Instituut voor Actuariaat & Econometrie

Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix.

by Jan R. Magnus

forthcoming in the JOURNAL OF ECONOMETRICS

January 1977
revised June 1977

Address: 23 Jodenbreestraat, Amsterdam

Universiteit van Amsterdam

Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix.

Jan R. Magnus*)
University of Amsterdam, Amsterdam, The Netherlands

1. Introduction

In this paper we consider the regression model $y = X\beta + \varepsilon$ with all the classical assumptions (including normality) but one, viz. we assume that the covariance matrix of the disturbances depends upon a finite number of unknown parameters $\theta_1 \ldots \theta_m$. If the parameters $\theta_1 \ldots \theta_m$ were known, the Aitken estimator would be the BLU and maximum likelihood estimator. Since we assume that the θ's are unknown, we are faced with the problem to estimate the β's and the θ's simultaneously. In sections three and four we derive the first and second order conditions and the information matrix for the ML estimators of β and θ. These appear to be surprisingly simple. The next two sections are devoted to the properties of the ML estimators and to an algorithm that leads, under general conditions, to a solution of the ML equations. In section seven we apply these formulae to a general case, which facilitates the derivation of the ML estimators and the information matrix in the last two sections which are devoted to the autocorrelated errors model and to Zellner-type regressions. It is known from the literature that iterative Zellner and iterative Cochrane-Orcutt are equivalent with the ML estimates. In the present paper these iterative estimators appear as corollaries of much more general cases.

---

*) I wish to express my gratitude to prof H. Neudecker, who advised and encouraged me in this research. I am also indebted to R.D.H. Heijmans for stimulating discussions on the statistical part of this paper.

2. The "vec"-function and the Kronecker product

Let $A = [a_{ij}]$ be an (m,n) matrix[1] and $A_{.j}$ the jth column of A, then vec A is the (mn) column vector

$$\text{vec } A = \begin{pmatrix} A_{.1} \\ \vdots \\ A_{.n} \end{pmatrix} .$$

Let further Q be an (s,t) matrix, then the Kronecker product $A \otimes Q$ is defined as the (ms,nt) matrix

$$A \otimes Q = [a_{ij}Q].$$

An important connection between the vec-function and the Kronecker product is[2]

$$\text{vec } ABC = (C' \otimes A) \text{ vec } B, \tag{1}$$

where A is (m,n), B is (n,p) and C is (p,q).
Special cases of (1) are

$$\text{vec } AB = (I_p \otimes A) \text{ vec } B = (B' \otimes I_m) \text{ vec } A = (B' \otimes A) \text{ vec } I_n . \tag{2}$$

The basic connection between the vec-function and the trace is

$$\text{tr } AZ = (\text{vec } A')'(\text{vec } Z), \tag{3}$$

where Z is an (n,m) matrix.
From (3) we derive the more complicated expressions

$$\begin{aligned} \text{tr } ABCD &= (\text{vec } B')'(A' \otimes C)\text{vec } D = (\text{vec } C')'(B' \otimes D)\text{vec } A \\ &= (\text{vec } D')'(C' \otimes A)\text{vec } B = (\text{vec } A')'(D' \otimes B)\text{vec } C, \end{aligned} \tag{4}$$

where D is a (q,m) matrix.
We now use (4) to establish the most general formula

$$\text{tr } ABCEF = (\text{vec } E')'(C' \otimes FA)\text{vec } B = (\text{vec } E)'(FA \otimes C')\text{vec } B', \tag{5}$$

where E and F are matrices of orders (q,r) and (r,m) respectively.

---

1) A matrix of order (m,n) is one having m rows and n columns.

2) A collection of theorems on Kronecker products and matrix differentiation has been given by Neudecker (1969).

For easy reference we state the following special case of (4):

$$\operatorname{tr} GVHV = (\operatorname{vec} G)'(V \otimes V)\operatorname{vec} H, \tag{6}$$

where G,H and V are symmetric matrices.
Finally,

$$x'AB = (\operatorname{vec} A)'(B \otimes x) = (\operatorname{vec} A')'(x \otimes B), \tag{7}$$

where x is an (m,1) vector.

3. The maximum likelihood equations

Consider the linear regression model

$$y = X\beta + \varepsilon, \tag{8}$$

where y is an (n,1) vector of observations on the dependent variable, X is an (n,k) matrix of the values of the regressors, β is a (k,1) vector of the regression coefficients, and ε is an (n,1) disturbance vector.
We shall make the following assumptions:

ASSUMPTION 1: ε is normally distributed .

ASSUMPTION 2: $E\varepsilon=0$, $E\varepsilon\varepsilon'=\Omega$, where Ω is a positive definite (hence nonsingular) matrix whose elements are twice differentiable functions of a finite and constant number of parameters $\theta_1,\theta_2,\ldots,\theta_m$, i.e. $\Omega=\Omega(\theta)$, $\theta \in \Theta$.

ASSUMPTION 3: X is a fixed matrix of full rank and $n > k$.

ASSUMPTION 4: The parameters in β are independent from those in θ.[3)] .

Theorem 1

The linear regression model (8) under the assumptions (1)-(4) has the following first-order ML conditions

$$\text{(i)} \quad \hat{\beta} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y \tag{9}$$

$$\text{(ii)} \quad \operatorname{tr}\left(\frac{\partial\Omega^{-1}}{\partial\theta_h}\right)_{\theta=\hat{\theta}} \hat{\Omega} = e'\left(\frac{\partial\Omega^{-1}}{\partial\theta_h}\right)_{\theta=\hat{\theta}} e, \quad \text{where } e=y-X\hat{\beta} \ . \quad (h=1\ldots m) \tag{10}$$

Further if $|\Omega|$, the determinant of Ω, does not depend upon θ_j, the jth equation in (10) reduces to

$$e'\left(\frac{\partial\Omega^{-1}}{\partial\theta_j}\right)_{\theta=\hat{\theta}} e = 0 \ .$$

---

3) Assumption 4 can be relaxed. See Magnus (1977a)

proof

The probability density of y takes the form

$$(2\pi)^{-n/2}|\Omega|^{-\frac{1}{2}} \exp -\tfrac{1}{2}\varepsilon'\Omega^{-1}\varepsilon \ .$$

Let $V=\Omega^{-1}$, then the loglikelihood is

$$\Lambda = \gamma + \tfrac{1}{2}\log|V| - \tfrac{1}{2}\varepsilon' V\varepsilon, \qquad (11)$$

where $\gamma = -\frac{n}{2}\log 2\pi$ is a constant.
Differentiating Λ we have[4)]

$$d\Lambda = \tfrac{1}{2}\mathrm{tr}V^{-1}dV - \varepsilon' V(d\varepsilon) - \tfrac{1}{2}\varepsilon'(dV)\varepsilon$$

$$= \varepsilon' VX(d\beta) + \tfrac{1}{2}\mathrm{tr}(V^{-1}-\varepsilon\varepsilon')(dV). \qquad (12)$$

Necessary for a maximum is that $d\Lambda=0$ for all $d\beta \neq 0$ and $d\theta \neq 0$. Thus:

(i) $\varepsilon' VX = 0$

(ii) $\mathrm{tr}(V^{-1}-\varepsilon\varepsilon')\frac{\partial V}{\partial\theta_h} = 0 \qquad (h=1...m)$,

which proves the first part of the theorem.
Now suppose that $|V|$ does not depend upon θ_j, then

$0 = \frac{\partial\log|V|}{\partial\theta_j} = \mathrm{tr}V^{-1}\frac{\partial V}{\partial\theta_j}$, which proves the second part. □

It will be convenient to write (12) explicitly as a function of $d\beta$ and $d\theta$.

$$\mathrm{tr}(V^{-1}-\varepsilon\varepsilon')(dV) = [\mathrm{vec}(V^{-1}-\varepsilon\varepsilon')]'\mathrm{vec}dV$$

$$= [\mathrm{vec}(V^{-1}-\varepsilon\varepsilon')]'\left(\frac{\partial\mathrm{vec}V}{\partial\theta}\right)'d\theta.$$

It follows that

$$d\Lambda = (d\beta)'X'V\varepsilon + \tfrac{1}{2}(d\theta)'\left(\frac{\partial\mathrm{vec}V}{\partial\theta}\right)\mathrm{vec}(V^{-1}-\varepsilon\varepsilon') \ . \qquad (13)$$

Remark

We shall refer to equation (10) as the θ-equation(s).

---

4) In what follows we shall use the definition of a matrix derivative as in Neudecker (1969). For instance the expression $\partial\mathrm{vec}V/\partial\theta$ describes an (m,n^2) matrix.

4. Derivation of the Hessian matrix and the information matrix

We recall from (11) and (13) that

$$\Lambda = \gamma + \tfrac{1}{2}\log|V| - \tfrac{1}{2}\,\varepsilon' V \varepsilon \quad \text{, and}$$

$$d\Lambda = (d\beta)'X'V\varepsilon + \tfrac{1}{2}\,(d\theta)'\left(\frac{\partial \mathrm{vec} V}{\partial \theta}\right)\ \mathrm{vec}(V^{-1}-\varepsilon\varepsilon').$$

Now, $d^2\Lambda = \left[(d\beta)',(d\theta)'\right]\ H \begin{bmatrix} d\beta \\ d\theta \end{bmatrix}$.

The structure of H is given by the following theorem.

