Title

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

Authors

Li, Libo
Bentler, Peter M.

Publication Date

2006-10-24

Peer reviewed

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

Libo Li and Peter M. Bentler*

University of California, Los Angeles

August 21, 2006

*Research supported in part by grants DA00017 and DA01070 from the National Institute on Drug Abuse.

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

Abstract

In the traditional model comparison procedure, two nested structural models are hypothesized to be equal under some constraints, e.g., equality constraints. A strict null hypothesis is then evaluated by statistical tests to decide on the acceptance or rejection of the restrictions that differentiate the models. We propose instead to evaluate model close match, using the distance between two models in terms of the Kullback-Leibler (1951) Information Criterion, either as important supplementary information or as a criterion for nested structured model comparison. Based on the results of Vuong (1989) and Yuan, Hayashi and Bentler (2005), we develop some ADF-like generalized RMSEA tests for inference on model closeness. Simulation studies show that our proposed tests have robust and desirable performance in spite of severe nonnormality across several examples when sample size is as large as 150, and its relevance to educational research is shown with models for some TOEFL data. Consequently, a two-stage procedure which combines the traditional nested model comparison and the additional inferential information regarding model close match is further suggested to improve the typical practice of structured model modification.

Key words: Likelihood ratio statistic, RMSEA, model close match, asymptotics

1. Introduction

Structural equation modeling (SEM) with latent variables is an important research tool in the behavioral and social sciences such as psychology, sociology, marketing research etc. In a typical study, the mean and covariance implications of a model based on some sets of hypothesized linear relationships among interesting variables is tested against sample means and covariances. A variety of statistics are used for evaluating the adequacy of the model, including the classical normal theory based likelihood ratio (NTLR) test, Browne's asymptotically distribution free test (Browne, 1984), the Satorra-Bentler scaled test (Satorra & Bentler, 1988, 1994) or the more recent residual-based tests (Yuan & Bentler, 1997, 1998, 1999). The distribution, and hence performance, of all of these statistics depends on meeting the various assumptions underlying these statistics. One of these assumptions is that the strict null hypothesis holds, namely, that the model is exactly correct in the population. Model evaluations are in principle carried out in a classical way by evaluating where a test statistic falls in the distribution under the null. In turn, the actual performance of these statistics may be evaluated by such features as their type I and type II errors.

Another standard statistical issue in SEM involves the comparison of alternative models, especially nested models that contain additional restrictions beyond those of the more general model. A standard approach to such a model comparison involves the chi-square difference test (e.g., Joreskog, 1971; Steiger, Shapiro, & Browne, 1985). This is often an NTLR test; for greater robustness of this test, Satorra (2000) and Satorra and Bentler (2001) extended the Satorra-Bentler (SB) corrections to nested models. Difference tests require estimation of both the general and restricted models. Because it is often more convenient to work with only the more general model or only the more restricted model, the Lagrange Multiplier (LM) and Wald (W) tests were introduced for such model comparisons (e.g., Chou & Bentler, 1990; Lee & Bentler, 1980; Lee, 1985; Sorbom, 1989). As when a single model is evaluated, the distribution of these various statistics depends on meeting various assumptions (Satorra, 1989). One of the most important assumptions is the strict correctness of the null hypothesis, namely, that the parameters that differentiate the general and restricted models are precisely zero in the population. Similarly, significance of each of these statistics is determined with

reference to the assumed distribution under the null. While such a procedure may not work perfectly in practice (e.g., Yuan & Bentler, 2004), especially when such a model comparison is post hoc rather than a priori (e.g., MacCallum, Roznowski & Necowitz, 1992), some type of model comparison can not be avoided in practice. Most a priori models are incorrect in some way, and the process of model modification to yield improved models remains an inevitable and important part in the application of SEM (Joreskog, 1993). One rationale for imposing constraints on a general model is that the estimates in the more restricted and parsimonious model will be more precise (Bentler & Mooijaart, 1989).

Ever since Joreskog (1969) developed confirmatory factor analysis, these types of statistical tests have been embraced in SEM because they provide scientific rigor to testing hypotheses with nonexperimental data. After some limitations were raised on the role of testing in exploratory factor analysis (Tucker & Lewis, 1973), Bentler and Bonett (1980) noted that tests of exact fit in general SEM can not on their own provide a sufficient basis for evaluation of models, especially in large samples where any restrictive null hypothesis is liable to be rejected. They proposed that a model also needed to be evaluated in terms of the extent to which it explains covariances better than a most restricted model of uncorrelated variables which explains no covariances. They provided several so-called fit indices to evaluate such an increment in fit, and also proposed to evaluate differences in model fit between two nested models by evaluating the associated increment in fit. In the meantime, additional fit indices such as the root mean square error of approximation (RMSEA, Steiger & Lind, 1980), comparative fit index (CFI, Bentler, 1990), goodness of fit index (GFI, Joreskog & Sorbom, 1981) etc. have been devised to provide a measure of the extent of approximate or close fit of a model.

Critiques of tests of exact fit were also made from two other perspectives, namely from a rejection of the basic null hypothesis, and from the point of view of statistical theory. It does not make sense to test a specific model null hypothesis if one does not in the first place believe that a specific model might exist in the population. Any particular model may be nothing more than an approximation to reality, and it may be said that the modeling enterprise should mainly aim to provide information about the relative performance of alter-

native plausible models, none of which may be precisely true (e.g., Bentler & Bonett, 1980; de Leeuw, 1988; Browne & Cudeck, 1993; MacCallum, 2003). From the point of view of statistical theory, questions have been raised on whether the distribution of a test statistic under the null hypothesis provides the most meaningful possible model evaluation when such a null hypothesis may not make a priori sense. To provide an alternative, recently researchers such as Ogasawara (2005) and Yuan, Hayashi, and Bentler (2005) investigated the general distribution of the NTLR test under model misspecification and weak distributional assumptions on the data. In addition, some asymptotically robust model close fit tests implemented via the sample RMSEA also have been introduced and studied by Li and Bentler (2006).

These critiques of hypothesis testing on exact fit of a given model apply directly to the comparison of nested models, but little statistical development has been done to provide an alternative approach for comparing such models. In this paper, we first review some relevant statistical theories and propose a measure of close match between two competing models and their corresponding estimators. Then, using the results of Vuong (1989) and Yuan, Hayashi and Bentler (2005), the asymptotic distribution of these estimators will be derived and some asymptotic robust tests of close match between competing models will be defined. Finally, numerical examples will be given.

2. Theoretical Background

In classical single population structural equation modeling, the relationship of p-observed variables in a $p \times 1$ random vector $X = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$ and m-unobserved factors may have many different specifications. Without loss of generality, we only consider two such model specifications at one time for simplicity. In one parameterization, M_1 has q free unknown parameters which are included in a $q \times 1$ parameter vector θ, while another competing parameterization M_2 has r free unknown parameters which are included in an $r \times 1$ parameter vector γ. As a result, the hypothesized model M_1 leads to the model-implied mean $\mu(\theta)$ and covariance matrix $\mathbf{\Sigma}(\theta)$ and M_2 leads to $\mu(\gamma)$ and $\mathbf{\Sigma}(\gamma)$.

For simplicity, we assume that sampling yields a complete data set. Now let $\mu = E(X)$, $\mathbf{\Sigma} = \text{cov}(X)$, $\bar{X}$ and $\mathbf{S}$ be the corresponding mean and unbiased sample estimator and

$\mathbf{S}^* = (n-1)\cdot\mathbf{S}/n$ 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 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 its MLE and unbiased estimator separately. Although two estimators are different, their difference will be 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. The normal theory based log likelihood function of the observations is then given by

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

where $f(X_i;\beta)$ 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 counterparts 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, $\sqrt{n}(\hat{\beta}^* - \beta_0) \stackrel{a}{=} \sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{L} N(0, \mathbf{\Gamma})$, where $\stackrel{a}{=}$ 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) with

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

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

When μ and $\mathbf{\Sigma}$ are parameterized as M_1 and M_2 as mentioned before, then the corresponding log likelihood functions become

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

separately. Estimators of θ and γ can be obtained by minimizing the well-known normal theory maximum likelihood discrepancy functions (Browne & Arminger, 1995)

$$\begin{aligned} F_{ML}(\bar{X}, \mathbf{S}^*; \mu(\theta), \mathbf{\Sigma}(\theta)) &= (\bar{X} - \mu(\theta))' \mathbf{\Sigma}^{-1}(\theta)(\bar{X} - \mu(\theta)) + \log|\mathbf{\Sigma}(\theta)| + \mathrm{tr}(\mathbf{S}^* \mathbf{\Sigma}^{-1}(\theta)) \\ &\quad -\log|\mathbf{S}^*| - p \qquad (1) \\ F_{ML}(\bar{X}, \mathbf{S}^*; \mu(\gamma), \mathbf{\Sigma}(\gamma)) &= (\bar{X} - \mu(\gamma))' \mathbf{\Sigma}^{-1}(\gamma)(\bar{X} - \mu(\gamma)) + \log|\mathbf{\Sigma}(\gamma)| + \mathrm{tr}(\mathbf{S}^* \mathbf{\Sigma}^{-1}(\gamma)) \\ &\quad -\log|\mathbf{S}^*| - p \qquad (2) \end{aligned}$$

respectively. Consequently, the minimizers of the equations above, $\hat{\theta}_{NML}$ and $\hat{\gamma}_{NML}$, are the maximum likelihood estimators of θ and γ respectively. Plugging $\hat{\theta}_{NML}$ and $\hat{\gamma}_{NML}$ into (1) and (2), we obtain the statistics

$$\begin{aligned} T_{NML\_M1} \equiv nF_{ML}(\bar{X}, \mathbf{S}^*; \mu(\hat{\theta}_{NML}), \mathbf{\Sigma}(\hat{\theta}_{NML})) &= 2\left[l_n(\hat{\beta}^*) - l_n(\hat{\theta}_{NML})\right] \\ &= 2\sum_{i=1}^{n} \log\left[\frac{f(X_i; \hat{\beta}^*)}{f(X_i; \hat{\theta}_{NML})}\right] \\ &\equiv 2LR_n(\hat{\beta}^*, \hat{\theta}_{NML}) \qquad (3) \\ T_{NML\_M2} \equiv nF_{ML}(\bar{X}, \mathbf{S}^*; \mu(\hat{\gamma}_{NML}), \mathbf{\Sigma}(\hat{\gamma}_{NML})) &= 2\left[l_n(\hat{\beta}^*) - l_n(\hat{\gamma}_{NML})\right] \\ &= 2\sum_{i=1}^{n} \log\left[\frac{f(X_i; \hat{\beta}^*)}{f(X_i; \hat{\gamma}_{NML})}\right] \\ &\equiv 2 \cdot LR_n(\hat{\beta}^*, \hat{\gamma}_{NML}) \qquad (4) \end{aligned}$$

for M_1 and M_2 respectively. Clearly, T_{NML_M1} and T_{NML_M2} are the well-known NTLR test statistics for testing the exact fit of M_1 and M_2 respectively.

