

The Spectral Maximum Likelihood Estimation of Econometric Models with Stationary Errors

by

Antoni Espasa



Vandenhoeck und Ruprecht in Göttingen

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Espasa, Antoni

The spectral maximum likelihood estimation of econometric models with stationary errors. – 1. Aufl. – Göttingen : Vandenhoeck & Ruprecht, 1977.

(Angewandte Statistik und Ökonometrie ; H. 3)

ISBN 3-525-11233-5

3

© Vandenhoeck & Ruprecht in Göttingen 1977. – Printed in Germany.
Ohne ausdrückliche Genehmigung des Verlages ist es nicht gestattet, das Buch oder Teile daraus auf foto- oder akustomechanischem Wege zu vervielfältigen. Druck: Hubert & Co., Göttingen

Contents

I Introduction	9
II Notation	13
III Stationary disturbances and asymptotic theory	18
IV Specfilm (Spectral full information maximum likelihood) Estimation	23
IV.1 The model and its likelihood funktion	23
IV.2 The first order conditions	33
IV.3 The second order conditions	42
IV.4 The Hannan-Terrell procedure	47
V The Specfilm estimation with inadequate sample size	49
VI The estimation of the multiple regression model with stationary errors and lagged endogenous variables	53
VI.1 Alternative likelihood functions	56
VI.2 The Spec LS estimator and its asymptotic distribution	58
VI.3 The spectral estimator and the maximum likelihood estimators	65
VII The Specfilm method as applied to models with lagged endogenous variables	73
VII.1 Alternative likelihood functions and first order conditions . . .	73
VII.2 The asymptotic distribution of the Specfilm estimators	78
VIII The asymptotic variance matrix of the structural estimators when the errors follow an AR process	83
IX Final Comment	97
Appendix. Two asymptotically equivalent expressions for the likelihood function	99
Bibliography	103
Abstract	107

Acknowledgements

The research presented here was done at the Econometric department of the London School of Economics, between 1973 and 1975. I wish to thank Professor J. Denis Sargan for the constant help, advice and encouragement he provided throughout the preparation of this study: indeed I owe him a considerable intellectual debt. The sections VII and VIII of this monograph are the result of a joint research with him. I have also benefited from innumerable useful discussions with various members of the L.S.E. econometric group, particularly Dr. David Hendry and Dr. Kenneth Wallis, and with Professor E.J. Hannan in this short visit to the L.S.E. in June 1974. I hold sole responsibility for any residual errors. The financial assistance provided by Fundació Jaume Bofill, Fundación Juan March and British Council are gratefully acknowledged. My thanks also to María del Socorro Aramburu for typing the manuscript.

I Introduction

In the construction of Econometric models it is common practice to specify a structure - more or less based on Economic Theory - relating the dependent and explanatory variables and then add to it a disturbance element, representing the effect of all the omitted variables. In the estimation of the parameters of these models the aim is to obtain estimators that are, at least, consistent and efficient. That is, estimators, that at least for sufficient large samples, will have as their mean the true parameters of the model, and their variance-covariance matrix will be minimum within a class of estimators. But in order to obtain estimators with these properties we need to assume that the explanatory variables and disturbance components satisfy a specific set of constraints. In order to concentrate our attention on the requirements of the error process we will start assuming that the explanatory variables are distributed independently of the disturbance term and then can be treated as a set of fixed regressors. For the asymptotic theory presented below we need to impose some restrictions on the asymptotic behaviour of the moments of the regressors such as stationarity and ergodicity, but to be more general we shall simply require what have been called 'Grenander's Conditions' (see Hannan [1970] section 2.6) under which the asymptotic theory presented below will apply (see Hannan [1970] Chapter 7). In the framework of multiple regression, if we assume that the errors are a white noise process we know that the Ordinary Least Squares (OLS) estimators are consistent and efficient and if the errors are Gaussian then OLS estimators are maximum likelihood estimators (see Goldberger [1964] section 6.2). But the

white noise hypothesis is, in many cases, not a realistic assumption and something more general as the requirement that the errors are stationary (with covariance matrix $K \neq \sigma^2 I$) is called for. In this case, if K is known, we can get consistency and efficiency by applying the Aitken's formula (see Goldberger [1964] 5.4). If the variance matrix is unknown - as is usually the case - but we have a consistent estimator \hat{K} say, we can still, in fixed regressor case, obtain efficiency by replacing K by \hat{K} in the Aitken expression (see Dhrymes [1970] 4.2).

It is obvious, though, that in the estimation of the variance matrix we need to introduce a priori information imposing strong restrictions in K , in order to make the estimation feasible with only one sample realization. The usual way to proceed in the Time Domain framework is to postulate a finite parametric process for the errors and then to estimate K consistently by consistently estimating the parameters of the process. In this approach the autoregressive (AR) formulation has been most thoroughly studied and applied and its application to simultaneous equation models has been considered by Sargan [1961]. More recently the moving average (MA) formulation has been appearing in the literature. (See Nicholls et. al. [1975] for a survey of the developments in the estimation and use of models with MA disturbance terms). So, we have that the departure from the white noise hypothesis constitutes a considerable complication in Time Domain estimation. In general the efficiency of the estimators proposed for the particular cases will be dependent on the validity of the approximating process used and the property of robustness will be sacrificed. In fact, Watson and Hannan [1956], studying the bounds for the efficiency of the regression estimators in cases where the disturbance terms are generated by certain stationary processes whose form are incorrectly prescribed a priori, found that 'in general the bounds are wide, indicating that the effect of not knowing the true form of the error process may be severe'. They also establish certain asymptotic results 'to show that, in many cases which occur in practice, the bounds may be attained'.

Another way of dealing with the problem of stationary errors is by spectral methods, and working with the spectra instead of the variance, i.e. by splitting the variance at different frequency bands and using the cumulative cross-periodogram matrix of the variables instead of their second moments. This method was first proposed, for multiple systems of regressions, by Hannan [1963] and later applied in Economics by K.F. Wallis [1966]. The advantage of the spectral framework is that we do not need either to assume "white-noise" errors or to specify a priori an error model, but only to postulate that a stationary stochastic process generates the error term.

Therefore in those cases where we do not have an a priori idea about the formation of the errors, and the specification of a parametric model is somewhat arbitrary, spectral estimation is a reasonable way to proceed. We will show that by this approach, in fixed regressor case, we obtain estimators that are both, robust and efficient. If the regressor set contains lagged endogenous variables then the proposed estimators are robust and consistent.

Certainly, it can be argued that experimentation with the form of the error process can occur, for instance, in the lines of Box and Jenkins [1971], but there is no doubt that this is going to be a harder and more expensive alternative, and what is worse, at the end, the significance of the structural estimators will be dependent on the previous tests on the error process. The literature on preliminary test or sequential estimators (see Bock et. al. [1973] and references there) is mainly concerned with preliminary tests on exact or stochastic linear restrictions (for the last case see Judge et. al. [1973]) on the structural parameters, but it is obvious that the same problem occurs when we test the parameters of the error process, even when for that case, as far as is known, no type of sequential estimator has been proposed.

The application of spectral estimation will be specially important for monthly models, since for monthly series Nerlove [1972] has challenged the view that they can be characterized

by a first order autoregressive process with root near unity (see Orcutt [1948] and Granger [1966]). He has shown that at least for some USA Price series a higher order representation is necessary, possibly involving some type of autoregressive moving average (ARMA) process.

Spectral methods of estimation are, indeed, very much connected with the name of E.J. Hannan and in many instances they have been presented in a way more familiar to statisticians than to econometricians. In order to introduce this estimation procedure in a way closer to the classical methods we state in section III how a block Toeplitz covariance matrix can be block diagonalized by means of a unitary matrix so that we can present in a more familiar context the advantages of the spectral approach. In preparation for this, section II presents the terminology and basic notations. Subsequently in section IV we derive the spectral estimators for a simultaneous model and in sections VI and VII we extend the results to models with lagged endogenous variables.

In this study we derive the spectral estimators beginning from the likelihood of the sample. The advantages in proceeding in this way are that the likelihood contains all the sample information and under suitable hypothesis on the stochastic variables the parameter values that maximise the likelihood are consistent and efficient estimators. An alternative approach is to seek simpler estimators by applying some kind of iterative Generalized Least Squares (GLS for short) method that applies the Aitken's formula using some consistent estimate of the covariance matrix. In many instances these estimators have the same asymptotic distribution as the maximum likelihood estimators, but this is not always necessarily the case (see for instance Maddala [1971] and Amemiya [1973-b]) then we loose efficiency in return for something that is supposed to be computationally cheaper. But it is going to be difficult to know how much efficiency we are loosing and therefore it is safer if we develop first the maximum likelihood estimators and try later to approximate them in a way that is cheaper in computer time but does not carry any loss in efficiency.

II Notation

We use small letters to represent scalars and small letters underlining by \sim for column vectors; the transposition sign (') is used to denote a row vector. Matrices will be represented by capital letters. Exceptions to these rules are in the representation of the periodogram and cross-periodogram (see below).

The Kronecker product between two matrices A and B is defined as

$$A \otimes B = (a_{ij} B) \quad (2.1)$$

Vec is an operator that transforms a matrix into the vector obtained by putting the columns of the matrix one below the other. Sometimes we also represent the vec of a matrix A by A^V .

We use \bar{A} for the complex conjugate and A^* for the transpose of the complex conjugate of the matrix A.

The Unitary Matrix Q :

$$Q = \left\{ \frac{1}{\sqrt{T}} \exp \left[\frac{2\pi i}{T} jt \right] \right\} \quad j, t = 0, 1, \dots, T-1 \quad (2.2)$$

is a unitary $T \times T$ matrix. Using

$$\omega_j = \frac{2\pi j}{T} \quad j = 0, 1, \dots, T-1 \quad (2.3)$$

we see that the rows of Q can be written as follows:

$$q'(\omega_j) = \left\{ \frac{1}{\sqrt{T}} \exp(i\omega_j \cdot 0) \frac{1}{\sqrt{T}} \exp(i\omega_j \cdot 1) \dots \frac{1}{\sqrt{T}} \exp[i\omega_j \cdot (T-1)] \right\} \quad (2.4)$$

We call $Q(\lambda_h)$ the $(m_h \times T)$ matrix obtained by taking m_h consecutive rows of Q. Then $\sqrt{T} Q(\lambda_h)$ is an alternant and

$$\Psi(\lambda_h) = Q^*(\lambda_h) Q(\lambda_h) = \frac{1}{T} \left[\begin{array}{c} \Sigma \\ \omega_j \in \lambda_h \end{array} \exp\{-i\omega_j(\ell-k)\} \right]_{\ell, k} \quad (2.5)$$

is a $(T \times T)$ circulant.

The matrix Q has the following properties:

$$Q = Q' \quad , \quad \bar{Q} = Q^* \quad , \quad QQ^* = I_T \quad , \quad (2.6)$$

where I_T is the $T \times T$ identity matrix.

When we apply Q to a system, S , we say that QS is a unitary transform of S . Note that because the latent roots of the unitary matrices are of modulus one (see Macduffee [1956] page 28), the modulus of the Jacobian of the transformation is one.

Fourier Transforms (FT):

Suppose $\underline{x} = [x_t]$, ($t = 0, \dots, T-1$) is a vector of T observations on a scalar variable x , then the FT at frequency ω_j is given by:

$$w_x(\omega_j) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} x_t \exp(i\omega_j t) = \underline{q}'(\omega_j) \underline{x} \quad . \quad (2.7)$$

If we note that $\exp(i\omega_j) = \overline{\exp(i\omega_{T-j})}$, and therefore

$$\bar{q}'(\omega_j) = \underline{q}'(\omega_{T-j}) \quad , \quad (2.8)$$

we see that for a real variable x , it is immaterial whether the FT is defined for ω_j ($j = 0, \dots, T-1$) or for ω_ℓ ($\ell = -(T/2) + 1, \dots, T/2$); only $\frac{T}{2} + 1$ of the FT's are independent, the other being the complex conjugate of these.

