DOCUMENT RESUME

ED 293 869 TM 011 449

AUTHOR Baldwin, Beatrice
TITLE The Effects of Model Misspecification and Sample Size on LISREL Maximum Likelihood Estimates.
PUB DATE Nov 87
NOTE 29p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (16th, Mobile, AL, November 11-13, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Computer Software; *Maximum Likelihood Statistics; Monte Carlo Methods; Predictive Measurement; Sample Size; *Statistical Analysis
IDENTIFIERS Error Analysis (Statistics); *LISREL Computer Program; Specification Error

ABSTRACT

The robustness of LISREL computer program maximum likelihood estimates under specific conditions of model misspecification and sample size was examined. The population model used in this study contains one exogenous variable; three endogenous variables; and eight indicator variables, two for each latent variable. Conditions of model misspecification included errors of omission of structural paths and simultaneous errors of omission and inclusion. All misspecifications were examined with sample sizes of 100 or 200. In general, the effects of omitting a path from one exogenous variable to another were less serious than were those of omitting a path from one endogenous variable to another. Adding a path from one endogenous variable to another endogenous variable affected the estimation of parameters and standard errors less than any other misspecification. Overall, however, few blanket recommendations as to classes of errors most specifically affecting robustness of parameter estimates could be made. This study confirms that LISREL-type modeling should be undertaken only when there is guiding substantive theory. (SLD)

THE EFFECTS OF MODEL MISSPECIFICATION AND SAMPLE SIZE ON LISREL MAXIMUM LIKELIHOOD ESTIMATES

Beatrice Baldwin
Southeastern Louisiana University

Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Mobile, AL, November 1987

THE EFFECTS OF MODEL MISSPECIFICATION AND SAMPLE SIZE ON LISREL MAXIMUM LIKELIHOOD ESTIMATES

Statement of the Problem

Covariance structure modeling using the LISREL program requires the researcher to make certain fundamental assumptions about the nature of the available data, the representativeness of the sample, and the plausibility of the theoretical model being tested. The statistical processes of specification, estimation, and testing of hypothetical models presuppose important statistical and theoretical conditions, and violations of these assumptions may seriously jeopardize the consistency of results. Such consistency, as it relates to the reliability of parameter estimates and test statistics, may also be termed robustness.

The robustness of LISREL to violations of assumptions should be known so that applied researchers can make more appropriate use of this sophisticated data analysis tool. The appropriateness of the LISREL model and the accuracy of results when known violations occur are subject to question. Most problems regarding robustness have largely been ignored by applied researchers, though not by choice. Because LISREL is a relatively new statistical technique, statistical researchers have not yet been able to answer most questions regarding the robustness of LISREL. Some recent Monte Carlo studies have examined the effects of using discrete variables or variables with skewed distributions; others have attempted to systematically review the effects of sample size and model misspecification. However, much work remains before statisticians or applied researchers can feel reasonably sure that LISREL results are robust to assumption violations or that violations will consistently distort results in known and predictable ways.

The purpose of this study is to examine the robustness of LISREL maximum likelihood estimates under specific conditions of model misspecification and sample size. The conditions of model misspecification include errors of omission of structural paths, errors of inclusion of structural paths, and simultaneous errors of omission and inclusion of structural paths. All misspecifications are examined under sample sizes of 100 and 200. By examining the values of parameter estimates and comparing them to the population values, we have more specific information about how such factors affect parameter estimates in applied research situations. Behavorial scientists need this information in order to make more informed decisions about model specification and its relation to substantive theory.

Research Questions

The results of previous simulation studies and the substantive knowledge available on the nature of model specification lead to several questions:

(1) Are certain types of specification error more serious in terms of parameter bias and/or model fit? This question has never been addressed adequately, although Gallini (1983) presented some preliminary observations and conclusions in the study of common specification errors in path analysis.

(2) Are compound specification errors more likely to lead to parameter bias than are single errors of omission or inclusion? MacCallum (1986) has attempted to gauge the effects of compound errors. This question, however, needs to be answered more systematically by developing a typology of

possible errors and testing the effects of such errors on the same structural model.

(3) Are sample sizes of 100 more likely to lead to parameter bias than are samples of 200? Boomsma (1983) considered the effects of sample size. Much of this work, however, dealt with the robustness of factor loadings. The same question in regard to the robustness of structural parameter estimates needs to be answered.

