DOCUMENT RESUME

ED 420 711	TM 028 433
AUTHOR	Nevitt, Johnathan; Hancock, Gregory R.
TITLE	Relative Performance of Rescaling and Resampling Approaches to Model Chi Square and Parameter Standard Error Estimation in Structural Equation Modeling.
PUB DATE	1998-04-00
NOTE	39p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Chi Square; Computer Software; *Error of Measurement; Estimation (Mathematics); Monte Carlo Methods; *Research Methodology; Sample Size; Sampling; *Scaling; *Structural Equation Models
IDENTIFIERS	AMOS Computer Program; Bootstrap Methods; EQS Computer Program; *Resampling Techniques

ABSTRACT

Though common structural equation modeling (SEM) methods are predicated upon the assumption of multivariate normality, applied researchers often find themselves with data clearly violating this assumption and without sufficient sample size to use distribution-free estimation methods. Fortunately, promising alternatives are being integrated into popular software packages. For estimating model chi square values and parameter standard errors, EQS (P. Bentler, 1996) combats the effects of nonnormality by rescaling these statistics. AMOS (J. Arbuckle, 1997), on the other hand, offers bootstrap resampling approaches to accurate model chi square and standard error estimation. The current study is a Monte Carlo investigation of these two methods under varied conditions of nonnormality, sample size, and model misspecification. Accuracy of the chi square statistic is evaluated in terms of model rejection rates, while accuracy of standard error estimates takes the form of bias and variability of the estimates themselves. An appendix provides data for the paper's figures. (Contains 2 tables, 5 figures, and 31 references.) (SLD)

Relative Performance of Rescaling and Resampling Approaches to Model χ^2 and Parameter Standard Error Estimation in Structural Equation Modeling

Jonathan Nevitt

University of Maryland, College Park

Gregory R. Hancock

University of Maryland, College Park

Paper presented at the annual meeting of the

American Educational Research Association

San Diego, CA

Tuesday April 14, 1997

Correspondence should be directed to the first author at:

1228 Benjamin Building
University of Maryland
College Park, MD 20742-1115

e-mail: jnevitt@wam.umd.edu
phone: (301)405-3629
fax: (301)314-9245

Abstract

Though common structural equation modeling (SEM) methods are predicated upon the assumption of multivariate normality, applied researchers often find themselves with data clearly violating this assumption and without sufficient sample size to utilize distribution-free estimation methods. Fortunately, promising alternatives are being integrated into popular software packages. For estimating model χ^2 values and parameter standard errors, EQS (Bentler, 1996) combats the effects of nonnormality by rescaling these statistics. AMOS (Arbuckle, 1997), on the other hand, offers bootstrap resampling approaches to accurate model χ^2 and standard error estimation. The current study is an investigation of these two methods under varied conditions of nonnormality, sample size, and model misspecification. Accuracy of the χ^2 statistic is evaluated in terms of model rejection rates, while accuracy of standard error estimates takes the form of bias and variability of the estimates themselves.

Introduction

Normal theory estimation methods, specifically maximum likelihood (ML), are widely employed in structural equation modeling (SEM). These methods rest upon the assumption of multivariate normality and are based on large sample theory. In practice, however, data often violate the normality assumption (Micceri, 1989) and relatively small sample sizes are not uncommon in applied research. Thus, there has been considerable interest in evaluating the robustness of ML estimators and other estimation methods with respect to violation of the distributional assumption (Anderson & Gerbing, 1984; Boomsma, 1983; Browne, 1982, 1984; Chou, Bentler, & Satorra, 1991; Finch, West, & MacKinnon, 1997; Harlow, 1985; Hu, Bentler, & Kano, 1992; Muthén & Kaplan, 1985, 1992; Tanaka, 1984). Model parameters estimated via ML have been shown to be relatively unaffected by (i.e., remain unbiased under) particular departures from normality. However, research has also evidenced that the χ^2 test statistic and parameter standard errors under ML are substantially affected when the data are nonnormal. Specifically, under nonnormal conditions the χ^2 test statistic tends to be inflated while parameter standard errors become attenuated (see Chou & Bentler, 1995, and West, Finch, & Curran, 1995).

West et al. (1995) note three different approaches that are used to address the problems associated with ML estimation under nonnormal conditions. The first approach is an asymptotically distribution free (ADF) estimation method developed by Browne (1982, 1984). This method has the same desirable properties as ML but relaxes distributional assumptions. Unfortunately, ADF has the disadvantage of being computationally intensive and is thus limited to relatively simple models (Bentler, 1996), as well as requiring extremely large samples to obtain stable estimates (Jöreskog & Sörbom, 1992).

The second approach reviewed by West et al. (1995) is to adjust the ML χ^2 and standard errors to account for the presence of nonzero kurtosis. The adjustment is a rescaling of the χ^2 statistic to yield a test statistic that more closely approximates the referenced χ^2 distribution (Browne, 1982, 1984). Satorra and Bentler (1988, 1994) developed a modification of this rescaled test statistic, the robust Satorra-Bentler (SB) χ^2 test statistic, that has been incorporated into the EQS program (Bentler, 1996). Research has demonstrated that this rescaled test statistic tends to be less affected by model complexity and sample size as compared to the ADF estimator (Chou et al., 1991; Chou & Bentler, 1995; Hu et al., 1992). Similar to the SB χ^2 test statistic, a correction to ML standard errors has also been developed by Browne (1982, 1984) and a variant of this correction procedure (Bentler & Dijkstra, 1985) is currently available in EQS. The correction involves generating a robust covariance matrix of the parameter estimates from which robust standard error estimates are computed.

A third approach to obtaining robust statistics in SEM, as discussed by West et al. (1995), is bootstrap resampling. Bollen and Stine (1992), in work similar to that of Beran and Srivastava (1985), proposed a bootstrap method for p-value adjustment of the model χ^2 statistic. In general, to obtain adjusted p-values under the bootstrap resampling approach the ML χ^2 statistic is referred to an empirical sampling distribution of the test statistic generated via bootstrap samples drawn from the original sample data. As noted by Bollen and Stine (1992), naïve bootstrapping of the model χ^2 statistic (i.e., resampling from the original sample data) for SEM models is inaccurate. To adjust for this inaccuracy, Bollen and Stine (1992) developed a transformation of the original data:

$$\underset{\sim}{Z} = \underset{\sim}{Y}\,\underset{\sim}{S}^{-\frac{1}{2}}\,\underset{\sim}{\hat{\Sigma}}^{\frac{1}{2}}$$

in which $\underset{\sim}{Y}$ is the original n by p data matrix, $\underset{\sim}{S}$ is the sample covariance matrix and $\underset{\sim}{\hat{\Sigma}}$ is the estimated model-implied covariance matrix. Bollen and Stine (1992) showed that bootstrap samples drawn from the transformed data matrix, $\underset{\sim}{Z}$, yield reasonably accurate bootstrapping results for the model χ^2 statistic.

With respect to parameter standard errors, bootstrap values are obtained by generating parameter estimates from bootstrap samples. A bootstrap standard error is then computed as the standard deviations of the bootstrap parameter estimates. Bootstrap standard errors and the Bollen and Stine method for bootstrap model p-value adjustment are both available in the AMOS SEM program (Arbuckle, 1997).

Of the three robust approaches mentioned, only the latter two appear viable for more realistic sample sizes. While the rescaling approach has been subjected to considerable investigation, West et al. (1995) note that there have been no studies investigating the performance of the bootstrapping approach under varied experimental conditions. As a consequence, little is currently known about the performance of the bootstrapping approach, or of the comparative performance of the rescaling and resampling approaches to model χ^2 and parameter standard error estimation. It is the purpose of this study to provide such an investigation. Specifically, a Monte Carlo simulation is used to investigate these two methods' performance under varying data distributions, sample sizes, and model specifications.

Methods

Model specifications

Figure 1 presents the base underlying population model in this study, an oblique confirmatory factor analysis (CFA) model with three factors, each factor having three indicator variables. This model is the same population model as previously examined by Curran, West, and Finch (1996). Population parameter values, as seen in Figure 1, are such that all factor variances are set to 1.0, all inter-factor covariances (correlations) are set to 0.30, all factor loadings are set to 0.70, and all error variances are set to 0.51 (thereby yielding unit variance for the variables).

Insert Figure 1 about here

Four model specifications as previously explored by Curran et al. (1996) were considered here. Model 1 is a properly specified model as shown in Figure 1. Model 2 has misspecifications of inclusion, cross-loadings in the sample model for V8 on F2 and V5 on F3 that were not in the base population model. Model 3 has misspecifications of exclusion. For this model specification, the base population model was modified to include cross-loadings of V7 on F2 and V6 on F3 (both variable-factor population cross-loadings are set to 0.35). These variable-factor cross-loadings in the modified base population model were then excluded in the sample model for Model 3. Model 4, which is a combination of Models 2 and 3, has misspecifications of both inclusion and exclusion. The modified base population model employed in Model 3 was used as the population model, while the sample model included the cross loadings found in Model 2.

