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

*Research supported in part by grants DA00017 and DA01070 from the National Institute on Drug Abuse.
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 \(X = (x_1, \ldots, x_p)'\) and \(m\)-unobserved factors are hypothesized to depend on \(q\) unknown structural parameters included in a \(q \times 1\) parameter vector \(\theta\). The hypothesized model leads to the model-implied mean \(\mu(\theta)\) and covariance matrix \(\Sigma(\theta)\). Now let \(\mu = E(X)\), \(\Sigma = \text{cov}(X)\), \(\bar{X}\) and \(S\) be the corresponding mean and unbiased sample estimator and \(S^* \equiv (n - 1) \cdot S/n\) be the MLE estimator of \(\Sigma\), where \(n\) is the sample size. Let \(\beta\) denote the parameter vector of the saturated model, then in this case \(\beta = (\mu', \text{vech}(\Sigma)')'\), where \(\text{vech}(...)\) is an operator which transforms a symmetric matrix into a vector by stacking the nonduplicated elements of the matrix. Further, \(\hat{\beta}^* \equiv (\bar{X}', \text{vech}(S^*)')'\) and \(\hat{\beta} \equiv (\bar{X}', \text{vech}(S')')'\) will be the MLE and unbiased estimator of \(\beta\) 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

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

where \(f(X_i; \mu, \Sigma)\) 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 \(\mu_0, \Sigma_0\) denote the population counterpart to \(\mu, \Sigma\) and \(\beta_0 \equiv (\mu_0', \text{vech}(\Sigma_0)')\). Let \(\Gamma\) 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,

\[
\sqrt{n}(\hat{\beta}^* - \beta_0) \overset{\text{a}}{=} \sqrt{n}(\hat{\beta} - \beta_0) \overset{L}{\rightarrow} N(0, \Gamma)
\]

(1)

where \(\overset{\text{a}}{=}\) refers to asymptotic equality (i.e., the difference between both sides of the equality tends to zero in probability as \(n \to \infty\)). Further, \(\Gamma\) 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

\[
A(\beta_0) = -E \left[ \frac{\partial^2 l_i(\beta_0)}{\partial \beta_0 \partial \beta_0'} \right] \quad B(\beta_0) = E \left[ \frac{\partial l_i(\beta_0)}{\partial \beta_0} \frac{\partial l_i(\beta_0)}{\partial \beta_0'} \right]
\]
where $E(\cdot)$ denotes the expectation with respect to the true distribution of $X$.

When $\mu$ and $\Sigma$ 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 $\theta$ can be obtained by minimizing the well-known normal theory maximum likelihood discrepancy function (Browne & Arminger, 1995)

$$F_{ML}(\bar{X}, S^*; \mu(\theta), \Sigma(\theta)) = (\bar{X} - \mu(\theta))' \Sigma^{-1}(\theta) (\bar{X} - \mu(\theta)) + \log |\Sigma(\theta)| + \text{tr}(S^* \Sigma^{-1}(\theta))$$

$$- \log |S^*| - p$$

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

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

In practice, the discrepancy function $F_{ML}(\bar{X}, S; \mu(\theta), \Sigma(\theta))$ is used in most literature instead of $F_{ML}(\bar{X}, S^*; \mu(\theta), \Sigma(\theta))$. Its minimizer, $\hat{\theta}_{ML}$, and the corresponding NTLR statistic $T_{ML} = nF_{ML}(\bar{X}, S; \mu(\hat{\theta}_{ML}), \Sigma(\hat{\theta}_{ML}))$ 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 $\theta_0$ 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

$$\sqrt{n}(\hat{\theta}_{ML} - \theta_0) \overset{a}{\rightarrow} \sqrt{n}(\hat{\theta}_{NML} - \theta_0) \overset{L}{\rightarrow} N(0, \Omega_{\theta_0})$$
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, $\beta_0$ is attainable and can be reproduced by a group of parameters $\theta_0$. However, the model parameterized by such $\theta_0$ may be hard to find or has no interpretability or substantive usefulness. In another possibility, $\beta_0$ is unattainable and can not be reproduced by any group of parameters, or in other words, $F_{ML}(\mu_0, \Sigma_0; \mu(\theta_0), \Sigma(\theta_0)) > 0$ for any possible value of $\theta$ except in a saturated model where $\theta_0$ is equal to, or is some transformation of, $\beta_0$.

