PSYCHOMETRIKA—VOL. 66, NO. 4, 507–514
DECEMBER 2001

A SCALED DIFFERENCE CHI-SQUARE TEST STATISTIC FOR MOMENT STRUCTURE ANALYSIS

ALBERT SATORRA

UNIVERSITAT POMPEU FABRA. BARCELONA

PETER M. BENTLER

UNIVERSITY OF CALIFORNIA, LOS ANGELES

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, that is, nonasymptotic and nonnormal, applications. In a recent paper, Satorra (2000) shows that the difference between two SB 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.

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 behavioural, social and economic studies to analyze structural relations between variables, some of which may be latent (i.e., unobservable); see, for example, 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 & Sörbom, 1994; EQS of Bentler, 1995). In multisample 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 nonnormal 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 & Bentler (1994; Satorra and Bentler, 1988a, 1988b) developed a family

This research was supported by the Spanish grants PB96-0300 and BEC2000-0983, and USPHS grants DA00017 and DA01070.

Requests for reprints should be sent to Albert Satorra, Universitat Pompeu Fabra, Ramon Trias Fargas, 25–27, 08005 Barcelona, SPAIN. E-Mail: satorra@upf.es

0033-3123/2001-4/1999-0759-A $00.75/0

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 & Satorra, 1991; Curran, West & Finch, 1996; Hu, Bentler & Kano, 1992). Bentler and Yuan (1999) provide a recent comparison of alternative test methods for small samples. Extension of Satorra-Bentler (SB) corrections to goodness-of-fit test statistics in the case of the analysis of augmented moment structures, multisamples 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 & 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 (2000) 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 $\sqrt{n}(s - \sigma)$ 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 σ, 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. 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, that is, when AO does not necessarily hold, T_0 is asymptotically distributed as a mixture of chi-square distributions of 1 degree of freedom (df) (see Satorra & Bentler, 1986); that is,

$$T_0 \xrightarrow{L} \sum_{j=1}^{r_0} \alpha_j \chi_j^2, \tag{2}$$

as $n \to \infty$, where the χ_j^2 are independent chi-square variables of 1 df, and the α_j are the nonnull 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. Clearly, when the α_i are all equal to α, say, then T/α is an asymptotic chi-square statistic. Conditions where that occurs are discussed in Satorra and Bentler (1986). In the context of structural models and for general types of distributions, Satorra and Bentler (1994; Satorra & Bentler, 1988a, 1988b) proposed replacing T by the scaled statistic

$$\overline{T} = T/\hat{c}, \tag{3}$$

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

$$c := \frac{1}{r_0}\text{tr}\, U_0\Gamma = \frac{1}{r_0}\text{tr}\,\{V\Gamma\} - \frac{1}{r_0}\text{tr}\,\left\{(\Delta' V \Delta)^{-1}\Delta' V \Gamma V \Delta\right\}. \tag{4}$$

In some situations, computations involving the large matrix V can be simplified using methods suggested in Kano (1992). Note that the SB scaled test statistic has the same asymptotic mean as the corresponding $\chi^2_{r_0}$ variate. The SB scaled goodness-of-fit test has been shown to outperform alternative test statistics in a variety of models and nonnormal distributions (e.g., Chou, Bentler & Satorra, 1991; Curran, West & Finch, 1996; Hu, Bentler & Kano, 1992). Of course, when AO 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 & Bentler, 1994; see also the illustration section of the present paper).

A goodness-of-fit statistic which can be used given any estimation method, is given by

$$T^\star := n(s-\hat{\sigma})'\{\hat{\Gamma}^{-1} - \hat{\Gamma}^{-1}\hat{\Delta}(\hat{\Delta}'\hat{\Gamma}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\hat{\Gamma}^{-1}\}(s-\hat{\sigma}). \tag{5}$$

When $\hat{\Gamma}$ is a (distribution-free) consistent estimator of Γ (as in (16) below), then $T^\star$ will be called an 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 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 $m \times 1$ known vector, and $\sigma^\star(.)$ and $a(.)$ are twice-continuously differentiable vector-valued functions of $\delta \in \Theta_1$, a compact subset of $\mathcal{R}^{q+m}$. 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

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

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 $P((q+m)\times(q+m)) := \Pi' V \Pi$ 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

$$c_1 := \frac{1}{r_1}\text{tr}\, U_1\Gamma = \frac{1}{r_1}\text{tr}\,\{V\Gamma\} - \frac{1}{r_1}\text{tr}\,\{P^{-1}\Pi' V \Gamma V \Pi\} \tag{6}$$