In SEM practice, instead of $F_{ML}(\bar{X}, \mathbf{S}^*; \mu(\theta), \mathbf{\Sigma}(\theta))$ and $F_{ML}(\bar{X}, \mathbf{S}^*; \mu(\gamma), \mathbf{\Sigma}(\gamma))$, the discrepancy functions $F_{ML}(\bar{X}, \mathbf{S}; \mu(\theta), \mathbf{\Sigma}(\theta))$ and $F_{ML}(\bar{X}, \mathbf{S}; \mu(\gamma), \mathbf{\Sigma}(\gamma))$ are used in most cases. Their minimizers, $\hat{\theta}_{ML}$ and $\hat{\gamma}_{ML}$, and the corresponding NTLR statistics $T_{ML\_M1} = NF_{ML}(\bar{X}, \mathbf{S}; \mu(\hat{\theta}_{ML}), \mathbf{\Sigma}(\hat{\theta}_{ML}))$ and $T_{ML\_M2} = NF_{ML}(\bar{X}, \mathbf{S}; \mu(\hat{\gamma}_{ML}), \mathbf{\Sigma}(\hat{\gamma}_{ML}))$ 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 these two sets of estimators and test statistics, such differences will become very slight as the sample size n increases (e.g., Bentler, 2006; Bollen, 1989; Browne & Arminger, 1995).

Let θ_* and γ_* be the minimizer of $F_{ML}(\mu_0, \mathbf{\Sigma}_0; \mu(\theta), \mathbf{\Sigma}(\theta))$ and $F_{ML}(\mu_0, \mathbf{\Sigma}_0; \mu(\gamma), \mathbf{\Sigma}(\gamma))$

respectively. Then it has been proved that under some standard regularity conditions (e.g., Kano, 1986; Shapiro, 1984), $\hat{\theta}_{NML}$ and $\hat{\gamma}_{NML}$, thus $\hat{\theta}_{ML}$ and $\hat{\gamma}_{ML}$, will be strongly consistent and asymptotically normally distributed (Yuan & Jennrich, 1998), and

$$\sqrt{n}(\hat{\theta}_{ML} - \theta_*) \stackrel{a}{=} \sqrt{n}(\hat{\theta}_{NML} - \theta_*) \xrightarrow{L} N(0, \Omega_{\theta_*}) \tag{5}$$

$$\sqrt{n}(\hat{\gamma}_{ML} - \gamma_*) \stackrel{a}{=} \sqrt{n}(\hat{\gamma}_{NML} - \gamma_*) \xrightarrow{L} N(0, \Omega_{\gamma_*}) \tag{6}$$

with $\Omega_{\theta_*} = A_{\theta_*}^{-1} B_{\theta_*} A_{\theta_*}^{-1}$ and $\Omega_{\gamma_*} = A_{\gamma_*}^{-1} B_{\gamma_*} A_{\gamma_*}^{-1}$.

3. Model Exact vs. Close Match

For convenience of illustration, we introduce the idea of model exact vs. close match by using an example by Curran, Bollen, Chen, Paxton and Kirby (2003) (see their population model 2). In this example, the population model underlying the data is as follows,

$$\mathbf{y} = \mathbf{\Pi}\eta + \epsilon \qquad\qquad \eta = \mathbf{B}\eta + \zeta$$

where ϵ and ζ are independent to each other with $E(\epsilon) = 0$, Cov(ϵ)=$\mathbf{\Psi}$, $E(\zeta) = 0$, Cov(ζ)=$\mathbf{\Xi}$. Moreover, $\mathbf{\Psi}$ = diag(.51, .51, .51, .51, .51, .2895, .51, .51, .51, .2895, .2895, .51, .51, .51, .51), $\mathbf{\Xi}$ = diag(.49, .3136, .3136),

$$\mathbf{\Pi} = \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & .30 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & .30 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & .30 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \end{bmatrix}'$$

$$\text{and} \qquad \mathbf{B} = \begin{bmatrix} .00 & .00 & .00 \\ .60 & .00 & .00 \\ .00 & .60 & .00 \end{bmatrix}.$$

For our illustration, we focus on four specifications used by Curran et al. (2003). They are: Specification 1 is properly specified, Specification 2 sets $\pi_{11,2}$ as zero, Specification 3 sets $\pi_{11,2}$ and $\pi_{10,3}$ as zero, and Specification 4 sets $\pi_{11,2}$, $\pi_{10,3}$ and $\pi_{6,1}$ as zero. During the model fitting of each specification, we set $\pi_{1,1}$, $\pi_{7,2}$ and $\pi_{12,3}$ to 1.0 for identification while all other nonzero parameters in the population model are set free. As a result, the models by Specification 1, 2, 3, and 4 have degrees of freedom equal to 85, 86, 87 and 88, respectively.

For model comparison, any pair of these four specifications can be a comparison pair. Let us denote the general model of a comparison pair as M_1 while the restricted one of the pair which is nested in M_1 as M_2. Then the equality constraints bridging M_1 and M_2 such as $\pi_{6,1} = 0$ for the comparison pair of Specification 3 vs. 4 can be reparameterized as the null hypothesis $H_0^E : \theta = \theta(\gamma)$. In this article, we consider $H_0^E : \theta = \theta(\gamma)$ as the null hypothesis of equality or exact match of the models M_1 and M_2. Clearly, when θ is equal to $\beta \equiv (\mu', \text{vech}(\mathbf{\Sigma})')'$, then $H_0^E : \theta = \theta(\gamma)$ becomes the null hypothesis of exact fit of M_2.

Since the omitted paths (cross loadings), $\pi_{11,2}$, $\pi_{10,3}$ and $\pi_{6,1}$, are equal to .30 in the population when compared all other loadings that are 1.0, we can consider these as minor cross loadings. In general, a simple cluster structure such as Specification 4 in this example will be very desirable from a theoretical perspective. However, real data hardly allow such a simple cluster structure, and typically may require a more complex factor loading structure, like the population model in this example where the cluster structure is compounded by some minor cross-loadings. As a result, the null hypothesis $H_0^E : \theta = \theta(\gamma)$ for any pair of the four specifications above is false, and related test statistics such as the NTLR statistic will reject this hypothesis if the sample size is large enough. Then the unwanted minor paths will be included in the final model, perhaps making it less interpretable.

What we illustrated here is a typical model comparison paradigm by the traditional approach. Like exact fit tests in model overall evaluation, the traditional approach to the model comparison involves choosing between the better fit of the general model M_1 and the parsimony or meaningfulness of the restricted model M_2 by examining a statistic assessing the equality or exact match of the nested models. Even though this approach is valuable, it may not be a complete one. In practice, $H_0^E : \theta = \theta(\gamma)$ may not hold because of some minor differences between two models, e.g., unexpected minor cross loadings as illustrated above. Even though these differences may be minor or unmeaningful substantively, the traditional exact match testing procedure would inevitably favor M_1 (especially in a large sample) because of the infeasibility of exact model equality. In practice, a more realistic approach to model comparison would decide between the better fit of M_1 and the parsimony or meaningfulness of M_2, using as a criterion the degree of close match instead of exact match

between two models[1]. In other words, like the concept of close fit in overall model evaluation, the concept of close match between M_1 and M_2 may yield an appropriate comparison of two models. As an additional approach to model comparison, this may yield a more practical criterion for model modification in substantive research.

Like many so-called fit indices in model close fit, we first need to define some measure of match between competing nested models. Let $F_1 = F_{ML}(\mu_0, \mathbf{\Sigma}_0; \mu(\theta_*), \mathbf{\Sigma}(\theta_*))$ and $F_2 = F_{ML}(\mu_0, \mathbf{\Sigma}_0; \mu(\gamma_*), \mathbf{\Sigma}(\gamma_*))$. Let $df_1 = p^* - q$ and $df_2 = p^* - r$ denote the degrees of freedom of M_1 and M_2 respectively, where $p^* = p + p(p+1)/2$. The well-known RMSEA indices (Browne & Cudeck, 1993; Steiger & Lind, 1980) corresponding to these models are defined as

$$\text{RMSEA}_1 = \sqrt{\frac{F_1}{df_1}} \qquad\qquad \text{RMSEA}_2 = \sqrt{\frac{F_2}{df_2}}$$

for M_1 and M_2 respectively. For this example, the true RMSEAs of the four specifications are 0, .0215, .0308 and .040 respectively.

In this article, we consider the RMSEAs above as sample size and model size independent measures of match between the preferred models and the saturated one. However, we need to generalize the concept behind this measure into one that provides a metric for comparing two different models M_1 and M_2. At the first glance, the difference between RMSEA_1 and RMSEA_2 would seem to be the obvious choice for such a measure. However, one problem with this difference is that when $H_0^E : \theta = \theta(\gamma)$ holds but neither model is exactly correct, $F_1 = F_2$ but the difference between RMSEA_1 and RMSEA_2 is not equal to zero. It varies with F_1, F_2, df_1 and df_2. Hence we conclude that this obvious difference does not accurately reflect the exact or close match of two competing models, and so we exclude it from further discussion.

Now let us look at these two quantities: $F_{12} = F_2 - F_1$ and $df_{12} = df_2 - df_1$. Like F_1 for M_1 and F_2 for M_2, F_{12} is a sample size independent measure of the degree of overmisspecification of M_2 compared to M_1 or the distance between M_1 and M_2. Further, in model comparison between M_1 and M_2, F_{12} is dependent on the number of the constraints df_{12} as well as the

[1]Another approach to the issue of the minor cross loadings or error covariances mentioned here is to use inequality constraints. We will address this approach with examples later.

model size df_1 or df_2.

