UCLA
Department of Statistics Papers

Title
A Scaled Difference Chi-square Test Statistic for Moment Structure Analysis

Permalink
https://escholarship.org/uc/item/23c604tb

Authors
Albert Satorra
Peter Bentler

Publication Date
2011-10-25

A Scaled Difference Chi-square Test Statistic for Moment Structure Analysis *

Albert Satorra
Universitat Pompeu Fabra
and
Peter M. Bentler
University of California, Los Angeles

August 3, 1999

Abstract

A family of scaling corrections aimed to improve the chi-square approximation of goodness-of-fit test statistics in small samples, large models, and nonnormal data was proposed in Satorra and Bentler (1994). For structural equations models, Satorra-Bentler's (SB) scaling corrections are available in standard computer software. Often, however, the interest is not on the overall fit of a model, but on a test of the restrictions that a null model say $\mathcal{M}_0$ implies on a less restricted one $\mathcal{M}_1$. If T_0 and T_1 denote the goodness-of-fit test statistics associated to $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively, then typically the difference $T_d = T_0 - T_1$ is used as a chi-square test statistic with degrees of freedom equal to the difference on the number of independent parameters estimated under the models $\mathcal{M}_0$ and $\mathcal{M}_1$. As in the case of the goodness-of-fit test, it is of interest to scale the statistic T_d in order to improve its chi-square approximation in realistic, i.e., nonasymptotic and nonnormal, applications. In a recent paper, Satorra (1999) shows that the difference between two Satorra-Bentler scaled test statistics for overall model fit does not yield the correct SB scaled difference test statistic. Satorra developed an expression that permits scaling the difference test statistic, but his formula has some practical limitations, since it requires heavy computations that are not available in standard computer software. The purpose of the present paper is to provide an easy way to compute the scaled difference chi-square statistic from the scaled goodness-of-fit test statistics of models $\mathcal{M}_0$ and $\mathcal{M}_1$. A Monte Carlo study is provided to illustrate the performance of the competing statistics.

Keywords: Moment-structures, goodness-of-fit test, chi-square difference test statistic, chi-square distribution, non-normality

1   Introduction

Moment structure analysis is widely used in behavioural, social and economic studies to analyse structural relations between variables, some of which may be latent (i.e., unobservable); see, e.g., Bollen (1989), Bentler and Dudgeon (1996), Yuan and Bentler (1997), and references therein. Commercial computer programs to carry out such analysis, for a general class of structural equation models, are available (e.g., LISREL of Jöreskog and Sörbom, 1994; EQS of Bentler, 1995). In multi-sample analysis, data from several samples are combined into one analysis, making it possible, among other features, to test for across-group invariance of specific model parameters. Statistics that are central in moment structure analysis are the overall goodness-of-fit test of the model and tests of restrictions on parameters.

Asymptotic distribution-free (ADF) methods which do not require distributional assumptions on the observable variables have been developed (Browne, 1984). The ADF methods, however, involve fourth-order sample moments, thus they may lack robustness to small and medium-sized samples. In the case of non-normal data, an alternative to the ADF approach is to use a normal-theory estimation method in conjunction with asymptotic robust standard errors and test statistics (see Satorra, 1992). Asymptotic robust test statistics, however, may still lack robustness to small and medium-sized samples. As an alternative to asymptotically robust test statistics, Satorra and Bentler (1994; Satorra and Bentler, 1988a,b) developed a family of corrected normal-theory test statistics which are easy to implement in practice, and which have been shown to outperform the asymptotic robust test statistics in small and medium-sized samples (e.g., Chou, Bentler and Satorra, 1991; Hu, Bentler and Kano, 1992; Curran, West and Finch, 1996). Bentler and Yuan (1999) provide a recent comparison of alternative test methods for small samples. Extension of Satorra-Bentler (SB)'s corrections to goodness-of-fit test statistics in the case of the analysis of augmented moment structures, multi-samples and categorical data, have been discussed respectively by Satorra (1992) and Muthén (1993).

