ON EFFICIENT ESTIMATION OF THE FINAL EQUATION FORM
OF A LINEAR MULTIPLE TIME SERIES PROCESS

F. PALM
Free University of Amsterdam

1. Introduction.

Linear vector time series models now belong to the standard tools of the econometrician. In recent years, many contributions have lead to a better understanding of the problems of specification, identification, testing and estimation of linear dynamic systems. Specification and testing of a multiple time series process have been considered for example by Granger and Newbold (1975), Hendry (1975), Mizon (1975), Wallis (1975) and Zellner and Palm (1974, 1975). Hannan (1969) gives necessary and sufficient conditions for the identification of vector mixed autoregressive-moving average systems. Among many others, Dunsmuir and Hannan (1976), Hannan (1975), Nelson (1973), Nicholls (1978), Palm and Zellner (1974), Wallis (1976) and Wilson (1973) have recently studied the problem of parameter estimation of vector autoregressive moving average models. For an excellent review of the literature on estimation of vector and scalar dynamic models, the reader is referred to Aigner (1971), Aström and Eykhoff (1971), and to Nicholls, Pagan and Terril (1975).

Estimation methods for dynamic models with autoregressive errors, which have a long tradition in econometrics, are reviewed by Hendry (1976).

In the present paper, we consider large sample estimation of the final equation form of a linear multiple time series process. Non-linear least squares (see e.g. Box and Jenkins (1970)) and maximum likelihood methods (see e.g. Osborn (1970)) are widely used to estimate single final equations and systems of final equations. Asymptotic properties of the estimation procedures are compared with those of the maximum likelihood estimator. In the framework of dynamic models with autocorrelated errors, iterative computation is generally the price to pay for asymptotic efficiency. It may be too high a price to pay, especially in the case of joint estimation.


In this paper, we develop and implement the procedure by Hataneke and Rainesal to obtain an efficient two-step estimator for the system of final equations. Of course, the final equations may be analyzed by single equation methods. This down to write a univariate ARMA process for each of the variables out of a multiple ARMA process. Consistent and asymptotically efficient univariate estimation methods are well-known (see
e.g. Box and Jenkins (1970). However, part of the available information on the process is neglected in the univariate analysis. Therefore, one may naturally be led to joint estimation of the final equation parameters to improve the precision of the results from the univariate analysis. The issue of using joint against single equation methods has to be considered as a matter of efficiency.

The final equation form is a useful parametrization of a multiple time series model as it explicitly brings out the characteristic equation of the dynamic model. In many situations, one may have rather accurate knowledge on the characteristic equation, e.g., the number of characteristic roots, the value of these roots... For example, the presence of seasonality, business cycles, trend... in econometric time series conditions the form of the characteristic equation. In the final equation form the simultaneity and interdependence in the multiple process are moved from the "deterministic" part to the error correlation structure. Joint estimation of the system of final equations allows to capture the simultaneity through taking into account the cross-equation error correlations.

Restrictions such as identical autoregressive coefficients for all the final equations may be incorporated in the two-step estimator, that we present in this paper. Other restrictions on the final equation parameters coming from the underlying structural form are ignored. The estimator is efficient in the sense that it has the same asymptotic covariance matrix as the maximum likelihood estimator for an unrestricted (except for possible identical autoregressive parts) final equation form. In section II, we specify the final equation form of a linear multiple time series process, we present an efficient estimator for the final equation parameters and develop a procedure to compute it. Finally in section III, we summarize the results and discuss problems that remain to be analyzed. Some of the analytical developments are given in the appendix.

II. Efficient estimation of the final equation form.

Consider a covariance stationary random vector \( \mathbf{z}_t \) and assume that it can be represented by the following multiple time series process \(^1\)

\[
H(L) \mathbf{z}_t = F(L) \mathbf{u}_t \\
p \times p \quad p \times p \quad p \times p
\]

where \( \mathbf{u}_t \) is a vector of random errors.

\( H(L) \) and \( F(L) \) are each matrix lag operators, assumed of full rank, with typical elements being finite degree polynomials in \( L \), namely \( h_{ij}(L) \) and \( f_{ij}(L) \).

