British Journal of Mathematical and Statistical Psychology (1991) 44, 347–357 Printed in Great Britain

Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study

Chih-Ping Chou†

Department of Preventive Medicine, University of Southern California, 1000 S. Fremont Avenue, No. 641, Alhambra, CA 91803-1358, USA

P. M. Bentler

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

Albert Satorra

Department d'Economia, Universitat Pompeu Fabra, Barcelona, Spain

Research studying robustness of maximum likelihood (ML) statistics in covariance structure analysis has concluded that test statistics and standard errors are biased under severe non-normality. An estimation procedure known as asymptotic distribution free (ADF), making no distributional assumption, has been suggested to avoid these biases. Corrections to the normal theory statistics to yield more adequate performance have also been proposed. This study compares the performance of a scaled test statistic and robust standard errors for two models under several non-normal conditions and also compares these with the results from ML and ADF methods. Both ML and ADF test statistics performed rather well in one model and considerably worse in the other. In general, the scaled test statistic seemed to behave better than the ML test statistic and the ADF statistic performed the worst. The robust and ADF standard errors yielded more appropriate estimates of sampling variability than the ML standard errors, which were usually downward biased, in both models under most of the non-normal conditions. ML test statistics and standard errors were found to be quite robust to the violation of the normality assumption when data had either symmetric and platykurtic distributions, or non-symmetric and zero kurtotic distributions.

1. Introduction

The maximum likelihood estimation procedure in covariance structure analysis is developed under the assumption of a multivariate normal distribution of observed variables. This assumption is usually violated in practice. The empirical robustness of normal theory maximum likelihood (ML) estimation has therefore been extensively studied (Anderson & Gerbing, 1984; Bentler, 1983*a*; Boomsma, 1983; Browne, 1980,

†Requests for reprints.

D

1984; Harlow, 1985; Muthén & Kaplan, 1985, in press; Tanaka, 1984). Muthén & Kaplan (1985) provided detailed comparisons among several robustness studies. In general, the ML estimates have been found to be fairly accurate, even when the assumption of normality is violated. The test statistics and standard errors, however, are usually biased. The bias is found to increase with the degree of non-normality as measured by the univariate skewness and kurtosis (Harlow, 1985).

Theoretically less restrictive estimators, test statistics, and standard errors are available via generalized least squares (GLS) procedures making no distributional assumption (ADF) (Browne, 1982, 1984). However, ADF test statistics, parameter estimators, and standard errors have been found to be affected by the sample size and the model size as well as the degree of non-normality (Harlow, 1985; Harlow *et al.*, 1986, Muthén & Kaplan, in press). Muthén & Kaplan (in press) concluded that, in general, the estimation procedures currently available, such as ML, normal theory GLS, and ADF, do not provide appropriate test statistics and standard errors for severely non-normal data. This conclusion holds in spite of the fact that recent research has verified that there exist conditions under which certain ML statistics (e.g. Anderson & Amemiya, 1988; Browne & Shapiro, 1988) and other normal theory statistics (Satorra & Bentler, 1986) are asymptotically robust to violation of assumptions. The simulation studies show that the theoretical robustness properties may not be observed in the type of models studied in practice.

Satorra & Bentler (1986, 1988) proposed an adjustment to the goodness of fit χ^2 statistic for non-normal data in covariance structure analysis. The adjustment involves a rescaling. A correlation to standard errors has also been suggested (Bentler, 1983*b*, Bentler & Dijkstra, 1985). Basically, both corrections on test statistic and standard errors are performed by taking the covariance matrix of the sample covariances into consideration. The purpose of this study is to compare the Satorra–Bentler adjusted test statistics and the corrected standard errors with the standard ML statistics; and to determine whether the proposed corrections are appropriate under some non-normality conditions. The ADF estimation method will also be carried out for comparison.

