[Amsterdam. University. Institute of actuarial science and econometrics]

H. Neudecker

A derivation of the Hessian of the (concentrated) likelihood function of the factor model employing the Schur product. AE5/76

Classification

AMS (MOS) subject classification scheme (1970): 15A18, 26A60, 62H25, 62P15.

Keywords: Hessian, likelihood function, Schur product

Abstract: An expression is derived for the Hessian of the (concentrated) likelihood function of the factor model, employing matrix differential calculus in conjunction with the Schur product. The results of Clarke are achieved by an alternative route.

(A short version of the paper appeared in the British Journal of math. statist. Psychology, 1975, p. 152-156).

Free copies are obtainable from
Interfakulteitsbureau Aktuariaat en Ekonometrie
Jodenbreestraat 23 - Amsterdam, 'phone: 020-5254212

H. Neudecker*)

A derivation of the Hessian of the (concentrated) likelihood function of the factor model, by employing the Schur product.

Introduction

Clarke [1] recently derived the typical element of the Hessian of the (concentrated) likelihood function of the factor model

$$x = \Lambda f + e$$

where e and f are multinormal with

$$Ee = 0,\ Ef = 0,\ Efe' = 0,\ Eee' = \Psi,\ Eff' = I_k,$$

Ψ is diagonal of order p, both Λ and Ψ are unknown parameter matrices. It is the aim of this paper to derive an expression for the Hessian by employing matrix differential calculus in conjunction with the Schur product.

We first prove a lemma on Schur multiplication.

A lemma on Schur multiplication

Lemma 1: If M is a diagonal matrix, then

$$\operatorname{tr} AMB'M = m'(A.B)m,$$

where $A.B = [a_{ij}b_{ij}]$ (the Schur product of A and B)

$m = Ms$, $s' = (1\ldots.1)$.

Proof: $\operatorname{tr} AMB'M = \sum_i (AMB'M)_{ii} = \sum_{ijk\ell} a_{ij}m_{jk}b_{\ell k}m_{\ell i}$

$$= \sum_{ij} a_{ij}m_{jj}b_{ij}m_{ii} = \sum_{ij} m_{ii}(a_{ij}b_{ij})m_{jj} = m'(A.B)m \quad .$$

The factor model and its (concentrated) likelihood function

Taking up the model again we find $Ex = 0$ and $Exx' = \Lambda\Lambda' + \Psi = \Sigma$, say. Let x_i be the vector of the i^{th} set of observations,

$$\bar{x} = \frac{1}{n}\sum_i x_i \quad (i=1\ldots n), \quad S = \frac{1}{n-1}\left(\sum_i x_i x_i' - n\bar{x}\bar{x}'\right) .$$

The log-likelihood of the sample variance S of the sample of independent vectors $x_1 \ldots x_n$ is

$\phi = \gamma - \frac{1}{2}(n-1)\{\log|\Sigma| + \operatorname{tr} S\Sigma^{-1}\}$, a function of (the unknown parameters) Λ and Ψ .

---

* Instituut voor aktuariaat en ekonometrie, Universiteit van Amsterdam

Jöreskog [2] has derived the concentrated log-likelihood as a function of Ψ, more precisely as a function of the k largest latent roots $\theta_1 \ldots \theta_k$ of $S^* = \Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}$. He has shown that maximizing this concentrated log-likelihood id equivalent with minimizing $\phi_k = \mathrm{tr}\Theta_2 - \log|\Theta_2| - (p-k)$, where Θ_2 is the digonal matrix of p-k smallest latent roots $\theta_{k+1} \ldots \theta_p$ of S^*.
Clarke [1] subsequently computed the second derivatives of ϕ_k with respect to the element ψ_{ii} of Ψ.

The derivation of the Hessian

Following Clarke we assure that S^* has distinct latent roots $\theta_1 \ldots \theta_k$ and $\theta_{k+1} \ldots \theta_p$ with associated (standard) latent vectors $w_1 \ldots w_k$ and $w_{k+1} \ldots w_p$. The following properties are needed:

I) $d\theta_r = -\theta_r w_r' \Psi^{-1}(d\Psi) w_r$

II) $dw_r = -\frac{1}{2} \sum\limits_{j \neq r} \frac{\theta_r + \theta_j}{\theta_r - \theta_j} w_j w_j' \Psi^{-1}(d\Psi) w_r$

III) $$d^2\theta_r = \theta_r [w_r' \Psi^{-1}(d\Psi) w_r]^2 + \theta_r \sum_{j \neq r} \frac{\theta_r + \theta_j}{\theta_r - \theta_j} [w_r' \Psi^{-1}(d\Psi) w_j]^2$$
$$+ \theta_r w_r' \Psi^{-1}(d\Psi) \Psi^{-1}(d\Psi) w_r .$$

These can be found in Clarke [1], but can be established in a more compact fashion. We shall do so.

