PSYCHOMETRIKA—VOL. 50, NO. 2, 229–242
JUNE 1985

NONCONVERGENCE, IMPROPER SOLUTIONS, AND STARTING VALUES IN LISREL MAXIMUM LIKELIHOOD ESTIMATION

ANNE BOOMSMA

UNIVERSITY OF GRONINGEN

In the framework of a robustness study on maximum likelihood estimation with LISREL three types of problems are dealt with: nonconvergence, improper solutions, and choice of starting values. The purpose of the paper is to illustrate why and to what extent these problems are of importance for users of LISREL. The ways in which these issues may affect the design and conclusions of robustness research is also discussed.

Key words: maximum likelihood estimation, LISREL, Davidon-Fletcher-Powell, starting values, nonconvergence, improper solutions, robustness, Monte Carlo, small sample results.

Introduction

In studies on the robustness of maximum likelihood estimation with LISREL, the effects of small sample size and those of nonnormal, discrete distributions were investigated (Boomsma, 1982, 1983). Three main conclusions from that research were made. (i) In LISREL modeling it is recommended not to use a sample size smaller than 100. (ii) LISREL is robust against symmetric, discrete distributions with normal kurtosis, but not against rather skewed, discrete distributions. (iii) It is not recommended to analyze correlation matrices instead of covariance matrices, because it may have serious effects on the estimated covariances of the parameter estimates.

The present paper deals with three problems encountered during the work just referred to: nonconvergence of the iterative maximum likelihood estimation, improper parameter estimates, and the choice of starting values in the estimation process.

The results discussed below indicate that users of LISREL should be aware of circumstances in which nonconvergence and improper solutions may occur, and their frequency of occurrence. They may also be interested in the effect of the starting values in iterative estimation. Some of the materials presented may serve as guidelines for researchers planning Monte Carlo studies. The emphasis in this paper will be on the practical implications of the findings for users of LISREL, illustrated by results from Monte Carlo work.

Monte Carlo design. The sampling design for the small sample part of our study is summarized as follows. In the LISREL framework for each specified model with a known population parameter vector $\boldsymbol{\omega}$, a population covariance matrix $\boldsymbol{\Sigma}$ is defined. These matrices $\boldsymbol{\Sigma}$ are the covariance structures of multivariate normal distributions from which independent samples were taken of size 25, 50, 100, 200, and 400. For each model and each *sample size N*, $\underline{\text{NRS}} \geq 300$ samples were taken (NRS = *number of replications in stock*). The samples were generated by using subroutine GGNMS from IMSL (1982) on a CDC Cyber 74/18 and a CDC Cyber 170/760 machine. Thus, for each model and each sample size, NRS *sample covariance matrices* S were obtained.

After this sampling process a LISREL analysis was done for each S until NR = 300 samples have led to a solution without numerical difficulties; for all *N* this *number of*

Requests for reprints should be sent to A. Boomsma, Vakgroep Statistiek en Meettheorie, Rijksuniversiteit Groningen, Oude Boteringestraat 23, 9712 GC Groningen, THE NETHERLANDS.

0033-3123/85/0600-5010$00.75/0

replications NR was fixed to 300. At the completion of these analyses for the parameters ω_i, their corresponding standard errors se_{ω_i}, as well as the likelihood ratio chi-square statistic for goodness-of-fit, 300 estimates were available. The Monte Carlo results indicate how closely the empirical sampling distributions follow known theoretical sampling distributions.

Model choice. In this paper two structural equation models (Models 1 and 2), and a number of factor analysis models (Models 3US through 4UL) are used. Model 1 (the stability of alienation) and Model 2 (peer influence on aspiration) are discussed by Jöreskog (1977, Figure IB & Table II). The population covariances chosen for these two models (Boomsma, 1983) are almost identical to the original sample covariances, also reproduced by Jöreskog and Sörbom (1981). From these references, it can be deduced that Model 1 has 6 observed variables, 17 parameters to be estimated, and 4 degrees of freedom; for Model 2 these numbers are 10, 17 and 17, respectively.

Twelve factor analysis models were studied, all having two factors. Each of these models is indexed by three symbols: (i) a 3 or 4, depending on the presence of three or four observed variables for each factor; (ii) a U or a C, depending on the size of the correlation ϕ between the two factors, chosen as 0 (Uncorrelated; ϕ fixed) or 0.3 (Correlated; ϕ free); (iii) an S, M or L, depending on the size of the factor loadings, chosen as Small (0.4, 0.6), Medium (0.6, 0.8) or Large (0.8, 0.9).