No matter of what kind of possibility it is, suppose now we have a model parameterized by a $q \times 1$ structural parameter vector $\theta$ as we assumed before. Due to the misspecification (parameter underrepresentation), the minimizer $\theta_*$ of $F_{ML}(\mu_0, \Sigma_0; \mu(\theta), \Sigma(\theta))$ differs from $\theta_0$ 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 $\theta_*$ and be asymptotically normally distributed (Arminger & Schoenberg, 1989; Vuong, 1989; Yuan & Jennrich, 1998), and

$$
\sqrt{n}(\hat{\theta}_{ML} - \theta_*) \overset{a}{=} \sqrt{n}(\hat{\theta}_{NML} - \theta_*) \overset{L}{\rightarrow} N(0, \Omega_{\theta_*})
$$

(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 $F_0 = F_{ML}(\mu_0, \Sigma_0; \mu(\theta_\ast), \Sigma(\theta_\ast))$. When the model is misspecified, it is well-known that the NTLR statistic $T_{ML} \overset{L}{\rightarrow} \chi^2_{d f}(NF_0)$ under the assumption of normality and the assumption of population drift, that is,

$$
\mu_0 - \mu(\theta_\ast) = O(1/\sqrt{n}) \quad \text{and} \quad \Sigma_0 - \Sigma(\theta_\ast) = O(1/\sqrt{n})
$$

(4)

Let $\sigma_\ast = (\mu(\theta_\ast)', \text{vech}(\Sigma(\theta_\ast))')'$ and $\dot{\sigma}_\ast = \partial \sigma_\ast / \partial \theta_\ast'$. Let $D_p$ be the duplication matrix as defined by Magnus and Neudecker (1988), $W = \text{diag}[\Sigma_0^{-1}, 2^{-1}D_p'(\Sigma_0^{-1} \otimes \Sigma_0^{-1})D_p]$, $W_\ast = \text{diag}[\Sigma_\ast^{-1}(\theta_\ast), 2^{-1}D_p'(\Sigma_\ast^{-1}(\theta_\ast) \otimes \Sigma_\ast^{-1}(\theta_\ast))D_p]$ and $U = W_\ast - W_\ast \dot{\sigma}_\ast (\dot{\sigma}_\ast' W_\ast \dot{\sigma}_\ast)^{-1} \dot{\sigma}_\ast' W_\ast$. Then it has been shown that under the population drift assumption (4) (Yuan & Marshall, 2004),

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

(5)

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

$$
\text{AE}(T_{ML}) = NF_0 + df
$$

(6)

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

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

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

**Proof.** 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

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

(7)
As mentioned above, $\hat{\theta}_{NML}$ is strongly consistent to $\theta_*$ 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 $\beta_0$ 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 $(\beta^*, \theta^*_{NML})'$, we can get

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

where

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

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

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

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

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

The term on the right side of the equation is the Kullback-Leibler (1951) Information Criterion in statistical theory. Suppose $E [\log [f(X_i; \beta_0)/f(X_i; \theta_*)]^2]$ 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),

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

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

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

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

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

**Proof.**

\[
E \left[ \log \frac{f(X_i; \beta_0)}{f(X_i; \theta_*)} \right] = E \left[ \log f(X_t; \beta_0) - \log f(X_t; \theta_*) \right] = \frac{1}{2} E \left[ \log |\Sigma(\theta_*)| + (X_t - \mu(\theta_*))^{\prime} \Sigma^{-1}(\theta_*)(X_t - \mu(\theta_*)) \right]
\]

\[= \frac{1}{2} \left[ \log |\Sigma(\theta_*)| + \text{tr}(\Sigma(\theta_*)^{-1}(\theta_*) + (\mu(\theta_*) - \mu(\theta_*))^{\prime} \Sigma^{-1}(\theta_*) \right]
\]

\[= \frac{1}{2} E \left[ \log |\Sigma_0| + (X_t - \mu_0)^{\prime} \Sigma_0^{-1}(X_t - \mu_0) \right]
\]

\[= \frac{1}{2} F_0
\]

Let \( \tilde{F}_{ML} \equiv \frac{1}{N} T_{ML} = F_{ML}(\tilde{X}, S; \mu(\hat{\theta}_{ML}), \Sigma(\hat{\theta}_{ML})) \). 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,

\[
\sqrt{n} \left( \tilde{F}_{ML} - F_0 - \frac{\text{tr}(UT)}{n} \right) \overset{L}{\to} N(0, 4\omega^2)
\]

when normality is assumed, this reduces to

\[
\sqrt{n} \left( \tilde{F}_{ML} - F_0 - \frac{df}{n} \right) \overset{L}{\to} N(0, 4\omega^2)
\]
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 $\omega^2$, that is,