Theorem 2

Define the symmetric (m,m) matrices

$$M^{ij} = \frac{\partial \Omega^{-1}_{ij}}{\partial\theta\partial\theta'} \qquad (i,j=1\ldots n) \tag{14}$$

and let $\Omega = [\omega_{ij}]$, then the Hessian matrix of the loglikelihood function (11) is

$$H = \begin{bmatrix} H_{11} & H_{12}' \\ H_{12} & H_{22} \end{bmatrix} \quad \text{, with}$$

$$H_{11} = -X'\Omega^{-1}X$$

$$H_{12} = \left(\frac{\partial \mathrm{vec}\Omega^{-1}}{\partial\theta}\right) \quad (X \otimes \varepsilon)$$

$$H_{22} = \tfrac{1}{2}\sum_{i,j}(\omega_{ij}-\varepsilon_i\varepsilon_j)M^{ij} - \tfrac{1}{2}\left(\frac{\partial \mathrm{vec}\Omega^{-1}}{\partial\theta}\right)\ (\Omega \otimes \Omega)\ \left(\frac{\partial \mathrm{vec}\Omega^{-1}}{\partial\theta}\right)'.$$

proof

Starting from (13) we have

$$d^2\Lambda = (d\beta)'X'V(d\varepsilon) + (d\beta)'X'(dV)\varepsilon + \tfrac{1}{2}(d\theta)'d\left(\frac{\partial \mathrm{vec}V}{\partial\theta}\right)\mathrm{vec}(V^{-1}-\varepsilon\varepsilon')$$

$$+ \tfrac{1}{2}(d\theta)'\left(\frac{\partial \mathrm{vec}V}{\partial\theta}\right)d\ \mathrm{vec}(V^{-1}-\varepsilon\varepsilon')$$

$$= -(d\beta)'X'VX(d\beta) + (d\beta)'X'(dV)\varepsilon + \tfrac{1}{2}(d\theta)'d\left(\frac{\partial \mathrm{vec}V}{\partial\theta}\right)\mathrm{vec}(V^{-1}-\varepsilon\varepsilon')$$

$$+ \tfrac{1}{2}(d\theta)'\left(\frac{\partial \mathrm{vec}V}{\partial\theta}\right)d\ \mathrm{vec}V^{-1} - \tfrac{1}{2}(d\theta)'\left(\frac{\partial \mathrm{vec}V}{\partial\theta}\right)d\ \mathrm{vec}(\varepsilon\varepsilon'). \tag{15}$$

The second term in (15) can be written as follows:

$$(d\beta)'X'(dV)\varepsilon = (d\beta)'\mathrm{vec}\left[X'(dV)\varepsilon\right] = (d\beta)'(\varepsilon' \otimes X')\mathrm{vec}\ dV$$

$$= (d\beta)'(\varepsilon' \otimes X')\ \left(\frac{\partial \mathrm{vec}V}{\partial\theta}\right)'d\theta.$$

From the definition of M^{ij} in (14) it follows that

$$d\left(\frac{\partial \mathrm{vec} V}{\partial \theta}\right) = \left[M^{11}d\theta, M^{12}d\theta, \ldots, M^{nn}d\theta\right] \text{ , so that}$$

$$d\left(\frac{\partial \mathrm{vec} V}{\partial \theta}\right) \mathrm{vec}(V^{-1}-\varepsilon\varepsilon') = \left[M^{11}d\theta, \ldots, M^{nn}d\theta\right]\mathrm{vec}(\Omega-\varepsilon\varepsilon')$$

$$= \left[\sum_{i,j} (\omega_{ij}-\varepsilon_i\varepsilon_j)M^{ij}\right] (d\theta) .$$

Further, since $dV^{-1} = -V^{-1}(dV)V^{-1}$, it follows that

$$d \text{ vec } V^{-1} = -(V^{-1} \otimes V^{-1})\mathrm{vec}\ dV = -(V^{-1} \otimes V^{-1}) \left(\frac{\partial \mathrm{vec} V}{\partial \theta}\right)' d\theta .$$

Also, $\quad d\mathrm{vec}\ \varepsilon\varepsilon' = \mathrm{vec}(d\varepsilon)\varepsilon' + \mathrm{vec}\,\varepsilon(d\varepsilon)'$

$$= -\mathrm{vec}X(d\beta)\varepsilon' - \mathrm{vec}\,\varepsilon(d\beta)'X'$$

$$= -(\varepsilon \otimes X)\ \mathrm{vec}(d\beta) - (X \otimes \varepsilon)\ \mathrm{vec}(d\beta)' = -\ (\varepsilon \otimes X + X \otimes \varepsilon)(d\beta) .$$

Collecting terms and inserting into (15) we find

$$d^2\Lambda = -(d\beta)'X'VX(d\beta) + (d\beta)'(\varepsilon' \otimes X')\left(\frac{\partial \mathrm{vec} V}{\partial \theta}\right)' d\theta$$

$$+ \tfrac{1}{2}(d\theta)'\left[\sum_{i,j} (\omega_{ij}-\varepsilon_i\varepsilon_j)M^{ij}\right](d\theta) - \tfrac{1}{2}(d\theta)'\left(\frac{\partial \mathrm{vec} V}{\partial \theta}\right)(V^{-1} \otimes V^{-1})\left(\frac{\partial \mathrm{vec} V}{\partial \theta}\right)'(d\theta)$$

$$+ \tfrac{1}{2}(d\theta)'\left(\frac{\partial \mathrm{vec} V}{\partial \theta}\right)(\varepsilon \otimes X + X \otimes \varepsilon)(d\beta).$$

We finally observe that, since dV is symmetric, it follows from (7) that

$$\varepsilon'(dV)X = (\mathrm{vec}\ dV)'(X \otimes \varepsilon) = (\mathrm{vec}\ dV)'\ (\varepsilon \otimes X).$$

This implies that

$$(d\theta)'\left(\frac{\partial \text{vec}V}{\partial \theta}\right)(X \otimes \varepsilon) = (d\theta)'\left(\frac{\partial \text{vec}V}{\partial \theta}\right)(\varepsilon \otimes X),$$

so that

$$d^2\Lambda = -(d\beta)'X'VX(d\beta) + 2(d\theta)'\left(\frac{\partial \text{vec}V}{\partial \theta}\right)(X \otimes \varepsilon)(d\beta)$$

$$+ \tfrac{1}{2}(d\theta)'\left[\sum_{i,j}(\omega_{ij}-\varepsilon_i\varepsilon_j)M^{ij} - \left(\frac{\partial \text{vec}V}{\partial \theta}\right)(V^{-1} \otimes V^{-1})\left(\frac{\partial \text{vec}V}{\partial \theta}\right)'\right](d\theta). \quad (16)$$

This concludes the proof.

Of particular interest is the information matrix Ψ, defined as minus the expectation of the Hessian matrix.[5)]

<u>Theorem 3</u>

The information matrix of the loglikelihood function (11) is

$$\Psi = \begin{bmatrix} X'\Omega^{-1}X & 0 \\ 0 & \tfrac{1}{2}\Psi_\theta \end{bmatrix}, \quad (17)$$

where Ψ_θ is a symmetric (m,m) matrix with typical element

$$(\Psi_\theta)_{ij} = \text{tr}\left(\frac{\partial \Omega^{-1}}{\partial \theta_i}\right)\Omega\left(\frac{\partial \Omega^{-1}}{\partial \theta_j}\right)\Omega \qquad (i,j=1\ldots m). \quad (18)$$

<u>proof</u>

Since $E\varepsilon=0$ and $E\varepsilon_i\varepsilon_j=\omega_{ij}$, it follows that

$$\Psi = -\begin{bmatrix} EH_{11} & EH_{12}' \\ EH_{12} & EH_{22} \end{bmatrix} = \begin{bmatrix} X'\Omega^{-1}X & 0 \\ 0 & \tfrac{1}{2}\left(\frac{\partial \text{vec}V}{\partial \theta}\right)(V^{-1} \otimes V^{-1})\left(\frac{\partial \text{vec}V}{\partial \theta}\right)' \end{bmatrix}.$$

Let $\Psi_\theta = \left(\frac{\partial \text{vec}V}{\partial \theta}\right)(V^{-1} \otimes V^{-1})\left(\frac{\partial \text{vec}V}{\partial \theta}\right)'$, then Ψ_θ is a symmetric (m,m) matrix whose ij-th element is

$$\left(\text{vec}\,\frac{\partial V}{\partial \theta_i}\right)'(V^{-1} \otimes V^{-1})\left(\text{vec}\,\frac{\partial V}{\partial \theta_j}\right) = \text{tr}\left(\frac{\partial V}{\partial \theta_i}\right)V^{-1}\left(\frac{\partial V}{\partial \theta_j}\right)V^{-1},$$

according to (6) □

---

5) Sometimes the information matrix is defined as $E\left(\frac{\partial \Lambda}{\partial \zeta}\right)\left(\frac{\partial \Lambda}{\partial \zeta}\right)'$, where $\zeta' = (\beta' \vdots \theta')$. We shall see in lemma (5) that this leads to the same expressions.

Now we have to ensure that Ψ_θ is a nonsingular matrix. We therefore need the following assumption:

ASSUMPTION 5: The m vectors $\text{vec}\,\frac{\partial\Omega^{-1}}{\partial\theta_1},\ldots,\text{vec}\,\frac{\partial\Omega^{-1}}{\partial\theta_m}$ are linearly independent.

lemma 1

Under the assumptions (1) - (5), the matrix Ψ as defined in (17) is positive definite.

Remark

Assumption five is also important for identification of the parameters. Suppose for example that $\Omega = (\theta_1+\theta_2)I$, then θ_1 and θ_2 are unidentified. Such a parametrization is made impossible by assumption five.

5. Finite properties of the two-step Aitken estimator and the ML estimator

In section three we defined the loglikelihood function

$$\Lambda = \gamma + \tfrac{1}{2}\log|\Omega^{-1}| - \tfrac{1}{2}\varepsilon'\Omega^{-1}\varepsilon \ , \tag{19}$$

and we found that Λ is maximized when

$$\begin{aligned} &\text{(i)} \quad \beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \\ &\text{(ii)} \quad \operatorname{tr}\left(\frac{\partial\Omega^{-1}}{\partial\theta_h}\right)\Omega = e'\left(\frac{\partial\Omega^{-1}}{\partial\theta_h}\right)e \qquad (h=1\ldots m)\ . \end{aligned} \tag{20}$$

Only in trivial cases, however, the system in (20) can be solved algebraically for the ML values of β and θ. We therefore consider the following iterative procedure[6)]

(i) Choose $\theta = \theta_0 \in \Theta$, the class of admissible values of θ.