---

\(^1\) For reasons of simplicity, we assume here \( E \mathbf{z}_t = 0 \).
Assume that $\varepsilon_t$ is normally distributed with

$$E \varepsilon_t = 0$$ and $E \varepsilon_t \varepsilon_{t'}' = \delta_{tt'} \Sigma \sigma^2$.

(2)

for all $t$ and $t'$, where $\delta_{tt'}$ is the Kronecker delta. Hannan [1986] gives necessary and sufficient conditions for identification of a model in the form of (1) with $H(0) = I$, in what follows, we assume that the system (1) satisfies Hannan's identification conditions, in particular that $H(L)$ is invertible, so that the roots of its determinant $|H(L)|$, a scalar polynomial in $L$ of degree $n$, lie outside the unit circle.

The system of linear equations (FES) associated with (1) is given by:

$$\phi(L) \varepsilon_t = \sum_{h=0}^{q} A_h \varepsilon_{t-h}, \quad t = 1, \ldots, T.$$  

(3)

where $\phi(L) = |H(L)|/c$ is a scalar polynomial in $L$ operating on each element of $\varepsilon_t$, $A(L) = H(0)/c$ is a matrix lag operator of order $q$, $H^*(L)$ is the adjoint matrix of $H(L)$ and the normalizing constant $c$ is the coefficient of $L^0$ in $|H(L)|$.

As pointed out by Zellner and Palm (1974) and Palm (1977), the autoregressive polynomial $\phi(L)$ operates on each element of $\varepsilon_t$. Unless there is cancelling, the autoregressive parts of the equations in (3) are identical. Since it is often of interest to test that the autoregressive parameters are the same in different equations and also for more generality, we shall discuss the problem of estimation of the following system

$$\begin{bmatrix}
\phi_1(L) & \ldots & 0 \\
0 & \ddots & \vdots \\
0 & \ddots & \phi_p(L) \\
0 & \ddots & 0 \\
\end{bmatrix} \begin{bmatrix}
\varepsilon_t \\
\varepsilon_{t-1} \\
\vdots \\
\varepsilon_{t-q} \\
\end{bmatrix} = \begin{bmatrix}
\varepsilon_t \\
\varepsilon_{t-1} \\
\vdots \\
\varepsilon_{t-q} \\
\end{bmatrix} + \sum_{h=1}^{q} B_h \varepsilon_{t-h}.$$  

(4)

where $\phi_i(L) = \phi_{i1} L^1 \ldots \phi_{in_i} L_{n_i}$, with $n_i$ given, $i = 1, \ldots, p$, and

$$E[\varepsilon_t \varepsilon_{t'}'] = \delta_{tt'} \Sigma \sigma^2.$$

Assuming initial conditions 2) to be zero, i.e. $\varepsilon_0 = \varepsilon_{-1} = \varepsilon_{-2} = \ldots = 0$,

we can write the model in matrix notation

$$E[\varepsilon_t \varepsilon_{t'}'] = \delta_{tt'} \Sigma \sigma^2.$$

Assuming initial conditions 2) to be zero, i.e. $\varepsilon_0 = \varepsilon_{-1} = \varepsilon_{-2} = \ldots = 0$,

we can write the model in matrix notation

$$E[\varepsilon_t \varepsilon_{t'}'] = \delta_{tt'} \Sigma \sigma^2.$$

(2) Of course, the starting values can be treated as parameters. (see for example Phillips (1988)). The starting values do not affect the asymptotic results. For the sake of simplicity, they are here set equal to zero. However, in most applications, the problem of starting values deserves special attention.
as follows:

\[ Z_t = W_{1t} \cdot \hat{\mu} + u_t \]  \hspace{1cm} (5)

with

\[
W_{1t} = \begin{bmatrix}
    z_{1t-1} & z_{2t-2} & \ldots & z_{t-n_1} & 0 & 0 & 0 & \ldots & 0 \\
    0 & 0 & 0 & z_{2t-1} & \ldots & z_{2t-n_2} & 0 & \ldots & 0 \\
    0 & 0 & \vdots & \vdots & \ddots & \vdots & 0 & \ldots & 0 \\
    0 & 0 & \ldots & \vdots & \ddots & \vdots & \vdots & \ldots & 0 \\
    0 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & z_{pt-n_p}
\end{bmatrix}
\]

