The robustness of a general method for the analysis of covariance structures against small sample sizes and departures from multivariate normality.*

A. Boomsma R.U. Groningen

Introduction

Some aspects of a study on the robustness of LISREL (Jöreskog, 1970, 1976; Jöreskog & Sörbom, 1978) against small sample sizes and departures from multivariate normality are discussed. We will confine ourselves to a general discussion of the background of our study and to an exposition of results on robustness against small samples studied for one particular model. It should be stressed beforehand that at this stage of investigation our conclusions will not be very general, but at least we can give some first impressions.

The general model considers a data matrix Z $(n \times k)$ of n observations on k random variables. It is assumed that the rows of Z are independently distributed, each having a multivariate normal N_k distribution with the same covariance matrix Σ $(k \times k)$, which has the following form

$$\Sigma = \begin{pmatrix} \Lambda_y (B^{-1}\Gamma\Phi\Gamma'B'^{-1} + B^{-1}\Psi B'^{-1}) \Lambda'_y + \Theta_\varepsilon & \Lambda_y B^{-1}\Gamma\Phi\Lambda'_x \\ \Lambda_x \Phi\Gamma'B'^{-1}\Lambda'_y & \Lambda_x \Phi\Lambda'_x + \Theta_\delta \end{pmatrix}. \quad (1)$$

It can be seen that the elements of Σ are functions of the elements of eight matrices. For a specified model the researcher wants estimates of the unknown coefficients of these eight matrices. The vector of all independent constrained and free parameters will be denoted by ω , which is of order s.

Assuming that the distribution of Z is multivariate normal it is possible to get maximum likelihood estimates of the elements of ω. That is what the LISREL-program does for us, if we have a data matrix Z or the corresponding sample covariance matrix S $(k \times k)$ and a specified model. It can be shown (cf.Cramér, 1946) that asymptotically, for large samples, the joint maximum likelihood estimators have nice distributional properties: consistency, efficiency, multivariate normality. Also once

*Paper presented at the European Meeting on Psychometrics and Mathematical Psychology, Uppsala, June 15-17, 1978 (partly rewritten).

maximum likelihood estimates have been obtained, in large samples the goodness of fit of the specified model may be tested by the likelihood ratio technique. Under sampling and distributional conditions previously described, it can be shown that asymptotically, in large samples, minus two times the particular likelihood ratio has a χ^2 distribution with a specified number of degrees of freedom. In large samples the goodness of fit of different models for the same data can be compared in an analogous way.

It will be clear that even if the multivariate normality assumption should hold, it remains of scientific interest to study how robust parameter estimates and the statistic for goodness of fit are against the use of small to moderate sample sizes.

Robustness against small sample size

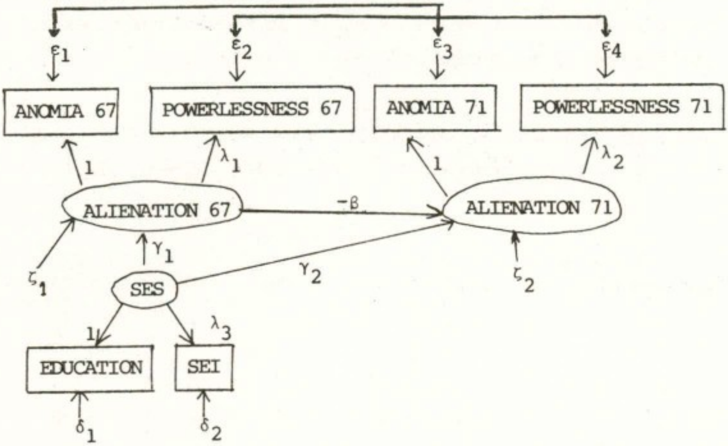
As a start an example was taken which has been discussed at several places in the literature (e.g. Jöreskog, 1976; Jöreskog & Sörbom, 1978). The example is a longitudinal study on the stability of alienation. Our point of departure is a model analyzed by Jöreskog (o.c., p.15 ff.), which proved to have a reasonable fit to the original data. To be certain that the model under investigation is the "true" model (fitting the data perfectly) we started from a slightly different covariance matrix than the original one. This matrix will be considered to be the population covariance matrix, where to sample from in the sequel.