Although SB corrections have been available for some time, formal derivations of SB corrections to the case of nested model comparisons have not been available. The obvious and generally accepted approach of computing separate SB-corrected test statistics for each of two nested models, and then computing the difference between them (e.g., Byrne and Campbell, 1999),

turns out to be an incorrect way to obtain a scaled SB difference test statistic. The difference could be even be negative, which is an improper value for a chi-square variate. In a recent paper, Satorra (1999) gives specific formulae for extension of SB corrections to score (Lagrange multiplier), difference and Wald test statistics. He showed that the difference between two SB-scaled test statistics does not necessarily correspond to the scaled chi-square difference test statistic. The purpose of the present paper is to provide a simple expression that allows a researcher to correctly compute the SB difference test statistic when the SB-scaled chi-square goodness of fit tests for the corresponding two nested models are available. The formula is simple to use and provides an alternative scaled test for evaluating a specific set of restrictions.

The paper is structured as follows. In Section 2 we describe goodness-of-fit tests in weighted least squares analysis, and the corresponding SB scaling corrections. In Section 3 we describe the proposed procedure for computing the SB scaled difference test statistic. Section 4 concludes with an illustration.

2   Goodness-of-fit tests

Let σ and s be p-dimensional vectors of population and sample moments respectively, where s tends in probability to σ as sample size $n \to +\infty$. Let $\sqrt{n}s$ be asymptotically normally distributed with a finite asymptotic variance matrix Γ $(p \times p)$. Consider the model $\mathcal{M}_0 : \sigma = \sigma(\theta)$ for the moment vector θ, where $\sigma(.)$ is a twice-continuously differentiable vector-valued function of θ, a q-dimensional parameter vector. Consider a WLS estimator $\hat{\theta}$ of θ defined as the minimizer of

$$F_V(\theta) := (s - \sigma)'\hat{V}(s - \sigma)$$

over the parameter space, where $\hat{V}$ $(p \times p$), converges in probability to V, a positive definite matrix. A typical test statistic used for testing the goodness-of fit-of the model $\mathcal{M}_0$ is $T_0 := nF_V(s, \hat{\sigma})$, where $\hat{\sigma} := \sigma(\hat{\theta})$. It is widely known that, when the model $\mathcal{M}_0$ holds and V satisfies the asymptotic optimality (AO) condition of $V = \Gamma^{-1}$, then T_0 is asymptotically chi-square distributed with degrees of freedom (df) $r_0 = p - q$. In practice, however, AO may not hold, and concern on the quality of the chi-square approximation do arise. For general types of distributions, i.e., when AO does not necessarily hold, T_0 is asymptotically distributed as a mixture of chi-square distributions

3

of 1 degree of freedom (df) (see Satorra and Bentler, 1986); that is

$$T_0 \xrightarrow{L} \sum_{j=1}^{r} \alpha_j \chi_j^2, \tag{1}$$

as $n \rightarrow \infty$, where the χ_j^2 are independent chi-square variables of 1 df, and the α_j are the non-null eigenvalues of the matrix $U_0 \Gamma$, with

$$U_0 := V - V\Delta(\Delta'V\Delta)^{-1}\Delta'V$$

and $\Delta := (\partial/\partial\theta')\sigma(\theta)$. When AO holds, then of course the α_j's are equal to 1 and the asymptotic exact chi-square distribution applies. In the context of structural models and for general types of distributions, Satorra and Bentler (1994; Satorra and Bentler, 1988a,b) proposed replacing T by the scaled statistic

$$\overline{T} = T/\hat{c}, \tag{2}$$

where $\hat{c}$ denotes a consistent estimator of

$$c := \frac{1}{r}\text{tr}\,U_0\Gamma = \frac{1}{r}\text{tr}\,\{V\Gamma\} - \frac{1}{r}\text{tr}\,\left\{(\Delta'V\Delta)^{-1}\Delta'V\Gamma V\Delta\right\}. \tag{3}$$