For all sample models, model identification was established by estimating the 3 factor variances and fixing one factor loading to 1.0 for each factor (V1 for F1, V4 for F2, V9 for F3). This approach to model identification was chosen (rather than fixing the factor variances to 1.0 and estimating all factor loadings) to ensure stability of the parameter estimates in the bootstrap resamplings. As noted by Arbuckle (1997), if model identification is achieved by fixing the factor variances, then the criterion for minimizing the model fit function may yield parameter estimates that are unique only up to a sign change. While the choice of approach to model identification is irrelevant in most applied settings, it has great importance with respect to bootstrap resampling. In bootstrapping, if the signs of some of the parameter estimates are arbitrary, these estimates will vary (some positive and some negative) from bootstrap sample to bootstrap sample thereby causing the resulting bootstrap standard errors to become artificially inflated. To avoid this inflation of standard errors in the bootstrapping methods, we fixed a factor loading and estimated factor variances to establish model identification.

Distributional forms and data generation

Three multivariate distributions were established through the manipulation of univariate skew and kurtosis. All manifest variables were drawn from the same univariate distribution for each data condition. Distribution 1 is multivariate normal with univariate skew and kurtosis both equal to 0. Note that we define normality, as is commonly done in practice, by using a shifted kurtosis value of 0 rather than a value of 3. Distribution 2 represents a moderate departure from normality with univariate skew of 2.0 and kurtosis of 7.0. Distribution 3 is severely nonnormal with univariate skew of 3.0 and kurtosis 21.0. Curran et al. (1996) report these levels of nonnormality to be reflective of real data distributions found in applied research.

Simulated data matrices consisting of n cases by nine variables were generated in GAUSS (Aptech Systems, 1996) to achieve the desired levels of univariate skew, kurtosis, and covariance structure. Multivariate normal and nonnormal data were generated via the algorithm developed by Vale and Maurelli (1983), which is a multivariate extension of the method for simulating nonnormal univariate data proposed by Fleishman (1978). Each simulated data matrix was obtained by first generating nine vectors of random deviates (of size n), each vector having a univariate standard normal distribution. The nine vectors of random normal deviates were concatenated to form an n by nine matrix of random normal deviates. This matrix was then pre-multiplied against the Cholesky factorization of an intermediate correlation matrix (described below) to yield a data matrix of multivariate normal random deviates with a correlational structure.

The intermediate correlation matrix mentioned above, based on the population correlation matrix drawn from the population model-implied covariance matrix, was generated using Equation 11 of Vale and Maurelli (1983, p. 467) and requires finding the roots of a third-degree polynomial. The purpose of this intermediate correlation matrix is to counteract the shift in the correlational structure of the multivariate normal data matrix that occurs when the univariate normal marginal distributions are transformed to nonnormal distributions via the transformation of Fleishman (Equation 1 of Vale & Maurelli, 1983). For the multivariate normal conditions in the study, the constructed intermediate correlation matrix is simply the original population correlation matrix.

After generating a simulated multivariate normal data matrix, Fleishman's transformation was applied to achieve the desired distributional form for each of the nine vectors in the data matrix. For a given population condition, all nine vectors in the simulated data matrix were transformed to the same degree of skew and kurtosis. As seen in Equation 1 of Vale and Maurelli (1983), the transformation to a nonnormal distribution is of the form:

$$\underset{\sim}{Y} = \underline{a} + \underline{b}\underset{\sim}{X} + \underline{c}\underset{\sim}{X}^2 + \underline{d}\underset{\sim}{X}^3$$

in which $\underset{\sim}{X}$ represents a matrix of multivariate normal deviates, and superscripts indicate raising each individual element in $\underset{\sim}{X}$ to the specified power. The scalar constants a, b, c, and d are chosen to yield transformed data with the desired univariate marginal skew and kurtosis.

Fleishman (1978) provides a table of these constants corresponding to varying combinations of skew and kurtosis.

For our investigation, we wished to investigate combinations of skew and kurtosis that were not provided in Fleishman's tables. Therefore, we developed an algorithm to yield the constants in the Fleishman transformation formula. This algorithm was based on a modified Newton-Raphson approach that solves for the Fleishman transformation constants found in Equations 2, 3, and 4 of Vale and Maurelli (1983). Using this algorithm, we were able to reproduce the constants from Fleishman's table (to 14 decimals) for a variety of skew and kurtosis combinations. Additionally, we generated the appropriate constants for the combinations of skew and kurtosis that are of primary interest in this investigation.

After applying the Fleishman transformation to the matrix of multivariate normal deviates, the resulting simulated data matrix conformed to the desired marginal distributional form with the desired correlational structure. The final step in the data generation process was to impose the model-implied covariance structure upon the simulated data matrix. This last step was accomplished by multiplying the n-by-nine data matrix by a nine-by-nine diagonal matrix of standard deviations taken from the population model-implied covariance matrix.

As a verification of our data generation mechanism, test data for the nine variables were simulated from the base population model. For each of the three distributional forms investigated in this study an n by nine data matrix was generated with n=100,000. From these simulated data matrices, sample estimates of skew and kurtosis were obtained for each of the nine vectors within each data matrix. Estimates of skew and kurtosis were computed using the Fisher g statistics (see, e.g., DeCarlo, 1997, p. 301). The sample covariance matrix was also obtained for each of the large sample data matrices.

Results for the large sample simulated data matrices suggest our data generation mechanism is effective for simulating data that conform to a specified covariance structure and distributional form. Inspection of the large sample covariance matrices shows only slight deviation from the population model-implied covariance matrix. For each distributional form, the average sample estimates of skew and kurtosis across the nine vectors in the data matrices were computed. Skew and kurtosis estimates for the three distributional forms are as follows: multivariate normal (skew=0, kurtosis=0) yielded -0.004, and 0.0025; moderately nonnormal (skew=2, kurtosis=7) yielded 2.0215, and 7.3211; severely nonnormal (skew=3, kurtosis=21) yielded 3.0140, and 21.8440. From these results, we are confident that our data generation mechanism yields simulated data that conform to both the target covariance structure and the target distributional form.

For each simulated data matrix replicated in our investigation (described below), we obtained from EQS a normalized estimate of multivariate kurtosis based on Mardia's (1970, 1974) coefficient. The average normalized estimate was computed for each of the distributional forms, at each level of sample size (described below). These values, which were averaged across the 800 data matrices (described below) for each distribution and sample size are as follows (for the multivariate normal, moderately nonnormal, and severely nonnormal distributions respectively): -0.6802, 17.7627, 35.9899 for n=100; -0.4787, 34.5262, 75.7690 for n=200; -0.3406, 67.4760, 166.0844 for n=500; -0.2319, 104.7899, 272.1365 for n=1000.

Design

Three conditions were manipulated in the current investigation: model specification (four models), distributional form of the simulated data (three types), and sample size (n=100, 200, 500, and 1000). The three manipulated conditions were completely crossed to yield 48 population conditions. Two-hundred random samples (i.e., data matrices) were analyzed in each of the 48 population conditions.

Model fittings and data collection

Both the EQS and AMOS programs require raw data (rather than sample covariance matrices) to obtain robust model χ^2 statistics and robust standard errors. For each simulated data matrix, models were fit in both EQS and AMOS to obtain the ML model χ^2 and parameter standard errors as well as the respective robust equivalents from each SEM program. The number of iterations to convergence for each model fitting was set to 200. This maximum was established for modeling each simulated data set in both EQS and AMOS, and was also established for bootstrap samples for the resampling estimators. Bootstrap samples drawn from each simulated data matrix were sampled with replacement and were of the same size as the original simulated data matrix.

From EQS the following information was collected: ML model χ^2 statistic and associated p-value, SB rescaled model χ^2 statistic and associated p-value, ML standard errors, and robust standard errors. From AMOS the information collected includes: ML model χ^2 statistic and associated p-value, Bollen and Stine bootstrapped model p-values from 250, 500, 1000, and 2000 bootstrap resamplings, ML standard errors, and bootstrap standard errors from 250, 500, 1000, and 2000 bootstrap resamplings. As we had found no guidelines regarding a sufficient number of bootstrap resamplings, we collected bootstrapped model p-values and standard errors from varying numbers of bootstrap resamplings to investigate the relative accuracy of these resampling estimators.

Summary measures

In the first part of our study, the ML model χ^2 test statistic and its robust alternatives are evaluated in terms of their model rejection rates (i.e., percentage of replications in which the target model is rejected at the α=0.05 level of significance). Model rejection rates were computed only under null conditions (Models 1 and 2) for which the expected value of the model χ^2 statistic equals the model degrees of freedom under multivariate normality (df=24 for Model 1 and df=22 for Model 2), and thus the desired rejection rate is 5%. Because Models 3 and 4 reflect model specifications in which factor loadings that truly existed in the population model were excluded in the sample, the percentage of model rejections is not a meaningful measure by which to judge the behavior of the χ^2 statistic.

