Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

A TESTING METHOD FOR COVARIANCE STRUCTURE ANALYSIS

BY P. M. BENTLER

University of California, Los Angeles

Covariance structure analysis is a well-known method for testing theories on nonexperimental data. Under the null hypothesis, the population covariance matrix Σ is hypothesized to be a function of a vector of more basic parameters θ, i.e., $\Sigma = \Sigma(\theta)$. An illustration is $\Sigma = \Lambda\Phi\Lambda' + \Psi$, the confirmatory factor analysis model. The null hypothesis is typically evaluated with test statistics that are presumed to have χ^2 distributions in large samples. Previous work by Satorra and Bentler (1986, 1988a, 1988b) has shown that the general null distribution of these statistics is not χ^2 (df), but rather a weighted sum of 1-*df* χ^2 statistics. In this paper, this mixture distribution is suggested to be approximated using a method proposed by Gabler and Wolff (1987). A sampling experiment evaluates the performance of this approximation. When applied to correcting the estimated probability of the maximum likelihood test statistic, it is found to work well under conditions of independence of latent variates underlying the model, except at the smallest sample sizes, but to perform poorly under conditions of dependence. When applied to correcting the Satorra-Bentler scaled test statistic, it is found to work well under independence, but to overcorrect under dependence. A theoretical basis for these divergent results remains to be found.

1. Introduction. Hu, Bentler, and Kano (1992) recently studied the performance of six goodness-of-fit test statistics in covariance structure analysis using Monte Carlo sampling under the null hypothesis. For an introduction to covariance structure analysis, see, e.g., Bollen (1989). Under an assumed distribution of variables and a hypothesized model $\Sigma(\theta)$ for the population covariance matrix Σ, these statistics have an asymptotic central χ^2

---

* Supported in part by USPHS grants DA01070 and DA00017.

AMS 1980 Subject Classifications: Primary 62E05, 62H10, 62N05.

Key words and phrases: Chi-square, dependent observations, factor analysis

distribution that describes the mean, variance, and tail performance of the statistics. Hu et al. investigated three ways of violating theoretical conditions relevant to the choice of χ^2 as the reference distribution: they violated distributional assumptions, assumed independence conditions, and asymptotic sample size requirements. The effects of these violations on normal theory maximum likelihood (ML) and generalized least squares (GLS), elliptical theory (ERLS), heterogeneous kurtosis (HK), asymptotic distribution-free (ADF), and scaling-corrected (SCALED) test statistics ($T_{\mathrm{ML}}, T_{\mathrm{GLS}}, T_{\mathrm{ERLS}}, T_{\mathrm{HK}}, T_{\mathrm{ADF}}, T_{\mathrm{SCALED}}$) were studied. They found that: the normal theory tests worked well under some conditions but completely broke down under other conditions; the elliptical test performed variably; the heterogeneous kurtosis test performed better; the asymptotic distribution free test performed very badly at all but the largest sample sizes; and the scaled test statistic performed best overall. In related work, Chou, Bentler, and Satorra (1991) and Muthén and Kaplan (1992) also found that standard test statistics in covariance structure analysis could perform badly under conditions of violation of assumptions. Since the statistical theory is asymptotic, an especially important practical problem continues to be how to improve the performance in small samples of the existing statistics.

The purpose of this study is to propose a method for more accurately approximating the distribution of the T_{ML} and T_{SCALED} test statistics and to evaluate the performance of this approximation. Satorra and Bentler (1986, 1988a, 1988b) had shown that test statistics used in covariance structure analysis are not in general χ^2 distributed, though under precise modeling conditions this reference distribution would be appropriate. These modeling conditions reduce a mixture distribution to that of the standard χ^2 variate. Satorra and Bentler provided no procedures for actually implementing their theory in practice. The contribution of this paper is to propose an implementation of the Satorra-Bentler theory using an approximation developed by Gabler and Wolff (1987) for a similar problem. This implementation is evaluated with a small sampling study based on the Monte Carlo conditions previously studied by Hu et al.

