

## ENSURING POSITIVENESS OF THE SCALED DIFFERENCE CHI-SQUARE TEST STATISTIC

ALBERT SATORRA

UNIVERSITAT POMPEU FABRA

PETER M. BENTLER

UNIVERSITY OF CALIFORNIA

A scaled difference test statistic  $\tilde{T}_d$  that can be computed from standard software of structural equation models (SEM) by hand calculations was proposed in Satorra and Bentler (Psychometrika 66:507–514, 2001). The statistic  $\tilde{T}_d$  is asymptotically equivalent to the scaled difference test statistic  $\bar{T}_d$  introduced in Satorra (Innovations in Multivariate Statistical Analysis: A Festschrift for Heinz Neudecker, pp. 233–247, 2000), which requires more involved computations beyond standard output of SEM software. The test statistic  $\tilde{T}_d$  has been widely used in practice, but in some applications it is negative due to negativity of its associated scaling correction. Using the implicit function theorem, this note develops an improved scaling correction leading to a new scaled difference statistic  $\tilde{T}_d$  that avoids negative chi-square values.

Key words: moment-structures, goodness-of-fit test, chi-square difference test statistic, chi-square distribution, nonnormality.

### 1. Introduction

Moment structure analysis is widely used in behavioral, social, and economic studies to analyze structural relations between variables, some of which may be latent; see, e.g., Bollen and Curran (2006), Grace (2006), Yuan and Bentler (2007), and references therein. In such analyses, it frequently happens that two nested models  $M_0$  and  $M_1$ , are compared using estimation methods that are nonoptimal (asymptotically) given the distribution of the data; e.g., maximum likelihood (ML) estimation is used when the data are not multivariate normal. In those circumstances, the usual chi-square difference test  $T_d = T_0 - T_1$ , based on the separate models' goodness of fit test statistics, is not  $\chi^2$  distributed. A correction to  $T_d$  by a scaling factor was proposed by Satorra (2000) and Satorra and Bentler (2001) to improve the chi-square approximation. The latter is the focus of this paper.

Satorra and Bentler's (2001) correction provided a simple procedure to obtain an approximate scaled chi-square statistic using hand calculations on regular output of structural equation modeling (SEM) analysis; it has the drawback, however, that a positive value for the scaling correction is not assured. The present paper develops a simple procedure by which a researcher can compute the exact scaling correction of difference test statistic based only on output from standard SEM programs.

\*Research supported by grants SEJ2006-13537 and PR2007-0221 from the Spanish Ministry of Science and Technology and by USPHS grants DA00017 and DA01070.

Requests for reprints should be sent to Albert Satorra, Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, Barcelona, Spain. E-mail: [albert.satorra@upf.edu](mailto:albert.satorra@upf.edu)

## 2. Setup and Notation

Throughout we adhere to the notation and results of Satorra and Bentler (2001). Let  $\sigma$  and  $s$  be  $p$ -dimensional vectors of population and sample moments, respectively, where  $s$  tends in probability to  $\sigma$  as sample size  $n \rightarrow +\infty$ . Let  $\sqrt{n}(s - \sigma)$  be asymptotically normally distributed with a finite asymptotic variance matrix  $\Gamma$  ( $p \times p$ ). Consider the model  $M_0 : \sigma = \sigma(\theta)$  for the moment vector  $\sigma$ , where  $\sigma(\cdot)$  is a twice-continuously differentiable vector-valued function of  $\theta$ , a  $q$ -dimensional parameter vector. Consider a discrepancy function  $F = F(s, \sigma)$  in the sense of Browne (1984), and the estimator  $\hat{\theta}$  based on  $F$  or on an (asymptotically equivalent) weighted least squares (WLS) analysis with weight matrix  $V = \frac{1}{2} \partial^2 F(s, \sigma) / \partial \sigma \partial \sigma'$  evaluated at  $\sigma = s$ .

Let  $M_0 : \sigma = \sigma^*(\delta)$ ,  $a(\delta) = 0$ , and  $M_1 : \sigma = \sigma^*(\delta)$  be two nested models for  $\sigma$ . Here,  $\delta$  is a  $(q + m)$ -dimensional vector of parameters, and  $\sigma^*(\cdot)$  and  $a(\cdot)$  (an  $m$ -valued function) are twice-continuously differentiable vector-valued functions of  $\delta \in \Theta_1$ , a compact subset of  $R^{q+m}$ . Our interest is in the test of a null hypothesis  $H_0 : a(\delta) = 0$  against the alternative  $H_1 : a(\delta) \neq 0$ .

