_0^{-1}(X_i - \mu_0), \qquad \frac{\partial l_i(\beta_0)}{\partial \text{vech}(\mathbf{\Sigma}_0)} = -\frac{1}{2}\mathbf{D}_p'\text{vec}\left[\mathbf{\Sigma}_0^{-1} - \mathbf{\Sigma}_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\mathbf{\Sigma}_0^{-1}\right]$$

, so

$$E\left[\frac{\partial l_i(\beta_0)}{\partial \beta_0}\right] = 0 \tag{A-8}$$

and

$$E\left[\frac{\partial l_i(\beta_0)}{\partial \mu_0} \cdot \frac{\partial l_i(\beta_0)}{\partial \mu_0'}\right] = \mathbf{\Sigma}_0^{-1} \tag{A-9}$$

and

$$E\left[\frac{\partial l_i(\beta_0)}{\partial \mu_0} \cdot \frac{\partial l_i(\beta_0)}{\partial \text{vech}(\mathbf{\Sigma}_0)'}\right]$$

$$= -\frac{1}{2}E\left\{\mathbf{\Sigma}_0^{-1}(X_i - \mu_0) \cdot \text{vec}\left[\mathbf{\Sigma}_0^{-1} - \mathbf{\Sigma}_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\mathbf{\Sigma}_0^{-1}\right]'\mathbf{D}_p\right\}$$

$$= \frac{1}{2}E\left\{\mathbf{\Sigma}_0^{-1}(X_i - \mu_0) \cdot \text{vec}\left[\mathbf{\Sigma}_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\mathbf{\Sigma}_0^{-1}\right]'\mathbf{D}_p\right\}$$

$$= \frac{1}{2}E\left\{\mathbf{\Sigma}_0^{-1}(X_i - \mu_0) \cdot \text{vec}\left[(X_i - \mu_0)(X_i - \mu_0)'\right]'(\mathbf{\Sigma}_0^{-1} \otimes \mathbf{\Sigma}_0^{-1})\mathbf{D}_p\right\}$$

$$= \frac{1}{2}\mathbf{\Sigma}_0^{-1}E\left\{(X_i - \mu_0) \cdot \text{vec}\left[(X_i - \mu_0)(X_i - \mu_0)'\right]'\right\}(\mathbf{\Sigma}_0^{-1} \otimes \mathbf{\Sigma}_0^{-1})\mathbf{D}_p$$

$$= \frac{1}{2}\mathbf{\Sigma}_0^{-1}\mathbf{\Gamma}_{\bar{X},\text{vech}(\mathbf{S})'}\mathbf{D}_p'(\mathbf{\Sigma}_0^{-1} \otimes \mathbf{\Sigma}_0^{-1})\mathbf{D}_p \tag{A-10}$$

24

and

$$E\left[\frac{\partial l_i(\beta_0)}{\partial \text{vech}(\boldsymbol{\Sigma}_0)} \cdot \frac{\partial l_i(\beta_0)}{\partial \text{vech}(\boldsymbol{\Sigma}_0)'}\right]$$

$$= \frac{1}{4}E\left[\mathbf{D}_p'\text{vec}\left[\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\boldsymbol{\Sigma}_0^{-1}\right]\cdot \right.$$
$$\left. \text{vec}\left[\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\boldsymbol{\Sigma}_0^{-1}\right]'\mathbf{D}_p\right]$$

$$= \frac{1}{4}\mathbf{D}_p'E\left[\text{vec}\left[\boldsymbol{\Sigma}_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\boldsymbol{\Sigma}_0^{-1}\right]\cdot \right.$$
$$\left. \text{vec}\left[\boldsymbol{\Sigma}_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\boldsymbol{\Sigma}_0^{-1}\right]' - \text{vec}(\boldsymbol{\Sigma}_0^{-1})\text{vec}(\boldsymbol{\Sigma}_0^{-1})'\right]\mathbf{D}_p$$

$$= \frac{1}{4}\mathbf{D}_p'E\left[(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\text{vec}\left[(X_i - \mu_0)(X_i - \mu_0)'\right]\text{vec}\left[(X_i - \mu_0)(X_i - \mu_0)'\right]'\cdot \right.$$
$$\left. (\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1}) - (\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\text{vec}(\boldsymbol{\Sigma}_0)\cdot\text{vec}(\boldsymbol{\Sigma}_0)'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\right]\mathbf{D}_p$$

