Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic

Authors
Satorra, Albert
Bentler, P M

Publication Date
2008-01-09

Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic *

Albert Satorra

Universitat Pompeu Fabra. Barcelona

and

Peter M. Bentler

University of California, Los Angeles

May 30, 2008

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

Abstract

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 (2001). The statistic $\tilde{T}_d$ is asymptotically equivalent to the scaled difference test statistic $\bar{T}_d$ introduced in Satorra (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 $\bar{T}_d$ that avoids negative chi-square values.

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

Introduction

Moment structure analysis is widely used in behavioural, social and economic studies to analyze structural relations between variables, some of which may be latent (i.e., unobservable, Bollen, 2002); see, e.g., Bollen and Curran (2006), Grace (2006), Lee (2007), Yuan and Bentler (2007), and references therein. A wide variety of computer programs can carry out analyses for a general class of structural equation models, including AMOS (SPSS, 2008), CALIS (SAS, 2008), EQS (Bentler, 2008), LISREL (Jöreskog & Sörbom, 1999), Mplus (Muthén & Muthén, 2007), Mx (Neale, 1997), and `sem` in `R` (Fox, 2006). As is well-known, statistics that are central in moment structure analysis are the overall goodness-of-fit test of the model and tests of restrictions on parameters. When the distributional assumptions of a relevant statistical method are not met, corrections and adjustments to test statistics proposed by Satorra and Bentler (1988, 1994) are widely used. An example is the scaled test statistic $T_{SB} = T_{ML}/c$ which scales the normal theory maximum likelihood (ML) goodness of fit test T_{ML} so that T_{SB} is closer in expectation to that of a χ^2 variate than the statistic T_{ML}. Here we focus on the problem of comparing two nested models M_0 and M_1 estimated with any non-optimal (asymptotically) method when the data violates that method's distributional assumption, e.g., when the estimator is ML but the data are not multivariate normal. Then the usual chi-square difference test $T_d = T_0 - T_1$, based on the separate models' goodness of fit test statistics, is not χ^2 distributed. A correction to T_d by a scaling factor was proposed by Satorra (2000) and Satorra and Bentler (2001). The latter is the focus of this paper.

Although Satorra-Bentler (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 approach of computing separate SB-corrected test statistics T_{SB} 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 χ^2 variate. Satorra (2000) provided 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. In a subsequent paper, Satorra and Bentler (2001) provided a simple procedure to obtain an approximate scaled chi-square statistic from the regular output of SEM analysis. Under the null hypothesis, the approximate scaled chi-square is asymptotically equal to the exact scaled test statistic of Satorra (2000), but it has the drawback that a positive value for the scaling correction is not assured. The present paper provides the technical development of a simple procedure by which a researcher can compute the exact SB difference test statistic based only on output from standard SEM programs.

Throughout we adhere to the notation and results of Satorra and Bentler

(2001). 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 $\sqrt{n}(s-\sigma)$ be asymptotically normally distributed with a finite asymptotic variance matrix Γ $(p \times p)$. Consider the model $M_0 : \sigma = \sigma(\theta)$ 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

$$F_V(\theta) := (s-\sigma)^{'}\hat{V}(s-\sigma) \tag{1}$$

over the parameter space, where $\hat{V}$ $(p \times p)$, converges in probability to V, a positive definite matrix. Alternatively, let $F = F(s, \sigma)$ be a discrepancy function between s and σ in the sense of Browne (1984). In the later case, V below will be the $\frac{1}{2}\partial^2 F(s,\sigma)/\partial\sigma\partial\sigma$ evaluated at $\sigma = s$.

Let $M_0 : \sigma = \sigma^\star(\delta), \quad a(\delta) = 0$, and $M_1 : \sigma = \sigma^\star(\delta)$ be two nested models for σ. Here δ is a $(q+m)$-dimensional vector of parameters, 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 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

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

which we assume to be regular at the true value of δ, say δ_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 δ_0) a one-to-one function $\delta = \delta(\theta)$ defined in an open and compact subset S of R^q, and a θ_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 θ_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 $a(\delta(\theta)) = a_0$, it holds that with A evaluated at δ_0, $AH = 0$ with $r(A) + r(H) = p$, and $r(.)$ denoting the rank of a matrix. Thus, H' is an orthogonal complement of A. Typically, the restrictions $a(.)$ are linear, so A and H do not vary with δ_0.