In the second part of our study, parameter standard error estimates are assessed in terms of bias and variance. Bias is an assessment of standard errors relative to a true standard error. In this case, the true standard error value for each parameter of interest under each population condition was approximated empirically as the standard deviation of 2000 parameter estimates parametrically bootstrapped from the original population covariance matrix under a given population condition. Bias associated with each standard error estimate was computed as the observed standard error minus the approximate "true" standard error. The average of such bias measures for a given estimation method across each condition's 200 replications represents an overall estimate of that method's bias. Table 1 presents the approximate true standard errors under Model 1 for the inter-factor covariance for F2 and F1, and for the variable-factor loading of V2 on F1. Our decision to present results for only these parameters under Model 1 is described in the Results section below.

Insert Table 1 about here

The variance associated with a method's standard error under each population condition was used to assess standard error stability. The mean squared error (MSE), which considers both estimator bias and variability simultaneously, was not employed in our investigation as this measure tends to mask estimator bias when estimator variability is large and vice versa. This was the case for the results in our investigation. Additionally, Chou and Bentler (1995) note that the MSE is not commonly used as a criterion for the comparison of standard errors. Thus our assessment of standard error behavior in this investigation is through standard error bias and standard error variance.

Non-convergence, EQS-AMOS discrepancies, and unusable bootstrap samples

The percentage of non-convergent model fittings was minimal across the 48 conditions of the study with overall convergence rates of about 99%. The lowest convergence rate was 97% for the case of n=100 drawn from the severely nonnormal distribution under Model 4. Attempted fittings that failed to converge within 200 iterations or yielded improper solutions in either EQS or AMOS were discarded and replaced with convergent runs.

Within a given simulated data matrix the ML model χ^2 statistics from EQS and AMOS were closely monitored for discrepancies. A discrepancy criterion of 0.1 between the model χ^2 from EQS and AMOS was established to identify, remove, and replace replications that did not converge to the same solution. For the 9600 total replications in the study (200 replications in each of 48 conditions), 144 yielded such discrepancies between EQS and AMOS model χ^2 values, most of which were from the n=100 sample size.

For the bootstrap resampling methods, bootstrap samples that did not converge to a solution within 200 iterations were considered unusable and were discarded by the AMOS program. The AMOS program draws bootstrap samples from the original data matrix until the specified number of usable bootstrap samples have been reached. For diagnostic purposes, the number of unusable bootstrap samples was monitored for the 2000 bootstrap sample standard error estimator. The frequency of unusable bootstrap samples was a function of sample size, distributional form, and model specification. For the larger sample sizes that we investigated (n=500 and n=1000) there were no bootstrap samples that were unusable. The most extreme levels of unusable bootstrap samples were found in the n=100, nonnormal, and misspecified model conditions. The largest percentage of unusable bootstrap samples under these conditions was 8.3%. Again, all such unusable samples were replaced.

Results

Model rejection rates

For evaluating the performance of the model χ^2 test statistics under null conditions, a quantitative measure of robustness as suggested by Bradley (1978) was utilized. Using Bradley's liberal criterion, an estimator is considered robust if it yields an empirical model rejection rate within the interval [0.5α, 1.5α]. Using α=0.05, this interval for robustness of rejection rates is [2.5%, 7.5%]. Note that this interval is actually narrower than an expected 95% confidence interval , which evaluates to [1.98%, 8.02%] given the 200 replications per condition in this investigation and α=0.05 (i.e., $0.05 \pm 1.96[(0.05)(0.95)/200]^{1/2}$).

It is important to note that estimator performance in this study is based on the p-value associated with the model χ^2 statistic, not on the test statistic itself. Unlike Curran et al. (1996), bias in the model χ^2 statistic could not be considered in this study because the Bollen and Stine (1992) bootstrapping method only provides an adjusted p-value for assessing overall model fit. Thus our assessment of estimator performance is based on the rates of model rejection for the various test statistics under our investigation.

Table 2 presents the model rejection rates under Models 1 and 2 for the ML estimators, the SB rescaled χ^2 statistic, and the Bollen and Stine bootstrapped p-values. Percentages of model rejections that fall outside the boundaries of our robustness interval are shown in boldface type. Results in Table 2 for the ML estimators are consistent with findings in previous research (Chou & Bentler, 1995; Curran et al., 1996). Under the normal distribution and Model 1, rejection rates even at the smallest sample sizes for the ML estimators are within our criterion for robustness. With departures from multivariate normality, the ML estimators are not robust even under the largest sample sizes, with percentages of model rejections ranging from about 20% to about 40%.

Insert Table 2 about here

Results in Table 2 for the SB rescaled estimator are also consistent with findings in previous research (Chou & Bentler, 1995; Curran et al., 1996). Under perfect multivariate normality, the SB rescaled estimator simplifies to the ML estimator and thus we expect the SB to perform much like the ML estimators in the samples drawn from a multivariate normal population. This similarity of performance between the SB estimator and the ML estimators holds with sufficiently large sample sizes as seen in Table 2 under the multivariate normal conditions. Notice that for the smallest sample size of $n=100$ under multivariate normality, the SB rescaled estimator exceeds the upper bound of our robustness criterion (unlike the ML estimators which show robustness under these conditions) with model rejection rates of 8% and 9.5% for Models 1 and 2, respectively.

With departures from multivariate normality, the SB estimator remains robust – again, given adequate sample sizes. Under Model 1, for the moderately nonnormal distributions, the SB estimator remains robust until the n=100 sample size, at which point the associated percentage of model rejections becomes intolerably high (10.5%). Under Model 1 with the severely nonnormal distribution, the SB rescaled estimator only remains robust down through the n=500 sample size. Beyond this sample size, model rejection rates again exceed the upper boundary of our robustness criterion with rejection rates of 12% and 11% for the n=100 and 200 sample sizes, respectively. Similar results for the SB rescaled estimator are evidenced under Model 2; however, the estimator shows robustness down to the smallest sample sizes under this model specification.

With respect to the Bollen and Stine bootstrap adjusted p-values, Table 2 shows these estimators to be robust under nearly every condition in Models 1 and 2 – even under the combination of extreme departures from multivariate normality and the smallest sample sizes. These bootstrap estimators tend toward low model rejection rates – with smaller sample sizes yielding generally lower percentages of model rejections. In Table 2, one finds three instances in which a bootstrapped model rejection rate falls beyond our interval of robustness. In each instance, the observed model rejection rate falls below the lower bound of 2.5% (2.0% for all three instances). Inspection of the results for the bootstrap estimators in Table 2 shows several instances of model rejection rates that fall right at our lower bound of robustness.

It is interesting to note from the results for the Bollen and Stine bootstrap estimators in Table 2, that there are very little differences in the model rejection rates for the varying number of model resamplings. Comparing the performance of the Bollen and Stine estimators for 250, 500, 1000, and 2000 resamplings, one does not note any particular advantage for using more than 250 resamplings. This pattern of consistency in the performance across the resampling estimators is seen under most of the combinations of model specification, distributional form, and sample size in Table 2.

Factor covariance standard errors

We collected data from a variety of parameters in the population model. When considering the standard errors associated with factor variances and error variances in the model, we do not assign a great degree of practical importance to their analysis. In applied settings, neither parameter value estimation nor significance testing of factor variances and error variances are usually of substantive interest. For this reason, analysis of the standard errors associated with these parameters will not be reported here.

Data were collected for two of the three factor covariances in the population model; Cov(F2,F1) and Cov(F3,F2). Inspection of biases and variances associated with the standard errors for these covariances show very similar patterns of results across the four model specifications. Additionally, the patterns of results are strikingly similar for both of the covariances investigated. Because the results are quite consistent across models and covariances, the results for Cov(F2,F1) under Model 1 are presented here to characterize the general behavior of the covariances in the four models. In Figure 2 each of the three charts corresponds to a distributional form investigated in our study. Bar groupings in each chart correspond to the four levels of sample size investigated. The numbers upon which this and all figures are based appear in the Appendix.

Insert Figure 2 about here

Notice from the bias results in Figure 2 that under all conditions, decreasing sample size yields increased levels of bias. Under the multivariate normal condition, bias levels are reasonably low and similar across standard error estimators under the n=500 and 1000 sample sizes. Under the smaller sample sizes, a pattern of small (tending toward negative) bias in the ML and rescaled standard errors, and larger positive bias in the bootstrapped standard errors begins to emerge. Under the nonnormal conditions (with the exception of the bootstrapped standard errors under the moderately nonnormal distribution and largest sample size) all estimators yield negative bias in the Cov(F2,F1) standard error. For the ML estimators, specifically, these results are consistent with previous research that has demonstrated that standard errors become attenuated with departures from multivariate normality (see West et al., 1995, for a review).

