KM 38 (1991)
pg.17-28

A comparison of four methods of calculating standard errors of Maximum Likelihood estimates in the analysis of covariance structure

Conor V. Dolan & Peter C.M. Molenaar

Abstract

Four methods of calculating standard errors (s.e.'s) in the analysis of covariance structure using normal theory Maximum Likelihood estimation are compared: LISREL s.e.'s, s.e.'s derived from the exact Hessian, s.e.'s derived from the finite difference approximation to the Hessian using exact gradients, and s.e.'s derived from the update of the Hessian obtained during quasi-Newton optimisation.

Two comparisons are made. The first is based on 30 data sets simulated according to an orthogonal common factor model. The second is based on the analysis of a well known data set concerning the Stability of Alienation.

The finite difference and exact s.e.'s are identical to at least four decimals. Little difference is observed between the remaining s.e.'s.

Faculteit Psychologie - Universiteit van Amsterdam.
Roeterstraat 15 1018 WB Amsterdam. Tel: 020 525 6830

1 Introduction.

In covariance structure analysis, a number of fitting functions can be used to estimate parameters and their standard errors (Jöreskog and Sörbom, 1988). Standard errors are indicative of the precision to which a parameter has been estimated and can be used to test a parameter's statistical significance. In the present paper, various ways of calculating standard errors are compared when estimation is carried out by minimisation of the likelihood ratio $-2\ln[L(\omega)/L(\Omega)]$, where $L(\Omega)$ and $L(\omega)$ denote the likelihood function in, respectively, Ω, the set of symmetric positive definite matrices, and ω, the subset for which the covariance structure model holds. Standard errors of such estimates are a function of the number of subjects and the positive definite matrix (i.e. the Hessian) of second order partial derivatives of the likelihood ratio with respect to the estimates.

The Hessian of the likelihood ratio in the analysis of covariance can be obtained by a number of methods: calculation of the exact Hessian, calculation of the Information matrix, a finite-difference approximation to the Hessian, and calculation of updates of the Hessian during quasi-Newton minimisation.

Generally the calculation of the exact Hessian is avoided as this can be complicated and time consuming depending on the complexity of the model under investigation. In the LISREL program (Jöreskog and Sörbom, 1984) standard errors are based on Fisher's Information matrix (Jöreskog, 1977, page 270). Due to the popularity of the LISREL program this method of calculating standard error is most often used. The approximation to the Hessian matrix by means of finite differences of the gradient is simple to implement (e.g. Dennis and Schnabel, 1983), but requires repeated evaluation of the gradients of the likelihood ratio. Finally, the Hessian matrix can be obtained during optimisation of the likelihood ratio by means of the quasi-Newton method. Here the elements of the inverse of the Hessian are built up from successive function and gradient evaluations using, for instance, the Davidon-Flechter-Powell (DFP) update (Gill, Murray and Wright, 1981; Dennis and Schnabel, 1983).

In this paper the LISREL standard errors (s.e.'s) of ML estimates are compared to those derived from the exact Hessian (referred to as exact s.e.'s), the finite difference (fd) approximation using analytic gradients (fd s.e.'s) and the DFP Hessian update (DFP s.e.'s). The comparison is made using data simulated according to an orthogonal common factor model. Ten data sets were simulated for 100, 200 and 400 cases. A second comparison is made by analysing a well known data set: the stability of alienation data (Jöreskog and Sörbom, 1988, page 171, model D). The latter comparison is limited to the LISREL, fd and DFP standard errors.

2 Methods.

In this section we will briefly described the methods used to calculate the standard errors of the ML estimates in the analysis of covariance structure. A Fortran 77 program was written incorporating the following covariance structure model (a LISREL sub-model):

$$y = \Lambda \eta + \varepsilon \tag{1a}$$

$$\eta = (I-B)^{-1} \zeta \tag{1b}$$

where **y** is a multinormal (p x 1) vector of i.i.d. manifest variables, which are a linear combination of common latent variables η (q x 1) and unique residuals ε (p x 1). The parameter matrix Λ (p x q) contains the loadings of the observed on the latent variables. The latent variables η may be interrelated by linear structural equations. The matrix B (q x q) and the vector ζ (q x 1) contain, respectively, the structural coefficients and the residual terms in these linear latent equations.