2. Statistical theory

Since no constraints were imposed among parameters in this study, the statistical summary will be reviewed under this assumption. See Satorra (1989) for additional details. Assume that $\boldsymbol{S}$ represents the $p \times p$ covariance matrix obtained from a random sample of size $N=(n+1)$ from a population with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The null hypothesis to be tested is H_0: $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$, where $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ is a function of a vector of q parameters specified for a covariance structure model. Let $F=F(\boldsymbol{S},\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$ be a function that measures the discrepancy between $\boldsymbol{S}$ and $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$, evaluated at an estimator $\hat{\boldsymbol{\theta}}$. The discrepancy function to be minimized with respect to $\boldsymbol{\theta}$ in normal theory ML is

$$F_{ML}=\log|\boldsymbol{\Sigma}|+\operatorname{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})-\log|\boldsymbol{S}|-p,$$

and under GLS it is

$$F=(s-\sigma(\theta))'W(s-\sigma(\theta)),$$

where s and $\sigma(\theta)$ are column vectors that consist of the p^*, or $p(p+1)/2$, non-redundant (or lower triangular) elements of S and $\Sigma(\theta)$, respectively. The weight matrix W tends in probability to V^{-1}, where V is the $p^* \times p^*$ covariance matrix of s and is determined by the distributional assumption and estimation approach. Suppose that

$$\sqrt{n}(s-\sigma(\theta))\xrightarrow{L}N(\delta, V)$$

where $\xrightarrow{L}$ refers to convergence in distribution, δ is a finite p^* vector. Let D be the derivative of σ with respect to θ, $D=(\partial\sigma/\partial\theta)$, then, in general, the standard errors of parameter estimators are available from inverting the 'information matrix' $I(\theta)=(D'V^{-1}D)$. Thus

$$\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{L}N(0,\Delta)$$

where $\Delta=(D'V^{-1}D)^{-1}$. The standard errors can be estimated as the square root of the diagonal elements of $\hat{\Delta}/n$.

In the ML approach, V is a function of $\Sigma(\theta)$:

$$V_{ML}=2K_p'(\Sigma\otimes\Sigma)K_p,$$

where K_p is a $p^2 \times p^*$ matrix (Browne, 1974). The elements of the K_p matrix are

$$K_p(ij,kl)=\begin{cases} 1/2, \text{ if } i\neq j \text{ but } (i=k,\ j=l), \text{ or } (i=l,\ j=k); \\ 1, \text{ if } i=j=k=l; \\ 0, \text{ else; where } i<p,\ j<p, k<l<p. \end{cases}$$

In the ADF approach, the covariance matrix for the sampling covariances V_{ADF} is a fourth-order multivariate product moment matrix:

$$V_{ADF}(ij,kl)=\sigma(ijkl)-\sigma_{ij}\sigma_{kl}$$

where $\sigma(ijkl)$ is the expected fourth-order multivariate moment

$$\sigma(ijkl)=E(X_{ti}-\mu_i)(X_{tj}-\mu_j)(X_{tk}-\mu_k)(X_{tl}-\mu_l),$$

and σ_{ij} is the ijth element of Σ. In practice, s_{ij} replaces σ_{ij}, the sample mean $\bar{X}_i$ replaces μ_i, and $E(\cdot)$ is replaced by the sample average. The asymptotic goodness of fit χ^2 statistic to evaluate the null hypothesis H_0 is defined as

$$T=n\,\mathrm{MIN}\,F(S,\Sigma(\hat{\theta})).$$

Table 1. Factor analytic model and population covariance matrix

Variable	1	2	3	4	5	6
Loadings						
F_1^*	0.649	0.709	0.373	0.000	0.000	0.000
F_2^*	0.000	0.000	0.000	0.831	0.778	0.897
Error variance	0.579	0.498	0.861	0.309	0.395	0.195
	1.0000					
	.4598	1.0000				
	.2421	.2647	1.0000			
	.1317	.1439	.0758	1.0000		
	.1232	.1347	.0709	.6469	1.0000	
	.1421	.1553	.0818	.7460	.6982	1.0000

*Correlation between F_1 and F_2 is 0.244.

Consequently, $T_{ML}=n\hat{F}_{ML}$, and $T_{ADF}=n\hat{F}_{ADF}$. The scaled test statistic T_S proposed by Satorra & Bentler (1988), is a function of the standard goodness-of-fit χ^2 statistic T:

$$T_S=c^{-1}T,$$

where $c=\text{trace}(UV_{ADF})/df$, with $U=V_{ML}^{-1}-V_{ML}^{-1}D\Delta D'V_{ML}^{-1}$, and $df=p^*-q$, the degrees of freedom for the model.