$$= \frac{1}{4}\mathbf{D}_p'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})E\left[\text{vec}((X_i - \mu_0)(X_i - \mu_0)')\text{vec}((X_i - \mu_0)(X_i - \mu_0)')\right.$$
$$\left. -\text{vec}(\boldsymbol{\Sigma}_0)\cdot\text{vec}(\boldsymbol{\Sigma}_0)'\right](\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\mathbf{D}_p$$

$$= \frac{1}{4}\mathbf{D}_p'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\mathbf{D}_p\boldsymbol{\Gamma}_{\text{vech}(\mathbf{S}),\text{vech}(\mathbf{S})'}\mathbf{D}_p'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\mathbf{D}_p \tag{A-11}$$

where $\text{vec}(\cdot)$ is the operator of transforming a matrix into a vector by stacking the columns of the matrix one underneath the other. Then under the population drift assumption (4),

$$A(\theta_*) = -E\left[\frac{\partial^2 l_i(\theta_*)}{\partial \theta_* \partial \theta_*'}\right] \stackrel{a}{=} \dot{\sigma}_*'\mathbf{W}\dot{\sigma}_* \stackrel{a}{=} \dot{\sigma}_*'\mathbf{W}_*\dot{\sigma}_* \tag{A-12}$$

by (A-5), (A-7) and (A-8), and

$$B(\theta_*)$$
$$= E\left[\frac{\partial l_i(\theta_*)}{\partial \theta_*}\frac{\partial l_i(\theta_*)}{\partial \theta_*'}\right]$$
$$\stackrel{a}{=} \dot{\sigma}_*'\left[\begin{array}{c}\boldsymbol{\Sigma}_0^{-1} \\ 2^{-1}\mathbf{D}_p'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\mathbf{D}_p\boldsymbol{\Gamma}_{\mu,\text{vech}(\Sigma)'}'\boldsymbol{\Sigma}_0^{-1} \\ 2^{-1}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Gamma}_{\mu,\text{vech}(\Sigma)'}\mathbf{D}_p'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\mathbf{D}_p \\ 4^{-1}\mathbf{D}_p'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\mathbf{D}_p\boldsymbol{\Gamma}_{\text{vech}(\Sigma),\text{vech}(\Sigma)'}\mathbf{D}_p'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\mathbf{D}_p\end{array}\right]\dot{\sigma}_*$$
$$= \dot{\sigma}_*'\mathbf{W}\boldsymbol{\Gamma}\mathbf{W}\dot{\sigma}_*$$
$$\stackrel{a}{=} \dot{\sigma}_*'\mathbf{W}_*\boldsymbol{\Gamma}\mathbf{W}_*\dot{\sigma}_* \tag{A-13}$$

by (A-6), (A-9), (A-10) and (A-11), and then

$$\mathbf{AE}(T_{NML}) = nF_0 + \text{tr}(\mathbf{W}\boldsymbol{\Gamma} - A^{-1}(\theta_*)B(\theta_*))$$

25

$$\overset{a}{=} \quad nF_0 + \text{tr}(\mathbf{W}_*\mathbf{\Gamma} - (\dot{\sigma}'_*\mathbf{W}_*\dot{\sigma}_*)^{-1}\dot{\sigma}'_*\mathbf{W}_*\mathbf{\Gamma}\mathbf{W}_*\dot{\sigma}_*)$$

$$= \quad nF_0 + \text{tr}(\mathbf{U}\mathbf{\Gamma})$$

by (A-3), (A-4), (A-12) and (A-13). When the model is correctly specified, the equation above can be proved similarly. In addition, under normality, $A(\beta_0) = B(\beta_0)$ and $A(\theta_*) = B(\theta_*)$, then

$$\mathbf{AE}(T_{NML}) \quad = \quad nF_0 + df$$

References

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). New York: John Wiley.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317-332.

Arminger, G., & Schoenberg, R. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika, 54,* 409-426.

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

Bentler, P. M. (2006). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software.