(4) Are sample sizes of 100 more likely to lead to problems with Heywood cases or nonconvergence of solutions than are samples of 200? This issue was also examined in the Boomsma (1983) study. The present study examines this problem more thoroughly for structural equation models.

(5) When the measurement model is not misspecified, will factor loadings be consistent from model to model? This question has not been specifically addressed in any previous study.

The Concept of the True Model

This study attempted to answer two questions: (1) how does sample size affect the maximum likelihood parameter estimates produced by a LISREL-type model; and (2) how do specific instances of structural model misspecification affect the parameter estimates. Robustness studies of this type are commonly made using a Monte Carlo method. The main advantage of this empirical method is that the true distribution is known, not assumed, as in most analytical methods of study (Hatch & Posten, 1966). The researcher knows the true probability distribution *a priori* because the researcher is free to specify the distribution and take samples from it (Acito & Anderson, 1984) Thus Monte Carlo methods involve the generation and analysis of artificial data.

The probability distribution from which to take samples is determined by the problem at hand. In this case, LISREL structural equation modeling requires a multivariate normal distribution. As it was not the aim of this study to examine the effects of non-normality or the effects of using categorical data, the indicator variables were assumed to follow a multivariate normal distribution.

Under typical applications of the LISREL model, a raw data matrix of size N X k would be obtained and a sample covariance or correlation matrix would be derived and used as input to the LISREL program. In the Monte Carlo study, the generation of raw data is unnecessary. Once the true model has been specified, true parameter values are determined. These parameter values are supplied to a computer program, and the population covariance matrix Σ is generated.

If we regard the finite population matrix Σ as a sample covariance matrix and analyze it using maximum likelihood estimation, the estimated parameter values would be exactly equal to the true parameter values. In fact, regardless of the estimation procedure used, the obtained solution would be identical to the true solution, and all solutions would have a perfect fit to the data with a chi-square of zero (Joreskog & Sorbom, 1984).

Sampling theory tells us that a sample taken from a population (known or unknown) can yield an estimate that is either close to the population value or widely discrepant from it due to sampling variability. In Monte Carlo studies, the population values are known, and subsequent sampling produces estimates within the wide range implied by the population. Thus a specific sample may yield covariances that are close to or widely deviant from population covariances, but on the average over numerous samples, the average sample values should approximate true population values (Hammerley & Handscomb, 1964).

The notion of sampling brings up two other questions: (1) what is the optimal sample size N; and (2) how many samples or number of replications NR should be generated. Sample size affects the generation of sample estimates as it is known from sampling theory that larger samples have a higher probability of yielding sample values which are close to the population values. Also the number of replications has an effect on this probability. Any one replication regardless of sample size may have sample specific characteristics. Thus multiple replications are necessary in Monte Carlo research.

The optimal sample size question is answered by consideration of past research. The LISREL likelihood ratio test assumes a large sample size and yet a definition of "large" was not examined until the work of Boomsma (1982a, 1982b, 1983). Boomsma initially experimented with models using sample sizes of 25, 50, 100, 200, 400, and 800. Analysis of his first model revealed that serious convergence problems often resulted from using sample sizes smaller than 100. For samples of size N= 100, convergence problems and improper solutions were fewer, but the distributions of sample parameter estimates were not normal. In subsequent analyses, the samples using N=800 were also dropped as estimates and chi-square statistics were not improved by the use of this sample size. The sample size N= 200 is a reference point established by Boomsma and considered in most Monte Carlo studies thereafter (Ethington, 1985; Gallini & Mandeville, 1983; Gerbing & Anderson, 1985; MacCallum, 1985).

The number of replications is also a primary issue that was first examined by Boomsma (1983). The number of replications is a determining factor in establishing the accuracy of sample parameters, standard errors, and chi-square statistics. As with sample size, more usually means better in terms of estimating population values. Monte Carlo studies however are often limited by the realities of using extensive computer time, the cost of such time, and the sheer difficulties of handling the massive amounts of numbers produced by a Monte Carlo study. For instance, Boomsma (1983, p. 46) states that use of a 99% confidence level would require 6643 replications. If a model were estimated that contained 20 parameters, producing the estimates, the standard errors, the t-values, the modification indices, the variances of the estimates, and the chi-square statistics would require handling 101 pieces of information for each replication or 670,943 numbers for all 6643 replications of one model. The amount of information is even more staggering when one considers that most studies involve comparisons of many models.