Let $P((q+m)\times(q+m)) := \Pi' V \Pi$. Associated to M_1, the less restricted model $\sigma = \sigma^\star(\delta)$, 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

$$c_1 := \frac{1}{r_1} tr U_1 \Gamma = \frac{1}{r_1} tr \left\{V\Gamma\right\} - \frac{1}{r_1} tr \left\{P^{-1}\Pi' V \Gamma V \Pi\right\}, \qquad (2)$$

where

$$U_1 := V - V\Pi P^{-1} \Pi' V. \qquad (3)$$

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 χ^2 distribution with degrees of freedom $m = r_0 - r_1$. This is the restricted test of M_0 within M_1. For general distribution of the data, the asymptotic chi-square approximation may not hold. To improve on the chi-square approximation, Satorra (2000) gave explicit formulae that extends the scaling corrections proposed by Satorra and Bentler (1994) to the case of difference, Wald, and score type 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:

$$\bar{T}_d := T_d/\hat{c}_d, \text{ where } c_d := \frac{1}{m} tr\, U_d \Gamma \tag{4}$$

with

$$U_d = V \Pi P^{-1} A' (A P^{-1} A')^{-1} A P^{-1} \Pi' V. \tag{5}$$

Here, $\hat{c}_d$ denotes c_d after substituting consistent estimates of V and Γ, and evaluating the Jacobians A and Π at the estimate $\hat{\delta}_1$ when fitting M_0 (or M_1).

A practical problem with the statistic $\bar{T}_d$ is that 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. 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 show 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 (cf., 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

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

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

The asymptotic equivalence of $\bar{T}_d$ and $\tilde{T}_d$ follows from the following matrix equality:

$$U_d = U_0 - U_1, \tag{6}$$

where U_d and U_1 are given in (5) and (3), and

$$U_0 := V - V\Pi H(H^{'}\Pi^{'}V\Pi H)^{-1}H^{'}\Pi^{'}V. \tag{7}$$

Since (6) implies

$$m\, c_d = tr U_d \Gamma = tr\,(U_0 - U_1)\Gamma = r_0 c_0 - r_1 c_1,$$

it follows that

$$c_d = (r_0 c_0 - r_1 c_1)/m.$$

This is the theoretical basis for Satorra and Bentler's (2001) proposal to scale the difference test by

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

where $\hat{c}_0 := T_0/\bar{T}_0$ and $\hat{c}_1 := T_1/\bar{T}_1$ are obtained by hand computation from the standard output of a SEM program, when fitting models M_0 and M_1 in turn.

A Problem with the Current Scaled Difference Test

Under Satorra and Bentler's (2001) proposal, $\tilde{c}_d$ evaluates U_0 and U_1 at the estimates $\hat{\dot{\delta}}_0$ and $\hat{\dot{\delta}}_1$ respectively. Since $\hat{\dot{\delta}}_1$ 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 fromM_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) had warned that "...even though, necessarily, $c_d > 0$,

$\tilde{c}_d$ may turn out to be negative in some extreme cases (leading then to an improper value for $\tilde{T}_d$)... an improper value of $\tilde{T}_d$ can be taken as indication that either M_0 is highly deviant from the true model, or the sample size is too small for relying on the test statistic; that is, as indication of a non-standard situation where the difference test statistic is not worth using". Clearly, under a sequence of local alternatives, $\bar{T}_d$ and $\tilde{T}_d$ are asymptotically equivalent, since both $\hat{c}_d$ and $\tilde{c}_d$ are consistent estimates of the population value c_d. Thus for large samples, and for not too-large misspecifications, an improper value of $\tilde{T}_d$ will not occur.

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 $\bar{T}_d$ using the formulae spelled out in Satorra (2000). Unfortunately this is impractical or impossible for most applied researchers, since Satorra's methodology involves statistics that are not computed in standard SEM software. Fortunately, as we show next, the exact value of $\bar{T}_d$ can also be obtained from the standard output of SEM software, using a new hand computation.