The factor pattern Λ ($k \times 2$), where k is the number of observed variables, was chosen such that half of the observed variables had a nonzero loading on the first factor and a zero loading on the second one, and the reverse for the other half. Let $\boldsymbol{\lambda} = [\lambda_i]$, $i = 1, \ldots, k$, denote the vector of nonzero loadings. Then $\boldsymbol{\lambda}'_{3US} = \boldsymbol{\lambda}'_{3CS} = (0.4\ 0.4\ 0.6\ 0.4\ 0.4\ 0.6)$, and $\boldsymbol{\lambda}'_{4US} = \boldsymbol{\lambda}'_{4CS} = (0.4\ 0.4\ 0.6\ 0.6\ 0.4\ 0.4\ 0.6\ 0.6)$. Analogously, $\boldsymbol{\lambda}'_{3UM} = \boldsymbol{\lambda}'_{3CM} = (0.6\ 0.6\ 0.8\ 0.6\ 0.6\ 0.8)$, and $\boldsymbol{\lambda}'_{3UL} = \boldsymbol{\lambda}'_{3CL} = (0.8\ 0.8\ 0.9\ 0.8\ 0.8\ 0.9)$.

Analytic details. In the study, starting some years ago, LISREL III was used. Today, most researchers have access to LISREL VI (Jöreskog & Sörbom, 1984) or LISREL V (Jöreskog & Sörbom, 1981), which we did not. It is not known to what extent the use of these latest versions would have changed the results obtained for the problems discussed here. Indications of no substantial differences between estimates from various LISREL versions are mentioned below.

All the results in this paper are based on the analysis of sample covariance matrices, not on those of sample correlation matrices.

The Problem of Nonconvergence

In LISREL the modified Davidon-Fletcher-Powell algorithm developed by Gruvaeus and Jöreskog (1970) has a convergence criterion EPS $= 0.5 \times 10^{-5}$. Under continuity conditions a solution converges if the absolute value of all derivatives of the function F to be minimized is less than EPS. For most practical problems the convergence criterion is met within the maximum of 250 iterations (MAXITE). In LISREL both EPS and MAXITE have a fixed value, which cannot be manipulated by the user.

Nevertheless, it may happen that the EPS-criterion is not met within 250 iterations. In our Monte Carlo work this led to specific problems. Because all those results are based on the algorithm just described, they need a conditional interpretation: for each N they are not based on strictly random samples $\mathbf{S}$ from $\boldsymbol{\Sigma}$, but on the first NR = 300 converging ones among a given number of random replications in stock (NRS). Since the asymptotic theory holds for the total set of random replications, the restriction to a subset of converging replications is theoretically unjustified. It could imply that one is dealing with a biased sample of size NR from a specified Wishart distribution. There is little choice, however, if one wants to study the robustness of LISREL with all its "restrictions" (EPS

TABLE 1.

The Percentage of Nonconvergence.
A * means NR=100, otherwise NR=300. A blank means 0%.

	sample size				
Model	25	50	100	200	400
1	22.1%	5.7%	0.7%		
2	6.5%	0.7%			
3US	46.7%	28.4%	12.8%	2.0%	
3CS	55.0%	35.3%	16.4%	2.9%	
3UM	11.2%	1.0%			
3CM	11.0%	1.3%			
3UL					
3CL	0.3%				
4US	27.0%*	8.5%	1.0%		
4CS	29.6%*	8.3%			
4UM	1.0%				
4CM	1.6%				
4UL					
4CL					

and MAXITE) imposed. Of course, the maximum number of iterations could have been raised, but it was uncertain whether extra iterations would quickly lead to a genuine maximum likelihood solution. This is the more questionable, since known population parameter values were used as starting points for the analyses.

Results on Nonconvergence

The phenomenon of nonconvergence is expressed in terms of the percentage of occurrence among NR replications. Recall that new replications were analyzed until 300 convergent ones were found (except for Models 4US and 4CS, where NR = 100). From Table 1 it can be seen that in two cases (Models 3US and 3CS) for $N = 25$ about 50% of the samples S did not converge within 250 iterations.

In general, nonconvergence decreases with sample size. For all models $N = 400$ was sufficient to raise no problems.

In comparing the different factor analysis models at least two other elements seem to influence the frequency of nonconvergence. (i) The size of the factor loadings (the population covariances). (ii) The ratio number of observed variables/number of factors (NV/NF).

Factor loading size. Nonconvergence increases if the population covariances get closer to zero (in Models 3US and 3CS, for variables linked to the same factor, covariances range from 0.16 to 0.24, in Models 4UL and 4CL from 0.64 to 0.81). A partial explanation of nonconvergence with these types of models can be given in terms of the sign pattern of the sample covariances of observed variables linked to the same factor. Inspection of nonconverging samples S may, for example, reveal that variable 1 has a positive sample correlation with variables 2 and 3, but variables 2 and 3 are negatively correlated, although in the population model all three variables are attached to the same factor. The general idea is that within such a set of variables the sign pattern of their covariances is incompatible with the signs of products of possible factor loadings in the population model. Random sampling may very well result in conflicts between the data and a perfectly specified model. For the frequency of such discrepancies the standard error of the product-moment correlation coefficient should also be taken into account.