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

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

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

(12)

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

Yuan, Hayashi and Bentler (2005) further derived the explicit form of $\omega^2$ 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)

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

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

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

where $\hat{\Gamma}$ is the consistent estimator of $\Gamma$ (e.g., Bentler, 2006) and $\hat{U}$ is a consistent estimator of $U$ obtained by replacing $\theta_*$ 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 $T_{ML} \overset{L}{\rightarrow} \chi^2_{df}(NF_0)$, then $\Pr[T_{ML} > \chi^2_{df, .95}(N \times df \times a^2)] \overset{L}{\rightarrow} .05$. 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 $\omega^2$ is given in (12). Based on these results, the following two test statistics are proposed: Vuong’s test statistic ($T_1$) which is:

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

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

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

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

$$T_1 \overset{a}{=} T_2 \overset{L}{\longrightarrow} N(\sqrt{n}\delta_1, 1)$$

and

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

where $a_0$ is the value of $\text{RMSEA}_0$. Further,

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

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

3. When $0 < \text{RMSEA}_0 < a$, 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 $H_0 : \text{RMSEA}_0 \leq a$, $\Pr[T_1 \text{ or } T_2 > \lambda_{.95}] \overset{L}{\longrightarrow} .05$. 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 $H_0 : \text{RMSEA}_0 \leq a$ 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 $\chi^2$ 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 $\text{RMSEA}_0$ is a square root transformation of the standard NCP $F_0$ divided by a constant $\sqrt{d_f}$. 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 ($\text{RMSEA}_0 \neq 0$),

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

or for the general case,

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

\[
\sqrt{n} \left( \text{RMSEA} - \text{RMSEA}_0 \right) = \sqrt{n} \left( \sqrt{\max(\hat{F}_{ML}/df - 1/n, 0)} - \sqrt{F_0/df} \right) \\
\approx \sqrt{n} \left( \sqrt{\hat{F}_{ML}/df} - \sqrt{F_0/df} \right) \\
\xrightarrow{L} N(0, \frac{\omega^2}{df \times F_0}) \quad \text{(Delta method)}
\]

The property of \( \tilde{\text{RMSEA}} \) 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

\[
T_3 = \frac{\sqrt{n} \left( \text{RMSEA} - a \right)}{\hat{\omega} / \sqrt{df \times (\hat{F}_{ML} - df/n)}}
\]

and the robust RMSEA test statistic \( T_4 \) which is

\[
T_4 = \frac{\sqrt{n} \left( \text{RMSEA} - a \right)}{\hat{\omega} / \sqrt{df \times (\hat{F}_{ML} - df/n)}}
\]

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

\[
T_5 = \frac{\sqrt{n} \left( \text{RMSEA} - a \right)}{\sqrt{\hat{\omega}^2 - \hat{c}} / \sqrt{df \times (\hat{F}_{ML} - df/n + \hat{c})}}
\]

and

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

Clearly, \( \hat{c} \) is an estimator of \( c_0 \equiv (\text{tr}(U \Gamma) - df)/n \) 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.
Another point which should be mentioned here is that when one or several quantities
among $\hat{F}_{M\ell} - df/n$, $\hat{F}_{M\ell} - df/n + \hat{c}$, $\hat{\omega}^2 - \hat{\omega}^2$ 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 ($\text{RMSEA}_0 \neq 0$),

$$T_3 \overset{\alpha}{=} T_4 \overset{\alpha}{=} T_5 \overset{\alpha}{=} T_6 \xrightarrow{L} N(\sqrt{n}\delta_2, 1) \quad \text{and} \quad \delta_2 = \frac{df}{\omega} \cdot (a_0^2 - a \cdot a_0)$$

Further,

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

2. When $\text{RMSEA}_0 > a$, then $\delta_2 > 0$ and $T_3 \overset{\alpha}{=} T_4 \overset{\alpha}{=} T_5 \overset{\alpha}{=} T_6 \xrightarrow{\rightarrow} +\infty$ as $n \xrightarrow{\rightarrow} +\infty$.

