



Published in final edited form as:

Struct Equ Modeling. 2013 January 1; 20(1): 148–156. doi:10.1080/10705511.2013.742403.

EVALUATION OF A NEW MEAN SCALED AND MOMENT ADJUSTED TEST STATISTIC FOR SEM

XIAOXIAO TONG and PETER M. BENTLER
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Abstract

Recently a new mean scaled and skewness adjusted test statistic was developed for evaluating structural equation models in small samples and with potentially nonnormal data, but this statistic has received only limited evaluation. The performance of this statistic is compared to normal theory maximum likelihood and two well-known robust test statistics. A modification to the Satorra-Bentler scaled statistic is developed for the condition that sample size is smaller than degrees of freedom. The behavior of the four test statistics is evaluated with a Monte Carlo confirmatory factor analysis study that varies seven sample sizes and three distributional conditions obtained using Headrick's fifth-order transformation to nonnormality. The new statistic performs badly in most conditions except under the normal distribution. The goodness-of-fit χ^2 test based on maximum-likelihood estimation performed well under normal distributions as well as under a condition of asymptotic robustness. The Satorra-Bentler scaled test statistic performed best overall, while the mean scaled and variance adjusted test statistic outperformed the others at small and moderate sample sizes under certain distributional conditions.

1. Introduction

Classical goodness-of-fit testing in factor analysis is based on the assumption that the test statistics employed are asymptotically chi-square distributed, but this property may not hold when the factors and errors and hence the observed variables are nonnormally distributed. Even when the factors and errors are normally distributed in the population, the performance of test statistics in small sample sizes may be compromised (Hu, Bentler and Kano, 1992; Curran, West, & Finch, 1996). Robust methods such as Satorra-Bentler's (1994) mean scaling and mean and variance adjusted statistics were developed to be robust to nonnormality. As is well known, the Satorra-Bentler scaled chi-square statistic scales a normal theory statistic such as the maximum likelihood (ML) so that the mean of the test statistic asymptotically has the same mean as the reference chi-square distribution. Recently, Lin and Bentler (2012) proposed an extension to this statistic which not only scales the mean but also adjusts the degrees of freedom based on the skewness of the obtained test statistic. This statistic was proposed primarily in order to improve its robustness under small samples. A small simulation was consistent with this expectation, but the statistic was not evaluated for its performance under a wider range of conditions.

The purpose of the study is to evaluate the new mean/skewness test in comparison to other well-known robust statistics. The performance of four goodness-of-fit chi-square test statistics is evaluated under small sample sizes as well as under violations of normality in order to evaluate the relative performance of these statistics under the correct structural model as well as under misspecification to evaluate power. The behavior of maximum likelihood goodness-of-fit chi-square test (T_{ML}) and its three robust extensions: Satorra-Bentler scaled chi-square statistic (T_{SB}), mean scaled and variance adjusted statistic (T_{MV}) and mean scaled and skewness adjusted statistic (T_{MS}) are examined in this study. The study also provides a comparison of the standard T_{SB} statistic to one that is corrected for small

sample size. Headrick's (2002; Headrick & Swailowsky, 1999) relatively unstudied methodology for generating nonnormal data is used due to its ability generate a wider range of skew and kurtosis as well as control higher order moments than the more standard Fleishman (1978) and Vale and Maurelli (1983) procedures.

2. Test Statistics

The discrepancy between S (the unbiased estimator of population covariance matrix $\Sigma_{p \times p}$ based on a sample of size n) and $\Sigma(\theta)$ (the structured covariance matrix based on a specified model of q parameters) is typically evaluated by the normal-theory maximum-likelihood (ML) or quadratic form discrepancy functions:

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

$$F_{QD} = (s - \sigma(\theta))' W (s - \sigma(\theta)) \quad (2)$$

where p is the number of observed variables, s and $\sigma(\theta)$ are $p(p + 1)/2$ dimensional vectors formed from the non-duplicated elements of S and $\Sigma(\theta)$. Assume $\sqrt{n}(s - \sigma(\theta)) \rightarrow N(0, \Gamma)$ in distribution as $n \rightarrow \infty$, where Γ is the asymptotic covariance matrix of s . The typical elements of Γ are given by

$$\gamma_{ijkl} = \sigma_{ijkl} - \sigma_{ij}\sigma_{kl} \quad (3)$$

where the multivariate product moment for four variables z_i, z_j, z_k and z_l is defined as

$$\sigma_{ijkl} = E(z_i - \mu_i)(z_j - \mu_j)(z_k - \mu_k)(z_l - \mu_l) \quad (4)$$

and σ_{ij} is the usual sample covariance. Under multivariate normality, a consistent estimator of W is given by

$$\widehat{V}^{-1} = 2K_p' \left(\widehat{\Sigma} \otimes \widehat{\Sigma} \right) K_p \quad (5)$$

where K_p is a known transition matrix. Furthermore, we define

$$U = W^{-1} - W^{-1} \dot{\sigma}' (\dot{\sigma}' W^{-1} \dot{\sigma}')^{-1} \dot{\sigma}' W^{-1} \quad (6)$$

where $\dot{\sigma}' = \partial \sigma(\theta) / \partial \theta$ is the Jacobian matrix evaluated at $\hat{\theta}$. In practice, U can be estimated by plugging in $W^{-1} = \widehat{V}$. Then the goodness-of-fit chi-square statistic is given as:

$$T_{ML} = (n-1) \widehat{F}_{ML} \quad (7)$$

where \widehat{F}_{ML} is the minimum of (1) evaluated at the maximum likelihood estimate of parameters. Under assumptions of multivariate normality, T_{ML} has a χ^2 distribution with degrees of freedom $d = p(p + 1)/2 - q$, and this holds asymptotically under specific nonnormal conditions (see e.g., Savalei, 2008). For example, in a confirmatory factor analysis, when all factors are independently distributed and the elements of the covariance matrices of common factors are free parameters, T_{ML} can be insensitive to violations of the normality assumption. The Satorra-Bentler scaled chi-square statistic:

$$T_{SB} = T_{ML} / k \quad (8)$$

where $k = \text{trace}(UT)/d$ is a scaling constant that corrects T_{ML} so that the sampling distribution of T_{ML} will be closer to the expected mean d . The scaling constant k is an estimate of the average of the nonzero eigenvalues of UT . However, when the sample size is smaller than the degrees of freedom ($N < d$), (8) is not the correct formula since there will not be d nonzero eigenvalues. Hence, when $N < d$, we propose the use of $k = \text{trace}(UT)/N$ instead. This new Satorra-Bentler scaled chi-square statistic is thus given by:

$$T_{SB(New)} = T_{ML} / k \quad (9)$$

where $k = \text{trace}(UT)/\min(d, N)$, and $T_{SB(New)}$ is referred to a χ^2 distribution with $\min(d, N)$ degrees of freedom. The Satorra-Bentler mean scaled and variance adjusted statistic:

$$T_{MV} = v T_{ML} / \text{trace}(UT) \quad (10)$$

where $v = [\text{trace}(UT)]^2 / \text{trace}[(UT)^2]$. T_{MV} involves both scaling the mean and a Saitterwarthe second moment adjustment of the degrees of freedom (Saitterwarthe, 1941), and the new reference distribution is a central χ^2 with degrees of freedom v . The mean scaled and skewness adjusted statistic T_{MS} , newly proposed by Lin and Bentler, is defined as:

$$T_{MS} = v^* T_{ML} / \text{trace}(UT) \quad (11)$$

where $v^* = \text{trace}[(UT)^2]^3 / \text{trace}[(UT)^3]^2$ is a function of the skewness of T_{ML} . In addition to scaling the mean as in T_{SB} and T_{MV} , T_{MS} adjusts the degrees of freedom such that asymptotically, the quadratic form of T has the same skewness with a new reference distribution $\chi^2(v^*)$. The goal of modifying the degrees of freedom in T_{MV} and T_{MS} is to downwardly adjust the obtained statistic such that its distributions is as close to a central chi-square as possible. Note the above test statistics are described in their population form, but in estimation, (7) – (11) can be implemented by replacing $\hat{U}\hat{T}$ for UT .