2. Test Statistics. In this section, the notation is introduced, some currently available test statistics in covariance structure analysis are reviewed, and the technical problem is defined. This review must necessarily be short; summaries of various aspects of the theory are provided, for example, by Bentler and Dijkstra (1985), Browne (1984), Satorra (1989), and Wakaki, Eguchi, and Fujikoshi (1990). The review is provided to provide a context for the current work, as well as to provide the definitions needed to compare the proposed method with previously existing methods.

Following Hu et al. (1992, Appendix), let S represent the usual unbiased estimator based on a sample of size n of a $p \times p$ population covariance matrix Σ, whose elements are functions of a $q \times 1$ parameter vector $\theta : \Sigma = \Sigma(\theta)$.

A discrepancy function $F = F(S, \Sigma(\theta))$ can be considered to be a measure of the discrepancy between S and $\Sigma(\theta)$ evaluated at an estimator $\widehat{\theta}$. The normal theory maximum-likelihood (ML) discrepancy function (Jöreskog, 1969) is :

$$F_{\mathrm{ML}} = \log|\Sigma| - \log|S| + \mathrm{tr}(S\Sigma^{-1}) - p.$$

At the minimum, $\widehat{\Sigma} = \Sigma(\widehat{\theta})$ and F_{ML} takes on the value $\widehat{F}_{ML}$, where $T_{\mathrm{ML}} = (n-1)\widehat{F}_{\mathrm{ML}}$ is distributed, under the null hypothesis, as an asymptotic goodness-of-fit χ^2 variate with $(p^* - q)$ degrees of freedom, where $p^* = p(p+1)/2$. T_{ML} can be used as a test statistic to evaluate the null hypothesis $\Sigma = \Sigma(\theta)$. The null hypothesis is rejected if T_{ML} exceeds a critical value in the χ^2 distribution at an α-level of significance.

A quadratic form discrepancy function is:

$$F_{\mathrm{QD}} = (s - \sigma(\theta))' W^{-1} (s - \sigma(\theta)),$$

where s and $\sigma(\theta)$ are $p^* \times 1$ column vectors formed from the nonduplicated elements of S and $\Sigma(\theta)$, respectively, and W is a $p^* \times p^*$ positive-definite weight matrix. The asymptotically distribution-free (ADF) covariance structure method used by Hu et al. minimizes F_{QD} under the choice of optimal weight matrix W with typical elements

$$w_{ij,kl} = \sigma_{ijkl} - \sigma_{ij}\sigma_{kl},$$

where $\sigma_{ijkl} = E(x_{ti} - \mu_i)(x_{tj} - \mu_j)(x_{tk} - \mu_k)(x_{tl} - \mu_l)$ is the fourth-order multivariate moment of variables x_i about their means μ_i, and σ_{ij} is an element of Σ. In practice, sample moment estimators $s_{ijkl} = \sum_1^n (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j)(x_{tk} - \bar{x}_k)(x_{tl} - \bar{x}_l)/n$ and $s_{ij} = \sum_1^n (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j)/(n-1)$ are used to consistently estimate σ_{ijkl} and σ_{ij}. The ADF estimator provides an asymptotically efficient estimator $\widehat{\theta}$ without the need for distributional assumptions on variables. Under the null hypothesis, the associated test statistic $T_{\mathrm{ADF}} = (n-1)\widehat{F}_{\mathrm{QD}}$ has an asymptotic χ^2 distribution based on $(p^* - q)$ degrees of freedom. See Browne (1984) or Chamberlain (1982).

The fitting function F_{QD} for normal theory GLS can be simplified to

$$F_{\mathrm{GLS}} = \frac{1}{2}\mathrm{tr}\big((S - \Sigma(\theta))V^{-1}\big)^2,$$

if $W = 2K_p'(V \otimes V)K_p$, where V is a positive definite matrix that converges to Σ probability, and K_p is a known transition matrix. At the minima of the respective functions, both T_{ML} and $T_{\mathrm{GLS}} = (n-1)\widehat{F}_{\mathrm{GLS}}$ have asymptotic χ^2 distributions with $(p^* - q)$ degrees of freedom; they are asymptotically equivalent when the model is correct. Browne (1974) has shown that if V converges in probability to Σ ($V = S$ is typically used in practice) then GLS estimators are asymptotically equivalent to maximum likelihood estimators. See also Lee and Bentler (1980).

