Published in final edited form as:

Multivariate Behav Res. 2012 January 1; 47(3): 448–462. doi:10.1080/00273171.2012.673948.

A Third Moment Adjusted Test Statistic for Small Sample Factor Analysis

Johnny Lin and Peter M. Bentler

University of California, Los Angeles

Abstract

Goodness of fit testing in factor analysis is based on the assumption that the test statistic is asymptotically chi-square; but this property may not hold in small samples even when the factors and errors are normally distributed in the population. Robust methods such as Browne's asymptotically distribution-free method and Satorra Bentler's mean scaling statistic were developed under the presumption of non-normality in the factors and errors. This paper finds new application to the case where factors and errors are normally distributed in the population but the skewness of the obtained test statistic is still high due to sampling error in the observed indicators. An extension of Satorra Bentler's statistic is proposed that not only scales the mean but also adjusts the degrees of freedom based on the skewness of the obtained test statistic in order to improve its robustness under small samples. A simple simulation study shows that this third moment adjusted statistic asymptotically performs on par with previously proposed methods, and at a very small sample size offers superior Type I error rates under a properly specified model. Data from Mardia, Kent and Bibby's study of students tested for their ability in five content areas that were either open or closed book were used to illustrate the real-world performance of this statistic.

Introduction

Maximum likelihood techniques used in confirmatory factor analysis rely on the assumption of multivariate normality and asymptotic theory. The estimation of parameters involves minimizing the difference between the sample covariance matrix S and the implied covariance matrix $\Sigma(\theta)$ using the fit function (Jöreskog, 1967, 1970)

$$F_{ML} = \log |\Sigma(\theta)| + \text{trace}(S \Sigma^{-1}(\theta)) - \log |S| - p \quad (1)$$

where p is the number of observed variables. The derivation of F_{ML} assumes that the vector of observed indicators y is independently sampled from a multivariate normal distribution (Bollen, 1989). Since the observed indicators are a function of parameters in the factor analytic model, non-normality in observed indicators is an implied consequence of non-normality in the distribution of factors and errors. Violations of normality due to excess skewness and kurtosis can still lead to consistent parameter estimates, but the standard errors and goodness of fit chi-square statistics are compromised (Browne, 1984; Bollen, 1989).

The confirmatory factor model begins with y , a vector of observed indicators that depends on Λ , a common factor loading matrix, η , a vector of latent factor scores, and ε , a vector of unique errors. Their linear relationship is defined as

$$y = \Lambda \eta + \varepsilon. \quad (2)$$

Assuming η is normally distributed and uncorrelated with ε the restricted model covariance of y is

$$\Sigma(\theta) = \Lambda\Phi\Lambda' + \Psi \quad (3)$$

where Ψ is a diagonal matrix of unique variances and Φ is a matrix of covariances among latent factor scores.

When the maximum likelihood fit function (1) is used to estimate the vector θ of parameters of the set $\{\Lambda, \Phi, \Psi\}$, the goodness of fit chi-square estimator is given as

$$T \equiv n\hat{F}_{ML} \quad (4)$$

where n is the sample size and $\hat{F}_{ML}$ is the minimum fit function (1) evaluated at the maximum likelihood estimate $\hat{\theta}$. This is -2 times the logarithm of the Wishart likelihood ratio criterion for testing the null hypothesis of (3) against the alternative hypothesis of unrestricted Σ (Amemiya & Anderson, 1990). If the latent factor scores η and unique errors ε are normally distributed, then T converges in law to a $\chi^2(d)$ distribution with degrees of freedom

$$d = \frac{1}{2}(p)(p+1) - t \quad (5)$$

where p is the number of observed variables and t is the number of free parameters in θ . Conceptually, (5) represents the number of non-redundant elements in the covariance matrix S minus the number of free parameters. This χ^2 approximation is based on the assumption that a) y has no kurtosis, b) the covariance rather than the correlation matrix is analyzed, c) the sample is sufficiently large, and d) $H_0: \Sigma = \Sigma(\theta)$ holds exactly (Bollen, 1989). Under violation of the first condition, there is an alternative procedure known as the asymptotically distribution-free (ADF) method (Browne, 1984). Empirically, this method has been shown to require large sample sizes to yield an accurate assessment of model adequacy since it requires the inversion of a sample fourth order moment based matrix that converges slowly to its population counterpart (Hu, Bentler, & Kano, 1992; Curran, West, & Finch, 1996). The current study utilized and extends a class of chi-square correction procedures known as mean scaling and degrees of freedom adjustment methods (Satorra & Bentler, 1988, 1994) that have been shown to perform better than the ADF method under small sample sizes.