Let Σ denote the model covariance matrix of which **S** is an unbiased estimate. Let $E[\zeta\zeta^t]$ equal Ψ and $E[\varepsilon\varepsilon^t]$ equal Θ_ε, where $E[\zeta] = E[\varepsilon] = 0$, E[.] denotes the expectation operator, and the superscript t denotes transposition. Then, given Eqs. 1a and 1b, we have:

$$\Sigma = \Lambda (I-B)^{-1} \ \Psi \ (I-B^t)^{-1} \ \Lambda^t + \Theta_\varepsilon \tag{2}$$

The model covariance matrix Σ depends upon a k-dimensional vector θ of parameters which are contained in the matrices on the right side of Eq. 2: $\Sigma=\Sigma(\theta)$. Estimates of θ are obtained by minimising the likelihood ratio $f_{ml}(\theta) = -2\ln[L(\omega)/L(\Omega)]$ which is defined as (Lawley and Maxwell, 1971; Jöreskog, 1977):

$$f_{ml}(\theta) = \log|\Sigma(\theta)| + \mathrm{tr}(S\Sigma^{-1}(\theta)) - \log|S| - p \tag{3}$$

Under the assumption that the specified model $\Sigma(\theta)$ is true, (N-1) times the minimum value of $f_{ml}(\theta)$, where N is the sample size, is distributed approximately as χ^2 with $p(p+1)/2 - k$ degrees of freedom.

The first order derivatives of f_{ml} for the full LISREL model are given in Jöreskog (1977) and are reproduced in the Appendix for the submodel defined by Eqs. 1a and 1b. The associated (k x k) matrix of second order partial derivatives is:

$$\mathbf{H}(\theta) = \partial^2 f_{ml} / \partial\theta \, \partial\theta^t \tag{4}$$

The relationship between the Hessian of the likelihood ratio given in Eq. 3 and the Hessian of the likelihood function, $L(\omega)$, is given in Bollen (1989, page 135). The covariance matrix, $\mathbf{G}$, of the parameter estimates in $\hat{\theta}$ is a function of the Hessian evaluated at $\hat{\theta}$, $\mathbf{H}(\hat{\theta})$, and the number of subjects, N:

$$\mathbf{G} = [2 / (N-1)]\ \mathbf{H}^{-1}(\hat{\theta}) \tag{5}$$

Let $[\mathbf{G}]_{ij}$ represent the element in the i-th row and j-th column of $\mathbf{G}$. The standard error of i-th parameter estimate $\hat{\theta}_i$ is defined as: $\sqrt{[\mathbf{G}]_{ii}}$. Apart from the standard errors, the correlations between the parameters can be obtained by standardising $\mathbf{G}$.

A Fortran 77 program was written to minimise f_{ml} by means of the quasi-Newton method using exact first order derivatives and DFP update of the inverse of the Hessian, $\mathbf{H}^{-1}(\theta)$ (see Dennis and Schnabel, 1976; Gill, Murray and Wright, 1981). Apart from the parameter estimates this program produces what were referred to above as the DFP standard errors. The approximation to the Hessian using forward differences of the gradients is carried out within the same Fortran program using algorithm 5.6.1 given in Dennis and Schnabel (1983). Let $\hat{\theta}$ denote the vector of parameter estimates that minimises Eq. 3 and let $\partial f_{ml}/\partial\hat{\theta}$ denote the associated vector of gradients. The approximation to the Hessian using forward differences of the gradient is calculated as follows:

$$\mathbf{w}_i = v^{-1}\ [\partial f_{ml}/\partial(\hat{\theta} + v\mathbf{e}_i) - \partial f_{ml}/\partial\hat{\theta}] \tag{6}$$

where $\mathbf{w}_i$ is the i-th column of the (k x k) matrix $\mathbf{W}$, v is the finite-difference interval, and the i-th unit vector $\mathbf{e}_i$ is the finite-difference vector (Gill, Murray, and Wright, 1981). As the matrix $\mathbf{W}$ will not necessarily be symmetric, the Hessian is calculated as $\mathbf{H}(\hat{\theta}) = 1/2(\mathbf{W} + \mathbf{W}^t)$.