Note that the SB scaled test statistic has the same mean as the corresponding χ_r^2 variate. The SB scaled goodness-of-fit test has been shown to outperform alternative test statistics in a variety of models and non-normal distributions (e.g., Chou, Bentler and Satorra, 1991; Hu, Bentler and Kano, 1992; Curran, West and Finch, 1996). Of course, when asymptotic optimality holds, this statistic will have the same asymptotic distribution as the unscaled statistic T_0. Note that a consistent estimator $\hat{\Gamma}$ of Γ under general distribution conditions is required to compute the scaling factor $\hat{c}$. In structural equation models, a consistent estimator of Γ is readily available from the raw data (e.g., Satorra and Bentler, 1994). A goodness-of-fit statistic which can be used given any estimation method, is given by

$$T^{\star} = n(s - \hat{\sigma})'\{\hat{\Gamma}^{-1} - \hat{\Gamma}^{-1}\hat{\Delta}(\hat{\Delta}'\hat{\Gamma}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\hat{\Gamma}^{-1}\}(s - \hat{\sigma}), \tag{4}$$

When $\hat{\Gamma}$ is the (distribution-free) consistent estimator of Γ in (16) below, then $T^{\star}$ will be called the asymptotic robust goodness-of-fit test statistic, since it is an asymptotic chi-square statistic regardless of the distribution of observable variables. In the context of single-sample covariance structure

4

analysis, this statistic was first introduced by Browne (1984). Its performance was studied by Yuan and Bentler (1998), who found that very large samples are required to obtain acceptable performance in models with intermediate to large degrees of freedom.

3   Testing a set of restrictions

Consider now the case of testing a specific set of restrictions on the model. Consider a re-parameterezation of $\mathcal{M}_0$ as $\sigma = \sigma^\star(\delta)$ with $a(\delta) = a_0$, where δ is a $(q + m)$-dimensional vector of parameters, a_0 is an $m \times 1$ vector of constants, and $\sigma^\star(.)$ and $a(.)$ are twice-continuously differentiable vector-valued functions of $\delta \in \Theta_1$, a compact subset of R^{q+m}. Our interest now is in the test of the null hypothesis $H_0 : a(\delta) = a_0$ against the alternative $H_1 : a(\delta) \neq a_0$. Define the Jacobian matrices

$$\Pi\left(p \times (q + m)\right) := (\partial/\partial \delta')\sigma^\star(\delta) \quad \text{and} \quad A\left(m \times (q + m)\right) := (\partial/\partial \delta')a(\delta),$$

which we assume to be regular at the value of δ associated with θ_0, say δ_0, with A of full row rank. Let $P\left((q + m) \times (q + m)\right) := \Pi'V\Pi$ and denote by $\mathcal{M}_1$ the less restricted model $\sigma = \sigma^\star(\delta)$. The goodness-of-fit test statistic associated with $\mathcal{M}_1$ is thus $T_1 = nF(s, \tilde{\sigma})$, where $\tilde{\sigma}$ is the fitted moment vector in model $\mathcal{M}_1$, now with associated degrees of freedom $r_1 = r_0 - m$ and scaling factor c_1 given by

$$c_1 := \frac{1}{r_1}\operatorname{tr} U_1\Gamma = \frac{1}{r_1}\operatorname{tr}\left\{V\Gamma\right\} - \frac{1}{r_1}\operatorname{tr}\left\{P^{-1}\Pi'V\Gamma V\Pi\right\} \tag{5}$$

where

$$U_1 := V - V\Pi P^{-1}\Pi'V.$$

When both models $\mathcal{M}_0$ and $\mathcal{M}_1$ are fitted, then we can test the restrictions $a(\theta) = a_0$ using the difference test statistic $T_d = T_0 - T_1$, where under the null hypothesis, it is intended that T_d have a chi-square distributed with degrees of freedom $m = r_0 - r_1$.