Boomsma first used 100 replications for his initial work. After studying the results, the number of replications was increased to 300. The demands of computer time, storage of information, and amount of information were heavy but not unreasonable. The results were greatly improved by using 300 replications instead of 100 as the standard errors were reduced by half. Therefore for this study 300 replications were used for each model tested.

Model Description

The population model used for this study was designed so as to represent a number of structural specifications commonly found in applications. Such specifications could then be manipulated in the models to be tested, and the effects on the models examined. The true model is presented in Figure 1, and the population parameters are given in Table 1.

The population model contains one exogenous variable, three endogenous variables, and eight indicator variables, two for each latent variable. Although the use of three indicator variables per latent variable would reduce the parameter bias that sometimes occurs when two or fewer indicator variables are used, it is useful to study the behavior of parameter estimates under the worst possible conditions. Also, the use of two indicator variables per latent variable is not uncommon in applied research (Gerbing & Anderson, 1985; Boomsma, 1986). For the present study, all model misspecification occurs in the structural model; the measurement model remains constant and consists of indicators generated from a multivariate normal distribution.

The following nine possible types of structural model misspecification were studied:

1. Errors of omission

   A. Omitted path from an exogenous variable to an endogenous variable (which now becomes an exogenous variable),

   B. Omitted recursive path from an endogenous variable to an endogenous variable, and

   C. Omitted non-recursive path between endogenous variables.

2. Errors of inclusion

   A. Included path from an exogenous variable to an endogenous variable,

   B. Included recursive path from an endogenous variable to an endogenous variable, and

c. Included non-recursive path between endogenous variables.

3. Simultaneous errors of omission and inclusion

a. Omitted path from an exogenous variable to an endogenous variable (which now becomes an exogenous variable) and included path from an exogenous variable to another endogenous variable,

b. Omitted recursive path from an endogenous variable to an endogenous variable and included recursive path from an endogenous variable to another endogenous variable, and

c. Omitted non-recursive path between endogenous variables and included a different non-recursive path between endogenous variables.

Generation of Sample Covariance Matrices and Estimation of the Misspecified Models

Using the assigned true parameter values, the model was specified and the population covariance matrix generated using a SAS PROC MATRIX program. The population covariance matrix Σ is used to generate 300 sample covariance matrices S for each model to be tested. The sample covariance matrices were produced by using a FORTRAN Wishart variate generator program (Smith & Hocking, 1972). This FORTRAN routine generates a sample covariance matrix from a multivariate normal population with mean vector $\underline{0}$ and the specified sample size.

LISREL programs were written for each of the nine models to be tested. In addition, each model was tested under two sample sizes, N= 100 and N=200. Thus 18 combinations of model misspecification and sample size were tested for a minimum total of 5400 replications over the entire study. The LISREL program computes starting values using an instrumental variables method (non-recursive models) or a least squares method (recursive models). Use of these starting values effectively cuts down on the computer time required for estimation and aids in reaching a convergent solution within the 250 iteration limit imposed by the LISREL program.

The relevant output from the program runs consisted of the maximum likelihood parameter estimates, the standard errors of the estimates, modification indices for all parameters that were not being estimated, and the chi-square goodness-of-fit value with the associated degrees of freedom. Means and average standard errors of parameter estimates were computed across replications for each combination of model and sample size N.

Assessment of the results was based on the following criteria:

(1) Average parameter estimates for each model and sample size combination across replications

A. Bias of sample estimates: Is the average sample parameter estimate different from the actual parameter value? This relative difference was judged by computing a difference statistic ω_d which is $[(\omega - \bar{\omega}) / \omega] \times 100$.

B. Standard errors of sample estimates: Is the root mean square error (RMSE) of a parameter estimate different from the expected standard error? The RMSE is an unbiased estimate of the average standard error for a restricted sample size. This statistic is the square root of the average uncorrected sum of squares for a parameter estimate. The actual standard errors would provide a biased estimate if averaged. The expected standard error is the standard error when Σ is used for a specified sample size. This relative difference was judged by computing a difference statistic se_d which is $[(RMSE - se)/ se] \times 100$.

(2) Average chi-square across replications: Would a misspecified model still be considered a good fit? What is the rate of rejection for misspecified models?

