INSTITUTE OF
FACULTY OF ACTUARIAL SCIENCE
&
ECONOMETRICS

REPORT AE 4/89

LISREL: Gradient and Hessian of the fitting function

H. Neudecker

A. Satorra

University of Amsterdam

Title: LISREL: Gradient and Hessian of the fitting function

Authors: H. Neudecker
A. Satorra*

Address: H. Neudecker
Institute of Actuarial Science and Econometrics
University of Amsterdam

A. Satorra
Department of Statistics and Econometrics
University of Barcelona

Corresponding address: Institute of Actuarial Science and Econometrics
University of Amsterdam
Jodenbreestraat 23
1011 NH AMSTERDAM

Date: May, 1989

Series and Number: Report AE 4/89

Pages: 10

Price: No charge

JEL Subject Classification: 211

Keywords: Covariance structure, Kronecker products, matrix differential calculus, (0,1)-matrices

Abstract:

Latent-variable models are nowadays frequently used in economic, social and behavioral studies to analyze relationships among variables. The LISREL model is a general model that integrates the classical simultaneous-equation model developed in econometrics with the factor-analysis model developed by psychometricians. The classical "errors-in-variables" model is also a particular case of LISREL. In this paper we obtain the hessian of a general type of fitting function for the LISREL model. Although the expressions of the first derivatives are known and widely used, the expressions obtained for the second derivatives are a novelty and may have practical implications. For instance, the expressions for the hessian would be needed to implement true Newton fitting algorithms, or when using observed hessians (instead of expected) in evaluating the asymptotic distribution of statistics of interest.

*The research of this author was made possible by a grant from Dutch Organization for Advancement of Pure Research (NWO)

1- Introduction

Latent-variable models are now frequently used in economic, social, and behavioral studies for analyzing structural relations among variables. Perhaps the most popular of these models is the so-called LISREL model (Jöreskog, 1977; Wiley, 1973), the analysis of which has become standard practice through Jöreskog and Sörbom (1983)'s computer program also named LISREL. The LISREL model integrates the classical simultaneous equation model developed in econometrics with the factor-analysis model developed by psychometricians. The classical "errors-in-variables" model is also a particular case of LISREL. Alternative formulations of latent-variable models, as Bentler and Weeks (1980)'s model implemented in the computer program EQS (Bentler, 1987), or the COSAN model of McDonald (1980) have been recognized to be equivalent specifications of the LISREL model. A particular type of analysis of the LISREL model suited for discrete data is implemented in Müthen (1988) 's program LISCOMP. A review of latent-variable models can be found in Anderson (1984), and Aigner, Hsiao, Kapteyn and Wansbeek (1984).

A maximum-likelihood analysis, under the assumption that the vector of observed variables is normally distributed, as well as more general distribution-free methods, are implemented in most of the above-mentioned computer programs. Covariance structure analysis is a general framework for fitting latent-variable models. Within this framework, if θ denotes the vector containing all the unknown parameters of the model, and Σ the population covariance matrix of the vector of all observed variables, then the covariance structure, $\Sigma = \Sigma(\theta)$, implied by the specified model is fitted to the corresponding sample covariance matrix S. The model is fitted by minimizing with respect to θ a (non-negative) real-valued function of $F=F(S,\Sigma(\theta))$ of θ. Numerical iterative methods are required to achieve such a fit. As far as we know, all the above-mentioned programs use only the first derivatives of F, thus quasi-Newton optimization methods and expected hessians are used. Browne (1984), Shapiro (1987) and Satorra (1989) deal with different theoretical aspects of covariance structure analysis. Lee and Jennrich (1979) study several algorithms which are used in the practice of covariance structure analysis.

This paper provides an expression for the second derivatives of a general type of fitting function used in the analysis of the LISREL model. Standard matrix differential calculus, as in Magnus and Neudecker (1986), will be used. Although the expressions of the first derivatives are known and widely used, they will also be given as being obtained on the way to the second derivatives. The expressions obtained for the second derivatives are a novelty and may have practical implications. For instance, the expressions for the hessian would be needed to implement true Newton fitting algorithms, or when using observed hessians (instead of expected) for evaluating the asymptotic distribution of statistics of interest.