In order to improve the chi-square approximation in the case of large values of m and moderate or small sample sizes, we are interested in the SB scaled difference test statistic, say $\bar{T}_d$. Extending his earlier work (Satorra, 1989), Satorra (1999) recently provided formulae for computing such scaled

statistics for the difference, Score and Wald test statistics. From Satorra's formulae it becomes aparent that the SB scaled difference test statistic does not coincide with the difference between the two SB scaled goodness-of-fit test statistics that arise when fitting the two nested models; that is, in general $\bar{T}_d \neq \bar{T}_0 - \bar{T}_1$, where by $\bar{T}_0$ and $\bar{T}_1$ we denote the SB scaled goodness-of-fit test statistics arising when fitting the models $\mathcal{M}_0$ and $\mathcal{M}_1$ respectively. In Satorra (1999), the SB scaled difference test statistic is defined as $\bar{T}_d = T_d/\hat{c}_d$ where $\hat{c}_d$ is a consistent estimator of

$$c_d := \frac{1}{m}\operatorname{tr} U_d \Gamma \tag{6}$$

with

$$U_d = V\Pi P^{-1} A'(AP^{-1}A')^{-1}AP^{-1}\Pi'V. \tag{7}$$

A practical problem with this expression for the scaled difference test statistic is it requires computations that are outside the standard output of current structural equation modeling programs. Furthermore, difference tests are usually hand computed from different modeling runs. Here we will show how to combine the scaling corrections c_0 and c_1 associated to the two fitted models $\mathcal{M}_0$ and $\mathcal{M}_1$ in order to compute the scaling correction c_d for the difference test statistic. It turns out that the computations are extremely simply and can be carried out using a hand calculator.

First we show that $U_d = U_0 - U_1$. Note that the model $\mathcal{M}_0$ implies a specific function $\delta = \delta(\theta)$, that by the implicit function theorem is continuous differentiable. Consider thus the matrix $H = \partial\delta(\theta)/\partial\theta'$. Clearly, it holds that $\Delta = \Pi H$ and $AH = 0$ (recall that A is a matrix $m \times (q + m)$), that is, the matrix A and H are orthogonal complements. We have

$$U_0 - U_1 = V\Pi(\Pi'V\Pi)^{-1}\Pi'V - V\Pi H(H'\Pi'V\Pi H)^{-1}H'\Pi'V$$

$$= V\Pi\left\{P^{-1} - H(H'PH)^{-1}H'\right\}\Pi'V$$

since

$$P^{-1} - H(H'PH)^{-1}H' = P^{-1}A'(AP^{-1}A')^{-1}AP^{-1},$$

as A and H are orthogonal complements (see Rao, 1973, p. 77). We thus have the basic result that

$$U_d = U_0 - U_1.$$

Now, since $r_0 c_0 - r_1 c_1 = \text{tr}\,(U_0 - U_1)\Gamma = \text{tr}\,U_d\Gamma = m c_d$, we obtain $c_d = (r_0 c_0 - r_1 c_1)/m$. This means that consistent estimation of c_d is available from consistent estimates of the scaling corrections c_1 and c_1 associated with the null and alternative model respectively.

Thus the proposed practical procedure is as follows. When fitting models $\mathcal{M}_0$ and $\mathcal{M}_1$, we obtain the unscaled and scaled goodness-of-fit tests, that is T_0 and $\bar{T}_0$ when fitting $\mathcal{M}_0$, and T_1 and $\bar{T}_1$ when fitting $\mathcal{M}_1$. Let r_0 and r_1 be the associated degrees of freedom of the goodness-of-fit test statistics. Then we compute the scaling corrections $\hat{c}_0 = T_0/\bar{T}_0$ and $\hat{c}_1 = T_1/\bar{T}_1$, and the usual chi-square difference $T_d = T_0 - T_1$. The SB scaled difference test can thus be computed as $\bar{T}_d = T/\hat{c}_d$, where

$$\hat{c}_d = (r_0 \hat{c}_0 - r_1 \hat{c}_1)/m.$$

When the two scaling corrections are equal, i.e. when $c_0 = c_1 = c$ then $c_d = c$ and thus $\bar{T}_d = \bar{T}_0 - \bar{T}_1$. This is the case, for example, when $c_0 = c_1 = 1$, i.e., when both goodness-of-fit tests are asymptotically chi-square statistics. In general, however, $c_0 \neq c_1$ and then the difference between two SB scaled goodness of fit test statistics does not yield the SB scaled difference test statistic.