(3) Modification indices for errors of omission

A. Average modification index for a particular error.

B. Percentage of cases in which the index is highest for the misspecification made: Does the modification index correctly indicate the model adjustment to be made?

(4) T-values for errors of inclusion

A. Average t-value for error: Is the t-value significant?

B. Percentage of cases in which the t-value is insignificant: Does the t-value correctly indicate that the parameter should be set equal to zero?

Improper Solutions and Nonconvergence

In addition to analyzing information about model misspecification, parameter bias, and resultant goodness-of-fit, it is also of interest to consider the occurrences of improper solutions and nonconvergence in a Monte Carlo study. Improper solutions result when maximum likelihood estimates of variances are nega . These negative variances indicate that the solution is unstable. Nonconvergence was defined as the inability of the program to find a unique solution which meets the convergence criteria within 250 iterations. Often it is uncertain whether raising the maximum number of iter ions will lead to a final solution. In cases of Monte Carlo study, it is often more efficient to simply terminate the program after 250 iterations (Boomsma, 1982a, 1985). Such was the case in the present study.

Negative estimates of variances or Heywood cases are problematic in that the solution is suspect. In Monte Carlo research however, the solutions are often regarded as plausible and the parameter estimates, standard errors, and chi-square statistics are analyzed as for admissible solutions (Boomsma, 1982, 1985; Rindskopf, 1984). Gerbing & Anderson (1985) disagree, and have emphasized that inclusion of improper solutions may lead to problems of interpretation and additional bias.

For this study, improper solutions were included in the analysis unless the improper solutions represented a sizeable percentage (over 10%, of the replications for any one model. It was planned that if some particular models had overly numerous improper solutions, the bias of parameter estimates would be calculated twice-- once with the improper solutions included and once with the improper solutions excluded. In this way, the influence of improper solutions on Monte Carlo results could be examined more thoroughly. These measures were not needed, however, as none of the misspecified models produced more than 10% of improper solutions.

A more serious problem in Monte Carlo research is that of nonconvergence. In general, nonconvergence problems are most often associated with small sample sizes. Because this study involved sample sizes of 100 and 200, it was expect d that nonconvergence problems would be infrequent. Since the solution in a nonconvergent LISREL analysis may deviate widely from a true solution, nonconvergent solutions were not included in the analysis. Any solutions that were nonconvergent were discarded and another computer analysis with a new random sample was used to take its place. In other words, there were at least 300 converged replications for each model tested.

Analysis

Each case of model misspecification was studied for each of two sample sizes, N=100 and N=200. Thus 18 cc ibinations of model misspecification and sample size were tested with at least 300 replications per combination. The program output was compiled and the PROC UNIVARIATE procedure of SAS was used to tabulate average parameter estimates, root mean square errors, average modification indices, and the frequencies of estimates across replications and within each combination. T-values were determined, and the relative difference statistics ω_d and se_d were calculated. These data provide information about the effects of model misspecification, parameter estimate bias, and the resultant goodness-of-fit of the model. In addition, the frequency of improper solutions and nonconvergence were examined.

Summary of the Results

In order to compare the rates of nonconvergence and improper solutions across models and sample sizes, Table 2 presents an overview of these results. In general, the incidence of nonconvergence and Heywood cases was the same for both sample sizes. However, for Models 1A, 2B, and 3B the difference in the rates of nonconvergence was larger. For Model 1A, a sample size of 200 produced 7% nonconvergent solutions as opposed to 13% for N=100. For Models 2B and 3B, the rates were 4% for N=200 and 13% for N=100. Models 1B, 2C, and 3A had the highest rates of nonconvergence- 49%, 37%, and about 50% respectively. There were no models in which the rates for improper solutions was higher than 10%.

Tables 3, 4, and 5 present summaries of the results for Models 1, 2, and 3. These tables give qualitative information on the relative performance of the models for each of the sample sizes considered. These tables are based on similar tables found in Boomsma (1983). Part I of each table indicates the degree of bias for the parameter estimates and standard errors. Part II indicates the degree of departure from optimal performance. For example, if a misspecified model had a high rate of acceptance and an average x^2 value less than the critical value, such performance could not be considered optimal as the goodness-of-fit is misleading.