A New Scaled Test Statistic $\bar{T}_d$

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 $\hat{c}_1{}^{(10)} := T_1{}^{(10)}/\bar{T}_1{}^{(10)}$, where $T_1{}^{(10)}$ and $\bar{T}_1{}^{(10)}$ 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 Π 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 (6) are evaluated at $\hat{\delta}_0$, the equality holds exactly, and not only asymptotically, as when some matrices are evaluated at $\hat{\delta}_0$ and some others at the estimate $\hat{\delta}_1$ under model M_1. The scaling correction that is now computed is

$$\hat{c}_d^{(10)} := (r_0\hat{c}_0 - r_1\hat{c}_1^{(10)})/m, \tag{8}$$

which now, necessarily, is a positive number. The new scaled difference statistic is thus defined as

$$\bar{T}_d^{(10)} = (T_0 - T_1)/\hat{c}_d^{(10)}, \tag{9}$$

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

In the next section, we will illustrate this procedure for two models with an empirical data set on which the original SB difference test produced negative values.

Empirical Examples: Effect of Smoking on Cancers

Fraumeni (1968) reported a pioneering epidemiological study of the effect of smoking on various cancers. He investigated the bivariate correlations and regressions between the per capita sales of cigarettes on the one hand, and variation in mortality from bladder, lung, kidney, and leukemia cancers, on the other. The original data, given at `http://lib.stat.cmu.edu/DASL/Datafiles/cigcancerdat.html`, represent smoking rates in 44 states in the USA and the associated age-adjusted death rates for the four cancers. Fraumeni did not use latent variable models, so here we use these interesting data to evaluate some variants of a simple one factor model for the various cancers, and its prediction by rates of smoking. The Bonett-Woodward-Randall (2002) test shows that these data have significant excess kurtosis indicative of non-normality at a one-tail .05 level, so test statistics derived from ML estimation may not be appropriate and we do the SB corrections.[1]

Prediction of Cancer as a One Factor Model

In the following, V_1 represents the quantity of cigarettes sold, while $V_2 - V_5$ represent bladder, lung, kidney, and leukemia cancers. A common factor F is hypothesized to explain the correlations among the four types of cancer, and this factor, in turn, is predicted by quantity of cigarette sales. It is a structured means model, with the mean cigarette sales indirectly affecting the mean rates of the various cancers. The specified model is

$$V_j = \lambda_j F + E_j, \quad j = 2, \ldots, 5$$

$$F = \beta V_1 + D_1,$$

$$V_1 = \mu + E_1,$$

[1] In fact the data have outliers, so case-robust methods may be more appropriate; see Bentler, Satorra, and Yuan (2008), who also provide a model closer to Fraumeni's original hypotheses.

where the V_j's denote observed variables; F, D_1, and E_j are the common, residual common, and unique factors respectively; λ_j denotes a factor loading parameter, β is the effect of cigarette smoking on the cancer factor, and μ 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 SB chi-square statistics are obtained

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

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.[2]

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

[2]Substantively, it may be interesting to note that the standardized factor loadings on the cancer factor are in the range .78-.88, and that the standardized effect of smoking rates on the cancer factor is .88.

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

which, with 1 degree of freedom (m), rejects the null hypothesis that the error variances for E4 and E5 are equal. Since the data is not normal, we compute the SB (2001) 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

$$[10(1.4322) - 9(1.6434)]/1 = 14.322 - 14.7906 = -.4686.$$

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 (8) and (9). The output that is missing in the prior runs is 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. In the Appendix we illustrate this procedure with EQS. The new result gives

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

where as expected, $T^{10} = T_0$ 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) 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 χ_1^2 variate for evaluation.