The exact Hessian for the matrix of factor loadings Λ in the common factor model is given by Lawley and Maxwell (1971, Equation A2.9) and is reproduced in the Appendix. A separate Fortran 77 program was written, incorporating the model $\Sigma = \Lambda\Lambda^t$, to calculate the exact Hessian.

The LISREL standard errors are obtained from the LISREL VI program (Jöreskog and Sörbom, 1984). These standard error are based on the asymptotic covariance matrix of the parameter estimates:

$$[2 / (N-1)]\ E[\mathbf{H}(\hat{\theta})]^{-1} \tag{7}$$

The matrix $E[\mathbf{H}(\hat{\theta})]$, Fisher's Information matrix, yields correct standard errors under assumption that the distribution of the elements in the sample covariance matrix, **S**, is asymptotically multinormal with mean $E[\mathbf{S}] = \Sigma(\theta)$. The latter implies that $\Sigma(\theta)$ is the true model, i.e., $\mathbf{S} - \Sigma(\hat{\theta})$ approaches zero as N approaches infinity.

3. Comparison 1.

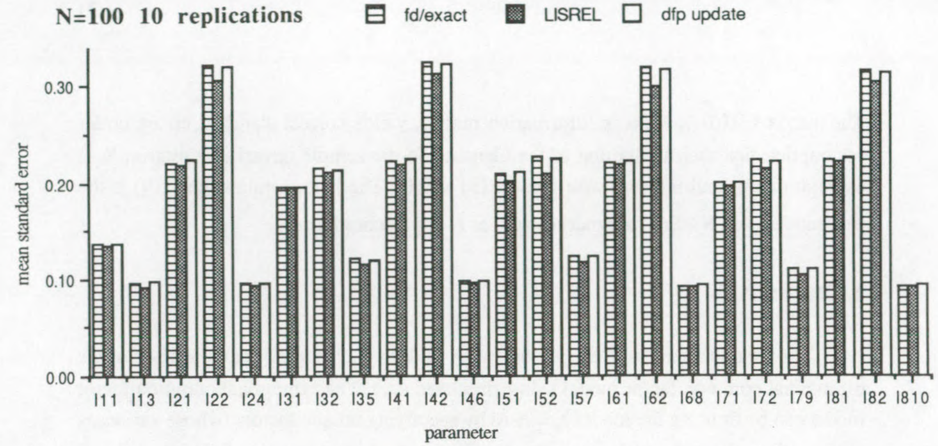
The first comparison of the various ways of deriving standard errors is made in the orthogonal common factor model using simulated data. The orthogonal common factor model can be fit using the model $\Sigma = \Lambda\,\Lambda^t$ by specifying unique factors (whose variances are usually estimated in Θ_ε) as components of η. In the present simulation there are 2 common factors and eight indicators. The matrix Λ equals:

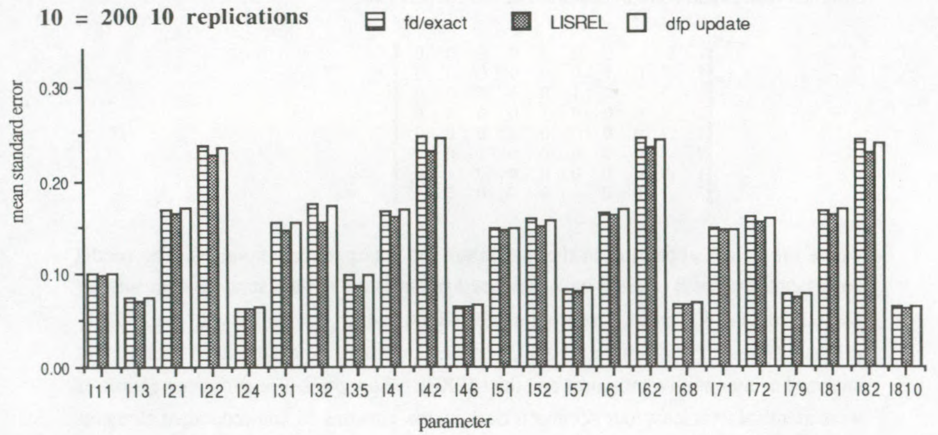
$$\Lambda = \left|\begin{array}{cccccccccc}
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & .7 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & .9 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
2 & .7 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & .9 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
2 & .7 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & .9 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
2 & .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right|$$