If we group the ω_j 's ($0 \leq j \leq T-1$) into $2M$ sets, the h^{th} of which denoted S_h , consists of m adjacent ω_j 's ($T = 2Mm$) clustered around

$$\lambda_h = 0 + \frac{\pi h}{M} \quad , \quad h = 0, \dots, 2M - 1 \quad (2.9)$$

then

$$\underline{w}_x(\lambda_h) = \{w_x(\omega_j)\} = Q(\lambda_h) \underline{x}^{(1)} \quad , \quad (2.10)$$

(1) For the cluster around $h = 0$ note that $w_x(\omega_{-j}) = w_x(\omega_{T-j})$, therefore $Q(\lambda_0)$ is obtained taking the $(m-1)/2$ last rows and the $(m+1)/2$ first rows of Q . Also note that for simplicity in the exposition we take m odd.

is a vector of m FT's such that $\omega_j \in S_h$. And

$$\underline{w}_x = \{w_x(\omega_j)\} = Q\underline{x} \quad , \quad (2.11)$$

is the vector of T Fourier transforms.

Suppose X is a $T \times k$ matrix of T observations of k variables (x_1, x_2, \dots, x_k) then

$$\underline{w}_{x_\ell} = \{w_{x_\ell}(\omega_j)\} \quad , \quad (2.12)$$

is a T vector, containing the T Fourier transforms of x_ℓ .

$$\underline{w}'_X(\omega_j) = \underline{q}'(\omega_j)X = [w_{x_1}(\omega_j) \ w_{x_2}(\omega_j) \ \dots \ w_{x_k}(\omega_j)] \quad , \quad (2.13)$$

is a k vector containing the FT at frequency ω_j for the variables x_1, x_2, \dots, x_k .

$$W_X = QX \quad , \quad (2.14)$$

is the $T \times k$ matrix of all FT of X , and has $\underline{w}'_X(\omega_j)$ in the ω_j th row. The submatrix corresponding to the set S_h is given by

$$W_X(\lambda_h) = Q(\lambda_h)X \quad . \quad (2.15)$$

In spectral estimation W_X is the analog of X in the time domain.

Periodogram and Crossperiodogram :

The periodogram of \underline{x} at frequency ω_j is defined by:

$$I_X(\omega_j) = \frac{1}{2\pi T} \left| \sum_t x_t \exp(i\omega_j \cdot t) \right|^2 = \frac{1}{2\pi} \bar{w}_X(\omega_j) w_X(\omega_j) \quad , \quad (2.16)$$

and is a scalar.

The matrix of periodograms and crossperiodograms for X is defined by

$$I_X(\omega_j) = \frac{1}{2\pi} \bar{w}_X(\omega_j) \underline{w}'_X(\omega_j) = \frac{1}{2\pi} X' \underline{q}(\omega_j) \underline{q}'(\omega_j)X \quad . \quad (2.17)$$

If $\underline{z} = [z_t]$, ($t = 0, \dots, T-1$) is a vector of T observations on a scalar variable z , then the vector of crossperiodograms

between z and X is

$$I'_{zX}(\omega_j) = \frac{1}{2\pi} \bar{w}_z(\omega_j) w'_X(\omega_j) = \frac{1}{2\pi} z' \bar{q}(\omega_j) q'(\omega_j) X. \quad (2.18)$$

We have kept the standard practice of using I for periodogram and crossperiodograms, but there will not be any danger of confusion with the identity matrix since the latter is subscripted by a constant such as T and the former by variables. Note that the $(k \times k)$ matrix $I_X(\omega_j)$ is a quadratic on X generated by a circulant $(\bar{q}(\omega_j) q'(\omega_j))$ of order one.

Average periodogram and crossperiodogram:

The average periodogram of x at the frequency band λ_h is defined by the scalar:

$$\begin{aligned} I_X(\lambda_h) &= \frac{1}{m_h} \sum_{\omega_j \in S_h} I_X(\omega_j) = \frac{1}{2\pi m_h} \bar{w}'_X(\lambda_h) w_X(\lambda_h) \\ &= \frac{1}{2\pi m_h} x' \Psi(\lambda_h) x, \end{aligned} \quad (2.19)$$

and the matrix of average periodograms and crossperiodograms for X is defined by:

$$\begin{aligned} I_X(\lambda_h) &= \frac{1}{2\pi m_h} W_X^*(\lambda_h) W_X(\lambda_h) = \frac{1}{2\pi m_h} X' Q^*(\lambda_h) Q(\lambda_h) X \\ &= \frac{1}{2\pi m_h} X' \Psi(\lambda_h) X. \end{aligned} \quad (2.20)$$

$I_X(\lambda_h)$ here is the analog of $X'X$ in the time domain, but now we have $2M$ $I_X(\lambda_h)$ matrices, but only $M + 1$ of them are independent. Note that if we cluster the $T w_X(\omega_j)$'s in just one group then the average periodogram is

$$I_X = \frac{1}{2\pi} \frac{1}{T} X' Q^* Q X = \frac{1}{2\pi} \frac{X' X}{T}, \quad (2.21)$$

i.e. $(2\pi)^{-1}$ times the sample covariance matrix. The representation of the matrix of average periodograms and crossperiodograms as in (2.20) is quite useful to derive the asymptotic distribution of $I_X(\lambda_h)$, (see Wahba [1968]) because

$\Psi(\lambda_h)$ is a circulant, idempotent, Hermitian matrix of rank m_h and its unique generalized inverse is itself, as with the idempotent symmetric matrices in the real field. It is shown in Wahba [1968], (see also Goodman [1963]) that M average periodogram matrices based on a sample realization of a multivariate Gaussian process, calculated at M appropriately spaced frequencies converge jointly in mean square to M independent complex Wishart matrices, even though M be large. Also, by making use of appropriate limiting theorems, Hannan [1970] shows that the joint distribution of average periodograms and crossperiodograms at an arbitrary frequency λ_h converge to the complex multivariate normal distribution. We make use of these last results to obtain the asymptotic distribution of the estimated parameters of the econometric models that we consider below.

III Stationary disturbances and asymptotic theory

The disturbances, \underline{u}_t , affecting the model considered in the next sections are assumed to follow a stationary vector process with finite dispersion and that can be represented as

$$\underline{u}_t = \sum_{j=0}^{\infty} A(j) \underline{\varepsilon}_{t-j} \quad , \quad (3.1)$$

and

$$\sum_{j=0}^{\infty} |j| \|A(j)\| < \infty \quad , \quad (3.2)$$

where $\underline{\varepsilon}_{t-j}$ are i.i.d. random vectors of order p , with zero mean and covariance matrix I_p . For $\|A(j)\|$ we mean a suitable matrix norm.

If we have a realization of \underline{u}_t for $t = 1, \dots, T$ the matrix $U = (\underline{u}_t)$ can be vectorized as $\underline{u} = \text{Vec}(U)$, and the covariance matrix for \underline{u} is

$$\text{cov}(\underline{u}) = \begin{pmatrix} G(0) & G(1) & \dots & G(T-1) \\ G(1) & G(0) & & G(T-2) \\ & & & G(1) \\ & & & G(0) \end{pmatrix} = \Omega \quad . \quad (3.3)$$

Each submatrix G is of order p hence Ω is of order pT .

It can be shown that

$$\Phi = (Q^* \otimes I_p) \Omega (Q \otimes I_p) + 2\pi F_U = 2\pi \begin{pmatrix} F_U(\omega_0) \\ \vdots \\ F_U(\omega_{T-1}) \end{pmatrix} \quad , \quad (3.4)$$

where

$$F_U(\omega_j) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} G(\tau) \{ \exp i \omega_j \tau \} = \left\{ f_{u_l u_h}(\omega_j) \right\} \quad ,$$

and so F_U is a block diagonal matrix with T blocks ($p \times p$) containing the spectra matrices of $\{u_t\}$ at frequencies ω_j .

The equation (3.4) relates Φ to F_U , the true spectra of the process, generally unknown. A consistent estimate of Φ is $2\pi I_U^{(2M)}$ defined as

$$I_U^{(2M)} = \begin{bmatrix} I_U(\lambda_0) & & & \\ & I_U(\lambda_1) & & \\ & & \ddots & \\ & & & I_U(\lambda_{2M-1}) \end{bmatrix} \quad (3.5)$$

where the λ_h 's are obtained by grouping the ω_j 's as in section II. (For the matrices F and I , sometimes a superscript (T) or (2M) is used to indicate the number of blocks in them. This index will be omitted when there is no danger of confusion). Then

$$I_U(\lambda_h) = \frac{1}{m} \sum_{\omega_j \in S_h} I_U(\omega_j) \quad \text{and} \quad (3.6)$$

$$I_U(\omega_j) = \frac{1}{2\pi} \bar{w}_U(\omega_j) w_U'(\omega_j)$$

and $w_U(\omega_j)$ is the ω_j row of QU .

Now, $2\pi I_U$ is asymptotically equivalent to Φ in the following sense. We assume that $F_U(\omega_j)$ is constant within each λ_h , ($\omega_j \in S_h$) and letting $T \rightarrow \infty$

$$p \lim I_U^{(2M)} \rightarrow F_U^{(2M)}, \quad (3.7)$$

i.e. the estimated matrix $I_U^{(2M)}$ will tend to the population one provided that the spectra are sufficiently smooth within bands of frequencies. Then if we allow $M \rightarrow \infty$

$$2\pi F_U^{(2M)} \rightarrow 2\pi F_U.$$

For this asymptotic theory we need

$$T \rightarrow \infty \quad (3.8)$$

$$M \rightarrow \infty \quad (3.9)$$

$$M/T \rightarrow 0 \quad (3.10)$$

and (3.10) is required for (3.7) to be true, because $\text{Var} [I(\lambda_h)] \sim \frac{M}{T} F(\lambda_h)$. In practice some compromise is needed in the selection of M , because the variance of $I_U(\lambda_h)$ is of $O(M/T)$ but the bias is of $O(M^{-q})$ where q is the exponent of (j) in (3.2), (see Hannan [1970] Chapter 5), i.e. we can only reduce the variance of $I_U(\lambda_h)$ (by using a bigger m) at the expense of the bias⁽²⁾. To ensure as small an order of magnitude as is possible for the mean squared error we choose

$$M = O(T^{1/(1+2q)})$$

In the present case that implies replacing (3.10) by $M/T^{1/2} \rightarrow 0$, which is also required for the asymptotic theory of the next sections.

In the following sections we present a method of obtaining efficient estimators for the structural parameters of an econometric model proceeding from the likelihood function but replacing Ω by their spectral diagonalization. The efficiency property arises in the following sense. If the error term (U) was normally distributed we could write the likelihood of U and obtain the parameter values that maximise it. We know that these maximum likelihood estimators for more restrictive hypotheses on U than that expressed in (3.1) give consistent and efficient estimators (see Kendall and Stuart [1967]). We can see, in a heuristic way, how a similar maximum likelihood theory could apply in our case by using the following double limiting procedure. First we assume that F_U (the spectra of

(2) Engel and Gardner [1976] have noted that in regression, both bias and variance of the spectra affect the mean square error and so it could be possible to find an optimum, although presumably it would be specific to the situation.

the error terms) can be divided in $2M$ bands between $[-\pi, \pi]$ such that for the h^{th} band, which we called S_h ,

$$f_{\underline{u}}(\omega) = \text{constant} \quad \text{for } \omega \in S_h .$$

With this assumption we first let T (the number of observations) tend to ∞ then we allow M to tend to ∞ . So when T tends to infinity the number of parameters in the likelihood is finite but since we also allow M to tend to infinity finally the $F_{\underline{u}}$ matrix converges to the diagonal form of Ω . So only in the limit will the likelihood contain all the information on the \underline{u} process, i.e. in the limiting process we not only increase the number of observations but also the information on \underline{u} by increasing M . This limiting process is similar to the one involved in the proof of the efficiency of two stage least squares by using principal components on the exogenous variables, where we not only increase T but also the number of principal components (see Dhrymes [1970]).