Under the assumption that all marginal distributions of a multivariate distribution are symmetric and have the same relative kurtosis, elliptical theory parameter estimators and test statistics can be obtained by readjusting the statistics derived from normal theory methods. Let $\kappa = \sigma_{iiii}/3\sigma_{ii}^2 - 1$ be the common kurtosis parameter of a distribution from the elliptical class. Multivariate normal distributions are members of this class with $\kappa = 0$. The fourth-order multivartiate moments σ_{ijkl} are related to κ by

$$\sigma_{ijkl} = (\kappa + 1)(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{jl}\sigma_{jk}),$$

where σ_{ij} is an element of Σ. As a result of this simplification, the discrepancy function for an elliptical distribution may be written as

$$F_E = \frac{1}{2}(\kappa + 1)^{-1}\mathrm{tr}((S - \Sigma(\theta))V^{-1})^2 - \delta(\mathrm{tr}(S - \Sigma(\theta))V^{-1})^2,$$

where as before V is any consistent estimator of Σ, and $\delta = \kappa/(4(\kappa + 1)^2 + 2p\kappa(\kappa+1))$ (Bentler (1983); Browne (1984)). The selection of V as a consistent estimator of Σ leads, under the model and assumptions, to an asymptotically efficient estimator of θ with $T_E = (n - 1)\widehat{F}_E$ at $\widehat{\theta}$ asymptotically distributed as a $\chi^2_{(p^*-q)}$ variate. A standard implementation is to choose $V = \widehat{\Sigma}$ at the minimum and $(\widehat{\kappa} + 1) = \sum_1^n((x - \bar{x})S^{-1}(x - \bar{x}))^2/np(p+2)$. Since the models to be investigated are invariant with respect to a constant scaling factor, at the minimum of F_E the second term drops out yielding $T_E = T_{\mathrm{ERLS}}$ as used in the Hu et al. study. See Shapiro and Browne (1987).

Heterogeneous kurtosis theory (Kano, Berkane, & Bentler, 1990) defines a more general class of multivariate distributions that allows marginal distributions to have heterogeneous kurtosis parameters. The elliptical distribution is a special case of this class of distributions. Let $\kappa_i^2 = \sigma_{iiii}/3\sigma_{ii}^2$ represent a measure of excess kurtosis of the i-th variable, and the fourth-order moments have the structure

$$\sigma_{ijkl} = (a_{ij}a_{kl})\sigma_{ij}\sigma_{kl} + (a_{ik}a_{jl})\sigma_{ik}\sigma_{jl} + (a_{il}a_{jk})\sigma_{il}\sigma_{jk},$$

where $a_{ij} = (\kappa_i + \kappa_j)/2$. If the covariance structure $\Sigma(\theta)$ is fully scale invariant and the modeling and distributional assumptions are met, the F_{QD} discrepancy function can be expressed as

$$F_{\mathrm{HK}} = \frac{1}{2}\mathrm{tr}((S - \Sigma(\theta))\widehat{C}^{-1})^2,$$

where $\widehat{C} = \widehat{A} * \widehat{\Sigma}$, and $*$ denotes the elementwise (Hadamard) product of the two matrices of the same order. Hu et al. used $\widehat{A} = (\widehat{a}_{ij}) = (\widehat{\kappa}_i + \widehat{\kappa}_j)/2$

based on the usual moment estimators $\widehat{\kappa}_i^2 = s_{iiii}/3s_{ii}^2$, with $\widehat{C} = \widehat{A} * S$. Kano et al. (1990) demonstrated that the simple adjustment of the weight matrix $\widehat{C}$ of the normal theory generalized least squares procedure (see F_{GLS} above) produces asymptotically efficient estimators. The associated test statistic $T_{\text{HK}} = (n-1)\widehat{F}_{\text{HK}}$ at the minimum has an asymptotic $\chi^2_{(p^*-q)}$ distribution under the assumed model.

