

Scaled and adjusted restricted tests in multi-sample analysis of moment structures *

Albert Satorra
Universitat Pompeu Fabra

July 15, 1999

*The author is grateful to Peter Bentler and Bengt Muthén for their very useful comments. Research supported by the Spanish DGES grant PB96-0300.

Abstract

We extend to score, Wald and difference test statistics the scaled and adjusted corrections to goodness-of-fit test statistics developed in Satorra and Bentler (1988a,b). The theory is framed in the general context of multi-sample analysis of moment structures, under general conditions on the distribution of observable variables. Computational issues, as well as the relation of the scaled and corrected statistics to the asymptotic robust ones, is discussed. A Monte Carlo study illustrates the comparative performance in finite samples of corrected score test statistics.

Keywords: Moment-structures, Goodness-of-fit test, score test, Wald test, scaling corrections, 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), 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. One issue which is central in moment structure analysis is the goodness-of-fit test of the model and the test of restrictions on parameters.

Asymptotic distribution-free (ADF) methods which do not require distribution assumptions on the observable variables have been developed (Browne, 1984; Muthén, 1989; Bentler, Lee and Weng, 1987). 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 may still lack robustness to small and medium-sized samples. As an alternative to the asymptotic 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). 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, 1993) and Muthén (1993). In recent personal communication, Bengt Muthén has kindly pointed to this author the need to extend SB's corrections to score and difference test statistics.

The purpose of the present paper is to extend SB's corrections to score (Lagrange multiplier), difference and Wald test statistics. This is undertaken in the general context of multi-sample analysis. We rely heavily on the notation and results of Satorra (1989) and Satorra and Bentler (1994; Satorra

and Bentler, 1988a,b).

The chapter is structured as follows. In Section 2 we review the general theory of normal-theory GLS analysis of multi-sample moment structures, placing special emphasis on goodness-of-fit test statistics. The extension of the scaling and adjusted corrections to the restricted tests is undertaken in Section 3. Finally, Section 4 illustrates, by means of a Monte Carlo study, the finite sample size performance of competing score test statistics developed in Section 3.

2 Goodness-of-fit tests

Let σ and s be p -dimensional vectors of population and sample moments respectively, with

$$\sqrt{n}(s - \sigma^0) \xrightarrow{L} \mathcal{N}(0, \Gamma), \quad (1)$$

where σ^0 is the probability limit of s , as sample size $n \rightarrow +\infty$, Γ is the asymptotic variance matrix of $\sqrt{n}s$, and “ \xrightarrow{L} ” denotes convergence in distribution as $n \rightarrow \infty$. Consider a moment structure for σ , $\sigma = \sigma(\theta)$, where $\sigma(\cdot)$ is a twice-continuously differentiable vector-valued function of a q -dimensional parameter vector $\theta \in \Theta$, a compact subset of R^q . Let $\hat{\theta}$ be an estimator of θ that satisfies

$$\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1), \quad (2)$$

with θ_0 in the interior of Θ . The usual notation of stochastic orders of magnitude is used, i.e., $O_p(1)$ and $o_p(1)$ denote, respectively, “bounded” and “tends to zero in probability”, as $n \rightarrow +\infty$. Usually, (2) is fulfilled whenever $\sigma = \sigma(\theta)$ is locally identified (viz., the function $\sigma(\theta)$ is injective in a neighbourhood of the true parameter value). Denote $\hat{\sigma} := \sigma(\hat{\theta})$ and $\sigma_0 := \sigma(\theta_0)$. It is assumed that $\sqrt{n}(\hat{\theta} - g(s)) = o_p(1)$, where $g(\cdot)$ is a q -dimensional continuously differentiable vector-valued function of s , with $g(\cdot)$ Fisher-consistent for θ , i.e. $g(\sigma(\theta)) = \theta$, $\theta \in \Theta$. The theory discussed involves the use of the Jacobian matrices $\Xi := (\partial/\partial s')g(s)$ and $\Delta := (\partial/\partial \theta')\sigma(\theta)$, which will be assumed to be regular at $\sigma_0 := \sigma(\theta_0)$ and θ_0 respectively. Note that Fisher-consistency of $g(s)$ implies that $\Xi\Delta = I_q$. In the developments that follow, Δ and Ξ denote the corresponding matrices evaluated at θ_0 and σ_0 respectively.

The above mentioned properties for $\hat{\theta}$ are satisfied by most of the estimators commonly in use in moment structure analysis. In particular, they

are satisfied by the ML and GLS estimators discussed below. They are also satisfied in the case of instrumental variable estimators (e.g., Satorra and Bentler 1991).