Discussion of the Results

The purpose of this study was to examine the robustness of LISREL maximum likelihood parameter estimates under specific conditions of model misspecification and sample size. These conditions of model misspecification are errors of omission of structural paths, errors of inclusion of structural paths, and simultaneous errors of omission and inclusion of structural paths. By examining the values of the parameter estimates and comparing them to the population values, we have some specific information about how such factors may affect parameter estimates in applied research situations. In addition, this study has investigated the effects of model misspecification and sample size on the estimates of standard errors, the t-values, the modification indices, the goodness-of-fit of the model, and the frequency of nonconvergent and improper solutions. These results will similarly provide us with information that may aid researchers in making informed decisions with regard to the theoretical specification of LISREL models.

Certain types of specification errors seem to be more serious in terms of parameter bias and/or model fit In general, the effects of omitting a path from an exogenous variable to an endogenous variable seem to be less serious than the effects of omitting a path from an endogenous variable to another endogenous variable. The estimation of parameters and standard errors was affected much less by the former error than by the latter. The marked bias of the structural parameter estimates when a path from one endogenous variable to another endogenous variable is omitted should be noted. Nonconvergence problems were also more frequent when such a path was omitted. The x^2 measures for goodness-of-fit appropriately indicated a lack of congruence between the data and the theoretical model except in the case of Model 1B with N=100

Adding a path from one endogenous variable to another endogenous variable affects the estimation of parameters and standard errors less than for any other misspecification, regardless of sample size. However, the misspecified model would be accepted as a good fit with a very high probability. The addition of a path from an exogenous variable to an endogenous variable presented problems only for the smaller sample size.

Omitting a single reciprocal path is also not serious. However, adding a reciprocal path when another reciprocal path already exists in the model seems to present severe estimation problems for the standard errors. This is probably due to the incorrect partioning of direct and indirect effects. Adding a reciprocal path to a model that has no other non-recursive path may not have the same effects as were noted in this study.

Simultaneous errors from an exogenous variable to an endogenous variable seem to be more problematic than single errors of omission or inclusion. The structural parameter estimates and the standard errors are more likely to be quite different from the population values. The problems of nonconvergence were also much more severe for the model with simultaneous errors.

Model goodness-of-fit was markedly affected for Models 3B and 3C in that these models fail to yield accurate results concerning goodness-of-fit, and the correct respecifications cannot be discerned from the modification indices and t-values. Moderate bias was noted for the parameter estimates.

As was expected, the factor loadings remained consistent from model to model with little or no bias detected. This finding confirms the fact that changes in the structural model have few effects on the measurement model.

Nonconvergence problems seem to be related to the type of error made in relation to the overall pattern of the true model. For example, the addition of a structural path from an endogenous variable to another endogenous variable presented numerous nonconvergence problems. One would expect that the model containing simultaneous errors might likewise be affected, but this was not the case. The particular problems for this misspecification may be model-specific, that is, due to the repartitioning of direct and indirect effects in contrast to such effects in the true model. Of the four latent variables in the model, three are endogenous. There is a certain dynamic to the flow of direct and indirect effects in this model. For the case in which reciprocal paths are present, an almost circular flow of effects could be assumed. Omitting a "major" path in the model may restrict this dynamic flow. Thus the nonconvergence problems may be more indicative of model-specific tendencies than due to the deletion of a particular type of path.

Sample size seemed to be a minor issue in many of the models studied if the models are examined on a case-by-case basis. There seems to be a general "rule of thumb" in that those models which fit well and are relatively "easy" to estimate can be estimated well whether the sample size is 100 or 200. On the other hand, models which have convergence problems or which do not fit well will usually have similar results for the different sample sizes as well. This does not mean, however, that a sample size of 100 necessarily gives the same results as a sample size of 200. In particular, the standard errors seem to be greatly affected by sample size. This was particularly true for Models 1A and 1C. Sample estimates of standard errors were much closer to population values when N=200. Even when the parameter estimates themselves were relatively unbiased for N=100, comparisons to the results of N=200 show that these estimates are even more accurate for the larger sample size. The incidence of rejecting a misspecified model also seems to improve when the sample size is increased. T-values more adequately reflect the inclusion of extraneous paths, and the modification indices more accurately flag those paths which should be included in the model.