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

Clearly, like $T_1$ and $T_2$ discussed before, $T_3$, $T_4$, $T_5$ or $T_6$ can be used to test the hy-
pothesis 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,

$$T_1 \overset{\alpha}{=} T_2 \xrightarrow{L} N(\sqrt{n}\delta_1, 1) \quad \quad T_3 \overset{\alpha}{=} T_4 \overset{\alpha}{=} T_5 \overset{\alpha}{=} T_6 \xrightarrow{L} N(\sqrt{n}\delta_2, 1)$$

where

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

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

16
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 $\text{RMSEA}_0$ to a specific value $a$ such as the widely accepted value $a = .05$. Instead, we set $a$ equal to the value of $\text{RMSEA}_0$ for all statistics since $\text{RMSEA}_0$ 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 $(\hat{\sigma}_*^\prime \hat{W} \hat{\sigma}_*)$, the Fisher information estimator, in $\hat{U}$ 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: $\Sigma_0 = \Lambda \Phi \Lambda^\prime + \Psi$ where

$$
\Lambda = \begin{bmatrix}
.70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 & .00 & .00 \\
.00 & .00 & .00 & .00 & .00 & .70 & .70 & .75 & .80 & .80 & .00 & .00 \\
.00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .70 & .70 & .75 & .80 \\
.00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00
\end{bmatrix}^\prime
$$

17
\[
\Phi = \begin{bmatrix}
1.0 & .50 & .40 \\
.50 & 1.0 & .30 \\
.40 & .30 & 1.0 \\
\end{bmatrix}
\]

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

In the fitted three-factor model, \( \phi_{21} \) and \( \phi_{12} \) 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 \( \text{RMSEA}_0 = 0.0489 \).

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 \( a = \text{RMSEA}_0 = 0.0489 \) 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 \( a > \text{RMSEA}_0 = 0.0489 \). 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, $\delta_1$ and $\delta_2$ 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

$$\begin{align*}
y &= \Pi \eta + \epsilon \\
\eta &= B \eta + \Lambda x + \zeta
\end{align*}$$

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

$$\begin{align*}
\Pi &= \begin{bmatrix}
1.0 & 1.0 & 1.0 & .39 & .00 & .00 & .00 & .00 \\
.00 & .00 & .00 & .91 & 1.0 & .53 & .00 & .00 \\
.00 & .00 & .00 & .00 & .00 & 1.95 & 1.0 & 1.0
\end{bmatrix}^T, \\
\Lambda &= \begin{bmatrix}
.40 & -.50 & -.60 \\
-.40 & .00 & .00 \\
-.20 & .25 & .35
\end{bmatrix}^T, \\
\Phi &= \begin{bmatrix}
.25 & .11 & .30 & .15 \\
.11 & .20 & .18 & .18 \\
.30 & .18 & 1.0 & .50 \\
.15 & .18 & .50 & 1.0
\end{bmatrix}
\end{align*}$$

In addition, the values of $\phi_{ii}$ 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, $\lambda_{21}$, $\lambda_{23}$, $\lambda_{31}$ and $\lambda_{33}$ are set to zero for misspecification. $\pi_{21}$, $\pi_{52}$ and $\pi_{83}$ 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 $\text{RMSEA}_0$ 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 $\text{RMSEA}_0$ 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 $\text{RMSEA}_0$ 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 $H_0 : \text{RMSE}_0 \leq a$ 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 $A \equiv \{ T_E > \chi^2_{d,f,a} \}$ and $B \equiv \{ T_C > \lambda_a \}$. 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 $\alpha_E$ and $\alpha_C$ 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 $\alpha_E$ and $\alpha_C$.

The theory of Vuong (1989) is based on likelihood ratio principles. In Theorem 1, we further demonstrate that $\text{tr}(U\Gamma)$, which is widely used in the Satorra and Bentler procedure and our RMSEA test statistics, is a special case of the more general term $\text{tr}(A^{-1}(\beta_0)B(\beta_0) - A^{-1}(\theta_0)B(\theta_0))$ 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),

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

where \(\text{tr}(A^{-1}(\beta_0)B(\beta_0) - A^{-1}(\theta_*)B(\theta_*))\) reduces to \(\text{tr}(\mathbf{U}\mathbf{T})\) 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

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

Similarly, we have

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

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

\[
2LR_n(\hat{\beta}^*, \hat{\theta}_{NML}) = 2LR_n(\beta_0, \theta_*) - n(\hat{\beta}^* - \beta_0)' \left[ \frac{1}{n} \sum_{i=0}^{n} \frac{\partial^2 l_i(\beta_0)}{\partial \beta_0 \partial \beta_0} \right] (\hat{\beta}^* - \beta_0)
\]