We shall define a relative RMSEA, RMSEA_{12}, as the measure of the match between M_1 and M_2. That is,

$$\text{RMSEA}_{12} = \sqrt{\frac{F_{12}}{df_1}}. \tag{7}$$

At first glance, df_2 or df_{12} instead of df_1 similarly could be chosen for (7), since then RMSEA_{12} will be reduced to RMSEA_2 when M_1 is the saturated model.

However, this is not the whole story. Let us look at df_2 first. In a typical setting of traditional model comparison, M_1 and M_2 are assumed to be equal under the constraint set. Under such an assumption, it is reasonable to start from M_1 to add the constraints, or from M_2 to free them. However, this logic does not necessarily make sense when this assumption is violated. When $H_0^E : \theta = \theta(\gamma)$ doesn't hold, the constraints should be freed according to the exact match criterion, and then the only thing left is to decide if M_2 is still close enough to M_1. In this case the starting point of consideration must be M_1 and correspondingly df_1 instead of df_2 should be selected for RMSEA_{12} in (7) on account of the model size.

The more difficult choice may be between df_1 and df_{12}. Let $\text{RM}\tilde{\text{S}}\text{EA}_{12}$ denote the relative RMSEA calculated by using df_{12} instead of df_1 in (7). For better illustration, we calculated RMSEA_{12} and $\text{RM}\tilde{\text{S}}\text{EA}_{12}$ for all comparison pairs based on the four specifications in the example. The results are presented in Table A1 and A2 respectively. From two tables, it is clear that RMSEA_{12} has a similar scale as RMSEA while $\text{RM}\tilde{\text{S}}\text{EA}_{12}$ is different. Besides the gross similarities, a detailed comparison in how the indices change with varying misspecification may be more important. In each of the last two rows of the two tables, the patterns of change in index size are different. RMSEA_{12} decreases along each of the rows in its table while $\text{RM}\tilde{\text{S}}\text{EA}_{12}$ increases instead. However, in this setup the distance between Specification 1 vs. 3 in this example should be reasonably considered to be larger than the distance between Specification 2 vs. 3. As a result, a measure of model difference that is useful for model comparison purposes should correspondingly have a larger value when comparing Specification 1 vs. 3 than when comparing Specification 2 vs. 3. A similar trend should be seen in the comparison of Specification 4 against other three specifications as listed in last

row of two tables. In Table A1, RMSEA_{12} performs consistent with these expectations. Its values decrease along the rows of Table A1. In contrast, the performance of $\tilde{\text{RMSEA}}_{12}$ in Table A2 contradicts these expectations. Its values increase instead of decrease along the rows of Table A2. For example, in Table A2 the comparison of Specification 1 vs. 3 with a larger model discrepancy has a smaller $\tilde{\text{RMSEA}}_{12}$ value than Specification 2 vs. 3. The same phenomenon can be seen in the comparison pairs of Specification 4 with the other three specifications in Table A2.

In theory, $\tilde{\text{RMSEA}}_{12}$ is a measure of the overmisspecification of M_2 averaged by the number of constraints. Even though such averaging is used in RMSEA when assessing exact or close fit, it may not be the best approach for our purposes. Usually, different individual constraints in a constraint set have a different impact on the excessive restrictions in M_2 compared to M_1. Even though the average excess restriction can be small, some individual constraints or subsets in a constraint set may generate more overmisspecification than others and may even make a big difference if considered alone. For example, a cross loading of some significant magnitude may make a big difference if ignored in a model. However, if this cross loading is omitted together with some other minor cross loadings, then the average overmisspecification due to the exclusion of these cross loadings can be low. So for these reasons, it seems appropriate when comparing two competing structured models that the total overmisspecification caused by the constraints be considered, rather than the average one. Hence in this study, df_1 instead of df_{12} is used to define (7).

Browne and Cudeck (1993) pointed out that the null hypothesis of exact fit of M_1 and M_2 can be expressed in the form of RMSEA. Given the nested relationship between M_1 and M_2, the null hypothesis $H_0^E : \theta = \theta(\gamma)$ can be expressed alternatively as

$$H_0^E : \ \text{RMSEA}_{12} = 0. \tag{8}$$

By further extending the close fit expression in term of RMSEA proposed by Browne and Cudeck (1993), we may express the hypothesis of close match of M_1 and M_2 as

$$H_0 : \ \text{RMSEA}_{12} \leq a \tag{9}$$

where a is an arbitrary small positive value. The idea behind (9) is clear. A restricted

model like M_2 may not be equal to the unrestricted one M_1 as assumed in (8). However, if the distance between two models is small and in some tolerable range, the restricted model with more parsimony or meaningfulness is still a good candidate to replace the unrestricted one. In the SEM literature, .05, .08 and .10 have been widely used as cutoff values for the population RMSEA when evaluating model close fit. Clearly, by the same idea, some small positive values can be used to define cutoff values of RMSEA_{12} for evaluating model close match.

In Table A1, all comparison pairs on the diagonal line have a one single minor cross loading difference in specification and their RMSEA_{12} values are less than .03. When there are two minor cross loading differences as in the comparison of Specification 1 vs. 3 or Specification 2 vs. 4, the RMSEA_{12} values are then between .03 and .04. When there are three minor cross loading differences, as in the comparison of Specification 1 vs. 4, the RMSEA_{12} value becomes over .04. So in this example the omission of a single minor path increases the RMSEA_{12} value no more than .03 while the omission of two minor paths will increase it to between .03 and .04. Now if we consider the omission of a single minor path as when comparing Specification 3 vs. 4 as a minor difference between two models, then .03 can be used as a cutoff value for RMSEA_{12}. Correspondingly, if the omission of two minor paths as in the comparison of Specification 2 vs. 4 can be considered as a minor difference, then .04 can be used for RMSEA_{12}.

The issue of whether model exact match or model close match is best used in model comparison probably depends on the truth or correctness of the more general model. When the general model is true there may be greater interest in whether the more restricted model is also exactly true, i.e., model exact match may be of greater concern. Unfortunately, in real data analysis it is quite likely that both models in a nested model comparison are misspecified. Certainly it has been argued that models are always an approximation to reality and hence that no model will ever be precisely true. In that case, the comparison of nested models involves comparing two models that are both technically incorrect. In such a situation, a null hypothesis of exact model match would seem to make little sense and the key issue is how tolerable the difference between the models may be. Given a small discrepancy

in fit in term of RMSEA_{12} between such nested models, the parsimony or meaningfulness of the more restricted but less well fitting model can be a reason to accept the restrictions.

4. General distribution of likelihood ratio statistics

Let $\hat{F}_1 = F_{ML}(\bar{X}, \mathbf{S}; \mu(\hat{\theta}_{ML}), \mathbf{\Sigma}(\hat{\theta}_{ML}))$ and $\hat{F}_2 = F_{ML}(\bar{X}, \mathbf{S}; \mu(\hat{\gamma}_{ML}), \mathbf{\Sigma}(\hat{\gamma}_{ML}))$. Then $T_{ML\_12} = N\hat{F}_2 - N\hat{F}_1 \xrightarrow{L} \chi^2_{df_{12}}$ under normality is the NTLR test statistic that is used to test H_0^E in (8) in a nested model comparison. When H_0^E in (8) doesn't hold, the inequality of two nested models becomes true and T_{ML_12} follows $\chi^2_{df_{12}}(NF_{12})$ under normality and the population drift assumption which is

$$\mu_0 - \mu(\theta_*) = O(1/\sqrt{n}) \text{ and } \mathbf{\Sigma}_0 - \mathbf{\Sigma}(\theta_*) = O(1/\sqrt{n}) \tag{10}$$

$$\mu_0 - \mu(\gamma_*) = O(1/\sqrt{n}) \text{ and } \mathbf{\Sigma}_0 - \mathbf{\Sigma}(\gamma_*) = O(1/\sqrt{n}) \tag{11}$$

(e.g., Satorra, 1989; Satorra & Saris, 1985; Steiger, Shapiro, & Browne, 1985). Although this noncentral chi-square distribution of T_{ML_12} can be used for testing $H_0: \text{RMSEA}_{12} \leq a$ in (9) under the inequality of two nested models, the assumptions of normality and population drift are hard to satisfy or verify in practice. These limitations prevent T_{ML_12} from being the appropriate statistic to use in such practical testing situations. Satorra (1989) further proposed a generalized score test and generalized wald test which drop the assumption of normality and are asymptotically noncentral chi-square distributed under the inequality of two nested models. However, the noncentrality parameters of their noncentral chi-square distributions contain $\mathbf{\Gamma}$ which is based on the distribution of the data and varies with its nonnormality. Such distributional dependence of the noncentrality parameters, along with the requisite population drift assumption, similarly raise questions about the appropriateness of using these statistics for testing model close match in (9).

Given the inadequacy of existing methods for testing of model close match, we turn our attention to some results of Yuan, Hayashi and Bentler (2005). They applied the theory of Vuong (1989) to mean and covariance structure analysis and derived the asymptotic distribution of T_{ML_12} under the alternative hypothesis of H_0^E in (8). Given the unfamiliarity of Vuong's theory in the SEM literature, we give a brief explanation of Vuong's theory. Then we apply it to the issue of model close match.