A Wald-type statistic for testing the goodness-of-fit of the model can be developed. Since

$$\sqrt{n}(s - \hat{\sigma}) = \sqrt{n}(s - \sigma^0) + \sqrt{n}(\sigma^0 - \sigma_0) - \sqrt{n}(\hat{\sigma} - \sigma_0),$$

a Taylor series expansion of $\sqrt{n}(\hat{\sigma} - \sigma_0)$ at σ_0 , gives

$$\sqrt{n}(s - \hat{\sigma}) = \sqrt{n}(s - \sigma^0) + \sqrt{n}(\sigma^0 - \sigma_0) - \Delta\sqrt{n}(\hat{\theta} - \theta_0) + o_p(1),$$

which combined with (1) yields

$$\sqrt{n}(s - \hat{\sigma}) \xrightarrow{L} \mathcal{N}(\tau, H), \quad (3)$$

where

$$\tau := \lim_{n \rightarrow +\infty} \sqrt{n}(\sigma^0 - \sigma_0) \quad (4)$$

and

$$H := (I - \Delta\Xi)\Gamma(I - \Delta\Xi)'. \quad (5)$$

We further assume

$$\tau = \lim_{n \rightarrow +\infty} \sqrt{n}(\sigma^0 - \sigma_0) < +\infty,$$

which is the usual assumption of a sequence of local alternatives (needed to ensure finite asymptotic distribution of test statistics when the model is not exactly true).

Let Δ_\perp be an orthogonal complement of Δ , i.e., Δ_\perp is a $p \times (p - q)$ matrix of full-column rank such that $\Delta'_\perp \Delta = 0$. From (3), we have

$$\sqrt{n}\Delta'_\perp(s - \hat{\sigma}) \xrightarrow{L} \mathcal{N}(\Delta'_\perp\tau, \Delta'_\perp\Gamma\Delta_\perp), \quad (6)$$

since $\Delta'_\perp H \Delta_\perp = \Delta'_\perp \Gamma \Delta_\perp$, as $\Delta'_\perp(I - \Delta G) = \Delta'_\perp$; hence, a residual-based goodness-of-fit test for the model is defined as the Wald-test statistic

$$T_R = n(s - \hat{\sigma})'\hat{\Delta}_\perp(\hat{\Delta}'_\perp\hat{\Gamma}\hat{\Delta}_\perp)^+\hat{\Delta}'_\perp(s - \hat{\sigma}), \quad (7)$$

where $\hat{\Delta}_\perp$ and $\hat{\Gamma}$ are consistent estimators of Δ_\perp and Γ , respectively. The matrix $\hat{\Delta}_\perp$ will be taken to be an orthogonal complement of $\hat{\Delta}$, the matrix

Δ evaluated at the estimated value $\hat{\theta}$. Clearly, from the standard theory of Wald-test statistics (e.g., Moore, 1977), T_R is asymptotically chi-square distributed with degrees of freedom $r := \text{rank } \Delta'_\perp \Gamma \Delta_\perp$ and non-centrality parameter $\lambda := \tau' \Delta_\perp (\Delta'_\perp \Gamma \Delta_\perp)^+ \Delta'_\perp \tau$. When $\hat{\Gamma}$ is non-singular, we have

$$T_R = n(s - \hat{\sigma})' \{ \hat{\Gamma}^{-1} - \hat{\Gamma}^{-1} \hat{\Delta} (\hat{\Delta}' \hat{\Gamma}^{-1} \hat{\Delta})^{-1} \hat{\Delta}' \hat{\Gamma}^{-1} \} (s - \hat{\sigma}), \quad (8)$$

When $\hat{\Gamma}$ is the (distribution-free) consistent estimator of Γ in (10) below, then T_R will be called the asymptotic robust goodness-of-fit test statistic, since it is an asymptotic chi-square statistic regardless of the distribution of z . 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), and found to be adequate only in very large samples.

Consider now multi-sample data from G population or groups, i.e., let $\{z_{gi}\}_{i=1}^{n_g}$, $g = 1, \dots, G$, where it is assumed that, for each group g , the z_{gi} are independent and identically distributed p_g -dimensional vectors of observable variables, with finite fourth-order moments. Define $\Sigma_g := E z_{gi} z'_{gi}$, and let the multi-sample vectors σ and s of population and sample moments, respectively, be given as $\sigma := (\sigma'_1, \dots, \sigma'_1)'$ and $s := (s'_1, \dots, s'_1)'$, where $\sigma'_g := \text{vec } \Sigma_g$, $s'_g = \text{vec } S_g$, 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). Clearly, when there is independence across samples, the asymptotic variance matrix of $\sqrt{n}s$ is of the form

$$\Gamma = \text{block diag} \left(\frac{n}{n_1} \Gamma_1, \dots, \frac{n}{n_G} \Gamma_G \right), \quad (9)$$

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

$$\hat{\Gamma} := \text{block diag} \left(\frac{n}{n_1} \hat{\Gamma}_1, \dots, \frac{n}{n_G} \hat{\Gamma}_G \right), \quad (10)$$

where

$$\hat{\Gamma}_g := \frac{1}{n_g - 1} \sum_{i=1}^{n_g} (d_{gi} - s_g)(d_{gi} - s_g)', \quad (11)$$