(ii) Calculate $\Omega_0^{-1} = \Omega^{-1}(\theta_0)$,

$$b_0 = (X'\Omega_0^{-1}X)^{-1}X'\Omega_0^{-1}y \ ,$$

6) This is by no means the only numerical method to find the roots of (20) The Newton-Raphson iteration, for example, does the same job. It involves, however, inversion of the Hessian matrix at each step of the algorithm. On the other hand, it does not need a solution of the θ-equation, as the procedure in

$$e_0 = y - Xb_0 .$$

(iii) Substitute e_0 into the θ-equation. This gives m (nonlinear) equations in m unknowns (the θ's). When it is possible to write the θ-equation explicitly as $\theta = \theta(e)$, we put $\theta_1 = \theta(e_0)$. When an explicit solution of the θ-equation does not exist, we may find more than one solution. In that case we select the solution with the highest likelihood. This is θ_1.

(iv) Calculate $\Omega_1^{-1} = \Omega^{-1}(\theta_1)$,

$$b_1 = (X'\Omega_1^{-1}X)^{-1}X'\Omega_1^{-1}y ,$$

and so forth, until convergence.

Oberhofer and Kmenta (1974) prove that, under very general conditions (their assumption 6), the above procedure converges to a solution of the first-order maximizing conditions.

The uniqueness of the ML solution is contained in the following

lemma 2

Suppose that the estimators obtained for β and θ are consistent at each step of the above iterative procedure. Then we have formed, upon convergence, a consistent root of the ML equations. This root is the unique ML estimator.

proof

See e.g. Cramèr (1946) or Dhrymes (1970, Chapter three)

The consistency of the estimators of β and θ is studied in the next section.

Definition 1

The two-step Aitken estimator $b_1(\theta_0)$ is the estimator b_1 defined by the above algorithm, based on the initial value θ_0.

Definition 2

The pure Aitken estimator b^* is $(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$, where Ω (or $\sigma^2\Omega$) is the true covariance matrix of the disturbances.

Lemma 3

The two-step Aitken estimator $b_1(\theta_0)$ is distributed symmetrically around β; it is unbiased if its mean exists.

proof

Since ε is symmetrically distributed, it follows from a line of thought applied by Kakwani (1967) that it is sufficient to show that Ω_1 is an even function of ε. Now, according to the algorithm, θ_1 is a solution of

$$\mathrm{tr}\left(\frac{\partial\Omega^{-1}}{\partial\theta_h}\right)\Omega = e_0'\left(\frac{\partial\Omega^{-1}}{\partial\theta_h}\right)e_0 \qquad (h=1...m),$$

where $e_0 = y - Xb_0 = [I-X(X'\Omega_0^{-1}X)^{-1}X'\Omega_0^{-1}]y = [I-X(X'\Omega_0^{-1}X)^{-1}X'\Omega_0^{-1}]\varepsilon$.

If ε changes sign, e_0 will change sign, but the expressions

$$e_0'\left(\frac{\partial\Omega^{-1}}{\partial\theta_h}\right)e_0$$

will not be affected. Thus θ_1 is an even function of ε, which implies that Ω_1 is an even function of ε.

Lemma 4

In so far as iteration leads to the ML estimator $\hat{\beta}$, it is unbiased, if its mean exists.

proof

In the proof of lemma (3) it was shown that Ω_1 is an even function of ε. This implies that $e_1 = [I-X(X'\Omega_1^{-1}X)^{-1}X'\Omega_1^{-1}]\varepsilon$ also is an even function of ε. But, since Ω_2 is an even function of e_1, it follows that Ω_2 is an even function of ε. Therefore $b_2(\theta_0)$ is unbiased if its mean exists. It is now clear that iteration does not affect the unbiasedness of the estimator of β. □

The existence of expectations is investigated in Swamy and Mehta (1969), Fuller and Battese (1973) and Mehta and Swamy (1976).

Lemma 5

The multivariate normal density of ε (with parameters β and θ) is regular with respect to its first and second derivatives, i.e.

$$E\ \partial\Lambda/\partial\zeta = 0 \tag{21}$$

$$-E\ \partial^2\Lambda/\partial\zeta\partial\zeta' = E(\partial\Lambda/\partial\zeta)(\partial\Lambda/\partial\zeta)' , \tag{22}$$

where $\zeta = \begin{pmatrix}\beta \\ \theta\end{pmatrix}$

proof

From (13) we have

$$\partial\Lambda/\partial\beta = X'\Omega^{-1}\varepsilon \quad \text{and} \quad \partial\Lambda/\partial\theta = \tfrac{1}{2}\left(\frac{\partial \mathrm{vec}\Omega^{-1}}{\partial\theta}\right) \mathrm{vec}(\Omega-\varepsilon\varepsilon').$$

Since $E\varepsilon = 0$ and $E\varepsilon\varepsilon' = \Omega$, it follows that

$E\partial\Lambda/\partial\beta = 0$ and $E\partial\Lambda/\partial\theta = 0$, which proves (21).

In order to establish (22) we note that $-E\ \partial^2\Lambda/\partial\zeta\partial\zeta'$ is the information matrix Ψ. According to (17) and (18) we then have to prove the following equalities:

$$\left[\begin{array}{ll}
\text{(i)} & E(\partial\Lambda/\partial\beta)(\partial\Lambda/\partial\beta)' = X'\Omega^{-1}X \\[2ex]
\text{(ii)} & E(\partial\Lambda/\partial\beta)(\partial\Lambda/\partial\theta)' = 0 \\[2ex]
\text{(iii)} & E\left[(\partial\Lambda/\partial\theta)(\partial\Lambda/\partial\theta)'\right]_{ij} = \tfrac{1}{2}\mathrm{tr}\ \dfrac{\partial\Omega^{-1}}{\partial\theta_i}\ \Omega\ \dfrac{\partial\Omega^{-1}}{\partial\theta_j}\ \Omega\ (i,j=1...m)
\end{array}\right.$$

Now,

$$E(\partial\Lambda/\partial\beta)(\partial\Lambda/\partial\beta)' = E\ X'\Omega^{-1}\varepsilon\varepsilon'\Omega^{-1}X = X'\Omega^{-1}\Omega\Omega^{-1}X = X'\Omega^{-1}X.$$

This proves (i). Further,

$$E(\partial\Lambda/\partial\beta)(\partial\Lambda/\partial\theta)' = \tfrac{1}{2}E\ X'\Omega^{-1}\varepsilon\ \left[\mathrm{vec}(\Omega-\varepsilon\varepsilon')\right]'\left(\frac{\partial \mathrm{vec}\Omega^{-1}}{\partial\theta}\right)'$$

$$= \tfrac{1}{2}\ X'\Omega^{-1}\left[E\ \varepsilon\{\mathrm{vec}(\Omega-\varepsilon\varepsilon')\}'\right]\left(\frac{\partial \mathrm{vec}\Omega^{-1}}{\partial\theta}\right)' .$$

Consider the (n,n^2)matrix

$$\varepsilon[\mathrm{vec}(\Omega-\varepsilon\varepsilon')]' ,$$

with typical element $\varepsilon_i[\omega_{jk} - \varepsilon_j\varepsilon_k]$.
Since $E\varepsilon_i\omega_{jk} = 0$ and $E\varepsilon_i\varepsilon_j\varepsilon_k = 0$ for all i,j,k, it follows that

$$E\varepsilon[\mathrm{vec}(\Omega-\varepsilon\varepsilon')]' = 0,$$

which proves (ii).
Finally,

$$(\partial\Lambda/\partial\theta)(\partial\Lambda/\partial\theta)' = \tfrac{1}{4}\left(\frac{\partial \mathrm{vec}\Omega^{-1}}{\partial\theta}\right)\mathrm{vec}(\Omega-\varepsilon\varepsilon')[\mathrm{vec}(\Omega-\varepsilon\varepsilon')]'\left(\frac{\partial \mathrm{vec}\Omega^{-1}}{\partial\theta}\right)' .$$

$$[(\partial\Lambda/\partial\theta)(\partial\Lambda/\partial\theta)']_{ij} = \tfrac{1}{4}(\mathrm{vec}\,\frac{\partial\Omega^{-1}}{\partial\theta_i})'\,\mathrm{vec}(\Omega-\varepsilon\varepsilon')[\mathrm{vec}(\Omega-\varepsilon\varepsilon')]'\,(\mathrm{vec}\,\frac{\partial\Omega^{-1}}{\partial\theta_j})$$

$$= \tfrac{1}{4}[\mathrm{tr}\,\frac{\partial\Omega^{-1}}{\partial\theta_i}(\Omega-\varepsilon\varepsilon')][\mathrm{tr}\,\frac{\partial\Omega^{-1}}{\partial\theta_j}(\Omega-\varepsilon\varepsilon')]$$

$$= \tfrac{1}{4}[\mathrm{tr}\,\frac{\partial\Omega^{-1}}{\partial\theta_i}\Omega - \varepsilon'\frac{\partial\Omega^{-1}}{\partial\theta_i}\varepsilon][\mathrm{tr}\,\frac{\partial\Omega^{-1}}{\partial\theta_j}\Omega - \varepsilon'\frac{\partial\Omega^{-1}}{\partial\theta_j}\varepsilon].$$

Now,

$$E\varepsilon'\frac{\partial\Omega^{-1}}{\partial\theta_i}\varepsilon = E\,\mathrm{tr}\,\varepsilon'\frac{\partial\Omega^{-1}}{\partial\theta_i}\varepsilon = E\,\mathrm{tr}\,\frac{\partial\Omega^{-1}}{\partial\theta_i}\varepsilon\varepsilon'$$

$$= \mathrm{tr}\,E\,\frac{\partial\Omega^{-1}}{\partial\theta_i}\varepsilon\varepsilon' = \mathrm{tr}\,\frac{\partial\Omega^{-1}}{\partial\theta_i}\Omega ,$$

so that the above expression can be written as

$$4[(\partial\Lambda/\partial\theta)(\partial\Lambda/\partial\theta)']_{ij} = (\varepsilon'\frac{\partial\Omega^{-1}}{\partial\theta_i}\varepsilon - E\varepsilon'\frac{\partial\Omega^{-1}}{\partial\theta_i}\varepsilon)(\varepsilon'\frac{\partial\Omega^{-1}}{\partial\theta_j}\varepsilon - E\varepsilon'\frac{\partial\Omega^{-1}}{\partial\theta_j}\varepsilon).$$

Taking expectations we find

$$4E[(\partial\Lambda/\partial\theta)(\partial\Lambda/\partial\theta)']_{ij} = \mathrm{cov}\,(\varepsilon'\frac{\partial\Omega^{-1}}{\partial\theta_i}\varepsilon,\ \varepsilon'\frac{\partial\Omega^{-1}}{\partial\theta_j}\varepsilon)$$

$$= 2\,\mathrm{tr}\,\frac{\partial\Omega^{-1}}{\partial\theta_i}\Omega\frac{\partial\Omega^{-1}}{\partial\theta_j}\Omega.$$

The last equality follows from Magnus and Neudecker (1977, p.16).
This concludes the proof.