The One Factor Model with More Restrictions

Since the number of degrees of freedom could impact the scaling factor and hence the performance of a χ^2 difference test, we consider another model on the same data. This model is identical in form to the previous one, but adds the constraint that all error variances on variables defining the factor are held equal. Specifically, model M_1 is the same as before, but a new model M_0 is computed with the following result

$$T_0 = 178.508, \quad \bar{T}_0 = 151.4442, \quad r_0 = 12, \quad \hat{c}_0 = 1.1787.$$

For reference purposes, we may compute the SB (2001) scaling factor $\tilde{c}_d := (r_0\hat{c}_0 - r_1\hat{c}_1)/m$, which equals -.2154. Hence the (2001) difference test again cannot be computed for this model. The new statistic, in contrast, behaves well. To obtain it, using procedures described earlier, we compute model M_{10} with the result

$$T^{(10)} = 178.508, \quad \bar{T}^{(10)} = 177.6320, \quad r_1 = 9, \quad \hat{c}^{(10)} = 1.0049.$$

Then we can compute

$$\hat{c}_d^{(10)} = (r_0\hat{c}_0 - r_1\hat{c}_1^{(10)})/m = [(12)(1.1787) - (9)(1.0049)]/3 = 1.70$$

and

$$\bar{T}_d = \bar{T}_d^{(10)} = (T_0 - T_1)/\hat{c}_d^{(10)} = (178.508 - 107.398)/1.10 = 29.179,$$

which is positive as expected, and can be evaluated by reference to a χ^2_3 distribution.

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 Satorra (2000). 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 χ^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 σ 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 Γ is singular, when the data is composed of various samples, as in multi-sample analysis, and to other estimation methods, e.g., pseudo ML estimation. It applies also to the case where the estimate of Γ 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.

References

Bentler, P. M. (2008). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., Satorra, A., & Yuan, K. –H. (2008). *Smoking and cancers: Case-robust analysis of a classic data set.* Under editorial review.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605-634.

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 Psychol-*

ogy, *37*, 62-83.

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.

Fox, J. (2006). Structural equation modeling with the sem package in R. *Structural Equation Modeling*, *13*, 465-486.

Fraumeni, J. F. Jr. (1968). Cigarette smoking and cancers of the urinary tract: Geographic variation in the United States. Journal of the National Cancer Institute, 41, 1205-1211.

Grace, J. B. (2006). *Structural equation modeling and natural systems.* New York: Cambridge.

Jöreskog, K., and Sörbom, D. (1999). *LISREL 8 user's reference guide* (2nd Ed). Lincolnwood, IL: Scientific Software International.

Lee, S. –Y. (Ed.) (2007). *Handbook of latent variable and related models.* Amsterdam: North-Holland.

Muthén, B. O., & Muthén, L.K. (2007). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.

Neale, M. C. (1997). Mx: *Statistical modeling.* Box 980126 MCV, Richmond, VA.

SAS (2008). *SAS/STAT 9.2 user's guide: The CALIS procedure.* Carey, NC: SAS Institute.

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. (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.

SPSS (2008). *Amos16.* Chicago: SPSS.

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.

Appendix

EQS Procedure for Implementing New Robust Difference Test

1. Set up the restricted model M_0 file as usual, and add the command

/PRINT

RETEST =”newfile.eqs”;

where *newfile.eqs* is the proposed name of a future model file. Run this model M_0 using METHOD = ML,ROBUST.

Find ML T_0, SB $\bar{T}_0$, and degrees of freedom r_0, and compute $\hat{c}_0 = T_0/\bar{T}_0$.

2. The *newfile.eqs* contains, in the top half, an echo of the model setup for M_0. In this section, delete everything from /EQUATIONS to /END. The remaining file contains the setup for model M_{10}. Keeping all start values and parameters as is, modify this file to become model M_1 (e.g., remove the constraints in M_0; if M_0 has fixed a zero parameter, add it as a free parameter.) Make sure all newly added parameters have start value zero. Add the statement

/TECH

ITER =0;

to the model file, save with a new name (say, *m01.eqs*) and run. Obtain the ML T^{10} and verify that it is identical to T_0 from the previous run. Obtain the SB $\bar{T}^{10}$, r_1, and $\hat{c}^{10}$. Then compute the new scaling factor $\hat{c}_d^{10} = (r_0\hat{c}_0 - r_1\hat{c}_1^{(10)})/m$.

3. Take the input file *m01.eqs*, delete the statements /TECH and ITER =0; and run. This is the typical run of model M_1, and gives the ML statistic T_1. Compute

the SB robust difference test as $\bar{T}_d = (T_0 - T_1)/\hat{c}_d^{10}$.