\[ \hat{\psi}' = [\hat{\phi}_{11}, \hat{\phi}_{12}, \ldots, \hat{\phi}_{1n_1}, \hat{\phi}_{21}, \ldots, \hat{\phi}_{pn_p}]' \]  \hspace{1cm} (6)

\[ 1 \times k \]

\[ k = \sum_{i=1}^{p} n_i \]

and

\[ u_t = v_t - \sum_{h=1}^{n_h} B_h \cdot v_{t-h} \]

and for a sample of \( T \) observations

\[ Z = W_1 \cdot \hat{\psi} + u \]  \hspace{1cm} (7)

where

\[ Z' = [z_1', z_2', \ldots, z_T'] \]

\[ W_1' = [w_{11}', w_{12}', \ldots, w_{1n_1}'] \]

\[ u' = [u_1', u_2', \ldots, u_T'] \]

Notice that \( u \) may be expressed in terms of \( v \)

\[ u = MV \]  \hspace{1cm} (8)

where

\[ M = \begin{bmatrix}
    I & 0 & \ldots & 0 \\
    0 & I & \ldots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \ldots & B_{n_1}
\end{bmatrix} \]

and \( v' = [v_1', \ldots, v_T'] \)

From the distributional assumptions on \( u \), we derive the joint density of the vector \( v \)

\[ D(v) = \frac{1}{\sqrt{2\pi}} |\Sigma_{v}^{1/2}| \exp \left( -\frac{1}{2} v' \left( I_{p} \otimes \Sigma_{v}^{-1} \right) v \right) \]  \hspace{1cm} (9)
As the jacobian of the transformation \( y \rightarrow u \rightarrow z \) is one, the data density can be written as:

\[
\begin{align*}
0 (z | z_0 = z_{-1} = \ldots = z_0 = \emptyset, M, \Sigma) &= \\
\| \Sigma \|^T/2 \exp \left[ -\frac{1}{2} (z - W_1 \emptyset)' M^{-1} (I_T \otimes \Sigma^{-1}) M^{-1} (z - W_1 \emptyset) \right].
\end{align*}
\] (10)

Through \( M^{-1} \) the model is clearly non-linear in the parameters. Computation of the maximum likelihood estimator naturally leads to an iterative procedure. The first order conditions for a maximum of the log-likelihood function are

\[
\frac{\partial S}{\partial \rho} = -2 \omega_2 M^{-1} (I_T \otimes \Sigma^{-1}) M^{-1} u = 0
\] (11)

where \( S = (z - W_1 \emptyset)' M^{-1} (I_T \otimes \Sigma^{-1}) M^{-1} (z - W_1 \emptyset) \).

In the appendix 1, we derive the following expression for the derivatives of \( S \) with respect to the element of \( M \):

\[
\frac{\partial S}{\partial \rho} = -2 \omega_2 M^{-1} (I_T \otimes \Sigma^{-1}) M^{-1} u = 0
\] (12)

where

\[
\begin{align*}
\emptyset &= \text{vec} [B_1, \ldots, B_q]' \\
\omega_2 &= |V_{k'}| \\
\rho &\equiv \rho_I \\
V_{k'} &= \left[ v_{ij} \right]_{i=1}^{kp} \\
\text{for } i &= \text{ith position, } j = \text{ith position}
\end{align*}
\]

The set of first order conditions for a maximum of the likelihood function with respect to \( \rho' = (\rho_1, \ldots, \rho_q)' \) can be summarized as

\[
\frac{\partial S}{\partial \rho} = -2 \omega_2 M^{-1} (I_T \otimes \Sigma^{-1}) M^{-1} u = 0
\] (13)

with

\[
\omega = (W_1 \omega_2).
\]

System (13) calls for several remarks:

1. Its solution gives a stationary point of the likelihood function for the multivariate ARMA process (4) with initial conditions set equal to zero.

2. For given \( \Sigma \), the set of equations in (13) is non-linear in the parameters of \( M \). Therefore solving (13) will usually require an iterative procedure so that the computational burden of the maximum likelihood estimator may become heavy.