6. Asymptotic properties of the two-step Aitken estimator and the ML estimator

The asymptotic properties of estimators are almost without exception based on random sampling, that is on the statistical independence of the y_i (or ε_i). In that case the central limit theorems apply. Our problem, however, consists in estimating β from a single (vector) observation on y.
A related complication is that Ω increases in size when n increases. We shall need the following assumptions.

ASSUMPTION 6: The elements of $Z_h = \frac{\partial \Omega^{-1}}{\partial \theta_h}$ (h=1...m) are continuous functions of θ in an open sphere S of θ_0, the true value of the parameter vector θ.

ASSUMPTION 7: $\lim_{n\to\infty} \frac{1}{n} X'\Omega^{-1}X$ exists as a positive definite matrix of fixed constants for all θ in S.

ASSUMPTION 8: $\lim_{n\to\infty} \frac{1}{n} X'Z_hX$ exists as a matrix whose elements are continous functions of θ (h=1...m).

These assumptions enable us to formulate the following theorem due to Fuller and Battese (1973).

Theorem 4

Suppose there exists an estimator $\tilde{\theta}$ for θ_0 such that $\Omega^{-1}(\tilde{\theta})$ exists for all n, and $\tilde{\theta} = \theta_0 + O(n^{-\delta})$, $\delta > 0$, then the assumptions (2) - (8) imply that

$$\tilde{\beta}_n - b_n^* = O(n^{-\frac{1}{2}-\delta}), \text{ where}$$

$$\tilde{\beta}_n = (X'\Omega^{-1}(\tilde{\theta})X)^{-1}X'\Omega^{-1}(\tilde{\theta})y \quad ,$$

and b_n^* is the pure Aitken estimator based on the true value θ_0.

proof

See Fuller and Battese (1973, p. 629)

Corollary

Under the same assumptions as in theorem 4 we have

$$\text{plim } \tilde{\beta}_n = \text{plim } b^*_n \text{ , and}$$

$\tilde{\beta}_n$ has the same asymptotic distribution as b^*_n,

that is

$$\text{plim } \tilde{\beta}_n = \beta \text{ , and}$$

$\sqrt{n}(\tilde{\beta}_n - \beta)$ has asymptotic distribution $N\left[0, \lim_{n\to\infty} n(X'\Omega^{-1}X)^{-1}\right]$.

We now turn to the ML estimators $\hat{\beta}$ and $\hat{\theta}$. In the standard case of random sampling the value of the ML method lies in the fact that it generates estimators with desirable asymptotic properties. Let $\hat{\zeta}$ be such a ML estimator. Then [7], under very general conditions, $\hat{\zeta}$ is consistent, asymptotically unbiased and asymptotically efficient. Further $\sqrt{n}(\hat{\zeta}-\zeta)$ has asymptotic distribution $N(0, \lim_{n\to\infty} n\Psi^{-1})$, where Ψ is the information matrix.

To deal with the more difficult non-standard case we need the following assumptions:

ASSUMPTION 9: Every element of $\frac{1}{n} X'\Omega^{-1}X$ converges as $n\to\infty$ to a finite function of θ, uniformly for θ in any compact set.

ASSUMPTION 10: Every diagonal element of $\frac{1}{n^2} X' \frac{\partial\Omega^{-1}}{\partial\theta_i} \Omega \frac{\partial\Omega^{-1}}{\partial\theta_i} X$ converges as $n\to\infty$ to zero, uniformly for θ in any compact set (i=1...m).

ASSUMPTION 11: $\frac{1}{n} \text{tr} \frac{\partial\Omega^{-1}}{\partial\theta_i} \Omega \frac{\partial\Omega^{-1}}{\partial\theta_j} \Omega$ converges as $n\to\infty$ to a finite function of θ, uniformly for θ in any compact set (i,j=1...m).

ASSUMPTION 12: $\frac{1}{n^2} \text{tr} \left(\frac{\partial^2\Omega^{-1}}{\partial\theta_i\partial\theta_j} \Omega\right)^2$ converges as $n\to\infty$ to zero, uniformly for θ in any compact set (i,j=1...m).

---

[7] See Kendall and Stuart (1967), Chapter 18.

Theorem 5 [8)]

The ML estimates $\hat{\beta}$ and $\hat{\theta}$ from the regression model (8) under the assumptions (1) - (5) and (9) - (12) are weakly consistent, asymptotically normally distributed, and asymptotically efficient in the maximum probability sense of Weiss and Wolfowitz (1967).

proof

It suffices to prove that the assumptions (2.1) and (2.2) of Weiss (1973) reduce to our assumptions (9) - (12). This is greatly facilitated by applying three theorems in Vickers(1977, section 1.4).

Let $H = (h_{ij})$ be the Hessian matrix from theorem 2 and $\Psi = (\psi_{ij})$ the information matrix from theorem 3. The implication of Vickers' theorems is that $\hat{\beta}$ and $\hat{\theta}$ are weakly consistent, asymptotically normally distributed and assymptotically efficient in the maximum probability sense, if

(a) $\frac{1}{n}\psi_{ij}$ converges as $n\to\infty$ to a finite function of β and θ, uniformly for values β and θ in any compact set $(i,j=1...k+m)$,

(b) $\frac{1}{n^2}\,\mathrm{var}(h_{ij})$ converges as $n\to\infty$ to zero, uniformly for values β and θ in any compact set $(i,j=1...k+m)$.

We shall now show that the assumption (a) and (b) reduce to (9)-(12). Partition

$$X = (x_1 ... x_k),$$

then

$$\frac{\partial^2\Lambda}{\partial\beta_i\partial\beta_j} = -x_i'\Omega^{-1}x_j \qquad (i,j=1...k),$$

$$\frac{\partial^2\Lambda}{\partial\theta_i\partial\beta_j} = x_j'\frac{\partial\Omega^{-1}}{\partial\theta_i}\varepsilon \qquad (i=1...m,j=1...k),$$

$$\frac{\partial^2\Lambda}{\partial\theta_i\partial\theta_j} = -\tfrac{1}{2}\left[\mathrm{tr}\,\frac{\partial\Omega^{-1}}{\partial\theta_i}\Omega\frac{\partial\Omega^{-1}}{\partial\theta_j}\Omega + \varepsilon'\frac{\partial^2\Omega^{-1}}{\partial\theta_i\partial\theta_j}\varepsilon - \mathrm{tr}\,\frac{\partial^2\Omega^{-1}}{\partial\theta_i\partial\theta_j}\Omega\right] \qquad (i,j=1...m).$$

---

8) I am grateful to professor Lionel Weiss and Dr. Kathleen Vickers for calling my attention to their work. Theorem five is a direct application of Dr. Vickers' Ph.D. thesis.

Thus,

$$E\left[-\frac{\partial^2\Lambda}{\partial\beta_i\partial\beta_j}\right] = x_i'\Omega^{-1}x_j ,$$

$$\mathrm{var}\left[-\frac{\partial^2\Lambda}{\partial\beta_i\partial\beta_j}\right] = 0 ,$$

$$E\left[-\frac{\partial^2\Lambda}{\partial\theta_i\partial\beta_j}\right] = 0 ,$$

$$\mathrm{var}\left[-\frac{\partial^2\Lambda}{\partial\theta_i\partial\beta_j}\right] = x_j' \frac{\partial\Omega^{-1}}{\partial\theta_i} \Omega \frac{\partial\Omega^{-1}}{\partial\theta_i} x_j ,$$

$$E\left[-\frac{\partial^2\Lambda}{\partial\theta_i\partial\theta_j}\right] = \tfrac{1}{2}\,\mathrm{tr}\,\frac{\partial\Omega^{-1}}{\partial\theta_i} \Omega \frac{\partial\Omega^{-1}}{\partial\theta_j} \Omega ,$$

$$\mathrm{var}\left[-\frac{\partial^2\Lambda}{\partial\theta_i\partial\theta_j}\right] = \tfrac{1}{4}\,\mathrm{var}\left[\varepsilon' \frac{\partial^2\Omega^{-1}}{\partial\theta_i\partial\theta_j} \varepsilon\right] = \tfrac{1}{2}\,\mathrm{tr}\left(\frac{\partial^2\Omega^{-1}}{\partial\theta_i\partial\theta_j}\Omega\right)^2 .$$

It is now clear that (a) reduces to the assumptions (9) and (11), and that (8) reduces to (10) and (12). This concludes the proof.

7. A general case

We shall apply the above theory to the autocorrelated errors model and to Zellner-type regressions[9], but before doing so we first study a more general case which will simplify the discussion in the next two sections.