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

It is clear that

\[
\left[ \frac{1}{n} \sum_{i=0}^{n} \frac{\partial^2 l_i(\beta_0)}{\partial \beta_0 \partial \beta_0} \right] \overset{L}{\rightarrow} -A(\beta_0) \quad \left[ \frac{1}{n} \sum_{i=0}^{n} \frac{\partial^2 l_i(\theta_*)}{\partial \theta_* \partial \theta_*} \right] \overset{L}{\rightarrow} -A(\theta_*) \quad (A-2)
\]

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

\[
\text{AE}(T_{NML}) = nF_0 + \text{tr}(A^{-1}(\beta_0)B(\beta_0) - A^{-1}(\theta_*)B(\theta_*)) \quad (A-3)
\]

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

\[
A^{-1}(\beta_0)B(\beta_0) = A(\beta_0)A^{-1}(\beta_0)B(\beta_0)A^{-1}(\beta_0) = \mathbf{W}\Gamma \quad (A-4)
\]

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

\[
\Gamma = \begin{bmatrix} \Sigma & \Gamma_{\hat{X}, vech(S)}' \\ \Gamma_{X, vech(S)}' & \Gamma_{vech(S), vech(S)}' \end{bmatrix}
\]
(see Browne & Arminger, 1995). In addition, it is also clear that

\[
\frac{\partial l_i(\theta)}{\partial \theta_i} = \frac{\partial \sigma_i'}{\partial \theta_i} \cdot \frac{\partial l_i(\sigma_i)}{\partial \sigma_i}
\]

\[
\frac{\partial^2 l_i(\theta)}{\partial \theta_i \partial \theta_i'} = \frac{\partial \sigma_i'}{\partial \theta_i} \cdot \frac{\partial^2 l_i(\sigma_i)}{\partial \sigma_i \partial \sigma_i'} + \frac{\partial \sigma_i'}{\partial \theta_i} \cdot \frac{\partial l_i(\sigma_i)}{\partial \sigma_i} + \frac{\partial l_i(\sigma_i)}{\partial \sigma_i} \cdot \frac{\partial l_i(\sigma_i)}{\partial \sigma_i'}
\]

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

\[
\frac{\partial^2 l_i(\beta_0)}{\partial \beta_i \partial \beta_i'} = \frac{\partial \sigma_i'}{\partial \beta_i} \cdot \frac{\partial^2 l_i(\beta_0)}{\partial \beta_i \partial \beta_i'} + \frac{\partial \sigma_i'}{\partial \beta_i} \cdot \frac{\partial l_i(\beta_0)}{\partial \beta_i} + O(1/\sqrt{n}) \quad (A-5)
\]

\[
\frac{\partial l_i(\beta_0)}{\partial \beta_i} = \Sigma_0^{-1} (X_i - \mu_0), \quad \frac{\partial l_i(\beta_0)}{\partial \text{vech}(\Sigma_0)} = -\frac{1}{2} \mathbf{D}_p' \text{vech} \left[ \Sigma_0^{-1} - \Sigma_0^{-1} (X_i - \mu_0)(X_i - \mu_0)' \Sigma_0^{-1} \right]
\]

so

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

and

\[
E \left[ \frac{\partial l_i(\beta_0)}{\partial \mu_0} \cdot \frac{\partial l_i(\beta_0)}{\partial \mu_0'} \right] = \Sigma_0^{-1} \quad (A-9)
\]

and

\[
E \left[ \frac{\partial l_i(\beta_0)}{\partial \mu_0} \cdot \frac{\partial l_i(\beta_0)}{\partial \text{vech}(\Sigma_0)'} \right] = -\frac{1}{2} E \left\{ \Sigma_0^{-1} (X_i - \mu_0) \cdot \text{vech} \left[ \Sigma_0^{-1} - \Sigma_0^{-1} (X_i - \mu_0)(X_i - \mu_0)' \Sigma_0^{-1} \right]' \mathbf{D}_p \right\}
\]

\[
= \frac{1}{2} E \left\{ \Sigma_0^{-1} (X_i - \mu_0) \cdot \text{vech} \left[ \Sigma_0^{-1} - \Sigma_0^{-1} (X_i - \mu_0)(X_i - \mu_0)' \Sigma_0^{-1} \right]' \mathbf{D}_p \right\}
\]

\[
= \frac{1}{2} E \left\{ (X_i - \mu_0) \cdot \text{vech} \left[ (X_i - \mu_0)(X_i - \mu_0)' \right]' \left( \Sigma_0^{-1} \otimes \Sigma_0^{-1} \right) \mathbf{D}_p \right\}
\]

\[
= \frac{1}{2} \Sigma_0^{-1} \Gamma_{X, \text{vech}(\Sigma)} \mathbf{D}_p' (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \mathbf{D}_p 
\]