Overall, few blanket recommendations as to the classes of specification errors which most seriously affect the robustness of parameter estimates can be made. From a theoretical point of view, it might seem that adding a path to an otherwise correctly specified model would be the least innocuous of all errors. Inspection of the parameter estimates for the models tested would seem to support this notion. However, it must be pointed out that the goodness-of-fit for all errors of inclusion was very good, a bit of information that is quite misleading. In addition the standard errors for all of these models were moderately to strongly biased. Such bias can also affect conclusions as to the significance or non-significance of individual parameters.

Which errors are of the most consequence from an applied standpoint? Exclusion of β paths, inclusion of reciprocal paths in models that already contain a reciprocal path, and multiple errors of any sort seem to have the most serious consequences for causal modeling. Such errors bias structural parameter estimates and severely distort standard errors. Recovery from such misspecifications is doubtful. T-values will not be reliable, modification indices cannot be guaranteed, and model goodness-of-fit is also affected.

The results of this study support some conclusions from past research. The recommendation by Boomsma (1983) to use sample sizes larger than 100 is supported. Although the average χ^2 statistics did not indicate any overall improvement from N=100 to N=200, the range of χ^2 values is much smaller for N=200; thus, the probability of obtaining a reliable test statistic is much improved for the larger sample size. As discussed previously, sample size has a definite effect on the estimation of parameters, standard errors, and modification indices.

MacCallum's (1985) research showed that models with one misspecification error often were not rejected. This research demonstrates that models containing errors of inclusion are particularly at high risk of being erroneously accepted as plausible. On the other hand, models containing errors of omission may be rejected or may fail to be rejected depending upon the type of path omitted. Contrary to MacCallum's study, two misspecification errors do not necessarily increase the chances of rejecting a model. The types of errors made seem to be more important than the number of errors. For instance, for Model 3C which omitted and included reciprocal paths, the model was accepted as plausible in most replications. This was true despite the fact that structural parameters were biased.

Gerbing & Anderson's (1985) Monte Carlo study demonstrated that the variability of parameter estimates decreases as sample size increases. This finding was likewise supported in this study. Their findings are based on research investigating the effects of model characteristics on parameter estimates in confirmatory factor analysis. However, it makes intuitive sense that parameter estimates should likewise be affected in structural equation models.

Gerbing & Anderson also noted that sample size and the number of indicators per latent variable had large effects on the structural parameter φ which relates two factors. This study found a somewhat similar effect on the variance terms of the matrix Φ and on the disturbance terms of the matrix Ψ. By misspecifying the model and by limiting the number of indicators to two, estimation is hindered. The incidence of bias for the parameters in these two matrices was much higher, even in cases in which the bias of other parameters was limited, and the model otherwise performed optimally. This was particularly true for sample sizes of 100. The bias of these parameters indicates that the misspecification introduces a degree of uncertainty into the estimation procedure. Much of the variance which is otherwise accounted for in the population model cannot be explained in the misspecified model. Thus these terms are apt to be affected.

Implications for Educational Research

The development of LISREL-type structural equation models has decidedly influenced the direction of educational research in the pa t few years. Causal modeling techniques allow researchers to hypothesize about the complex relationships among theoretical variables in a manner that is not possible with path analysis or multiple regression.

The popular usage of any statistical technique leads many researchers to speculate on the utility of that technique. While the robustness of more traditional methodologies against violations of assumptions has been tested, the effects of such violations are not clear when using LISREL-type models. These procedures have only recently been under study.

This study adds to the literature by having examined the effects of model misspecification and sample size. Although the generalizations of the study are restricted due to the particular class of models and the particular types of misspecification examined, such generalizations may lead us to a more "informed guessing" of how such model misspecifications may affect results when working in an applications context. Although linear structural equation modeling is never recommended to be used in the total absence of substantive theory, the presence of equally plausible yet conflicting theories may result in various different model specifications for the same theoretical research question. Knowledge of how models behave under alternate misspecifications is a valuable asset in such situations.

In general, the results of this study strongly confirm the idea that LISREL-type modeling must be undertaken only when there is guiding substantive theory. In many cases, misspecified models were inaccurately described as having acceptable goodness-of-fit. Alternative models may have equally good fits. The ultimate decision to accept a model must lay with the researcher. The numbers themselves cannot be used as the sole criteria for judging the quality of a model.

References

Acito, F. & Anderson, R. D. (1984) On simulation methods for investigating structural modeling. Journal of Marketing Research, 21, 107-112.