Define $s = (s_{11}, s_{21}, \dots, s_{kk})'$ as the vector of elements in the lower triangle of the sample covariance matrix S , and $\sigma(\theta) = (\sigma_{11}, \sigma_{21}, \dots, \sigma_{kk})'$ the elements of the model restricted covariance matrix. Browne (1974, 1984) showed that minimizing the maximum likelihood fit function described in (1) is asymptotically equivalent to minimizing the quadratic form

$$F(\theta, V) = [s - \sigma(\theta)]' V [s - \sigma(\theta)]. \quad (6)$$

Define $\hat{V}$ to be the consistent estimate of V as

$$\hat{V} = 2^{-1} D' (A_n^{-1} \otimes A_n^{-1}) D \quad (7)$$

where A_n is a matrix that converges to Σ with probability one (e.g., $\widehat{\sum}(\theta)$), and D is the 0–1 duplication matrix that eliminates the supra-diagonal elements of A_n (Magnus & Neudecker, 1986). Assume that $\sqrt{n}(s-\sigma)$ converges in distribution to $N(0, \Gamma)$ as $n \rightarrow \infty$, where Γ is the asymptotic covariance matrix of s . Typically, Γ is estimated via the covariance matrix (Browne, 1984; Satorra & Bentler, 1994):

$$\widehat{\Gamma} = \frac{1}{n-1} \sum_{i=1}^n (b_i - \bar{b})(b_i - \bar{b})' \quad (8)$$

where $b_i = D^+ \text{vec}(y_i - \bar{y})(y_i - \bar{y})'$, D^+ is the Moore-Penrose inverse of the duplication matrix of D described in (7), $\bar{b}$ and $\bar{y}$ are the means of b and y , and vec is the operator that vectorizes a matrix column-wise. Let the Jacobian matrix be $\Delta \equiv \frac{\partial \sigma}{\partial \theta'}$, which is estimated in the sample by $\widehat{\Delta}$, and define

$$\widehat{U} = \widehat{V} - \widehat{V} \widehat{\Delta} (\widehat{\Delta}' \widehat{V} \widehat{\Delta})^{-1} \widehat{\Delta}' \widehat{V}. \quad (9)$$

Asymptotically,

$$T \rightarrow n(s - \sigma(\theta))' U (s - \sigma(\theta)) \quad (10)$$

where the matrix U is the asymptotic true form of $\widehat{U}$. Using Theorem 2.1 of Box (1954)

$$T \xrightarrow{L} \sum_{i=1}^d \lambda_i \chi_i^2(1) \quad (11)$$

where λ_i 's are the positive eigenvalues of the matrix $U\Gamma$. Typically population values are not accessible, and we estimate T using $\widehat{T}$.

Mean Scaling and Moment Adjustment Methods

The Satorra Bentler (1994) scaled chi-square statistic, here called the Mean Scaled statistic, is

$$T_M \equiv \frac{T}{m} \quad (12)$$

where $m \equiv \text{trace}(U\Gamma)/d$. Note that for brevity, the robust statistics are described in its population form, but can be implemented by replacing $\widehat{U}\widehat{\Gamma}$ for $U\Gamma$.

Satorra and Bentler (1994) also proposed an extension of T_M which involves both scaling the mean and a Saitterwaithe second moment adjustment of the degrees of freedom (Satterthwaite, 1941). This statistic is called the Mean Scaled and Variance Adjusted statistic, and is defined as

$$T_{MV} \equiv \frac{v}{\text{trace}(U\Gamma)} T \quad (13)$$

where $v \equiv [\text{trace}(U\Gamma)]^2 / \text{trace}[(U\Gamma)^2]$.