So we have a population covariance matrix

$$\Sigma = \begin{array}{c} y_1 \\ y_2 \\ y_3 \\ y_4 \\ x_1 \\ x_2 \\ \\ \end{array} \begin{array}{c} \left[\begin{array}{rrrrrr} 11.832 & & & & & \\ 6.946 & 9.364 & & & & \\ 6.820 & 5.085 & 12.534 & & & \\ 4.790 & 5.028 & 7.497 & 9.985 & & \\ -3.913 & -3.830 & -3.919 & -3.613 & 9.610 & \\ -20.424 & -19.990 & -20.455 & -18.858 & 35.522 & 450.297 \end{array}\right] \\ \begin{array}{rrrrrr} y_1 \quad & y_2 \quad & y_3 \quad & y_4 \quad & x_1 \quad & x_2 \end{array} \end{array} \qquad (2)$$

We have a specified model, which is given in figure 1, and we know what the true parameter values, exposed in table 1, are.

Figure 1. Model for the stability of alienation.



Now one hundred samples were taken from a multivariate normal ditribution with a covariance matrix given by (2). This was done for varying sample sizes, ranging from 25 to 400; this range seems realistic enough to keep things near to everyday practice. After the sampling was done, on each sample a LISREL-analysis was done for the model just specified. The end result is that we have for each of the 17 parameters 100 estimated values, and this for varying sample size. The same holds for the corresponding standard errors and the goodness of fit statistic.

We know that the sampling distribution of the estimates is normal and that the sampling distribution of the goodness of fit statistic is chi-square in very large samples. What are these distributions like for the sample sizes we consider? It will be clear that given a specified model some parameters are of greater interest than other ones, but let us concentrate on an overall picture first.

Parameter estimates

In table 2 the difference between the mean value of the parameter estimates (each based on 100 observations) and the true parameter value is shown for five sample sizes. To keep things clear all values are rounded to one decimal place.

Table 1. Difference between the mean value of parameter estimates and their corresponding true value, $(\bar{\bar{\omega}}_i - \omega_i)$.

	sample size					
parameter	25	50	100	200	400	true value
λ_1	.1	.1	.1	.0	.0	.98
λ_2	.2	.1	.1	.0	.0	.92
λ_3	.9	.3	.6	.0	.0	5.22
$-\beta$	.0	.0	.0	.0	.0	-.61
γ_1	-.1	.0	.0	.0	.0	-.57
γ_2	.0	.0	.0	.0	.0	-.23
ϕ	2.3	1.6	.2	.2	.1	6.81
ψ_{11}	.9	.4	-.1	.1	.1	4.85
ψ_{22}	1.1	.1	-.3	-.1	-.1	4.09
$\theta^{\varepsilon}_{11}$	-1.1	-.7	.0	-.1	-.1	4.73
$\theta^{\varepsilon}_{22}$	-.5	-.5	-.2	.0	-.1	2.57
$\theta^{\varepsilon}_{31}$	-.9	-.3	.0	-.1	-.1	1.62
$\theta^{\varepsilon}_{33}$	-2.4	-.7	.1	-.2	.0	4.40
$\theta^{\varepsilon}_{42}$	-.2	-.2	-.2	.0	.0	.34
$\theta^{\varepsilon}_{44}$	-1.1	-.5	-.5	.0	.0	3.07
θ^{δ}_{11}	-1.7	-1.6	.0	-.2	-.1	2.80
θ^{δ}_{22}	-58.2	-9.3	-19.0	-3.5	-2.6	264.89

It can be seen that, roughly speaking, the mean of the estimated parameter values is closer to the true parameter value with increasing sample size. For a sample size n = 200 the mean parameter estimate is already close to the true value. There is only slight or no improvement if these results are compared with those of n=400. (Though we do not present figures for sample sizes as large as 800, one would find that there is hardly any improvement in the mean parameter estimates going from n = 200 to n = 800).

We thus looked at the empirical distribution of the parameter estimates and compared the mean of each of the estimates with the corresponding true value. The figures of table 1 show the amount of bias when the mean is used as an estimator for the location parameter ω_i. It is well known that the mean is not a robust estimator (Andrews, et al., 1972). In the final publication of this research detailed attention will be given to so-called M-estimates (o.c., p.14).