While all estimators of standard errors show negative bias under the nonnormal distributions, notice in Figure 2 the robust methods – both the rescaled and the bootstrap resampled – yield standard errors that exhibit less bias than those using ML estimation. For the robust standard errors under departures from normality, the bootstrap standard errors yield smaller bias than the rescaled standard error. This comparison becomes more dramatic with decreasing sample size. Finally, comparing the bootstrap standard errors against one another, one finds very little difference in bias across the four resampling estimators. Thus, from the perspective of bias, there appears to be no real advantage to drawing greater numbers of bootstrap samples.

Figure 3 presents the analysis of standard errors for the Cov(F2,F1) parameter in terms of variance. From inspection of these results, one sees the variability in the standard errors increases with decreasing sample size under normal and nonnormal conditions. Under severely nonnormal conditions, however, variability in the standard errors for the robust methods is highest at the n=200 sample size rather than with n=100, a seemingly anomalous outcome.

Insert Figure 3 about here

From the charts in Figure 3, notice also that the ML estimators yield the smallest standard error variability. This pattern of results holds across all distributional forms and is most pronounced under the smaller sample sizes. Comparing the rescaled standard error to the bootstrap resampled standard errors shows the rescaled estimator yields only marginally smaller standard error variability than that of the bootstrap standard errors. As previously seen in the analysis of standard error bias, comparing the bootstrap standard errors against one another shows little difference in the variability of the standard errors with varying numbers of bootstrap samples.

Variable-factor loading standard errors

Data were collected for the variable-factor loadings of V2 on F1, V5 on F2, and V6 on F2. As with the factor covariances, patterns of results across models for each of the factor loadings were reasonably similar. Also, the patterns of results from one factor loading to another were also quite comparable. Thus, for simplicity, only the results for the loading standard errors of V2 on F1 (V2,F1) within Model 1 are presented here.

Figure 4 presents bias in the standard errors for the V2,F1 factor loading. As seen previously in the standard errors for the factor covariances, bias in the factor loading standard errors increases with decreasing sample size. This pattern exists across all three distributional forms. From the charts in Figure 4, notice again that increasing departures from multivariate normality lead to increased negative bias in the ML standard errors. Decreasing sample size also exacerbates this increase in negative bias associated with the ML estimators.

Insert Figure 4 about here

In Figure 4, we also see very different behavior between the robust approaches to estimating loading standard errors. The rescaled standard errors exhibit some degree of negative bias. However, it is substantially less than that exhibited in the ML standard errors, especially under the nonnormal conditions. As for the bootstrap standard errors for the loadings, these exhibit relatively small bias given sufficient sample size. Under all distributional forms for the n=1000, 500, and 200 sample sizes, the bootstrap standard errors yield very small bias as compared to the other estimators – often exhibiting the smallest bias of any of the standard errors examined. Under the smallest sample size, however, all bootstrapped standard errors yield large positive bias – substantially larger in absolute magnitude than the bias associated with any other estimator.

Variances associated with the V2,F1 standard errors are presented in Figure 5. Generally, variances in the factor loading standard errors for all estimators increased with decreasing sample size. With respect to the three distributional forms, a pattern of increased standard error variance with increasing departure from multivariate normality is evidenced, with the exception of the bootstrap standard errors under the smallest sample size. For these resampled standard errors under the n=100 sample size, standard error variances are larger under the moderately nonnormal condition than under the severely nonnormal condition. This anomaly is not readily interpretable or explained.

Insert Figure 5 about here

Comparing the standard error estimators against each other, the variances in the standard errors for the ML estimators tend to be the smallest, while the variances for the bootstrap resampling methods are the largest. This pattern exists across the three distributional forms and appears to be exacerbated by decreasing sample size. For the robust standard errors, the rescaled standard error variances are only slightly larger than the variances associated with the ML estimators. Considering the rescaled and resampled standard error variances, Figure 5 shows the robust estimators are comparable with respect to standard error variance (with the rescaled standard errors showing only marginally smaller variance) down to the smallest sample size. At the n=100 sample size and under all three distributional forms, the resampled standard error variances are considerably larger than the variances for the rescaled standard errors.

Comparing the bootstrap standard error variances against each other, one again sees very little difference in the behavior of the resampled standard errors with varying numbers of bootstrap samples. An exception to this pattern is under the n=100 sample size and nonnormal conditions. Under these conditions, larger variances tend to be associated with larger numbers of bootstrap samples drawn. Thus, from the perspective of standard error variability, there again appears to be no real advantage to drawing greater numbers of bootstrap samples.

Discussion

The results of this investigation replicate previous findings as well as expand our understanding of robust estimation procedures. As expected based on the literature (e.g., Bentler & Chou, 1987), the current study shows that under violations of multivariate normality the normal theory ML estimation procedures yield inflated model χ^2 values for correctly specified models as well as attenuated parameter standard errors. Also, the previous finding that the SB rescaled χ^2 statistic is robust under departures from multivariate normality given sufficient sample sizes (West et al., 1995) has been corroborated.

New findings from the current study center around the relative performance of the various robust procedures for parameter standard error and model χ^2 estimation, and will be discussed in terms of behavior under nonnormal conditions (as these methods would not likely be selected if one's data met standard distributional assumptions). Regarding standard errors, we first note that there were virtually identical results for all bootstrapping methods. That is, increasing the number of bootstrapped samples beyond 250 (the smallest number examined herein) did not appear to afford any change in quality of the bootstrapped standard error estimator; even fewer bootstrapped samples may work as well. Compared to other methods, under all but the smallest sample size the bootstrapped standard errors for both factor covariances and factor loadings tended to exhibit the least amount of bias - less than both the ML standard errors and the rescaled standard errors. For these sample sizes of $n \geq 200$ the magnitude of the bootstrap superiority over the rescaled approach decreased as sample size got larger. Under the smallest sample size of $n=100$, the bootstrapped standard errors yielded relatively small negative bias for the factor covariances, while the rescaled standard errors tended to exhibit over twice as much negative bias. However, when estimating factor loadings with $n=100$, bootstrapping methods showed a large amount of positive bias while the rescaled statistic had considerably smaller negative bias. These patterns of results were quite consistent across all models, those that were correct as well as those including specification errors.

Thus far, findings seem to indicate that using rescaling or bootstrapping methods is unwise with the smallest sample size of $n=100$, while for $n \geq 200$ results behave systematically for estimating covariance or loading standard errors. For these larger sample sizes both robust methods tend toward negative bias in standard error estimation, which means that a z statistic for parameter testing would often be slightly inflated. Again, bootstrapping's edge over the rescaled statistic is largest for the smaller sample sizes, but virtually disappearing as the sample size reaches n=1000.

Before championing the resampling approach, however, one much also consider the variability in the resampled statistics: small bias in the long run is of little use if individual results behave erratically. Consider first the estimation of a covariance parameter standard error, using a worst case scenario of n=200 under severe nonnormality as an example. The empirical true standard error is approximately .0734, while bootstrap methods (and the rescaled method as well) yielded a mean standard error variance near .008. Taking the square root yields an expected standard deviation for standard error estimates around .09, a value larger than the standard error itself. Under moderate nonnormality things look somewhat better, with a standard deviation determined to be near .02 when estimating a true standard error of .0628. This pattern improves slightly as sample size increases. The best case scenario of n=1000 under moderate nonnormality is estimating a true covariance standard error near .0264; the standard deviation of the standard error estimates is approximately .0045.

As for the estimation of a loading parameter standard error using the resampling based methods, the worst case of n=200 under severe nonnormality is estimating a true loading standard error near .2902; the standard deviation of the standard error estimates is approximately .14, roughly half the parameter standard error itself. The best case, on the other hand, with n=1000 under moderate nonnormality is estimating a true loading standard error near .0927; the standard deviation of the standard error estimates is approximately .015. In short, then, the stability of any given bootstrap standard error estimate seems to be better for loadings than for covariances, and improves with sample size and decreasing nonnormality as expected; for smaller sample sizes and/or greater nonnormality, and in particular for covariance standard errors, instability reaches levels that may well be considered intolerable .

Turning now toward assessing the overall fit of the model, the current study did not attempt to determine bias in estimates of the model χ^2 as has been done elsewhere (e.g., Curran et al., 1996); this is because the Bollen and Stine resampling approach yields a p-value rather than a χ^2 statistic. For this reason the quality of a method for assessing model fit was gauged by the Type I error rate when fitting a correct model. Given that the prior discussion regarding parameter standard error estimation recommended against the $n=100$ case, the results for nonnormal conditions with $n \geq 200$ suggest both the rescaling and resampling approaches to assessing overall model fit are preferable to ML estimation under violations of the normality assumption, and are quite comparable to each other. Generally, though not without exceptions, the bootstrapping approaches' model rejection rate seem to be slightly lower within the robustness interval (.025 to .075) than the SB rescaled method. On one occasion with $n \geq 200$ the resampling method even yields an error rate below the interval. Further, although the case of $n=100$ has been discounted here, it remains interesting that the SB scaled statistic yields unilaterally liberal error rates under all nonnormal (and even normal) conditions.