For the developments that follow, we require the Jacobian matrices  $\Pi(p \times (q + m)) := \partial \sigma^*(\delta) / \partial \delta'$  and  $A(m \times (q + m)) := \partial a(\delta) / \partial \delta'$ , which we assume to be regular at the true value of  $\delta$ , say  $\delta_0$ . We also assume that  $A$  is of full row rank. By using the implicit function theorem, associated to  $M_0$  (more specifically, to the restrictions  $a(\delta) = 0$ ), there exists (locally in a neighborhood of  $\delta_0$ ) a one-to-one function  $\delta = \delta(\theta)$  defined in an open and compact subset  $S$  of  $R^q$ , and a  $\theta_0$  in the interior of  $S$  such that  $\delta(\theta_0) = \delta_0$  and  $\sigma(\delta(\theta))$  satisfies the model  $M_0$ . Let  $H = \partial \delta(\theta) / \partial \theta'$  ( $(q + m) \times q$ ) be the corresponding Jacobian matrix evaluated at  $\theta_0$ . Hence, by the chain rule of differentiation,  $\Delta = \partial \sigma / \partial \theta' = (\partial \sigma(\delta) / \partial \delta') (\partial \delta(\theta) / \partial \theta') = \Pi H$ . Since  $a(\delta(\theta)) = 0$ , it holds that with  $A$  evaluated at  $\delta_0$ ,  $AH = 0$  with  $r(A) + r(H) = q + m$ , and  $r(\cdot)$  denoting the rank of a matrix. Thus,  $H$  is an orthogonal complement of  $A'$ . Let  $P((q + m) \times (q + m)) := \Pi' V \Pi$ . Associated to  $M_1$ , the less restricted model  $\sigma = \sigma^*(\delta)$ , the goodness-of-fit test statistic is  $T_1 = nF(s, \tilde{\sigma})$ , where  $\tilde{\sigma}$  is the fitted moment vector in model  $M_1$  with associated degrees of freedom  $r_1 = r_0 - m$  and scaling factor  $c_1$  given by

$$c_1 := \frac{1}{r_1} \text{tr}\{U_1 \Gamma\} = \frac{1}{r_1} \text{tr}\{V \Gamma\} - \frac{1}{r_1} \text{tr}\{P^{-1} \Pi' V \Gamma V \Pi\}, \quad (1)$$

where

$$U_1 := V - V \Pi P^{-1} \Pi' V. \quad (2)$$

We refer to Satorra and Bentler (2001) for further details.

## 3. Scaling Correction for the Difference Test

When both models  $M_0$  and  $M_1$  are fitted, for example, by ML, then we can test the restriction  $a(\delta) = 0$ , assuming  $M_1$  holds, using the chi-square difference test statistic  $T_d := T_0 - T_1$ . Under the null hypothesis, we would like  $T_d$  to have a  $\chi^2$  distribution with degrees of freedom  $m = r_0 - r_1$ . This is the restricted test of  $M_0$  within  $M_1$ . For general distributions of the data, the asymptotic chi-square approximation may not hold. To improve on the chi-square approximation, Satorra (2000) gave explicit formulae that extend the scaling corrections proposed by Satorra and Bentler (1994) to the case of difference, Wald, and score types of test statistics. General expressions for those corrections were also put forward in Satorra (1989, p. 146). Specifically, for the test statistic  $T_d$  we are considering, Satorra (2000, p. 241) proposed the following scaled test statistic:

$$\bar{T}_d := T_d / \hat{c}_d, \quad \text{where } c_d := \frac{1}{m} \text{tr}\{U_d \Gamma\} \quad (3)$$

with

$$U_d = V\Pi P^{-1}A'(AP^{-1}A')^{-1}AP^{-1}\Pi'V. \quad (4)$$

Here,  $\hat{c}_d$  denotes  $c_d$  after substituting consistent estimates of  $V$  and  $\Gamma$ , and evaluating the Jacobians  $A$  and  $\Pi$  at the estimate  $\hat{\delta}_0$  when fitting  $M_0$  (or  $M_1$ ). Since  $\text{tr}\{U_d\Gamma\}$  can be expressed as the trace of the product of two positive definite matrices,  $\text{tr}\{U_d\Gamma\} > 0$ , and thus  $c_d > 0$ ; the same for  $\hat{c}_d > 0$ . Consequently,  $\bar{T}_d$  is ensured to be a nonnegative number.