We start from $$S^* w_r = \theta_r w_r \tag{1}$$

with $$w_r' w_r = 1 \tag{2}$$

or equivalently $$w_r' dw_r = 0 \tag{2*}$$

where dw_r is the differential of w_r.

Further $$w_j' w_r = 0 \quad (j \neq r) \tag{3}$$

and $$\sum_j w_j w_j' = I \tag{4}$$

We then have:

$$(dS^*) w_r + S^* dw_r = (d\theta_r) w_r + \theta_r dw_r$$

$$w_r'(dS^*) w_r + w_r' S^* dw_r = (d\theta_r) w_r' w_r + \theta_r w_r' dw_r ,$$

from which follows

$$d\theta_r = w_r'(dS^*)w_r \tag{5}$$

because of (1) and (2).

Further:

$$w_j'(dS^*)w_r + w_j'S^*dw_r = (d\theta_r)w_j'w_r + \theta_r w_j'dw_r \quad (j\neq r)$$

or $\theta_r w_j'dw_r = w_j'(dS^*)w_r + \theta_j w_j'dw_r$

because of (1) and (3),

$$w_j'dw_r = \frac{1}{\theta_r-\theta_j} w_j'(dS^*)w_r$$

$$\sum_{j\neq r} w_j w_j'dw_r = \sum_{j\neq r} \frac{1}{\theta_r-\theta_j} w_j w_j'(dS^*)w_r ,$$

hence $dw_r = \sum_{j\neq r} \frac{1}{\theta_r-\theta_j} w_j w_j'(dS^*)w_r$ (6)

because of (4) and (2^*).

From $S^* = \Psi^{-\frac{1}{2}}S\Psi^{-\frac{1}{2}}$, Ψ diagonal, it follows that

$$dS^* = -\tfrac{1}{2}\Psi^{-1}(d\Psi)S^* - \tfrac{1}{2}S^*\Psi^{-1}d\Psi.$$

We rewrite (5):

$$d\theta_r = -\tfrac{1}{2}w_r'\Psi^{-1}(d\Psi)S^*w_r - \tfrac{1}{2}w_r'S^*\Psi^{-1}(d\Psi)w_r$$

$$= - w_r'S^*\Psi^{-1}(d\Psi)w_r = - \theta_r w_r'\Psi^{-1}(d\Psi)w_r \tag{I}$$

because of (1).

We rewrite (6):

$$dw_r = \sum_{j\neq r} \frac{1}{\theta_r-\theta_j} w_j w_r'(dS^*)w_r$$

$$= -\tfrac{1}{2} \sum_{j\neq r} \frac{1}{\theta_r-\theta_j} w_j w_j'\Psi^{-1}(d\Psi)S^*w_r - \tfrac{1}{2} \sum_{j\neq r} \frac{1}{\theta_r-\theta_j} w_j w_j'S^*\Psi^{-1}(d\Psi)w_r$$

$$= -\tfrac{1}{2} \sum_{j\neq r} \frac{\theta_r+\theta_j}{\theta_r-\theta_j} w_j w_j'\Psi^{-1}(d\Psi)w_r \tag{II}$$

because of (1).

For $d^2\theta_r$ we find:

$$d^2\theta_r = -(d\theta_r)w_r'\Psi^{-1}(d\Psi)w_r - 2\theta_r w_r'\Psi^{-1}(d\Psi)dw_r$$

$$+\ \theta_r w_r'\Psi^{-1}(d\Psi)\Psi^{-1}(d\Psi)w_r, \text{ because of (I)}$$

$$= \theta_r\left[w_r'\Psi^{-1}(d\Psi)w_r\right]^2 + \theta_r w_r'\Psi^{-1}(d\Psi)\sum_{j\neq r}\frac{\theta_r+\theta_j}{\theta_r-\theta_j}\, w_j w_j'\Psi^{-1}(d\Psi)w_r$$