Practically speaking, all the attention to the χ^2 values' ability to control Type I error rates may seem at odds with the current practice of utilizing fit indices rather than a model χ^2 to gauge model acceptability. However, because many fit indices are constructed from model (and null model) χ^2 values, their behavior relative to distributional expectations is entirely relevant. Given the model rejection rate comparability of the SB and resampling approaches to assessing model fit, an immediate advantage to the former is that its rescaled χ^2 statistic can easily be incorporated into common fit indices (along with a rescaled null model χ^2 for incremental fit indices such as the CFI, as provided automatically in recent versions of EQS). Still, similar robust fit indices could be derived from the information provided by Bollen and Stine's approach. Specifically, the resulting p-value could be used to obtain the corresponding normal theory χ^2 statistic from the distribution with the proper number of degrees of freedom. This "bootstrapped" χ^2 statistic could then be incorporated into fit measures, along with a similarly derived statistic for the null model when required, just as with the SB rescaled statistics. To ensure sufficient precision in bootstrapped p-values, and hence in the corresponding χ^2 and the derived robust fit indices, a relatively large number of bootstrapped samples may be required even though 250 resamplings was seen to be perfectly adequate for standard error estimation. Given resampling methods' promise as shown in the current study, the theoretical development and empirical evaluation of such robust fit indices may be of great interest to the SEM community.

References

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

Aptech Systems. (1996). GAUSS System and Graphics Manual. Aptech Systems, Inc, Maple Valley, WA.

Arbuckle, J. L. (1997). AMOS Users' Guide, Version 3.6. SPSS.

Bentler, P. M. (1996). EQS Structural Equations Program Manual. Encino, CA: Multivariate Software, Inc.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. Sociological Methods & Research, 16, 78-117.

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.

Beran, R., & Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. Annals of Statistics, 13, 95-115.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. Sociological Methods & Research, 21, 205-229.

Boomsma, A. (1983). On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality. Unpublished doctoral dissertation, University of Gröningen, Gröningen.

Bradley, J. V. (1978). Robustness? British Journal of Mathematical and Statistical Psychology, 31, 144-152.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), Topics in applied multivariate analysis (pp. 72-141). Cambridge, England: Cambridge University Press.

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

Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), Structural equation modeling: Issues and applications (pp. 37-55). Newbury Park, CA: Sage.

Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. British Journal of Mathematical and Statistical Psychology, 44, 347-357.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychological Methods, 1, 16-29.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. Psychological Methods, 2, 292-307.

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. Structural Equation Modeling: A Multidisciplinary Journal, 4, 87-107.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. Psychometrika, 43, 521-532.

Harlow, L. L. (1985). Behavior of some elliptical theory estimators with non-normal data in a covariance structures framework: A Monte Carlo study. Unpublished doctoral dissertation, University of California, Los Angeles.

Hu, L. -T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? Psychological Bulletin, 112, 351-362.

Jöreskog, K. G., & Sörbom, D. (1992). LISREL VIII: A guide to the program and applications. Mooresville, IN: Scientific Software.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. Biometrika, 57, 519-530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. Sankhya, B36, 115-128.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105, 156-166.

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

Muthén, B, & Kaplan, D. (1992). A comparison of 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, 45, 19-30.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. American Statistical Association 1988 proceedings of the Business and Economics Sections (pp. 308-313). Alexandria, VA: American Statistical Association.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), Latent variables analysis: Applications for developmental research (pp.399-419). Thousand Oaks, CA: Sage.

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

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

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equations with non-normal variables: Problems and remedies. In R. H. Hoyle (Ed.), Structural equation modeling: Issues and applications (pp. 56-75). Newbury Park, CA: Sage.

Table 1. Empirically Based True Standard Errors for Properly Specified Model

Distribution	Sample Size	Cov(F2,F1)	Factor Loading for V2 on F1
Normal	100	0.07389620	0.21719432
	200	0.05064775	0.15377571
	500	0.03171152	0.08980573
	1000	0.02250867	0.06471447
Modately Nonnormal	100	0.09346542	0.31654237
	200	0.06278746	0.21547391
	500	0.03845778	0.13104480
	1000	0.02640384	0.09266848
Severely Nonnormal	100	0.10235749	0.41877029
	200	0.07340406	0.29015204
	500	0.04436619	0.19171851
	1000	0.03175627	0.13169042

Table 2. Model Rejection Rates for ML and Robust Estimators
Properly Specified and Inclusion Error Models

Model	Distribution	Sample Size	EQS ML	AMOS ML	Satorra Bentler	BS 250	BS 500	BS 1000	BS 2000
1	1	100	5.5	5.5	8.0	3.5	3.5	3.5	3.0
1	1	200	7.0	7.0	7.0	3.0	4.0	4.0	4.0
1	1	500	6.0	5.5	6.5	5.0	5.0	5.0	5.0
1	1	1000	7.5	7.0	7.0	5.5	6.0	6.5	7.0
1	2	100	23.5	23.0	10.5	3.5	3.5	4.0	3.5
1	2	200	20.5	20.0	5.5	4.0	4.5	3.0	4.0
1	2	500	20.0	20.0	6.5	4.5	4.0	4.0	3.5
1	2	1000	22.0	22.0	4.0	4.5	4.5	4.0	4.0
1	3	100	30.0	29.5	12.0	2.5	2.5	3.0	3.5
1	3	200	40.0	40.0	11.0	6.5	5.5	5.5	5.5
1	3	500	37.0	37.0	4.0	2.5	2.0	2.5	3.0
1	3	1000	36.5	36.5	3.5	2.5	3.0	3.5	3.5
2	1	100	7.5	7.5	9.5	2.5	2.0	3.0	2.5
2	1	200	5.5	5.5	6.5	3.5	3.5	4.0	3.5
2	1	500	8.5	8.5	7.5	6.0	6.0	6.5	6.5
2	1	1000	8.5	8.5	8.0	5.5	5.5	5.0	6.0
2	2	100	22.0	22.0	8.5	2.0	2.5	2.5	2.5
2	2	200	19.5	19.5	7.5	6.0	6.0	6.0	6.0
2	2	500	20.0	20.0	7.0	6.0	7.5	6.5	6.5
2	2	1000	20.5	20.5	5.0	4.0	3.5	3.5	4.0
2	3	100	27.5	27.5	11.5	2.5	2.5	2.5	2.5
2	3	200	30.5	30.5	5.5	2.5	3.0	3.0	2.5
2	3	500	38.5	38.5	5.5	2.5	3.0	2.5	2.5
2	3	1000	42.0	42.0	6.5	4.0	4.5	4.5	4.5

Note: Model 1 = properly specified model. Model 2 = inclusion error. Distribution 1 = Multivariate Normal.
Distribution 2 = Moderately Nonnormal. Distribution 3 = Severely Nonnormal.
BS250 = Bollen and Stine Bootstrapped p-value, 250 resamplings
BS500 = Bollen and Stine Bootstrapped p-value, 500 resamplings
BS1000 = Bollen and Stine Bootstrapped p-value, 1000 resamplings
BS2000 = Bollen and Stine Bootstrapped p-value, 2000 resamplings