Boomsma, A. (1982a). Facts and failures of Monte Carlo work with LISREL. Bulletin HB-82-577-EX, Heymans Psychological Institute, University of Groningen, The Netherlands.

Boomsma, A. (1982b). The robustness of LISREL against small sample sizes in factor analysis models. In K. Joreskog & H. Wold (Eds.), Systems under indirect observation. Amsterdam: North-Holland.

Boomsma, A. (1983). On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality. Amsterdam: Sociometric Research Foundation.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. Psychometrika, 50, 229-242.

Ethington, C. (1985). The robustness of LISREL estimates in structural equation models with categorical data. Unpublished dissertation, Virginia Polytechnical Institute and State University.

Gallini, J. (1983). Misspecifications that can result in path analysis structures. Applied Psychological Measurement, 7, 125-137.

Gallini, J. & Mandeville, G. K. (1983). An investigation of the effect of sample size and specification error on the fit of structural equation models. Journal of Experimental Education, 51, 9-19.

Gerbing, D. W. & Anderson, J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. Multivariate Behavorial Research, 20, 225-271.

Hammersley, J. M. & Handscomb, D. C. (1964). Monte Carlo methods. New York: John Wiley & Sons.

Hatch, L. O. & Posten, H. O. (1969). A quantitative approach to robustness. Research Report No. 42, Department of Statistics, University of Connecticut.

Joreskog, K. G. & Sorbom, D. (1984). LISREL VI User's Guide. Mooresville, Indiana: Scientific Software, Inc.

MacCallum, R. (1986). Specification searches in covariance structure modeling. Psychological Bulletin, 100, 107-120.

Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems. Sociological Methods & Research, 13, 109-119.

Smith, W. B. & Hocking, R. R. (1972). Wishart variate generator. Applied Statistics, 21, 341-345.

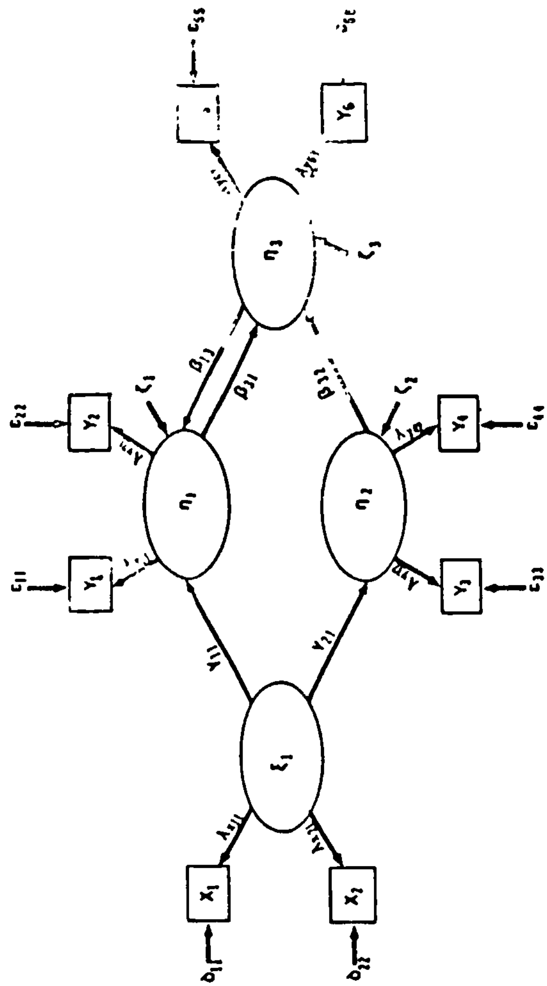

FIGURE 1. TRUE MODEL

PARAMETER	ω
λ_{y11}	1.0*
λ_{y21}	.9
λ_{y32}	1.0
λ_{y42}	.7
λ_{y53}	1.0
λ_{y63}	.8
λ_{x11}	1.0
λ_{x21}	.9
β_{13}	.3
β_{31}	.7
β_{32}	.5
γ_{11}	.6
γ_{21}	.8
ϕ_{11}	.9
ψ_{11}	.5
ψ_{22}	.2
ψ_{33}	.1
$\theta_{\epsilon 11}$	.1
$\theta_{\epsilon 22}$	.2
$\theta_{\epsilon 33}$	.1
$\theta_{\epsilon 44}$	.4
$\theta_{\epsilon 55}$	.1
$\theta_{\epsilon 66}$	.3
$\theta_{\delta 11}$	.1
$\theta_{\delta 22}$	.2