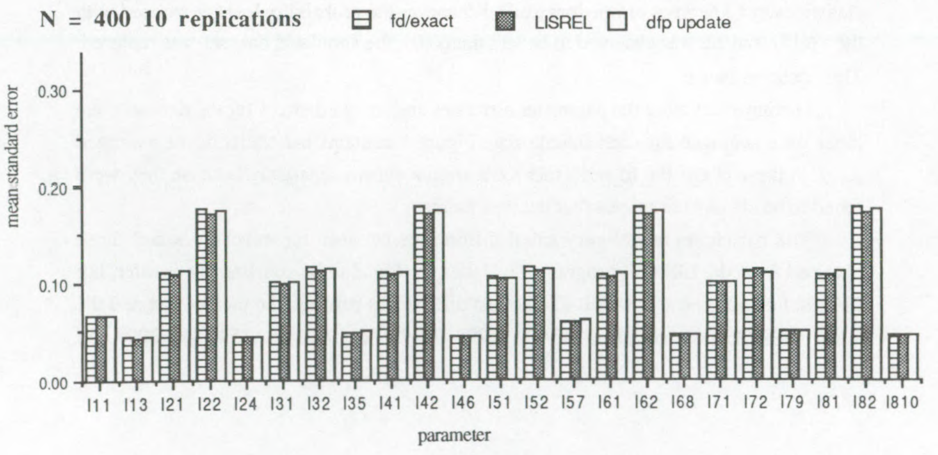
where the values appearing in the matrix are the true parameters values. The model covariance matrix Σ implied by Λ was used to simulate 30 data sets: 10 data sets for N=100, 200 and 400 cases. The program GENRAW, which is supplied with the PRELIS program (Jöreskog and Sörbom, 1986), was used to simulate the data. Each data set was analysed using our own programs and the LISREL VI program. The parameter estimates were identical to at least four decimals. Each matrix contains 36 non-redundant elements which are modelled using 23 parameters, so each analysis yielded a χ^2 goodness-of-fit statistic with 13 degrees of freedom (d.f.). Whenever the probability level associated with the $\chi^2(13)$ statistic was observed to be less than 0.05, the simulated data set was replaced. This occurred twice.

Having calculating the parameter estimates and standard errors for the data sets, the latter were averaged for each sample size. Figure 1 contains bar charts of the averaged s.e.'s. In these charts the fd and exact s.e.'s are not shown separately because they were found to be identical to at least four decimal points.

The bar charts reveal very small differences between the exact s.e.'s and those obtained from the LISREL program. The latter are found to be consistently smaller, but the absolute difference is small. The largest differences between the mean exact and the mean LISREL s.e.'s equal: .3183 vs. .2994 (N=100), .1752 vs. .1544 (N=200) and


N=100 10 replications
fd/exact
LISREL
dfp update
mean standard error
0.00
0.10
0.20
0.30
l11 l13 l21 l22 l24 l31 l32 l35 l41 l42 l46 l51 l52 l57 l61 l62 l68 l71 l72 l79 l81 l82 l810
parameter


10 = 200 10 replications
fd/exact
LISREL
dfp update
mean standard error
0.00
0.10
0.20
0.30
l11 l13 l21 l22 l24 l31 l32 l35 l41 l42 l46 l51 l52 l57 l61 l62 l68 l71 l72 l79 l81 l82 l810
parameter


N = 400 10 replications
fd/exact
LISREL
dfp update
mean standard error
0.00
0.10
0.20
0.30
l11 l13 l21 l22 l24 l31 l32 l35 l41 l42 l46 l51 l52 l57 l61 l62 l68 l71 l72 l79 l81 l82 l810
parameter

.1798 vs. .1727 (N=400). The s.e.'s obtained from the DFP updates also come very close to the exact ones.

Correlations of parameters estimates can be obtained by standardising the covariance matrix **G**. These correlations are a multiplicative function of the elements in **G**, hence errors inherent in approximations of the Hessian may thus become augmented. To illustrate, the average correlation between the estimates of the common factor loadings $[\Lambda]_{11}$ and $[\Lambda]_{21}$ were found to equal .2463 (exact & fd), .2708 (LISREL), and .2194 (DFP update) in the N= 100 samples. This magnitude of difference is typical of what was found in the simulation.