Since the theory of Vuong (1989) focuses on T_{NML_12} instead of T_{ML_12}, let $T_{NML\_12} = T_{NML\_M2} - T_{NML\_M1}$. Then from (3) and (4), it can be easily found that

$$\frac{1}{2n}T_{NML\_12} = \frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{f(X_i;\hat{\theta}_{NML})}{f(X_i;\hat{\gamma}_{NML})}\right] \equiv \frac{1}{n}LR_n(\hat{\theta}_{NML},\hat{\gamma}_{NML}) \tag{12}$$

As mentioned before, $\hat{\theta}_{NML}$ and $\hat{\gamma}_{NML}$ are strongly consistent and $\hat{\theta}_{NML} - \theta_* = O_p(1/\sqrt{n})$, $\hat{\gamma}_{NML} - \gamma_* = O_p(1/\sqrt{n})$ under some standard regularity conditions. Given these properties of $\hat{\theta}_{NML}$ and $\hat{\gamma}_{NML}$, by using a Taylor expansion of $LR_n(\theta_*,\gamma_*)$ around $(\hat{\theta}'_{NML},\hat{\gamma}'_{NML})'$, we can get

$$\begin{aligned}
&\frac{1}{n}LR_n(\theta_*,\gamma_*) \\
&\quad\equiv\quad \frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{f(X_i;\theta_*)}{f(X_i;\gamma_*)}\right] \\
&\quad=\quad \frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{f(X_i;\hat{\theta}_{NML})}{f(X_i;\hat{\gamma}_{NML})}\right] + \frac{1}{n}\cdot V\cdot\left[(\theta'_*,\gamma'_*)' - (\hat{\theta}'_{NML},\hat{\gamma}'_{NML})'\right] + O_p(1/n)
\end{aligned} \tag{13}$$

where

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

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

$$\frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{f(X_i;\hat{\theta}_{NML})}{f(X_i;\hat{\gamma}_{NML})}\right] \quad=\quad \frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{f(X_i;\theta_*)}{f(X_i;\gamma_*)}\right] + O_p(1/n) \tag{14}$$

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

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

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

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

By the Central Limit Theorem,

$$\sqrt{n}\left\{\frac{1}{n}LR_n(\theta_*,\gamma_*) - E\left[\log\left[\frac{f(X_i;\theta_*)}{f(X_i;\gamma_*)}\right]\right]\right\} \quad\xrightarrow{L}\quad N(0,\omega^2) \tag{15}$$

Lemma 1. The following identity holds

$$E\left[\log\left[\frac{f(X_i;\theta_*)}{f(X_i;\gamma_*)}\right]\right] = \frac{1}{2}F_{12}$$

Proof.

$$\begin{aligned}
&E\left[\log\left[\frac{f(X_i;\theta_*)}{f(X_i;\gamma_*)}\right]\right] \\
&= E\left[\log f(X_t;\theta_*)\right] - E\left[\log f(X_i;\gamma_*)\right] \\
&= -\frac{1}{2}E\left[\log|\Sigma(\theta_*)| + (X_t-\mu(\theta_*))'\Sigma^{-1}(\theta_*)(X_t-\mu(\theta_*))\right] \\
&\quad +\frac{1}{2}E\left[\log|\Sigma(\gamma_*)| + (X_t-\mu(\gamma_*))'\Sigma^{-1}(\gamma_*)(X_t-\mu(\gamma_*))\right] \\
&= -\frac{1}{2}\left[\log|\Sigma(\theta_*)| + \mathrm{tr}(\Sigma_0\Sigma^{-1}(\theta_*)) + (\mu_0-\mu(\theta_*))'\Sigma^{-1}(\theta_*)(\mu_0-\mu(\theta_*))\right. \\
&\quad \left.-\log|\Sigma_0| - p\right] + \frac{1}{2}\left[\log|\Sigma(\gamma_*)| + \mathrm{tr}(\Sigma_0\Sigma^{-1}(\gamma_*)) + (\mu_0-\mu(\gamma_*))'\Sigma^{-1}(\gamma_*)(\mu_0-\mu(\gamma_*))\right. \\
&\quad \left.-\log|\Sigma_0| - p\right] \\
&= \frac{1}{2}(F_2 - F_1) \\
&= \frac{1}{2}F_{12}
\end{aligned}$$

Combining (12), (13), (14), (15) and Lemma 1, we obtain

$$\sqrt{n}\left\{\frac{1}{2n}T_{NML\_12} - \frac{1}{2}F_{12}\right\} \xrightarrow{L} N(0,\omega^2) \tag{16}$$

One point which should be mentioned is that this asymptotic approximation holds only when $\omega^2 \neq 0$. Vuong (1989) pointed out that the equivalence between $\omega^2 = 0$ and $f(X_i;\theta_*) = f(X_i;\gamma_*)$ holds in general (see Lemma 4.1 by Vuong). For nested models, Vuong (1989) showed that $f(X_i;\theta_*) = f(X_i;\gamma_*)$ and $\theta_* = \theta(\gamma_*)$ are equivalent to each other under standard regularity conditions (see Lemma 7.1 by Vuong). So a rejection of the equality or exact match of two nested models: $\theta_* = \theta(\gamma_*)$ or $\mathrm{RMSEA}_{12} = 0$, which is equivalent to $f(X_i;\theta_*) = f(X_i;\gamma_*)$, is a way to establish $\omega^2 \neq 0$ and should be conducted before the use of (16).

Now let $\hat{F}_{12} = \hat{F}_2 - \hat{F}_1$. By (16) and the asymptotic equivalence between T_{NML_12} and T_{ML_12}, we obtain the following corollary (Yuan, Hayashi & Bentler, 2005, Corollary 1)

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

$$\sqrt{n}\left(\hat{F}_{12} - F_{12}\right) \xrightarrow{L} N(0,4\omega^2)$$

if $\omega^2 \neq 0$ or equivalently if $\text{RMSEA}_{12} \neq 0$ when two models are nested.

Let $\sigma_{1*} = (\mu(\theta_*)', \text{vech}(\boldsymbol{\Sigma}(\theta_*))')'$, $\sigma_{2*} = (\mu(\gamma_*)', \text{vech}(\boldsymbol{\Sigma}(\gamma_*))')'$, $\dot{\sigma}_{1*} = \partial\sigma_{1*}/\partial\theta_*$ and $\dot{\sigma}_{2*} = \partial\sigma_{2*}/\partial\gamma_*$ respectively and let $\mathbf{D}_p$ be the duplication matrix as defined by Magnus and Neudecker (1988). Then we define

$$\begin{aligned}
\mathbf{W}_{1*} &\equiv \text{diag}\left[\boldsymbol{\Sigma}^{-1}(\theta_*), 2^{-1}\mathbf{D}_p'(\boldsymbol{\Sigma}^{-1}(\theta_*) \otimes \boldsymbol{\Sigma}^{-1}(\theta_*))\mathbf{D}_p\right] \\
\mathbf{W}_{2*} &\equiv \text{diag}\left[\boldsymbol{\Sigma}^{-1}(\gamma_*), 2^{-1}\mathbf{D}_p'(\boldsymbol{\Sigma}^{-1}(\gamma_*) \otimes \boldsymbol{\Sigma}^{-1}(\gamma_*))\mathbf{D}_p\right] \\
\mathbf{U}_1 &\equiv \mathbf{W}_{1*} - \mathbf{W}_{1*}\dot{\sigma}_{1*}(\dot{\sigma}_{1*}'\mathbf{W}_{1*}\dot{\sigma}_{1*})^{-1}\dot{\sigma}_{1*}'\mathbf{W}_{1*} \\
\mathbf{U}_2 &\equiv \mathbf{W}_{2*} - \mathbf{W}_{2*}\dot{\sigma}_{2*}(\dot{\sigma}_{2*}'\mathbf{W}_{2*}\dot{\sigma}_{2*})^{-1}\dot{\sigma}_{2*}'\mathbf{W}_{2*}
\end{aligned}$$

Then it has been shown that under the population drift assumption (10) and (11) (Yuan & Marshall, 2004),

$$\mathbf{AE}(T_{ML\_M1}) = NF_1 + \text{tr}(\mathbf{U}_1\boldsymbol{\Gamma}) \tag{17}$$

$$\mathbf{AE}(T_{ML\_M2}) = NF_2 + \text{tr}(\mathbf{U}_2\boldsymbol{\Gamma}) \tag{18}$$

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

$$\mathbf{AE}(T_{ML\_M1}) = NF_1 + df_1 \tag{19}$$

$$\mathbf{AE}(T_{ML\_M2}) = NF_2 + df_2 \tag{20}$$

Combining (17), (18), (19), (20) and Corollary 1, we obtain the following Corollary (Yuan, Hayashi & Bentler, 2005, Corollary 2 and 3)

Corollary 2. Under standard regularity conditions as in Yuan and Bentler (1997),

$$\sqrt{n}\left(\hat{F}_{12} - F_{12} - \frac{\text{tr}(\mathbf{U}_2\boldsymbol{\Gamma})}{n} + \frac{\text{tr}(\mathbf{U}_1\boldsymbol{\Gamma})}{n}\right) \xrightarrow{L} N(0, 4\omega^2)$$

if $\omega^2 \neq 0$ or equivalently if $\text{RMSEA}_{12} \neq 0$ when two models are nested. When normality is assumed, this reduces to

$$\sqrt{n}\left(\hat{F}_{12} - F_{12} - \frac{df_2}{n} + \frac{df_1}{n}\right) \xrightarrow{L} N(0, 4\omega^2)$$

Notice that Corollary 2 has no conflict with Corollary 1 because the extra term $\text{tr}(\mathbf{U}_1\mathbf{\Gamma})/n$, $\text{tr}(\mathbf{U}_2\mathbf{\Gamma})/n$, df_1/n and df_2/n in Corollary 2 approach zero as n goes to infinity. Vuong (1989) further gave a consistent estimator of ω^2, that is,

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

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

$$\hat{\omega}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left[\frac{f(X_i;\hat{\theta}_{ML})}{f(X_i;\hat{\gamma}_{ML})}\right]\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{f(X_i;\hat{\theta}_{ML})}{f(X_i;\hat{\gamma}_{ML})}\right]\right]^2 \tag{21}$$

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

Yuan, Hayashi and Bentler (2005) further derived an explicit form for ω^2 under various conditions and gave the corresponding estimators. Although their work is valuable, our 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 (21). 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 use $\hat{\omega}^2$ for the tests that follow.

5. Tests of model close match

In section 3, we proposed RMSEA_{12} as a measure of model close match and the null hypothesis $H_0 : \text{RMSEA}_{12} \leq a$ against its alternative $H_1 : \text{RMSEA}_{12} > a$ to be one way to test model close match. Clearly, by (7), the null and alternative hypotheses above can be written as $H_0 : F_{12} \leq df_1 \cdot a^2$ against $H_1 : F_{12} > df_1 \cdot a^2$. Suppose now for two nested models, $H_0 : F_{12} \leq df_1 \cdot a^2$ is true and $T_{ML\_12} \xrightarrow{L} \chi^2_{df_{12}}(NF_{12})$, then $\Pr\{T_{ML\_12} > \chi^2_{df_{12},.95}(N \times df_1 \times a^2)\} \to .05$. So a test of close match can be proposed for nested model comparison. The null hypothesis will be rejected in favor of the alternative if T_{ML_12} is greater than $\chi^2_{df_{12},.95}(N \times df_1 \times a^2)$. Otherwise, the null hypothesis can not be rejected.