The paper is structured as follows. Section 2 describes the model. Sections 3 and 4 obtain respectively the first and second derivatives of the matrix valued function $\Sigma = \Sigma(\theta)$. Finally, section 5 integrates the results of the previous sections deriving the desired gradient and hessian expressions.

2. The model

A general linear statistical relation that combines measurement and structural equations is the following :

$$(2.1) \quad \begin{cases} z = \Lambda\eta + \varepsilon \\ \eta = B_0\eta + \xi , \end{cases}$$

where z represents a p-vector of observable variables, and η , ε and ξ are random vectors such that ε is uncorrelated with ξ. The matrices Λ and B_0, and the variance matrices of ε and ξ, Ψ and Φ respectively, contain the parameters of the model. The specific form of these parameter matrices gives rise to particular models.

It can be shown that (2.1) is both a specific case and a generalization of the LISREL model (Jöreskog, 1977; Wiley, 1973), which has the following specification

$$(2.2) \quad \begin{cases} y = \Lambda_y\eta + \varepsilon \\ x = \Lambda_x\xi + \delta \\ \eta = B_0\eta + \Gamma\xi + \zeta, \end{cases}$$

where ε, δ , ξ and ζ are mutually uncorrelated random variables (and correspond to new notation with respect to the one of (2.1)) with variance matrices Θ_ε, Θ_δ , Φ and Ψ respectively. Λ_y, Λ_x, B_0 and Γ represent matrices of appropriate dimension. Rewriting (2.2) as

$$(2.3\text{-}1) \quad \begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} \Lambda_y & 0 \\ 0 & \Lambda_x \end{pmatrix} \begin{pmatrix} \eta \\ \xi \end{pmatrix} + \begin{pmatrix} \varepsilon \\ \delta \end{pmatrix}$$

$$(2.3\text{-}2) \quad \begin{pmatrix} \eta \\ \xi \end{pmatrix} = \begin{pmatrix} B_0 & \Gamma \\ 0 & I \end{pmatrix} \begin{pmatrix} \eta \\ \xi \end{pmatrix} + \begin{pmatrix} \zeta \\ 0 \end{pmatrix} ,$$

we see that the LISREL model is a special case of (2.1). Further, one sees immediately that (2.1) is of the form (2.2), where the second equation and the component $\Gamma\xi$ on the right-hand side of the third equation in (2.2) have been dropped.

Equations (2.1) with the assumptions of zero correlation between ε and ξ, imply the following <u>covariance structure</u> for the matrix Σ of variances of z:

$$(2.4) \qquad \Sigma = \Lambda B^{-1} \Phi B^{-T} \Lambda' + \Psi,$$

where $B \equiv (I - B_0)$ is supposed to be non-singular.

Prior information with regard to the form of B_0, Λ, Φ and Ψ, yields more specific models; for instance, in LISREL one may restrict some of the elements of the matrices B_0, Λ, Φ and Ψ to have equal specific values, or to be equal among themslves. Given a specific model, the distinct and functionally unrelated unknown elements of the matrices B_0, Λ, Φ and Ψ will be assembled into a say q-dimensional parameter vector θ.

We will also consider the following vector δ of parameters of model (2.1) unrestricted:

$$\delta = [\, (\text{vec } \Lambda)' \mid (\text{vech } \Phi)' \mid (\text{vec } B)' \mid (\text{vech } \Psi)' \,]'$$

Usually, without further restrictions added to (2.1), the parameter vector δ will not be identified by Σ. A specific model will induce a function $\delta = \delta(\theta)$ expressing δ in terms of a parameter vector θ of smaller dimension. For model (2.1) with the type of restrictions considered, the function $\delta = \delta\,(.)$ will be regular enough in order to be twice continuously differentiable. In fact, in LISREL the derivative matrix $A \equiv \partial\delta \,/\, \partial\,\theta'$ is a constant matrix of zeroes, ones and minus ones. Moreover, equality (2.4) implies that $\sigma \equiv \mathrm{vec}\,\Sigma$ is a function of δ, hence of θ.