We can see now that, for the fixed regressor case, with this double limiting procedure, we have that for large T and M the spectral maximum likelihood estimates for the case with stationary errors will be equivalent to the any (time or frequency domain) likelihood estimates using an ARMA process for \underline{u} when it exists. In these cases only in the limit will the spectral estimates, presented below, contain all the information on \underline{u} used by the other methods. But if the finite parametrization of \underline{u} is wrong or if it does not exist, the only efficient method is the spectral approach. Thus it seems that the spectral method is a general robust method of estimation with no asymptotic loss in efficiency in relation to other particular methods where their application is justified. What order of magnitude we need for T and M is another question that could be explored by Monte Carlo studies.

The generality of the hypothesis (3.1) makes the investigation on models with stationary errors quite valuable at a time when estimation methods for particular AR, MA or ARMA process are proliferating. Quite a complete bibliography can be found in

Hendry [1976] and Nicholls et. al. [1975] for the AR and MA cases respectively.

The covariance sequence of a stationary process with continuous spectral density can be approximated arbitrarily close by the covariance of an AR or MA process (see Anderson [1971] pp 411), therefore the analogue of the spectral method of this study in the time domain, will be a procedure that increases with T the number of parameters used in defining the U process. (See Amemiya [1973a]).

In deriving the asymptotic properties of the spectral maximum likelihood estimates we can drop the normality assumption on \underline{u} and still be able to obtain the values of the structural parameters that maximise the same function (no longer a likelihood function) and obtain their asymptotic distribution by the use of some central limit theorems. If this distribution is the same as that which we obtain under the Gaussian hypothesis we say that our estimates are efficient. We will proceed in this way using theorems due to Hannan [1970] . For this asymptotic theory, on dropping the normality requirement, we need to assume that the distribution of $\underline{\varepsilon}$ has finite moments up to the fourth order and that the \underline{u} process is stationary to the same order.

IV Specfiml (Spectral Full Information Maximum Likelihood)
Estimation.

IV.1 The model and its likelihood function.

The spectral estimation for the simultaneous model has been done by Hannan and Terrell [1973]. In this section we propose an alternative derivation of these estimators from the likelihood of the sample and we show that from the Specfiml formulae we can derive those corresponding to different time domain procedures.

We will consider the equation system given by:

$$\begin{aligned} BY' + \Gamma Z' &= AX' = U' \\ A &= (B \ : \ \Gamma) \\ X &= (Y \ : \ Z) \end{aligned} \tag{4.1}$$

where,

B is $p \times p$ matrix of unknown coefficients
Y is $T \times p$ matrix of observations of the
endogenous variables
 Γ is $p \times k$ matrix of unknown coefficients
Z is $T \times k$ matrix of observations of the exogenous
variables and
U is $T \times p$ matrix of unobserved disturbance terms.

We assume that model is identified by zero restrictions and one normalizing restriction per equation and there are no identities.⁽³⁾

We require that $U = (u_t)$ satisfies conditions (3.1) and (3.2) but in this section we use $A(0) = I$ and G, a $(p \times p)$ positive definite matrix, for the covariance of ε_t . We have then that

(3) The last condition is assumed for the ease of exposition, but the derivations that follow can be modified to include systems containing identities.

the spectral density matrix of \underline{u}_t is of the form

$$F_U(\lambda) = \frac{1}{2\pi} \left(\sum_j A(j)e^{-ij\lambda} \right) G \left(\sum_j A(j)e^{-ij\lambda} \right)^* ,$$

and we require that

$$\det F_U(\lambda) > 0 , \quad \lambda \in [-\pi, \pi] .$$

In this section we restrict ourselves to the case where the z_j ($j = 1, \dots, k$) variables are exogenous, so that they are totally independent of \underline{u} . The theorems that we will present are valid only asymptotically and we will assume that the following limits exist:

$$G_Z(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \underline{z}_t \underline{z}'_{t+\tau} \quad (\tau = 0, 1, 2, \dots)$$

then

$$G_Z(\tau) = \int_{-\pi}^{\pi} e^{-ik\lambda} dM(\lambda) , \quad (4.2)$$

where $M(\lambda)$ is a matrix-valued function whose increments are Hermitian non-negative definite. To simplify notation we will assume that the derivative of $M(\lambda)$ exists, denoted by $F(\lambda)$, but this is not necessary for the existence of the theorems presented below - on replacing $F(\lambda)$ by $dM(\lambda)$ these still hold. We will also require

$$G_Z(0) \quad \text{to be non-singular.}$$

These conditions on \underline{z}_t can be relaxed to the simpler requirements of what has been called Grenander's conditions, (see Hannan and Terrell [1973] page 313).

If we apply to (4.1) the operation of vectorizing by columns and using

$$\text{Vec}(ABC) = (ABC)^V = (C' \otimes A)\text{Vec } B ,$$

(4.1) can be expressed as

$$(Y \otimes I_p)\text{Vec } B + (Z \otimes I_p)\text{Vec } \Gamma = \text{Vec } (U') . \quad (4.3)$$

Using

$$\text{Cov \{Vec(U')\}} = \begin{pmatrix} G(0) & G(1) & \dots & G(T-1) \\ G'(1) & G(0) & & \\ & & \ddots & \\ G'(T-1) & \dots & G'(1) & G(0) \end{pmatrix} = \Omega = \Omega_U, \quad (4.4)$$

the likelihood of $\text{Vec}(U')$ can be written as

$$\mathcal{L}(\text{Vec}(U')) = (2\pi)^{-pT/2} |\Omega_U|^{-1/2} \{ \exp -\frac{1}{2} \text{Vec}(U')' \Omega_U^{-1} \text{Vec}(U') \}. \quad (4.5)$$

In this case Ω_U contains $p^2 \times T$ parameters (of course they can not be estimated with only pT data points; we will deal with this below). In the white noise case Ω_U is block diagonal with the same matrix in each block of the main diagonal (i.e. we deal with only p^2 parameters). In the first order AR case Ω_U depends only on $2 \times p^2$ parameters⁽⁴⁾. Thus (4.5) is the general form representing the probability law of U'^V if the covariance matrix of $\text{Vec } U'$ has a block Toeplitz form. And the different time domain estimation procedures consist in postulating particular probability laws for u_t in order to reduce the $p^2 \times T$ parameters of Ω_U to a manageable number considerably less than pT .

In the Specfiml estimation instead of using Ω^{-1} we work with $(2\pi)^{-1}(Q \otimes I_p) F_U^{-1} (Q^* \otimes I_p)$ and we assume that the spectra of $\text{Vec}(U')$ can be divided in $2M$ frequency bands in the interval $(0, 2\pi)$ such that they are constant in each band. This assumption allows us to work with $F_U^{(2M)}$ instead of $F_U^{(T)}$, i.e. we can reduce the $p^2 \times T$ parameters in F_U to $p^2 \times 2M$ which makes the method feasible for certain values of p and k . For efficiency we need that in the limiting process both T and M tend to infinity, because if M does not, $2\pi F_U^{(2M)}$ will not be equivalent to Φ , as defined in (3.4).

(4) Note that in both cases not all the parameters are different. In fact we have $p(p+1)/2$ different parameters in the white noise case and $p^2 + p(p+1)/2$ in the first order AR case.

We saw that as $T \rightarrow \infty$ (3.4) is true and from it

$$(Q^* \otimes I_p) \Omega^{-1} (Q \otimes I_p) \rightarrow \frac{1}{2\pi} F_U^{-1} . \quad (4.6)$$

Using (4.6) the exponential term in (4.5) can be expressed as follows:

$$\exp\left\{-\frac{1}{4\pi} (\text{Vec } U')'(Q \otimes I_p) F_U^{-1} (Q^* \otimes I_p)(\text{Vec } U')\right\} . \quad (4.7)$$

In (4.7)

$$(Q^* \otimes I_p) \text{Vec}(U') = \text{Vec}(I_p U' Q^*) = \text{Vec}(U' Q^*) = \text{Vec } W_U^* \quad (4.8)$$

and

$$(\text{Vec } U')'(Q \otimes I_p) = [\text{Vec}(U'Q)]' = [\text{Vec}(W_U^*)]' ,$$

therefore (4.7) can be expressed

$$\exp\left\{-\frac{1}{4\pi} [\text{Vec}(W_U^*)]' F_U^{-1} \text{Vec}(W_U^*)\right\} \quad (4.9)$$

and since F_U^{-1} is the block diagonal with $F_U^{-1}(\omega_j)$ in the j^{th} block (4.9) can be written as

$$\exp\left\{-\frac{1}{2} \frac{1}{2\pi} \sum_{j=0}^{T-1} \tilde{w}_U'(\omega_j) F_U^{-1}(\omega_j) \tilde{w}_U(\omega_j)\right\} ,$$

where $\tilde{w}_U'(\omega_j)$ is the ω_j row of W_U or as

$$\exp\left\{-\frac{1}{2} \sum_j \text{tr } F_U^{-1}(\omega_j) I_U(\omega_j)\right\} , \quad (4.10)$$

where $I_U(\omega_j)$ is the cross periodogram matrix of U at ω_j .

Using (4.10), (4.5) can be written as:

$$\mathcal{L}(U^V) = (2\pi)^{-pT/2} |\Omega_U|^{-1/2} \exp\left\{-\frac{1}{2} \sum_{\omega_j} \text{tr } F_U^{-1}(\omega_j) I_U(\omega_j)\right\}. \quad (4.11)$$

Note that (4.11) is simpler than (4.5) since the latter contains the inverse of Ω (a $pT \times pT$ matrix) and the former only contains $T/2 + 1$ matrices F_U^{-1} of order $p \times p$. To obtain the likelihood of Y^V we note that the Jacobian is:

$$J = \left| \frac{\partial (U'V)_\ell}{\partial (Y'V)_\ell} \right| = |B^\circ| = T|B| \quad (4.12)$$

where B° is a block diagonal matrix with B in each non-empty block.

Now, since we assume that $F_U(\omega_j)$ is constant in each S_h , we can sum over a range of frequencies and this allows substitution of $I_U(\omega_j)$ by $I_U(\lambda_h)$ for the frequency band $S_h(\omega_j \in S_h)$. We have

$$T = 2Mm \quad (4.13)$$

where $2M$ is the number of frequency bands we use in the interval 0 to 2π and m is the number of periodogram ordinates used in each average. Note that with this summation the diagonalization of Ω_U does not contain T different blocks but only $2M$, and this is the crucial assumption in this estimation method to reduce the number of unknowns in Ω_U .

Making use of the above we can write

$$\begin{aligned} \mathcal{L}(Y'V) &= (2\pi)^{-pT/2} |2\pi(Q^* \otimes I_p) F_U(Q \otimes I_p)|^{-1/2m} T|B| \\ &\exp \left\{ -\frac{m}{2} \sum_{\lambda_h} \text{tr} F_U^{-1}(\lambda_h) I_U(\lambda_h) \right\}. \end{aligned} \quad (4.14)$$

From now onwards F_U refers to a $2M$ block diagonal matrix with $F_U(\lambda_h)$, ($F_U(\lambda_h) = F_U(\omega_j)$ for $\omega_j \in S_h$) in each block. This is the reason for the factor m appearing in the exponent of the first determinant. The factor m appears in the exponential term because $I_U(\lambda_h)$ has been defined as an average.