Note that the above procedure applies to a general modeling setting. The vector of parameters σ to be modeled may contain various types of moments: means, product-moments, frequencies (proportions), and so forth. Thus, the procedure applies to a variety of techniques, such as factor analysis, simultaneous equations for continuous variables, log-linear multinomial parametric models, etc.. It can easily be seen that the procedure applies also in the case where the matrix Γ is singular, and when the data is composed of various samples, as in multi-sample analysis. The results apply also to other estimation methods, e.g., pseudo ML estimation.

It is important to recognize that a competitor to the statistic $\bar{T}_d$ will be the difference between the robust goodness-of-fit test statistics associated with the models $\mathcal{M}_0$ and $\mathcal{M}_1$; that is, an asymptotic chi-square test statistic for H_0 is just $T_d^\star := T_0^\star - T_1^\star$, where $T_0^\star$ and $T_1^\star$ are the goodness-of-fit test statistics associated to the models $\mathcal{M}_0$ and $\mathcal{M}_1$ respectively. In the next section, we will illustrate using Monte Carlo simulation the small sample size performance of the competing test statistics for the above mentioned null hypothesis H_0.

4 Illustration

In this section we ilustrate in a simple model context of a regression with errors in variables the performance in finite samples of three test statistics. We consider a regression equation

$$y_{gi}^\star = \beta x_{gi} + v_{gi}, \quad i = 1, \ldots, n_g,\tag{8}$$

where for case i in group g $(g = 1, 2)$, $y_{gi}^\star$ and x_{gi} are the values of the response and explanatory variables, respectively, v_{gi} is the value of the disturbance term, and β is the regression coefficient. The model assumes that x_{gi} is unobservable, but there are two observable variables $x_{1gi}^\star$ and $x_{2gi}^\star$ related to x_{gi} by the following measurement-error equations

$$x_{1gi}^\star = x_{gi} + u_{1gi}, \quad x_{2gi}^\star = x_{gi} + u_{2gi},\tag{9}$$

where u_{1gi} and u_{2gi} are mutually independent and also independent of v_{gi} and x_{gi}. It is assumed that the observations are independent and identically distributed within each group. Equations (8) and (9) with the associated assumptions yield an identified model (see Fuller (1987) for a comprehensive overview of measurement-error models in regression analysis). Inference is usually carried out in this type of model under the assumption that the observable variables are normally distributed. Write the model of (8) and (9) as

$$z_{gi} = \Lambda \xi_{gi}, \quad i = 1, 2, \ldots, n,\tag{10}$$

where

$$z_{gi} := \begin{pmatrix} y_{gi}^\star \\ x_{1gi}^\star \\ x_{2gi}^\star \end{pmatrix}, \quad \xi_{gi} := \begin{pmatrix} x_{gi} \\ v_{gi} \\ u_{1gi} \\ u_{2gi} \end{pmatrix}$$

and

$$\Lambda := \begin{pmatrix} \beta & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.\tag{11}$$

Define

$$\Phi := E\xi_{gi}\xi_{gi}' = \begin{pmatrix} \sigma_{xx} & 0 & 0 & 0 \\ 0 & \sigma_{vv} & 0 & 0 \\ 0 & 0 & \sigma_{uu} & 0 \\ 0 & 0 & 0 & \sigma_{uu} \end{pmatrix},\tag{12}$$

and the parameter vector $\theta := (\sigma_{vv}, \sigma_{xx}, \sigma_{uu}, \beta)'$. Under this set-up, we obtain the moment structure