3- First derivatives of $\sigma = \sigma\,(\,\delta)$

Vectorizing (2.4), and taking differentials, we get

$$\begin{aligned} d\,\mathrm{vec}\,\Sigma \;=\; & (I + K\,)(\Lambda\, B^{-1}\Phi\, B^{-T} \;\otimes\; I\,)\; d\,\mathrm{vec}\,\Lambda \\ & - (I \;+ K\,)(\Lambda\, B^{-1}\Phi\, B^{-T} \;\otimes\; \Lambda\, B^{-1})\;\; d\,\mathrm{vec}\,B \\ & + \;(\Lambda\, B^{-1} \otimes\; \Lambda\, B^{-1})\;\; d\,\mathrm{vec}\,\Phi \\ & \qquad\qquad + \;d\,\mathrm{vec}\,\Psi\;\;. \end{aligned}$$

where "d" denotes differential, "$\otimes$" denotes kronecker product, I is the identity matrix of appropriate order, and K is a commutation matrix. Using now vec A = D vech A and vech A = L vec A , for symmetric A, where D and L are respectively the duplication and elimination matrices of Magnus & Neudecker (1986), we get

$$\begin{aligned} d\,\mathrm{vech}\,\Sigma \;=\; & L(I + K)\,(\Lambda\, B^{-1}\Phi\, B^{-T} \;\otimes\; I\,)\;\; d\,\mathrm{vec}\,\Lambda \\ & - L(I \;+ K)\;\;(\Lambda\, B^{-1}\Phi\, B^{-T} \;\otimes\; \Lambda\, B^{-1})\;\; d\,\mathrm{vec}\,B \;\;+ \\ & + L\,(\Lambda\, B^{-1} \otimes\; \Lambda\, B^{-1})\;\; D\;\; d\,\mathrm{vech}\,\Phi \\ & + \quad d\,\mathrm{vech}\,\Psi\;\;; \end{aligned}$$

which says that

$$d\,\mathrm{vech}\,\Sigma \;=\; G\; d\delta\;,$$

with

$$\begin{aligned} G = \;[\; & L(I + K)\;\;(\Lambda\, B^{-1}\Phi\, B^{-T} \;\otimes\; I\,) \;\;|\;\;\; L\,(\Lambda\, B^{-1} \otimes\; \Lambda\, B^{-1})\;\; D \;\;\;| \\ & - L(I + K)\,(\Lambda\, B^{-1}\Phi\, B^{-T} \;\otimes\; \Lambda\, B^{-1})\,|\;\, I\;\;] \end{aligned}$$

4- Second derivatives of $\Sigma = \Sigma(\theta)$

Consider

$$\partial^2 \Sigma_{ij} / \partial\delta\, \partial\delta' = \begin{array}{c} \begin{array}{cccc} (\Lambda & \Phi & B & \Psi) \end{array} \\ \left[\begin{array}{cccc} H_{11} & H_{12} & H_{13} & 0 \\ & 0 & H_{23} & 0 \\ \mathit{Sym} & & H_{33} & 0 \\ & & & 0 \end{array}\right] \end{array} ,$$

Differentiating the expression of d vec Σ obtained in the section above, we get

$$H_{11} = B^{-1}\Phi B^{-T} \otimes T_{ij}$$

$$H_{12} = (B^{-1} \otimes T_{ij} \Lambda B^{-1}) D$$

$$H_{13} = -(B^{-1}\Phi B^{-T} \otimes T_{ij} \Lambda B^{-1}) - K(T_{ij} \Lambda B^{-1}\Phi B^{-T} \otimes B^{-1})$$

$$H_{23} = -D'(B^{-T} \otimes B^{-T}\Lambda' T_{ij} \Lambda B^{-1})$$

$$\begin{aligned} H_{33} = \; & (B^{-1}\Phi B^{-T}\Lambda' T_{ij} \Lambda B^{-1} \otimes B^{-T})K + \\ & K(B^{-T}\Lambda' T_{ij} \Lambda B^{-1}\Phi B^{-T} \otimes B^{-1}) + \\ & (B^{-1}\Phi B^{-T} \otimes B^{-T}\Lambda' T_{ij} \Lambda B^{-1}), \end{aligned}$$

where $E_{ij} = e_i e_j'$ and $T_{ij} = E_{ij} + E_{ji}$, and zero matrices for the elements of the partition above which are marked with "0".