Of course it is possible to present similar results for the standard errors as we did for the parameter estimates in table 1 (in effect it was found that for the standard errors M-estimates behave excellent compared to the mean of the 100 estimated standard errors). However, in practice it is a good statistical custom to calculate confidence intervals for some model parameters. That is where the standard errors of the estimates [se($\hat{\omega}_{ij}$)] play their role. It is possible to count the number of times out of 100 samples in which the true value ω_i is outside the 95% confidence interval, $\hat{\omega}_{ij} \pm 1.96\ \mathrm{se}(\hat{\omega}_{ij})$.

Certainly, it is not easy to summarize the results from table 2, but for n = 200 the figures are what we should expect them to be. Approximately five times out of 100 the intervals do not cover the true parameter values. For smaller sample size the range of numbers is more unfavorable.

Above, for every parameter ω_i 100 confidence intervals were considered. In theory $q_{ij} = (\hat{\omega}_{ij} - \omega_i)/\mathrm{se}\ (\hat{\omega}_{ij})$, j = 1, 2, ..., 100 has a standard normal distribution. A different way to look at the results is by calculating the 95% confidence interval for the mean of q_{ij}, based on 100 estimated values. Now the question is whether this interval $\bar{q}_{ij} \pm 1.98\ s\ (q_{ij})/\ 100^{\frac{1}{2}}$, where $\bar{q}_{ij}$ and $s(q_{ij})$ are the mean and the standard deviation of q_{ij}, respectively, covers the true parameter value of zero.

Table 2. The number of times out of 100 the true value ω_i is outside the confidence interval $\hat{\omega}_{ij} \pm 1.96$ se $(\hat{\omega}_{ij})$.

	sample size				
parameter	25	50	100	200	400
λ_1	8	8	2	8	2
λ_2	13	5	3	4	2
λ_3	5	8	6	6	5
$-\beta$	3	5	6	7	4
γ_1	3	11	7	4	1
γ_2	11	6	10	6	7
ϕ	6	5	9	9	6
ψ_{11}	10	8	8	8	3
ψ_{22}	10	13	12	8	5
$\theta^{\varepsilon}_{11}$	2	2	2	3	3
$\theta^{\varepsilon}_{22}$	2	2	1	4	2
$\theta^{\varepsilon}_{31}$	5	3	7	5	3
$\theta^{\varepsilon}_{33}$	4	4	7	4	4
$\theta^{\varepsilon}_{42}$	4	7	2	4	5
$\theta^{\varepsilon}_{44}$	2	3	0	5	2
θ^{δ}_{11}	2	5	10	4	2
θ^{δ}_{22}	10	3	8	9	7
min. no.	2	2	1	3	1
max. no.	13	13	12	9	7

The results given in table 3 look promising. For n = 200 and n = 400 all but one interval covers the true parameter value. For a sample size smaller than 200 things look worse.

Table 3. Does the confidence interval $\bar{q}_{ij} \pm 1.98\ s(q_{ij})/100^{\frac{1}{2}}$ cover the true parameter value of zero? 1 = yes, -0 = true value to the right, +0 = true value to the left.

	sample size				
parameter	25	50	100	200	400
λ_1	-0	1	1	1	1
λ_2	-0	1	1	1	1
λ_3	1	1	1	1	1
$-\beta$	1	1	1	1	1
γ_1	1	+0	1	1	1
γ_2	+0	1	1	1	1
ϕ	1	-0	1	1	1
ψ_{11}	-0	-0	-0	1	1
ψ_{22}	-0	-0	-0	-0	-0
$\theta^{\varepsilon}_{11}$	1	1	1	1	1
$\theta^{\varepsilon}_{22}$	1	1	1	1	1
$\theta^{\varepsilon}_{31}$	-0	1	1	1	1
$\theta^{\varepsilon}_{33}$	-0	1	1	1	1
$\theta^{\varepsilon}_{42}$	1	1	1	1	1
$\theta^{\varepsilon}_{44}$	1	1	1	1	1
θ^{δ}_{11}	1	1	+0	1	1
θ^{δ}_{22}	-0	1	-0	1	1

In addition to these overall views we looked in more detail at results for specific model parameters; for example at the sampling distribution of parameter $-\beta$ which is one of the salient ones in the model, being indicative for stability of alienation over time. By graphical methods (QQ- and/or PP-plots, histograms and density functions) it is possible to see how close the sampling distribution of q_1 based on 100 observations comes to the theoretical distribution. By lack of space we cannot illustrate the attractiveness of these methods: they add information otherwise undiscovered.