Equation (4.14) is, apart from the Jacobian and for large T , the likelihood of $\text{Vec}(U'Q)$,

$$(Q \otimes I_p) \text{Vec } U' = \text{Vec}(U'Q) = (W_Y \otimes I_p) \text{Vec } B + (W_Z \otimes I_p) \text{Vec } \Gamma, \quad (4.15)$$

and it will be interesting to consider the model

$$\text{Vec}(U'Q) = (W_Y \otimes I_p) \text{Vec } B_h + (W_Z \otimes I_p) \text{Vec } \Gamma_h, \quad (4.16)$$

where now the matrices B and Γ can be different for different frequency bands, to test the hypothesis that

$$B_h = B$$

and

$$\Gamma_h = \Gamma \quad \forall h$$

as a way to ensure that (4.1) is well specified. In practice we will only be able to deal with very low values for 2M.

Working with the log-likelihood we have:

$$L_1(Y^V) = \frac{L(Y^V)}{T} = -p \log(2\pi) - \frac{m}{2T} \sum_{\lambda_h} \log \det F_U(\lambda_h) - \frac{m}{2T} \sum_{\lambda_h} \text{tr} F_U^{-1}(\lambda_h) I_U(\lambda_h) + \log \det B \quad (4.17)$$

and

$$\frac{\partial L_1(Y^V)}{\partial F_U(\lambda_h)} = -\frac{m}{2T} F_U^{-1}(\lambda_h) + \frac{m}{2T} F_U^{-1}(\lambda_h) I_U(\lambda_h) F_U^{-1}(\lambda_h)$$

for each λ_h , $h = 0, \dots, 2M-1$. Therefore

$$\begin{aligned} \hat{F}_U(\lambda_h) &= I_U(\lambda_h) = A I_{XX}(\lambda_h) A' = A X' \Psi(\lambda_h) X A' \\ &= B I_{YY}(\lambda_h) B' + \Gamma I_{ZZ}(\lambda_h) \Gamma' + B I_{YZ}(\lambda_h) \Gamma' + \Gamma I_{ZY}(\lambda_h) B' \end{aligned} \quad (5) \quad (4.18)$$

and concentrating L_1 with respect to $F_U(\lambda_h)$ we get

$$\begin{aligned} L_2 &= c + \log \det B - \frac{1}{2 \cdot 2M} \sum_{\lambda_h} \log \det I_U(\lambda_h) \\ &= c + \log \det B - \frac{1}{2 \cdot 2M} \sum_{\lambda_h} \log \det A X' \Psi(\lambda_h) X A' \end{aligned} \quad (4.19)$$

Now since the frequency bands are symmetric we have

$$L_2 = c + \log \det B - \frac{1}{2M} \sum_0^M \delta \log \det I_U(\lambda_h) \quad (4.20)$$

where the summation is between 0 and π , and the factor δ means that frequency bands around zero and π that appear only once must be multiplied by $\frac{1}{2}$.

(5) From now on and unless otherwise specified, we define $\Psi(\lambda_h)$ and the FT's including the factors $1/2\pi m_h$ and $1/\sqrt{2\pi m_h}$ respectively.

The Specfiml estimates of the structural parameters are those values of B and Γ that maximize L_2 . But to see more clearly what is involved in the maximization of L_2 , it is convenient to note that for a error vector process of the form we are considering here, Rozanov [1967] section 2.6, shows that

$$\log \det G = p \log (2\pi) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \det F_U(\lambda) d\lambda . \quad (4.21)$$

A finite approximation to it is

$$\log \det \hat{G} = p \log (2\pi) + \frac{1}{2M} \sum_{-M+1}^M \log \det I_U(\lambda_h) \quad (4.22)$$

therefore the maximization of the likelihood, L_2 , amounts to minimizing a finite approximation to the generalized variance of the regular process $\underline{\varepsilon}_t^{(6)}$ underlying the disturbance term \underline{u}_t , subject to the normalization rules introduced by the Jacobian. Det G is also the generalized variance of the error in the prediction of \underline{u}_t by the best linear unbiased (b.l.u.) one step ahead predictor, see Rozanov [1967], therefore in obtaining the Specfiml estimates of A we replace \underline{u}_t by some feasible b.l.u. predictor.

In fact L_2 is related to the likelihood of the regular process in the following way. For large T, since the dependence of \underline{u}_t on $\underline{\varepsilon}_{t-s}$ tends to zero at $s \rightarrow \infty$, we can consider the $\underline{\varepsilon}_t$'s for the same sample range as the \underline{u}_t 's by just ignoring the end effects. Therefore we could write

$$\frac{L(E^V)}{T} = - \frac{p}{2} \log (2\pi) - \frac{1}{2} \log \det G - \frac{1}{2T} \text{tr } G^{-1} E^V E. \quad (4.22a)$$

where $E = (\underline{\varepsilon}_t)$. Now

$$\frac{1}{T} \text{tr } G^{-1} E^V E \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr}[I_U(\lambda) F_U^{-1}(\lambda)] d\lambda \quad (4.22b)$$

(see Whittle [1953] theorem 5). So using (4.21) and (4.22b) we can put $\{L(E^V)/T\}$ in terms of the u's as

(6) Note that in the vector case the regular process is defined up to multiplication by an arbitrary unitary matrix, but since the determinant of a unitary matrix is unity, the generalised variance is the same in all cases.

$$-p \log(2\pi) - \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \det F_U(\lambda) d\lambda - \frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr} [F_U^{-1}(\lambda) I_U(\lambda)] d\lambda. \quad (4.22c)$$

The Jacobian of the transformation from the $\underline{\xi}$'s to the \underline{u} 's is unity, therefore (4.22c) can be seen as the log likelihood of U^V/T . In fact (4.22c) is a general expression for different hypotheses on \underline{u}_t , in particular for the white noise and AR cases. In those instances (4.22c) or (4.22a) are valid for any $T^{(7)}$. What we have done in (4.17) is to replace the integrals by summations and λ by λ_h , as a result of the assumption that $F_U(\lambda)$ is constant within each S_h band.

If we assume that F_U is constant at all frequency bands and therefore we obtain the cumulative periodogram matrix by summing over all ω_j 's, the concentrated likelihood then is

$$\begin{aligned} L_f &= k + \log \det B - \frac{1}{2} \log \det \frac{1}{T} W_U^* W_U \quad (8) \\ &= k + \log \det B - \frac{1}{2} \log \det \frac{1}{T} U' Q^* Q U \\ &= k + \log \det B - \frac{1}{2} \log \det A \left[\frac{X' X}{T} \right] A' , \quad (4.23) \end{aligned}$$

which is the concentrated likelihood used in time domain estimation under the white noise hypothesis (FIML estimation). Therefore an appealing way to see if the difference between L_2 and L_f is significant is by comparing

$$\eta_f = 2T(L_2 - L_f) , \quad (4.24)$$

with the corresponding $\chi^2(\ell)$ value at certain significance level and reject L_f (white noise) if η_f exceeds that value. The number of degrees of freedom, ℓ , can be taken as the difference between the number of parameter used to estimate Ω under L_2 , $[(M - 1) p^2 + p(p + 1)]$, and the number used under

(7) In the A.R. case, assuming that the initial observations are known to be zero.

(8) The W_U 's matrices here are defined without including the factor $\frac{1}{\sqrt{2\pi m_h}}$.

L_f , $[p(p + 1)/2]$, i.e.,

$$l = (M - 1) p + \frac{p(p + 1)}{2} . \quad (4.25)$$

Similarly L_2 can be compared with the concentrated likelihood, L_0 , under the hypothesis that \tilde{u}_t follows a vector AR, MA or ARMA process. But in these cases L_2 and L_0 are nested only for large T and M and therefore for small samples η_0 , ($\eta_0 = 2T(L_2 - L_0)$), can be negative. In the small sample case it may also be necessary to correct the likelihood ratio tests for degrees of freedom.⁽⁹⁾

Of the parametric processes mentioned, the AR one is the most easy to manipulate because we can apply an autoregressive transformation (the one postulated for the errors) to the model and have it then expressed in terms of a white noise vector process such that the standard FIML procedure (see Koopmans and Hood [1953]) can be applied to it. (We call ARFIML the estimation procedure for that case). So, if we consider the first order process

$$\tilde{u}_t = R \tilde{u}_{t-1} + \varepsilon_t , \quad \text{Var}(\varepsilon_t) = G \quad (4.26)$$

or in matrix notation

$$U' = RU'_1 + E' ,$$

where R is the matrix of autoregressive parameters, and is unrestricted and we have used the subscript 1 to denote the lagged value of the corresponding matrix, the transformation applied to (4.1) is $(I - RL)$, where L is the lag operator. For that case T^{-1} times the log likelihood after concentrating respect with G and R is (see Sargan [1961] and also Hendry [1970] chapter 3)

$$\begin{aligned} L_{a_1} &= k_a + T \log \det B - \frac{1}{2} T \log \det \hat{E}'\hat{E} \\ &= k_a + T \log \det B - \frac{1}{2} T \log \det (AX'D(1)XA') , \end{aligned} \quad (4.27)$$

(9) If the above tests are performed using the kernels of L_f , L_0 and L_2 we must be sure that we are disregarding the same constant term in each hypothesis.

where

$$\hat{E}' = U' - \hat{R}U_1' \quad , \quad (4.28)$$

$$\hat{R} = U'U_1(U_1'U_1)^{-1} \quad (4.29)$$

and

$$D(1) = [I - X_1A'(AX_1'X_1A')^{-1} AX_1'] \quad . \quad (4.30)$$

In (4.27) the kernel of the generalized variance of the regular process is approximated by $AX'D(1)XA'$, where $D(1)$ is an idempotent matrix. L_{a_1} can, of course, be generalized for the r^{th} order error process by substituting $D(1)$ by $D(r)$ in (4.27) and defining $D(r)$ as:

$$D(r) = [I - U^+(U^{+'}U^+)^{-1} U^{+'}] \quad , \quad (4.31)$$

where

$$U^+ = (X_1A' : X_2A' \dots : X_rA') \quad .$$

Now the generality of the formulation in (4.19) is clear. In Specfiml \hat{G} is defined by means of the unitary idempotent matrices $\Psi(\lambda_h)$ and FIML and ARFIML are special cases where $\Psi(\lambda_h)$ is assumed constant over the frequency bands and is taken as the unity or $D(r)$ matrix respectively.

In the invertible MA case we need an infinite number of lags to express E' in terms of U_1' and the analysis is more cumbersome, but one possible way to proceed is by taking only a finite number of terms.

In defining L_2 we have used $2M$ bands, but we could use a different number $2M^+$ and test if the difference in the likelihood is significant. But also in this case both hypothesis are not necessarily nested for finite M and the ratio test can be negative.

So far we have assumed the same m for the $2M$ bands, but in practice it will be more convenient to use a different m_h for each band in which case we have

$$L_2 = c + \log |\det B| - \frac{1}{2M} \sum_0^M \delta \frac{m_h}{m} \log \det I_U(\lambda_h) \quad . \quad (4.32)$$

Using (4.19)

$$L_2 = c + \log |\det B| - \frac{1}{2M} \sum_0^M \delta \frac{m_h}{m} \log \det A I_{XX}(\lambda_h) A'. \quad (4.33)$$

IV.2 The first order conditions.

The values of A for which L_2 attains its global maximum will be the set of efficient estimates of A under the general assumption of a stationary vector error process. There are several ways to obtain these estimates.

The one that we have followed in constructing a Fortran program for that purpose directly minimizes $-L_2$ by means of the Powell [1964] subroutine that calculates the minimum of a function without using derivatives. Another way would be to obtain the first order conditions for the maximum of L_2 and solve the resulting equations for the unrestricted values of A. Of course, these equations are non-linear and some sort of iterative procedure is still required if we want a coherent solution for all the parameters (structural and spectral) involved. Nevertheless, we will see below that provided we sacrifice the internal coherency between both sets of parameters, if the predetermined variables are non-stochastic and we work with large samples, then the first derivative equations can be linearized and no iteration is needed for asymptotically equivalent estimators. As mentioned this was not the way in which the numerical calculations were done in a Fortran program constructed by the author, but the first order conditions are needed to derive the asymptotic properties of the estimator.