The robust covariance matrix of the ML estimator (Bentler & Dijkstra, 1985, equation 1.4.5; Bentler, 1983*b*, equation 3.2; Satorra & Bentler, 1986; also Arminger & Schoenberg, 1989, equation 12) considering the distribution of variables, is

$$\Delta_R=\Delta(D'V_{ML}^{-1}V_{ADF}V_{ML}^{-1}D)\Delta.$$

When variables are in fact multivariate normally distributed, $\Delta_R=\Delta$. The standard errors can be estimated as the square root of the diagonal elements of $\hat{\Delta}_R/n$. These standard errors are called 'robust' standard errors in this study.

3. Methodology

The factor analytic model and some non-normal conditions obtained from Harlow (1985) are used for data simulation with Vale & Maurelli's (1983) algorithm. The factor analytic model contains two correlated latent factors; each factor has three observed indicators, Table 1 presents the model and the population covariance matrix of the observable variables. The non-normality conditions are reflected by

Table 2. Six conditions and population values for univariate skewness and kurtosis (Harlow, 1985)

Non-normality condition	Skewness						Kurtosis					
1. Symmetric with equal negative kurtoses	0.0	0.0	0.0	0.0	0.0	0.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0
2. Symmetric with equal positive kurtoses	0.0	0.0	0.0	0.0	0.0	0.0	6.0	6.0	6.0	6.0	6.0	6.0
3. Symmetric with unequal kurtoses	0.0	0.0	0.0	0.0	0.0	0.0	2.0	5.0	3.0	6.0	7.0	8.0
4. Unequal negative skewness, zero kurtoses	−.6	−.7	−.8	−.5	−.6	−.7	0.0	0.0	0.0	0.0	0.0	0.0
5. Unequal positive skewness, equal kurtoses	1.2	1.4	1.6	1.5	1.75	2.0	6.0	6.0	6.0	6.0	6.0	6.0
6. Unequal skewness, unequal kurtoses	−.5	1.5	−2.0	0.0	−1.0	2.0	0.0	4.0	8.0	−1.0	3.0	7.0

various combinations of skewnesses and kurtoses of the variables. The six conditions with 'extreme' non-normality in Harlow (1985) were selected for this study and are reported in Table 2. This study considered a wider range of skewnesses and kurtoses than that in Muthén and Kaplan (in press) with skewnesses ranging from −2.0 to 2.0 and the kurtoses ranging from −1.0 to 8.0. Two sample sizes, 200 and 400, are used for simulation. For each condition, 100 samples were generated for each sample size considered. Each sample was then evaluated by two versions of the model in Table 1. The first version, model A, contains 13 parameters, 6 factor loadings, 6 measurement error variances, and 1 factor covariance. The second version, model B, only includes the measurement error variances and the factor covariance as free parameters while fixing the factor loadings at the true values, as given in Table 1.

Although the six conditions selected cannot cover the wide range of non-normal distributions, they represent some non-normality conditions which are frequently encountered in practice. As described in Table 2, the six non-normality conditions included symmetric versus non-symmetric, and platykurtic (flat) versus leptokurtic (high kurtotic) distributions. Some conditions assumed the same distribution for all the variables as in Muthén & Kaplan (in press), and others assumed that variables have different distributions.

The ML and ADF test statistics under each condition are to be compared to their theoretically expected values as well as to the Satorra–Bentler scaled test statistic, and the standard errors are similarly to be compared to the expected values and with the robust standard errors. Three values of standard error are reported for ML and ADF approaches. The population values are the standard errors obtained from the 'population', which includes all 100 samples after being scaled for the appropriate sample size. The mean estimates are the means of the formula-based standard errors across 100 samples. Finally, the empirical values are the standard deviations of the parameter estimates. For robust standard errors, since corrections are made on the standard errors but not the parameter estimates, only the first two values are available.

4. Results

Table 3 reports all the χ^2 goodness-of-fit statistics for both models A and B at samples sizes 200 and 400 across the six conditions. The expected χ^2 value is equal to the degrees of freedom, say r, of a model and its variance is $2r$. The first panel summarizes the mean χ^2 statistics for model A with sample size 200 over 100 replications. In general, the mean χ^2 statistics are all close to the expected value, i.e. 8, for model A; and the standard deviation of χ^2 statistics (in the second row of each section) is also close to the expected value, namely 4. The non-normality conditions considered do not cause any significant difference among T_{ML}, T_{ADF}, or T_S. The frequencies of rejecting the null hypothesis at $\alpha = 0.5$, shown in the third row of each segment, seem to yield some variations among different conditions. However, the difference may be accounted for by sample variations. Harlow's (1985) study with more replications indicated that the rejection rates under all these six conditions are close to 5 per cent. Both ML and ADF test statistics can be concluded to be behaving very well under model A with a sample size of 200. The scaled test statistic also behaves as expected. A similar conclusion can be drawn for model A with a sample size of 400: the frequencies of rejecting the null hypothesis are again slightly higher than those in Harlow's study, and all three χ^2 statistics behave rather well.