3. Method

The confirmatory factor model is specified as $y = \Lambda \eta + \varepsilon$, where y is a vector of observed indicators that depends on Λ , a common factor loading matrix, η is a vector of latent factor scores (common factors) and ε is a vector of unique errors (unique factors). Typically, we assume that η is normally distributed and uncorrelated with ε . Hence, the restricted covariance structure of y is $\Sigma(\theta) = \Lambda \Phi \Lambda^T + \Psi$, where Φ is the covariance matrix of the latent factors and Ψ is a diagonal matrix of variances of errors. Since the observed indicators are a function of parameters in the factor analytic model, nonnormality in observed indicators is an implied consequences of nonnormality in the distributions of factors and errors.

In this study, a confirmatory factor model with 15 observed variables and 3 common factors is used to generate a model-based simulation. A simple structure of Λ is used where each set of five observed variables load onto a single factor with loadings of 0.7, 0.7, 0.75, 0.8 and 0.8 respectively, as shown in (12). Under each condition, the common and unique factors are generated using Headrick's fifth-order transformation (Headrick, 2002), and then the 15 observed variables are generated by a linear combination of these factors.

$$\Lambda^T = \begin{pmatrix} 0.7 & 0.7 & 0.75 & 0.8 & 0.8 & 0/0.8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.7 & 0.7 & 0.75 & 0.8 & 0.8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0.7 & 0.75 & 0.8 & 0.8 \end{pmatrix} \quad (12)$$

After generation of the population covariance matrix Σ , random samples of a given size from the population are taken. In each sample, the parameters of the model are estimated and the above four test statistics are computed by calling EQS using the REQS function in R (Mair, Wu, & Bentler, 2010) and specifying METHOD = ML, ROBUST in EQS. In estimation, the factor loading of the last indicator of each factor is fixed for identification at 0.8, and all the remaining nonzero parameters are free to be estimated. The behavior of T_{ML} , T_{SB} , T_{MV} and T_{MS} are observed at sample sizes of 50, 100, 250, 500, 1,000, 2,500 and 5,000. Particularly, when $N = 50 < d = 87$, the behavior of $T_{SB(New)}$ is also observed. At each sample size, 500 replications are drawn from the population. A statistical summary of the mean value and standard error of T under the confirmatory factor analysis model across the 500 replications, and the empirical rejection rate (Type I Error) at significance levels of $\alpha = 0.05$ on the basis of the assumed χ^2 distribution, are reported in Tables 2–4. An ideal type I error rate should approach 5% rejection of the null hypothesis, with a deviation of less than $2[(.05)(.95)/500]^{0.5} = .0195$.

To measure the empirical power of these test statistics, a misspecified model with an additional path from η_1 to y_6 is used for hypothesis testing. The loading of this path is fixed at 0.8 in estimation. The observed variables are still generated under the correct model, but are then analyzed under the incorrectly specified model. The empirical power, reported in the fourth row for each cell in Tables 2–4, is defined as the proportion of rejections of the null hypothesis for 500 simulated trials. A high rejection rate typically implies ideal performance of the test statistic, but this is not the case when simultaneously a high type I error rate exists (e.g., larger than 0.0695).

Three different conditions of distributions of factors and errors are simulated to examine the robustness of the above test statistics. In Condition 1, both common and unique factors are identically independently distributed as $\mathcal{N}(0, 1)$, resulting in a multivariate normal distribution of the observed variables. Condition 2 is designed to be consistent with asymptotic robustness theory, where the common and unique factors are independently generated nonnormal distributions. The common factors are correlated with specified first six moments and intercorrelations as in Table 1, while the unique factors are independent with arbitrarily chosen first six moments. In Condition 3, based on the distributions in Condition 2, the factors and error variates are divided by a random variable $Z = [\chi^2(5)]^{1/2} / \sqrt{3}$ that is distributed independently of the original factors and errors. This division results in the dependence of factors and errors, even though they remain uncorrelated. Because of the dependence, asymptotic robustness of normal-theory statistics is not to be expected under Condition 3.