$$+\ \theta_r w_r'\Psi^{-1}(d\Psi)\Psi^{-1}(d\Psi)w_r$$

$$= \theta_r\left[w_r'\Psi^{-1}(d\Psi)w_r\right]^2 + \theta_r\sum_{j\neq r}\frac{\theta_r+\theta_j}{\theta_r-\theta_j}\left[w_r'\Psi^{-1}(d\Psi)w_j\right]^2$$

$$+\ \theta_r w_r'\Psi^{-1}(d\Psi)\Psi^{-1}(d\Psi)w_r\ . \qquad \text{(III)}$$

We shall now differentiate ϕ_k. This leads to

$$d\phi_k = \mathrm{tr}\,d\Theta_2 - \mathrm{tr}\,\Theta_2^{-1}\,d\Theta_2\ ,$$

and
$$d^2\phi_k = \mathrm{tr}\,d^2\Theta_2 + \mathrm{tr}\,\Theta_2^{-1}(d\Theta_2)\Theta_2^{-1}d\Theta_2 - \mathrm{tr}\,\Theta_2^{-1}d^2\Theta_2$$

$$= \sum_{r=k+1}^{p}\left(1-\frac{1}{\theta_r}\right)d^2\theta_r + \sum_{r=k+1}^{p}\frac{1}{\theta_r^2}(d\theta_r)^2$$

$$= \sum_r\left(1-\frac{1}{\theta_r}\right)\theta_r\left[w_r'\Psi^{-1}(d\Psi)w_r\right]^2$$

$$+ \sum_r\left(1-\frac{1}{\theta_r}\right)\theta_r\sum_{j\neq r}\frac{\theta_r+\theta_j}{\theta_r-\theta_j}\left[w_r'\Psi^{-1}(d\Psi)w_j\right]^2 + \sum_r\left(1-\frac{1}{\theta_r}\right)\theta_r w_r'\Psi^{-1}(d\Psi)\Psi^{-1}(d\Psi)w_r$$

$$+ \sum_r\frac{1}{\theta_r^2}\,\theta_r^2\left[w_r'\Psi^{-1}(d\Psi)w_r\right]^2$$

$$= \sum_r\theta_r\left[w_r'\Psi^{-1}(d\Psi)w_r\right]^2 + \sum_r\sum_{j=r}(\theta_r-1)\frac{\theta_r+\theta_j}{\theta_r-\theta_j}\left[w_r'\Psi^{-1}(d\Psi)w_j\right]^2$$

$$+ \sum_r(\theta_r-1)w_r'\Psi^{-1}(d\Psi)\Psi^{-1}(d\Psi)w_r$$

$$= \sum_r\mathrm{tr}\,\theta_r\Psi^{-1}w_r w_r'\Psi^{-1}(d\Psi)w_r w_r'd\Psi$$

$$+ \sum_r\sum_{j\neq r}\mathrm{tr}(\theta_r-1)\frac{\theta_r+\theta_j}{\theta_r-\theta_j}\,\Psi^{-1}w_r w_r'\Psi^{-1}(d\Psi)w_j w_j'd\Psi$$

$$+ \sum_r\mathrm{tr}(\theta_r-1)w_r w_r'\Psi^{-1}(d\Psi)\Psi^{-1}d\Psi$$

$$= (d\psi)'\left[\sum_r \theta_r \Psi^{-1} w_r w_r' \Psi^{-1} . w_r w_r'\right] d\psi$$

$$+ (d\psi)'\left[\sum_r \sum_{j\neq r} (\theta_r - 1) \frac{\theta_r + \theta_j}{\theta_r - \theta_j} \Psi^{-1} w_r w_r' \Psi^{-1} . w_j w_j'\right] d\psi$$

$$+ (d\psi)'\left[\sum_r (\theta_r - 1) w_r w_r' \Psi^{-1} . \Psi^{-1}\right] d\psi \quad ,$$

where $\psi = \Psi s$.

We now employ the

Lemma 2: Let ϕ be a double differentiable scalar function of the vector x. From $d^2\phi = (dx)'\,Adx$ follows that the Hessian of ϕ is $\frac{1}{2}(A+A')$. See [4].

This leads to the results

$$H_\psi = \Psi^{-1}(\sum_r \theta_r w_r w_r' . w_r w_r')\Psi^{-1}$$

$$+ \Psi^{-1}(\sum_r \sum_{j\neq r} (\theta_r - 1) \frac{\theta_r + \theta_j}{\theta_r - \theta_j} w_r w_r' . w_j w_j')\Psi^{-1}$$

$$+ \Psi^{-1}(\sum_r (\theta_r - 1) w_r w_r' . I)\Psi^{-1},$$ where H_ψ denotes the Hessian.