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

Bollen, K. A. (1986). Sample size and Bentler and Bonett's nonnormed fit index. *Psychometrika, 51,* 375-377.

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

Bollen, K. A., & Curran, P. J. (2006). *Latent Curve Models: A Structural Equation Approach.* Wiley Series on Probability and Mathematical Statistics. John Wiley & Sons: New Jersey.

Browne, M. W. (1974). Generalized least squares in the analysis of covariance structures. *South African Statistical Journal, 8,* 1-24.

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

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean and covariance structure models. In G. Arminger, C.C., Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for social and behavioral science* (pp. 185-249). New York: Plenum.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-62). Newbury Park, CA: Sage.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming.* Mahwah, NJ: Lawrence Erlbaum.

Chen W. W., & Deo R. S. (2004). Power transformations to induce normality and their applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66,* no. 1, pp. 117-130.

Clarke, K. A. (2003). Nonparametric model discrimination in international relations. *Journal of Conflict Resolution, 47,* 72-93.

Clarke, K. A. (2005). *A Simple Distribution-Free Test for Non-Nested Model Selection.* Manuscript.

Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. (2003). The finite sampling properties of the RMSEA: Point estimates and confidence intervals. *Sociological Methods and Research, 32,* 208-252.

De Leeuw, J. (1988). Model selection in multinomial experiments. In T. K. Dijkstra (Ed.), *On model uncertainty and its statistical implications* (pp. 118-138). Berlin: Springer.

Goldberger, A. S., & Duncan, O. (Eds.) (1973). *Structural Equation Models in the Social Sciences.* New York: Seminar Press.

Golden, R. M. (2003). Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. *Psychometrika, 68,* 229-49.

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika, 66,* 373-388.

Hays, R. D., Revicki, D., & Coyne, K. (2005). Application of structural equation modeling to health outcomes research. *Evaluation and the Health Professions, 28,* 295-309.

Holbert, R. L., & Stephenson, M. T. (2002). Structural equation modeling in the communication sciences, 1995-2000. *Human Communication Research, 28,* 531-551.

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

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

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

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Models, assumptions, and data.* Beverly Hills: Sage Publications, Inc.

Joreskog, K. G., & Sorbom, D. (1981). *LISREL V: Analysis of linear structural relationships by the method of maximum likelihood.* Chicago, IL: National Educational Resources.

Kano, Y. (1986). Conditions on consistency of estimators in covariance structure model. *Journal of the Japan Statistical Society, 16,* 75-80.

Kano, Y. (2001). Structural equation modeling for experimental data. In R. Cudeck, S.H.C. du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 381-402). Lincolnwood, IL: Scientific.

Kim, K. (2003). *The relationship among fit indices, power, and sample size in structural equation modeling.* Ph.D thesis, UCLA.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22,* 76-86.

Lee, S.-Y. (Ed.). (in press). *Handbook of structural equation models.* Amsterdam: Elsevier.

Liang, J., & Bentler, P. M. (2004). An EM algorithm for fitting two-level structural equation models. *Psychometrika, 69,* 101-122.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research, 38,* 113-139.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201-226.

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

McDonald, R.P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6,* 97-103.

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

Magnus, J. R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics.* New York: Wiley.

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

Martens, M. P. (2005). The use of structural equation modeling in counseling psychology research. *Counseling Psychologist, 33,* 269-298.

Muthen, L. & Muthen, B. (2003). *Mplus Version 2.13: Addendum to the Mplus User's Guide.* Available at www.statmodel.com/support/index.html.

Olsson U. H., Foss T., & Breivik E. (2004). Two equivalent discrepancy functions for maximum likelihood estimation: Do their test statistics follow a non-central chi-square distribution under model misspecification? *Sociological Methods & Research, 32,* 453-500.

Peek, M. K. (2000). Structural equation modeling and rehabilitation research. *American Journal of Physical Medicine and Rehabilitation, 79,* 301-309.

Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004). Comparing dynamic causal models. *NeuroImage 22,* 1157-72.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *Proceedings of the American Statistical Association,* 308-313.

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

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507-514.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50,* 83-90.

Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *South African Statistical Journal, 17,* 33-81.