Figure 1. Base Population Model

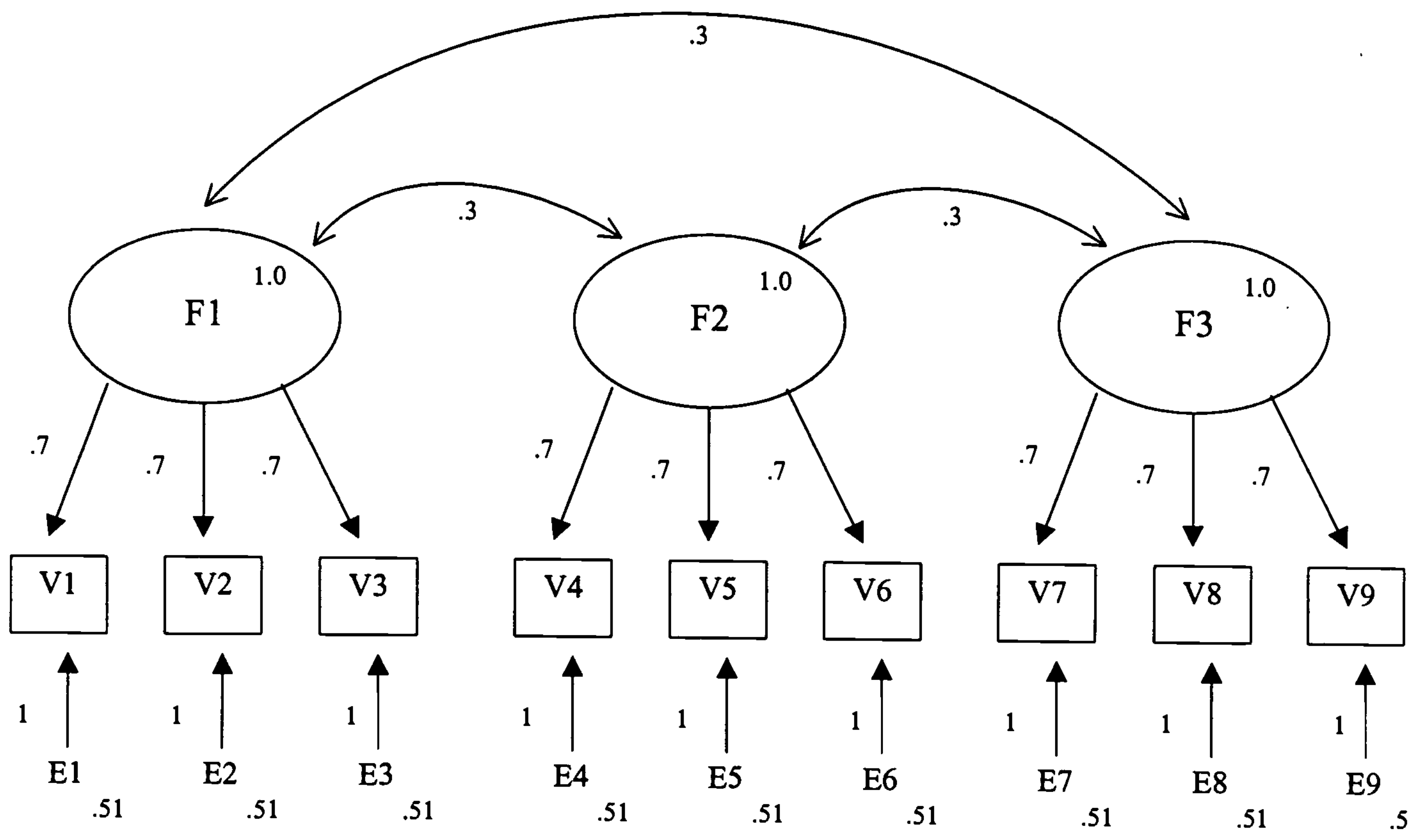
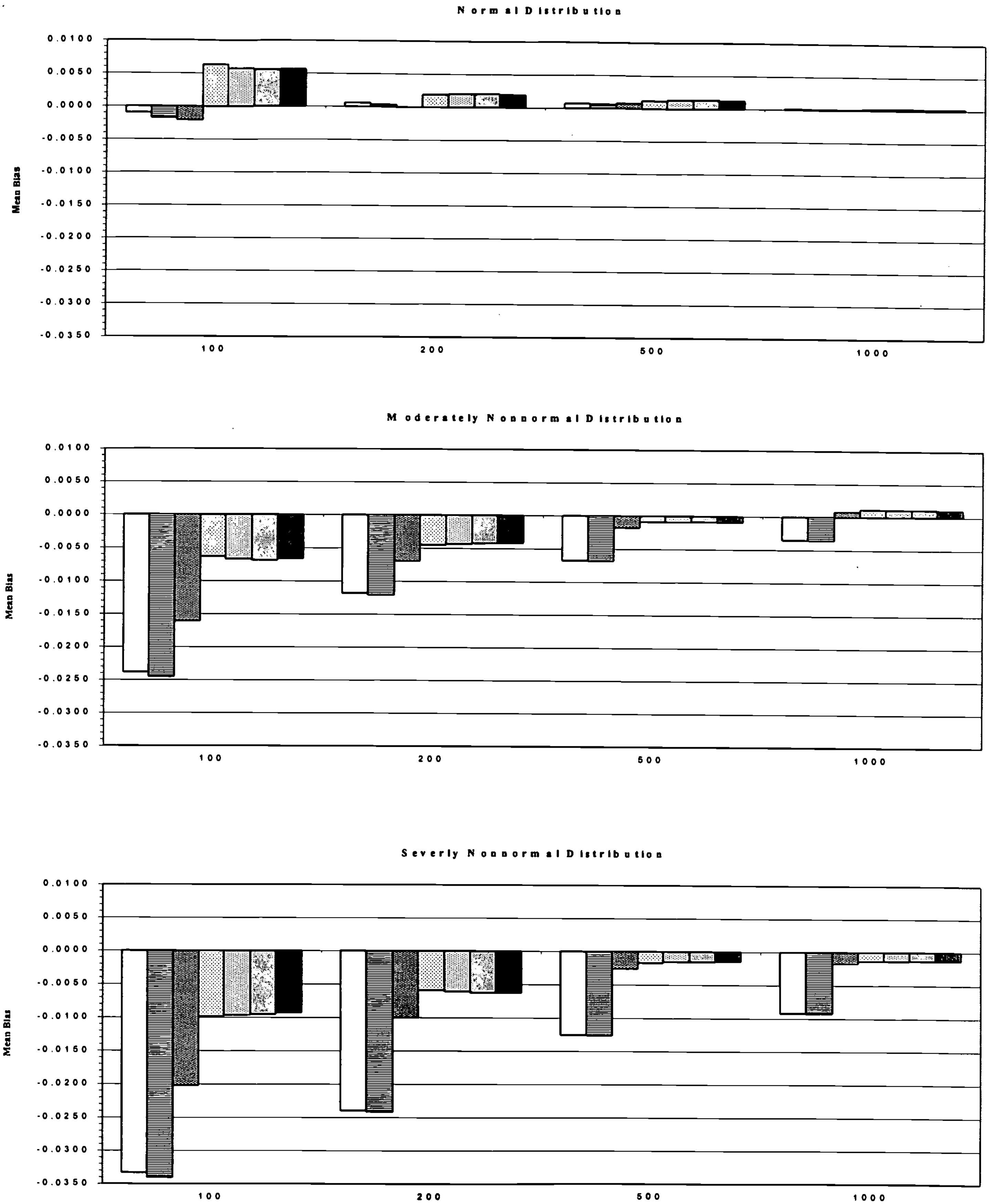
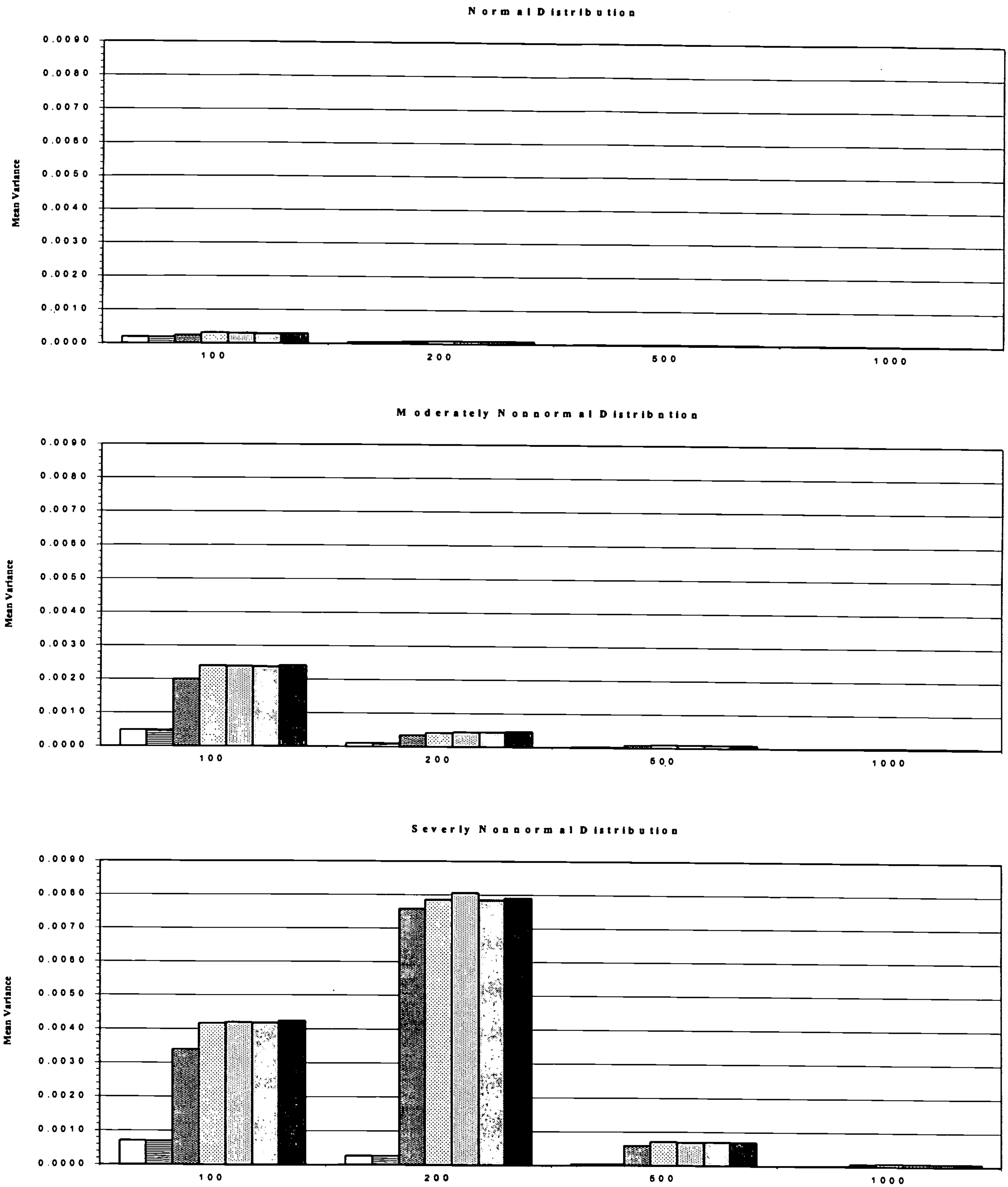