A practical problem with the statistic  $\bar{T}_d$  is that it requires computations that are outside the standard output of current SEM programs. Furthermore, difference tests are usually hand computed from different modeling runs. Satorra and Bentler (2001) proposed a procedure to combine the estimates of the scaling corrections  $c_0$  and  $c_1$  associated to the chi-square goodness-of-fit test for the two fitted models  $M_0$  and  $M_1$  in order to compute a consistent estimate of the scaling correction  $c_d$  for the difference test statistic. A modified (easy to compute) scaled test statistic  $\tilde{T}_d$  with the same asymptotic distribution as  $\bar{T}_d$  was proposed. Both statistics were shown to be asymptotically equivalent under a sequence of local alternatives (so they have the same asymptotic local power). Their procedure to compute  $\tilde{T}_d$  is as follows (see Satorra and Bentler, 2001, p. 511).

1. Obtain the unscaled and scaled goodness-of-fit tests when fitting  $M_0$  and  $M_1$ , respectively; that is,  $T_0$  and  $\bar{T}_0$  when fitting  $M_0$ , and  $T_1$  and  $\bar{T}_1$  when fitting  $M_1$ ;
2. Compute the scaling corrections  $\hat{c}_0 = T_0/\bar{T}_0$ ,  $\hat{c}_1 = T_1/\bar{T}_1$ , and the unscaled chi-square difference  $T_d = T_0 - T_1$  and its degrees of freedom  $m = r_0 - r_1$ ;
3. Compute the scaled difference test statistic as

$$\tilde{T}_d := T_d/\tilde{c}_d \quad \text{with } \tilde{c}_d = (r_0\hat{c}_0 - r_1\hat{c}_1)/m.$$

Here,  $r_0$  and  $r_1$  are the respective degrees of freedom of the models  $M_0$  and  $M_1$ .

The basis for computing the scaling correction for the difference test statistic is the following alternative expression for  $U_d$  of (4) (see Satorra and Bentler, 2001, p. 510)

$$U_d = U_0 - U_1, \quad (5)$$

where  $U_1$  is given in (2) and  $U_0 := V - V\Pi H(H'\Pi'V\Pi H)^{-1}H'\Pi'V$ . Since (5) implies

$$mc_d = \text{tr}\{U_d\Gamma\} = \text{tr}\{(U_0 - U_1)\Gamma\} = r_0c_0 - r_1c_1, \quad (6)$$

it follows that  $c_d = (r_0c_0 - r_1c_1)/m$ . This is the theoretical basis for Satorra and Bentler's (2001) proposal as given in steps 1–3 above.

#### 4. Problem with the Current Scaled Difference Test

For an arbitrary matrix  $V > 0$ , (5) and (6) are exact equalities when the matrices  $U_d$ ,  $U_0$  and  $U_1$  are evaluated at a common value  $\delta$  (as, e.g., the fitted value under model  $M_0$  or under  $M_1$ ); however, they are just asymptotic equalities when  $U_d$ ,  $U_0$  and  $U_1$  are evaluated at different estimates that converge to the same true value under the null hypothesis. Under Satorra and Bentler's (2001) proposal,  $\tilde{c}_d$  evaluates  $U_0$  and  $U_1$  at the estimates  $\hat{\delta}_0$  and  $\hat{\delta}_1$ , respectively. Since  $\hat{\delta}_1$  will in general not satisfy the null model  $M_0$  (i.e., it will not be of the form  $\delta = \delta(\theta)$  for the function implied by the implicit function theorem), when it deviates highly from  $M_0$ , the estimated difference  $\tilde{c}_d = (r_0\hat{c}_0 - r_1\hat{c}_1)/m$  may turn out to be negative. This may happen in small samples, or when  $M_0$  is highly incorrect; a result can be an improper value for  $\tilde{T}_d$ . Satorra and Bentler (2001, p. 511) warned on the possibility that an improper value of  $\tilde{T}_d$  could arise.

In order to be sure to avoid a negative value for  $\tilde{c}_d$ , and hence  $\tilde{T}_d$ , currently one would need to resort to computing  $\tilde{T}_d$  using (3). Unfortunately, this is impractical or impossible for most applied researchers who only have access to standard SEM software.

Fortunately, as we show next, the exact value of  $\tilde{T}_d$  can also be obtained from the standard output of SEM software, using a new hand computation.