In an attempt to predict convergence, using incompatible sign patterns of sample covariances both between variables 1, 2, 3 and between variables 4, 5, 6 as a criterion, in Model 3US the following results were found: for $N = 25$ (NR = 400) the correct-prediction rate of (non)convergence was 98%; for $N = 50$ (NR = 400) it was 99%. For Model 3CS ($N = 25$, 50; NR = 400) these percentages were 62 and 56, respectively. However, using similar prediction criteria for Model 4US was unsuccessful.

NV/NF ratio. Nonconvergence increases if the NV/NF ratio decreases: 8-variable two-factor models lead to more convergence than their 6-variable counterparts.

Most of the results reported here are in accordance with the findings of Anderson and Gerbing (1984), who conclude for a number of factor analysis models (NR = 100) "that the proportion of convergent solutions increased as: Sample size increased, the number of indicators per factor increased, loadings varied from mixed to all .6 to all .9, and factor correlations increased" (p. 162). Table 1 shows that for most factor analysis models convergence occurred somewhat less frequent when the two factors were correlated than when they were uncorrelated. Note, however, that in our orthogonal models ϕ was fixed to zero, while for the models used by Anderson and Gerbing ϕ had to be estimated.

The question can be posed whether convergence problems may be due to violations of the assumption that the function F to be minimized is continuously differentiable (Gruvaeus & Jöreskog, 1970, p. 1). It could be that difficulties of nondifferentiability occur if the rank of the model is close to the dimension of the covariance matrix. More specifically, it is of interest to know how often it occurs that two or more eigenvalues of S are almost equal. This could explain the model dependency of nonconvergence. However, inspection of the eigenvalues of converging and nonconverging samples gave no explanation along these lines.

Summary. The seriousness of the problem of nonconvergence depends heavily on the sample size, and may broadly vary with the model under study. A general and most direct way to avoid the problem is to have samples of at least moderate size, $N > 100$ say.

As indicated earlier, our results are based on the use of LISREL III. It should be noticed, however, that nonconvergence (or "serious problems during minimization") may also occur with LISREL V and VI. This conclusion is primarily based on two studies. First, a number of LISREL V analyses of our random samples (Model 3CS; $N = 25$) were kindly performed by Jöreskog. Compared with LISREL III the results were very much the same. Secondly, we analyzed 50 random samples (Model 3US; $N = 25$) with LISREL VI. No differences in convergence were found between LISREL III and VI. In comparing both versions using the same starting values, differences between parameter estimates were mostly in the fifth decimal place, which is due to the fact that EPS is 10 times as large in LISREL III as it is in LISREL VI. It thus seems that our simulation results would have been hardly different if most recent program versions would have been used.

In practice, one can always restart the iteration process by using the "solution" attained after the first 250 iterations as starting values in a continued analysis. For Model 3CS ($N = 25$) it appeared, however, that the discrepancy between the model under study and specific random samples can be so large that even with an *unrestricted* number of iterations, no convergence could be attained with LISREL V.

The Problem of Improper Solutions

A second but not less serious issue is that of negative estimates of variances (in factor analysis similar problems are known as "Heywood cases"). The occurrence of such improper solutions has been studied by Mattson, Olsson, and Rosén, (1966); see also Jöreskog (1967). Discussions of the problem in maximum likelihood factor analysis are given by Tumura and Fukutomi (1970), and Van Driel (1978). Here, the so-called ultra Heywood case is considered only; no specific treatment of boundary estimates (zero variances or exact Heywood cases) is given.

The LISREL algorithm does not handle Heywood cases by imposing constraints on the variances (see Jöreskog & Sörbom, 1984), although in principle methods like those developed by Lee (1980) could be used to modify an inconstrained LISREL program into a constrained one.

In LISREL modeling there are at least four strategies in dealing with sample covariance matrices leading to improper solutions. (i) Do not bother about them. This leaves the researcher with a problem of interpretation if the model would not be rejected. (ii) Given a number of negative estimates of variances for a specific model, fix the corresponding parameters to zero or to small positive values, and reanalyze the sample under the modified model. This practice is disputable, certainly if dependencies among parameter estimates are considered. (iii) Use clever model respecifications. Recent developments are presented by Rindskopf (1983, 1984), and Kelderman (in press). (iv) Instead of a maximum likelihood LISREL approach, choose for an alternative type of analysis, e.g., weighted least squares, and see whether that is helpful.