Figure 2. Bias in Cov(F2,F1) Standard Errors



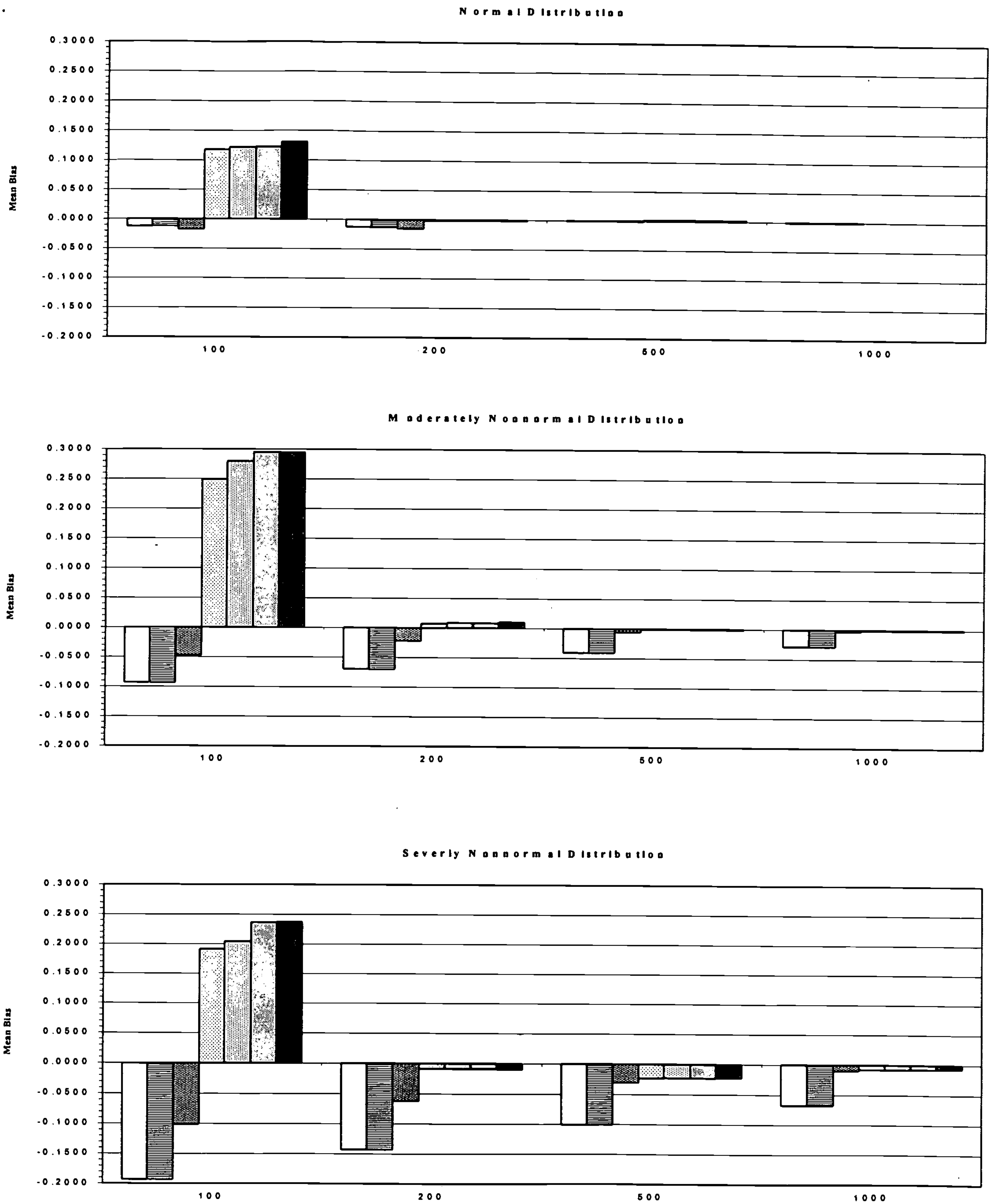
Note: Bar ordering within each cluster (from left to right): ML – EQS, ML – AMOS, rescaled, boostrap:250, bootstrap:500, bootstrap:1000, bootstrap:2000

Figure 3. Variance in Cov(F2,F1) Standard Errors



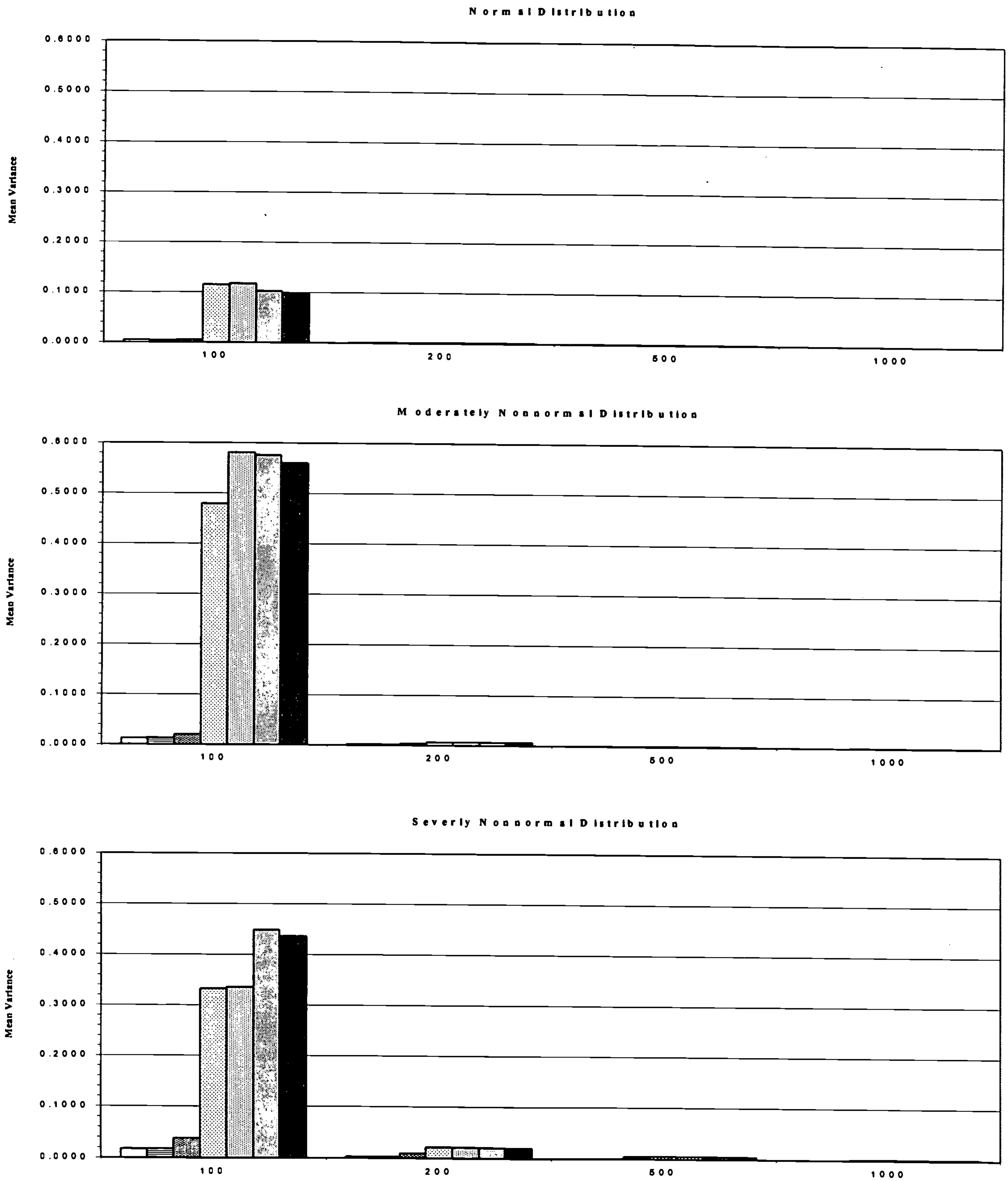
Note: Bar ordering within each cluster (from left to right): ML – EQS, ML – AMOS, rescaled, boostrap:250, bootstrap:500, bootstrap:1000, bootstrap:2000