The third and fourth panels of Table 3 provide the results for model B. The expected χ^2 value for model B is 14 with variance 28. Under extreme non-normality, both ML and ADF test statistics performed rather badly at both sample sizes. The ADF test statistics, as expected, behaved better with the larger sample size of 400. The rejection rates, however, show that the null hypothesis is rejected more often than expected. The ML test statistics are found to perform rather badly under conditions 2, 3, and 5. They appeared to remain robust if the data are symmetric with negative kurtoses (condition 1), or non-symmetric with zero kurtoses (condition 4). The performance of scaled test statistic T_S apparently exceeds that of both T_{ML} and T_{ADF} under the conditions investigated.

It has been found that the ML, GLS, and ADF estimated standard errors do not show bias when used with approximately normal data (Boomsma, 1983; Browne, 1982; Muthén & Kaplan, 1985). To study the behaviour of the robust standard errors, we shall only concentrate on the conditions 1, 3, 4, and 6, which provide quite a good representation of non-normality. Since the behaviour of estimates was similar at different sample sizes, only the results at a sample size of 400 are presented.

The results of ML, ADF, and robust (R) standard errors for model A are summarized in Table 4. The population robust standard errors, estimated by combining all 100 samples with $N = 400 * 100 = 40\,000$ and rescaling to the sample size of 400, are considered as the expected standard errors and used as the criteria to study the performance of various mean standard errors. The population standard errors for ADF and ML are also provided. Both ADF and R have almost the same

Table 3. Summary of χ^2 goodness-of-fit test statistics[a]

Condition	1	2	3	4	5	6
Model A ($N=200$):						
T_{ML}	8.193	8.321	8.198	8.256	8.781	7.609
	(4.150)	(4.253)	(4.136)	(3.836)	(5.193)	(4.171)
	4	9	7	7	9	5
T_{ADF}	8.738	8.725	8.721	8.513	8.393	8.185
	(4.354)	(4.237)	(4.153)	(3.868)	(3.846)	(4.279)
	7	8	9	5	7	6
T_S	8.259	8.381	8.264	8.173	8.231	7.766
	(4.159)	(4.003)	(3.888)	(3.803)	(4.140)	(4.153)
	4	6	6	6	7	4
Model A ($N=400$):						
T_{ML}	8.396	8.809	8.761	8.623	9.171	8.102
	(4.224)	(4.767)	(4.738)	(4.606)	(4.757)	(4.243)
	7	12	8	10	9	8
T_{ADF}	8.602	8.552	8.520	8.654	8.402	8.369
	(4.318)	(4.459)	(4.359)	(4.547)	(4.218)	(4.300)
	5	10	10	11	8	6
T_S	8.418	8.626	8.578	8.540	8.302	8.109
	(4.233)	(4.472)	(4.449)	(4.592)	(3.808)	(4.226)
	6	11	10	10	5	6
Model B ($N=200$):						
T_{ML}	13.346	20.167	20.180	14.965	21.612	15.053
	(5.276)	(9.673)	(9.326)	(5.711)	(10.917)	(6.315)
	4	27	30	8	32	9
T_{ADF}	16.537	21.462	21.337	17.121	21.933	18.855
	(6.867)	(13.043)	(12.742)	(8.663)	(12.097)	(10.007)
	10	35	34	16	32	23
T_S	14.633	14.837	14.866	14.813	14.508	14.740
	(5.860)	(7.318)	(7.326)	(6.035)	(6.779)	(6.598)
	7	12	10	8	10	9
Model B ($N=400$):						
T_{ML}	13.301	20.946	21.234	15.346	22.645	15.642
	(5.261)	(9.259)	(10.004)	(6.201)	(9.966)	(6.377)
	1	35	35	6	41	13
T_{ADF}	15.109	18.126	18.157	15.820	19.758	16.903
	(6.261)	(9.148)	(9.226)	(7.479)	(10.837)	(7.833)
	9	21	20	8	25	19
T_S	14.469	14.515	14.547	15.029	14.649	14.912
	(5.816)	(5.828)	(6.039)	(6.313)	(6.370)	(6.407)
	6	9	9	5	9	8