Each of these strategies has its disadvantages, not only for the regular user of LISREL, but also for designers of Monte Carlo research. First, results will be presented on the occurrence of improper solutions. In the next section two strategies are compared in a simulation study.

Results on Improper Solutions

Beforehand, it should be remarked that in our research improper solutions cannot be caused by misspecification of the model, because the population structures are fully known.

From Table 2 (Model 1; 8 estimates of variances) it follows that even for $N = 200$ improper solutions may occur.

Within a single replication more than one variance can have a negative estimate, which may or may not be due to dependencies between such estimates. This can be seen from the relative frequency distribution of the number K of negative estimates of variances within a single replication. Table 3 gives results for Model 1.

In Model 2 among 17 parameters 6 variances have to be estimated. The percentages in Table 4 show that there remains a serious danger of negative estimates for $N \leq 50$, but not for $N \geq 100$. In comparing Tables 2 and 3 it is noticed that the population values are closer to zero for Model 2, but the mean percentage of negative estimates for this model is smaller than for Model 1. In general, also given the results from Table 5, the range of population values does not offer a sufficient explanation here.

Only for $N = 25$ were solutions found with more than one out of six of these variances being negative: 2.7% of the replications had two improper estimates, 42.7% had only one negative estimate.

Table 5 shows global results for the twelve factor analysis models. Table 6 gives details of the specific variances in Model 3UM.

Summary. It is concluded that there is a real danger of improper solutions with small N. The frequency of its occurrence depends on a combination of three factors. (i) The sample size. More frequent with decreasing N. (ii) The population values of the variances. a) Across comparable models, e.g. the factor analysis models, more frequent with small population values of the variances (Van Driel, 1978, the Close to Zero case). b) Among variances within a single model, more frequent for those parameters with relatively small population values. (iii) The model under study. For example, more frequent in

TABLE 2.

The Percentage of Negative Estimates of Variances, and the Minimum of $\hat{\omega}_{ij}$. Model 1, NR=300. A blank means 0%.

	sample size										
	25		50		100		200		400		
prmt	%	min	%	min	%	min	%	min	%	min	ω_i
ψ_{11}	1.3	-5		1		2		2		3	5
ψ_{22}	3.7	-6	0.3	-1		1		2		2	4
$\theta^{\varepsilon}_{11}$	10.7	-26	4.0	-37	0.3	-1		2		3	5
$\theta^{\varepsilon}_{22}$	19.3	-31	13.7	-27	7.0	-13	1.7	-3		1	3
$\theta^{\varepsilon}_{33}$	18.3	-58	7.7	-13	1.0	-1		0		1	4
$\theta^{\varepsilon}_{44}$	15.3	-57	13.7	-9	5.0	-19	0.3	-1		1	3
θ^{δ}_{11}	24.7	-108	15.7	-140	9.3	-11	2.7	-2		0	3
θ^{δ}_{22}	7.3	-2396	1.3	-242	0.3	-71		124		194	265

the 6-variable factor analysis models than in 8-variable ones, which can also be formulated in terms of the NV/NF ratio.

Given these results, the user of LISREL should in general be urged to avoid samples of size $N \leq 50$.

Comparison of the Inclusion and Exclusion Strategy

The strategies mentioned previously were also of importance in considering the design of the Monte Carlo study and for evaluating results on the robustness of LISREL.

TABLE 3.

The Relative Frequency Distribution of the Number K out of 8 Variances with a Negative Estimate. Model 1, NR=300. A blank means 0%.

	sample size				
K	25	50	100	200	400
0	32.0%	60.0%	80.3%	95.7%	100%
1	41.3%	25.3%	16.3%	4.0%	
2	20.7%	13.0%	3.3%	0.3%	
3	6.0%	1.7%			
≥ 4					

TABLE 4.

The Percentage of Negative Estimates of Variances, and the Minimum of $\hat{\omega}_{ij}$. Model 2, NR=300. A blank means 0%.

	sample size										
	25		50		100		200		400		
prmt	%	min	%	min	%	min	%	min	%	min	ω_i
ψ_{11}	5.7	-0.2		0.0		0.1		0.1		0.2	.28
ψ_{22}	8.7	-0.2	0.7	-0.0		0.1		0.1		0.2	.27
$\theta^{\varepsilon}_{11}$	6.3	-2.1	0.7	-4.0		0.1		0.2		0.3	.42
$\theta^{\varepsilon}_{22}$	13.3	-2.8	6.0	-2.4		0.0		0.1		0.2	.33
$\theta^{\varepsilon}_{33}$	9.3	-12.4	1.0	-0.2	0.3	0.1		0.1		0.2	.30
$\theta^{\varepsilon}_{44}$	4.7	-1.7	0.3	-0.0		0.1		0.2		0.3	.41

In the following, two approaches in dealing with samples **S** leading to improper solutions are being compared. These strategies are (i) *include* replications with improper solutions (IN), and (ii) *exclude* replications with improper solutions (EX).