Under the model $\Sigma(\theta)$, the degrees of freedom is 87. According to asymptotic robustness theory, we expect the normal-theory based test statistics to be valid for nonnormal data in Condition 2, in addition to the standard normal data in Condition 1. The expected mean of T_{ML} is 87 under Condition 1 and 2, while T_{ML} might break down in Condition 3. The anticipated mean of T_{SB} is 87, regardless of the three types of distributions and conditions considered. Particularly, when $N < d$, the expected mean of $T_{SB(New)}$ is corrected to N . The predicted means of T_{MV} and T_{MS} depend on the variables and are to be estimated during implementation.

4. Results

The simulation results for Conditions 1–3 are reported in Tables 2–4, one table per condition. The columns of each table give the sample size used for a particular set of 500 replications from the population. At each sample size, a sample was drawn, and each of the four test statistics shown in the rows of the table (ML, SB, MV, MS) was computed; this process was repeated 500 times. Then the resulting T statistics were used to compute (a) the mean of the 500 T statistics, (b) the standard deviation of the 500 T statistics, (c) the frequency of rejecting the null hypothesis at the 0.05 level under the correct model, i.e., the type I error, and (d) the frequency of rejecting the null hypothesis at the 0.05 level under the incorrect model, i.e., the empirical power. These are the four entries in each cell of each table.

Condition 1 in Table 2 is the baseline condition in which the factors and errors, and hence the observed variables, are multivariate normally distributed. Asymptotically, T_{ML} and T_{SB} yield a mean test statistic T of about 87, and the standard deviations are around 13.19. The means and standard deviations of T_{MV} and T_{MS} increase as the sample size gets larger, which turns out not to be a constant as we anticipate. All test statistics yields ideal type I error (within ± 0.0195 deviation from the 0.05 level) when the sample size reaches 1,000. Under small and moderate sample sizes, T_{MV} performs the best, followed by T_{ML} and T_{SB} , while T_{MS} tends to accept the null hypothesis too readily. It is clear that the adjustment on the Satorra-Bentler scaled test statistic, $T_{SB(New)}$, demonstrates improvement to some extent. The empirical power of all the test statistics reaches 100% when sample size is as large as 500. At smaller sample sizes, T_{SB} and T_{ML} perform on par in rejecting the misspecified model, while T_{MV} loses its advantage. Again, T_{MS} accepts the wrong model too frequently and yields very low rejection rates.

Condition 2 is designed to be consistent with asymptotic robustness theory. As we can see from Table 3, the behavior of the four test statistics is very similar to that in Condition 1. Asymptotically, T_{ML} and T_{SB} perform almost exactly as we expected, whose type I error lie within 0.008 deviation from 0.05. T_{MV} begins to approach T_{ML} and T_{SB} when sample size exceeds 500, while T_{MS} requires a sample size of 5,000 to demonstrate an ideal rejection rate. At smaller sample sizes, T_{MV} still outperforms the other test statistics while T_{MS} accepts the null hypothesis even more frequently than in Condition 1. The empirical power repeats the pattern we have observed in Condition 1, with T_{SB} and T_{ML} performing the best, followed by T_{MV} , and T_{MS} still performing the worst.

Condition 3 simulates a situation when the asymptotic robustness of normal-theory based test statistics is no longer valid. The empirical robustness of all test statistics except T_{SB} completely breaks down in this case: T_{ML} tends to always reject the correct model while T_{MV} and T_{MS} tend to always accept the null hypothesis. In either case, the empirical power of the test statistics can not be trusted. T_{SB} performs the best across all sample sizes, though the type I error rates are not so close to 0.05 level as those under Condition 1 and 2. The expected mean, standard deviation and empirical power of T_{SB} are retained asymptotically, indicating that T_{SB} should be a reliable test statistic under nonnormal distributions. The advantage of T_{MV} at small sample sizes disappears in this case, and T_{MS} continues giving unsatisfying results.