From (4.17) we have:

$$L_0 = k_0 - \frac{m}{2} \sum_{\lambda_h} \log \det F_U(\lambda_h) - \frac{m}{2} \sum_{\lambda_h} \text{tr} F_U^{-1}(\lambda_h) A I_{XX}(\lambda_h) A' + T \log \det B. \quad (4.34)$$

Throughout this study a superscript u attached to (i) a Vec means that only those elements which correspond to unconstrained coefficients in Vec A are to be considered, (ii) a square matrix means that only those rows and columns which correspond to unconstrained elements of Vec A are to be retained; the other

rows and columns are deleted. Let us call

$$\phi = (\text{Vec } A)^u \quad (4.35)$$

where ϕ is a $(n \times 1)$ vector, n being the number of unrestricted coefficients in A . Now

$$\frac{\partial L_0}{\partial F_U(\lambda_h)} = 0 \Rightarrow F_U(\lambda_h) = AI_{XX}(\lambda_h)A' \quad (4.36)$$

and

$$\frac{\partial L_0}{\partial \phi} = q = \{T[(B')^{-1} : 0] - m \sum_{\lambda_h} \hat{F}_U^{-1}(\lambda_h) AI_{XX}(\lambda_h)\}^v u = 0. \quad (4.37)$$

So the first order conditions are given by equations (4.36) and (4.37). But following a similar approach to Durbin [1963], (4.36) and (4.37) can be expressed in a more useful form. The equation (4.36) can be rewritten as

$$\frac{1}{2M} \sum_{\lambda_h} F_U^{-1}(\lambda_h) AI_{XX}(\lambda_h)A' = I, \quad (4.38)$$

where

$$\begin{aligned} I_{XX}(\lambda_h) &= \begin{bmatrix} I_{YY}(\lambda_h) & I_{YZ}(\lambda_h) \\ I_{ZY}(\lambda_h) & I_{ZZ}(\lambda_h) \end{bmatrix} \\ &= \sum_{\omega_j} \bar{w}_X(\omega_j) \bar{w}_X'(\omega_j) = \bar{w}_X^*(\lambda_h) \bar{w}_X(\lambda_h). \end{aligned} \quad (4.39)$$

So we have:

$$\frac{1}{2M} \sum_{\lambda_h} [\hat{F}_U^{-1}(\lambda_h) AI_{XY}(\lambda_h)B' + \hat{F}_U^{-1}(\lambda_h) AI_{XZ}(\lambda_h)\Gamma'] = I$$

and postmultiplying by $(B')^{-1}$ we have

$$\frac{1}{2M} \sum_{\lambda_h} [\hat{F}_U^{-1}(\lambda_h) AI_{XY}(\lambda_h) - \hat{F}_U^{-1}(\lambda_h) AI_{XZ}(\lambda_h)\Pi'] = (B')^{-1}, \quad (4.40)$$

where

$$\Pi = -B^{-1}\Gamma. \quad (4.41)$$

Thus

$$(B')^{-1} - \frac{1}{2M} \sum_{\lambda_h} \hat{F}_U^{-1}(\lambda_h) AI_{XY}(\lambda_h) = -\frac{1}{2M} \sum_{\lambda_h} \hat{F}_U^{-1}(\lambda_h) AI_{XZ}(\lambda_h)\Pi'. \quad (4.42)$$

Note that using (4.13), (4.37) can be written as

$$\{[(B')^{-1} : 0] - \frac{1}{2M} \sum_{\lambda_h} \hat{F}_U^{-1}(\lambda_h) A I_{XX}(\lambda_h)\}^V u = 0 \quad , \quad (4.43)$$

and substituting (4.42) into (4.43) we have

$$\begin{aligned} & \{- \frac{1}{2M} \sum_{\lambda_h} \hat{F}_U^{-1}(\lambda_h) \hat{A} \sum_{\omega_j} \bar{w}_X(\omega_j) w'_Z(\omega_j) \hat{P}'\}^V u \\ & = \{- \frac{1}{2M} \sum_{\lambda_h} \hat{F}_U^{-1}(\lambda_h) A I_{XZ}(\lambda_h) \hat{P}'\}^V u = 0 \quad , \quad (4.44) \end{aligned}$$

where

$$P' = (\Pi' : I) \quad . \quad (4.45)$$

The Specfiml estimators of A and $F_U(\lambda_h)$ must satisfy (4.38) and (4.44), subject to (4.45) and (4.41). Note that (4.44) can be written as

$$\{- \frac{1}{2M} \sum_{\lambda_h} \hat{F}_U^{-1}(\lambda_h) I_{UZ}(\lambda_h) \hat{P}'\}^V u = 0 \quad , \quad (4.46)$$

therefore, if the Z matrix only contains exogenous variables, though $\hat{F}_U(\lambda_h)$ and \hat{P}' are functions of A, changes in them resulting from changes in A are asymptotically negligible because

$$p \lim W_U^*(\lambda_h) W_Z(\lambda_h) = 0 \quad . \quad (4.47)$$

The consequence is that consistent estimates of $F_U(\lambda_h)$ and P are enough to obtain efficient estimates for A. We will see below how this applies in the ARFIML case.

In the time domain estimation of simultaneous equation models under the white noise hypothesis a large number of estimators have been proposed, see for instance Zellner and Theil [1962] , Rothenberg and Leenders [1964] , Brundy and Jorgenson [1971] , Dhrynes [1973] , and in a recent paper Hendry [1976] shows that these estimators can be seen as different numerical ways of solving an expression for the Full Information Maximum Likelihood (FIML) estimators. Similarly from our equations (4.44), (4.18) and (4.45) we can define a class of spectral estimators for the simultaneous model with stationary errors,

as follows:

$$\left\{ \frac{1}{2M} \sum_{\lambda_h} F_{U(1)}^{-1}(\lambda_h) A_{(2)} I_{XZ}(\lambda_h) P'_{(3)} \right\}^V u = 0 \quad (i)$$

$$F_{U(1)}(\lambda_h) = I_{U(1)}(\lambda_h) = A_{(1)} I_{XX}(\lambda_h) A'_{(1)} \quad (ii) \quad (4.48)$$

$$P'_{(3)} = (\Pi'_{(3)} : I) \quad (iii)$$

where the subscript numbers mean that different estimators could be used in each equation (in fact the class can be enlarged by using other estimation procedure for spectra then the proposed cumulative periodogram; for a variety of these spectra estimates see Hannan [1970] section 5.4). Any of the estimators defined in (4.48) can be seen as a way of approximating the first order conditions for Specfiml and these estimators will have, in general, different statistical properties for small samples, but provided $F_{U(1)}^{-1}(\lambda_h)$ and $P'_{(3)}$ are consistent estimates, the asymptotic distribution of $A_{(2)}$ will be the same, i.e. there is no need to iterate. In Specfiml we iterate upon (4.44), (4.18) and (4.45) to obtain estimates of A , $F_U(\lambda_h)$ and P that are mutually consistent for any sample size.

The equations (4.48) appear to be a generalization of those equations corresponding to quite a variety of Time Domain procedures. If the errors are white noise and therefore

$$F_U^{-1}(\lambda_h) = 2\pi G^{-1} \quad , \quad \forall \lambda_h \quad (4.49)$$

(4.48(i)) can be written as

$$\begin{aligned} & \{ 2\pi G_{(1)}^{-1} A_{(2)} \left[\sum_{\lambda_h} X' \Psi(\lambda_h) Z \right] P'_{(3)} \}^V u = 0 \\ & = \{ 2\pi G_{(1)}^{-1} A_{(2)} \frac{X' Z}{2\pi} P'_{(3)} \}^V u = 0 \quad . \end{aligned} \quad (4.50)$$

Similarly (4.48(ii)) can be written in terms of ω_j as

$$\begin{aligned} \sum_{\omega_j} F_{U(1)}(\omega_j) &= \frac{T}{2\pi} G_{(1)} = \sum_{\omega_j} I_{U(1)}(\omega_j) = A_{(1)} \left[\sum_{\omega_j} X' \Psi(\omega_j) X \right] A'_{(1)} \\ &= A_{(1)} \frac{X' X}{2\pi} A'_{(1)} \quad , \end{aligned}$$

and so

$$G_{(1)} = A_{(1)} \frac{X'X}{T} A'_{(1)} \quad , \quad (4.51)$$

and (4.50), (4.51) and (4.48(iii)) correspond to the equations given by Hendry [1976] for the FIML case.

For the first order AR case as defined in (4.26) we have that

$$\begin{aligned} F_U^{-1}(\lambda) &= 2\pi(I - R' e^{i\lambda}) G^{-1}(I - Re^{-i\lambda}) \\ &= 2\pi[G^{-1} - R'G^{-1} e^{i\lambda} - G^{-1} Re^{-i\lambda} + e^{i\lambda}R'G^{-1} Re^{-i\lambda}] \quad . \quad (4.52) \end{aligned}$$

Noting

$$I_{XZ}(\omega_j)e^{-i\omega_j} = \bar{w}_X(\omega_j) w'_Z(\omega_j)e^{-i\omega_j} \quad ,$$

we have that

$$I_{XZ}(\omega_j)e^{i\omega_j} = I_{XZ_1}(\omega_j) + o(1/\sqrt{T}) \quad , \quad (4.53)$$

$$I_{XZ}(\omega_j)e^{-i\omega_j} = I_{X_1Z}(\omega_j) + o(1/\sqrt{T}) \quad (4.54)$$

and

$$e^{-i\omega_j} I_{XZ}(\omega_j)e^{i\omega_j} = I_{X_1Z_1}(\omega_j) + o(1/T) \quad . \quad (4.55)$$

If we substitute (4.52) in (4.48(i)), make use of (4.52), (4.53) and (4.54), do the summation over ω_j and multiply by T we have

$$\begin{aligned} \{G_{(1)}^{-1}A_{(2)}X'ZP'_{(3)} - R'_{(4)}G_{(1)}^{-1}A_{(2)}X'Z_1P'_{(3)} - G_{(1)}^{-1}R_{(4)}A_{(2)}X'_1ZP'_{(3)} \\ + R'_{(4)}G_{(1)}^{-1}R_{(4)}A_{(2)}X'_1Z_1P'_{(3)}\}^V u = 0 \quad , \quad (4.56) \end{aligned}$$

where the asymptotically vanishing terms have been dropped.

Using $A_{(2)}X' - R_{(4)}A_{(2)}X'_1 = E'_{(2)}$, we have

$$\{G_{(1)}^{-1}E'_{(2)}ZP'_{(3)} - R'_{(4)}G_{(1)}^{-1}E'_{(2)}Z_1P'_{(3)}\}^V u = 0 \quad , \quad (4.57)$$

or using

$$\begin{aligned} R'(\ell) &= I \quad \text{for } \ell = 0 \\ &= -R' \quad = 1 \end{aligned}$$

(4.57) can be written as

$$\left\{ \sum_{\ell} R'(\ell) G_{(1)}^{-1} E'_{(2)} Z_{\ell} P'_{(3)} \right\} V_u = 0 \quad . \quad (4.58)$$

Throughout this study a subscript ℓ on a U, V, X, Y or Z matrix denotes an ℓ period lag.

In fact (4.58) is valid for the r^{th} order case if we define

$$R^{(0)} = -I, R^{(1)} = R'_1, \dots, R^{(r)} = R'_r \quad (4.58a)$$

where R_j are the parameter matrices of the autoregressive process and they are assumed to be unrestricted.