Theoretically the inclusion strategy seems to be the most attractive one, because it comes closest to a full set of random replications (the nonconvergent ones are excluded).

TABLE 5.

The Percentage of Improper Solutions in Factor Analysis Models. A * means NR=100, otherwise NR=300. A blank means 0%.

	sample size				
Model	25	50	100	200	400
3US	51.0%	41.3%	22.0%	10.7%	2.7%
3CS	47.3%	32.7%	18.3%	6.3%	2.7%
3UM	37.7%	21.0%	11.3%	1.3%	
3CM	43.7%	25.0%	5.7%	1.3%	
3UL	25.7%	9.0%		0.3%	
3CL	24.3%	7.3%	0.7%		
4US	47.0%*	19.3%	3.0%		
4CS	37.0%*	15.3%	3.7%	0.3%	
4UM	22.3%	5.0%			
4CM	27.3%	1.0%			
4UL	7.3%	0.3%			
4CL	7.0%				

TABLE 6.

The Percentage of Negative Estimates of Variances, and the Minimum of $\hat{\omega}_{ij}$. Model 3UM, NR=300. A blank means 0%.

	sample size										
	25		50		100		200		400		
prmt	%	min	%	min	%	min	%	min	%	min	ω_i
θ^{δ}_{11}	2.3	-0.7	0.3	-0.1		0.3		0.4		0.5	.64
θ^{δ}_{22}	5.7	-8.2	0.3	-0.1		0.3		0.4		0.5	.64
θ^{δ}_{33}	15.7	-14.7	11.0	-3.2	2.7	-0.4	0.7	-0.1		0.1	.36
θ^{δ}_{44}	2.3	-2.8	0.3	-0.0		0.3		0.4		0.5	.64
θ^{δ}_{55}	2.3	-0.5		0.0		0.3		0.4		0.5	.64
θ^{δ}_{66}	22.6	-9.4	13.3	-6.1	3.0	-1.4	0.7	-0.0		0.1	.36

Compared to the inclusion strategy the exclusion of improper cases means that the sampling distribution of S is once again different from a Wishart distribution. On the other hand, the disadvantage with inclusion is that strictly the estimates are not maximum likelihood estimates, so the asymptotic theory cannot apply to them.

The comparison of IN and EX is of interest, because the question may be raised to what extent both approaches give different Monte Carlo results. Therefore, for Model 1 simulation results using IN and EX were compared for $N = 25, 50, 100$.

In comparing 25IN with 25EX, 50IN with 50EX, and 100IN with 100EX, it should be realized that for 25EX, 50EX, and 100EX the *number of replications* is 96, 180, and 241, respectively, and not 300 as for IN (see Table 3).

With respect to the bias of parameter estimates the exclusion strategy leads to a decrease in bias, and also to a relative decrease in variance and mean square error (especially for $N = 25, 50$). The same is found for the bias of the estimated standard errors.

For two-sided 95% confidence intervals of separate parameters there are no striking differences between IN and EX (see Table 7, showing the observed minus expected percentage of 95% confidence intervals covering the population value ω_i among NR replications). By excluding improper replications there is a tendency towards conservative interval estimation.

When empirical correlations between parameter estimates are inspected, the differences with asymptotic correlations are somewhat smaller when the inclusion strategy is used.

Results on the chi-square estimate for goodness-of-fit show no substantial differences (see Table 8).

Summary. For $N = 25, 50, 100$ a separate comparison of the inclusion and exclusion strategy revealed that the Monte Carlo results are not systematically closer to theory when improper cases are included rather than excluded. An exception should be made for the bias and variance of both parameter estimates and their corresponding standard

TABLE 7.