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

We consider the generalised least-square (GLS) estimation method with normality as a working assumption; viz., the estimator $\hat{\theta}$ is defined as the minimiser over the parameter space Θ of the fitting function

$$F_{GLS}(\theta) := (s - \sigma)' \hat{V} (s - \sigma),$$

with

$$\hat{V} := \text{block diag}\left(\frac{n_1}{n} \hat{V}_1, \dots, \frac{n_G}{n} \hat{V}_G\right) \quad (12)$$

and $\hat{V}_g = \frac{1}{2}(S_g^{-1} \otimes S_g^{-1})$; clearly, $F_{GLS}(\theta) = \frac{1}{2} \sum_{g=1}^G \frac{n_g}{n} \text{tr}\{(S_g - \Sigma_g)S_g^{-1}\}^2$. For single-sample covariance structure analysis, this method of estimation was first proposed by Browne (1974). When in the minimization process the matrices S_g in (12) are replaced iteratively by the fitted $\hat{\Sigma}_g$, the so-called normal-theory reweighted GLS estimates are obtained.

Differentiation with respect to θ yields the following first-order condition for the estimator $\hat{\theta}$:

$$(s - \hat{\sigma})' \hat{V} \hat{\Delta} = 0, \quad \left(\sum_{g=1}^G \frac{n_g}{n} (s_g - \hat{\sigma}_g)' \hat{V}_g \hat{\Delta}_g = 0 \right). \quad (13)$$

A related estimation method which is asymptotically equivalent to normal-theory GLS, and which produces estimators numerically equal to the reweighted GLS estimators, is pseudo-ML estimation; viz., $\hat{\theta}$ is the minimiser over Θ of

$$F_{ML}(\theta) = \sum_{g=1}^G \frac{n_g}{n} F_{ML}^{(g)}(\theta),$$

with

$$F_{ML}^{(g)}(\theta) := \ln |\Sigma_g| + \text{tr } S \Sigma_g^{-1} - \ln |S_g| - p_g^*,$$

and p_g^* denoting the dimension of the z_{gi} .

A goodness-of-fit test statistic associated with GLS estimation is defined as $T_{GLS} := n F_{GLS}(s, \hat{\sigma})$, which, using (13), can be written as

$$T_{GLS} = (s - \hat{\sigma})' \left(\hat{V} - \hat{V} \hat{\Delta} (\hat{\Delta}' \hat{V} \hat{\Delta})^{-1} \hat{\Delta}' \hat{V} \right) (s - \hat{\sigma}).$$

The goodness-of-fit test associated with pseudo-ML estimation is in turn defined as $T_{ML} := n F_{ML}(\hat{\theta})$, which corresponds to the Log-Likelihood Ratio test

statistic for testing the model $\sigma = \sigma(\theta)$, against the general alternative in which the Σ_g are unrestricted. Under normality, the statistics T_R , T_{GLS} and T_{ML} are asymptotically equivalent (i.e., they have the same asymptotic distribution as that stated above for T_R). Under non-normality of the z_g , T_{GLS} and T_{ML} are asymptotically equivalent, though not necessarily asymptotic chi-square. See Satorra (1989) for more details on the above goodness-of-fit test statistics, with discussion of their equivalence.

Under general distribution of the z_g , when the model holds exactly, T_{GLS} and T_{ML} are asymptotically distributed as a mixture of chi-square distributions of 1 degree of freedom (df) (see Satorra and Bentler, 1986); that is

$$T \xrightarrow{L} \sum_{j=1}^r \alpha_j \chi_j^2, \quad (14)$$

as $n \rightarrow \infty$, where T denotes either of the statistics T_{GLS} or T_{ML} . Here the χ_j^2 are independent chi-square variables of 1 df, the α_j are the non-null eigen-values of the matrix $U\Gamma$, with

$$U := V - V\Delta J^{-1}\Delta'V,$$

$J := \Delta'V\Delta$ ($= \sum_{g=1}^G f_g \Delta'_g V_g \Delta_g$) and $r := \text{rank}(\Delta'_\perp \Gamma \Delta_\perp)$. If the z_g are normally distributed, then the α_j are equal to 1 and thus one obtains an asymptotic exact chi-square distribution for the goodness-of-fit test statistics. This result motivated the scaling correction to the goodness-of-fit test statistic proposed by Satorra and Bentler (1994; Satorra and Bentler, 1988a,b), the specialization of which to T_{GLS} and T_{ML} is as follows. Let

$$\begin{aligned} c &:= \frac{1}{r} \text{tr} U\Gamma = \frac{1}{r} \text{tr} \left\{ V\Gamma - J^{-1}\Delta'V\Gamma V\Delta \right\} \quad (15) \\ &= \frac{1}{r} \sum_{g=1}^G \text{tr} \left\{ V_g \Gamma_g - f_g J^{-1} \Delta'_g V_g \Gamma_g V_g \Delta_g J^{-1} \right\} = \frac{1}{r} \text{tr} \left\{ \left(\Delta'_\perp V^{-1} \Delta_\perp \right)^{-1} \left(\Delta'_\perp \Gamma \Delta_\perp \right) \right\} \\ &= \frac{1}{r} \text{tr} \left\{ \left(\sum_{g=1}^G f_g^{-1} H'_g V_g^{-1} H_g \right)^{-1} \left(\sum_{g=1}^G f_g^{-1} H'_g \Gamma_g H_g \right) \right\}, \end{aligned}$$