Satorra and Bentler (1988a, b) developed two modifications of any standard goodness of fit statistic test $T = (T_{\text{ML}}, T_{\text{HK}}$ etc.) so that its distributional behavior should more closely approximate χ^2. One of these, the scaled test statistic, is available in the computer program EQS (Bentler, 1989, p.218) and is studied here. Satorra and Bentler (1986) noted that the general distribution of T is in fact not χ^2, but rather a mixture

$$T \xrightarrow{\ell} \sum_1^{df} \alpha_i T_i,$$

where α_i is one of the df (=degrees of freedom) nonnull eigenvalues of the matrix UV_{ss}, V_{ss} is the asymptotic covariance matrix of $\sqrt{n}(s-\sigma(\theta))$, T_i is one of the df independent χ_1^2 variates, and, when there are no constraints on free parameters (as in this study)

$$U = W^{-1} - W^{-1}\dot{\sigma}(\dot{\sigma}'W^{-1}\dot{\sigma})^{-1}\dot{\sigma}'W^{-1}$$

is the residual weight matrix under the model and the weight matrix W used in the estimation. The scaled statistic used by Hu et al. was based on T_{ML}, with $W = 2K'_p(\widehat{\Sigma} \otimes \widehat{\Sigma})K_p$, the normal theory ML weight matrix at the minimum of F_{ML} and $\dot{\sigma} = \partial F_{\text{ML}}/\partial\theta'$ evaluated at $\widehat{\theta}$. The mean of the asymptotic distribution of T_{ML} is given by $\text{tr}(UV_{ss})$. Then, defining the scaling estimate $k = \text{tr}(\widehat{U}\widehat{V}_{ss})/df$, where $\widehat{U}$ is a consistent estimator of U based on $\widehat{\theta}$, and $\widehat{V}_{ss}$ is the distribution-free estimator with elements $s_{ijkl} - s_{ij}s_{kl}$ (see above), the scaled ML statistic

$$T_{\text{SCALED}} = T_{\text{ML}}/k$$

defines Satorra and Bentler's SCALED test statistic as applied by Hu et al. This statistic is easier to implement than the general form.

3. The Hu, Bentler, and Kano (1992) Study. The Hu et al. sampling study is now described further to provide a more detailed summary of prior findings, and also because their conditions are replicated in the current investigation to study the tests proposed below. They used the confirmatory factor model $x = \Lambda\xi + \varepsilon$ to generate measured variables x under various conditions on the common factors ξ and unique variates ("errors") ε. They used 15 measured variables in x, with Λ being a 15×3 matrix having a simple cluster structure. See Hu et al. for details. In the standard approach to factor analysis, factors and errors are assumed to be normally distributed, factors are allowed to correlate with covariance matrix $\varepsilon(\xi\xi') = \Phi$, errors are uncorrelated with factors, i.e., $\varepsilon(\xi\varepsilon') = 0$ and various error variates are uncorrelated and have a diagonal covariance matrix $\varepsilon(\varepsilon\varepsilon') = \Psi$. As a result, $\Sigma = \Sigma(\theta) = \Lambda\Phi\Lambda' + \Psi$, and the elements of θ are the unknown parameters in Λ, Φ, and Ψ. Hu et al. had one condition in which factors and errors were multivariate normally distributed, so that the latent variates that are uncorrelated in the factor model are also independent of each other. They also used conditions in which the factors and/or errors were not normally distributed. In some of these conditions, factor/error variates that are uncorrelated under the model also were independent, while in other conditions these variables were uncorrelated but not independent. Independence of latent variates is a key condition in so-called asymptotic robustness theory that describes conditions under which normal theory ML and GLS test statistics are robust to violations of normality (e.g., Amemiya & Anderson (1990); Anderson & Amemiya (1988); Browne & Shapiro (1988); Mooijaart & Bentler (1991); Satorra & Bentler (1990)).

After generation of the population covariance matrix Σ under the assumed conditions, random samples of a given size from the population were taken, the null model was estimated, and the statistics $T = (n-1)\widehat{F}$ (for $T = T_{ML}, T_{GLS}, T_{ERLS}, T_{HK}, T_{ADF}$ and T_{SCALED}) were computed. The performance of these statistics across the sampling replications at a given sample size were the main data of their study. They used sample sizes of 150, 250, 500, 1000, 2500, and 5000 to evaluate the effects of sample size. In each condition at each sample size, 200 replications (samples) were drawn from the population, and the various estimators and goodness-of-fit tests were computed. The mean values and standard deviations of T across the 200 replications, and the empirical rejection rates at the $\alpha = .05$ level based on the assumed χ^2 distribution, were used to compare their methods.