Once we use a finite parametrization of the error process (that does not change with T), the number of unknowns in the likelihood function is constant for any sample size. Therefore in the AR case the standard maximum likelihood theory applies and we will show that (4.58) are the maximum likelihood normal equations and that they allow a simplification [compare Hendry [1976] formula (55(i))] of the normal equations compared to the more general case with lagged endogenous variables, this simplification consists of replacing X_{ℓ} (in the general case) by $Z_{\ell} P'_{(3)}$ (in the fixed predetermined variables case) for $\ell \geq 1$, because the omitted terms are asymptotically negligible.

Proof

In order to show that, consider the reduced form model in the r^{th} order AR case,

$$Y' = -B^{-1} \Gamma Z' + B^{-1} U' = \Pi Z' + V' \quad (4.59)$$

where

$$U' = R_1 U'_1 + \dots + R_r U'_r + E' \quad , \quad (4.60)$$

$$V' = B^{-1} U' = S_1 V'_1 + \dots + S_r V'_r + E^0 \quad , \quad (4.61)$$

$$E^0 = B^{-1} E' \quad , \quad (4.62)$$

and

$$\text{Var}(\tilde{\varepsilon}_t^0) = G^0 = B^{-1} G B^{-1} \quad , \quad (4.63)$$

where $\tilde{\varepsilon}_t^{\circ}$ is the t^{th} column in E° . If $\tilde{\varepsilon}_t^{\circ}$ are Gaussian the log likelihood function can be written as

$$L_r = K_r - \frac{1}{2} \log \det G^{\circ} - \frac{1}{2} \text{tr} \{G^{\circ-1} E^{\circ} E^{\circ}\} . \quad (4.64)$$

Note that it does not matter to write L_r in terms of ϕ , G° and the S_k 's or in terms of ϕ , G and R_k 's, because since we assume no restrictions on the G and R_k 's matrices, we have that for any set of $\hat{\phi}$, \hat{G}° , and \hat{S}_k 's estimates we can determine uniquely the \hat{G} and \hat{R}_k 's matrices by

$$\hat{R}_k = \hat{B} \hat{S}_k \hat{B}^{-1} \quad (4.65)$$

and

$$\hat{G} = \hat{B} \hat{G}^{\circ} \hat{B}' . \quad (4.66)$$

From (4.64) we have that the first order conditions for the maximum likelihood estimates of ϕ are given by

$$\text{tr} \{ \hat{G}^{\circ-1} \hat{E}^{\circ}, \frac{\delta \hat{E}^{\circ}}{\delta \phi_j} \} = 0, \quad j = 1, \dots, n \quad (4.67)$$

and (4.67) can be written as

$$\text{tr} \{ \hat{G}^{\circ-1} (\hat{E}^{\circ} - E^{\circ}), \frac{\delta \hat{E}^{\circ}}{\delta \phi_j} \} = - \text{tr} \{ \hat{G}^{\circ-1} E^{\circ}, \frac{\delta \hat{E}^{\circ}}{\delta \phi_j} \} . \quad (4.68)$$

Using (4.61) and (4.59) replacing the true values by the estimates we have that

$$\frac{\delta \hat{\varepsilon}_t^{\circ}}{\delta \phi_j} = - \frac{\delta \hat{\Pi}}{\delta \phi_j} z_t + \sum_{k=1}^r \hat{S}_k \frac{\delta \hat{\Pi}}{\delta \phi_j} z_{t-k} = \sum_{k=0}^r \hat{S}^{(k)} \frac{\delta \hat{\Pi}}{\delta \phi_j} z_{t-k} , \quad (4.69)$$

where $\hat{S}^{(k)}$ is defined as in (4.58a) replacing $R^{(k)}$ by $\hat{S}^{(k)}$ and R_k by \hat{S}_k , and

$$\frac{\delta \hat{\varepsilon}_{st}^{\circ}}{\delta S_{k(\ell, q)}} = - v_{q(t-k)} \quad \text{if } s = \ell^{(10)} \\ = 0 \quad \text{otherwise.} \quad (4.70)$$

Using a Taylor series expansion for (4.68) we get:

(10) $\hat{v}_{q(t-k)}$ is the (q, t) element in \hat{V}_k' .

$$\begin{aligned} & \text{tr} \left\{ \hat{G}^{O-1} (\hat{E}^{O'} - E^{O'}) \frac{\delta \hat{E}^O}{\delta \phi_j} \right\} \\ &= \sum_{j=1}^n \left\{ \text{tr} \hat{G}^{O-1} \left[\frac{\delta E^{O+'}}{\delta \phi_j} \frac{\delta \hat{E}^O}{\delta \phi_i} \right] \right\} \Delta \phi_j \\ &+ \sum_{k=1}^r \sum_{\ell, q=1}^p \left\{ \text{tr} \hat{G}^{O-1} \frac{\delta E^{O+'}}{\delta S_{k(\ell, q)}} \frac{\delta \hat{E}^O}{\delta \phi_i} \right\} \Delta S_{k(\ell, q)}, \quad (4.71) \end{aligned}$$

where the + represents some point between $[\phi, S_1, \dots, S_r]$ and $[\hat{\phi}, \hat{S}_1, \dots, \hat{S}_r]$.

Now since we assume that the model is identified, the maximum likelihood estimators are consistent (see corollary I-2 Sargan [1972]) and we have

$$p \lim_{T \rightarrow \infty} \left[\text{tr} \left\{ \frac{1}{T} \hat{G}^{O-1} \left[\frac{\delta E^{O+'}}{\delta \phi_j} \frac{\delta \hat{E}^O}{\delta \phi_i} \right] \right\} \right] = p \lim_{T \rightarrow \infty} \left[\text{tr} \left\{ G^{O-1} \frac{1}{T} \frac{\delta E^{O'}}{\delta \phi_j} \frac{\delta E^O}{\delta \phi_i} \right\} \right].$$

On the other hand, using (4.69) noting that $\left[\frac{\delta \hat{E}^O}{\delta \phi_i} \right]'$ is a row in $\frac{\delta \hat{E}^O}{\delta \phi_i}$ we can write

$$\begin{aligned} & \text{tr} \left\{ \hat{G}^{O-1} \frac{\delta E^{O+'}}{\delta S_{k(\ell, q)}} \frac{\delta \hat{E}^O}{\delta \phi_i} \right\} = - \text{tr} \left\{ \hat{G}^{O-1} \frac{\delta E^{O+'}}{\delta S_{k(\ell, q)}} Z \frac{\delta \hat{\Pi}'}{\delta \phi_i} \right\} \\ &+ \sum_{k=1}^r \text{tr} \left[\hat{G}^{O-1} \frac{\delta E^{O+'}}{\delta S_{k(\ell, q)}} Z_k \frac{\delta \hat{\Pi}'}{\delta \phi_i} \hat{S}'_k \right]. \quad (4.73) \end{aligned}$$

Using (4.70) and the convention in (4.58a), but for S_k , we have that (4.73) equals to

$$- \sum_{k=0}^r \text{tr} \left\{ \frac{\delta \hat{\Pi}'}{\delta \phi_i} \hat{S}^{(k)'} \hat{G}^{O-1} \underset{\sim}{d}_q \underset{\sim}{v}_q^{+(k)'} Z_k \right\}, \quad (4.74)$$

where $\underset{\sim}{d}_q$ is a $(p \times 1)$ column vector with zeros everywhere except in the q^{th} element that is unity, and $\underset{\sim}{v}_q^{+(k)'} is the q^{th} row in $V_k^{+'}$, where the + means that the v's have been calculated for $\phi = \phi^+$, and the subscript k that the corresponding matrix has been lagged k times.$

Now $\underset{\sim}{v}_q^{+(k)'} Z_k$ is a row vector and its j^{th} element is given by

$$\sum_{t=1}^T (v_q^+ z_j(t-k)) / T = o(1/\sqrt{T}) \quad (4.75)$$

so

$$\text{tr} \left\{ \hat{G}^{o-1} \frac{\delta E^{o+}}{\delta S_{k(\ell, q)}} \frac{\delta \hat{E}^o}{\delta \phi_i} \right\} = o(1/\sqrt{T}) . \quad (4.76)$$

Therefore for large T (4.68) can be written as:

$$\begin{aligned} \sqrt{T} \sum_{j=1}^n \Delta \phi_j \text{tr} \left[\frac{\hat{G}^{o-1}}{T} \frac{\delta E^{o+}}{\delta \phi_j} \frac{\delta E^o}{\delta \phi_i} \right] = \\ \sum_{k=0}^r \text{tr} \left[\hat{G}^{o-1} \frac{E^{o'} z_k}{\sqrt{T}} \frac{\delta \hat{\Pi}'}{\delta \phi_i} \hat{S}(k)' \right], \quad i = 1, \dots, n \end{aligned} \quad (4.77)$$

where the LHS has been written using (4.71) and (4.72) and dropping the asymptotically vanishing terms according to (4.76) and the RHS has been written using (4.69).

Reversing the Taylor series expansion we have that the first order conditions for the maximum, according to (4.77) can be written as

$$\sum_{k=0}^r \text{tr} \left\{ \hat{S}(k)' \hat{G}^{o-1} \hat{E}^{o'} z_k \frac{\delta \hat{\Pi}'}{\delta \phi_j} \right\} = 0, \quad j = 1, \dots, n \quad (4.78)$$

For our case, in which the model is identified by zero restrictions, and therefore the elements of ϕ are the unrestricted elements of A we have that

$$\frac{\delta \hat{\Pi}}{\delta B_{\ell, j}} = \hat{B}^{-1} \frac{\delta B}{\delta B_{\ell, j}} \hat{B}^{-1} \Gamma \quad \text{and}$$

$$\frac{\delta \hat{\Pi}}{\delta \Gamma_{s, k}} = - \hat{B}^{-1} \frac{\delta \Gamma}{\delta \Gamma_{s, k}}$$

$$\text{so} \quad \begin{bmatrix} \delta \hat{\Pi} \\ \delta \phi \end{bmatrix} = \begin{bmatrix} - \hat{P}' & \hat{\Theta} \hat{B}^{-1} \end{bmatrix} u . \quad (4.79)$$

Now the equations in (4.78) can be written as

$$\sum_{k=0}^r \left[\text{vec} \left(\frac{\delta \hat{\Pi}}{\delta \phi_j} \right) \right]' \text{vec}(\hat{S}(k)' \hat{G}^{o-1} \hat{E}^{o'} z_k) = 0, \quad j = 1, \dots, n$$

$$\text{or } \left\{ \sum_{k=0}^r (\hat{P} \otimes \hat{B}^{-1}) \text{Vec}(\hat{S}^{(k)'} \hat{G}^{o-1} \hat{E}^{o'} z_k) \right\}^u = 0,$$

$$\text{or } \left\{ \sum_k \hat{B}^{-1} \hat{S}^{(k)'} \hat{G}^{o-1} \hat{E}^{o'} z_k \hat{P}' \right\}^{vu} = 0$$

$$= \left\{ \sum_k \hat{B}^{-1} \hat{B}' \hat{R}^{(k)'} \hat{B}^{-1} \hat{B}' \hat{G}^{-1} \hat{B} \hat{B}^{-1} \hat{E}' z_k \hat{P}' \right\}^{vu} = 0$$

and cancelling \hat{B} with \hat{B}^{-1} we get (4.58).

End of proof.