Recall in Equation 11 that T asymptotically approaches a quadratic form which is a weighted sum of independent chi-square variates. Based on the cumulant generating function of a central chi-square quadratic form

$$c_k = \sum_{i=1}^d \lambda_i^k \quad (14)$$

where k is an index of its moment; the mean, standard deviation and skewness of T are $\mu_T = c_1$, $\sigma_T = 2c_2$, $\beta_T = \sqrt{8}(c_3/c_2^{3/2})$ respectively (Liu, Tang, & Zhang, 2009). In practice, sample estimates $\mu_{\hat{T}}$, $\sigma_{\hat{T}}$, and $\beta_{\hat{T}}$ are used.

The proposed statistic, to be called the Mean Scaled and Skewness Adjusted statistic, is defined as

$$T_{MS} \equiv \frac{v^*}{\text{trace}(UT)} T \quad (15)$$

with estimated degrees of freedom

$$v^* \equiv \frac{c_2^3}{c_3^2} = \frac{(\sum_{i=1}^d \lambda_i^2)^3}{(\sum_{i=1}^d \lambda_i^3)^2} = \frac{\text{trace}[(UT)^2]^3}{\text{trace}[(UT)^3]^2} \quad (16)$$

where v^* is a function of the skewness of T . Motivated by the Pearson-Imhof three-moment central χ^2 approximation method (Pearson, 1959; Imhof, 1961), T_{MS} scales the mean as does T_M and T_{MV} but also adjusts the degrees of freedom such that the asymptotic quadratic form of T has the same skewness as a candidate $\chi^2(v^*)$, the new reference distribution. The goal of modifying the degrees of freedom in addition to scaling the mean is to downwardly adjust the obtained statistic such that its distribution is as close to a central chi-square as possible, despite a potentially large disparity in the population eigenvalues of the matrix UT .

Theoretical Conditions for Approaching χ^2

Consider the case in the population where the eigenvalues of UT are constant, notably if $\lambda_i = \lambda$ for $i = 1, \dots, d$. Then Equation 16 simplifies to

$$v^* = \frac{(\sum_{i=1}^d \lambda_i^2)^3}{(\sum_{i=1}^d \lambda_i^3)^2} = \frac{(\sum_{i=1}^d \lambda^2)^3}{(\sum_{i=1}^d \lambda^3)^2} = \frac{(d\lambda^2)^3}{(d\lambda^3)^2} = \frac{d^3 \lambda^6}{d^2 \lambda^6} = \frac{d^3}{d^2} = d. \quad (17)$$

Substituting d into Equation 15,

$$T_{MS} = \frac{v^*}{\text{trace}(UT)} T = \frac{d}{\text{trace}(UT)} T = T_M \quad (18)$$

which makes T_{MS} equivalent to T_M . In the special case where $\lambda_i = 1$ for $i = 1, \dots, d$, the

asymptotic distribution of T would approach χ^2 since $\sum_{i=1}^d \lambda \chi_i^2(1) = \chi^2(d)$. Then by modification of Equation 18, T_{MS} would be equivalent to T . When λ_i is not constant, the distribution of T cannot be determined, and the assumption is that scaling the mean and

adjusting the degrees of freedom of T will result in a better approximation of a chi-square variate.

The Effect of Sample Size on the Obtained Test Statistic

The previous sections introduced well-known robust methods and the proposed third moment adjusted statistic based on the assumption of asymptotic theory with given population parameters. In practice however, large sample sizes and population parameters may not be available to the researcher, which means that the properties of these robust methods may not hold. As stated in the introduction, these methods were developed to address the issue of population non-normality in factors and errors and do not directly address the issue of non-normality due to sampling error. Consider the situation in which the population factors and errors are normal but because of small sample size, the eigenvalues of the matrix $\hat{U}\hat{T}$ are not all one and hence the obtained test statistic is not chi-square distributed. The latter half of the paper marks a diversion away from asymptotic theory to show the empirical relation between sample size and the test statistics at hand.