As mentioned above, the derivation of the LISREL s.e.'s is based on the assumption that $\mathbf{S} - \Sigma(\theta)$ tends to zero asymptotically. It was pointed out by Jöreskog (1978, page 448) that this may not be a realistic assumption in the analysis of real data. Hence, it is of some interest to evaluate the standard errors for a model which contains a misspecification. To this end the 10 data sets comprising N=400 cases were analysed using a single common factor model. The removal of the second common factor resulted in a gain of 7 d.f. so that there are now 20 d.f. associated with the χ^2 goodness-of-fit statistic. In the analyses the $\chi^2(20)$ was found to range between 60 and 115. Given the associated probability levels (never exceeding 0.001), the hypothesis of a single common factor model would be rejected consistently. Again the standard errors were averaged over the 10 replications and are shown in Figure 2. Clearly a violation of the assumption that asymptotically $\mathbf{S} - \Sigma(\theta)$ is zero does not affect the estimates of the standard error based on the Information matrix.

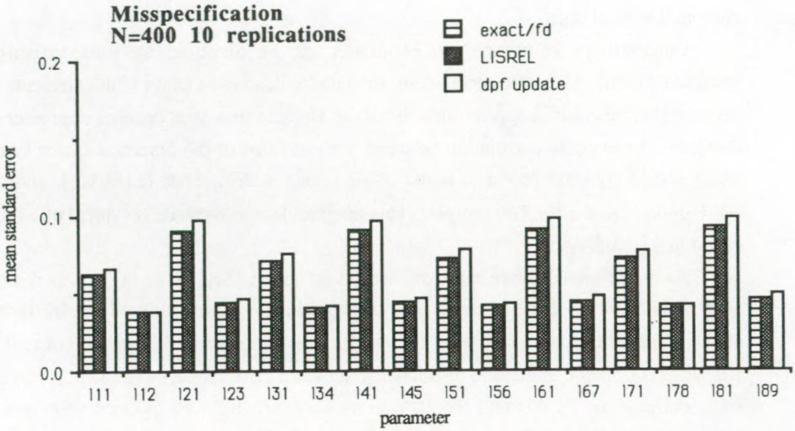


Figure 2: Bar chart of the mean standard error given a misspecified model: a single common factor instead of two common factors. Standard errors are averaged over 10 replications (N=400).

4. Comparison 2.

A second comparison is made by analyzing a well known data set concerning the stability of alienation. The stability of alienation study was originally reported by Wheaton, Muthén, Alwin and Summers (1977) and features as an illustrative LISREL analysis in Jöreskog (1977) and Jöreskog and Sörbom (1988). This analysis is carried out using the LISREL sub-model defined by Eqs. 1a and 1b. The exact standard errors of the parameter estimates are not included in the present comparison.

The stability of alienation data consist of 6 variables: two indicators (y_1 and y_2) of social economic status (SES), two indicators (y_3 and y_4) of alienation (A67) obtained in 1967 and the same indicators (y_5 and y_6) of alienation (A71) obtained in 1971. These variables were measured in a sample of 932 subjects. The sample covariance matrix is given in Jöreskog and Sörbom (1988, page 169). The following model is tested (discarding the subject subscript):

measurement model:

$$y_1 = \lambda_1 SES + \varepsilon_1 \qquad y_2 = \lambda_2 SES + \varepsilon_2$$
$$y_3 = \lambda_3 A67 + \varepsilon_3 \qquad y_4 = \lambda_4 A67 + \varepsilon_4$$
$$y_5 = \lambda_5 A71 + \varepsilon_5 \qquad y_6 = \lambda_6 A71 + \varepsilon_6$$

structural equation model:

$$A67 = \beta_1 \, SES + \zeta_{A67}$$
$$A71 = \beta_2 \, SES + \beta_3 \, A67 + \zeta_{A71}$$

Table 1: Results of the covariance structure analysis stability of alienation ($\chi^2(4)$=4.73).
Parameter estimates and standard errors (s.e.).