Consider a covariance matrix of the following form

$$\Omega = \begin{bmatrix} \sigma_{11}Q_1\Gamma Q_1' & \cdots & \sigma_{1p}Q_1\Gamma Q_p' \\ \\ \sigma_{p1}Q_p\Gamma Q_1' & \cdots & \sigma_{pp}Q_p\Gamma Q_p' \end{bmatrix} = Q(\Sigma \otimes \Gamma)Q' , \qquad (23)$$

9) Applications to the heteroskedastic model and to error component analysis are studied in Magnus (1977 a,b).

where Σ and Γ are symmetric positive definite matrices of order p and T respectively, $p \leq T$[10], p fixed. The number of observations is n=pT. Further,

$$Q = \begin{bmatrix} Q_1 & & & 0 \\ & Q_2 & & \\ & & \ddots & \\ 0 & & & Q_p \end{bmatrix} ,$$

where the Q_i (i=1...p) are nonsingular matrices of order T.
The covariance matrix (23) is clearly a generalization of Zellner's case of seemingly unrelated regressions. It is also an extension of Ω itself, as can be seen by putting p=1. The matrix then reduces to

$$\Omega = \sigma^2 Q\Gamma Q' ,$$

where Γ is positive definite and Q is nonsingular.
In the next section, where we study the autocorrelated errors model, we shall work with $\Omega = QQ'$, which simplifies matters greatly.
We suppose that Σ is completely unknown, thus containing $\frac{1}{2}p(p+1)$ parameters, $Q=Q(\zeta)$, and $\Gamma=\Gamma(\xi)$, where ζ and ξ are parameter vectors containing q and r components respectively.
Thus θ consists of the elements of ζ and ξ and of the $\frac{1}{2}p(p+1)$ distinct elements σ_{hk} of Σ.

$$\theta' = [\zeta', \xi', \tilde{\sigma}'] \quad , \text{ where}$$

$$\zeta' = [\zeta_1 \dots \zeta_q] ,$$

$$\xi' = [\xi_1 \dots \xi_r] ,$$

$$\tilde{\sigma}' = [\sigma_{11}, \sigma_{12}, \dots, \sigma_{1p}, \sigma_{22}, \dots, \sigma_{2p}, \dots, \sigma_{pp}] .$$

---

10) This is necessary to ensure the nonsingularity of the estimator of Σ.

To derive the θ-conditions we proceed as follows:

$$\Omega^{-1} = (Q^{-1})'(\Sigma^{-1} \otimes \Gamma^{-1})Q^{-1} .$$

$$d\Omega^{-1} = (dQ^{-1})'(\Sigma^{-1} \otimes \Gamma^{-1})Q^{-1} + (Q^{-1})'(\Sigma^{-1} \otimes \Gamma^{-1})(dQ^{-1})$$

$$+ (Q^{-1})' \left[(d\Sigma^{-1}) \otimes \Gamma^{-1}\right]Q^{-1} + (Q^{-1})'\left[\Sigma^{-1} \otimes (d\Gamma^{-1})\right] Q^{-1} .$$

From this it follows that

$$\left.\begin{aligned}
&\text{(i)} \quad \frac{\partial\Omega^{-1}}{\partial\zeta_i} = \left(\frac{\partial Q^{-1}}{\partial\zeta_i}\right)'(\Sigma^{-1} \otimes \Gamma^{-1})Q^{-1} + (Q^{-1})'(\Sigma^{-1} \otimes \Gamma^{-1})\frac{\partial Q^{-1}}{\partial\zeta_i} \quad (i=1\ldots q)\\
&\text{(ii)} \quad \frac{\partial\Omega^{-1}}{\partial\xi_j} = (Q^{-1})'\left(\Sigma^{-1} \otimes \frac{\partial\Gamma^{-1}}{\partial\xi_j}\right) Q^{-1} \quad (j=1\ldots r)\\
&\text{(iii)} \quad \frac{\partial\Omega^{-1}}{\partial\sigma_{hk}} = -(Q^{-1})'\left[(\Sigma^{-1}Y^{hk}\Sigma^{-1}) \otimes \Gamma^{-1}\right] Q^{-1} \quad (1 \leq h \leq k \leq p),
\end{aligned}\right\}$$

where Y^{hk} is a square matrix of order p with zeros everywhere except in the hk-th and kh-th position where it has unity (24)

and (iii) follows from the fact that

$$d\Sigma^{-1} = -\Sigma^{-1}(d\Sigma)\Sigma^{-1} ,$$

so that

$$\frac{\partial\Sigma^{-1}}{\partial\sigma_{hk}} = -\Sigma^{-1}Y^{hk}\Sigma^{-1}.$$

Now define the following matrices:

$$\left.\begin{aligned}
&\text{(i)} \quad G_\zeta^i = \frac{\partial\Omega^{-1}}{\partial\zeta_i}\Omega = \left(\frac{\partial Q^{-1}}{\partial\zeta_i}\right)'Q' + (Q^{-1})'(\Sigma^{-1} \otimes \Gamma^{-1})\frac{\partial Q^{-1}}{\partial\zeta_i}Q(\Sigma \otimes \Gamma)Q' \quad (i=1\ldots q)\\
&\text{(ii)} \quad G_\xi^j = \frac{\partial\Omega^{-1}}{\partial\xi_j}\Omega = (Q^{-1})'\left(I \otimes \frac{\partial\Gamma^{-1}}{\partial\xi_j}\Gamma\right)Q' \quad (j=1\ldots r)\\
&\text{(iii)} \quad G_\sigma^{hk} = \frac{\partial\Omega^{-1}}{\partial\sigma_{hk}}\Omega = -(Q^{-1})'(\Sigma^{-1}Y^{hk} \otimes I)Q' \quad (1 \leq h \leq k \leq p) .
\end{aligned}\right\}$$

The traces of these matrices are:

$$\left.\begin{aligned}
&\text{(i)} \quad \operatorname{tr} G_\zeta^i = 2 \operatorname{tr} \frac{\partial Q^{-1}}{\partial\zeta_i}Q \quad (i=1\ldots q)\\
&\text{(ii)} \quad \operatorname{tr} G_\xi^j = p \operatorname{tr} \frac{\partial\Gamma^{-1}}{\partial\xi_j}\Gamma \quad (j=1\ldots r)\\
&\text{(iii)} \quad \operatorname{tr} G_\sigma^{hk} = -T \operatorname{tr}(\Sigma^{-1}Y^{hk}) \quad (1 \leq h \leq k \leq p).
\end{aligned}\right\}$$

Let $e' = [e_1', e_2', \ldots, e_p']$,

$$z_h = Q_h^{-1} e_h \quad , \quad \tilde{z}_{hi} = \frac{\partial Q_h^{-1}}{\partial \zeta_i} e_h \quad (h=1\ldots p; i=1\ldots q),$$

$$Z = [z_1, z_2, \ldots, z_p] \quad , \quad \tilde{Z}_i = [\tilde{z}_{1i}, \tilde{z}_{2i}, \ldots, \tilde{z}_{pi}] \quad (i=1\ldots q).$$

Then $Q^{-1}e = \text{vec } Z$ and $\frac{\partial Q^{-1}}{\partial \zeta_i} e = \text{vec } \tilde{Z}_i$,

so that

$$\left.\begin{aligned}
&\text{(i)} \quad e'\frac{\partial \Omega^{-1}}{\partial \zeta_i} e = 2e'\left(\frac{\partial Q^{-1}}{\partial \zeta_i}\right)'(\Sigma^{-1} \otimes \Gamma^{-1})Q^{-1}e = 2(\text{vec } \tilde{Z}_i)'(\Sigma^{-1} \otimes \Gamma^{-1})\text{vec } Z \\
&\qquad = 2\text{tr}\Gamma^{-1}Z\Sigma^{-1}\tilde{Z}_i' \qquad (i=1\ldots q) \\
&\text{(ii)} \quad e'\frac{\partial \Omega^{-1}}{\partial \xi_j} e = (\text{vec } Z)'(\Sigma^{-1} \otimes \frac{\partial \Gamma^{-1}}{\partial \xi_j})\text{vec } Z = \text{tr}\frac{\partial \Gamma^{-1}}{\partial \xi_j} Z\Sigma^{-1}Z' \qquad (j=1\ldots r) \\
&\text{(iii)} \quad e'\frac{\partial \Omega^{-1}}{\partial \sigma_{hk}} e = -(\text{vec } Z)'\left[(\Sigma^{-1}Y^{hk}\Sigma^{-1}) \otimes \Gamma^{-1}\right] \text{vec } Z = -\text{tr } \Sigma^{-1}Z'\Gamma^{-1}Z\Sigma^{-1}Y^{hk} \\
&\qquad (1 \leq h \leq k \leq p).
\end{aligned}\right\}$$

The θ-conditions (10) are in the present case:

$$\left.\begin{aligned}
&\text{(i)} \quad \text{tr}\frac{\partial Q^{-1}}{\partial \zeta_i} Q = \text{tr}\Gamma^{-1}Z\Sigma^{-1}\tilde{Z}_i' \qquad (i=1\ldots q) \\
&\text{(ii)} \quad p\ \text{tr}\frac{\partial \Gamma^{-1}}{\partial \xi_j} \Gamma = \text{tr}\frac{\partial \Gamma^{-1}}{\partial \xi_j} Z\Sigma^{-1}Z' \qquad (j=1\ldots r) \\
&\text{(iii)} \quad \Sigma = \frac{1}{T} Z'\Gamma^{-1}Z ,
\end{aligned}\right\} \qquad (25)$$

where (i) and (ii) are obvious and (iii) follows from

$$-T\ \text{tr}\Sigma^{-1}Y^{hk} = -\text{tr}\Sigma^{-1}Z'\Gamma^{-1}Z\Sigma^{-1}Y^{hk} \qquad (1 \leq h \leq k \leq p)$$

or

$$\text{tr}(T\Sigma^{-1} - \Sigma^{-1}Z'\Gamma^{-1}Z\Sigma^{-1})Y^{hk} = 0 \qquad (1 \leq h \leq k \leq p).$$

This is equivalent with

$$T\Sigma^{-1} = \Sigma^{-1}Z'\Gamma^{-1}Z\Sigma^{-1} ,$$

which in turn is equivalent with (iii).