In conclusion, T_{SB} performs the best across three types of conditions. In particular, T_{SB} shows superior performance when all the other test statistics break down under Condition 3, in which case the asymptotic robustness theory is invalid. T_{ML} performs at least as well as T_{SB} under Conditions 1 and 2, and gives a slightly better type I error rate at small and moderate sample sizes. Under Conditions 1 and 2, T_{MV} significantly outperforms the other

test statistics at small and moderate sample sizes, in terms of the frequency of rejecting the null hypothesis under the correct model. The performance of T_{MS} improves as sample size increases under Condition 1, while it tends to accept the null hypothesis too frequently under Condition 2 and 3. This indicates that T_{MS} may downwardly overcorrect T_{ML} and thus cannot be trusted in testing when the data is nonnormally distributed.

5. Discussion

The behavior of the recently proposed mean scaled and skew-adjusted statistic was evaluated through a Monte Carlo study. To provide an appropriate comparison, two additional classical robust extensions of the standard maximum likelihood goodness-of-fit chi-square test statistic were utilized. As we can see from equations (8)-(11), the performance of these scaled and adjusted test statistics will mainly be affected by the eigenvalues of the product matrix UT . Yuan and Bentler (2010) evaluated the type I error and mean-square error of T_{MV} and T_{SB} under different coefficients of variation in the eigenvalues of UT , and found that T_{MV} will perform better than T_{SB} when the disparity of eigenvalues is large. This might lead to the situations we observed at small and moderate sample sizes under Condition 1 and 2. Lin and Bentler (2012) pointed out that when the eigenvalues of UT are constant, $v^* = d$, and T_{MS} will be equivalent to T_{ML} . This equivalence was not observed in the three conditions simulated in this study. It seems likely, as noted by Lin and Bentler, that the distribution of sample eigenvalues of UT may depart substantially from those of the population, especially in smaller samples. While it is clear T_{ML} and T_{SB} have tail behavior consistent with the asymptotic chi-square distribution under Condition 1 and 2, T_{MS} does not provide a better approximation of the chi-square variate and does not perform ideally. Also, the performance of T_{MS} under normality assumptions improves with an increasing sample size instead of a decreasing size as Lin and Bentler hypothesized.

We also proposed a modification to the Satorra-Bentler scaled statistic for the case of sample size smaller than degrees of freedom. In each of the conditions studied, this modification performed better than the standard version of the scaled statistic. However, at the smallest sample size this modification is still inadequate as model overacceptance remains a problem. Nonetheless, our overall results imply that in practice it may be beneficial to always apply the Satorra-Bentler scaled test statistic when we have little information about the distributions of observed variables. However, when we have sufficient confidence in the assumptions of normality or asymptotic robustness with a small or moderate size of observations, T_{MV} is recommended as an addition to T_{ML} and T_{SB} . T_{MS} could be taken into consideration when we want to be more conservative in confirming the fit of a model, but with limitation to normally distributed data.

References

1. Bentler, PM. EQS 6 structural equations program manual. Encino, CA: Multivariate Software, Inc; 2006.
2. Curran PJ, West SG, Finch JF. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*. 1996; 1:16–29.
3. Fleishman AI. A method of simulating non-normal distributions. *Psychometrika*. 1978; 43:521–532.
4. Headrick TC. Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics and Data Analysis*. 2002; 40:685–711.
5. Headrick TC, Swailowsky SS. Simulating correlated multivariate nonnormal distributions: Extending the fleishman power method. *Psychometrika*. 1999; 64:25–34.
6. Hu L, Bentler PM, Kano Y. Can test statistics in covariance structure analysis be trusted. *Psychological Bulletin*. 1992; 112(2):351–362. [PubMed: 1454899]

7. Lin J, Bentler PM. A third moment adjusted test statistic for small sample factor analysis. *Multivariate Behavior Research*. 2012; 47:448–462.
8. Mair P, Wu E, Bentler PM. EQS goes R: Simulations for SEM using the package REQS. *Structural Equation Modeling*. 2010; 17:333–349.
9. Savalei V. Is the ML chi-square ever robust to nonnormality? A cautionary note with missing data. *Structural Equation Modeling*. 2008; 15:1–22.
10. Vale CD, Maurelli VA. Simulating multivariate nonnormal distributions. *Psychometrika*. 1983; 48:465–471.
11. Yuan KH, Bentler PM. Two simple approximations to the distribution of quadratic forms. *British Journal of Mathematical and Statistical Psychology*. 2010; 63:273–291. [PubMed: 19793410]