### 5. A New Scaled Test Statistic $\tilde{T}_d$

Denote by  $M_{10}$  the fit of model  $M_1$  to a model setup with starting values taken as the final estimates obtained from model  $M_0$ , and with number of iterations set to 0. Consider  $\hat{c}_1^{(10)} := T_1^{(10)}/\tilde{T}_1^{(10)}$ , where  $T_1^{(10)}$  and  $\tilde{T}_1^{(10)}$  are the standard unscaled and scaled test statistic of this additional run. Note that the estimate  $\hat{c}_1^{(10)}$  uses model  $M_1$  but the matrices  $\Pi$  and  $A$  are now evaluated at  $\hat{\delta}_0 := \delta(\hat{\theta})$ , where  $\hat{\theta}$  is the estimate under  $M_0$ . Since now all the matrices involved in (5) are evaluated at  $\hat{\delta}_0$ , the equality (6) holds exactly, and not only asymptotically, as when  $U_0$  is evaluated at  $\hat{\delta}_0$  and  $U_1$  at  $\hat{\delta}_1$ . The scaling correction that is now computed is

$$\hat{c}_d^{(10)} := (r_0\hat{c}_0 - r_1\hat{c}_1^{(10)})/m, \tag{7}$$

which, in view of (6), is the scaling correction of (3) when  $U_d$  is evaluated at the estimate  $\hat{\delta}_0$  when fitting  $M_0$ , and thus it is a positive number. The new scaled difference statistic is thus defined as

$$\tilde{T}_d^{(10)} := (T_0 - T_1)/\hat{c}_d^{(10)}. \tag{8}$$

Clearly,  $\tilde{T}_d^{(10)} = \tilde{T}_d$ ; that is,  $\tilde{T}_d^{(10)}$  coincides numerically with the scaled statistic (3) proposed in Satorra (2000).

### 6. An Illustration

We use these data just for illustrative purposes, and because they provide an example where the standard scaling correction fails to be positive. We use a latent variable model discussed for data by Bentler, Satorra, and Yuan (2009). The Bonett–Woodward–Randall (2002) test shows that these data have significant excess kurtosis indicative of nonnormality at a one-tail 0.05 level, so test statistics derived from ML estimation may not be appropriate and we do the scaling corrections.

The model considered is a structured means model, with the mean cigarette sales indirectly affecting the mean rates of the various cancers. The specified model is

$$V_j = \lambda_j F + E_j, \quad j = 2, \dots, 5, \quad F = \beta V_1 + D_1, \quad V_1 = \mu + E_1,$$

where the  $V_j$ 's denote observed variables;  $F$ ,  $D_1$ , and  $E_j$  are the common, residual common, and unique factors, respectively;  $\lambda_j$  denotes a factor loading parameter,  $\beta$  is the effect of cigarette smoking on the cancer factor, and  $\mu$  is the mean parameter for rates of smoking. The units of measurement for the factor were tied to  $V_2$ , with  $\lambda_2 = 1$ . The following values for the ML and Satorra–Bentler (1994) (SB) scaled chi-square statistics are obtained

$$T_1 = 107.398, \quad \tilde{T}_1 = 65.3524, \quad r_1 = 9, \quad \hat{c}_1 = 1.6434,$$

along with the degrees of freedom  $r_1$  and the scaling correction  $\hat{c}_1$ . The model does not fit, though for the sake of the illustration we are aiming for, this is not of concern to us.

*Restricted Model  $M_0$*  The same model is now fitted with the added restriction that the error variances of the kidney and leukemia cancers,  $E_4$  and  $E_5$ , are equal. This model gives the following statistics

$$T_0 = 139.495, \quad \bar{T}_0 = 97.4034, \quad r_0 = 10, \quad \hat{c}_0 = 1.4322.$$

*Difference Test* Our main interest lies in testing the difference between  $M_0$  and  $M_1$ , which we do with the chi-square difference test. The ML difference statistic is

$$T_d = 139.495 - 107.398 = 32.097,$$

which, with 1 degree of freedom ( $m$ ), rejects the null hypothesis that the error variances for  $E_4$  and  $E_5$  are equal. Since the data are not normal, we compute the SB (2001)  $\bar{T}_d$  scaled difference statistic. This requires computing the scaling factor  $\tilde{c}_d = (r_0\hat{c}_0 - r_1\hat{c}_1)/m$  given by

$$[10(1.4322) - 9(1.6434)]/1 = 14.322 - 14.7906 = -0.4686.$$

The scaling factor  $\tilde{c}_d$  is negative, so the SB difference test cannot be carried out; or if carried out, it results in an improper negative chi-square value.