Figure 1 illustrates the effect of small, medium, and large sample size on the distribution of the eigenvalues of $\hat{U}\hat{T}$ from a simulated run of a case where factors and errors were normally distributed. Despite the theoretical equality of eigenvalues in the population, the figure shows that the *estimated* eigenvalues under $n = 25$ are markedly more dispersed compared to other sample size conditions, with eigenvalues ranging from 0 to 12 and positive values dropping off at $i = 25$. This increased dispersion of sample eigenvalues compared to population counterparts has been previously noted for cases when the ratio of estimated parameters to sample size is large (Ledoit & Wolf, 2004). As shown in the figure, the dispersion decreases with moderate sample size ($n = 100$), with eigenvalues now ranging from 0 to about 3.5. Asymptotically ($n = 5000$), the positive eigenvalues stabilize at a constant point near one, and drop off to zero at $i = 54$, the expected degrees of freedom. The increasing dispersion of eigenvalues with decreasing sample size highlights the rationale of using mean scaling and degrees of freedom adjustments for small sample factor analysis.

To demonstrate the relation of sample size to the skewness of the obtained statistic, calculate the skewness using the formula

$$\beta_{\hat{T}} = \sqrt{8(\hat{c}_3^2 / \hat{c}_2^3)} = \sqrt{8/\nu^*}. \quad (19)$$

For a sample replicate of $n = 25$, the skewness of $\hat{T}$ was calculated to be 1.01. Given $n = 100$, the skewness decreased to 0.70, and for $n = 5000$ it was 0.39, approaching the optimal skewness of 0.38 given that the degrees of freedom of the reference chi-square is 54. Since both the sample size and the estimated degrees of freedom are inversely proportional the skewness of the obtained test statistic $\hat{T}$, it makes intuitive sense to adjust the estimated degrees of freedom downward when the skewness of the obtained statistic may be higher due to smaller sample size. In the empirical example to follow (Table 2), notice that the degrees of freedom for T_{MS} are consistently lower than that of other statistics.

Monte Carlo Simulation

The current study evaluates the performance of two existing robust statistics against the proposed Mean Scaled and Skewness Adjusted statistic, T_{MS} . Yuan and Bentler (2010) evaluated the Type I and mean-square error of T_M and T_{MV} under different coefficients of variation in the eigenvalues of UT , and found that T_{MV} performs better than T_M when the disparity of eigenvalues is large. Based on this finding, the current hypothesis proposes that asymptotically, the performance of T_{MS} would not differ from T_M and T_{MV} given that the

factors and errors are normally distributed in the population. Due to the inverse relationship between sample size and skewness of the obtained test statistic, we hypothesize that a third moment based adjustment of the obtained statistic would be better than a second moment adjustment or mean scaling alone and that the performance of T_{MS} would improve as sample size decreases. As such, the test statistics T , T_M , T_{MV} and T_{MS} were evaluated for their Type I error and empirical power under asymptotic sample size ($n = 5000$), moderate sample size ($n = 100$), and very small sample size ($n = 25$).

Confirmatory Factor Model

A confirmatory factor model with 12 observed variables and three orthogonal factors was used to generate a model based simulation study in R 2.13.0 for Windows (R Development Core Team, 2011). Factors and errors were generated independently and distributed $N(0, 1)$. A simple structure of Λ was used where each set of four observed variables loaded onto a single factor with loadings of 0.80, as shown in (20). Random variates for factor and errors were generated and entered into Equation 3, with $\Phi = I$. Model 1 is defined as the data generation model with $\lambda_{51} = 0$, shown as the loading to the left of the slash in (20). Model 2 is defined as the data generation model in which $\lambda_{51} = 0.8$, shown to the right of the slash. Each analysis was replicated 100 times, since at $n = 25$, convergence issues made it difficult to perform more replications. The simulated datasets were analyzed using the sem package in R (Fox, 2006) and the mean and scaling statistics obtained by modifying source code from sem.additions (Byrnes, 2010) and CompQuadForm (Duchesne & Lafaye De Micheaux, 2010). The parametrization of sem.additions involved specifying useFit=T so that $A_n = \sum(\theta)$ in Equation 7, and useGinv=T so that a generalized inverse is used in place of the inverse of $\hat{\Delta}' \hat{V} \hat{\Delta}$ (see Equation 9). These parametrizations were chosen so that the computed values of T , T_M and T_{MV} match those of EQS 6.2 (Bentler, 2006).

$$\Lambda' = \begin{bmatrix} 0.8 & 0.8 & 0.8 & 0.8 & 0/0.8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.8 & 0.8 & 0.8 & 0.8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0.8 & 0.8 & 0.8 \end{bmatrix} \quad (20)$$