5- Gradient and Hessian of a (weighted) least squares fitting function F

In the practice of covariance structure it is usual to minimize the (weighted) least-squares fitting function

(5.1) $F(\theta) = (s - \sigma(\theta))' W (s - \sigma(\theta))$;

where W is a possibly stochastic (positive semi-definite) matrix. This is the fitting function used in generalized least-squares estimation implemented in most of the computer programs for covariance structure analysis (EQS, LISREL, LISCOMP, etc.). Also maximum-likelihood estimation (under the assumption that z is normally distributed) has been shown to yield the same estimators as a weighted least-squares analysis where W is updated after each iteration (see, e.g., Browne, 1973; Lee and Jennrich, 1979; Fuller, 1987, Section 4.2.2)

The weighted least-squares function (5.1), with W supposed not to depend on θ, implies the following expression for the gradient of F

(3) $\partial F / \partial \delta' = -2(s - \sigma)' W (\partial\sigma / \partial\delta')$.

Differentiating dF, we get

$$d^2F = 2(d\sigma)' W d\sigma - 2 (s - \sigma)' W d^2\sigma =$$

$$2(d\sigma)' W d\sigma - 2 \Sigma_r ((s - \sigma)' W)_r (d^2\sigma)_r ,$$

where the sum Σ_r runs over the indexes (i,j) corresponding to the $p^* = p(p+1)/2$ distinct elements of Σ.

Using now $d\sigma = G\, d\delta$ and $(d^2\sigma)_r . = (d\delta)' H_r\, d\delta$, with the matrices G and H_r as obtained in the sections above, we get

$$d^2 F = 2(d\delta)' [G'W G - \Sigma_r ((s - \sigma)' W)_r H_r](d\delta) ;$$

hence, the hessian of F is expressed as

$$\partial^2 F / \partial\delta \partial\delta' = 2 [G'W G - \Sigma_r \{ (\text{vech}(S - \Sigma))' W \}_r H_r],$$

where

$$\Sigma_r \{ (\text{vech}(S - \Sigma))' W \}_r H_r \quad = \quad \Sigma_r \ (\Sigma_h (s_h - \sigma_h) w_{hr}) H_r ,$$

where w_{hr} is the hr-th element of W, and $H_r \equiv \partial \Sigma_{ij} / \partial\delta \partial\delta'$ for corresponding index ij.

Note that from the gradient and hessian given above, we obtain immediately the gradient and hessian of a specific model. Effectively, such specific model has associated a function $\delta = \delta(\theta)$ with corresponding derivative matrix $A \equiv \partial \delta(\theta) / \partial \theta'$.

References

Aigner, D.J., Hsiao, C., Kapteyn, A., & Wansbeek, T. (1984). Latent variable models in econometrics. In Z. Griliches & M. D. Intriligator (Eds.)., Handbook of econometrics, Vol. 2 (pp. 1321-1393). Amsterdam: North-Holland.

Anderson, T.W. (1984). Estimating Linear Statistical Relationships. The Annals of Statistics, 12, 1, 1-45.

Bentler, P.M. (1985). Theory and implementation of EQS, a structural equations program. Los Angeles: BMDP Statistical Software.

Bentler, P.M., & Weeks, D. G. (1980). Linear structural equations with latent variables. Psychometrika, 45, 289-308.

Browne, M. W. (1974). Generalized Least Squares Estimators in the Analysis of Covariance Structures. South African Statistical Journal, 8, 1-24.

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

Fuller, W. (1987). Measurement Error Models. New York: Wiley.

Jöreskog, K.G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P.R. Krishnaiah (Ed.), Application of statistics (p. 265-287). Amsterdam: Norh-Holland

Jöreskog, K. G. and D. Sörbom, (1983). LISREL V User's Guide. Analysis of Structural Equation Relationship by Maximum Likelihood and Least Squares Methods. Chicago: International Educational Services.

Lee, S. -Y., and Jennrich, R. I. (1979). A Study of Algorithms for Covariance Structure Analysis with Specific Comparison Using Factor Analysis. Psychometrika 44, 99-113.

Magnus, J.R. & Neudecker, H. (1986). Symmetry, 0-1 matrices and jacobians, Econometric Theory, 2, 157-190

McDonald, R.P. (1980). A simple comprehensive model for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 31, 59-72.

Muthén, B. (1987). LISCOMP: Software for advanced analysis of L.S.E. with a comprehensive measurement model. Mooresville, IN: Scientific Software.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: a unified approach, Psychometrika, 54, 1, 131-151.

Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. Journal of the American Statistical Association, 81, 142-149.

Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.) Structural equation models in the social sciences (pp. 69-83). New York: Seminar.