Going back to the Specfiml estimators, whose normal equations are those defined in (4.44), it is possible to specify a Spectral Three-Stage Least Squares (Spec 3SLS) estimator that will be computationally cheaper than $\tilde{\text{Specfiml}}$ but with the same asymptotic characteristics. If \hat{P} and $\hat{I}_U(\lambda_h)$ are consistent estimators of P and $F_U(\lambda_h)$ - for example they can be obtained from a Two Stage Least Squares (2SLS) estimation of A -, if we introduce the normalizing conditions

$$A = (B^\dagger : \Gamma) - (I : 0) = A^\dagger - (I : 0) \quad , \quad (4.80)$$

and if we vectorize (4.44) and cancel $(2M)^{-1}$, the Spec 3SLS, ϕ_{3S} , are given by

$$\left[\begin{array}{c} \Sigma \\ \lambda_h \end{array} \tilde{P} I_{ZX}(-\lambda_h) \otimes \tilde{I}_U^{-1}(\lambda_h) \right]^u \phi_{3S} = \left[\begin{array}{c} \Sigma \\ \lambda_h \end{array} \tilde{P} \otimes \tilde{I}_U^{-1}(\lambda_h) \text{Vec} I_{YZ}(\lambda_h) \right]^u \quad . \quad (4.81)$$

IV.3 The second order conditions.

To obtain the asymptotic distribution of the Specfiml estimators, \hat{A} , note that from (4.44), (4.46) and (4.18) we have

$$\begin{aligned} & \frac{1}{2M} \left[\begin{array}{c} \Sigma \\ \lambda_h \end{array} \hat{P} I_{ZX}(-\lambda_h) \otimes I_U^{-1}(\lambda_h) \right]^u \sqrt{T}(\hat{\phi} - \phi) \\ & = \frac{\sqrt{m}}{\sqrt{2M}} \left[\begin{array}{c} \Sigma \\ \lambda_h \end{array} \hat{P} \otimes I_U^{-1}(\lambda_h) \text{Vec} I_{UZ}(\lambda_h) \right]^u \quad . \quad (4.82) \end{aligned}$$

Noting that

$$X' = PZ' + \begin{bmatrix} V' \\ 0 \end{bmatrix}, \quad (4.83)$$

where V is defined in (4.59) using (3.1) for U , we have

$$\begin{aligned} I_{ZX}(\lambda_h) &= W_Z^*(\lambda_h) W_X(\lambda_h) \\ &= W_Z^*(\lambda_h) W_Z(\lambda_h) P' + [W_Z^*(\lambda_h) W_V(\lambda_h) : 0], \end{aligned} \quad (4.84)$$

and using (4.47), the last term in the RHS of (4.84) is asymptotically negligible. Making use of all that and following Hannan [1970] pp. 487-90 we have that

$$\begin{aligned} &P \lim \frac{1}{2M} \begin{bmatrix} \Sigma & \hat{P} \\ \lambda_h & I_{ZX}(-\lambda_h) \otimes I_U^{-1}(\lambda_h) \end{bmatrix}^u \\ &= \frac{1}{2\pi} \left[\int_{-\pi}^{\pi} P F_{ZZ}(-\lambda) P' \otimes F_U^{-1}(\lambda) d\lambda \right]^u. \end{aligned} \quad (4.85)$$

By the Hannan theorem (see Hannan [1970], pp. 442 and 487-492)

$$\frac{\sqrt{m}}{\sqrt{2M}} \begin{bmatrix} \Sigma & I_U^{-1}(\lambda_h) \text{Vec } I_{UZ}(\lambda_h) \\ \lambda_h & \end{bmatrix}^u \quad (4.86)$$

has an asymptotic multivariate normal distribution with variance-covariance matrix given by

$$\frac{1}{2\pi} \left[\int_{-\pi}^{\pi} F_{ZZ}(-\lambda) \otimes F_U^{-1}(\lambda) d\lambda \right]^u, \quad (4.87)$$

therefore $\sqrt{T}(\hat{\phi} - \phi)$ has an asymptotically multivariate normal distribution with zero means and variance covariance matrix as

$$\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} P F_{ZZ}(-\lambda) P' \otimes F_U^{-1}(\lambda) d\lambda \right]^u^{-1}. \quad (4.88)$$

The covariance matrix can be consistently estimated by

$$\left[\begin{array}{c} \frac{1}{2M} \quad \Sigma \quad \hat{P} \quad I_{ZZ}(-\lambda_h) \hat{P}' \quad \otimes \quad I_U^{-1}(\lambda_h) \\ \lambda_h \end{array} \right]^{u^{-1}}. \quad (4.89)$$

The equation (4.89) can be derived from the Hessian as follows:

$$\begin{aligned} \Delta q &= \left[-\frac{1}{2M} \quad \Sigma \quad I_U^{-1}(\lambda_h) \quad [\Delta A] I_{ZX}(\lambda_h) \hat{P}' \right]^{v_u} \\ &= \left[-\frac{1}{2M} \quad \Sigma \quad \left[\hat{P} \quad I_{ZX}(-\lambda_h) \quad \otimes \quad I_U^{-1}(\lambda_h) \right] \quad \text{vec}(\Delta A) \right]^u, \end{aligned} \quad (4.90)$$

so

$$\frac{\partial q}{\partial \phi'} = \left[-\frac{1}{2M} \quad \Sigma \quad \hat{P} \quad I_{ZX}(-\lambda_h) \quad \otimes \quad I_U^{-1}(\lambda_h) \right]^u, \quad (4.91)$$

and using (4.84) we can omit the asymptotically negligible terms in (4.91) and we get

$$\begin{aligned} \text{Hessian} &= \frac{\partial q}{\partial \phi'} = \frac{\partial^2 L_2}{\partial \phi \partial \phi'} = \left[-\frac{1}{2M} \quad \Sigma \quad \hat{P} \quad I_{ZZ}(-\lambda_h) \hat{P}' \quad \otimes \quad I_U^{-1}(\lambda_h) \right]^u \\ &= \left\{ -\frac{1}{2M} (\hat{P} \otimes I_P) \left[\begin{array}{c} \Sigma \quad I_{ZZ}(-\lambda_h) \quad \otimes \quad I_U^{-1}(\lambda_h) \\ \lambda_h \end{array} \right] (\hat{P} \otimes I_P)' \right\}^u \end{aligned} \quad (4.92)$$

and the proposed consistent estimator of (4.88) is minus the inverse of the Hessian.

There is no difficulty in proving that we get the same asymptotic distribution for $\sqrt{T}(\phi_{3S} - \phi)$. Also, and because of (4.47), (4.88) is the asymptotic variance of the Specfiml estimators when the $F_U(\lambda_h)$ matrices are known.

Once again, we have that the Specfiml expression for the asymptotic variance is a general one and from it we can obtain the variances for the cases when the errors are white noise or follow an AR process, by just substituting (4.49) or (4.52) into (4.88) and observing (4.2). Similarly (4.89) is a general expression for a consistent estimator of those variance-

covariance matrices, if we make use of expressions similar to those in (4.53) to (4.55) and that the sum of the periodogram-crossperiodogram matrices over all the frequencies is $T/2\pi$ times the sample covariance matrix.

In particular, for the first order AR case - a generalization to a higher order does not present additional problems - if we do in (4.88) the substitutions mentioned above we get

$$\begin{aligned} & \left[P G_Z(0)P' \otimes G^{-1} - P G_Z'(-1)P' \otimes R'G^{-1} - P G_Z(-1)P' \otimes G^{-1}R \right. \\ & \left. + P G_Z(0)P' \otimes R'G^{-1}R \right]^{-1} . \end{aligned} \quad (4.93)$$

We will prove now that (4.93) is the asymptotic variance of the maximum likelihood estimator when the errors follow a first order AR process.

Proof

From our previous discussion of the general AR case, we have that if we call this variance \bar{E}^{-1} , the (i, j) elements of \bar{E} is

$$- p \lim \left[\text{tr } G^{O-1} \frac{\partial E^{O'}}{\partial \phi_j} \frac{\partial E^O}{\partial \phi_i} \right] = \xi_{ij} , \quad (4.94)$$

and using (4.69) and the convention set up in (4.58a)

$$\begin{aligned} \text{tr} \left[G^{O-1} \frac{\partial E^{O'}}{\partial \phi_j} \frac{\partial E^O}{\partial \phi_i} \right] &= \text{tr} \left[G^{O-1} \sum_{k=0}^r S^{(k)} \frac{\partial \Pi}{\partial \phi_j} Z_k' \sum_{\ell=0}^r Z_\ell \frac{\partial \Pi'}{\partial \phi_i} S^{(\ell)'} \right] \\ &= \text{tr} \left[\sum_{\ell} \frac{\partial \Pi'}{\partial \phi_i} S^{(\ell)'} G^{O-1} \sum_k S^{(k)} \frac{\partial \Pi}{\partial \phi_j} Z_k' Z_\ell \right] \\ &= \sum_{\ell=0}^r \sum_{k=0}^r \left[\text{vec } \frac{\partial \Pi}{\partial \phi_i} \right]' Z_k' Z_\ell \otimes S^{(\ell)'} G^{O-1} S^{(k)} \left[\text{vec } \frac{\partial \Pi}{\partial \phi_j} \right] . \end{aligned} \quad (4.95)$$

Now using (4.79)

$$\begin{aligned}
 - \Xi &= p \lim \left[\begin{matrix} r & r \\ \Sigma & \Sigma \end{matrix} (\hat{P} \otimes \hat{B}^{-1}) \left(\frac{Z_k' Z_k}{T} \right) \otimes \left[S^{(l)'} G^{-1} S^{(k)} \right] (P' \otimes \hat{B}^{-1}) \right]^u \\
 &= p \lim \left[\begin{matrix} r & r \\ \Sigma & \Sigma \end{matrix} \hat{P} \frac{Z_k' Z_k}{T} P' \otimes R^{(l)'} G^{-1} R^{(k)} \right]^u \\
 &= \left[\begin{matrix} r & r \\ \Sigma & \Sigma \end{matrix} P G_Z (\ell-k) P' \otimes R^{(l)'} G^{-1} R^{(k)} \right]^u, \quad (4.96)
 \end{aligned}$$

and we can see that (4.93) is the expression for Ξ^{-1} when $r = 1$.

End of proof.

Summing up these results we have that an important characteristic of the simultaneous model with only exogenous variables in the set of predetermined variables, is that the asymptotic variance of the structural estimators does not depend on the variance of the estimators of the parameters of the error process i.e. for large T the Hessian is block-diagonal, and we have seen how this applies for Specfiml, FIML and ARFIML. This result is a generalization of the asymptotic efficiency of the GLS with a consistent estimator for the covariance matrix, in the context of the multiple regression system. On the other hand we have that for the variance of the parameters of the covariance matrix of u_t we only need to consider the south-east corner of the Hessian. In particular for the AR case, if we use

$$\rho = \text{Vec}(R_1 \ R_2 \ \dots \ R_r) \ ,$$

the general formula for the variance of ρ is:

$$\begin{aligned}
 \widehat{\text{Var}}(\rho) &= - \left[\frac{\partial^2 L}{\partial \rho \partial \rho'} - \frac{\partial^2 L}{\partial \rho \partial \phi'} \left(\frac{\partial L}{\partial \phi \partial \phi'} \right)^{-1} \frac{\partial^2 L}{\partial \phi \partial \rho'} \right]^{-1} \\
 &= - \left\{ \left[\frac{\partial^2 L}{\partial \rho \partial \rho'} \right]^{-1} + \left[\left(\frac{\partial^2 L}{\partial \rho \partial \rho'} \right)^{-1} \frac{\partial^2 L}{\partial \rho \partial \phi'} \left(\frac{\partial^2 L}{\partial \phi \partial \phi'} \right)^{-1} \frac{\partial^2 L}{\partial \phi \partial \rho'} \left(\frac{\partial^2 L}{\partial \rho \partial \rho'} \right)^{-1} \right] \right\} \\
 &\hspace{15em} (4.97)
 \end{aligned}$$

and in the present case since

$$p \lim \frac{\partial^2 L}{\partial \rho \partial \phi'} \rightarrow 0 \quad , \quad (\text{see (4.76)})$$

(4.97) collapses to

$$- \left[\frac{\partial^2 L}{\partial \rho \partial \rho'} \right]^{-1} \quad . \quad (4.98)$$

For the first order case the typical element $\frac{\partial^2 L}{\partial \rho \partial \rho'}$ is given by

$$\begin{aligned} & - \text{tr} \left[G^{O-1} \frac{\partial E^{O'}}{\partial S(\ell, q)} \quad \frac{\partial E^O}{\partial S(\ell', q')} \right] \\ & = - \left[\text{Vec} \left[B \frac{\partial E^{O'}}{\partial S(\ell', q')} \right] \right] I \otimes G^{-1} \left[\text{Vec} \left[B \frac{\partial E^{O'}}{\partial S(\ell, q)} \right] \right] , \quad (4.99) \end{aligned}$$

and using (4.70) and passing from the reduced to the structural form we end up with

$$- \left[\frac{\partial^2 L}{\partial \rho \partial \rho'} \right]^{-1} = (U_1' U_1)^{-1} \otimes G \quad . \quad (4.100)$$

Our expression (4.100) can be compared with formula (11) in Hendry [1971] to see the gain in simplicity that we get when all the predetermined variables are exogenous.