(A-10)
and

\[
E \left[ \frac{\partial l_i(\beta_0)}{\partial \text{vech}(\Sigma_0)} \cdot \frac{\partial l_i(\beta_0)}{\partial \text{vech}(\Sigma_0)'} \right] \\
= \frac{1}{4}E \left[ D_p' \text{vec} \left[ \Sigma_0^{-1} - \Sigma_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\Sigma_0^{-1} \right] \cdot \text{vec} \left[ \Sigma_0^{-1} - \Sigma_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\Sigma_0^{-1} \right]' D_p \right] \\
= \frac{1}{4}D_p' E \left[ \text{vec} \left[ \Sigma_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\Sigma_0^{-1} \right] \cdot \text{vec} \left[ \Sigma_0^{-1}(X_i - \mu_0)(X_i - \mu_0)'\Sigma_0^{-1} \right]' - \text{vec}(\Sigma_0^{-1})\text{vec}(\Sigma_0^{-1})' \right] D_p \\
= \frac{1}{4}D_p' E \left[ (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \text{vec} \left[ (X_i - \mu_0)(X_i - \mu_0)' \right] \text{vec} \left[ (X_i - \mu_0)(X_i - \mu_0)' \right]' - (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \text{vec}(\Sigma_0) \cdot \text{vec}(\Sigma_0)'(\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \right] D_p \\
= \frac{1}{4}D_p' (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) E \left[ \text{vec}((X_i - \mu_0)(X_i - \mu_0)')(X_i - \mu_0)' \right) \text{vec}((X_i - \mu_0)(X_i - \mu_0)') \\
- \text{vec}(\Sigma_0) \cdot \text{vec}(\Sigma_0)' \right] (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) D_p \\
= \frac{1}{4}D_p' (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) D_p \Gamma \text{vech}(S),\text{vech}(S)' D_p' (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) D_p 
\tag{A-11}
\]

where vec(·) 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),

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

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

\[
B(\theta_*) \\
= E \left[ \frac{\partial l_i(\theta_*)}{\partial \theta_i} \cdot \frac{\partial l_i(\theta_*)}{\partial \theta_i'} \right] \\
\overset{\text{a}}{=} \dot{\sigma}_* \\
\overset{\text{a}}{=} \dot{\sigma}_* W \Gamma W \dot{\sigma}_* \\
\overset{\text{a}}{=} \dot{\sigma}_* W_* \Gamma W_* \dot{\sigma}_* \tag{A-13}
\]

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

\[
AE(T_{NML}) = nF_0 + \text{tr}(W \Gamma - A^{-1}(\theta_*)B(\theta_*))
\]
\[ \begin{align*}
\alpha = \ &= nF_0 + \text{tr}(W_1\Gamma - (\hat{\sigma}' W_1 \hat{\sigma})^{-1}\hat{\sigma}' W_1 \Gamma W_1 \hat{\sigma}) \\
\ &= nF_0 + \text{tr}(U\Gamma)
\end{align*} \]

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

\[ \text{AE}(T_{NML}) = nF_0 + df \]

References


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.


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

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

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

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

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

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>150</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>1000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{ML}$</td>
<td>1272</td>
<td>1113</td>
<td>1024</td>
<td>903</td>
<td>718</td>
<td>514</td>
</tr>
<tr>
<td>$T_1$</td>
<td>324</td>
<td>270</td>
<td>282</td>
<td>282</td>
<td>271</td>
<td>252</td>
</tr>
<tr>
<td>$T_2$</td>
<td>17</td>
<td>6</td>
<td>11</td>
<td>14</td>
<td>21</td>
<td>46</td>
</tr>
<tr>
<td>$T_3$</td>
<td>682/1998</td>
<td>491</td>
<td>448</td>
<td>416</td>
<td>364</td>
<td>285</td>
</tr>
<tr>
<td>$T_4$</td>
<td>129/1998</td>
<td>52</td>
<td>51</td>
<td>47</td>
<td>54</td>
<td>71</td>
</tr>
<tr>
<td>$T_5$</td>
<td>235</td>
<td>98</td>
<td>90</td>
<td>83</td>
<td>75</td>
<td>90</td>
</tr>
<tr>
<td>$T_6$</td>
<td>274</td>
<td>132</td>
<td>112</td>
<td>96</td>
<td>95</td>
<td>94</td>
</tr>
</tbody>
</table>
Table 4. Rejection rate of $T_{ML,0.05}$, $T_{1,0.05}$ to $T_{6,0.05}$
with for Example 1, normal condition, NR=2000