Type I Error and Empirical Power

Type I error performance was measured by dividing the observed rejection rate of the null hypothesis by the total number of trials ($N = 100$) under the correctly specified model (Model 1 generated, Model 1 analyzed). Setting the probability of rejection criterion at $\alpha = 0.05$,

$$1 - P(T \leq \chi^2(d) | \sum(\theta) = \sum_0) = \alpha \quad (21)$$

where an ideal Type I error rate should approach 5% rejection of the null hypothesis. Under an incorrectly specified model $\Sigma(\theta) \neq \Sigma_0$ (Model 2 generated, Model 1 analyzed), the empirical power is defined as the proportion of rejection of the null hypothesis for $N = 100$ simulated trials.

Simulation Results

Table 1 shows the results of $N = 100$ replications of normal factors and errors under asymptotic sample size ($n = 5000$), moderate sample size ($n = 100$), and a very small sample size ($n = 25$) across $\alpha = \{0.01, 0.05, 0.1\}$. As expected, at $n = 5000$ and under correct model specification (Model 1 generated, Model 1 analyzed), the Type I error performance (left

columns) was nearly identical across test statistics, with rates ranging from 0.02 to 0.04 at $\alpha = 0.05$. Under incorrect model specification (Model 1 generated and Model 2 analyzed), empirical power (right column) was excellent, as every statistic rejected the model 100% of the time across alpha levels. The results give evidence that asymptotically, T_{MS} performs on par with the other statistics in its Type I error and empirical power.

The advantage of T_{MS} is more evident at a very small sample size ($n = 25$). Under correct model specification, T_M had the highest Type I error rates at $\alpha = 0.05$, rejecting the null hypothesis 74% of the time, followed by T at 45%, T_{MV} at 19%, and finally T_{MS} , which offered the most optimal rejection rate at 4%. A similar pattern emerged for $\alpha = 0.01$, except that T_{MV} performed on par with T_{MS} . At $\alpha = 0.1$, all statistics had higher than expected rejection rates, but T_{MS} still offered the most optimal rate at 25% rejection compared to 82% for T_M . Figure 2 shows the density of p -values for the test statistics considered and a simulated chi-square variate with 54 degrees of freedom under correct model specification at the region of $p < 0.05$. Compared to the reference chi-square probability density, T_M was overly peaked (rejecting the correctly specified model too frequently at $n = 25$), whereas T_{MS} was closest to the reference chi-square probability density. Under incorrect model specification, the lower right part of Table 1 shows that T_M had the most correct rejections of the null hypothesis at 90% and T_{MS} the least at 27% for $\alpha = 0.05$. Similar patterns emerged for $\alpha = 0.01$ and $\alpha = 0.1$. The numbers suggest that the empirical power of T , T_M and T_{MV} under a very small sample exceeded that of T_{MS} when the model is misspecified.

To test the threshold at which the Type I error rate and power of T_{MS} converges with the other statistics, a third sample size condition at $n = 100$ was added. The results show that at this moderate sample size, the Type I performance of T_{MV} begins to approach T_{MS} (and in fact may have more optimal performance at $\alpha = 0.10$), while T and T_M continue to have inflated Type I rates. The underpowered phenomenon of T_{MS} from the much smaller sample size condition ($n = 25$) disappears as an empirical power of 0.99 is achieved with $\alpha = 0.05$. The results suggest that across a range of sample sizes T_{MS} may be a good choice for achieving optimal Type I error rates, although T_{MV} may be better for moderate sample sizes.

Illustrative Application of T_{MS}

Analysis of the Open-book Closed-book Dataset

To illustrate the real-world performance of these statistics, Mardia, Kent, and Bibby's (1980) Open-book Closed-book dataset was chosen. The dataset can be freely downloaded from the Comprehensive R Archive Network (CRAN) as the `scor` component in the bootstrap R package (Tibshirani & Leisch, 2009; Efron & Tibshirani, 1993). A total of $n = 88$ students were tested for their ability in five content areas: mechanics (closed book), vectors (closed book), algebra (open book), analysis (open book), and statistics (open book). Following Cai and Lee (2009), the first model was analyzed using a two correlated factor structure, in which the two closed book skills grouped together, and the four remaining open book skills grouped together to create a model with 11 free parameters and four degrees of freedom. As with the simulation study, the data were analyzed using a source code modification of the `sem` R package and `sem.additions` extension, and the results of T , T_M , T_{MV} were verified with EQS 6.2 software using the `METHOD = ML, ROBUST` specification.