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 distribution 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 $\overline{T}_d$. Extending his earlier work (Satorra, 1989), Satorra (2000) recently provided formulae for computing such scaled statistics for the difference, Score and Wald test statistics. From Satorra's formulae it becomes apparent 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 $\overline{T}_d \neq \overline{T}_0 - \overline{T}_1$, where by $\overline{T}_0$ and $\overline{T}_1$ 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 (2000), the SB scaled difference test statistic is defined as $\overline{T}_d := T_d/\hat{c}_d$ where $\hat{c}_d$ is a consistent estimate of

$$c_d := \frac{1}{m}\operatorname{tr} U_d \Gamma \tag{7}$$

with

$$U_d = V\Pi P^{-1}A'(AP^{-1}A')^{-1}AP^{-1}\Pi'V. \tag{8}$$

As follows from Satorra (1989), when the nonnull eigen-values of $U_d\Gamma$ are equal, not necessarily equal to 1, then the scaled statistic $\overline{T}_d$ is asymptotically a chi-square statistic. Note that necessarily $c_d \geq 0$, since the eigen-values of $U_d\Gamma$ are non-negative. Let $\hat{\delta}_0$ and $\hat{\delta}_1$ be the estimated values of δ when fitting models $\mathcal{M}_0$ and $\mathcal{M}_1$ respectively. A consistent estimate of c_d is obtained when Γ is replaced by a distribution-free consistent estimate, as in (16) below, and U_d of (8) is evaluated at one of the estimates $\hat{\delta}_0$ or $\hat{\delta}_1$. Note that as with c_d, we have that $\hat{c}_d > 0$.

A practical problem with the statistic $\overline{T}_d$ 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. Here we will show how to combine the estimates of 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 a consistent estimate of the scaling correction c_d for the difference test statistic. It turns out that the computations are extremely simple 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 $m \times (q + m)$ matrix), that is, the matrix A' is an orthogonal complement of H. We have

$$\begin{aligned} U_0 - U_1 &= V\Pi(\Pi'V\Pi)^{-1}\Pi'V - V\Pi H(H'\Pi'V\Pi H)^{-1}H'\Pi'V \\ &= V\Pi\{P^{-1} - H(H'PH)^{-1}H'\}\Pi'V \end{aligned}$$

since

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

as $AH = 0$ (see Rao, 1973, p. 77). We thus have the basic result that

$$U_d = U_0 - U_1.$$

Now, since $r_0c_0 - r_1c_1 = \operatorname{tr}(U_0 - U_1)\Gamma = \operatorname{tr} U_d\Gamma = mc_d$, we obtain $c_d = (r_0c_0 - r_1c_1)/m$. This means that consistent estimation of c_d is available from consistent estimates of the scaling corrections c_0 and c_1 associated with the models $\mathcal{M}_0$ and $\mathcal{M}_1$ 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 $\overline{T}_0$ when fitting $\mathcal{M}_0$, and T_1 and $\overline{T}_1$ 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 $\hat{c}_0 = T_0/\overline{T}_0$ and $\hat{c}_1 = T_1/\overline{T}_1$, and the usual chi-square difference $T_d = T_0 - T_1$. The SB scaled difference test is then given as $\tilde{T}_d := T_d/\tilde{c}_d$, where

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

Note that $\tilde{c}_d$ is obtained evaluating U_0 and U_1 at the estimates $\hat{\delta}_0$ and $\hat{\delta}_1$ respectively. Thus, even though, necessarily, $c_d \geq 0$, $\tilde{c}_d$ may turn out to be negative in some extreme cases (leading then to an improper value for $\tilde{T}_d$). Such an improper value requires that $\hat{\delta}_0$ and $\hat{\delta}_1$ deviate substantially from each other, as a result of a small sample size artifact, or the null model $\mathcal{M}_0$ being too deviant from the true model. In fact, our asymptotic theory of the difference test statistic uses the classical assumption of a sequence of local alternatives (see, e.g., Satorra, 1989; Satorra, 2000, p. 235), which requires $\mathcal{M}_0$ and $\mathcal{M}_1$ to be non-grossly misspecified. Thus, an improper value of $\tilde{T}_d$ can be taken as indication that either $\mathcal{M}_0$ 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. (Of course, if needed, more complex computations involving only $\hat{\delta}_0$ or $\hat{\delta}_1$ produce the statistic $\overline{T}_d$ described above, which value is certainly non-negative.) Clearly, under a sequence of local alternatives, $\overline{T}_d$ 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.