The last section gave some asymptotic results for $\hat{F}_{12}$, and a consistent estimator of ω^2 was given in (21). Based on these results, we propose two test statistics for model close match. The first is Vuong's test statistic (T_1), which is

$$T_1 = \frac{\sqrt{n}\left(\hat{F}_{12} - df_1 \cdot a^2 - df_2/n + df_1/n\right)}{2\hat{\omega}}$$

as well as a robust version of Vuong's test statistic (T_2), which is

$$T_2 = \frac{\sqrt{n}\left(\hat{F}_{12} - df_1 \cdot a^2 - \text{tr}(\hat{\mathbf{U}}_2\hat{\mathbf{\Gamma}})/n + \text{tr}(\hat{\mathbf{U}}_1\hat{\mathbf{\Gamma}})/n\right)}{2\hat{\omega}}$$

where $\hat{\mathbf{\Gamma}}$ is the consistent estimator of $\mathbf{\Gamma}$ (e.g., Bentler, 2006), and $\hat{\mathbf{U}}_1$ and $\hat{\mathbf{U}}_2$ are consistent estimators of $\mathbf{U}_1$ and $\mathbf{U}_2$ obtained by replacing θ_* and γ_* by $\hat{\theta}_{ML}$ and $\hat{\gamma}_{ML}$ respectively.

Corollary 3. Given $\omega^2 \neq 0$ or equivalently $\text{RMSEA}_{12} \neq 0$ when two models are nested, then under some standard regularity conditions as in Yuan and Bentler (1997)

$$T_1 \overset{a}{=} T_2 \overset{L}{\longrightarrow} N(\sqrt{n}\delta_1, 1) \qquad \text{and} \qquad \delta_1 = \frac{df_1}{\omega} \cdot \left[\frac{a_0^2 - a^2}{2}\right]$$

where a_0 is the value of RMSEA_{12}. Further,

1. When $\text{RMSEA}_{12} = a$, then $\delta_1 = 0$ and $T_1 \overset{a}{=} T_2 \overset{L}{\longrightarrow} N(0, 1)$.

2. When $\text{RMSEA}_{12} > a$, then $\delta_1 > 0$ and $T_1 \overset{a}{=} T_2 \longrightarrow +\infty$ as $n \longrightarrow +\infty$.

3. When $\text{RMSEA}_{12} < a$, then $\delta_1 < 0$ and $T_1 \overset{a}{=} T_2 \longrightarrow -\infty$ as $n \longrightarrow +\infty$.

Let $\lambda_{.95}$ be 95 percent quantile of the standard normal distribution, then $\Pr\{T_1 \text{ or } T_2 > \lambda_{.95}\} \rightarrow .05$. Clearly, T_1 and T_2 can be used to test (9) if $\omega^2 \neq 0$ or equivalently if $\text{RMSEA}_{12} \neq 0$ when two models are nested. For each of them, (9) will be rejected if it is greater than $\lambda_{.95}$. Otherwise, it can not be rejected.

In an earlier section, we proposed RMSEA_{12} as a measure of model close match. Since the true value of RMSEA_{12} is unknown, we define a sample RMSEA_{12} as its estimate. That is,

$$\widehat{\text{RMSEA}}_{12} = \sqrt{\max\left(\frac{\hat{F}_{12}}{df_1} - \frac{df_2}{n \cdot df_1} + \frac{1}{n}, 0\right)}.$$

Notice that in the definition above we use $1/n$ instead of $1/N$ as suggested by (19) and (20). However, the difference between $1/n$ and $1/N$ will be tiny when n is reasonably large.

From (17) and (18), the asymptotic bias of T_{ML_M1} and T_{ML_M2} differs from df_1 and df_2 in nonnormal conditions. So it is not hard to define a robust sample RMSEA_{12} as follows

$$\widetilde{\text{RMSEA}}_{12} = \sqrt{\max\left(\frac{\hat{F}_{12}}{df_1} - \frac{\text{tr}(\hat{\mathbf{U}}_2\hat{\mathbf{\Gamma}})}{n \cdot df_1} + \frac{\text{tr}(\hat{\mathbf{U}}_1\hat{\mathbf{\Gamma}})}{n \cdot df_1}, 0\right)}.$$

Obviously, $\widehat{\text{RMSEA}}_{12}$, $\widetilde{\text{RMSEA}}_{12}$ and RMSEA_{12} are some power transformations of $\hat{F}_{12}$ and F_{12}. By Corollary 1 and the Delta method, we obtain the following approximation.

Corollary 4. Given $\omega^2 \neq 0$ or equivalently if $\text{RMSEA}_{12} \neq 0$ when two models are nested, then under some standard regularity conditions as in Yuan and Bentler (1997)

$$\sqrt{n}\left(\widehat{\text{RMSEA}}_{12} - \text{RMSEA}_{12}\right) \xrightarrow{L} N\left(0, \frac{\omega^2}{df_1 \cdot F_{12}}\right)$$

and

$$\sqrt{n}\left(\widetilde{\text{RMSEA}}_{12} - \text{RMSEA}_{12}\right) \xrightarrow{L} N\left(0, \frac{\omega^2}{df_1 \cdot F_{12}}\right)$$

Proof.

$$\begin{aligned}
\sqrt{n}\left(\widehat{\text{RMSEA}}_{12} - \text{RMSEA}_{12}\right) &\overset{a}{=} \sqrt{n}\left(\sqrt{\hat{F}_{12}/df_1} - \sqrt{F_{12}/df_1}\right) \\
&\xrightarrow{L} N\left(0, \frac{\omega^2}{df_1 \cdot F_{12}}\right) \qquad \text{(Delta method)}
\end{aligned}$$

The distribution of $\widetilde{\text{RMSEA}}_{12}$ can be proved in the same way.

By Corollary 4, we define the RMSEA_{12} test statistic T_3 which is

$$T_3 = \frac{\sqrt{n}\left(\widehat{\text{RMSEA}}_{12} - a\right)}{\hat{\omega}\Big/\sqrt{df_1 \cdot \left(\hat{F}_{12} - df_2/n + df_1/n\right)}}$$

and the robust RMSEA_{12} test statistic T_4 which is

$$T_4 = \frac{\sqrt{n}\left(\widetilde{\text{RMSEA}}_{12} - a\right)}{\hat{\omega}\Big/\sqrt{df_1 \cdot \left(\hat{F}_{12} - df_2/n + df_1/n\right)}}$$

Let $\hat{c} = (\text{tr}(\hat{\mathbf{U}}_2\hat{\mathbf{\Gamma}}) - \text{tr}(\hat{\mathbf{U}}_1\hat{\mathbf{\Gamma}}) - df_2 + df_1)/n$. Following Li and Bentler (2006), we further define another two RMSEA_{12} test statistics T_5 and T_6 as

$$T_5 = \frac{\sqrt{n}\left(\widetilde{\text{RMSEA}}_{12} - a\right)}{\sqrt{\hat{\omega}^2 - \hat{c}}\Big/\sqrt{df_1 \cdot \left(\hat{F}_{12} - df_2/n + df_1/n + \hat{c}\right)}}$$

and

$$T_6 = \frac{\sqrt{n}\left(\widetilde{\text{RMSEA}}_{12} - a\right)}{\sqrt{\hat{\omega}^2 - 2.5 \cdot \hat{c}} \Big/ \sqrt{df_1 \cdot \left(\hat{F}_{12} - df_2/n + df_1/n + \hat{c}\right)}}$$

Clearly, $\hat{c}$ is an estimator of $c_0 = (\text{tr}(\mathbf{U}_2\mathbf{\Gamma}) - \text{tr}(\mathbf{U}_1\mathbf{\Gamma}) - df_2 + df_1)/n$ and converges to c_0 in the order of $O_p(n^{-3/2})$. When the data is normal, 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 information on nonnormality of the data. So compared to T_3, T_4 has a correction in the numerator while 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 two models are nested, one or several quantities among $\hat{F}_{12} - df_2/n + df_1/n$, $\hat{F}_{12} - df_2/n + df_1/n + \hat{c}$, $\hat{\omega}^2 - \hat{c}$ and $\hat{\omega}^2 - 2.5 \cdot \hat{c}$ can be less than or equal to zero especially in a small sample. 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 a problem will be discarded.

Corollary 5. Given $\omega^2 \neq 0$ or equivalently if $\text{RMSEA}_{12} \neq 0$ when two models are nested, then under some standard regularity conditions as in Yuan and Bentler (1997)

$$T_3 \stackrel{a}{=} T_4 \stackrel{a}{=} T_5 \stackrel{a}{=} T_6 \stackrel{L}{\longrightarrow} N(\sqrt{n}\delta_2, 1) \qquad \text{and} \qquad \delta_2 = \frac{df_1}{\omega} \cdot (a_0^2 - a_0 \cdot a)$$

where a_0 is the value of RMSEA_{12}. Further,

1. When $\text{RMSEA}_{12} = a$, 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 $\text{RMSEA}_{12} > a$, 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 $\text{RMSEA}_{12} < a$, 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, after a rejection of exact match, like T_1 and T_2 discussed before, T_3, T_4, T_5 or T_6 also can be used to test the hypothesis of close match in (9). The null hypothesis will be rejected for each statistic if its estimate is greater than $\lambda_{.95}$. Otherwise, it can not be rejected.

Corollary 6. 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 reject the hypothesis of model close match.

Proof. By Corollary 3 and 5,

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

where

$$\delta_1 = \frac{df_1}{\omega} \cdot \left[\frac{a_0^2 - a^2}{2}\right] \qquad\qquad \delta_2 = \frac{df_1}{\omega} \cdot (a_0^2 - a_0 \cdot a)$$

It is not hard to prove that $\delta_2 > \delta_1$ when $a_0 > a > 0$.

6. Examples

In the sections above, we discussed seven test statistics and their corresponding critical values for testing. They are T_{ML_12}, T_1, T_2, T_3, T_4, T_5 and T_6. In order to establish these statistics as reliable tools for testing $H_0 : \text{RMSEA}_{12} \leq a$, we first need to look at the asymptotic approximation and one-sided type I errors of these statistics when $\text{RMSEA}_{12} = a$. It is hard to manipulate the level of RMSEA_{12} to a specific value a such as the suggested cutoff value $a = .03$. Instead, we set a equal to the value of RMSEA_{12} for all statistics since RMSEA_{12} 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 should not be used.