Figure 4. Bias in Standard Errors for Variable-Factor Loading of V2 on F1



Note: Bar ordering within each cluster (from left to right): ML – EQS, ML – AMOS, rescaled, bootstrap:250, boostrap:500, bootstrap:1000, bootstrap:2000

Figure 5. Variance in Standard Errors for Variable-Factor Loading of V2 on F1



Note: Bar ordering within each cluster (from left to right): ML – EQS, ML – AMOS, rescaled, bootstrap:250, boostrap:500, bootstrap:1000, bootstrap:2000

Appendix

Data for Figures

Bias in Standard Errors for Cov(F2,F1)

Model	Distribution	Sample Size	EQS ML	AMOS ML	rescaled	bootstrap 250	bootstrap 500	bootstrap 1000	bootstrap 2000
1	1	100	-0.00096215	-0.00172600	-0.00204793	0.00630900	0.00570400	0.00557400	0.00564400
1	1	200	0.00058373	0.00031700	-0.00004655	0.00188700	0.00202200	0.00203700	0.00193700
1	1	500	0.00070359	0.00060300	0.00077003	0.00119300	0.00133800	0.00134800	0.00123800
1	1	1000	0.00009617	0.00005600	0.00003354	0.00016100	0.00023100	0.00017600	0.00016600
1	2	100	-0.02382435	-0.02453000	-0.01606580	-0.00630500	-0.00669500	-0.00686000	-0.00659000
1	2	200	-0.01178749	-0.01205200	-0.00692223	-0.00450700	-0.00431700	-0.00424700	-0.00416700
1	2	500	-0.00672598	-0.00677300	-0.00177072	-0.00088300	-0.00091800	-0.00084800	-0.00089300
1	2	1000	-0.00350643	-0.00352900	0.00079721	0.00113600	0.00110600	0.00108600	0.00104600
1	3	100	-0.03330533	-0.03405200	-0.02020035	-0.00986200	-0.00965700	-0.00953700	-0.00926200
1	3	200	-0.02393200	-0.02415400	-0.00998588	-0.00590400	-0.00606900	-0.00626400	-0.00624900
1	3	500	-0.01255375	-0.01261100	-0.00256341	-0.00170600	-0.00154100	-0.00154600	-0.00146100
1	3	1000	-0.00920604	-0.00923600	-0.00171175	-0.00134600	-0.00134100	-0.00130600	-0.00129100

Variance in Standard Errors for Cov(F2,F1)

Model	Distribution	Sample Size	EQS ML	AMOS ML	rescaled	bootstrap 250	bootstrap 500	bootstrap 1000	bootstrap 2000
1	1	100	0.00019201	0.00018850	0.00023599	0.00031069	0.00030046	0.00029382	0.00029536
1	1	200	0.00005469	0.00005420	0.00006549	0.00008049	0.00008119	0.00008082	0.00007948
1	1	500	0.00000727	0.00000747	0.00001072	0.00001348	0.00001219	0.00001177	0.00001136
1	1	1000	0.00000224	0.00000235	0.00000293	0.00000464	0.00000454	0.00000409	0.00000342
1	2	100	0.00048020	0.00047046	0.00199199	0.00239668	0.00238504	0.00237080	0.00241051
1	2	200	0.00009937	0.00009784	0.00033965	0.00040968	0.00043437	0.00043779	0.00045864
1	2	500	0.00001730	0.00001729	0.00007947	0.00009065	0.00008560	0.00008613	0.00008743
1	2	1000	0.00000392	0.00000393	0.00002012	0.00002045	0.00002059	0.00002142	0.00002170
1	3	100	0.00070263	0.00069011	0.00339243	0.00416641	0.00419348	0.00417472	0.00423503
1	3	200	0.00026886	0.00026649	0.00758154	0.00784907	0.00805091	0.00783766	0.00790403
1	3	500	0.00003655	0.00003627	0.00058634	0.00070624	0.00068324	0.00069699	0.00068925
1	3	1000	0.00001009	0.00000988	0.00006769	0.00007658	0.00007496	0.00007439	0.00007413

Note: Model 1 = properly specified model. Distribution 1 = Multivariate Normal.
Distribution 2 = Moderately Nonnormal. Distribution 3 = Severely Nonnormal.

Appendix (continued)

Data for Figures

Bias in Standard Errors for Factor Loading of V2 on F1

Model	Distribution	Sample Size	EQS ML	AMOS ML	rescaled	bootstrap 250	bootstrap 500	bootstrap 1000	bootstrap 2000
1	1	100	-0.01265215	-0.01199400	-0.01747716	0.11758100	0.12234100	0.12268600	0.13069100
1	1	200	-0.01252441	-0.01250600	-0.01523267	-0.00153100	-0.00143100	-0.00125600	-0.00129100
1	1	500	-0.00055298	-0.00050600	-0.00108783	0.00178900	0.00197900	0.00193400	0.00187400
1	1	1000	-0.00072149	-0.00072900	-0.00083724	0.00007600	-0.00008400	-0.00002900	0.00000600
1	2	100	-0.09349429	-0.09342200	-0.04756999	0.24904300	0.27969300	0.29544300	0.29510300
1	2	200	-0.06977493	-0.06978900	-0.02174037	0.00712600	0.00838100	0.00828100	0.00992600
1	2	500	-0.04037575	-0.04036500	-0.00541728	-0.00122500	-0.00077500	-0.00083500	-0.00112000
1	2	1000	-0.02864266	-0.02865800	-0.00254852	-0.00115300	-0.00110800	-0.00096800	-0.00096800
1	3	100	-0.19344861	-0.19296500	-0.10122301	0.19106000	0.20389000	0.23676000	0.23727500
1	3	200	-0.14320653	-0.14317700	-0.06195104	-0.00882200	-0.00866700	-0.00906200	-0.00973700
1	3	500	-0.10070715	-0.10068900	-0.03052504	-0.02304400	-0.02249900	-0.02286400	-0.02282400
1	3	1000	-0.06749721	-0.06748000	-0.00959810	-0.00780000	-0.00763500	-0.00741500	-0.00719500

Variance in Standard Errors for Factor Loading of V2 on F1

Model	Distribution	Sample Size	EQS ML	AMOS ML	rescaled	bootstrap 250	bootstrap 500	bootstrap 1000	bootstrap 2000
1	1	100	0.00422598	0.00419741	0.00474683	0.11458234	0.11670930	0.10229007	0.09709874
1	1	200	0.00066588	0.00066763	0.00074157	0.00130829	0.00126159	0.00123686	0.00120367
1	1	500	0.00012641	0.00012601	0.00013353	0.00019758	0.00018310	0.00017339	0.00017208
1	1	1000	0.00003563	0.00003556	0.00003958	0.00005429	0.00005051	0.00004663	0.00004674
1	2	100	0.01261960	0.01270140	0.01995371	0.47906006	0.58104858	0.57607383	0.56010433
1	2	200	0.00154965	0.00154701	0.00325130	0.00600427	0.00604097	0.00595211	0.00679440
1	2	500	0.00020045	0.00020075	0.00067193	0.00074593	0.00073917	0.00072553	0.00070755
1	2	1000	0.00005881	0.00005922	0.00020382	0.00023683	0.00022426	0.00021958	0.00021985
1	3	100	0.01704205	0.01709042	0.03751906	0.33241595	0.33525167	0.44857189	0.43593636
1	3	200	0.00253971	0.00254571	0.00952510	0.02109303	0.02100801	0.02043223	0.01974029
1	3	500	0.00044802	0.00044688	0.00483715	0.00472887	0.00485844	0.00471300	0.00480366
1	3	1000	0.00010132	0.00010231	0.00109147	0.00115449	0.00114919	0.00112388	0.00113148

Note: Model 1 = properly specified model. Distribution 1 = Multivariate Normal.
Distribution 2 = Moderately Nonnormal. Distribution 3 = Severely Nonnormal.

TM 028433

ERIC

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Relative Performance of Rescaling and Resampling Approaches to Model χ^2 and Parameter Standard Error Estimation in Structural Equation Modeling	
Author(s): Jonathan Nevitt / Gregory R. Hancock	
Corporate Source: University of Maryland, College Park	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☒	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Jonathan Nevitt	Printed Name/Position/Title: Jonathan Nevitt / Graduate Student	
Organization/Address: 7656 Stony Creek Lane Ellicott City, MD 21043	Telephone: (410) 796-3720	FAX:
	E-Mail Address: jnevitt@wam.umd.edu	Date: 4/15/98

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com