[a]The expected goodness-of-fit statistics for models A and B are 8 and 14, respectively.

population standard errors for model A, whereas ML in general yields a relatively smaller value. Mean ML standard errors in model A under non-normality conditions 1 and 4 have rather acceptable preformance, in contrast to results under other conditions; the differences between mean ML standard errors and the 'expected' standard errors are relatively small, and all three types of estimated standard errors are very similar. The mean ML standard errors under condition 6 are generally

Table 4. Summary of ML, ADF, and robust standard errors for model A[a] ($N=400$)

Condition	1			3			4			6		
ϕ	0.063	0.063	0.063	0.063	0.064	0.064	0.063	0.064	0.064	0.063	0.064	0.064
	0.063	0.062	0.063	0.062	0.061	0.062	0.062	0.063	0.063	0.061	0.061	0.063
	0.058	0.059	–	0.063	0.064	–	0.062	0.063	–	0.061	0.061	–
ε_1	0.084	0.082	0.082	0.083	0.095	0.095	0.083	0.086	0.086	0.084	0.081	0.083
	0.085	0.082	0.083	0.085	0.091	0.095	0.086	0.086	0.088	0.085	0.081	0.083
	0.076	0.075	–	0.086	0.093	–	0.085	0.085	–	0.076	0.079	–
ε_2	0.096	0.096	0.096	0.096	0.118	0.119	0.097	0.100	0.100	0.097	0.114	0.114
	0.100	0.097	0.100	0.100	0.107	0.114	0.101	0.100	0.104	0.100	0.106	0.112
	0.089	0.086	–	0.118	0.115	–	0.104	0.101	–	0.115	0.111	–
ε_3	0.067	0.055	0.055	0.066	0.130	0.131	0.067	0.069	0.069	0.067	0.133	0.133
	0.067	0.054	0.055	0.067	0.109	0.121	0.067	0.067	0.068	0.067	0.114	0.126
	0.052	0.052	–	0.122	0.119	–	0.057	0.060	–	0.120	0.122	–
ε_4	0.033	0.035	0.035	0.034	0.061	0.061	0.033	0.035	0.035	0.033	0.038	0.038
	0.033	0.033	0.034	0.033	0.047	0.051	0.033	0.034	0.035	0.033	0.032	0.034
	0.034	0.034	–	0.058	0.055	–	0.037	0.037	–	0.039	0.036	–
ε_5	0.036	0.036	0.036	0.036	0.063	0.063	0.036	0.038	0.038	0.036	0.058	0.058
	0.036	0.035	0.035	0.035	0.053	0.058	0.036	0.036	0.037	0.036	0.050	0.055
	0.036	0.036	–	0.068	0.061	–	0.038	0.038	–	0.062	0.051	–
ε_6	0.032	0.033	0.033	0.032	0.049	0.049	0.032	0.034	0.034	0.032	0.069	0.069
	0.032	0.032	0.033	0.032	0.042	0.045	0.032	0.033	0.033	0.032	0.050	0.057
	0.030	0.031	–	0.047	0.043	–	0.032	0.033	–	0.068	0.057	–
λ_1	0.072	0.069	0.069	0.072	0.080	0.080	0.072	0.073	0.074	0.072	0.070	0.070
	0.073	0.068	0.070	0.072	0.075	0.078	0.073	0.073	0.074	0.072	0.069	0.070
	0.067	0.067	–	0.079	0.079	–	0.078	0.078	–	0.069	0.072	–
λ_2	0.076	0.073	0.073	0.076	0.088	0.088	0.076	0.078	0.078	0.076	0.079	0.079
	0.078	0.072	0.074	0.077	0.084	0.087	0.078	0.077	0.079	0.078	0.078	0.080
	0.070	0.066	–	0.087	0.087	–	0.081	0.077	–	0.073	0.073	–
λ_3	0.061	0.059	0.059	0.060	0.067	0.067	0.061	0.062	0.063	0.061	0.062	0.063
	0.061	0.058	0.059	0.060	0.063	0.066	0.061	0.061	0.062	0.061	0.069	0.062
	0.062	0.060	–	0.065	0.065	–	0.065	0.062	–	0.064	0.061	–
λ_4	0.043	0.036	0.036	0.043	0.074	0.074	0.043	0.044	0.044	0.043	0.036	0.036
	0.043	0.035	0.036	0.043	0.066	0.070	0.043	0.043	0.044	0.043	0.034	0.035
	0.033	0.033	–	0.068	0.062	–	0.043	0.042	–	0.036	0.034	–
λ_5	0.044	0.037	0.037	0.044	0.069	0.069	0.044	0.045	0.045	0.044	0.049	0.049
	0.044	0.037	0.037	0.044	0.063	0.066	0.044	0.044	0.044	0.044	0.047	0.048
	0.039	0.040	–	0.072	0.070	–	0.048	0.049	–	0.052	0.051	–
λ_6	0.042	0.034	0.034	0.042	0.079	0.079	0.042	0.042	0.042	0.042	0.054	0.054
	0.042	0.033	0.034	0.042	0.069	0.072	0.042	0.041	0.042	0.042	0.051	0.053
	0.032	0.033	–	0.080	0.073	–	0.046	0.047	–	0.051	0.052	–