3. As the information matrix is block-diagonal for \( \rho \) and \( \Sigma \), the use of a consistent estimator for an unknown \( \Sigma \) in (13) will not affect the asymptotic properties of the solution of (13) for further discussion see Hendry (1976), Maddala (1971).
An alternative to an exact iterative solution of (13) consists in approximating the first order conditions. Using the mean-value theorem, Chyuens and Taylor (1976) and Raivais (1976) show that if \( \hat{\beta} \) is a consistent estimator of \( \beta \) such that \( \sqrt{T}(\hat{\beta} - \beta) \) has a limiting distribution and \( \Gamma(\beta) \) is a nonsingular matrix such that

\[
\lim_{T \to \infty} \frac{1}{T} \sqrt{\sum_{t=1}^{T} (\hat{\beta} - \beta)^2} \to \sigma
\]

then the estimator

\[
\hat{\beta} = \hat{\beta} - \frac{1}{\sigma} (\hat{\beta} - \beta) \cdot \frac{\partial \hat{\beta}}{\partial \beta} \cdot \hat{\beta}
\]

has the same limiting distribution as the maximum likelihood estimator \( \hat{\beta}_{ML} \). In the appendix, we show that

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} (\hat{\beta} - \beta)^2 = \frac{1}{T} \sum_{t=1}^{T} (\hat{\beta} - \beta)^2 \cdot \hat{\beta} - \beta
\]

If we evalute (13) and (16) at some consistent estimates \( \hat{\beta} \) and \( \hat{\Sigma} \), the estimator

\[
\hat{\beta} = \hat{\beta} - \frac{1}{\sigma} (\hat{\beta} - \beta) \cdot \frac{\partial \hat{\beta}}{\partial \beta} \cdot \hat{\beta}
\]

has the same asymptotic properties as the maximum likelihood estimator (for its properties, see for example Dunsmuir and Hennem (1976) solving (13), i.e. \( \hat{\beta} \) is a consistent estimator of \( \beta \), asymptotically normally distributed and efficient. Its covariance matrix is consistently estimated by:

\[
\hat{\Sigma} = \left( \hat{\beta} - \beta \right)^{-1} \hat{\Sigma} \left( \hat{\beta} - \beta \right)^{-1}
\]

The estimator \( \hat{\beta} \) in (17) can also be written in the form of Hatemak's "residual adjusted" estimator.

Expression (5) may be transformed into

\[
\hat{\Sigma}^{-1} \cdot \hat{\beta} = \hat{\Sigma}^{-1} \cdot \hat{\beta} + \sum_{t=1}^{T} \hat{\beta}_t \cdot \hat{\beta}_t^T - \hat{\Sigma}^{-1}
\]

where \( \hat{\Sigma}_h \) is some estimate of \( \Sigma \) and \( \hat{\Sigma}_h \) is the corresponding computed residuals. Ignoring the last right-hand side term of (19), we get an approximate expression

\[
\hat{\Sigma} = \hat{\Sigma}^{-1} \cdot \hat{\beta} + \sum_{t=1}^{T} \hat{\beta}_t \cdot \hat{\beta}_t^T - \hat{\Sigma}^{-1}
\]

where \( \hat{\Sigma} = \hat{\Sigma}^{-1} \cdot \hat{\beta} \cdot \hat{\beta}_t \cdot \hat{\beta}_t^T \).

With initial values set equal to zero, we have for the sample

\[
\hat{\Sigma} = \hat{\Sigma}^{-1} \cdot \hat{\beta} + \hat{\Sigma} \cdot \hat{\beta}_t \cdot \hat{\beta}_t^T
\]

\[
\hat{\Sigma} = \hat{\Sigma}^{-1} \cdot \hat{\beta} + \hat{\Sigma} \cdot \hat{\beta}_t \cdot \hat{\beta}_t^T
\]
with \( W = (\tilde{W}_1, \tilde{W}_2) \).