Observed minus Expected Percentage of 95% Confidence Intervals Covering ω_i. Model 1. A blank means 0.

prmt	25IN NR=300	25EX NR=96	50IN NR=300	50EX NR=180	100IN NR=300	100EX NR=241
λ_1	-3	3	-4	-2		
λ_2	-6	1		3		1
λ_3	1	2	-1	3	-1	3
$-\beta$	1	1		1	-1	-1
γ_1	-3	2	-5	4	-3	
γ_2	-4	3	-1	2	-3	1
ϕ	-6	-4		1	-2	-2
ψ_{11}	-7	-10	-5	-2	-4	-2
ψ_{22}	-9	-6	-8	-7	-6	-4
$\theta^{\varepsilon}_{11}$	3	3	3	3	1	3
$\theta^{\varepsilon}_{22}$	4	5	2	4	1	2
$\theta^{\varepsilon}_{31}$	1	3	2	3	2	
$\theta^{\varepsilon}_{33}$	2	3	3	4		3
$\theta^{\varepsilon}_{42}$	1	3		1	2	3
$\theta^{\varepsilon}_{44}$	3	4	2	4	2	3
θ^{δ}_{11}	3	3	1	-1	-3	-4
θ^{δ}_{22}	-4	-3	-3	-6	-3	-3

errors. On the basis of the limited comparison, however, it cannot be concluded that LISREL is generally more robust under an exclusion strategy.

The results of this section have probably little practical implications for users of LISREL, because for nearly all practical purposes they have to deal with model estimates based on one set of data, and not on replicated sets. Once again, however, the regular user of LISREL, who wants to avoid improper solutions in the only sample available, is recommended to be especially concerned about small sample size.

The Problem of Choosing Starting Values

In our Monte Carlo work the population values of the parameters, ω_i, were chosen as initial estimates in each replication. These values can be regarded as *ideal starting values* (ISV). In everyday practice a researcher will start with estimates at some distance from those true values. This raises the question whether the results from our robustness study give too favorable an impression. Apart from the limited question whether and how

TABLE 8.

Characteristics of the Observed Sampling Distribution of the χ^2 Statistic for Goodness-of-Fit compared to the Expected Values of the χ^2_4 Distribution. Model 1, NR=300.

N	NR	observed minus expected value: median	mean	st.dev.	skewn.	kurto.	% $\chi^2 > 9.49$
25IN	300	-.2	-.3	-.4	-.3	-1.9	-2%
25EX	96	-.4	-.6	-.6	-.6	-3.1	-3%
50IN	300	-.2	-.3	-.1	.2	1.4	-1%
50EX	180	-.2	-.3	-.2	-.1	-0.8	-2%
100IN	300	.1	-.0	.0	.3	1.8	-1%
100EX	241	-.1	-.2	-.0	.5	3.3	-2%

	median	mean	st.dev.	skewn.	kurto.	% $\chi^2 > 9.49$
expected value	3.4	4.0	2.8	1.4	6.0	5%
st.error obs.value NR=300	0.2	0.2	0.2	0.3	2.5	1%

the choice of other sets of starting values (SV) might affect Monte Carlo conclusions, there is a general, intrinsic interest to study the effect of such choices.

In this section two questions are answered. (i) What is the effect of a fixed set of *alternative starting values* (ASV) compared to ISV? (ii) What is the effect of sample dependent SV on maximum likelihood estimation? One of the important improvements of LISREL V was the feature of "automatic" starting values, using noniterative instrumental variable and *two stage least squares* (TSLS) methods (Hägglund, 1982). These automatic initial estimates are dependent on sample fluctuations. More specifically, the second question is whether the use of ISV would lead to different results than that of TSLS estimates.

To answer both questions Model 1 was used. In Boomsma (1983) it was recommended not to use a sample size smaller than 100 for this model. Therefore, it was decided to study ISV vs. ASV, and ISV vs. TSLS for $N = 100, 200, 400$. Note, that LISREL III was used for ISV and ASV, while LISREL VI was used for TSLS.

With respect to the first problem a partly arbitrary fixed set of ASV was chosen: the ISV plus or minus 1.5 times the standard error for $N = 100$ (see Table 9, last column). The choice for plus or minus was made at random, except that no sign differences between ISV and ASV were allowed for. The *same* ASV were also used for $N = 200$ and $N = 400$. For all sample sizes the LISREL analyses thus start at the same distance from ω. Since the large sample estimates of the standard errors, se_∞, are proportional to $N^{-1/2}$ this means for $N = 200$ and $N = 400$ choosing $\underline{\text{ASV}}$ as ISV $\pm$ 2.1 and ISV $\pm$ 3.0 times their respective standard error.

Due to convergence problems *the number of common replications* (NCR) in comparing ISV with ASV was smaller than 300. For example, for $N = 100$ the number of non-converging replications was 2 for ISV, and 11 for ASV. For such reasons, NCR was 291, 299, and 299 for $N = 100$, 200, 400, respectively. In comparing ISV with TSLS initial estimates, NCR was fixed to 100 for all three sample sizes.

Before turning to results the following remark is made. If different starting values are used, given the same S it is possible that the convergence criterion is met both for ISV and ASV (or for ISV and TSLS), while the solution does not arrive exactly at the same (local) maximum. Large differences between solutions being compared might indicate that for the two sets of SV no absolute maximum was found but a local one. Rubin and Thayer (1982) suggest the possibility of multiple local maxima of the likelihood function, which would make the choice of starting values highly important. (See Bentler and Tanaka, 1983, and Rubin and Thayer, 1983, for a discussion.) The literature shows little empirical evidence how often maximum likelihood solutions converge to local maxima.