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>150</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>1000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{ML,0.05}$</td>
<td>132</td>
<td>83</td>
<td>73</td>
<td>56</td>
<td>47</td>
<td>27</td>
</tr>
<tr>
<td>$T_{1,0.05}$</td>
<td>51</td>
<td>34</td>
<td>36</td>
<td>35</td>
<td>38</td>
<td>24</td>
</tr>
<tr>
<td>$T_{2,0.05}$</td>
<td>62</td>
<td>35</td>
<td>41</td>
<td>35</td>
<td>39</td>
<td>23</td>
</tr>
<tr>
<td>$T_{3,0.05}$</td>
<td>103/1980</td>
<td>57</td>
<td>58</td>
<td>47</td>
<td>44</td>
<td>26</td>
</tr>
<tr>
<td>$T_{4,0.05}$</td>
<td>126/1980</td>
<td>67</td>
<td>60</td>
<td>50</td>
<td>44</td>
<td>27</td>
</tr>
<tr>
<td>$T_{5,0.05}$</td>
<td>117/1969</td>
<td>62</td>
<td>58</td>
<td>50</td>
<td>44</td>
<td>27</td>
</tr>
<tr>
<td>$T_{6,0.05}$</td>
<td>110/1969</td>
<td>61</td>
<td>58</td>
<td>49</td>
<td>44</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 5. Rejection rate of $T_{ML,0.05}$, $T_{1,0.05}$ to $T_{6,0.05}$
with for Example 1, mild nonnormal condition, NR=2000

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>150</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>1000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{ML,0.05}$</td>
<td>420</td>
<td>276</td>
<td>223</td>
<td>209</td>
<td>134</td>
<td>48</td>
</tr>
<tr>
<td>$T_{1,0.05}$</td>
<td>102</td>
<td>61</td>
<td>72</td>
<td>66</td>
<td>63</td>
<td>34</td>
</tr>
<tr>
<td>$T_{2,0.05}$</td>
<td>35</td>
<td>19</td>
<td>21</td>
<td>17</td>
<td>24</td>
<td>17</td>
</tr>
<tr>
<td>$T_{3,0.05}$</td>
<td>208/1991</td>
<td>129</td>
<td>122</td>
<td>104</td>
<td>80</td>
<td>39</td>
</tr>
<tr>
<td>$T_{4,0.05}$</td>
<td>112/1991</td>
<td>40</td>
<td>45</td>
<td>32</td>
<td>33</td>
<td>22</td>
</tr>
<tr>
<td>$T_{5,0.05}$</td>
<td>136/1993</td>
<td>59</td>
<td>59</td>
<td>44</td>
<td>37</td>
<td>23</td>
</tr>
<tr>
<td>$T_{6,0.05}$</td>
<td>153/1993</td>
<td>65</td>
<td>68</td>
<td>53</td>
<td>41</td>
<td>26</td>
</tr>
</tbody>
</table>

Table 6. Rejection rate of $T_{ML,0.05}$, $T_{1,0.05}$ to $T_{6,0.05}$
with for Example 1, severe nonnormal condition, NR=2000

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>150</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>1000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{ML,0.05}$</td>
<td>1222</td>
<td>1027</td>
<td>894</td>
<td>775</td>
<td>558</td>
<td>282</td>
</tr>
<tr>
<td>$T_{1,0.05}$</td>
<td>295</td>
<td>224</td>
<td>229</td>
<td>213</td>
<td>185</td>
<td>125</td>
</tr>
<tr>
<td>$T_{2,0.05}$</td>
<td>15</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>$T_{3,0.05}$</td>
<td>637/1998</td>
<td>415</td>
<td>367</td>
<td>332</td>
<td>233</td>
<td>142</td>
</tr>
<tr>
<td>$T_{4,0.05}$</td>
<td>108/1998</td>
<td>40</td>
<td>42</td>
<td>30</td>
<td>28</td>
<td>23</td>
</tr>
<tr>
<td>$T_{5,0.05}$</td>
<td>204</td>
<td>71</td>
<td>67</td>
<td>56</td>
<td>44</td>
<td>31</td>
</tr>
<tr>
<td>$T_{6,0.05}$</td>
<td>245</td>
<td>98</td>
<td>87</td>
<td>74</td>
<td>55</td>
<td>32</td>
</tr>
</tbody>
</table>
Table 7. Rejection rate of different statistics with $\alpha = .05$ for Example 2, normal condition, NR=2000