where $\Delta'_\perp = [H'_1, \dots, H'_G]$ is partitioned conformably with the partition given above of V as a block-diagonal matrix. From the expression (10) of a

consistent estimator of Γ , a consistent estimator of c is

$$\hat{c} := \sum_{g=1}^G \sum_{i=1}^{n_g} \frac{n}{n_g(n_g - 1)} (d_{gi} - s_g)' \left\{ H_g \left(\sum_{g=1}^G \frac{n}{n_g} H_g' V_g^{-1} H_g \right)^{-1} H_g' \right\} (d_{gi} - s_g); \quad (16)$$

thus, the scaled (adjusted for mean) goodness-of-fit test statistic is

$$\bar{T} = T/\hat{c}, \quad (17)$$

where T denote either of the test statistics T_{GLS} or T_{ML} .

We now discuss the adjusted test statistic (adjusted for mean and variance) introduced in structural equation models by Satorra and Bentler (1994; Satorra and Bentler, 1988a,b); viz.,

$$\bar{\bar{T}} = \frac{d}{\text{tr}(U\Gamma)} T,$$

where d is the integer closest to

$$d' := \frac{(\text{tr } U\Gamma)^2}{\text{tr}(U\Gamma)^2},$$

and probability levels of $\bar{\bar{T}}$ computed using a chi-square distribution of d' degrees of freedom. The statistic $\bar{\bar{T}}$ is a Satterthwaite type test statistic (Satterthwaite, 1941). When tables from a chi-square distribution with fractional degrees of freedom are available, then d' is used instead of d .

For the sake of completeness, we also review the asymptotic variance matrix of estimators. In the current set-up, the estimator $\hat{\theta}$ is asymptotically normal with asymptotic variance matrix (e.g., Satorra and Neudecker, 1994)

$$\text{avar}(\hat{\vartheta}) = n^{-1} J^{-1} \Delta' V \Gamma V \Delta J^{-1}, \quad (18)$$

where $\Delta' V \Gamma V \Delta = \sum_{g=1}^G f_g \Delta_g' V_g \Gamma_g V_g \Delta_g$. Obviously, standard errors of parameter estimators are produced by replacing population matrices for consistent estimators in the above formulae, and replacing the f_g by the $\frac{n_g}{n}$. Clearly, when

$$\Delta' V \Gamma V \Delta = \Delta' V \Delta, \quad (19)$$

then $\text{avar}(\hat{\vartheta}) = \text{avar}^*(\hat{\theta}) := n^{-1} J^{-1}$. Note that a sufficient condition for (19) to hold is

$$V_g \Gamma_g V_g = V_g, \quad g = 1, \dots, G, \quad (20)$$

which is verified when the z_g are normally distributed.

3 Restricted tests

In this section we focus on testing a specific set of restrictions on the parameters of the model, when the alternative hypothesis is that the restrictions do not hold (rather than the general alternative that the Σ_g are unrestricted).

Let us parameterise the fitted model $\sigma = \sigma(\theta)$ as $\sigma = \sigma^*(\delta)$ with the constraint $a(\delta) = a_0$, where δ is a $(q+m)$ -dimensional vector of parameters, a_0 is an $m \times 1$ vector of constants, and $\sigma^*(\cdot)$ and $a(\cdot)$ are twice-continuously differentiable vector-valued functions of $\delta \in \Theta_1$, a compact subset of R^{q+m} . Our interest is in testing $H_0 : a(\delta) = a_0$ against $H_1 : a(\delta) \neq a_0$. In multi-sample analysis, a typical example arises when testing across-sample invariance of model parameters (see the example in the illustration below).

Define the Jacobian matrices

$$\Pi := (\partial/\partial\delta')\sigma^*(\delta) \quad \text{and} \quad A := (\partial/\partial\delta')a(\delta),$$

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

$$P := \Pi'V\Pi \left(= \sum_{g=1}^G f_g \Pi_g' V_g \Pi_g \right)$$

and

$$h' := (s - \hat{\sigma})' \hat{V} \hat{\Pi} \left(= \sum_{g=1}^G \frac{n_g}{n} (s_g - \hat{\sigma}_g)' \hat{V}_g \hat{\Pi}_g \right).$$

Generalised score and Wald-test statistics for restricted tests were investigated by Satorra (1989, Section 5); viz., the generalised score test statistic is

$$T_S := nh' \hat{P}^{-1} \hat{A}' \left(\hat{A} \hat{P}^{-1} \hat{\Pi}' \hat{V} \hat{\Gamma} \hat{V} \hat{\Pi} \hat{P}^{-1} \hat{A}' \right)^{-1} \hat{A} \hat{P}^{-1} h,$$

where \hat{A} , $\hat{\Pi}$ and \hat{P} are the matrices A , Π and P evaluated at $\hat{\theta}$, and $\hat{\Gamma}$ is the (distribution-free) consistent estimator of Γ defined in (10).