Generalized least-squares applied to (21) with estimated \( \Sigma \) yields
\[
\theta_{GLS} = (\tilde{W} \cdot \tilde{M}^{-1} (I_{\tilde{T}} \otimes \tilde{\Sigma}^{-1}) \tilde{M}^{-1} \tilde{W})^{-1} (\tilde{W} \cdot \tilde{M}^{-1} (I_{\tilde{T}} \otimes \tilde{\Sigma}^{-1}) \tilde{M}^{-1} \tilde{W})^{-1} \tilde{Y}
\] (22)

which is a "residual adjusted" estimator. Further, it has the same form as the estimator given in (17) because
\[
\tilde{Y} = \tilde{W} \cdot \tilde{B} + \tilde{M} \cdot \tilde{Y} = \tilde{W} \cdot \tilde{B} + \tilde{Y}.
\]

The procedure proposed in Palm and Zellner (1974) can be used to compute the estimator in (17):

**Step I:**

a) Estimate the parameter of the single final equations using non-linear least squares or the single equation maximum likelihood method.
\[
\hat{q}_i = \hat{q}_{i(L)} \hat{z}_{it} = \hat{v}_{it} + \sum_{j=1}^{p} \hat{\beta}_{ij} \hat{v}_{it-j}, \quad i = 1, \ldots, p
\] (23)

These methods yield consistent estimates for \( \hat{q}_{i(L)} \) and the \( \hat{\beta}_{ij} \)'s. Compute the residuals \( \hat{v}_{it} \).

b) Apply ordinary least squares to each equation
\[
\hat{z}_{it} = \hat{w}_{1it} \hat{d}_1 + \hat{w}_{2it} \hat{d}_2 + \hat{v}_{it}
\] (24)

where
\[
\begin{align*}
\hat{w}_{1it} & = (\hat{z}_{it-1}, \ldots, \hat{z}_{it-n_1}) \\
\hat{d}_1 & = (\hat{y}_{11}, \ldots, \hat{y}_{in_1}) \\
\hat{w}_{2it} & = (\hat{w}_{1i-1}, \hat{w}_{1i-2}, \ldots, \hat{w}_{1i-n_2}) \\
\hat{d}_2 & = \text{ith row of } B = (B_1, B_2, \ldots, B_q)
\end{align*}
\]

We get a consistent estimate of \( \hat{d}_1 \) and \( \hat{d}_2 \).

c) Estimate \( \Sigma \) using
\[
\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \hat{v}_{il} \hat{v}_{it}'
\]
where the \( \hat{v}_{il} \)'s are the residuals obtained in step I-b.

**Step II:**

Use the estimated \( \hat{d}_1, \hat{d}_2, \hat{z} \) and the \( \hat{v}_{it} \)'s to form
\[
\hat{W} = (\hat{W}_1, \hat{W}_2) \quad \hat{M} \quad \text{and to evaluate expression (17)}.
\]
III. Some final remarks and tentative conclusions

1. The "residual adjusted" estimator has the same asymptotic properties as the maximum likelihood estimator under initial conditions equal to zero. There is however room for a deeper treatment of the problem of initial values.

2. The estimator is justified on the basis of its large sample properties. Small sample properties have to be investigated and compared with those of other estimators, such as the maximum likelihood estimator or single equation procedures.

3. In order to compute the estimator in (17), we have to invert the covariance matrix of a vector moving average process, \( M(I_q \otimes \Sigma M) \), which involves the inverse of the large triangular matrix \( M \).

\( M^{-1} \) can be written as

\[
M^{-1} = \begin{bmatrix}
I & 0 & \cdots & 0 \\
D_1 & I \\
D_2 & D_4 \\
\vdots & \vdots & \ddots & \vdots \\
D_{q-1} & D_q & \cdots & I
\end{bmatrix}
\]

where

\[
D_j = -[\sum_{i=1}^{j} B_i D_{j-i}] \quad 1 \leq j \leq q,
\]

with

\( D_0 = I \)

and

\[
D_j = -[\sum_{i=1}^{q} B_i D_{j-i}] \quad j > q
\]

so that the computation of \( M^{-1} \) only requires multiplication and addition of matrices of order \( p \times p \).

4. The "residual adjusted" estimator facilitates the incorporation of restrictions on the parameters. For example, if we use the restrictions that the initial equations have identical autoregressive part, the matrix \( W_1 \) has the following form:

\[
W_1 = \begin{bmatrix}
\bar{z}_0 & \bar{z}_{-1} & \cdots & \bar{z}_{-n+1} \\
\bar{z}_1 & \bar{z}_0 & \cdots & \bar{z}_{-n+2} \\
\vdots & \vdots & \ddots & \vdots \\
\bar{z}_n & \bar{z}_{n-1} & \cdots & \bar{z}_{n-2} \\
\end{bmatrix}
\]

so that the restrictions are easily imposed. They can be tested using a likelihood ratio test. Further, setting some of the elements of \( \beta \) equal to zero is also straightforward.
5. It should be noticed that there is a class of efficient two-step estimators. Each member of this class is characterized by a particular choice of the matrix \( \gamma \), subject to the requirement on its probability limit in (14).