Hu et al. found that when the latent common and unique factors were *independently* distributed, the anticipated asymptotic robustness properties of the χ^2 test were retained for normal theory methods when the sample size was relatively large. That is, the test statistics behaved as χ^2 variates even though assumed distributional assumptions were violated. Asymptotic robustness, however, could not be guaranteed at smaller sample sizes with ML. ERLS slightly, and HK somewhat more, overcorrected the test statistics when some or all the latent variates were nonnormal. The ADF method was very sensitive to sample size ; except under normality, it did not even perform acceptably with a sample size as large as 2500. The SCALED statistic outperformed ADF at all but the largest sample sizes. They also found that under conditions of *dependency* among latent factors and unique variates, normal theory methods could not be trusted, HK worked substantially better, ADF performed well only at very large sample sizes, and, across all sample sizes, the Satorra-Bentler SCALED statistic performed at closest to nominal levels of all the methods considered.

The question is whether a new approximation to the distribution of statistics $T = T_{\text{ML}}, T_{\text{GLS}}, T_{\text{ERLS}}, T_{\text{HK}}, T_{\text{ADF}}$ and T_{SCALED} could perform better under the same conditions. For simplicity, this study is limited to the performance of T_{ML} and T_{SCALED} when evaluated not by χ^2 but rather by the proposed approximation to the weighted sum of 1-df χ^2 variates. In addition, the performance of these statistics at smaller sample sizes than considered by Hu et al. is evaluated.

4. Approximating the General Distribution. As noted above, under general conditions the distribution of T is the distribution of a weighted sum of chi-square variables

$$T \xrightarrow{\ell} \sum_{1}^{\text{df}} \alpha_i T_i,$$

where α_i is one of the nonnull eigenvalues of UV_{ss}, V_{ss} is the asymptotic covariance matrix of $\sqrt{n}(s - \sigma(\theta))$, U is the residual matrix $U = W^{-1} - W^{-1}\dot{\sigma}(\dot{\sigma}'W^{-1}\dot{\sigma})^{-1}\dot{\sigma}'W^{-1}$, and T_i is one of the df independent χ_1^2 variates. Specific implementation depends on the estimator, here taken to be ML. Thus, $W = 2K'_p(\Sigma \otimes \Sigma)K_p$ and $\dot{\sigma} = \partial F_{\text{ML}}/\partial\theta'$. In practice these matrices are evaluated at the ML estimator $\widehat{\theta}$, yielding $\widehat{\alpha}_i$.

An explicit expression for the distribution of T is given in Johnson and Kotz (1970), but it is difficult to evaluate in practice. Gabler and Wolff (1987) proposed to do this by constructing a random variable Y that has the same first three moments as those of T, and has only minor differences in the higher moments. They standardized the problem so that $\Sigma\alpha_i = 1$, and took Y to be the positive random variable with density function

$$g(y) = \frac{1}{2}\sum_{i=1}^{\text{df}} \left(\frac{y}{2\alpha_i}\right)^{\frac{1}{2\alpha_i}-1} e^{-\frac{y}{2\alpha_i}} \Big/ \Gamma\left(\frac{1}{2\alpha_i}\right).$$

For $m = 1, 2, \cdots$, they verified that $E(T^m) \leq E(Y^m)$, derived the distribution function of Y, obtained the laplace transforms of T and Y, and showed that for small $t > 0$

$$H(t) = \gamma\left(\frac{\text{df}}{2}, \frac{t}{2\delta}\right) \Big/ \Gamma\left(\frac{\text{df}}{2}\right) \quad \text{with} \quad \delta = \left(\prod_{i=1}^{\text{df}} \alpha_i\right)^{\frac{1}{\text{df}}}$$

is a good approximation for the distribution function $F(t)$ of T where γ denotes the incomplete gamma function. Based on this theory they proposed an algorithm for approximating the probability $\Pr(\sum_1^{df} \alpha_i T_i \leq t)$ by the minimum of the functions $G(t)$ and $H(t)$, where

$$G(t) = \sum_{1}^{\mathrm{df}} \alpha_i \frac{\gamma\left(\frac{1}{2\alpha_i}, \frac{1}{2\alpha_i}t\right)}{\Gamma\left(\frac{1}{2\alpha_i}\right)}$$

and

$$H(t) = \Pr\left(\sum_{1}^{\mathrm{df}} \mathcal{T}_i \le \frac{t}{\delta}\right).$$

They indicate that comparisons made for those cases in which exact results are available showed extremely good accuracy to their approximation.