* factor loadings of 1.0 are fixed for manifest variables

TABLE 1. VALUES OF POPULATION PARAMETERS

	Model 1A		Model 1B		Model 1C	
N=	200	100	200	100	200	100
Part I.[a]						
Bias of factor loadings	-	-	-	-	-	-
Bias of structural parameters	-	-	**	**	-	*
Bias of phi or psi matrices	-	*	**	**	*	*
Bias of meas. error terms	-	-	-	-	-	-
Bias of standard errors	*	**	**	**	-	*
Part II.[b]						
Goodness-of-fit	-	-	-	*	**	**
Modification index for error	-	-	-	*	**	**
T-value for err -	NA	NA	NA	NA	NA	NA
Nonconvergence	*	*	**	**	-	-
Improper solutions	*	*	-	-	-	*

[a]Part I indicates the degree of bias from strongest (**) to no bias (-). The degree of bias was based on the number of parameters affected and the severity of the bias.

[b]Part II indicates degree of departure from optimal performance; it ranges from strongest departure (**) to no departure (-).

TABLE 2. SUMMARY OF RESULTS FOR MODEL 1

	Model 2A		Model 2B		Model 2C	
N:	200	100	200	100	200	100
Part I. [a]						
Bias of factor loadings	-	-	-		-	-
Bias of structural parameters	-	*	-	-	-	*
Bias of phi or psi matrices	-	*	-	-	*	*
Bias of meas. error terms	-	-	-	-	-	*
Bias of standard errors	**	**	*	*	**	**
Part II. [b]						
Goodness-of-fit	**	**	**	**	**	**
Modification index for error	NA	NA	NA	NA	NA	NA
T-value for error	-	-	-	-	-	*
Nonconvergence	*	*	-	*	**	**
Improper solutions	-	-	-	-	-	-

[a] Part I indicates degree of bias from strongest (**) to no bias (-). The degree of bias was based on the number of parameters affected and the severity of the bias.

[b] Part II indicates degree of departure from optimal performance; it ranges from strongest departure (**) to no departure (-)

TABLE 3. SUMMARY OF RESULTS FOR MODEL 2

	Model 3A		Model 3B		Model 3C	
N=	200	100	200	100	200	100
Part I.[a]						
Bias of factor loadings	-	-	-	-	-	-
Bias of structural parameters	*	*	**	**	*	*
Bias of phi or psi matrices	*	*	**	**	*	*
Bias of meas. error terms	-	*	-	-	-	-
Bias of standard errors	**	**	**	**	**	*
Part II.[b]						
Goodness-of-fit	-	-	**	-	**	**
Modification index for error	-	-	*	**	*	**
T-value for error	-	-	*	**	*	**
Nonconvergence	**	**	-	*	-	-
Improper solutions	-	-	-	*	-	-

[a] Part I indicates the degree of bias from strongest (**) to no bias (-). The degree of bias was based on the number of parameters affected and the severity of the bias.

[b] Part II indicates degree of departure from optimal performance; it ranges from strongest departure (**) to no departure (-).

TABLE 4. SUMMARY OF RESULTS FOR MODEL 3

	Total Solutions	Nonconvergence	Improper Solutions
Model 1A			
N:200	327	22(7%)	28(9%)
N:100	350	44(13%)	29(9%)
Model 1B			
N:200	597	291(49%)	7(3%)
N:100	600	291(49%)	9(3%)
Model 1C			
N:200	350	0(0%)	17(5%)
N:100	350	0(0%)	21(6%)
Model 2A			
N:200	339	27(8%)	10(3%)
N:100	350	26(7%)	17(5%)
Model 2B			
N:200	343	14(4%)	8(2%)
N:100	350	44(13%)	5(2%)
Model 2C			
N:200	480	176(37%)	0(0%)
N:100	487	187(37%)	0(0%)
Model 3A			
N:200	636	324(51%)	2(1%)
N:100	642	322(50%)	1(0%)
Model 3B			
N:200	351	13(4%)	6(2%)
N:100	372	50(13%)	25(8%)
Model 3C			
N:200	322	0(0%)	1(0%)
N:100	351	0(0%)	16(5%)

TABLE 5. IMPROPER SOLUTIONS AND NONCONVERGENCE