Now define $K_i = \frac{\partial Q^{-1}}{\partial \zeta_i} Q$ $(i=1\ldots q)$ and $C_j = \frac{\partial \Gamma^{-1}}{\partial \xi_j} \Gamma$ $(j=1\ldots r)$, and let

$$g(h,k) = k + (h-1)(p-\tfrac{1}{2}h) \qquad (1 \leq h \leq k \leq p),$$

then

$$\theta_{g(h,k)+q+r} = \sigma_{hk},$$

and the symmetric matrix Ψ_θ from the information matrix (17) takes the form

$$\Psi_\theta = \begin{matrix} q \\ r \\ \tfrac{1}{2}p(p+1) \end{matrix} \begin{bmatrix} \Psi_{\zeta\zeta} & \Psi_{\zeta\xi} & \Psi_{\zeta\sigma} \\ . & \Psi_{\xi\xi} & \Psi_{\xi\sigma} \\ . & . & \Psi_{\sigma\sigma} \end{bmatrix}$$

where [11])

(26)

(i) $(\Psi_{\zeta\zeta})_{ij} = \mathrm{tr} G_\zeta^i G_\zeta^j = 2\mathrm{tr} K_i K_j + 2\mathrm{tr} K_i'(\Sigma^{-1} \otimes \Gamma^{-1}) K_j (\Sigma \otimes \Gamma) \qquad (i,j=1\ldots q)$

(ii) $(\Psi_{\zeta\xi})_{ij} = \mathrm{tr} G_\zeta^i G_\xi^j = 2\mathrm{tr} K_i'(I \otimes C_j) \qquad (i=1\ldots q, j=1\ldots r)$

(iii) $(\Psi_{\xi\xi})_{ij} = \mathrm{tr} G_\xi^i G_\xi^j = p\,\mathrm{tr} C_i C_j \qquad (i,j=1\ldots r)$

(iv) $(\Psi_{\zeta\sigma})_{i,g(h,k)} = \mathrm{tr} G_\zeta^i G_\sigma^{hk} = -2\mathrm{tr} K_i'(\Sigma^{-1} Y^{hk} \otimes I) \quad (i=1\ldots q; g(h,k)=1\ldots\tfrac{1}{2}p(p+1))$

(v) $(\Psi_{\xi\sigma})_{j,g(h,k)} = \mathrm{tr} G_\xi^j G_\sigma^{hk} = -(\mathrm{tr}\Sigma^{-1} Y^{hk}) \mathrm{tr} C_j = \begin{cases} -2\sigma^{hk} \mathrm{tr} C_j & \text{if } h \neq k \\ -\sigma^{hh} \mathrm{tr} C_j & \text{if } h = k \end{cases}$

$(j=1\ldots r; g(h,k)=1\ldots\tfrac{1}{2}p(p+1))$

(vi) $(\Psi_{\sigma\sigma})_{g(i,j),g(h,k)} = \mathrm{tr} G_\sigma^{ij} G_\sigma^{hk} = T\mathrm{tr}(\Sigma^{-1} Y^{ij})(\Sigma^{-1} Y^{hk})$

$$= \begin{cases} 2T(\sigma^{ih}\sigma^{jk} + \sigma^{ik}\sigma^{jh}) & \text{if } i \neq j \text{ and } h \neq k \\ 2T\,\sigma^{ih}\sigma^{jh} & \text{if } i \neq j \text{ and } h = k \\ 2T\,\sigma^{ih}\sigma^{ik} & \text{if } i = j \text{ and } h \neq k \\ T(\sigma^{ih})^2 & \text{if } i = j \text{ and } h = k \end{cases}$$

$(g(i,j), g(h,k) = 1\ldots\tfrac{1}{2}p(p+1))$

---

11) σ^{hk} denotes the typical element of Σ^{-1}.

Finally, we give the expressions for the case p=1:

$$\Omega = \sigma^2 Q\Gamma Q' \quad , \tag{27}$$

where $\Gamma=\Gamma(\xi)$ is positive definite and $Q=Q(\zeta)$ is nonsingular, both of order n.

$$\theta' = [\zeta', \xi', \sigma^2]$$

The θ-conditions are

$$\left[\begin{array}{lll} \text{(i)} & \sigma^2 \mathrm{tr}K_i = e'\left(\dfrac{\partial Q^{-1}}{\partial \zeta_i}\right)' \Gamma^{-1} Q^{-1} e & (i=1\ldots q) \\ \text{(ii)} & \sigma^2 \mathrm{tr}C_j = e'(Q^{-1})' \dfrac{\partial \Gamma^{-1}}{\partial \xi_j} Q^{-1} e & (j=1\ldots r) \\ \text{(iii)} & \sigma^2 = \dfrac{1}{n} e'(Q^{-1})' Q^{-1} e \ , & \end{array}\right. \tag{28}$$

and

$$\Psi_\theta' = \begin{matrix} q \\ r \\ 1 \end{matrix} \begin{bmatrix} \Psi_{\zeta\zeta} & \Psi_{\zeta\xi} & \Psi_{\zeta\sigma} \\ . & \Psi_{\xi\xi} & \Psi_{\xi\sigma} \\ . & . & \Psi_{\sigma\sigma} \end{bmatrix} \quad , \tag{29}$$

where

$$\left[\begin{array}{lll} (\Psi_{\zeta\zeta})_{ij} = 2\mathrm{tr}K_i K_j + 2\mathrm{tr}K_i' \Gamma^{-1} K_j \Gamma & & (i,j=1\ldots q) \\ (\Psi_{\zeta\xi})_{ij} = 2\mathrm{tr}K_i' C_j & & (i=1\ldots q; j=1\ldots r) \\ (\Psi_{\xi\xi})_{ij} = \mathrm{tr}C_i C_j & & (i,j=1\ldots r) \\ (\Psi_{\zeta\sigma})_i = 2\sigma^{-2} \mathrm{tr}K_i & & (i=1\ldots q) \\ (\Psi_{\xi\sigma})_j = \sigma^{-2} \mathrm{tr}C_j & & (j=1\ldots r) \\ \Psi_{\sigma\sigma} = n\sigma^{-4} . & & \end{array}\right. \tag{30}$$

8. The autocorrelated errors model

Let $y = X\beta + \varepsilon$, $\varepsilon_t = \rho\varepsilon_{t-1} + \zeta_t$, $E\zeta = 0$, $E\zeta\zeta' = \sigma^2 I_n$, $|\rho| < 1$. (31)

These conditions donot specify the standard first-order autocorrelated errors model completely; one more assumption is needed as to the initial value of the disturbances.
For the moment we shall only assume that

$$\varepsilon_1 - \rho\varepsilon_0 = \phi\varepsilon_1 ,$$

where ϕ may depend on ρ; $\phi > 0$.

This implies that $\zeta_1 = \phi\varepsilon_1$, so that we can write

$$\zeta = A\varepsilon \text{ , where } A = \begin{bmatrix} \phi & & & \\ -\rho & 1 & & \\ & \ddots & \ddots & \\ & & -\rho & 1 \end{bmatrix} .$$

Now $\sigma^2 I = E\zeta\zeta' = EA\varepsilon\varepsilon' A' = A(E\varepsilon\varepsilon')A'$, so

$$\Omega = E\varepsilon\varepsilon' = \sigma^2 A^{-1}(A^{-1})' .$$

According to (28) the θ-conditions are

$$\left[\begin{array}{ll} \text{(i)} & \sigma^2 \mathrm{tr} K = e'\left(\frac{\partial A}{\partial \rho}\right)' Ae \\ \text{(ii)} & \sigma^2 = \frac{1}{n} e'A'Ae \end{array}\right. , \qquad (32)$$

where $K = \frac{\partial A}{\partial \rho} A^{-1}$. Let $\phi' = \partial\phi/\partial\rho$,

then $$\frac{\partial A}{\partial \rho} = \begin{bmatrix} \phi' & & & \\ -1 & 0 & & \\ & \ddots & \ddots & \\ & & -1 & 0 \end{bmatrix} .$$

Since

$$A^{-1} = \left[\begin{array}{l|llll} \phi^{-1} & 0 & \cdots & & 0 \\ \hline \phi^{-1}\rho & 1 & & & 0 \\ \vdots & \rho & \ddots & & \\ \phi^{-1}\rho^{n-1} & \rho^{n-2} & \cdots & \rho & 1 \end{array}\right]$$

We find that

$$K = -\left[\begin{array}{c|cccc} -\phi'\phi^{-1} & 0 & \cdots & & 0 \\ \hline \phi^{-1} & 0 & & & 0 \\ \phi^{-1}\rho & 1 & \ddots & & \\ & \rho & \ddots & \ddots & \\ \vdots & \vdots & \ddots & \ddots & \\ \phi^{-1}\rho^{n-2} & \rho^{n-3} & \cdots\ \rho & 1 & 0 \end{array}\right] .$$

Clearly, $\mathrm{tr}K = \phi'\phi^{-1}$.

Further $e'A'Ae = \sum_{i=2}^{n} (e_i - \rho e_{i-1})^2 + \phi^2 e_1^2 = \sum_{i=1}^{n} (e_i - \rho e_{i-1})^2$,

where $e_0 = \dfrac{1-\phi}{\rho} e_1$.

$$e'\left(\frac{\partial A}{\partial \rho}\right)' Ae = \phi\phi' e_1^2 + \rho \sum_1^{n-1} e_i^2 - \sum_1^{n-1} e_i e_{i+1} .$$

The θ-conditions (32) boil down to

$$\begin{cases} \text{(i)} & \hat{\sigma}^2 \phi'\phi^{-1} = \phi\phi' e_1^2 + \hat{\rho} \sum_1^{n-1} e_i^2 - \sum_1^{n-1} e_i e_{i+1} \\ \text{(ii)} & \hat{\sigma}^2 = \frac{1}{n} \sum_1^{n} (e_i - \hat{\rho} e_{i-1})^2 . \end{cases}$$

In order to compute the information matrix we need

$$\mathrm{tr}K^2 = (\phi'\phi^{-1})^2 ,$$

$\mathrm{tr}K'K$ = the sum of the squared elements of K

$$= (\phi'\phi^{-1})^2 + \phi^{-2} \sum_{i=0}^{n-2} \rho^{2i} + \sum_{k=0}^{n-3} \sum_{i=0}^{k} \rho^{2i}$$

$$= (\phi'\phi^{-1})^2 + \phi^{-2} \frac{1-\rho^{2(n-1)}}{1-\rho^2} + \sum_{k=0}^{n-3} \frac{1-\rho^{2(k+1)}}{1-\rho^2}$$

$$= (\phi'\phi^{-1})^2 + \frac{1}{1-\rho^2}\left[\phi^{-2}(1-\rho^{2(n-1)}) + (n-2) - \frac{\rho^2}{1-\rho^2}(1-\rho^{2(n-2)})\right]$$

$$= (\phi'\phi^{-1})^2 + \frac{1}{1-\rho^2}\left[n-(1-\phi^{-2})(1-\rho^{2(n-1)}) - \frac{1}{1-\rho^2}(1-\rho^{2n})\right].$$