In the application of this approach to covariance structure analysis, specific test statistics T must be chosen. Here, only T_{ML} and T_{SCALED} are studied. Further, the population eigenvalues α_i have to be replaced by their sample counterparts $\widehat{\alpha}_i$. The accuracy of the procedure thus depends on several features. First, is the distribution of T_{ML} and T_{SCALED} well described as a mixture of 1-df χ^2 statistics under the conditions of the study (small sample size, dependence, nonnormality). Second, is the Gabler-Wolff approximation a good one to the theoretical distribution. Third, what is the quality of the estimators of the population eigenvalues. The proposed procedure can break down due to problems at any of these points. Regarding the eigenvalues, only the simplest estimator is considered based on the eigenvalues of the sample moment matrix product $\widehat{U}\widehat{V}_{ss}$. These eigenvalues can be computed using standard approaches for obtaining the eigenvalues of nonsymmetric matrices.

5. Simulation Results. Seven conditions were studied using the Hu, Bentler, Kano (1992) procedures. These conditions depended on the distribution of the common factors ξ and the unique factors ε, as well as the mutual dependence/independence of these sets of factors. The conditions were as follows:

1. Normal ξ	Normal ε	Mutual independence of ξ and ε
2. Nonnormal ξ	Nonnormal ε	Mutual independence of ξ and ε
3. Nonnormal ξ	Nonnormal ε	Mutual independence of ξ and ε
4. Normal ξ	Nonnormal ε	Mutual independence of ξ and ε
5. Nonnormal ξ	Nonnormal ε	Dependent ξ and ε
6. Nonnormal ξ	Nonnormal ε	Dependent ξ and ε
7. Nonnormal ξ	Nonnormal ε	Dependent ξ and ε

The conditions varied in several ways beyond that stated above, see Hu et al., but for the current purpose the main points are that: condition 1 is the standard case, leading to multivariate normal measured variables; conditions 3–4 with mutually independent latent variables are consistent with the theory of asymptotic robustness for normal theory test statistics; and conditions 5–7 violate asymptotic robustness conditions. In all conditions except condition 1, the measured variables were symmetrically but nonnormally distributed. Marginal kurtoses for factors and errors were in the range 0–28. In conditions 2 and 3, the true kurtoses for the nonnormal factors were −1.0, 2.0, and 5.0. In conditions 2–4, the kurtoses of the unique variates ranged from −1.0 to 7.5. In conditions 5–7, similarly defined nonnormal factors and errors were further divided by a rescaled chi random variable that was independent of the factors and errors. In addition to the sample sizes 150–5000 as studied by Hu et al., samples of size 50 and 100 also were drawn in the current study.

The behavior of the statistics T_{ML} and T_{SCALED} were evaluated in the simulation using proposed procedure to obtain estimated probability values. With 200 replications, at a nominal alpha level of .05, the true model should be rejected about 10 times. The following table summarizes the rejection rate under the seven conditions for the statistic T_{ML}.

Table 1. Summary of Simulation Results for T_{ML}: Number of Model Rejections

Condition	Sample Size 50	100	150	250	500	1000	2500	5000
1	195	7	7	12	6	5	4	8
2	198/199	9	7	8	9	17	14	10
3	195	4	8	6	9	15	10	10
4	191	5	8	8	7	15	9	9
5	177/177	148	158	175	195	197	200	199
6	--	140	147	172	190	196	199	200
7	--	134	141	169	188	195	199	199

Several major results can be seen in Table 1. When the data are normal (condition 1), the proposed procedure works well at all sample sizes except the smallest. At $n = 50$, the true model is rejected in 195 out of 200 replications. Essentially the same results are obtained even if the factors and errors are not normally distributed but they are mutually independent (conditions 2–4). At the smallest sample size, fewer than 200 replications sometimes yielded converged solutions. For example in condition 2, 199 out of 200 replications yielded converged solutions; all but one of these yielded a probability value suggesting model rejection. Most dramatically, the true model was rejected in an extremely high proportion of model tests under factor and error dependence (conditions 5–7). At the smallest sample sizes in conditions 6 and 7, no converged solutions were obtained, while at larger sample sizes about 3/4 to almost all true models were rejected. Results for the T_{SCALED} statistic are given in Table 2.