Results on the Effect of Different Starting Values

Magnitude of Differences for Model Estimates. All differences are expressed as a single number, the *decimal place* in which *absolute* differences occur. This is exemplified in Table 9. For $N = 100$ and parameter λ_1, among 291 replications max $|\hat{\omega}_{ij}^{\text{ISV}} - \hat{\omega}_{ij}^{\text{ASV}}| =$ 0.001068. Therefore, a 3 is reported.

For *parameter estimates* Table 9 shows that the maximum absolute difference depends on the size of the population value (ISV), and that it does not affect the second decimal place for most parameters. For two-thirds of the parameters the median absolute difference is in the fourth decimal place. Hardly any differences between the use of ASV and TSLS are observed.

No detailed results are given for the *estimates of the corresponding standard errors.* Considering the size of their population values the findings are about the same as for the parameter estimates, showing no substantial differences between the ISV, ASV and TSLS approach.

The latter is all the more clear for the *chi-square estimate for goodness-of-fit.* Studying ISV vs. ASV, for all N the maximum, minimum, and median absolute difference is in the 5th, 9th and 6th decimal place, respectively. Here also, the effect of TSLS is about the same as for ASV.

Finally, it is noted that for all three types of estimates there are no striking differences between the sample sizes N.

It can be *concluded* that when using a standard LISREL output, which on default gives final maximum likelihood estimates in three decimal places, for most replications the results would be exactly the same for ISV, ASV, and TSLS. This suggests that LISREL is quite robust against different starting values, and that its users should not bother too much about their choice.

Differences in Distributional Properties of Estimates. The findings of the previous section are also of importance for a question more specifically relevant in the context of evaluating Monte Carlo results: would the use of ASV have affected the conclusions from our robustness study? An answer can be found by inspecting the distributional properties of the LISREL estimates. For this purpose ISV has been compared with ASV only.

Given the previous results, it cannot be expected that an analysis of distributional properties of the estimates gathered with ASV will lead to other conclusions than those based on an analysis with ISV. This expectation was fully confirmed. For example, no differences were found in the percentage of improper solutions, only minor differences in

TABLE 9.

The Effect of Starting Values in Model 1.

$\underset{j}{\mathrm{Max}}|\hat{\omega}_{ij}^{ISV}-\hat{\omega}_{ij}^{ASV}|$, $\underset{j}{\mathrm{Med}}|\hat{\omega}_{ij}^{ISV}-\hat{\omega}_{ij}^{ASV}|$, $\underset{j}{\mathrm{Max}}|\hat{\omega}_{ij}^{ISV}-\hat{\omega}_{ij}^{TSLS}|$, and $\underset{j}{\mathrm{Med}}|\hat{\omega}_{ij}^{ISV}-\hat{\omega}_{ij}^{TSLS}|$, among NCR Replications. Decimal Place Difference Indication.

	N=100				N=200				N=400				starting values	
	max		med		max		med		max		med			
prmt	ASV	TSLS	ASV	TSLS	ASV	TSLS	ASV	TSLS	ASV	TSLS	ASV	TSLS	ISV	ASV
λ_1	3	4	5	5	4	4	5	5	4	4	5	5	0.98	0.70
λ_2	3	4	5	5	4	4	5	5	4	4	5	5	0.92	1.20
λ_3	2	3	4	4	3	3	4	4	3	3	4	4	5.22	7.16
$-\beta$	4	4	5	5	4	4	5	5	4	4	5	5	-0.61	-0.37
γ_1	4	4	5	5	4	4	5	5	4	4	5	5	-0.57	-0.83
γ_2	4	4	5	5	4	4	5	5	4	4	5	5	-0.23	-0.47
ϕ	2	3	4	4	3	3	4	4	3	3	4	4	6.81	9.80
ψ_{11}	3	3	4	4	3	3	4	4	3	3	4	4	4.85	2.69
ψ_{22}	3	3	4	4	3	3	4	4	3	3	4	4	4.09	5.95
$\theta_{11}^{\varepsilon}$	3	3	4	4	3	3	4	4	3	3	4	4	4.73	2.65
$\theta_{22}^{\varepsilon}$	3	3	4	4	3	3	4	4	3	3	4	4	2.57	0.71
$\theta_{31}^{\varepsilon}$	3	3	4	4	3	3	4	4	3	3	4	4	1.62	3.07
$\theta_{33}^{\varepsilon}$	3	3	4	4	3	3	4	4	3	3	4	4	4.40	6.78
$\theta_{42}^{\varepsilon}$	3	3	4	4	3	3	4	4	3	3	4	4	0.34	1.54
$\theta_{44}^{\varepsilon}$	3	3	4	4	3	3	4	4	3	3	4	4	3.07	5.07
θ_{11}^{δ}	2	3	4	4	3	3	4	4	3	3	4	4	2.80	5.14
θ_{22}^{δ}	1	1	2	3	1	2	2	3	1	2	2	3	264.98	181.37