parameter	estimate	LISREL s.e.	fd s.e.	DFP s.e.
λ_1	0.9787	0.0616	0.0620	0.0612
λ_2	0.9221	0.0595	0.0598	0.0593
λ_3	0.5220	0.0422	0.0426	0.0421
β_1	-.5750	0.0564	0.0580	0.0594
β_2	-.2268	0.0524	0.0531	0.0541
β_3	0.6070	0.0511	0.0513	0.0519
var ζ_{A67}	4.8466	0.4681	0.4630	0.4608
var ζ_{A71}	4.0875	0.4048	0.4044	0.4003
var SES	6.8048	0.6500	0.6537	0.6431
var e_1	4.7357	0.4538	0.4570	0.4557
var e_2	2.5662	0.4037	0.4068	0.4044
var e_3	4.4040	0.5158	0.5179	0.5120
var e_4	3.0731	0.4349	0.4369	0.4373
var e_5	2.8052	0.5078	0.5125	0.4996
var e_6	2.6489	0.1816	0.1825	0.1785
cov[e_1e_3]	1.6247	0.3140	0.3159	0.3091
cov[e_2e_4]	0.3391	0.2614	0.2632	0.2570

The factor loadings λ_1, λ_3 and λ_5 were fixed to equal unity and the latent variances of SES, A67, and A71 were estimated. Furthermore the covariances among the error terms ε_1 and ε_3 (cov[$\varepsilon_1\varepsilon_3$]) and ε_2 and ε_4 (cov[$\varepsilon_2\varepsilon_4$]) were estimated. Overall goodness of fit was $\chi^2(4)$ = 4.73, p=.32. Table 1 contains the parameter estimates and the standard errors. The standard errors are also displayed in bar charts in Figure 3. The results in the present analysis again reveal very minor differences in the three methods of calculating standard errors.

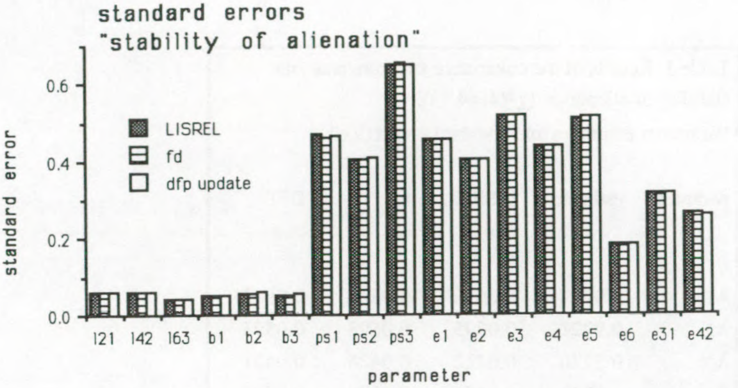


Figure 3: Bar chart of the standard errors obtained from the analysis of stability of alienation. Parameters estimates and standard errors given in Table 1.

6. Conclusion

In the present paper we have compared various ways of calculating standard errors in the analysis of covariance structure using normal theory ML estimation. It has been shown using simulated data that the fd approximation using exact gradients gives a good approximation to the exact Hessian. Furthermore the standard errors based on the DFP update of the Hessian and those provided by the LISREL program were very close to the fd standard errors. In each analysis using our own programs the starting values were set to equal 1.0 and all estimates were obtained within 30 to 40 iteration (LISREL VI generally converged much more rapidly). Thus it does not require an excessive number of iterations to obtain a good approximation of the Hessian using DFP updates. However some caution should be taken in generalizing the results regarding the DFP standard errors to other analyses and to other optimisation routines. The accuracy of these standard errors depends on the distance of the starting values to the minimum of the likelihood ratio. In the LISREL program, for instance, a modified DFP algorithm is used in which updates calculated during optimisation yield an approximation to the Hessian. Nonetheless LISREL standard errors are based on the Information matrix. It is likely that the generally good automatic starting values generated by LISREL, combined with a very efficient optimisation routine, render the required number of iterations so small that a reliable approximation to the Hessian matrix cannot be guaranteed. This and the fact that different updating algorithms may yield different standard errors (see Gill, Murray and Wright, 1981, page 120) probably prompted Lawley and Maxwell (1971, page 92) to maintain that it is theoretically better to recalculate the Information matrix at the minimum of the likelihood ratio, rather than rely on the the Hessian provided by the DFP algorithm.