$$\Sigma := \Lambda \Phi \Lambda' = \Sigma(\theta), \tag{13}$$

where $\Sigma(.)$, $\Lambda(.)$ and $\Phi(.)$ are (twice-continuously differentiable) matrix-valued functions of θ, as deduced from (11), (12) and (13). Note that the model restricts the variances of u_1 and u_2 by equality. This is a setting of two-sample data, where the population and sample vectors σ and s are defined as $s = (\sigma_1', \sigma_2')'$ and $s = (s_1', s_2')'$, where $\sigma_g' = \text{vec} \, S_g$ and $s_g' = \text{vec} \, S_g$, with

$$S_g := \frac{1}{n_g} \sum_{i=1}^{n_g} z_{gi} z_{gi}'.$$

Here "vec" denotes the column-wise vectorisation operator (see Magnus and Neudecker, 1999, for full details on this operator). We consider the estimation of the model using weighted least squares under the assumption of normality. That is, the matrix V (see above) has the form

$$\hat{V} := \text{block diag}(\frac{n_1}{n} \hat{V}_1, \frac{n_2}{n} \hat{V}_2) \tag{14}$$

and $\hat{V}_g = \frac{1}{2}(S_g^{-1} \otimes S_g^{-1})$, $g = 1, 2$. Clearly, when there is independence across samples, the asymptotic variance matrix of $\sqrt{n}s$ is of the form

$$\Gamma = \text{block diag}(\frac{n}{n_1} \Gamma_1, \frac{n}{n_2} \Gamma_2), \tag{15}$$

where Γ_g is the asymptotic variance of $\sqrt{n_g} s_g$, $g = 1, 2$. We further assume that the matrices S_g and Σ_g are positive definite, and that $n_g/n \to f_g > 0$, as $n \to +\infty$ $(g = 1, 2)$; in this case, a distribution-free consistent estimator of Γ is

$$\hat{\Gamma} := \text{block diag}(\frac{n}{n_1} \hat{\Gamma}_1, \frac{n}{n_2} \hat{\Gamma}_2) \tag{16}$$

where

$$\hat{\Gamma}_g := \frac{1}{n_g - 1} \sum_{i=1}^{n_g} (d_{gi} - s_g)(d_{gi} - s_g)', \tag{17}$$

with $d_{gi} := \text{vec} \, z_{gi} z_{gi}'$.

The Monte Carlo study generates two-sample data from the above model. Two models are fitted. Model $\mathcal{M}_0$ has the parameters restricted across groups, and model $\mathcal{M}_1$ has parameters that are unrestricted across groups.

9

For each of the estimated models, we compute the goodness-of-fit test statistics T_0 and T_1, the SB scaled statistics $\bar{T}_0$ and $\bar{T}_1$, and the robust test statistics $T_0^\star$ and $T_1^\star$. To test the hypothesis of parameter invariance across-samples, we consider the competing statistics $T_d = T_0 - T_1$, $T_d^\star = T_0^\star - T_1^\star$, $d\bar{T} = \bar{T}_0 - \bar{T}_1$ and finally, the proposed statistic $\bar{T}_d$. Note that only $T_d^\star$ is asymptotically an exact chi-square statistic. Our conjecture is that for non-normal data, small samples and/or models with large degrees of freedom, the statistic $\bar{T}_d$ will perform the best.

We obtained replications of the above statistics for various combinations of sample sizes, ranging from a small sample size to an intermediately large sample size. Results are reported in Table 1 and also in Figure 1, where the empirical p-values of the various statistics are fitted against the theoretical ones corresponding to a uniform distribution. In all the replications we used $\theta_0 = (1, 1, .3, 2)'$. The distributions of v and x were independent conveniently scaled zero mean and unit variance chi-squared of 1 df (i.e., a highly non-normal distribution); the distribution of u_1 and u_2 were set to be normal, mutually independent, and independent of v and x. The normal-theory GLS estimation method described in Section 2 was used. The restricted model imposed across-group invariance of model parameters. In each replication, we computed the statistics mentioned above, corresponding to the null hypothesis of invariance of cross-sample model parameters. Clearly, in our Monte Carlo set-up, the null hypothesis holds true, with the null distribution of the statistics being chi-square with $m = 4$ degrees of freedom. Note that in our Monte Carlo set-up, severe non-normality of random constituents of the model requires the use of robust and/or corrected versions of the difference test statistic. We note that the normal-theory chi-square goodness-of-fit T_1 of the unrestricted model (i.e., the model that does not restrict parameters across groups) is an asymptotic chi-square statistic despite non-normality of the data (this follows from the asymptotic robustness theory for multi-samples; cf., Satorra, 1992). In contrast, the normal-theory chi-square goodness of fit T_0 of the restricted model (i.e., the model that imposes parameter invariance across-samples) is not necessarily an asymptotic chi-square statistic (since variances of non-normal constituents of the model are restricted by equality across-groups; cf., Satorra, 1992).