[a]The three entries in each cell are: population value, mean estimate, and empirical value.
– denotes value not applicable.

underestimated; they are further from the expected values, and mean ADF and R standard errors are closer to the expected values. Under condition 3, the mean ML standard errors are consistently negatively biased. Although negative bias is also observed for both mean ADF and R standard errors, with R performing slightly better than ADF, they are obviously much closer to the expected values than ML.

Table 5. Summary of ML, ADF, and robust standard errors for model B[a] ($N=400$)

Condition	1			3			4			6		
ϕ	0.063	0.063	0.063	0.062	0.064	0.065	0.062	0.064	0.064	0.063	0.063	0.064
	0.062	0.062	0.063	0.062	0.055	0.063	0.062	0.061	0.063	0.062	0.058	0.063
	0.057	0.057	–	0.064	0.067	–	0.061	0.061	–	0.063	0.065	–
ε_1	0.056	0.043	0.045	0.056	0.078	0.079	0.056	0.056	0.056	0.056	0.058	0.058
	0.056	0.041	0.045	0.056	0.069	0.076	0.056	0.053	0.056	0.056	0.055	0.057
	0.042	0.042	–	0.073	0.077	–	0.058	0.061	–	0.061	0.062	–
ε_2	0.054	0.042	0.044	0.054	0.095	0.102	0.054	0.055	0.055	0.054	0.092	0.101
	0.054	0.041	0.044	0.054	0.077	0.093	0.054	0.052	0.055	0.054	0.076	0.093
	0.041	0.043	–	0.103	0.082	–	0.052	0.056	–	0.104	0.090	–
ε_3	0.065	0.047	0.048	0.064	0.127	0.135	0.065	0.065	0.065	0.065	0.129	0.138
	0.065	0.046	0.048	0.064	0.103	0.126	0.065	0.062	0.065	0.065	0.106	0.131
	0.045	0.045	–	0.126	0.116	–	0.052	0.059	–	0.124	0.116	–
ε_4	0.030	0.027	0.028	0.030	0.058	0.066	0.030	0.031	0.031	0.030	0.029	0.031
	0.030	0.026	0.028	0.030	0.041	0.055	0.030	0.029	0.030	0.030	0.025	0.029
	0.028	0.028	–	0.064	0.053	–	0.030	0.032	–	0.032	0.030	–
ε_5	0.034	0.029	0.032	0.034	0.060	0.066	0.034	0.036	0.036	0.035	0.054	0.064
	0.034	0.029	0.031	0.034	0.047	0.062	0.034	0.034	0.035	0.035	0.044	0.060
	0.033	0.032	–	0.072	0.66	–	0.036	0.036	–	0.069	0.050	–
ε_6	0.027	0.024	0.025	0.027	0.047	0.054	0.027	0.028	0.028	0.027	0.056	0.081
	0.027	0.023	0.025	0.027	0.037	0.049	0.027	0.027	0.028	0.027	0.037	0.069
	0.025	0.024	–	0.051	0.049	–	0.028	0.030	–	0.081	0.056	–

[a]The three entries in each cell are: population value, mean estimate, and empirical value.
– denotes value not applicable.

The mean ML standard errors for the factor covariance ϕ seems to be the only exception. Its estimates performed rather well across all the non-normal conditions considered in this research.

Similar conclusions can be drawn from the results for model B summarized in Table 5. Standard errors of error variances again yielded greater bias under ML than under ADF or R estimation, except in condition 4. Although some robust standard errors are not as close to the expected values in model B as they are in model A, in general they provided better estimates than both ML and ADF.