Clearly, when

$$\Pi'VTV\Pi = \Pi'V\Pi,$$

i.e., when

$$\sum_{g=1}^G \frac{n_g}{n} \Pi_g' V_g \Gamma_g V_g \Pi_g = \sum_{g=1}^G \frac{n_g}{n} \Pi_g' V_g \Pi_g,$$

then T_S is asymptotically equivalent to the (non-robust, asymptotically) score test statistic

$$T_S^* := nh' \hat{P}^{-1} \hat{A}' (\hat{A} \hat{P}^{-1} \hat{A}')^{-1} \hat{A} \hat{P}^{-1} h.$$

The statistic T_S^* is asymptotic chi-square under the condition that the estimation method is asymptotically optimal (Satorra, 1989). Specifically, when V has the normal-theory form given above and the z_{gi} are normally distributed, then T_S^* corresponds to the usual score test statistic. In the case of one-dimensional restrictions, T_S^* is the ‘‘Modification Index’’ of Jöreskog and Sörbom and Sörbom (1989). T_S^* is the ‘‘Lagrange Multiplier’’ test available in EQS (Bentler, 1995) for general multi-dimensional restrictions.

When the fitted model is the less restricted one, that is when the model analysed is $\sigma = \sigma^*(\delta)$, then the generalised Wald-test statistic for $H_0 : a(\delta) = a_0$ against $H_0 : a(\delta) \neq a_0$ is (see Satorra, 1989, Section 5)

$$T_W := n\hat{a}' \left(\hat{A} \hat{P}^{-1} \hat{\Pi}' \hat{V} \hat{\Gamma} \hat{V} \hat{\Pi} \hat{P}^{-1} \hat{A}' \right)^{-1} \hat{a},$$

where the matrices are now evaluated at the estimated value $\hat{\delta}$ under the unrestricted model and $\hat{a} := a(\hat{\delta})$. When $\Pi' V \Gamma V \Pi = J$, then T_W is asymptotically equivalent to the (non-robust, asymptotically) Wald-test statistic

$$T_W^* := n\hat{a}' \left(\hat{A} \hat{P}^{-1} \hat{A}' \right)^{-1} \hat{a}.$$

Under the assumption of normality of the z_{gi} , when we use ML or normal-theory GLS estimation methods then T_W^* is the typical Wald-test statistic.

When both the restricted and unrestricted models are analyzed using GLS or ML estimation methods, the restricted test of $H_0 : a(\delta) = a_0$ can be carried out using the difference-test statistic

$$DT := T_0 - T_1, \tag{21}$$

where T_0 and T_1 denote the goodness-of-fit test statistics (T_{GLS} or T_{ML}) associated with the restricted and unrestricted models respectively. Under an arbitrary distribution of the observable variables, the three test statistics T_S^* , T_W^* and DT are asymptotically equivalent (see Satorra, 1989, Theorem 4.1), though not necessarily chi-square distributed; when the estimation method is asymptotically optimal, then all the statistics T_S^* , T_W^* , DT , T_S and T_W are asymptotically equivalent, with the same asymptotic chi-square distribution (Satorra, 1989).

We will now obtain the expression of T_S and T_S^* for the particular case in which $a(\delta) = a_0$ corresponds to equating parameters to specific values; in this case, without loss of generality, we have $A = [0_{m \times q}, I_m]$ and $h' = [0_{1 \times q}, h'_2]$; consequently,

$$h' \hat{P}^{-1} \hat{A}' = h'_2 \hat{P}^{22} \quad \text{and} \quad (\hat{A} \hat{P}^{-1} \hat{A}')^{-1} = (\hat{P}^{22})^{-1} = \hat{P}_{22} - \hat{P}_{21} \hat{P}_{11}^{-1} \hat{P}_{12},$$

where we used

$$\hat{P} := \begin{pmatrix} \hat{P}_{11} & \hat{P}_{12} \\ \hat{P}_{21} & \hat{P}_{22} \end{pmatrix} \quad \text{and} \quad \hat{P}^{-1} := \begin{pmatrix} \hat{P}^{11} & \hat{P}^{12} \\ \hat{P}^{21} & \hat{P}^{22} \end{pmatrix}$$