Goodness of fit

How well does the chi-square statistic for goodness of fit behave in small samples when the model is true? The results are summarized in table 4.

Table 4. Distributional characteristics of the chi-square statistic with four degrees of freedom from 100 replications.

n	min.	max.	range	med.	mean	var.	ske.	kur.	$\chi_4^2 > 9.49$
25	.2	10.6	10.4	2.9	3.4	5.0	1.0	3.5	# 2
50	.3	12.3	12.0	3.1	3.8	7.8	1.1	3.5	6
100	.3	19.0	18.7	3.6	4.1	8.0	2.0	10.1	3
200	.2	14.2	14.0	4.0	4.6	8.3	.9	3.5	4
400	.1	10.5	10.4	3.5	3.8	5.8	.9	3.3	3
true value				3.4	4.0	8.0	.7	3.0	

There is no clear trend in these figures, but for n = 200 the picture is rather satisfying. For n = 400 the variance is small. From the last column of table 4 it appears that for n = 50 the number of χ_4^2 - values larger than 9.49, which is the 95-th quantile of the chi- square distribution with four degrees of freedom, is close to 5, the number we should expect. Comparing histograms of the sampling distributions with the theoretical density function with four degrees of freedom did not give a definite answer to the question for which sample size the goodness of fit statistic has a distribution which is reasonably close to the theoretical distribution. It might well be that the number of samples is just too small; research has been started to see whether the picture improves by enlarging that number.

Discussion

We could not present all interesting results from this pilot-study; one thing not mentioned for example is the dependence between parameter estimates. Nevertheless, suppose we were a Bayesian and suppose we had to formulate our expectation with respect to the robustness of LISREL-procedures for a variety of models against small sample size, what would our answer be, given that we had prior information from this particular model?

Our formulation would be that we expect LISREL to be fairly robust for sample sizes as large as 200, expecting also little improvement if n was four times as large. It might seem questionable to use LISREL with samples smaller than 200, but more research need to be done here.

Earlier it was mentioned that it would be worthwhile to investigate the robustness of LISREL against non-normality. If samples are taken from studies in the social sciences, the distributional properties of the variables social scientists use should strike us, unless we are ignorant. Non-normality and discreteness are the outstanding properties of many variables.

What ought to be done is, first to take samples from non-normal distributions, for example by varying the skewness and study its effect on parameter estimates and the goodness of fit statistic. Secondly, it seems realistic to vary things on a discreteness dimension, for example by putting the sampled values in 3, 10 or 20 classes (cf. Olsson, 1978). Although we cannot present any results yet, it is sure a challenge to study this topic in the hope that eventually we might be able to reassure users of the LISREL-program not to worry to much about non-normality and discreteness of observations. But even if that appears to be a false hope, we certainly want to know more than we have learned sofar.

Literature

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. & Tukey, J.W. (1972) Robust estimates of location: survey and advances. Princeton, New Jersey: Princeton University Press.

Cramér, H. (1946) Mathematical methods of statistics. Princeton: Princeton University Press.

Jöreskog, K.G. (1970) A general method for analysis of covariance structures. Biometrika, 57, 239-251.

Jöreskog, K.G. (1976) Structural equation models in the social sciences: specification, estimation and testing. Research Report 76-9, University of Uppsala: Department of Statistics.

Jöreskog, K.G. & Sörbom, D. (1978) LISREL IV. A general computer program for estimation of linear structural equation systems by maximum likelihood methods. University of Uppsala: Department of Statistics.

Olsson, U. (1978) On the robustness of factor analysis against crude classification of the observations. Research Report 78-6, University of Uppsala: Department of Statistics.