5. Discussion

From the results of this study, we may conclude that the ML test statistics are quite robust to the conditions of non-normality considered in the study of model A. However, it should be stressed that, with larger models, ML has been found to be not as robust (Harlow, *et al.*, 1986; Muthén & Kaplan, in press). Although ML performed less well for model B under extreme non-normality, it still yields reliable results if the distributions of observable variables are in general symmetric and equally negative kurtotic, or have zero kurtoses if non-symmetric. Considering the minimal distributional assumption, the ADF test statistics did not perform as well as expected. In general, they are rejected too often. Although the ADF test statistics also behave badly in model B, they seemed to improve considerably when sample sizes became larger. The Satorra–Bentler scaled test statistic undoubtedly provided more satisfying results in all the non-normality conditions considered in this study. It should be noted that the model studied in this research is a very simple one and the non-normality conditions investigated are limited. The performance of the scaled test statistic with more complex models and different conditions of non-normality still requires further investigation. Nevertheless this research has demonstrated for the first time that the scaled test statistic is very promising in practice.

The ML 'information matrix' standard error estimates are obviously downward, or negatively, biased under extremely non-normal conditions. They were found to be off by a factor of two or more. In a single sample analysis, Arminger & Schoenberg (1989) found the ML standard errors to be downward biased by a factor of up to ten. The practical effect would be that ML parameters are far too often mistakenly judged to be significantly different from zero. Again the ML standard errors were found to perform slightly better under symmetric and platykurtic distributions, or non-symmetric and zero kurtotic distributions. The ADF standard errors performed comparatively better than the ML, and the robust standard errors showed the most acceptable behaviour under the conditions and models investigated in this study.

The practitioner may be interested to know that the Satorra–Bentler scaled test statistic and the robust standard errors are available in programs EQS, version 3.0. This program also contains a simulation feature that permits Monte Carlo studies, such as that reported above, to be implemented easily.

Acknowledgement

This research was supported by grants DA03976 and DA01070 from the US National Institute on Drug Abuse.

References

Anderson, J. C. & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analyses. *Psychometrika*, **49**, 155–173.

Anderson, T. W. & Amemiya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics*, **16**, 759–771.

Arminger, G. & Schoenberg, R. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika*, **54**, 409–425.

Bentler, P. M. (1983*a*). Simultaneous equation system as moment structure models. *Journal of Econometrics*, **22**, 13–42.

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

Bentler, P. M. & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), *Multivariate Analysis VI*, pp. 9–42. Amsterdam: North-Holland.

Boomsma, A. (1983). On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality. Doctoral dissertation, Rijksuniversiteit Groningen.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, **8**, 1–24.

Browne, M. W. (1980). The maximum-likelihood solution in interbattery factor analysis. *British Journal of Mathematical and Statistical Psychology*, **32**, 75–86.

Browne, M. W. (1982). Covariance structure. In D. M. Hawkins (Ed.), *Topics in Applied Multivariate Analysis*, pp. 72–141. Cambridge: Cambridge University Press.

Browne, M. W. (1984). Asymptotic distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62–83.

Browne, M. W. & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, **41**, 193–208.

Harlow, L.L. (1985). Behavior of some elliptical theory estimations with nonnormal data in a covariance structures framework: A Monte Carlo study. Doctoral dissertation, University of California, Los Angeles, USA.

Harlow, L. L., Chou, C.-P. & Bentler, P. M. (1986). Performance of chi-square statistic with ML, ADF, and elliptical estimators. Paper presented at the Psychometric Society meeting, Totonto, Canada.

Muthén, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, **38**, 171–189.

Muthén, B. & Kaplan, D. (in press). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology.*

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

Satorra, A. & Bentler, P.M. (1986). Some robustness properties of goodness of fit statistics in covariance structure analysis. *American Statistical Association, Proceedings of the Business and Economic Statistics Section*, pp. 549–554.

Satorra, A. & Bentler, P. M. (1988). Scaling corrections for statistics in covariance structure analysis. Los Angeles: *UCLA statistics series 2.*

Tanaka, J. S. (1984). Some results on the estimation of covariance structure models. Doctoral dissertation, University of California, Los Angeles, USA, unpublished.

Vale, C. D. & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, **48**, 465–471.

Received 25 October 1989; revised version received 26 July 1990