*New Scaled Difference Test* As described above, to compute the scaled statistic  $\bar{T}_d$ , we implement (7) and (8). The output that is missing in the prior runs is  $\bar{T}_1^{(10)}$  the value of the SB statistic obtained at the final parameter estimates for model  $M_0$  when model  $M_1$  is evaluated. This can be obtained by creating a model setup  $M_{10}$  that contains the parameterization of  $M_1$  with start values taken from the output of model  $M_0$ . Model  $M_{10}$  is run with zero iterations, so that the parameter values do not change before output including test statistics is produced.<sup>1</sup> The new result gives

$$T_1^{(10)} = 139.495, \quad \bar{T}_1^{(10)} = 94.9551, \quad r_1 = 9, \quad \hat{c}^{(10)} = 1.4691,$$

where as expected,  $T_1^{(10)} = T_0$  as reported above (i.e., the ML statistics are identical), and the value  $\hat{c}^{(10)}$  is hand computed. As a result, we can compute

$$\hat{c}_d^{(10)} = (r_0\hat{c}_0 - r_1\hat{c}_1^{(10)})/m = [(10)(1.4322) - (9)(1.4691)] = 1.10,$$

which, in contrast to the SB (2001)  $\tilde{c}_d$  computations, is positive. Finally, we can compute the proposed new SB corrected chi-square statistic as

$$\bar{T}_d = \bar{T}_d^{(10)} = (T_0 - T_1)/\hat{c}_d^{(10)} = (139.495 - 107.398)/1.10 = 29.179,$$

which can be referred to a  $\chi_1^2$  variate for evaluation.

## 7. Discussion

The implicit function theorem was used to provide a theoretical basis for the development of a practical version of the computationally more difficult scaled difference statistic proposed by

<sup>1</sup>For this particular example, the Appendix of Satorra and Bentler (2008) illustrates this procedure with EQS (Bentler, 2008). In the same reference, there is a second illustration with a larger degree of freedom.

Satorra (2000).<sup>2</sup> The proposed method is only marginally more difficult to compute than that of Satorra and Bentler (2001) and solves the problem of an uninterpretable negative  $\chi^2$  difference test that applied researchers have complained about for some time.

Like the method it is replacing, the proposed procedure applies to a general modeling setting. The vector of parameters  $\sigma$  to be modeled may contain various types of moments: means, product-moments, frequencies (proportions), and so forth. Thus, this scaled difference test applies to methods 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  $\Gamma$  is singular, when the data is composed of various samples, as in multisample analysis, and to other estimation methods. It applies also to the case where the estimate of  $\Gamma$  reflects the fact that we have intraclass correlation among observations, as in complex samples. Hence, this new statistic should be useful in a variety of applied modeling contexts. Simulation work will be needed to understand its virtues and limitations, relative to other alternatives, in such contexts.

#### References

- Bentler, P.M. (2008). *EQS 6 structural equations program manual*. Encino: Multivariate Software.
- Bentler, P.M., Satorra, A., & Yuan, K.-H. (2009). Smoking and cancers: case-robust analysis of a classic data set. *Structural Equation Modeling*, *16*, 382–390.
- Bollen, K.A., & Curran, P.J. (2006). *Latent curve models: a structural equation perspective*. New York: Wiley.
- Bonett, D.G., Woodward, J.A., & Randall, R.L. (2002). Estimating p-values for Mardia's coefficients of multivariate skewness and kurtosis. *Computational Statistics*, *17*, 117–122.
- Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.
- Grace, J.B. (2006). *Structural equation modeling and natural systems*. New York: Cambridge University Press.
- 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 D.D.H. Heijmans & D.S.G. Pollock, A. Satorra (Eds.), *Innovations in multivariate statistical analysis: a festschrift for Heinz Neudecker* (pp. 233–247). Dordrecht: Kluwer Academic.
- 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: Sage.
- Satorra, A., & Bentler, P.M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507–514.
- Satorra, A., & Bentler, P.M. (2008). *Ensuring positiveness of the scaled difference chi-square test statistic*. Department of Statistics, UCLA Department of Statistics Preprint. <http://repositories.cdlib.org/uclastat/papers/2008010905>.
- Yuan, K.-H., & Bentler, P.M. (2007). Structural equation modeling. In C.R. Rao & S. Sinharay (Eds.) *Handbook of statistics 26: Psychometrics* (pp. 297–358). Amsterdam: North-Holland.

*Manuscript Received: 4 JUN 2008*

*Final Version Received: 25 MAY 2009*

*Published Online Date: 20 JUN 2009*

<sup>2</sup>Satorra (2000) provides also Monte Carlo evidence—on a specific model context and various sample sizes—of the superiority of the scaling correction over other alternatives such as the adjusted (mean and variance corrected) statistic.