The first model fit the data very well with an estimated $RMSEA < 0.01$, $CFI = 1.00$, and $\chi^2(4) = 2.07$, $p = 0.72$. All test statistics showed that the two correlated factor model was retained at $p > 0.5$. The difference between the degrees of freedom adjusted statistics and the others is evident in the lower estimated degrees of freedom (see Table 2 top left column). The second model fit the same factor loading pattern as the first model except that the

factors were uncorrelated. This model did not fit the data well with an estimated RMSEA = 0.324, CFI = 0.76, and $\chi^2(5) = 50.53$, $p < 0.01$. Similar patterns of model rejection were seen for the uncorrelated factor model across all statistics, except that T_{MS} again resulted in the lowest mean scaling and degrees of freedom after adjustment (see Table 2, top right column). The results show that under a typical data analytic setting, all test statistics perform comparably well.

Analysis of a Sub-sample of the Open-book Closed-book Dataset

The second analysis demonstrates the test statistics at a much smaller sample size. A total of 15 students were randomly selected from the original Open-Book Closed-Book dataset, with the particular order of student cases chosen as: 2, 3, 5, 15, 21, 26, 33, 57, 58, 68, 74, 75, 76, 84, and 86. Table 2 (lower left column) shows that at $\alpha = 0.05$, all statistics except for T_M retained the correlated factor model, whereas all statistics rejected the uncorrelated factor model at $p < 0.01$ (Table 2 lower right column). The results are consistent with those from the simulation study in that T_{MS} was shown to be relatively more conservative in retaining the specified model than T_M at a very small sample size, although the differences between them were not as definitive as in the simulation study. The authors suspect that the efficacy

of T_{MS} may depend on the magnitude of the ratio $\sqrt{d/v^*}$, which is $\sqrt{4/2.17} = 1.35$ in this empirical example, but was greater at $\sqrt{54/7.84} = 2.62$ in the simulation study.

Conclusion and Discussion

The current paper proposed a new statistic that scales the mean and adjusts the degrees of freedom of the obtained test statistic to match the skewness of the reference chi-square variate. The rationale is that this adjustment will allow the obtained statistic to approach a central chi-square variate, even under small sample sizes, so that it is possible to obtain p -values by looking up a standard chi-square table. A Monte Carlo simulation study assessed the performance of the proposed statistic compared with existing methods across a range of sample sizes, and an example using the Open-book Closed-book dataset was provided.

The simulation study gives evidence that T_{MS} performs on par asymptotically with existing methods, and under a very small sample size ($n = 25$) can offer superior Type I error rates by as much as 1360% when compared to T_M at $\alpha = 0.05$, and by 260% compared with T_{MV} . However, the optimum Type I error rate of T_{MS} under a correctly specified model comes with the cost of being underpowered for a misspecified model at a small sample size. The T_M procedure that currently exists in structural equation modeling programs such as EQS (Bentler, 2006), Mplus (Muthen & Muthen, 2006) and LISREL (Jöreskog, Sorbom, du Toit, & du Toit, 2000) may be a better choice, although the current study suggests that the power of T_{MS} approaches T_M when sample size reaches $n = 100$. In sum, it may be beneficial to use the more conservative T_{MS} when confirming the fit of the model under a very small sample size, but to use the more powerful T_M as a criterion for rejecting an incorrect model. If the researcher finds that the model is rejected by other robust methods but retained by T_{MS} , it may be a sign that the model is correctly specified, and it may be worth the extra effort to continue recruiting subjects.