(partitions corresponding to $h' = [0_{1 \times q}, h'_2]$), and standard formulae for the inverse of partitioned matrices (e.g., Magnus and Neudecker, 1999, p.11). Consequently,

$$\begin{aligned} T_S &= n h'_2 \hat{P}^{22} \left\{ \begin{pmatrix} \hat{P}^{21} & \hat{P}^{22} \end{pmatrix} \hat{\Pi}' V \Gamma V \hat{\Pi} \begin{pmatrix} \hat{P}^{12} \\ \hat{P}^{22} \end{pmatrix} \right\}^{-1} \hat{P}^{22} h_2 \\ &= n h'_2 \left\{ \begin{pmatrix} -\hat{P}_{21} \hat{P}_{11}^{-1} & I_m \end{pmatrix} \hat{\Pi}' V \Gamma V \hat{\Pi} \begin{pmatrix} -\hat{P}_{11}^{-1} \hat{P}_{12} \\ I_m \end{pmatrix} \right\}^{-1} h_2, \end{aligned}$$

where we used $\hat{P}^{21} = -\hat{P}^{22} \hat{P}_{21} \hat{P}_{11}^{-1}$ and $(\hat{P}^{21})' = \hat{P}^{12}$. Similarly, we can express T_S^* as

$$\begin{aligned} T_S^* &= n h'_2 \hat{P}^{22} (\hat{P}^{22})^{-1} \hat{P}^{22} h_2 \\ &= n h'_2 (\hat{P}_{22} - \hat{P}_{21} \hat{P}_{11}^{-1} \hat{P}_{12})^{-1} h_2. \end{aligned}$$

Note that h_2 corresponds to the components of h that are not equal to zero when evaluated at the restricted parameter value $\hat{\theta}$. Note also that \hat{P}_{11}^{-1} is the ‘‘information’’ matrix associated with the restricted model, and it does not change with the restrictions being tested.

The SB scaled versions of T_S^* , T_W^* and DT will thus be obtained by dividing the corresponding statistic by a consistent estimate of c in (15), where U is now given by (cf., Satorra 1989, p. 146)

$$U = V \Pi P^{-1} A' (A P^{-1} A')^{-1} A P^{-1} \Pi' V, \quad (22)$$

with m (the number of rows of A) replacing r ; that is, the scaled statistic is

$$\bar{T} := T/\hat{c},$$

where here T denotes any of the statistics T_S^* , T_W^* and DT , $\hat{c} = \frac{1}{m} \text{tr} \hat{U} \hat{\Gamma}$ and \hat{U} is the matrix U of (22) with the matrices Π , P and A evaluated at estimated values, and \hat{V} replaces V .

We will now obtain an explicit expression for c in the case of restricted tests and multi-samples. Clearly,

$$\begin{aligned} c &= \frac{1}{m} \text{tr} U \Gamma = \frac{1}{m} \text{tr} \left(AP^{-1} \Pi' V \Gamma V \Pi P^{-1} A' \right) \left(AP^{-1} A' \right)^{-1} \quad (23) \\ &= \frac{1}{m} \text{tr} (V \Gamma V) \left\{ \Pi P^{-1} A' (AP^{-1} A')^{-1} AP^{-1} \Pi' \right\} \\ &= \frac{1}{m} \sum_{g=1}^G f_g \text{tr} (V_g \Gamma_g V_g) \left\{ \Pi_g P^{-1} A' (AP^{-1} A')^{-1} AP^{-1} \Pi_g' \right\}, \end{aligned}$$

where $\Pi_g := (\partial/\partial\delta')\sigma_g^*(\delta)$. Note that this scaling factor does not necessarily coincide with the scaling factor associated with T_0 , nor with that associated with T_1 ; i.e., the SB scaled DT is not the difference of the SB scaled goodness-of-fit tests T_1 and T_0 .

We now show that when $m = 1$, then the scaled score statistic $\overline{T}_S (= T_S^*/\hat{c})$ coincides with the robust score statistic T_S . When $m = 1$, the scaling factor c defined above is

$$c = \frac{AP^{-1} \Pi' V \Gamma V \Pi P^{-1} A'}{AP^{-1} A'}, \quad (24)$$

and the robust and non-robust score statistics can be expressed as

$$T_S = n \frac{h' \hat{P}^{-1} \hat{A}' \hat{A} \hat{P}^{-1} h}{\hat{A} \hat{P}^{-1} \hat{\Pi}' \hat{V} \hat{\Gamma} \hat{V} \hat{\Pi} \hat{P}^{-1} \hat{A}'} \quad (25)$$

and

$$T_S^* = n \frac{h' \hat{P}^{-1} \hat{A}' \hat{A} \hat{P}^{-1} h}{\hat{A} \hat{P}^{-1} \hat{A}'}. \quad (26)$$

Consequently, we have

$$T_S = T_S^*/\hat{c}, \quad (27)$$

where

$$\hat{c} = \frac{\hat{A} \hat{P}^{-1} \hat{\Pi}' \hat{V} \hat{\Gamma} \hat{V} \hat{\Pi} \hat{P}^{-1} \hat{A}'}{\hat{A} \hat{P}^{-1} \hat{A}'}, \quad (28)$$

is a consistent estimator of the SB scaling factor c of (23). Moreover, if in addition to $m = 1$ we have that $a(\delta) = a_0$ restricts a parameter to a

specific value, then \hat{c} of (28) is the ratio of two standard errors of the restricted parameter: the asymptotic robust standard error (computed from (18), unrestricted model), and the NT standard error (computed from J^{-1} , unrestricted model).