Although derived under the assumption that the true model is being tested, the LISREL standard errors have been found to be excellent given a fairly gross model misspecification. This finding is corroborated by the analysis of the stability of alienation data set where very small differences were found between the fd and the LISREL standard errors.

It seems reasonable to advise anyone who has occasion to program his or her own routines for structural equation modeling to use either the fd approximation, or the Information matrix. The finite difference method has the advantage that it combines very light programming requirements with a very good approximation to the exact Hessian.

References.

Bollen, K. A. (1989). *Structural equations with latent variables.* John Wiley and Sons.

Dennis, Jr. J.E. and Schnabel, R.B. (1983). *Numerical methods for unconstrained optimization and nonlinear equations.* Prentice Hall.

Gill, P.G., Murray, W. and Wright, M.H. (1981). *Practical Optimization.* Academic Press.

Jöreskog, K.G. (1977). Structural equation models in the social sciences: specification, estimation and testing. In: P.R.Krishnaiah (Ed.).*Applications of Statistics.* North-Holland Publishing Company.

Jöreskog, K.G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, **43**, no. 4, 443-477.

Jöreskog, K.G. and Sörbom, D. (1988). *LISREL VII: A guide to the program and applications.* SPSS Inc.

Jöreskog, K.G. and Sörbom, D. (1986). *Prelis: A preprocessor for LISREL.* Scientific Software, Inc. Mooresville, USA.

Jöreskog, K.G. and Sörbom, D. (1984). *LISREL VI: Analysis of Linear structural relations by the method of maximum likelihood.* National Education resources. Chicago.

Lawley, D.N. and Maxwell, A.E. (1971). *Factor analysis as a statistical method.* London Butterworths.

Wheaton, B., Muthén, B., Alwin, D. and Summers, G. (1977). Assessing reliability and stability in panel models. In D.R. Heise (Ed.): *Sociological Methodology 1977.* San Fransisco: Jossey-Bass.

Appendix

The first order derivatives of the likelihood ratio function given in Eq. 3 for the model defined by Eqs. 1a and 1b (Jöreskog, 1977):

$$1/2\partial f/\partial \Lambda \quad = \Omega\, \Lambda\, (I-B)^{-1}\, \Psi\, (I-B^t)^{-1}$$

$$1/2\partial f/\partial B \quad = -(I-B^t)^{-1}\Lambda^t\, (\Omega\, \Lambda\, (I-B)^{-1}\, \Psi\, (I-B^t)^{-1})$$

$$\partial f/\partial \Psi \quad = (I-B^t)^{-1}\Lambda^t\, \Omega\, \Lambda\, (I-B)^{-1}$$

$$\partial f/\partial \Theta_\varepsilon \quad = \Omega$$

where $\Omega = \Sigma^{-1}(\Sigma\text{-}S)\, \Sigma^{-1}$ and matrices B, Λ, Ψ and Θ_ε are defined in the text.

For the model $\Sigma = \Lambda\Psi\Lambda^t + \Theta_\varepsilon$, the second derivative of the ML function with respect to Λ is obtained as follows: Let u_{ij} be the element on the i-th row and the j-th column of Ω (defined above), denoted $u_{ij} = [\Omega]_{ij}$. Furthermore: $v_{ir} = [\Omega\Lambda\Psi]_{ir}$, $w_{rs} = [\Psi\Lambda^t\Omega\Lambda\Psi]_{rs}$, $\sigma^{ij} = [\Sigma^{-1}]_{ij}$, $\eta_{ir} = [\Sigma^{-1}\, \Lambda\Psi]_{ir}$, $\mu_{rs} = [\Psi\Lambda^t\Sigma^{-1}\Lambda\Psi]_{rs}$ and $\psi_{rs} = [\Psi]_{rs}$. The Hessian matrix consists of the elements (Lawley and Maxwell, 1971, page 145):

$$1/2\, \partial^2 f_{ml}/\partial\lambda_{ir}\lambda_{js} = \sigma^{ij}\, [\mu_{rs} - w_{rs}] + \eta_{is}\eta_{jr} - \eta_{is}v_{jr} - \eta_{jr}v_{is} + [\psi_{rs} - \mu_{rs}]\, u_{ij}$$

ontvangen 15-2-1990
geaccepteerd 1-7-1991