When the population values of the two scaling corrections c_0 and c_1 are equal, that is, $c_0 = c_1 = c$, then $c_d = c$; thus, this is a case where the simple difference of SB scaled chi-square test statistics, $d\overline{T} := \overline{T}_0 - \overline{T}_1$ is asymptotically equivalent to $\tilde{T}_d$. This is the case, for example, when $c_0 = c_1 = 1$, that is, when both un-scaled 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, loglinear 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 multisample analysis. The results apply also to other estimation methods, for example, pseudo ML estimation.

It is important to recognize that a competitor to the statistic $\tilde{T}_d$ will be the difference between asymptotic 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 of (5) (with $\hat{\Gamma}$ a distribution-free consistent estimate of Γ) associated to the models $\mathcal{M}_0$ and $\mathcal{M}_1$ respectively (e.g., Satorra, 1989). In the next section, we will illustrate using Monte Carlo methods the small sample size performance of the competing test statistics for H_0.

4. Illustration

In this section a simple model context is used to illustrate the performance in finite samples of the described test statistics. The model context is a regression with errors-in-variables. We consider a regression equation

$$y^\star_{gi} = \beta x_{gi} + v_{gi}, \quad i = 1, \ldots, n_g, \tag{9}$$

where for case i in group g ($g = 1, 2$), $y^\star_{gi}$ 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 $x^{\star}_{1gi}$ and $x^{\star}_{2gi}$ related to x_{gi} by the following measurement-error equations

$$x^{\star}_{1gi} = x_{gi} + u_{1gi}, \quad x^{\star}_{2gi} = x_{gi} + u_{2gi}, \tag{10}$$

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 (9) and (10) 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 (9) and (10) as

$$z_{gi} = \Lambda \xi_{gi}, \quad i = 1, 2, \ldots, n_g, \; g = 1, 2, \tag{11}$$

where

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

and

$$\Lambda := \begin{pmatrix} \beta & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{12}$$

Define

$$\Phi := E\xi_{gi}\xi'_{gi} = \begin{pmatrix} \sigma_{xx} & 0 & 0 & 0 \\ 0 & \sigma_{vv} & 0 & 0 \\ 0 & 0 & \sigma_{uu} & 0 \\ 0 & 0 & 0 & \sigma_{uu} \end{pmatrix}, \tag{13}$$

and the parameter vector $\theta := (\sigma_{vv}, \sigma_{xx}, \sigma_{uu}, \beta)'$. Under this set-up, $\Sigma_g := Ez_{gi}z'_{gi}$ has the moment structure

$$\Sigma_g = \Lambda\Phi\Lambda' = \Sigma(\theta), \tag{14}$$

where $\Sigma(.)$, $\Lambda(.)$ and $\Phi(.)$ are (twice-continuously differentiable) matrix-valued functions of θ, as deduced from (12), (13) and (14). 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 $\sigma'_g = \text{vec}\Sigma_g$ 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 & Neudecker, 1999, for full details on this operator). We consider normal theory WLS estimation, where $\hat{V}$ of (1) has the form

$$\hat{V} := \text{block diag}\left(\frac{n_1}{n}\hat{V}_1, \frac{n_2}{n}\hat{V}_2\right), \tag{15}$$

with $n = n_1 + n_2$ and $\hat{V}_g = \frac{1}{2}(S_g{}^{-1} \otimes S_g{}^{-1})$, $g = 1, 2$. We further assume that the matrices S_g and Σ_g are positive definite, and that $n_g/n \to f_g > 0$, as $n \to +\infty$, $g = 1, 2$. Clearly, when there is independence across groups, the asymptotic variance matrix of $\sqrt{n}(s - \sigma)$ is of the form $\Gamma = \text{block diag}\,(f_1^{-1}\Gamma_1, f_2^{-1}\Gamma_2)$ where Γ_g is the asymptotic variance of $\sqrt{n_g}(s_g - \sigma_g)$. In this case, a distribution-free consistent estimator of Γ is