As shown in Table 1, in our specific model context, in the smaller sample, the SB scaled statistic, $\bar{T}_d$, seems to outperform the alternative robust test statistic $T_d^\star$. As expected from theory, in the case of the large sample, $T_d^\star$

Table 1: Monte Carlo results:
empirical significance levels of test statistics

nominal significance levels:	1%	5%	10%	20%
$n_1 = 100$ and $n_2 = 120$				
$\bar{T}_d$	2.7	8.5	13.7	23.1
$T_d^\star$	3.6	10.6	19.3	32.0
T_d	24.8	39.5	50.1	62.1
$d\bar{T}$	76.8	77.5	78.2	78.9
$n_1 = 800$ and $n_2 = 900$				
$\bar{T}_d$	1.2	7.0	11.4	19.9
$T_d^\star$	.5	4.3	11.0	21.7
T_d	29.8	43.1	52.5	62.8
$d\bar{T}$	67.5	68.4	69.1	70

outperforms the alternative test statistics. Especially interesting is that the statistic $d\bar{T} = \bar{T}_0 - \bar{T}_1$ performs very badly indeed. That is, doing the presumably natural thing, simply computing the difference between two SB scaled chi-square statistics, yields a very poorly performing test when evaluated by the chi-square distribution.

References

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

[2] Bentler, P. M., and Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology*, 47, 541–570.

[3] Bentler, P. M., and Yuan, K. -H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, 34, 183–199.

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

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

[6] Byrne, B. M. and Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-cultural Psychology, 30*, 557–576.

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

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

[9] Fuller, W. A. (1987). *Measurement error models*. New York: Wiley.

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

[11] Jöreskog, K., and Sörbom, D. (1994). *LISREL 8 user's reference guide*. Mooresville, IN: Scientific Software.

[12] Magnus, J., and Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.

[13] Muthén, B. (1993). Goodness of fit test with categorical and other non-normal variables. In K. A. Bollen and J. S. Long (Eds.), *Testing structural equation models* (pp. 205–234). Newbury Park: Sage.

[14] Rao, C.R., (1973). *Linear Statistical Inference and Its Applications*, 2nd. edit. New York: Wiley.

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

[16] Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology,* 22, 249–278.

[17] Satorra, A. (1999). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. To appear in *Innovations in Multivariate Statistical Multivariate Analysis: A Festschrift for Heinz Neudecker* (R.D.H. Heijmans, D.S.G. Pollock, and A. Satorra, edts.), Dordrecht: Kluwer Academic Publishers

[18] Satorra, A., and Bentler, P. M. (1986). Some robustness properties of goodness of fit statistics in covariance structure analysis. *1986 ASA Proceedings of the Business and Economic Statistics Section,* 549–554, Alexandria, VA: American Statistical Association.

[19] Satorra, A., and Bentler, P. M. (1988a). Scaling corrections for chi-square statistics in covariance structure analysis. *ASA 1988 Proceedings of the Business and Economic Statistics Section,* 308–313, Alexandria, VA: American Statistical Association.

[20] Satorra, A., and Bentler, P. M. (1988b). Scaling corrections for statistics in covariance structure analysis. *UCLA Statistics Series # 2,* University of California, Los Angeles.

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

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

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