The proposed statistic and the methodological design of the paper is not without limitations. First, since the population factors and errors were normally distributed, the theoretical distribution of the test statistic should be exactly chi-square. This is not the situation for which robust statistics were developed, as non-normality of factors and errors is typically assumed. Given the focus of the paper, the case where population eigenvalues are not equal to one was not empirically assessed, although it is reasonable to assume that the current theory would still hold. Second, rather than perform mean scaling and degrees of freedom

adjustments to allow the obtained statistic to approach a central chi-square variate, it may be possible to test the probability of a mixture chi-square directly (Bentler, 1994; Lo, Mendell, & Rubin, 2001). The drawback of the direct approach is that under small samples, the obtained test statistic may not actually be distributed mixture chi-square, which would lead to incorrect testing conclusions. Third, the current study considered the robustness of T_{MS} only for continuous data. Lei (2009) showed that for ordinal variables, T_M may reject too frequently at large sample sizes compared to the mean scaled and variance adjusted procedure, and that contrary to our findings for moderately large samples T_M was less powerful. However, the conclusions do not map one-to-one to our current paper, since the variance adjusted procedure was estimated using weighted least squares whereas T_M was estimated using maximum likelihood. In addition, the smallest sample size considered in that paper was $n = 100$, which is larger than the $n = 25$ considered in this paper. Finally, the current paper developed its new test statistic under the assumption of complete data with no mean structure. It is possible to extend the current statistic to the case of incomplete data with mean structure using the work of Yuan and Bentler (2000) by adapting their analogous matrices of U to our methodology.

Acknowledgments

We thank the reviewers of this paper, Yutaka Kano, and Jianmin Wu for their invaluable suggestions. Part of this research is made possible by a pre-doctoral advanced quantitative methodology training grant (#R305B080016) awarded to UCLA by the Institute of Education Sciences of the US Department of Education and by grants 2P01DA01070-36 and 2K05DA000017-33 from the National Institute on Drug Abuse. The views expressed in this paper are the authors' alone and do not reflect the views/policies of the funding agencies.

References

- Amemiya Y, Anderson T. Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*. 1990; 18(3):1453–1463.
- Bentler, PM. A testing method for covariance structure analysis. In: Anderson, TW.; Fang, KT.; Olkin, I., editors. *Multivariate analysis and its applications*. Vol. 24. Institute of Mathematical Statistics; 1994. p. 123-135.
- Bentler, PM. EQS 6 structural equations program manual. Encino, CA: Multivariate Software; 2006.
- Bollen, KA. *Structural equations with latent variables*. John Wiley & Sons; 1989.
- Box G. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics*. 1954; 25(2):290–302.
- Browne M. Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*. 1974; 8:1–24.
- Browne M. Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*. 1984; 37(1):62–83. [PubMed: 6733054]
- Byrnes, JE. The sem.additions package for R. 2010. <http://r-forge.r-project.org/projects/sem-additions>
- Cai L, Lee T. Covariance structure model fit testing under missing data: An application of the Supplemented EM algorithm. *Multivariate Behavioral Research*. 2009; 44(2):24.
- Curran P, West S, Finch J. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*. 1996; 1(1):16–29.
- Duchesne P, Lafaye De Micheaux P. Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Computational Statistics & Data Analysis*. 2010; 54(4):858–862.
- Efron, B.; Tibshirani, R. *An introduction to the bootstrap*. Vol. 57. Chapman & Hall/CRC; 1993.
- Fox J. Teacher's corner: Structural equation modeling with the sem package in R. *Structural Equation Modeling*. 2006; 13(3):465–486.
- Hu L, Bentler PM, Kano Y. Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*. 1992; 112(2):351–362. [PubMed: 1454899]