We now provide an explicit expression for c when $A = [0_{m \times q}, I_m]$. In this case we have

$$\begin{aligned} c &= \frac{1}{m} \text{tr} V \Gamma V \Pi \left\{ P^{-1} A' (A P^{-1} A')^{-1} A P^{-1} \right\} \Pi' \\ &= \frac{1}{m} \text{tr} V \Gamma V \Pi \left\{ \begin{pmatrix} \hat{P}^{21} & \hat{P}^{22} \end{pmatrix} (\hat{P}^{22})^{-1} \begin{pmatrix} \hat{P}^{12} \\ \hat{P}^{22} \end{pmatrix} \right\} \Pi' = \frac{1}{m} \text{tr} V \Gamma V \Pi M \Pi', \end{aligned}$$

where

$$M := \begin{pmatrix} P^{12} (P^{22})^{-1} P^{21} & P^{12} \\ P^{21} & P^{22} \end{pmatrix} = \begin{pmatrix} P_{11}^{-1} P_{12} P^{22} P_{21} P_{11}^{-1} & -P_{11}^{-1} P_{12} P^{22} \\ -P^{22} P_{21} P_{11}^{-1} & P^{22} \end{pmatrix}.$$

Finally, the specialization of the SB adjusted (for mean and variance) goodness-of-fit test to T_S^* , T_W^* and DT test statistics is as follows. We define

$$\overline{\overline{T}} := (d / \text{tr} \hat{U} \hat{\Gamma}) T,$$

where T denotes any of the statistics T_S^* , T_W^* or DT , and d is the integer closest to

$$d' := \frac{(\text{tr} \hat{U} \hat{\Gamma})^2}{\text{tr} (\hat{U} \hat{\Gamma})^2},$$

with \hat{U} being the matrix of (22) evaluated at estimated values, and $\hat{\Gamma}$ given in (10). The null distribution of the adjusted test statistic $\overline{\overline{T}}$ is taken to be a chi-square distribution with d degrees of freedom. Here d' also replaces d when tables of a chi-square distribution with fractional degrees of freedom are available.

4 Illustration

To illustrate the performance of the above statistics in finite samples, a specific model context is used to simulate the competing score-test statistics introduced in the previous section for various sample sizes. The performance

of the test statistics is evaluated by comparing Monte Carlo rejection frequencies with the corresponding nominal values.

We use a very simple model context. We consider a regression equation

$$y_{gi}^* = \beta x_{gi} + v_{gi}, \quad i = 1, \dots, n_g, \quad (29)$$

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

$$x_{1gi}^* = x_{gi} + u_{1gi}, \quad x_{2gi}^* = x_{gi} + u_{2gi}, \quad (30)$$

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 (29) and (30) 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 (29) and (30) as

$$z_{gi} = \Lambda \xi_{gi}, \quad i = 1, 2, \dots, n, \quad (31)$$

where

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

and

$$\Lambda := \begin{pmatrix} \beta & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \quad (32)$$

Define

$$\Phi := E \xi_{gi} \xi_{gi}' = \begin{pmatrix} \sigma_{xx} & 0 & 0 & 0 \\ 0 & \sigma_{vv} & 0 & 0 \\ 0 & 0 & \sigma_{uu} & 0 \\ 0 & 0 & 0 & \sigma_{uu} \end{pmatrix}, \quad (33)$$

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

$$\Sigma := \Lambda \Phi \Lambda' = \Sigma(\theta), \quad (34)$$

where $\Sigma(\cdot)$, $\Lambda(\cdot)$ and $\Phi(\cdot)$ are (twice-continuously differentiable) matrix-valued functions of θ , as deduced from (32), (33) and (34). Note that the model restricts the variances of u_1 and u_2 by equality.

The simulation of two-sample data from the above model and computation of score test statistics was replicated 1000 times, for various combinations of two-sample sizes as shown in Table 1. 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 model imposed across-group invariance of model parameters. In each replication, we computed the various versions of the score-test statistic of Section 3, for the null hypothesis of across-group invariance of all model parameters. Clearly, in our Monte Carlo set-up, the null hypothesis holds true, with the asymptotic distribution of T_S 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 score-test statistic.

We note that the normal-theory chi-square goodness-of-fit T_1 (T_{GLS} or T_{ML}) 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-sample; cf., Satorra, 1993); 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, 1993). Consequently, since DT ($= T_1 - T_0$) is asymptotically equal to T_S^* (Satorra, 1989), the scaling factor c for T_S^* does not need to coincide with the scaling factor c associated with T_1 (which is 1, as deduced from the mentioned theory of asymptotic robustness), nor with the scaling factor c associated to T_0 .

As shown in Table 1, in our specific model context, in the case of small

samples, the SB scaled statistic, $\overline{T_S^*}$, seems to outperform the alternative robust test statistic T_S . As expected from theory, in the case of large samples, T_S outperforms the alternative test statistics. The adjusted test statistic, $\overline{\overline{T_S^*}}$, shows acceptable performance in various sample sizes. Clearly, the normal-theory score-test statistic T_S^* performs very poorly for all sample sizes considered in this Monte Carlo study.