Shapiro, A. (1984). A note on the consistency of estimators in the analysis of moment structures. *British Journal of Mathematical and Statistical Psychology, 37,* 84-88.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models.* Boca Raton, FL: Chapman & Hall/CRC.

Smith, D., & Langfield-Smith, K. (2004). Structural equation modeling in management accounting research: Critical analysis and opportunities, *Journal of Accounting Literature, 23,* pp. 49-86.

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

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

Tanaka, J. S. (1987). How big is enough? Sample size and goodness-of fit in structural equation models with latent variables. *Child Development, 58,* pp 134-46.

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology, 38,* 197-201.

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

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica, 57,* 307-333.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society, 54,* 426-482.

Wilson, E. B., & Hilferty, M. M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences of the United States of America, 17,* 684-688.

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

Yuan, K.-H., & Bentler, P. M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association, 92,* 767-774.

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

Yuan, K.-H., & Bentler, P. M. (1999). F-tests for mean and covariance structure analysis. *Journal of Educational and Behavioral Statistics, 24,* 225-243.

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In Sobel, M.E. & Becker, M.P. (Eds.), *Sociological Methodology* (pp. 165-200). Washington, D.C.: ASA.

Yuan, K.-H., & Bentler, P. M. (2004). On chi-square difference and z tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement, 64,* 737-757.

Yuan, K.-H., & Bentler, P. M. (in press). Structural equation modeling. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Psychometrics.* Amsterdam: North-Holland.

Yuan, K.-H., Hayashi, K., & Bentler, P. M. (2005). Normal theory likelihood ratio statistic for mean and covariance structure analysis under alternative hypothesis. Under review.

Yuan, K.-H., & Jennrich, R. I. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis, 65,* 245-260.

Yuan, K.-H., & Marshall, L. L. (2004). A new measure of misfit for covariance structure models. *Behaviormetrika, 31,* 67-90.

Yuan, K.-H., Marshall, L. L., & Bentler, P. M. (2003). Assessing the effect of model misspecifications on parameter estimates in structural equation models. *Sociological Methodology, 33,* 241–265.

Table 1. Rejection rate of different statistics
with $\alpha = .05$ for Example 1, normal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
T_{ML}	154	112	110	88	93	97
T_1	59	50	58	53	67	91
T_2	79	52	60	54	67	92
T_3	129/1980	89	82	77	85	103
T_4	148/1980	91	89	75	86	103
T_5	139/1969	90	88	77	86	104
T_6	137/1969	90	88	77	85	104

Table 2. Rejection rate of different statistics
with $\alpha = .05$ for Example 1, mild nonnormal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
T_{ML}	464	336	307	278	236	174
T_1	118	84	100	101	108	112
T_2	44	25	30	24	44	52
T_3	236/1991	169	153	149	138	133
T_4	127/1991	62	68	59	70	71
T_5	160/1993	81	87	81	76	74
T_6	174/1993	91	99	91	85	82

Table 3. Rejection rate of different statistics
with $\alpha = .05$ for Example 1, severe nonnormal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
T_{ML}	1272	1113	1024	903	718	514
T_1	324	270	282	282	271	252
T_2	17	6	11	14	21	46
T_3	682/1998	491	448	416	364	285
T_4	129/1998	52	51	47	54	71
T_5	235	98	90	83	75	90
T_6	274	132	112	96	95	94

Table 4. Rejection rate of $T_{ML,.05}$, $T_{1,.05}$ to $T_{6,.05}$

with for Example 1, normal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
$T_{ML,.05}$	132	83	73	56	47	27
$T_{1,.05}$	51	34	36	35	38	24
$T_{2,.05}$	62	35	41	35	39	23
$T_{3,.05}$	103/1980	57	58	47	44	26
$T_{4,.05}$	126/1980	67	60	50	44	27
$T_{5,.05}$	117/1969	62	58	50	44	27
$T_{6,.05}$	110/1969	61	58	49	44	27

Table 5. Rejection rate of $T_{ML,.05}$, $T_{1,.05}$ to $T_{6,.05}$

with for Example 1, mild nonnormal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
$T_{ML,.05}$	420	276	223	209	134	48
$T_{1,.05}$	102	61	72	66	63	34
$T_{2,.05}$	35	19	21	17	24	17
$T_{3,.05}$	208/1991	129	122	104	80	39
$T_{4,.05}$	112/1991	40	45	32	33	22
$T_{5,.05}$	136/1993	59	59	44	37	23
$T_{6,.05}$	153/1993	65	68	53	41	26