According to (29) and (30) the matrix Ψ_θ is

$$\Psi_\theta = \begin{bmatrix} \psi_{\rho\rho} & \psi_{\rho\sigma} \\ . & \psi_{\sigma\sigma} \end{bmatrix} , \qquad (33)$$

where

$$\begin{cases} \psi_{\rho\rho} = 2\mathrm{tr}K^2 + 2\mathrm{tr}K'K = 4(\phi'\phi^{-1})^2 + \frac{2}{1-\rho^2}\left[n-(1-\phi^{-2})(1-\rho^{2(n-1)}) - \frac{1}{1-\rho^2}(1-\rho^{2n})\right] \\ \psi_{\rho\sigma} = 2\sigma^{-2}\mathrm{tr}K = 2\sigma^{-2}\phi'\phi^{-1} \\ \psi_{\sigma\sigma} = n\sigma^{-4} . \end{cases}$$

Two cases are of particular interest:

Case (i): iterative Cochrane-Orcutt

When $\phi=1$, the ML conditions are simply

$$\left.\begin{aligned} \hat{\rho} &= \frac{\sum_1^{n-1} e_i e_{i+1}}{\sum_1^{n-1} e_i^2} \\ \hat{\beta} &= (X'\hat{A}'\hat{A}X)^{-1}X'\hat{A}'\hat{A}y \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_1^n (e_i - \hat{\rho}e_{i-1})^2 , \quad e_0 = 0 . \end{aligned}\right\} \qquad (34)$$

Application of the algorithm of section five gives the unique ML estimators of β, ρ and σ^2.

The information matrix reduces to

$$\Psi = \begin{pmatrix} \frac{1}{\sigma^2} X'A'AX & 0 & 0 \\ 0 & \frac{1}{1-\rho^2}\left(n - \frac{1-\rho^{2n}}{1-\rho^2}\right) & 0 \\ 0 & 0 & \frac{n}{2\sigma^4} \end{pmatrix} . \tag{35}$$

Case (ii)

Kadiyala (1968) suggested $\phi = \sqrt{1-\rho^2}$, which thereafter appeared in the textbooks (e.g. Theil (1971), p.253).
The condition for ρ is then

$$\rho \sum_2^{n-1} e_i^2 = \sum_1^{n-1} e_i e_{i+1} - \frac{\rho}{1-\rho^2}\sigma^2 . \tag{36}$$

Define $a = \sigma^{-2}\sum_2^{n-1} e_i^2$, $c = \sigma^{-2}\sum_1^{n-1} e_i e_{i+1}$, and $f(\rho) = a\rho + \frac{\rho}{1-\rho^2}$,

then (36) reduces to $f(\rho) = c$.
On the interval $(-1,1)$ $f(\rho)$ is a monotonically increasing function of ρ. Moreover $\lim_{\rho \uparrow 1} f(\rho) = \infty$ and $\lim_{\rho \downarrow -1} f(\rho) = -\infty$.

Thus for every c there is one unique solution of $f(\rho) = c$ in the interval $(-1,1)$. The algorithm of section five thus leads to the unique ML estimators of β, ρ and σ^2. The information matrix is

$$\Psi = \begin{bmatrix} \frac{1}{\sigma^2} X'A'AX & 0 & 0 \\ 0 & \frac{1}{1-\rho^2}\left(n-1+\frac{2\rho^2}{1-\rho^2}\right) & \frac{-\rho}{\sigma^2(1-\rho^2)} \\ 0 & \frac{-\rho}{\sigma^2(1-\rho^2)} & \frac{n}{2\sigma^4} \end{bmatrix} . \tag{37}$$

Of course, asymptotically the two cases are equivalent.

9. Zellner-type regressions

The formulae (25) and (26) are readily applied to the following two well-known cases:

Case (i): iterated Zellner

In Zellner's (1962) case of seemingly unrelated regressions we have[12)]

$$\Omega = \Sigma \otimes I. \tag{38}$$

We therefore put $\Gamma = I$ and $Q = I$ in (25) and find

$$\hat{\Sigma} = \frac{1}{T} E'E, \tag{39}$$

where

$$E = [e_1, e_2, \ldots, e_p].$$

This shows again[13)] that continuing Zellner's estimation procedure until convergence yields the ML estimator.
The information matrix is

$$\Psi = \begin{bmatrix} X'\Omega^{-1}X & 0 \\ 0 & \frac{1}{2}\Psi_{\sigma\sigma} \end{bmatrix}, \tag{40}$$

where $\Psi_{\sigma\sigma}$ is defined in (26.vi).

Case (ii): iterated Parks

Parks (1967) investigated a system of regression equations where the disturbances are both serially and contemporaneously correlated, and he proposed a three-step estimator for β, which he proved to be consistent and asymptotically efficient. The covariance matrix in this case[14)] is

$$\Omega = Q(\Sigma \otimes I)Q', \tag{41}$$

---

12) In some applications we have $\Omega = I \otimes \Sigma$ or $\Omega = \Gamma \otimes \Sigma$. The formulae for these cases may be derived in a similar fashion.

13) See Dhrymes (1971).

14) Our model differs slightly from Parks', viz. in the specification of the initial value of the disturbances. See the discussion in the previous section.

where $Q = \begin{bmatrix} Q_1 & & 0 \\ & \ddots & \\ 0 & & Q_p \end{bmatrix}$ and $Q_i^{-1} = \begin{bmatrix} 1 & & & \\ -\rho_i & \ddots & & \\ & \ddots & \ddots & \\ & & -\rho_i & 1 \end{bmatrix}$ (i=1...p) .

Clearly, $|Q|$ does not depend upon the ρ_i, which implies that

$$\mathrm{tr}\,\frac{\partial Q^{-1}}{\partial \rho_i} Q = 0 \qquad (i=1\ldots p) .$$

Further, $\dfrac{\partial Q_i^{-1}}{\partial \rho_j} = \begin{cases} A\ , & \text{if } i=j \\ 0\ , & \text{if } i\neq j \end{cases}$,

where $A = \begin{bmatrix} 0 & & & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ & & -1 & 0 \end{bmatrix}$ (42)

It then follows from (25) that the θ-conditions are

$$\left.\begin{array}{ll} \text{(i)} & \mathrm{tr}\, Z\Sigma^{-1}\tilde{Z}_i' = 0 \qquad (i=1\ldots p) \\ \text{(ii)} & \Sigma = \frac{1}{T} Z'Z \end{array}\right\} , \qquad (43)$$

where $Z = [Q_1^{-1}e_1,\ldots,Q_p^{-1}e_p]$, and

$$\tilde{Z}_i = [0\ \ldots\ 0,\ Ae_i,\ 0\ \ldots\ 0] \qquad (i=1\ldots p) .$$

Let σ^{ij} be the typical element of Σ^{-1}, then the condition (43.i) reduces to

$$\sum_{j=1}^{p} \sigma^{ij} e_i' A' Q_j^{-1} e_j = 0 \qquad (i=1\ldots p) . \qquad (44)$$

Now $A'Q_j^{-1} = \begin{bmatrix} \rho_j & -1 & & \\ & \ddots & \ddots & \\ & & \rho_j & -1 \\ & & & 0 \end{bmatrix} = \rho_j R_1 - R_2$, where $R_1 = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ 0 & & & 0 \end{bmatrix}$ and $R_2 = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}$.

Then $e_i'A'Q_j^{-1}e_j = \rho_j e_i'R_1 e_j - e_i'R_2 e_j$,

so that (44) can be written as

$$\sum_{j=1}^{p} (\sigma^{ij} e_i'R_1 e_j)\rho_j = \sum_{j=1}^{p} \sigma^{ij} e_i'R_2 e_j \qquad (i=1...p),$$

or in matrix notation

$$(\Sigma^{-1} \circ E'R_1E)\rho = (\Sigma^{-1} \circ E'R_2E)s ,$$

where $\rho' = (\rho_1 ... \rho_p)$, s is a vector consisting of p 1's and $C \circ D = [c_{ij}d_{ij}]$ is the Schur product.
The θ-conditions (43) may now be written as

$$\begin{cases} \text{(i)} & \hat{\rho} = (\hat{\Sigma}^{-1} \circ E'R_1E)^{-1}(\hat{\Sigma}^{-1} \circ E'R_2E)s \\ \text{(ii)} & \hat{\Sigma} = \frac{1}{T}\hat{Z}'\hat{Z} . \end{cases} \qquad (45)$$

We have to make sure that the expressions in (45) exist, i.e. that $\hat{\Sigma}$ and $\hat{\Sigma}^{-1} \circ E'R_1E$ have rank p. Since the Schur product of two positive definite matrices is also positive definite (see Bellman (1970, p.95)), sufficient for $\hat{\Sigma}^{-1} \circ E'R_1E$ to be positive definite is that $\hat{\Sigma}$ is positive definite and $E'R_1E$ is nonsingular. Let E be the (T-1,p) matrix that is derived from E by deleting its last row, then sufficient for $\hat{\Sigma}$ and $\hat{\Sigma}^{-1} \circ E'R_1E$ to be positive definite is that

$$\text{rank } (\hat{Z}) = \text{rank } (\tilde{E}) = p .$$

When we now add the condition for β to the θ-conditions (45) , it is clear that we have formed three well-defined functions:

$$\begin{cases} \text{(i)} & \hat{\beta} = \beta(\hat{\rho},\hat{\Sigma}) = [X'(\hat{Q}^{-1})'(\hat{\Sigma}^{-1} \otimes I)\hat{Q}^{-1}X]^{-1}X'(\hat{Q}^{-1})'(\hat{\Sigma}^{-1} \otimes I)\hat{Q}^{-1}y \\ \text{(ii)} & \hat{\rho} = \rho(\hat{\beta},\hat{\Sigma}) = (\hat{\Sigma}^{-1} \bullet E'R_1E)^{-1}(\hat{\Sigma}^{-1} \bullet E'R_2E)s \\ \text{(iii)} & \hat{\Sigma} = \Sigma(\hat{\rho},\hat{\beta}) = \frac{1}{T}\hat{Z}'\hat{Z}. \end{cases} \qquad (46)$$