Since the statistics we proposed are asymptotically distribution free, we generated data under three distribution conditions for each of three 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 data generation. In the severe nonnormal condition, they are set to 2.0 and 7.0. For all examples, the sample size levels are set to 150, 300, 500 and 1000. So there are $3 \times 4 = 12$ 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 $(\hat{\dot{\sigma}}_*{}'\hat{\mathbf{W}}_*\hat{\dot{\sigma}}_*)$, the Fisher information estimator, in $\hat{\mathbf{U}}_1$ and $\hat{\mathbf{U}}_2$ is replaced by the estimator of the Hessian or observed information matrix.

Example 1: Unwanted Paths. The example in section 3 is our first example. It contains some unwanted paths, which occurs frequently in SEM practice. For this example, we only study the performance of the comparison pairs on the diagonal line of Table A1. For all these pairs, a in the statistics is set to the RMSEA_{12} value of that pair. Then the rejection rates of all statistics for Specification 1 vs. 2, Specification 2 vs. 3 and Specification 3 vs. 4 are presented in Table 1A-1C, Table 2A-2C and Table 3A-3C respectively.

In Table 1A, under the normal condition, T_{ML_12} performs well across the sample sizes. However, in Tables 1B and 1C, across the sample sizes, the inflation of the rejection rates of T_{ML_12} increases as the nonnormality of the data increases. Its performance is poor, especially with severe nonnormal data. Unlike T_{ML_12}, T_1 and T_2 perform poorly across all sample sizes in the three tables. They overaccept in all conditions. As we mentioned before, T_3, T_4, T_5 and T_6 can be undefined if some elements in the denominators of their definitions are less than or equal to zero, and then will be discarded from the analysis. We put the number of undefined replications into parenthesis after the rejection rates in each cell. In Tables 1A-1C, T_3, T_4, T_5 and T_6 are undefined over ten percent of the time when $n = 150$. Their failures increase as nonnormality increases, but failure reduces dramatically when n increases to 300 for all data conditions. Their rejection performance is consistent across the sample sizes in the three tables. The rejection rates are close to the target .05 for all conditions, even though they, and especially T_4, have a slight tendency to overaccept across the tables.

The results on these seven statistics for Specification 2 vs. 3 in Tables 2A-2C, and Specification 3 vs. 4 in Tables 3A-3C, are very similar. When the data is normal, T_{ML_12} performs well for Specification 2 vs. 3 and Specification 3 vs. 4 as long as the sample size is 150. However, for both specification pairs, overrejections of the null hypothesis occurs for all sample sizes as the nonnormality of the data increases. Its performance becomes especially

poor when the data is severely nonnormal. As before, T_1 and T_2 perform poorly across the sample sizes in all six tables. With the same data, the rejection rates of T_3, T_4, T_5 and T_6 for Specification 2 vs. 3 and Specification 3 vs. 4 reach the target level of .05 more closely than in Specification 1 vs. 2 in most conditions across the six tables. This may be due to the increased RMSEA_{12} values of these two specification pairs (see Table A1). Also in three tables of each specification pair in this example, the number of undefined cases for the four statistics generally decreases in the corresponding cells along three specification pairs. This also may be due to the increased RMSEA_{12} values along these pairs.

In total, in this example, T_{ML_12} in normal data and T_3, T_5 and T_6 in general have a desirable rejection performance across specification pairs as long as the sample size reaches 150. Based on these results, we believe that the new statistics should be helpful in practice for testing the magnitude of paths and rejecting the unwanted or unnecessary minor ones in a model.

Until this point, we have not focused on inequality constraints. Theoretically, imposing an inequality constraint on unwanted or unnecessary minor paths such as $\pi_{6,1} \leq .4$ and testing this by the likelihood ratio test (see Dijkstra, 1992; Shapiro, 1985) is also a possible way of handling minor paths. Unfortunately to our knowledge, there is no development of an appropriate methodology for such purpose, with existing approaches to inequality constraints requiring a correctly specified fitting function. This requirement limits their application to close fitting models, especially when the data is nonnormal.

Example 2: Model Uncertainty. In previous example, a clean structure is compounded with some undesirable paths. In these types of cases, a researcher may have a strong a priori reason to reject unwanted paths in spite of the lack of support from exact match based test statistics such as the NTLR test. Clearly, in this situation, our close match based test statistics can provide some help for deciding between models.

A perhaps more typical situation occurs when a researcher does not have a strong substantive preference for a specific model. In SEM practice, there often may be many models that can be considered for one single data that are meaningful substantively. This is certainly true in exploratory factor analysis. For example, both a two factor model and a three factor

model may be interpretable for some psychological data. Typically, the NTLR test will give support to the model with more parameters if the extra factor in the three factor model can capture some extra characteristics of the population. On the other hand, methods like AIC and BIC, due to their assigning a penalty for more parameters in the model, sometimes may yield the opposite result. Similarly, other indicators such as χ^2/df ratio or various fit indices may indicate only a small distance between the models. Even though these supplementary criteria are valuable, they are not probabilistic criteria and hence they do not optimally allow inference on the discrepancy between models in the population. As a remedy, our close match based test statistics are probabilistic decision criteria like the traditional exact testing or NTLR tests. By their very definition, like AIC and BIC, our close match approach already includes a tradeoff between model goodness of fit and model parsimony.

In this example, we illustrate our close match approach to solving model uncertainty by using an example from a TOEFL® iBT test[2] developed by the Educational Testing Service. We treat an empirical data set as a population, sample from this population under several circumstances, and evaluate the performance of our approach under this somewhat more realistic situation.

For this TOEFL® iBT test, the variables in the original data (n=774) were grouped within each of four test sections: Speaking, Writing, Reading and Listening. After some parceling, there are 22 variables: six variables for Speaking, two variables for Writing, eight variables for Reading and six variables for Listening. In the language assessment area, there is not a consensus on the number of factors underlying data such as this. As a result, models with a different number of factors have been hypothesized and studied (Bachman, Davidson, Ryan, & Choi, 1995; Carroll, 1983; Hale, Rock, & Jirele, 1989; Kunnan, 1995; Swinton & Powers, 1980). In our example, we focus on only three specifications. In Specification 1, there are three factors: one factor is for all variables in the Speaking section, one factor for all variables in the Writing section, and one factor for all variables in the Reading and Listening section. In Specification 2, there are two factors: one factor is for all variables in

[2]TOEFL® is a registered trademark of Educational Testing Service (ETS), which provided the data for this study. This publication is not endorsed or approved by ETS.

the Speaking section and another for all other variables in the test. In Specification 3, there are also two factors: one factor is for all variables in the Speaking and Writing sections and another for all variables in the Reading and Listening sections. In each specification above, all factors are hypothesized to be correlated with each other.

In traditional simulation studies, the data is generated by a predefined true model, much as we did in Example 1. In reality, a true model may not exist, or if it does, it may be disturbed or distorted by many other sources. Thus many models can be close fitted to such a population. As stated, in this example the sample covariance matrix of the 22 variables from the TOEFL® iBT test data is treated as the population matrix. We do not know its true structure, but whatever its structure, normal, mild and severely nonnormal samples with different sample sizes are generated from this matrix. The rationale behind this research paradigm is that the sample is a representative of the population underlying the TOEFL® iBT test and the samples generated from the sample covariance matrix actually represent something like parametric bootstrap-like samples. Then, as in typical bootstrap analysis, fitting the three specifications in last paragraph to the sample and the obtained bootstrap-like samples mimics the fitting to the unknown population and its many possible samples. However, unlike the bootstrap, we are able to control the distribution of the variables in our samples.

These three specifications have 206, 208, and 208 degrees of freedoms, respectively, and their population RMSEAs are .061, 0.067 and 0.072 respectively. By the widely-used cutoff values for the population RMSEA, they all have some mild misspecifications. However, the differences among these three true RMSEAs are minor. Presumably, the sample RMSEAs (or other indices mentioned before) also will imply minor misspecification, and model uncertainty will result if there is not a strong substantive preference for a particular parameterization.

Clearly, Specifications 2 and 3 are nested in Specification 1. We present the RMSEA_{12} values of each nested pair in Table B. Although three specifications in terms of RMSEA are not very distinguishable, in terms of RMSEA_{12} one could choose between models using our cutoff value .03 for model close match. Using that cutoff, the difference between Specification 1 vs. 2 can be considered as minor while the difference between Specification 1 vs. 3 is not

ignorable. Although the Specification 1 will have a better fit due to more parameters, the ignorable difference between Specifications 1 and 2 makes Specification 2 a good candidate to replace Specification 1, while the nonignorable difference between Specification 1 and 3 would propose a rejection of Specification 3. Like Example 1, we first examine how our proposed statistics would evaluate two nested pairs in terms of type I error when confronted with a sample from this population. The simulation results regarding these statistics on the two comparison pairs above are presented in Tables 4A-5C.

One surprising result in Tables 4A-5C is that, in contrast to Example 1, T_{ML_12} comparing two nested pairs now performs poorly under both normal and nonnormal conditions. One possible reason may be a violation of the population drift assumption in this example. As before, T_1 and T_2 perform poorly most of the time in all six tables. When the data is normal or mildly nonnormal, T_3, T_4, T_5 and T_6 have similar rejection patterns that are close to the target one, even when $n = 150$. However, when the data is severely nonnormal, T_3 still performs very well at all sample sizes while T_4 overaccepts the null hypothesis exclusively. The performance of T_5 and T_6 under the severe nonnormality is somewhere between those of T_3 and T_4. They are better than T_4 but have some general tendency to overaccept.

So far, we set $a = \text{RMSEA}_{12}$ in all statistics for all specification pairs. Given that the RMSEA_{12} values of two specification pairs in this example are different from our proposed cutoff value .03, so in the next step we set a in all seven statistics to .03 for all specification pairs and look at the acceptance or rejection performance of these statistics when $a \neq \text{RMSEA}_{12}$. Let $T_{ML\_12,0.03}$, $T_{1,0.03}$, $T_{2,0.03}$, $T_{3,0.03}$, $T_{4,0.03}$, $T_{5,0.03}$, and $T_{6,0.03}$ denote the corresponding test statistics when a is set to .03. Ideally, we expect the close match hypothesis to be always accepted or rejected, i.e., hardly ever rejected or accepted, depending on if the RMSEA_{12} value of the specification pair is less than or greater than .03. The rejection rates of these statistics in the simulation are presented in Tables 6A-6C (Specification 1 vs. 2) and Tables 7A-7C (Specification 2 vs. 3).