Table 6. Rejection rate of $T_{ML,.05}$, $T_{1,.05}$ to $T_{6,.05}$

with for Example 1, severe nonnormal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
$T_{ML,.05}$	1222	1027	894	775	558	282
$T_{1,.05}$	295	224	229	213	185	125
$T_{2,.05}$	15	4	7	6	11	11
$T_{3,.05}$	637/1998	415	367	332	233	142
$T_{4,.05}$	108/1998	40	42	30	28	23
$T_{5,.05}$	204	71	67	56	44	31
$T_{6,.05}$	245	98	87	74	55	32

Table 7. Rejection rate of different statistics
with $\alpha = .05$ for Example 2, normal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
T_{ML}	92	91	88	67	76	78
T_1	54	72	64	56	75	81
T_2	72	81	68	58	75	80
T_3	90	94	86	71	89	89
T_4	112	99	93	74	89	89
T_5	102/1999	97	92	73	89	89
T_6	95/1999	97	92	73	89	89

Table 8. Rejection rate of different statistics
with $\alpha = .05$ for Example 2, mild nonnormal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
T_{ML}	166	129	121	115	108	110
T_1	82	70	85	81	100	115
T_2	64	53	64	58	82	94
T_3	127	106	107	109	111	128
T_4	102	74	84	85	96	107
T_5	108/1999	83	84	87	98	107
T_6	114/1997	88	91	90	100	109

Table 9. Rejection rate of different statistics
with $\alpha = .05$ for Example 2, severe nonnormal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
T_{ML}	390	317	256	272	190	150
T_1	145	139	129	146	134	130
T_2	81	49	37	51	66	86
T_3	233	199	165	188	150	139
T_4	139	86	68	84	91	97
T_5	152/1993	105	80	102	102	100
T_6	170/1988	116	93	109	108	105

Table 10. Rejection rate of $T_{ML,.05}$, $T_{1,.05}$ to $T_{6,.05}$

with for Example 2, normal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
$T_{ML,.05}$	1561	1952	1997	2000	2000	2000
$T_{1,.05}$	1280	1905	1986	2000	2000	2000
$T_{2,.05}$	1368	1909	1987	2000	2000	2000
$T_{3,.05}$	1518	1942	1997	2000	2000	2000
$T_{4,.05}$	1579	1943	1997	2000	2000	2000
$T_{5,.05}$	1550/1999	1942	1997	2000	2000	2000
$T_{6,.05}$	1533/1999	1943	1997	2000	2000	2000

Table 11. Rejection rate of $T_{ML,.05}$, $T_{1,.05}$ to $T_{6,.05}$

with for Example 2, mild nonnormal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
$T_{ML,.05}$	1679	1975	1996	2000	2000	2000
$T_{1,.05}$	1342	1937	1989	2000	2000	2000
$T_{2,.05}$	1287	1911	1980	2000	2000	2000
$T_{3,.05}$	1601	1965	1992	2000	2000	2000
$T_{4,.05}$	1560	1954	1992	2000	2000	2000
$T_{5,.05}$	1581/1999	1961	1991	2000	2000	2000
$T_{6,.05}$	1585/1997	1962	1991	2000	2000	2000

Table 12. Rejection rate of $T_{ML,.05}$, $T_{1,.05}$ to $T_{6,.05}$

with for Example 2, severe nonnormal condition, NR=2000

Sample Size	150	300	400	500	1000	2500
$T_{ML,.05}$	1818	1987	1999	2000	2000	2000
$T_{1,.05}$	1413	1956	1995	2000	2000	2000
$T_{2,.05}$	1109	1858	1973	1996	2000	2000
$T_{3,.05}$	1703	1980	1998	2000	2000	2000
$T_{4,.05}$	1512	1942	1993	2000	2000	2000
$T_{5,.05}$	1585/1993	1964	1995	2000	2000	2000
$T_{6,.05}$	1627/1988	1970	1997	2000	2000	2000