bias of parameter estimates and that of estimates for standard errors. Also, the agreement in estimates for 95% confidence intervals, as well as in the chi-square estimates for goodness-of-fit, was very close.

Summary. For Model 1 it is concluded that the use of ideal starting values does not give too favorable a picture of distributional properties of the estimates, compared to the use of alternative nonideal starting values.

Recommendations

Given the results of the first two problems, for many models users of LISREL, who in most cases base their conclusions on just one sample, should be advised to have a sample size larger than 100. Considering the fact that they often do exploratory research, there is a strong need for replication and cross-validation. In such circumstances a sample size of at least 200 is no statistical luxury: the statistician is then in a position to explore data on one half of the sample, while in principle the other half could be used for confirmatory purposes. This recommendation is not weakened if many other statistical properties of model estimates in covariance structure analysis are being considered (Boomsma, 1983).

Researchers designing Monte Carlo studies for LISREL should be well aware of the problems to be expected in studying the effects of small sample size. The number of replications in stock needed is larger with decreasing sample size. It is still disputable whether replications with improper solutions should be included or excluded in such research.

The choice of ideal starting values did not seem to have influenced Monte Carlo results. In large sample studies the use of ideal starting values in robustness research is recommended, not only because of the fact that different starting values have very little effect in estimation, but because it saves some computer time. In converging samples, compared to the use of TSLS estimates, on average ISV analyses are faster (LISREL VI). Also, TSLS starting values may not lead to convergence where ISV do; e.g. among 50 replications (Model 3US, $N = 25$) this happened three times. On the other hand, when both ISV and TSLS end up in nonconvergence, ISV take more computer time. More important, in practical applications where population values are unknown, the use of automatic starting values is highly attractive, also because no indications for convergence to local maxima were found.

References

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

Bentler, P. M., & Tanaka, J. S. (1983). Problems with EM algorithms for ML factor analysis. *Psychometrika, 48*, 247–251.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: causality, structure, prediction* (Part 1, pp. 149–173). Amsterdam: North-Holland.

Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality.* Unpublished doctoral dissertation, University of Groningen, Groningen.

Gruvaeus, G. T., & Jöreskog, K. G. (1970). *A computer program for minimizing a function of several variables* (Research Bulletin 70-14). Princeton, NJ: Educational Testing Service.

Hägglund, G. (1982). Factor analysis by instrumental variable methods. *Psychometrika, 47*, 209–222.

IMSL (1982). *IMSL Library. Reference Manual.* (Vol. 2, 9th ed.). Houston, TX: International Mathematical and Statistical Libraries.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32*, 443–482.

Jöreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation, testing. In P.R. Krishnaiah (Ed.), *Applications of statistics* (pp. 265–287). Amsterdam: North-Holland.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V. Analysis of linear structural relationships by maximum likelihood and least squares methods* (Research Report 81-8). Uppsala: University of Uppsala, Department of Statistics.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI. Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods. User's guide.* Uppsala: University of Uppsala, Department of Statistics.

Kelderman, H. (in press). LISREL models for inequality constraints in factor and regression analysis. In P. F. Cuttance & J. R. Ecob (Eds.), *Structural modeling.* Cambridge: Cambridge University Press.

Lee, S. Y. (1980). Estimation of covariance structure models with parameters subject to functional restraints. *Psychometrika, 45*, 309–324.

Mattson, A., Olsson, U., & Rosèn, M. (1966). *The maximum likelihood method in factor analysis with special consideration to the problem of improper solutions* (Research Report). Uppsala: University of Uppsala, Department of Statistics.

Rindskopf, D. (1983). Parameterizing inequality constraints on unique variances in linear structural models. *Psychometrika, 48*, 73–83.

Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika, 49*, 37–47.

Rubin, D. B., & Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika, 47*, 69–76.

Rubin, D. B., & Thayer, D. T. (1983). More on EM for ML factor analysis. *Psychometrika, 48*, 253–257.

Tamura, Y., & Fukutomi, K. (1970). On the improper solutions in factor analysis. *TRU Mathematics, 6*, 63–71.

Van Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika, 43*, 225–243.

Manuscript received 8/5/82
Final version received 1/24/85