One iterative scheme to find the solution of (46) would be as follows:

(i) Choose the initial values $\rho^{(0)}$ and $\Sigma^{(0)}$

(ii) $\hat{\beta}^{(0)} = \beta(\rho^{(0)},\Sigma^{(0)})$ and $e^{(0)} = y-X\hat{\beta}^{(0)}$

(iii) $\hat{\rho}^{(1)} = \rho(\hat{\beta}^{(0)},\Sigma^{(0)})$

(iv) $\hat{\beta}^{(1)} = \beta(\hat{\rho}^{(1)},\hat{\Sigma}^{(0)})$ and $e^{(1)} = y-X\hat{\beta}^{(1)}$

(v) $\hat{\Sigma}^{(1)} = \Sigma(\hat{\rho}^{(1)},\hat{\beta}^{(1)})$

(vi) $\hat{\beta}^{(2)} = \beta(\hat{\rho}^{(1)},\hat{\Sigma}^{(1)})$ and $e^{(2)} = y-X\hat{\beta}^{(2)}$

etcetera until convergence.

Another scheme, equally feasible, would be to choose $\rho^{(0)}$ and $\Sigma^{(0)}$ and then calculate $\hat{\beta}^{(0)}$, $\hat{\rho}^{(1)}$, $\hat{\Sigma}^{(1)}$, $\hat{\beta}^{(1)}$, $\hat{\rho}^{(2)}$, $\hat{\Sigma}^{(2)}$, $\hat{\beta}^{(2)}$, etcetera.
There are several other possibilities.

Parks' procedure can be completely described in terms of the first scheme, based on the initial values $\rho^{(0)}=0$ and $\Sigma^{(0)}=I$.
In that case

$$\hat{\beta}^{(0)} = (X'X)^{-1}X'y \quad \text{and} \quad e^{(0)} = y-X\hat{\beta}^{(0)}$$

$$\hat{\rho}_j^{(1)} = \left(\sum_{k=1}^{T-1} e_{jk}e_{j,k+1}\right) \Big/ \sum_{k=1}^{T-1} e_{jk}^2 \qquad (j=1\ldots p) \text{ and } \hat{Q}= Q(\hat{\rho}^{(1)})$$

$$\hat{\beta}^{(1)} = [X'(\hat{Q}^{-1})'\hat{Q}^{-1}X]^{-1}X'(\hat{Q}^{-1})'\hat{Q}^{-1}y \; ; \; e^{(1)} = y-X\hat{\beta}^{(1)}; \; \hat{Z}= Z(\hat{Q},e^{(1)})$$

$$\hat{\Sigma}^{(1)} = \frac{1}{T}\hat{Z}'\hat{Z}$$

$$\hat{\beta}^{(2)} = [X'(\hat{Q}^{-1})'(\hat{\Sigma}^{(1)-1} \otimes I)\hat{Q}^{-1}X]^{-1}X'(\hat{Q}^{-1})'(\hat{\Sigma}^{(1)-1} \otimes I)\hat{Q}^{-1}y.$$

$\hat{\beta}^{(2)}$ is Parks' so-called 'three-step' estimator. However, continuing the procedure until convergence yields the ML estimator.

To construct the information matrix we proceed as follows:

$$K_i = \frac{\partial Q^{-1}}{\partial \rho_i} Q = Y^{ii} \otimes AQ_i \qquad (i=1...p),$$

where Y^{ii} is defined in (24) and A in (42).

$$K_iK_j = \begin{cases} 0 & \text{if} \quad i \neq j \\ Y^{ii} \otimes (AQ_i)^2 & \text{if} \quad i=j \; . \end{cases}$$

$$AQ_i = \begin{bmatrix} 0 & & & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ & & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & & & \\ \rho_i & \ddots & & \\ \vdots & \ddots & \ddots & \\ \rho_i^{n-1} & \cdots & \rho_i & 1 \end{bmatrix} = - \begin{bmatrix} 0 & & & & 0 \\ 1 & \ddots & & & \\ \rho & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ \rho^{n-2} & \cdots & \rho & 1 & 0 \end{bmatrix} .$$

It is clear that

$$tr(AQ_i) = tr(AQ_i)^2 = trK_iK_j = 0$$

and

$$tr(AQ_i)'AQ_j = \sum_{k=0}^{n-2} \sum_{h=0}^{k} (\rho_i\rho_j)^h .$$

Now let $\lambda = \rho_i\rho_j$, then

$$\sum_{k=0}^{n-2} \sum_{h=0}^{k} \lambda^h = \sum_{k=0}^{n-2} \frac{1-\lambda^{k+1}}{1-\lambda} = \frac{1}{1-\lambda} \left[n-1 - \frac{\lambda(1-\lambda^{n-1})}{1-\lambda}\right]$$

$$= \frac{n}{1-\lambda} - \frac{1-\lambda^n}{(1-\lambda)^2} .$$

From this it follows that

$$trK_i'(\Sigma^{-1} \otimes I)K_j(\Sigma \otimes I)$$

$$= tr(Y^{ii} \otimes Q_i'A')(\Sigma^{-1} \otimes I)(Y^{ii} \otimes AQ_j)(\Sigma \otimes I)$$

$$= tr(Y^{ii}\Sigma^{-1}Y^{jj}\Sigma) \otimes (Q_i'A'AQ_j)$$

$$= \mathrm{tr}(Y^{ii}\Sigma^{-1}Y^{jj}\Sigma) \otimes (Q_i'A'AQ_j)$$

$$= \sigma^{ij}\sigma_{ij}\left[\frac{n}{1-\rho_i\rho_j} - \frac{1-(\rho_i\rho_j)^n}{(1-\rho_i\rho_j)^2}\right].$$

Further

$$\mathrm{tr}K_i'(\Sigma^{-1}Y^{hk} \otimes I) = \mathrm{tr}(Y^{ii} \otimes Q_i'A')(\Sigma^{-1}Y^{hk} \otimes I)$$

$$= \mathrm{tr}(Y^{ii}\Sigma^{-1}Y^{hk})\ \mathrm{tr}(Q_i'A') = 0 .$$

Using the formulae in (26) we have

$$\Psi_\theta = \begin{bmatrix} \Psi_{\rho\rho} & 0 \\ 0 & \Psi_{\sigma\sigma} \end{bmatrix} , \qquad (47)$$

where $(\Psi_{\rho\rho})_{ij} = 2\sigma^{ij}\sigma_{ij}\left[\frac{n}{1-\rho_i\rho_j} - \frac{1-(\rho_i\rho_j)^n}{(1-\rho_i\rho_j)^2}\right]$ $(i,j=1...p)$,

and $\Psi_{\sigma\sigma}$ is defined in (26.vi).

References

Bellman, Richard, 1970, Introduction to matrix analysis, 2nd ed. (McGraw-Hill, New York).

Cramèr, H., 1946, Mathematical methods of statistics (Princeton Un. Press, Princeton).

Dhrymes, Phoebus J., 1970, Econometrics: Statistical foundations and applications (Harper and Row, New York).

Dhrymes, Phoebus J., 1971, Equivalence of iterative Aitken and maximum likelihood estimators for a system of regression equations, Australian Economic Papers 10, 20-24.

Fuller, W.A. and G.E. Battese, 1973, Transformations for estimation of linear models with nested-error structure, Journal of the American Statistical Association 68, 626-632.

Kadiyala, K.R., 1968, A transformation used to circumvent the problem of autocorrelation, Econometrica 36, 93-96.

Kakwani, N.C., 1967, The unbiasedness of Zellner's seemingly unrelated regression equation estimators, Journal of the American Statistical Association 62, 141-142.

Kendall, M.G. and A. Stuart, 1967, The advanced theory of statistics, vol. II, 2nd ed. (Charles Griffin, London).

Magnus, Jan R., 1977 a, Maximum likelihood estimation of the GLS model: A general approach and an application to the heteroskedastic model (technical report, Instituut voor Actuariaat en Econometrie, Universiteit van Amsterdam).

Magnus, Jan R., 1977 b, Maximum likelihood estimation of error component models (technical report, Instituut voor Actuariaat en Econometrie, Universiteit van Amsterdam).

Magnus, Jan R. and H. Neudecker, 1977, The commutation matrix: some theorems and applications, submitted for publication.

Mehta, J.S. and P.A.V.B. Swamy, 1976, Further evidence on the relative efficiencies of Zellner's seemingly unrelated regressions estimator, Journal of the American Statistical Association 71, 634-639.

Neudecker, H., 1969, Some theorems on matrix differentiation with special reference to Kronecker matrix products, Journal of the American Statistical Association 64, 953-963.

Oberhofer, W. and J. Kmenta, 1974, A general procedure for obtaining maximum likelihood estimates in generalized regression models, Econometrica 42, 579-590

Parks, Richard W., 1967, Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated, Journal of the American Statistical Association 62, 500-509.

Swamy, P.A.V.B. and J.S. Mehta, 1969, On Theil's mixed regression estimator, Journal of the American Statistical Association 64, 273-276.

Theil, H., 1971, Principles of Econometrics (Wiley, New York).

Vickers, M. Kathleen, 1977, Optimal Asymptotic properties of maximum likelihood estimators for parameters of some econometric models (technical report, Cornell University).

Weiss, Lionel, 1973, Asymptotic properties of maximum likelihood estimators in some nonstandard cases, II, Journal of the American Statistical Association 68, 428-430.

Weiss, L. and J. Wolfowitz, 1967, Maximum probability estimators, Annals of the Institute of Statistical Mathematics 19, no.2, 193-206.

Zellner, Arnold, 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, Journal of the American Statistical Association 57, 348-368.