Table 2. Summary of Simulation Results for T_{SCALED}: Number of Model Rejections

	Sample Size							
Condition	50	100	150	250	500	1000	2500	5000
1	11	10	11	16	4	5	5	8
2	13/199	11	11	9	9	15	11	7
3	9	4	9	8	7	15	9	10
4	10	8	8	7	8	15	11	9
5	2/177	14	0	0	0	0	1	2
6	– –	12	0	0	0	1	1	1
7	– –	24	0	0	0	0	1	1

When evaluating the results using the Gabler-Wolff procedure as implemented here, it appears that as compared to the T_{ML} statistic, the T_{SCALED} statistic performs better under conditions 1–4 at sample size 50. At $n = 100$ or beyond, in conditions 1–4 the results for T_{ML} and T_{SCALED} are approximately the same. On the other hand, under conditions of variate dependence (conditions 5–7), T_{SCALED} generally yields probability estimates that are too high. At sample sizes 150 and larger, the true model is almost always accepted rather than rejected at the nominal alpha level. An exception seems to be occur at $n = 100$, where performance closer to nominal is found. But this trend does not hold for $n = 50$, where nonconvergence is a major problem.

6. Conclusions. The proposed method for evaluating the fit of covariance structure methods seems to be only marginally better than previously reported methods. Hu et al. (1992) had reported that the ML method under conditions of variate independence performed well when sample size was

500 or greater, but that at samples size 150 and 250 the statistic T_{ML}, when evaluated according to the χ^2 reference distribution, rejected models about twice as often as would be expected nominally. When evaluating T_{ML} using the weighted sum of $\chi^2(1)$ variates as proposed here, the proportions of model rejections are much closer to nominal at samples sized of 100 or above. This is a clear improvement. On the other hand, at sample size 50, model rejections are so frequent as to make the proposed procedure useless. At this time, there is no explanation for this surprising decrement in performance at the smallest sample size. It is likely that it has to do with the inadequate estimation of the eigenvalues of UV_{ss}, especially the smallest nonzero population eigenvalues which may be null at the smallest sample size . Further research will have to evaluate this hypothesis.

On the other hand, the proposed method when applied to evaluating T_{ML} under conditions of dependence of factors and errors was no better than when T_{ML} was evaluated with reference to the χ^2 distribution. Since this method was inadequate at even the largest sample size, and since the method does work well under conditions of variate independence, the poor performance must stem from either the fact that the distribution of T_{ML} is not well described as a mixture of $\chi^2(1)$ variates, or the fact that the Gabler-Wolff approximation breaks down under these conditions. Clearly, the derivations of Satorra and Bentler (1986, 1988a, 1988b) and Gabler and Wolff (1987) will have to be examined to determine whether there is a failure in some theoretical assumption that may invalidate either or both of their results under conditions of dependence.

Hu et al. had already shown that the T_{SCALED} statistic performed better on average than any other statistic they considered, when evaluated against a χ^2 distribution. In the present study, this statistic was referred to the Gabler-Wolff approximating distribution. As with the T_{ML} statistic, the T_{SCALED} statistic performed better at smaller sample sized under conditions of variate independence. However, when evaluated under conditions of variate dependence, T_{SCALED} under Gabler-Wolff performed substantially worse than when evaluated by the χ^2 distribution. Thus, there does not seem to be much virtue to the use of T_{SCALED} using the probability calculations based on Gabler and Wolff (1987) as applied in this study.

As noted by a reviewer, there is some non-smoothness in results across sample sizes for both test statistics. In Table 1, an example occurs at $n = 1000$ for condition 3–4. In Table 2, an example occurs at $n = 100$ for conditions 5–7. An obvious problem with the simulation is that the number of replications is undoubtedly too small to accurately describe the tail behavior of these test statistics. However, it does not seem likely that this design feature explains the apparent anomalies in trends of results, so there may be other technical difficulties with the simulation that we have not been able to locate.