6. The "residual adjusted" estimator can also be generalized and applied to estimate the structural, reduced or transfer function form of a multivariate time series model with exogenous variables. For the transfer function form, the single equation approach is an analysis of the marginal density of one endogenous variable conditionally on the exogenous variables, whereas the systems methods apply to the parameters of the joint density of the vector of endogenous variables conditionally on the exogenous variables.

In conclusion, we have presented an efficient two-step estimator for the final equation form of a dynamic model with moving average errors. Instead of an iterative solution, the estimator approximates the first order conditions for a maximum of the likelihood function by the first step of the Newton-Raphson method starting from consistent estimates of the parameters. Asymptotically, the approximation holds true. It requires however the inverse of the usually large error covariance matrix. To invert the error covariance matrix, we fully exploit the assumption that the errors are generated by a moving average process.

Appendix.

1. Differentiation of \( S \) with respect to \( b_{ij}^k \) where \( b_{ij}^k \) is the \( ij \)th element of \( B_k \) leads to:

\[
\frac{\partial S}{\partial b_{ij}^k} = (z - w_i \phi) \left[ \left( I_r \otimes \Sigma^{-1} \right) M^{-1} \right] \frac{\partial M^{-1}}{\partial b_{ij}^k} (z - w_i \phi) \\
+ (z - w_i \phi) M^{-1} \left[ I_r \otimes \Sigma^{-1} \right] \frac{\partial M^{-1}}{\partial b_{ij}^k} (z - w_i \phi) \\
= -2 (z - w_i \phi) M^{-1} \left[ I_p \otimes \Sigma^{-1} \right] \frac{\partial M}{\partial b_{ij}^k} \left[ I_r \otimes \Sigma^{-1} \right] M^{-1} (z - w_i \phi) \\
\]  

(A.1)

(A.2)

where we use the matrix differentiation rules

\[
\frac{\partial \Sigma}{\partial x} = \Sigma \frac{\partial \Sigma}{\partial x} \Sigma \\
\frac{\partial \Sigma^{-1}}{\partial x} = -\Sigma \frac{\partial \Sigma}{\partial x} \Sigma \\
\]  

(A.3)

and the fact that \( S \) is a scalar.

The derivative of \( M' \) with respect to \( b_{ij}^k \) is a \( T \times p \) matrix with the \( [ (k+q)\rho + 1, \rho + j ] \)th element equal to one for \( l = 0, 1, \ldots, q-1 \), and zero otherwise. As \( (z - w_i \phi)' M^{-1} y' = y' M^{-1} y \), we can write (A.2) as:

\[
\frac{\partial S}{\partial b_{ij}^k} = -2y' \left[ I_p \otimes \Sigma^{-1} \right] M^{-1} (z - w_i \phi) \\
= -2y_{ij}^k M^{-1} \left[ I_p \otimes \Sigma^{-1} \right] M^{-1} (z - w_i \phi) \\
\]  

(A.4)

(A.5)
with

\[ \phi_{ij}^k = 0 \ldots 0 \ 0 \ldots 0 \ 0 \ldots 0 \ v_{ij} \ 0 \ldots 0 \ v_{ij} \ 0 \ldots 0 \ v_{ij-k} 0 \]

\[ 1 \times p^T \]
\[ 1 \times p^T \]
\[ k \text{ times} \]
\[ p \]
\[ p \]
\[ i^{th} \text{ position} \]

where \( i = 1, \ldots, q; j = 1, \ldots, p \).

Define the \( (p^q \times p^7) \) matrix

\[ w_2^i = [v_{ij}^k] \]  \hspace{1cm} (A.8)

with typical row being \( \phi_{ij}^k \).