References

- [1] Bentler, P.M., (1995), *EQS Structural Equations Program Manual*, Multivariate Software Inc., Encino, CA.
- [2] Bentler, P. M., Lee, S.-Y., and Weng, J., (1987), Multiple population covariance structure analysis under arbitrary distribution theory, *Communications in Statistics - Theory*, 16, 1951-1964
- [3] Bollen, K.A., (1989), *Structural Equations with Latent Variables*, John Wiley, New York.
- [4] Browne, M.W., (1974), Generalized Least Squares Estimators in the Analysis of Covariance Structures, *South African Statistical Journal*, 8, 1-24.
- [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] Chou, C.-P., P.M. Bentler and A. Satorra, (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.
- [7] 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.
- [8] Fuller, W.A., (1987), *Measurement Error Models*, John Wiley, New York.

Table 1: Monte Carlo results

nominal significance levels:	1%	5%	10%	20%
$n_1 = 50$ and $n_2 = 60$				
T_S	2.20	9.90	16.60	28.30
$\overline{T_S^*}$	1.00	5.60	11.00	24.10
$\overline{\overline{T_S^*}}$	0.80	3.00	7.40	18.90
T_S^*	14.60	33.40	45.20	59.80
$n_1 = 100$ and $n_2 = 120$				
T_S	1.40	7.60	14.90	24.70
$\overline{T_S^*}$	1.40	5.40	11.00	20.70
$\overline{\overline{T_S^*}}$	0.60	3.00	6.30	18.10
T_S^*	22.00	40.20	49.20	60.60
$n_1 = 200$ and $n_2 = 240$				
T_S	1.00	4.20	10.10	22.40
$\overline{T_S^*}$	1.60	6.60	11.30	20.40
$\overline{\overline{T_S^*}}$	0.90	3.40	8.50	18.40
T_S^*	25.70	41.50	49.90	60.30
$n_1 = 800$ and $n_2 = 600$				
T_S	0.80	5.40	10.00	20.30
$\overline{T_S^*}$	2.60	7.70	11.60	20.80
$\overline{\overline{T_S^*}}$	1.10	4.50	9.50	17.20
T_S^*	29.00	45.90	54.50	66.70

- [9] Hu, L., Bentler, P. M., and Kano, Y. (1992), Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112, 351–362.
- [10] Jöreskog, K., and D. Sörbom, (1994), *8 User's Reference Guide*, Scientific Software Inc., Mooresville, In
- [11] Magnus J. and H. Neudecker, (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley, New York.
- [12] Moore, D.S., (1977), Generalized Inverses, Wald's Method, and the Construction of Chi-squared Tests of Fit, *Journal of the American Statistical Association*, 72, 131–137.
- [13] Muthén, B., (1989), Multiple Group Structural Modelling with Non-normal Continuous Variables, *British Journal of Mathematical and Statistical Psychology*, 42, 55–62.
- [14] Muthén, B., (1993), Goodness of Fit Test with Categorical and Other Non-normal Variables, in *Testing Structural Equation Models*, K.A. Bollen and J. Scott Long (eds.), Sage Publications Inc., Newbury Park.
- [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., and P.M. Bentler, (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, American Statistical Association, Alexandria (Virginia).
- [18] Satorra, A., and P.M. Bentler, (1988a), Scaling Corrections for Chi-square Statistics in Covariance Structure Analysis, *ASA 1988 Proceedings of the Business and Economic Statistics Section*, 308–313, American Statistical Association: Alexandria (Virginia).

- [19] Satorra, A., and P.M. Bentler, (1988b), Scaling Corrections for Statistics in Covariance Structure Analysis, *UCLA Statistics Series # 2*, University of California, Los Angeles.
- [20] Satorra, A., and P.M. Bentler, (1991), Goodness-of-Fit Test under IV Estimation; Asymptotic Robustness of a NT Test Statistic. In Gutierrez, R. and Valderrama, M.J. (eds.), *Applied Stochastic Models and Data Analysis*, World Scientific, London.
- [21] Satorra, A., and Bentler, P. M. (1994), Corrections to test statistics and standard errors in covariance structure analysis. In von Eye, A. and Clogg, C.C. (eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419), Sage, Thousand Oaks, CA.
- [22] Satorra, A., and H. Neudecker, (1994), On the Asymptotic Optimality of Alternative Minimum-Distance Estimators in Linear Latent-Variable Models, *Econometric Theory*, 10, 867–883.
- [23] Satterthwaite, F.E., (1941), Synthesis of Variance, *Psychometrika*, 6, 309–316.
- [24] Sörbom, D., (1989), Model Modification, *Psychometrika*, 54, 371-384.
- [25] Yuan, K.-H., and P.M. Bentler, (1997), Mean and Covariance Structure Analysis: Theoretical and Practical Improvements, *Journal of The American Statistical Association*, 92, 767–774.
- [26] 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.