Acknowledgment. The generous assistance of Eric J.C. Wu and Shinn Wu in programming, Li-tze Hu in performing the simulations, and several reviewers, is gratefully acknowledged.

REFERENCES

AMEMIYA, Y. and ANDWESON, T. W. (1990). Asymptotic chi-square test for a large class of factor analysis models, *Ann. Statist.* **18**, 1453–1463.

ANDERSON, T. W. and AMEMIYA, Y.(1988). The asymptotic normal distribution of estimators in factor analysis under general conditions, *Ann. Statist.* **16**, 759–771.

BENTLER, P. M. (1983). Some contributions to efficient statistics for structural models: Specification and estimation of moment structures, *Psychometrika* **48**, 493–517.

BENTLER, P. M. (1989). *EQS Structural Equations Program Manual.* BMDP Statistical Software, Los Angeles.

BENTLER, P. M. and DIJKSTRA, T. (1985). Efficient estimation via linearization in structural models. In *Multivariate Analysis VI* (P.R. Krishnaiah, ed.) 9–42. North-Holland, Amsterdam.

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables.* Wiley, New York.

BROWNE, M. W. (1974), Generalized least squares estimators in the analysis of covariance structure, *S. African Statist. J.* **8**, 1–24.

BROWNE, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures, *Brit. J. Math. Statist. Psychol.* **37**, 62–83.

BROWNE, M. W. and SHAPIRO, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models, *Brit. J. Math. Statist. Psychol.* **41**, 193–208.

CHAMBERLAIN, G. (1982). Multivariate regression models for panel data, *J. Econometrics* **18**, 5–46

CHOU, C.-P., BENTLER, P. M. and SATORRA, A. (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: A Monte Carlo study, *Brit. J. Math. Statist. Psychol.* **44**, 347–357.

GABLER, S. and WOLFF, C. (1987). A quick and easy approximation to the distribution of a sum of weighted chi-square variables, *Statist. Hefte* **28**, 317–325.

HU, L., BENTLER, P. M. and KANO, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychol. Bull.* **112**, 351–362.

JOHNSON, N. L. and KOTZ, S. (1970). *Continuous Univariate Distributions -2.* Houghton Mifflin, Boston.

JÖRESKOG, K. G., (1969). A general approach to confirmatory maximum likelihood factor analysis, *Psychometrika* **34**, 183–202.

KANO, Y., BERKANE, M and BENTLER, P. M. (1990). Covariance structure analysis with heterogeneous kurtosis parameters, *Biometrika* **77**, 575–585.

LEE, S.-Y and BENTLER, P. M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models, *S. African Statist. J.* **14**, 121–136.

MOOIJAART, A. and BENTLER, P. M. (1991) Robustness of normal theory statistics in structural equation models, *Statist. Neerlandica* **45**, 159–171.

MUTHÉN, B. and KAPLAN, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model, *Brit. J. Math. Statist. Psychol.* **45**, 19–30.

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

SATORRA, A. and fred Bentler, P.M. (1986). Some robustness properties of goodness of fit statistics in covariance structure analysis, *Proc. Bus. Econ. Stat. Amer. statist. Assoc.* 549–554.

SATORRA, A. and BENTLER, P. M. (1988a). Scaling corrections for chi-square statistics in covariance structure analysis, *Proc. Amer. Statist. Assoc.* 308–313.

SATORRA, A. and BENTLER, P. M. (1988b). Scaling corrections for statistics in covariance structure analysis. Technical Report, UCLA Statistics Series # 2.

SATORRA, A. and BENTLER, P. M. (1990), Model conditions for asymptotic robustness in the analysis of linear relations, *Comp. Statist. Data Anal.* **10**, 235–249.

SHAPIRO, A. and BROWNE, M. (1987) Analysis of covariance structures under elliptical distributions, *J. Amer. Statist. Assoc.* **82**, 1092–1097.

WAKAKI, H., EGUCHI, S. and FUJIKOSHI, Y. (1990). A class of tests for a general covariance structure, *J. Multiv. Anal.* **32**, 313–325.

Department Of Psychology
University Of California, Los Angeles
Los Angeles, CA 90024-1563
USA