Table 1

Specified Distributions of Factors ($\mu = 0, \sigma^2 = 1$)

| | Skew | Kurtosis | Fifth | Sixth | Correlations |
|----------|------|----------|-------|-------|--------------|
| η_1 | 0 | -1 | 0 | 28 | 1.0 0.3 0.4 |
| η_2 | 1 | 2 | 4 | 24 | 1.0 0.5 |
| η_3 | 2 | 6 | 24 | 120 | 1.0 |

Table 2

Summary of Simulation Results for Condition 1. (Factors and errors are independently distributed normal variates.)

| Test Statistics | Sample Size | | | | | | |
|------------------------|------------------|--------|--------|--------|--------|--------|--------|
| | 50 | 100 | 250 | 500 | 1,000 | 2,500 | 5,000 |
| ML | | | | | | | |
| Mean | 102.099 | 92.891 | 89.353 | 89.289 | 87.414 | 86.092 | 86.746 |
| SD | 14.888 | 14.706 | 13.271 | 13.331 | 14.214 | 12.631 | 12.504 |
| Type I Error | 0.29 | 0.118 | 0.074 | 0.066 | 0.066 | 0.046 | 0.04 |
| Empirical Power | 0.474 | 0.512 | 0.886 | 1.00 | 1.00 | 1.00 | 1.00 |
| SB scaled / new | | | | | | | |
| Mean | 108.518 / 62.367 | 95.755 | 90.415 | 89.778 | 87.709 | 86.203 | 86.801 |
| SD | 15.85 / 9.109 | 15.028 | 13.278 | 13.395 | 14.262 | 12.654 | 12.505 |
| Type I Error | 0.48 / 0.274 | 0.162 | 0.084 | 0.072 | 0.068 | 0.048 | 0.038 |
| Empirical Power | 0.618 / 0.436 | 0.59 | 0.902 | 1.00 | 1.00 | 1.00 | 1.00 |
| MV | | | | | | | |
| Mean | 30.535 | 41.937 | 59.432 | 71.121 | 77.581 | 81.923 | 84.59 |
| SD | 4.795 | 6.344 | 8.317 | 10.533 | 12.429 | 11.954 | 12.143 |
| Type I Error | 0.078 | 0.048 | 0.04 | 0.05 | 0.062 | 0.036 | 0.036 |
| Empirical Power | 0.162 | 0.29 | 0.83 | 1.00 | 1.00 | 1.00 | 1.00 |
| MS | | | | | | | |
| Mean | 16.639 | 22.969 | 36.836 | 50.894 | 63.256 | 74.562 | 80.501 |
| SD | 3.631 | 4.059 | 5.276 | 7.246 | 9.933 | 10.767 | 11.479 |
| Type I Error | 0.01 | 0.008 | 0.012 | 0.028 | 0.042 | 0.034 | 0.038 |
| Empirical Power | 0.028 | 0.1 | 0.716 | 1.00 | 1.00 | 1.00 | 1.00 |

Table 3

Summary of Simulation Results for Condition 2. (Factors and errors are nonnormally distributed, and are independent.)