The results, shown in Tables 6A-7C, basically match our expectation. Overall, all statistics intend to accept Specification 1 vs. 2 completely, while rejecting Specification 1 vs. 3 completely as n increases. Compared to the other six statistics, $T_{ML\_12,0.03}$ performs poorly

for acceptance in Tables 6A-6C, while it does better for rejection in Tables 7A-7C. For Specification 1 vs. 2, $T_{1,0.03}$ and $T_{2,0.03}$ perform better than $T_{3,0.03}$, $T_{4,0.03}$, $T_{5,0.03}$ and $T_{6,0.03}$ in general. But $T_{3,0.03}$, $T_{4,0.03}$, $T_{5,0.03}$ and $T_{6,0.03}$ have more power to reject than $T_{1,0.03}$ and $T_{2,0.03}$, as expected when comparing Specification 1 vs. 3, although they reach rejection rates above 90% only in the normal condition with $n = 1000$.

7. Discussion

The close match approach and related statistics provides a new approach to model comparison. We believe that the methods provide an additional tool for evaluating a preferred model which may be rejected by a traditional exact match based test. In addition to avoiding limitations of the traditional exact match approach, they provide a new alternative in SEM to such common model comparison methods as AIC, BIC, χ^2/df ratio and the difference between fit indices.

Our simulation results in two examples show that T_3, T_5 and T_6 perform well in terms of type I error rates across different data conditions when n is as large as 150. As to the power to reject, the simulation results are consistent with Corollaries 3, 5 and 6, although sometimes a large sample size is needed to achieve complete rejection (e.g., Tables 7A-7C). Additional power analysis of the proposed test statistics can be conducted based on these Corollaries. And obviously, cutoff values other than .03 may also be interesting to explore.

One important point we want to emphasize again is that $\text{RMSEA}_{12} = 0$, i.e., model exact match, must be rejected in order to appropriately use T_3, T_5 and T_6 for comparing specification pairs. Since in practice one does not know exactly whether this requirement is satisfied or not, it makes sense that in a specific study one should first conduct some evaluation of the exact match hypothesisis. Of course there are many possible tests for evaluating exact match, including the NTLR test, LM test or Wald test under normality, or an asymptotically distribution free test such as the Satorra-Bentler scaled difference test (Satorra, 2000; Satorra & Bentler, 2001) or generalized score or Wald tests (Satorra, 1989). Thus we propose to use a sequential two-stage procedure for overall nested model comparison: accept the restricted model if it satisfies an exact match test; or, accept the model if it is

rejected by the exact match test but still satisfies one of the close match tests such as T_3, T_5 and T_6.

One potential problem of the two-stage procedure above is its significance level during overall nested model comparison. Notice that $H_0 : \text{RMSEA}_{12} \leq a$ is a composite of H_0^E and $H_0 - H_0^E$. Let T_E denote some reliable exact match test statistic such as the Satorra-Bentler scaled difference test, and let T_C denote some reliable close match test statistic such as $T_{3,0.03}$, $T_{5,0.03}$ and $T_{6,0.03}$. Further, let $A \equiv \{\ T_E > \chi^2_{df_{12},\alpha}\}$ and $B \equiv \{\ T_C > \lambda_\alpha\}$. Then Pr[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$(A|H_0^E) \to \alpha_E$ and Pr$(B|H_0 - H_0^E) \to \alpha_C$. 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 Li and Bentler (2006), we further demonstrate that tr$(\mathbf{U}_1\mathbf{\Gamma})$ and tr$(\mathbf{U}_2\mathbf{\Gamma})$, which are widely used in the Satorra-Bentler procedure and our RMSEA test statistics, are special cases of a more general term based on the likelihood ratio principle. Given this result and the generality of the likelihood ratio, it seems that our close match test statistics, and hence the two-stage procedure of nested model comparison, may be extendable to a wide variety of situations where the likelihood ratio principle applies. Clearly, this would tremendously increase the scope of application of the proposed methodology.

References

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

Bachman, L. F., Davidson, F., Ryan, K., & Choi, I-C. (1995). *An investigation into the comparability of two tests of English as a foreign language.* Cambridge, England: Cambridge University Press.

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.

Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin, 106*(2), 315-317.

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

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.

Carroll, J. B. (1983). Psychometric theory and language testing. In J. W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 80-107). Rowley, MA: Newbury House.

Chou, C.-P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange multiplier, and Wald tests. *Multivariate Behavioral Research, 25,* 115-136.

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.

Dijkstra, T. K. (1992). On statistical inference with parameter estimates on the boundary of the parameter space. *British Journal of Statistical and Mathematical Psychology, 45,* 289-309.

Hale, G. A., Rock, D. A., & Jirele, T. (1989). *Confirmatory factor analysis of the Test of English as a Foreign Language* (TOEFL Research Rep. No. RR-32; ETS RR-89-42). Princeton, NJ: ETS.

Joreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183-202.

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

Joreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294-316). Newbury Park, CA: Sage.

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.

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

Kunnan, A. J. (1995). *Test taker characteristics and test performance: A structural modeling approach.* Cambridge, England: Cambridge University Press

Lee, S. Y. (1985). Analysis of covariance and correlation structure. *Computational Statistics and Data Analysis, 2,* 279-295.

Lee, S. Y. & Bentler, P. M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *South African Statistical Journal, 14,* 121-136.

Li, L. & Bentler, P. M. (2006). Robust Statistical Tests for Evaluating the Hypothesis of Close Fit of Misspecified Mean and Covariance Structural Models. UCLA Statistics Preprint #494.

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

MacCallum, R., Roznowski, M., & Necowitz, L.B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490-504.

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

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

Ogasawara, H. (2005). Approximations to the distributions of fit indexes under fixed alternatives in normal and nonnormal samples. Under review.

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

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In Heijmans, R.D.H., Pollock, D.S.G. & Satorra, A. (eds.), *Innovations in multivariate statistical analysis. A Festschrift for Heinz Neudecker* (pp.233-247). London: Kluwer Academic Publishers.

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

Shapiro, A. (1985), Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints, *Biometrika, 72,* 133-140.

Sorbom, D. (1989). Model modification. *Psychometrika, 54,* 371-384.

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.

Swinton, S. S., & Powers, D. E. (1980). *Factor analysis of the Test of English as a Foreign Language for several language groups* (TOEFL Research Rep. No. RR-06; ETS RR-80-32). Princeton, NJ: ETS.

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.

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. (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., 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.

Table A1. $RMSEA_{12}$s by specification pairs in Example 1

Specification of	Specification of unrestricted model		
restricted model	Specification 1	Specification 2	Specification 3
Specification 2	.0216	-	-
Specification 3	.0312	.0224	-
Specification 4	.0408	.0343	.0258

Table A2. $\tilde{RMSEA}_{12}$s by specification pairs in Example 1

Specification of	Specification of unrestricted model		
restricted model	Specification 1	Specification 2	Specification 3
Specification 2	.1990	-	-
Specification 3	.2034	.2077	-
Specification 4	.2166	.2249	.2409

Table B. $RMSEA_{12}$s by specification pairs in Example 2

Specification of	Specification of the baseline model
the competing model	Specification 1
Specification 2	.0279
Specification 3	.0377

Table 1A. Rejection rate of different statistics with $\alpha = .05$
for Specification 1 vs. 2 in Example 1, normal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.051	0.061	0.051	0.061
T_1	0.011	0.021	0.022	0.026
T_2	0.012	0.021	0.022	0.028
T_3	0.040(203)	0.045(23)	0.037(1)	0.045(0)
T_4	0.041(203)	0.045(23)	0.037(1)	0.045(0)
T_5	0.040(201)	0.044(22)	0.037(1)	0.045(0)
T_6	0.038(204)	0.045(22)	0.037(1)	0.045(0)

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

Sample Size	150	300	500	1000
T_{ML}	0.072	0.074	0.077	0.073
T_1	0.011	0.016	0.017	0.021
T_2	0.009	0.013	0.013	0.021
T_3	0.042(220)	0.042(29)	0.037(5)	0.037(0)
T_4	0.038(220)	0.037(29)	0.033(5)	0.035(0)
T_5	0.040(202)	0.039(29)	0.034(3)	0.035(0)
T_6	0.042(223)	0.041(31)	0.036(3)	0.036(0)

Table 1C. Rejection rate of different statistics with $\alpha = .05$
for Specification 1 vs. 2 in Example 1, severe nonnormal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.125	0.112	0.124	0.115
T_1	0.013	0.013	0.014	0.019
T_2	0.008	0.008	0.011	0.013
T_3	0.054(283)	0.036(77)	0.044(8)	0.043(0)
T_4	0.038(283)	0.028(77)	0.036(8)	0.037(0)
T_5	0.044(242)	0.033(59)	0.039(2)	0.038(0)
T_6	0.051(299)	0.036(76)	0.041(6)	0.039(0)

Table 2A. Rejection rate of different statistics with $\alpha = .05$
for Specification 2 vs. 3 in Example 1, normal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.060	0.053	0.054	0.059
T_1	0.013	0.015	0.021	0.029
T_2	0.013	0.016	0.022	0.029
T_3	0.045(135)	0.038(21)	0.040(1)	0.045(0)
T_4	0.047(135)	0.037(21)	0.040(1)	0.045(0)
T_5	0.046(134)	0.038(22)	0.040(1)	0.045(0)
T_6	0.043(134)	0.038(22)	0.040(1)	0.045(0)

Table 2B. Rejection rate of different statistics with $\alpha = .05$
for Specification 2 vs. 3 in Example 1, mild nonnormal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.082	0.085	0.083	0.076
T_1	0.008	0.018	0.026	0.029
T_2	0.004	0.017	0.025	0.025
T_3	0.050(170)	0.047(32)	0.052(4)	0.048(0)
T_4	0.042(170)	0.043(32)	0.048(4)	0.045(0)
T_5	0.047(165)	0.044(30)	0.051(4)	0.046(0)
T_6	0.050(176)	0.047(31)	0.052(4)	0.048(0)

Table 2C. Rejection rate of different statistics with $\alpha = .05$
for Specification 2 vs. 3 in Example 1, severe nonnormal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.118	0.113	0.11	0.118
T_1	0.012	0.012	0.015	0.026
T_2	0.006	0.006	0.009	0.020
T_3	0.053(271)	0.041(64)	0.037(13)	0.05(0)
T_4	0.037(271)	0.030(64)	0.032(13)	0.043(0)
T_5	0.045(231)	0.034(47)	0.035(6)	0.046(0)
T_6	0.053(298)	0.040(58)	0.036(10)	0.048(0)