- Imhof J. Computing the distribution of quadratic forms in normal variables. *Biometrika*. 1961; 48(3/4): 419–426.
- Jöreskog K. Some contributions to maximum likelihood factor analysis. *Psychometrika*. 1967; 32(4): 443–482.
- Jöreskog K. A general method for analysis of covariance structures. *Biometrika*. 1970; 57(2):239.
- Jöreskog, K.; Sorbom, D.; du Toit, S.; du Toit, M. LISREL 8: New statistical features. Lincolnwood, IL: Scientific Software International; 2000.
- Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*. 2004; 88(2):365–411.
- Lei P. Evaluating estimation methods for ordinal data in structural equation modeling. *Quality and Quantity*. 2009; 43(3):495–507.
- Liu H, Tang Y, Zhang H. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Computational Statistics and Data Analysis*. 2009; 53(4):853–856.
- Lo Y, Mendell N, Rubin D. Testing the number of components in a normal mixture. *Biometrika*. 2001; 88(3):767.
- Magnus J, Neudecker H. Symmetry, 0–1 matrices and Jacobians: A review. *Econometric Theory*. 1986; 2(2):157–190.
- Mardia K, Kent J, Bibby J. *Multivariate analysis*. 1980
- Muthen, L.; Muthen, B. *Mplus users guide*. Los Angeles, CA: Muthen & Muthen; 2006.
- Pearson E. Note on an approximation to the distribution of non-central χ^2 . *Biometrika*. 1959; 46(3): 364–364.
- R Development Core Team. *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria: 2011. Available from <http://www.R-project.org>
- Satorra A, Bentler PM. Scaling corrections for chi-square statistics in covariance structure analysis. 1988 *Proceedings of Business and Economics Sections*. 1988; 36:308–313.
- Satorra, A.; Bentler, PM. Corrections to test statistics and standard errors in covariance structure analysis. von Eye, A.; Clogg, CC., editors. Newbury Park, CA: Sage; 1994. p. 399–419.
- Satterthwaite F. Synthesis of variance. *Psychometrika*. 1941; 6(5):309–316.
- Tibshirani, R.; Leisch, F. Functions for the book an introduction to the bootstrap. 2009. <http://cran.r-project.org/web/packages/bootstrap/index.html>
- Yuan KH, Bentler PM. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological methodology*. 2000; 30(1):165–200.
- Yuan KH, Bentler PM. Two simple approximations to the distributions of quadratic forms. *British Journal of Mathematical and Statistical Psychology*. 2010; 63:273–291. [PubMed: 19793410]

Figure 1. Distribution of the ranked eigenvalues of $\hat{U}\hat{T}$ for one sample replicate at $n = 25$, $n = 100$, and $n = 5000$

Figure 2.

Density distribution of probabilities for a $\chi^2(54)$ variate and test statistics at $n = 25$, $N = 100$ replications, $p = 0.05$ under Model 1.

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

Performance of Test Statistics under N=100 Replications

Test Statistic	Type I Error				Empirical Power			
	0.01	0.05	0.10	0.10	0.01	0.05	0.10	0.10
Asymptotic Performance $n = 5000$								
T	0.00	0.03	0.1	1.00	1.00	1.00	1.00	1.00
T_M	0.00	0.04	0.1	1.00	1.00	1.00	1.00	1.00
T_{MV}	0.00	0.03	0.1	1.00	1.00	1.00	1.00	1.00
T_{MS}	0.00	0.02	0.1	1.00	1.00	1.00	1.00	1.00
Medium Sample Performance $n = 100$								
T	0.03	0.10	0.18	1.00	1.00	1.00	1.00	1.00
T_M	0.05	0.12	0.23	1.00	1.00	1.00	1.00	1.00
T_{MV}	0.01	0.05	0.10	0.98	1.00	1.00	1.00	1.00
T_{MS}	0.00	0.03	0.05	0.90	0.99	1.00	1.00	1.00
Very Small Sample Performance $n = 25$								
T	0.20	0.45	0.58	0.67	0.79	0.88	0.88	0.88
T_M	0.45	0.74	0.82	0.81	0.90	0.95	0.90	0.95
T_{MV}	0.01	0.19	0.41	0.14	0.56	0.79	0.56	0.79
T_{MS}	0.00	0.04	0.25	0.00	0.27	0.58	0.27	0.58

Table 2

Estimates of Test Statistics for the Open-Book Closed-Book Dataset

Test Statistic	Correlated Factor Model		Uncorrelated Factor Model	
	Estimate (df)	Probability	Estimate (df)	Probability
Original Sample $n = 88$				
T	2.07 (4)	0.72	50.53 (5)	< 0.01
T_M	2.21 (4)	0.70	41.03 (5)	< 0.01
T_{MV}	1.89 (3.42)	0.67	39.33 (4.79)	< 0.01
T_{MS}	1.55 (2.82)	0.64	28.86 (3.52)	< 0.01
Sub-sample $n = 15$				
T	8.73 (4)	0.07	20.77 (5)	< 0.01
T_M	10.16 (4)	0.04	23.24 (5)	< 0.01
T_{MV}	7.24 (2.85)	0.06	17.84 (3.84)	< 0.01
T_{MS}	5.50 (2.17)	0.07	11.64 (2.50)	< 0.01