| Test Statistics | Sample Size | | | | | | |
|------------------------|-----------------|--------|--------|--------|--------|--------|--------|
| | 50 | 100 | 250 | 500 | 1,000 | 2,500 | 5,000 |
| ML | | | | | | | |
| Mean | 101.317 | 93.059 | 89.735 | 87.148 | 88.238 | 87.042 | 86.464 |
| SD | 15.589 | 14.601 | 13.939 | 13.438 | 12.830 | 14.012 | 13.547 |
| Type I Error | 0.296 | 0.116 | 0.065 | 0.054 | 0.052 | 0.054 | 0.054 |
| Empirical Power | 0.456 | 0.504 | 0.89 | 1.00 | 1.00 | 1.00 | 1.00 |
| SB scaled / new | | | | | | | |
| Mean | 107.84 / 61.977 | 95.879 | 90.939 | 87.661 | 88.534 | 87.141 | 86.894 |
| SD | 16.111 / 9.259 | 15.116 | 14.005 | 13.280 | 12.874 | 13.977 | 13.547 |
| Type I Error | 0.45 / 0.272 | 0.156 | 0.082 | 0.058 | 0.052 | 0.058 | 0.05 |
| Empirical Power | 0.622 / 0.438 | 0.598 | 0.90 | 1.00 | 1.00 | 1.00 | 1.00 |
| MV | | | | | | | |
| Mean | 26.048 | 35.284 | 52.248 | 63.321 | 74.048 | 80.744 | 83.581 |
| SD | 5.104 | 6.788 | 8.5772 | 9.657 | 10.788 | 12.833 | 12.985 |
| Type I Error | 0.056 | 0.038 | 0.038 | 0.034 | 0.042 | 0.044 | 0.05 |
| Empirical Power | 0.126 | 0.208 | 0.81 | 1.00 | 1.00 | 1.00 | 1.00 |
| MS | | | | | | | |
| Mean | 12.682 | 16.457 | 27.149 | 40.114 | 53.378 | 69.21 | 77.486 |
| SD | 3.964 | 5.0237 | 6.603 | 7.776 | 8.879 | 11.006 | 11.975 |
| Type I Error | 0.002 | 0.004 | 0.002 | 0.008 | 0.022 | 0.034 | 0.05 |
| Empirical Power | 0.014 | 0.032 | 0.532 | 0.994 | 1.00 | 1.00 | 1.00 |

Table 4

Summary of Simulation Results for Condition 3. (Factors and errors are nonnormally distributed, and they are dependent.)

| Test Statistics | Sample Size | | | | | | |
|-----------------|------------------|---------|---------|---------|---------|---------|---------|
| | 50 | 100 | 250 | 500 | 1,000 | 2,500 | 5,000 |
| ML | | | | | | | |
| Mean | 149.251 | 159.211 | 176.434 | 197.834 | 202.859 | 217.761 | 224.404 |
| SD | 33.036 | 38.562 | 56.070 | 76.511 | 72.881 | 91.106 | 67.324 |
| Type I Error | 0.936 | 0.94 | 0.98 | 0.988 | 1.00 | 0.994 | 0.998 |
| Empirical Power | 0.972 | 0.995 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| SB scaled | | | | | | | |
| Mean | 109.717 / 63.056 | 97.687 | 91.077 | 87.562 | 87.218 | 86.712 | 86.489 |
| SD | 15.509 / 8.913 | 13.739 | 12.793 | 11.828 | 11.189 | 13.817 | 12.688 |
| Type I Error | 0.446 / 0.286 | 0.17 | 0.06 | 0.028 | 0.026 | 0.032 | 0.036 |
| Empirical Power | 0.648 / 0.476 | 0.52 | 0.582 | 0.856 | 0.992 | 0.998 | 0.998 |
| MV | | | | | | | |
| Mean | 13.321 | 14.403 | 17.527 | 19.030 | 24.708 | 30.199 | 34.724 |
| SD | 4.757 | 6.298 | 8.824 | 10.302 | 11.936 | 15.927 | 17.941 |
| Type I Error | 0.016 | 0.002 | 0.004 | 0.00 | 0.00 | 0.01 | 0.006 |
| Empirical Power | 0.028 | 0.035 | 0.098 | 0.396 | 0.758 | 0.956 | 0.984 |
| MS | | | | | | | |
| Mean | 5.476 | 5.148 | 5.501 | 5.356 | 6.985 | 8.636 | 10.279 |
| SD | 2.724 | 3.698 | 3.698 | 3.904 | 5.536 | 7.604 | 9.304 |
| Type I Error | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.002 | 0.00 |
| Empirical Power | 0.006 | 0.005 | 0.002 | 0.072 | 0.306 | 0.832 | 0.962 |