Table 3A. Rejection rate of different statistics with $\alpha = .05$
for Specification 3 vs. 4 in Example 1, normal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.053	0.056	0.054	0.046
T_1	0.016	0.024	0.026	0.026
T_2	0.017	0.025	0.026	0.026
T_3	0.039(67)	0.043(2)	0.042(0)	0.036(0)
T_4	0.039(67)	0.044(2)	0.041(0)	0.037(0)
T_5	0.039(64)	0.043(2)	0.041(0)	0.036(0)
T_6	0.039(66)	0.043(2)	0.041(0)	0.036(0)

Table 3B. Rejection rate of different statistics with $\alpha = .05$
for Specification 3 vs. 4 in Example 1, mild nonnormal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.07	0.079	0.073	0.075
T_1	0.018	0.028	0.027	0.030
T_2	0.015	0.021	0.021	0.028
T_3	0.047(83)	0.056(3)	0.045(0)	0.048(0)
T_4	0.040(83)	0.049(3)	0.043(0)	0.041(0)
T_5	0.044(83)	0.051(3)	0.043(0)	0.041(0)
T_6	0.046(88)	0.055(3)	0.045(0)	0.043(0)

Table 3C. Rejection rate of different statistics with $\alpha = .05$
for Specification 3 vs. 4 in Example 1, severe nonnormal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.119	0.122	0.111	0.117
T_1	0.020	0.023	0.022	0.026
T_2	0.013	0.013	0.013	0.021
T_3	0.056(156)	0.056(17)	0.048(0)	0.048(0)
T_4	0.039(156)	0.040(17)	0.036(0)	0.039(0)
T_5	0.046(146)	0.050(12)	0.040(0)	0.042(0)
T_6	0.054(173)	0.054(15)	0.044(0)	0.044(0)

Table 4A. Rejection rate of different statistics with $\alpha = .05$
for Specification 1 vs. 2 in Example 2, normal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.062	0.069	0.067	0.073
T_1	0.021	0.024	0.025	0.039
T_2	0.021	0.024	0.026	0.037
T_3	0.042(2)	0.044(0)	0.043(0)	0.054(0)
T_4	0.039(2)	0.043(0)	0.041(0)	0.053(0)
T_5	0.039(0)	0.044(0)	0.042(0)	0.053(0)
T_6	0.041(0)	0.043(0)	0.042(0)	0.054(0)

Table 4B. Rejection rate of different statistics with $\alpha = .05$
for Specification 1 vs. 2 in Example 2, mild nonnormal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.099	0.09	0.1	0.098
T_1	0.025	0.022	0.036	0.036
T_2	0.015	0.015	0.029	0.034
T_3	0.052(3)	0.042(0)	0.054(0)	0.048(0)
T_4	0.041(3)	0.032(0)	0.046(0)	0.044(0)
T_5	0.047(2)	0.035(0)	0.047(0)	0.045(0)
T_6	0.05(2)	0.037(0)	0.049(0)	0.045(0)

Table 4C. Rejection rate of different statistics with $\alpha = .05$
for Specification 1 vs. 2 in Example 2, severe nonnormal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.146	0.162	0.142	0.149
T_1	0.025	0.027	0.029	0.03
T_2	0.01	0.013	0.017	0.022
T_3	0.054(10)	0.053(0)	0.045(0)	0.045(0)
T_4	0.031(10)	0.035(0)	0.032(0)	0.032(0)
T_5	0.036(9)	0.041(0)	0.035(0)	0.034(0)
T_6	0.042(15)	0.044(0)	0.036(0)	0.035(0)

Table 5A. Rejection rate of different statistics with $\alpha = .05$
for Specification 1 vs. 3 in Example 2, normal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.066	0.074	0.069	0.065
T_1	0.029	0.037	0.034	0.04
T_2	0.028	0.037	0.034	0.04
T_3	0.042(0)	0.05(0)	0.045(0)	0.048(0)
T_4	0.04(0)	0.049(0)	0.045(0)	0.048(0)
T_5	0.04(0)	0.049(0)	0.045(0)	0.048(0)
T_6	0.041(0)	0.049(0)	0.045(0)	0.048(0)

Table 5B. Rejection rate of different statistics with $\alpha = .05$
for Specification 1 vs. 3 in Example 2, mild nonnormal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.095	0.09	0.092	0.101
T_1	0.034	0.029	0.033	0.036
T_2	0.025	0.022	0.026	0.029
T_3	0.052(0)	0.046(0)	0.044(0)	0.048(0)
T_4	0.042(0)	0.035(0)	0.038(0)	0.041(0)
T_5	0.045(0)	0.038(0)	0.039(0)	0.041(0)
T_6	0.047(0)	0.04(0)	0.04(0)	0.042(0)

Table 5C. Rejection rate of different statistics with $\alpha = .05$
for Specification 1 vs. 3 in Example 2, severe nonnormal condition, NR=2000

Sample Size	150	300	500	1000
T_{ML}	0.14	0.148	0.133	0.153
T_1	0.028	0.037	0.034	0.035
T_2	0.013	0.021	0.018	0.022
T_3	0.052(0)	0.058(0)	0.047(0)	0.046(0)
T_4	0.026(0)	0.033(0)	0.035(0)	0.034(0)
T_5	0.033(0)	0.04(0)	0.036(0)	0.037(0)
T_6	0.038(0)	0.044(0)	0.036(0)	0.037(0)

Table 6A. Rejection rate of $T_{ML,0.03}$, $T_{1,0.03}$ to $T_{6,0.03}$
for Specification 1 vs. 2 in Example 2, normal condition, NR=2000

Sample Size	150	300	500	1000
$T_{ML,0.03}$	0.029	0.018	0.013	0.011
$T_{1,0.03}$	0.009	0.007	0.005	0.003
$T_{2,0.03}$	0.009	0.007	0.005	0.003
$T_{3,0.03}$	0.018(2)	0.014(0)	0.009(0)	0.003(0)
$T_{4,0.03}$	0.019(2)	0.014(0)	0.009(0)	0.003(0)
$T_{5,0.03}$	0.018(0)	0.014(0)	0.009(0)	0.003(0)
$T_{6,0.03}$	0.018(0)	0.014(0)	0.009(0)	0.003(0)

Table 6B. Rejection rate of $T_{ML,0.03}$, $T_{1,0.03}$ to $T_{6,0.03}$
for Specification 1 vs. 2 in Example 2, mild nonnormal condition, NR=2000

Sample Size	150	300	500	1000
$T_{ML,0.03}$	0.059	0.041	0.038	0.025
$T_{1,0.03}$	0.011	0.009	0.01	0.004
$T_{2,0.03}$	0.007	0.005	0.008	0.004
$T_{3,0.03}$	0.027(3)	0.015(0)	0.015(0)	0.009(0)
$T_{4,0.03}$	0.016(3)	0.009(0)	0.013(0)	0.005(0)
$T_{5,0.03}$	0.02(2)	0.01(0)	0.015(0)	0.006(0)
$T_{6,0.03}$	0.023(2)	0.011(0)	0.015(0)	0.006(0)

Table 6C. Rejection rate of $T_{ML,0.03}$, $T_{1,0.03}$ to $T_{6,0.03}$
for Specification 1 vs. 2 in Example 2, severe nonnormal condition, NR=2000

Sample Size	150	300	500	1000
$T_{ML,0.03}$	0.095	0.096	0.058	0.045
$T_{1,0.03}$	0.014	0.015	0.007	0.006
$T_{2,0.03}$	0.005	0.005	0.004	0.004
$T_{3,0.03}$	0.031(10)	0.024(0)	0.013(0)	0.011(0)
$T_{4,0.03}$	0.017(10)	0.014(0)	0.007(0)	0.007(0)
$T_{5,0.03}$	0.02(9)	0.017(0)	0.007(0)	0.007(0)
$T_{6,0.03}$	0.025(15)	0.018(0)	0.009(0)	0.007(0)

Table 7A. Rejection rate of $T_{ML,0.03}$, $T_{1,0.03}$ to $T_{6,0.03}$
for Specification 1 vs. 3 in Example 2, normal condition, NR=2000

Sample Size	150	300	500	1000
$T_{ML,0.03}$	0.382	0.58	0.77	0.95
$T_{1,0.03}$	0.21	0.429	0.633	0.904
$T_{2,0.03}$	0.208	0.424	0.629	0.903
$T_{3,0.03}$	0.294(0)	0.5(0)	0.697(0)	0.926(0)
$T_{4,0.03}$	0.291(0)	0.495(0)	0.69(0)	0.924(0)
$T_{5,0.03}$	0.292(0)	0.496(0)	0.69(0)	0.924(0)
$T_{6,0.03}$	0.293(0)	0.497(0)	0.692(0)	0.924(0)

Table 7B. Rejection rate of $T_{ML,0.03}$, $T_{1,0.03}$ to $T_{6,0.03}$
for Specification 1 vs. 3 in Example 2, mild nonnormal condition, NR=2000

Sample Size	150	300	500	1000
$T_{ML,0.03}$	0.422	0.577	0.735	0.927
$T_{1,0.03}$	0.194	0.351	0.559	0.841
$T_{2,0.03}$	0.159	0.319	0.534	0.837
$T_{3,0.03}$	0.29(0)	0.44(0)	0.623(0)	0.867(0)
$T_{4,0.03}$	0.258(0)	0.412(0)	0.601(0)	0.861(0)
$T_{5,0.03}$	0.269(0)	0.421(0)	0.604(0)	0.863(0)
$T_{6,0.03}$	0.28(0)	0.428(0)	0.611(0)	0.863(0)

Table 7C. Rejection rate of $T_{ML,0.03}$, $T_{1,0.03}$ to $T_{6,0.03}$
for Specification 1 vs. 3 in Example 2, severe nonnormal condition, NR=2000

Sample Size	150	300	500	1000
$T_{ML,0.03}$	0.425	0.565	0.702	0.885
$T_{1,0.03}$	0.157	0.275	0.413	0.691
$T_{2,0.03}$	0.091	0.202	0.353	0.661
$T_{3,0.03}$	0.249(0)	0.359(0)	0.505(0)	0.739(0)
$T_{4,0.03}$	0.182(0)	0.305(0)	0.451(0)	0.712(0)
$T_{5,0.03}$	0.207(0)	0.32(0)	0.469(0)	0.718(0)
$T_{6,0.03}$	0.223(0)	0.327(0)	0.477(0)	0.721(0)