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

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

We now consider a Monte Carlo study (with 2000 replications), in which, for each replication, two-sample data is generated from the above regression model, with population values of the parameters given by $\theta_0 = (1, 1, .3, 2)'$. Regarding the distribution of random constituents of the model, the values of v and x are generated as i.i.d. independent χ_1^2 (i.e., a highly nonnormal distribution), conveniently scaled to have zero mean and unit variance. The variables u_1 and u_2 are generated as normal variables, mutually independent, and independent also of v and x. Two models are fitted for each simulated (two-sample) data set. Model $\mathcal{M}_0$ that restricts the parameter vector θ to be invariant across groups, and model $\mathcal{M}_1$ that allows θ to vary across groups. For each simulated data set, we compute the usual chi-square goodness-of-fit test statistics, T_0 and T_1; the SB scaled statistics, $\overline{T}_0$ and $\overline{T}_1$; and the robust test statistics $T_0^\star$ and $T_1^\star$. Three combinations of samples sizes are used (small samples: $n_1 = 100$ and $n_2 = 120$; moderate: $n_1 = 300$ and $n_2 = 400$; large samples: $n_1 = 800$ and $n_2 = 900$).

Summary results of the Monte Carlo study are shown in Table 1.[1] From this table we see that in the case of small samples, and our specific model context, the new statistic $\tilde{T}_d$ outperforms the alternative asymptotic robust test statistic $T_d^\star$. In the case of large samples, $T_d^\star$ outperforms the

TABLE 1.
Monte Carlo results: empirical significance levels of test statistics

nominal significance levels:	1%	5%	10%	20%
	$n_1 = 100$ and $n_2 = 120$			
$\tilde{T}_d$	1.35	5.45	10.10	20.70
$T_d^\star$	2.40	9.00	15.45	28.30
T_d	21.40	39.10	49.15	61.25
$d\overline{T}$	76.70	77.20	78.00	78.80
	$n_1 = 300$ and $n_2 = 400$			
$\tilde{T}_d$	1.55	5.80	10.60	20.25
$T_d^\star$	0.95	5.10	11.85	23.15
T_d	27.90	43.70	51.70	64.10
$d\overline{T}$	71.60	72.40	73.00	74.30
	$n_1 = 800$ and $n_2 = 900$			
$\tilde{T}_d$	1.65	6.55	11.35	20.70
$T_d^\star$	0.80	4.75	10.50	21.10
T_d	30.45	44.05	53.75	64.20
$d\overline{T}$	70.05	70.08	71.25	72.45

[1]Under the current Monte Carlo design of 2000 replications, the standard error of a cell of the table is $(p(100-p)/2000)^{1/2}$, where p denotes the percentage reported in the cell. Under the null hypothesis that the nominal values are the true ones, we obtain the values .22, .49, .67 and .89 for the standard errors of cells in columns 1, 5, 10 and 20, respectively. In the use of these standard errors, we should warn on the statistical dependence of percentages across columns, since they are based on the same set of 2000 replications (though, of course, there is independence across the various sample sizes experiments).

alternative test statistics, though $\tilde{T}_d$ shows also an accurate performance. Especially interesting is that the statistic $d\overline{T} = \overline{T}_0 - \overline{T}_1$ 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. We should note that no improper value of $\tilde{T}_d$ (i.e., a negative value) was found in all replications in the study (in clear contrast with dT, for which a substantial proportion of improper (negative) values were found). The present Monte Carlo study is just an illustration of the comparative performance of $\tilde{T}_d$ with alternative difference test statistics discussed in this paper. It would be of interest, of course, a more intensive Monte Carlo study that compares the small sample size performance of $\tilde{T}_d$ with other scaled restricted test discussed in Satorra (2000). Such an extended Monte Carlo evaluation, however, exceeds the scope of the present paper.

References

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

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

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

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

Byrne, B.M., & 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.

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

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

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

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

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

Kano, Y. (1992). Robust statistics for test-of-independence and related structural models. *Statistics and Probability Letters, 15*, 21–26.

Magnus, J., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. (rev. ed.) New York, NY: Wiley.

Muthén, B. (1993). Goodness of fit test with categorical and other nonnormal variables. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 205–234). Newbury Park, CA: Sage Publications.

Rao, C.R., (1973). *Linear statistical inference and its applications* (2nd. ed.). New York, NY: Wiley.

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

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

Satorra, A. (2000). Scaled and adjusted restricted tests in multisample 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, The Netherlands: Kluwer Academic Publishers.

Satorra, A., & 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.

Satorra, A., & 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.

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

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.

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 modelling. *British Journal of Mathematical and Statistical Psychology, 51*, 289–309.

Manuscript received 23 AUG 1999
Final version received 1 DEC 2000