Then

\[ \frac{\Delta S}{\Delta t} = -2w_2^i M^{-1} (I_T \otimes \Sigma^T) M^{-1} \ y = 0 \]  \hspace{1cm} (A.7)

where \( \delta \times \text{vec} \{ \delta_{12}, \delta_{23}, \ldots, \delta_{1 q} \} \).

2. Now we derive the second order derivatives of \( S \) with respect to \( \delta \).

\[ \frac{\Delta^2 S}{\Delta ^2 \delta} = 2w_2^i M^{-1} (I_T \otimes \Sigma^T) M^{-1} \ y \]  \hspace{1cm} (A.8)

\[ \begin{align*}
\frac{\Delta^2 S}{\delta^2} & = 2w_2^i M^{-1} \left( I_T \otimes \Sigma^T \right) M^{-1} y \\
& + 2w_2^i M^{-1} \left( I_T \otimes \Sigma^T \right) M^{-1} \frac{\Delta M}{\delta \delta_{ij}} M^{-1} y
\end{align*} \]  \hspace{1cm} (A.9)

where \( w_{rs} = [0 \ldots 0 z_{r-s} 0 \ldots 0 z_{r-s-1} \ldots 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \)

\[ \begin{array}{c}
\text{r}\text{-th position} \\
\text{p+r}\text{-th position}
\end{array} \]

is the \( (r + 1) \times (n_s + s) \)th row of \( w_2^i \).

Notice that \( M^{-1} \frac{\Delta M}{\delta \delta_{ij}} M^{-1} \) is an upper triangular matrix (with many zero elements in the upper triangle) and that \( z_{r \times t} \) and \( v_{rt} \) are independent for \( t > 0 \) and for all \( t \). For \( \delta > 1 \), the first right-hand side term in (A.9) is a martingale whose terms have a variance of order \( T \), so that by the law of large numbers for martingales we have:

\[ p \lim_{T \to \infty} 2w_2^i M^{-1} \frac{\Delta M}{\delta \delta_{ij}} M^{-1} \ y = 0 \]  \hspace{1cm} (A.10)

therefore

\[ p \lim_{T \to \infty} \frac{\Delta^2 S}{\delta \delta_{ij}} = p \lim_{T \to \infty} 2w_2^i M^{-1} \left( I_T \otimes \Sigma^T \right) M^{-1} w_2 \]  \hspace{1cm} (A.11)
Further, we have

\[
\frac{\delta S}{\delta \beta_{i,j}} = 2 \mu' \Omega^{-1} \frac{\delta \mu'}{\delta \beta_{i,j}} \Omega^{-1} \mu' \Omega^{-1} (I_T \otimes \Sigma^{-1}) \Omega^{-1} \mu + 2 \mu' \Omega^{-1} \frac{\delta \mu'}{\delta \beta_{i,j}} \Omega^{-1} \mu' \Omega^{-1} (I_T \otimes \Sigma^{-1}) \Omega^{-1} \mu + 2 \mu' \Omega^{-1} \frac{\delta \mu'}{\delta \beta_{i,j}} \Omega^{-1} (I_T \otimes \Sigma^{-1}) \Omega^{-1} \mu \Omega^{-1} \mu' \Omega^{-1} \mu .
\]

(A.12)

Using a similar argument as for (A.10), it can be shown that the first two right-hand side terms of (A.12) divided by \( T \) have zero probability limit, so that we have:

\[
p \lim_{T \to \infty} \frac{2 \Sigma^T}{\mu \Sigma} = p \lim_{T \to \infty} \frac{2}{T} \mu (I_T \otimes \Sigma^{-1}) \Omega^{-1} \mu = 0 .
\]

(A.13)

Expressions (A.8), (A.11) and (A.13) together yield:

\[
p \lim_{T \to \infty} \frac{2 \Sigma^T}{\mu \Sigma} = p \lim_{T \to \infty} \frac{2}{T} \mu (I_T \otimes \Sigma^{-1}) \Omega^{-1} \mu .
\]

(A.14)

where \( \mu = (\mu_1, \mu_2) \).

Acknowledgements.

Thanks are due to J.G. de Goijer for his helpful comments. The author is presently at the Free University of Amsterdam. Part of the work was done while he was a research associate at the center for Operations Research and Econometrics, University of Louvain.

Financial support from both institutions is gratefully acknowledged.

REFERENCES