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

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

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

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

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>150</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>1000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{ML}$</td>
<td>390</td>
<td>317</td>
<td>256</td>
<td>272</td>
<td>190</td>
<td>150</td>
</tr>
<tr>
<td>$T_1$</td>
<td>145</td>
<td>139</td>
<td>129</td>
<td>146</td>
<td>134</td>
<td>130</td>
</tr>
<tr>
<td>$T_2$</td>
<td>81</td>
<td>49</td>
<td>37</td>
<td>51</td>
<td>66</td>
<td>86</td>
</tr>
<tr>
<td>$T_3$</td>
<td>233</td>
<td>199</td>
<td>165</td>
<td>188</td>
<td>150</td>
<td>139</td>
</tr>
<tr>
<td>$T_4$</td>
<td>139</td>
<td>86</td>
<td>68</td>
<td>84</td>
<td>91</td>
<td>97</td>
</tr>
<tr>
<td>$T_5$</td>
<td>152/1993</td>
<td>105</td>
<td>80</td>
<td>102</td>
<td>102</td>
<td>100</td>
</tr>
<tr>
<td>$T_6$</td>
<td>170/1988</td>
<td>116</td>
<td>93</td>
<td>109</td>
<td>108</td>
<td>105</td>
</tr>
</tbody>
</table>
Table 10. Rejection rate of $T_{ML,0.05}$, $T_{1,0.05}$ to $T_{6,0.05}$ with for Example 2, normal condition, NR=2000

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>150</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>1000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{ML,0.05}$</td>
<td>1561</td>
<td>1952</td>
<td>1997</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{1,0.05}$</td>
<td>1280</td>
<td>1905</td>
<td>1986</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{2,0.05}$</td>
<td>1368</td>
<td>1909</td>
<td>1987</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{3,0.05}$</td>
<td>1518</td>
<td>1942</td>
<td>1997</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{4,0.05}$</td>
<td>1579</td>
<td>1943</td>
<td>1997</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{5,0.05}$</td>
<td>1550/1999</td>
<td>1942</td>
<td>1997</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{6,0.05}$</td>
<td>1533/1999</td>
<td>1943</td>
<td>1997</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 11. Rejection rate of $T_{ML,0.05}$, $T_{1,0.05}$ to $T_{6,0.05}$ with for Example 2, mild nonnormal condition, NR=2000

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>150</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>1000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{ML,0.05}$</td>
<td>1679</td>
<td>1975</td>
<td>1996</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{1,0.05}$</td>
<td>1342</td>
<td>1937</td>
<td>1989</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{2,0.05}$</td>
<td>1287</td>
<td>1911</td>
<td>1980</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{3,0.05}$</td>
<td>1601</td>
<td>1965</td>
<td>1992</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{4,0.05}$</td>
<td>1560</td>
<td>1954</td>
<td>1992</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{5,0.05}$</td>
<td>1581/1999</td>
<td>1961</td>
<td>1991</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{6,0.05}$</td>
<td>1585/1997</td>
<td>1962</td>
<td>1991</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 12. Rejection rate of $T_{ML,0.05}$, $T_{1,0.05}$ to $T_{6,0.05}$ with for Example 2, severe nonnormal condition, NR=2000

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>150</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>1000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{ML,0.05}$</td>
<td>1818</td>
<td>1987</td>
<td>1999</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{1,0.05}$</td>
<td>1413</td>
<td>1956</td>
<td>1995</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{2,0.05}$</td>
<td>1109</td>
<td>1858</td>
<td>1973</td>
<td>1996</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{3,0.05}$</td>
<td>1703</td>
<td>1980</td>
<td>1998</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{4,0.05}$</td>
<td>1512</td>
<td>1942</td>
<td>1993</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{5,0.05}$</td>
<td>1585/1993</td>
<td>1964</td>
<td>1995</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$T_{6,0.05}$</td>
<td>1627/1988</td>
<td>1970</td>
<td>1997</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
</tbody>
</table>