It can be shown easily that this expression is equivalent to the matrix version of Clarke's formula 7. [1]:

$$\Psi^{-1}(\Omega_2\Omega_2' \,.\, \Omega_2\Omega_2')\Psi^{-1} - 2\Psi^{-1}(\Omega_2(I-\Omega_2)\Omega_2' . I)\Psi^{-1}$$

$$+ \;\; 2\Psi^{-1}(\sum_r \sum_j \frac{\theta_j(\theta_r - 1)}{\theta_r - \theta_j} w_r w_r' . w_j w_j' \;)\Psi^{-1} \; .$$

Write $w_r w_r' = W_r$, $w_j w_j' = W_j$.

Hence $\sum_r \sum_{j\neq r} (\theta_r - 1) \frac{\theta_r + \theta_j}{\theta_r - \theta_j} W_r \, . \, W_j$

---

1) A factor $\frac{1}{\psi_{ii}\psi_{\ell\ell}}$ is missing in the last term of Clarke's expression.

$$= \sum_r \sum_j \frac{\theta_r(\theta_r-1)}{\theta_r-\theta_j} W_r \cdot W_j + 2 \sum_r \sum_j \frac{\theta_j(\theta_r-1)}{\theta_r-\theta_j} W_r \cdot W_j$$

$$+ \sum_r \sum_{j\neq r} \frac{\theta_r(\theta_r+\theta_j)}{\theta_r-\theta_j} W_r \cdot W_j - \sum_r \sum_j \frac{\theta_j(\theta_r-1)}{\theta_r-\theta_j} W_r \cdot W_j$$

$$= 2 \sum_r \sum_j \frac{\theta_j(\theta_r-1)}{\theta_r-\theta_j} W_r \cdot W_j + \sum_r \sum_j (\theta_r-1) W_r \cdot W_j$$

$$+ \sum_r \sum_{j\neq r} \frac{\theta_r^{\,2}}{\theta_r-\theta_j} W_r \cdot W_j$$

$$= 2 \sum_r \sum_j \frac{\theta_j(\theta_r-1)}{\theta_r-\theta_j} W_r \cdot W_j + \sum_r \sum_j (\theta_r-1) W_r \cdot W_j$$

$$+ \tfrac{1}{2} \sum_r \sum_{j\neq r} (\theta_r+\theta_j) W_r \cdot W_j$$

$$= 2 \sum_r \sum_j \frac{\theta_j(\theta_r-1)}{\theta_r-\theta_j} W_r \cdot W_j + \Omega_2(\Theta_2-1)\, \Omega_2' \cdot \Omega_1\Omega_1'$$

$$+ \Omega_2\Theta_2\Omega_2' \cdot \Omega_2\Omega_2' - \sum_r \theta_r W_r \cdot W_r \,.$$

$$\text{So} \sum_r \sum_{j\neq r} (\theta_r-1) \frac{\theta_r+\theta_j}{\theta_r-\theta_j} W_r \cdot W_j + \sum_r \theta_r W_r \cdot W_r$$

$$+ \sum_r (\theta_r-1) W_r \cdot I = 2 \sum_r \sum_j \frac{\theta_j(\theta_r-1)}{\theta_r-\theta_j} W_r \cdot W_j$$

$$+ \Omega_2(\Theta_2 - I)\Omega_2' \cdot \Omega_1\Omega_1' + \Omega_2(\Theta_2 - I)\Omega_2' \cdot I + \Omega_2\Theta_2\Omega_2' \cdot \Omega_2\Omega_2'$$

$$= 2 \sum_r \sum_j \frac{\theta_j(\theta_r - 1)}{\theta_r - \theta_j} W_r \cdot W_j - 2\Omega_2(I - \Theta_2)\Omega_2' \cdot I$$

$$+ \Omega_2\Omega_2' \cdot \Omega_2\Omega_2'.$$

Literature

(1) Clarke, M.R.B. (1970). A rapidly convergent method for maximum-likelihood factor analysis. Br. J. math.statist.psychol. 23, 43-52.

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

(3) Lawley, D.N. & Maxwell, A.E. (1971). Factor Analysis as a Statistical Method. London: Butterworths.

(4) Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. JASA 64, 953-963.