IV.4 The Hannan and Terrell procedure.

The Hannan-Terrell procedure consists of solving the 1st order conditions for Specfiml. Their method can be explained as follows. To solve (4.44) we will use a first order Taylor series expansion approximation as

$$H(\hat{\phi}_{r+1} - \hat{\phi}_r) = (-q)_r \quad (4.101)$$

or

$$\hat{\phi}_{r+1} = \hat{\phi}_r - H^{-1}(q)_r \quad ,$$

where H is the Hessian, and $(q)_r$ the first derivatives calculated using the results from the r^{th} iteration. For H they use

$$\frac{1}{2M} \sum_{\lambda_h} \hat{P}' I_{ZZ}(-\lambda_h) \hat{P}' \otimes \hat{F}_U^{-1}(\lambda_h)$$

and previously we have seen that this is a valid approximation when Z only contains exogenous variables. q is as (4.44) but reformulated as

$$\begin{aligned} & \left(-\frac{1}{2M} \sum_{\lambda_h} \hat{F}_U^{-1}(\lambda_h) \hat{B}' I_{YZ}(\lambda_h) \hat{P}' \right)' v_u \\ & + \left(-\frac{1}{2M} \sum_{\lambda_h} \hat{F}_U^{-1}(\lambda_h) \hat{\Gamma}' I_{ZZ}(\lambda_h) \hat{P}' \right)' v_u = 0 \end{aligned} \quad (4.102)$$

or using $\begin{bmatrix} 0 \\ \hat{\Gamma} \end{bmatrix} \hat{P} = \hat{\Gamma}$, as

$$\begin{aligned} & -\frac{1}{2M} \sum_{\lambda_h} \hat{P}' \otimes \hat{F}_U^{-1}(\lambda_h) \hat{B}' \text{Vec}(I_{YZ}) \\ & -\frac{1}{2M} \sum_{\lambda_h} \hat{P}' I_{ZZ}(-\lambda_h) \hat{P}' \otimes \hat{F}_U^{-1}(\lambda_h) \text{Vec} \begin{bmatrix} 0 \\ \hat{\Gamma} \end{bmatrix} \quad (11) (4.103) \end{aligned}$$

(11) Also note that instead of $\hat{F}_U^{-1}(\lambda_h)$ they use $B'^{-1} \hat{F}_V^{-1}(\lambda_h) B^{-1}$.

V The Specfiml estimation with inadequate sample size.

In the appendix to a recent paper Sargan [1975] shows that for the simultaneous model as defined in (4.1) but with white noise errors, the FIML estimators are not well defined when $T < p + k$. Here we show that for the general case of stationary errors, the Specfiml estimators are not well defined if for some band S_h , $m_h < p + k$.

We consider

$$g = \log \det (A_0 X' Q^*(\lambda_h) Q(\lambda_h) X A_0')$$

that is the log det term in the concentrated likelihood (L_2) replacing A by A_0 . If g is $-\infty$ for any h then L_2 becomes $+\infty$, and of course is very large for A close to A_0 . Now

$$\det(A_0 X' Q^*(\lambda_h) Q(\lambda_h) X A_0') = 0$$

if and only if for some complex vector η

$$Q(\lambda_h) X A_0' \eta = 0 \quad ,$$

and writing

$$A_0' \eta = \alpha \quad , \tag{5.1}$$

we have that

$$Q(\lambda_h) X \alpha = 0 \quad . \tag{5.2}$$

There will exist a suitable complex α if

$$m_h < p + k \quad , \tag{5.3}$$

and then if we can find an A_0 matrix satisfying the restrictions of the model such that

$$\det B_0 \neq 0 \tag{5.4}$$

and (5.3) holds, the first term, after the constant, of L_2 is finite and the second is infinite, so that the likelihood has no maximum. If we can also show that there are an infinite set of A_0 satisfying the conditions then the estimator can be considered to be unidentified.

If (5.3) is true for $h = 0$ or M we can show that $\underline{\alpha}$ is real. For example if $h = 0$ we have that the frequencies concerned are ω_j , $j = -(m_0 - 1)/2, \dots, 0, \dots, +(m_0 - 1)/2$ ⁽¹²⁾. If we now take a square matrix J which reverses the order of the rows of a matrix (J has all its elements zero except those on its transverse diagonal), then

$$J Q(\lambda_0) = \overline{Q(\lambda_0)} .$$

Thus

$$J Q(\lambda_0) X \underline{\alpha} = \overline{Q(\lambda_0)} X \underline{\alpha} = 0 .$$

Now taking the conjugate of this equation

$$Q(\lambda_0) X \overline{\underline{\alpha}} = 0 . \tag{5.5}$$

The vector $\underline{\alpha}$ may be purely imaginary. If so, multiply (5.2) by i , and one then has an equation with a real $\underline{\alpha}$. If, on the other hand, $\underline{\alpha}$ has a non-zero real part, then adding (5.2) and (5.5) we have

$$Q(\lambda_0) X (\underline{\alpha} + \overline{\underline{\alpha}}) = 0 ,$$

and since $(\underline{\alpha} + \overline{\underline{\alpha}})$ is real one can take (5.2) for $\underline{\alpha}$ real. Then we can solve (5.1) as in Sargan [1975] or as we do below for the other bands but considering only a set of equations. Note that with $\underline{\alpha}$ real, (5.2) can be split in $2 m_h$ equations m_h for the real part of $Q(\lambda_h)$ and m_h for the imaginary part, but since the band is symmetric around zero only m_h equations are different.

Similarly for $h = M$.

For other values of h we have to consider $\underline{\alpha}$ as complex, and it seems simpler to separate real and imaginary parts. We have from (5.1)

$$A'_0 \underline{\eta}_R = \underline{\alpha}_R \tag{5.6}$$

$$A'_0 \underline{\eta}_I = \underline{\alpha}_I .$$

We need to consider given $\underline{\alpha}_R$ and $\underline{\alpha}_I$, whether we can find A'_0 , $\underline{\eta}_R$ and $\underline{\eta}_I$ such that (5.6) is satisfied. This is a more difficult

(12) We are assuming that the m_h 's are odd.

problem than the previous one since we must satisfy both sets of equations. To deal with it we split $A_0 = (B_0 : \Gamma_0)$, $\alpha'_R = (\beta'_R : \gamma'_R)$, $\alpha'_I = (\beta'_I : \gamma'_I)$ and assume that every column of the Γ_0 matrix has two elements which may be non-zero.

We attempt to choose $A_0 = (B^+ : \Gamma_0)$ where B^+ is the true B matrix. This has the advantage that it satisfies the a priori constraints and $\det B^+ \neq 0$.

Then we must have

$$\begin{aligned} B^{+'} \eta_R &= \beta_R & (5.7) \\ B^{+'} \eta_I &= \beta_I \end{aligned}$$

and this determines η_R and η_I as linear functions of β_R and β_I .

Now, take Γ_0 as any matrix satisfying the zero restrictions and consider

$$\begin{aligned} \Gamma'_0 \eta_R &= \gamma_R & (5.8) \\ \Gamma'_0 \eta_I &= \gamma_I \end{aligned}$$

Now we can take the j^{th} row of Γ'_0 and without loss in generality we can assume that two non-zero elements in it⁽¹³⁾ appear in columns one and two and so, noting $\Gamma_0 = [c_{0lj}]$, we have

$$\begin{aligned} \eta_{R1} c_{01j} + \eta_{R2} c_{02j} &= \gamma_{Rj} \\ \eta_{I1} c_{01j} + \eta_{R2} c_{02j} &= \gamma_{Ij} \end{aligned} \quad (5.9)$$

In (5.9) we have two linear equations with (at least) two unknowns c_{01j} and c_{02j} to be determined. The solution is only impossible if

$$\det \begin{pmatrix} \eta_{R1} & \eta_{R2} \\ \eta_{I1} & \eta_{I2} \end{pmatrix} = 0$$

(13) To ensure at least one solution it is enough to consider two non-zero elements.

and since the η 's have an absolutely continuous distribution that will happen with probability zero.

If one row of Γ'_0 has more than two non-zero elements we have an infinite number of solutions for the corresponding c_{0kj} 's and therefore an infinite number of A matrices satisfying (5.1).

The treatment of $h \neq 0, M$ is unnecessary if $m_h = m, h=0, \dots, M$, for in these cases it is enough that the determinants corresponding to $h = 0$ and M can be made zero by feasible choices of the A matrix.

Summing up the preceding discussion we can state the following theorem:

Theorem 5.1: If in the model (4.1) each exogenous variable appears at least in two equations, the reduced form errors have every conditional distribution absolutely continuous, and if (5.3) is true, the probability that (5.4) and (5.1) are not satisfied is zero.

If at least one exogenous variable appears in more than two equations there exists an infinite number of A_0 matrices for which theorem 5.1 is true.

VI The estimation of the multiple regression model with stationary errors and lagged endogenous variables.

In this section we derive the spectral estimators for a multiple regression model that besides a set of exogenous variables also contains r lagged values of the dependent variable as regressors. This model can be expressed as

$$\alpha_0 \underline{y} = -\alpha_1 \underline{y}_1 - \alpha_2 \underline{y}_2 - \dots - \alpha_r \underline{y}_r + X\underline{\beta} + u \quad ,$$

$$\alpha_0 = 1 \quad (6.1)$$

where \underline{y} is the observation vector of the endogenous variable, \underline{y}_ℓ is the vector y lagged ℓ periods, and X is the observation matrix ($T \times k$) of the exogenous variables, ⁽¹⁴⁾ $\alpha_1, \dots, \alpha_r$ and $\underline{\beta}$ are the coefficients to be estimated and $\underline{u} = [u_t]$ is a purely non-deterministic stationary process with finite dispersion and can be represented as

$$u_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \quad , \quad a_0 = 1 \quad (6.1a)$$

where ε_t is an independent and identically distributed random variable with variance σ^2 . We require that

$$\sum_{j=0}^{\infty} j |a_j| < \infty \quad . \quad (6.1b)$$

The spectral density of u is

$$f_u(\lambda) = \frac{\sigma^2}{2\pi} \begin{bmatrix} \sum_0^{\infty} a_j e^{ij\lambda} \\ 0 \end{bmatrix} \begin{bmatrix} \sum_0^{\infty} a_j e^{-ij\lambda} \\ 0 \end{bmatrix} \quad (6.1c)$$

and we also require that

$$f_u(\lambda) \geq c > 0 \quad , \quad \lambda \in [-\pi, \pi] \quad . \quad (6.1d)$$

If the model had no lagged endogenous variables, we could require that the exogenous variables satisfy the Grenander

(14) Note the difference in terminology with respect the simultaneous equations model where X is the observation matrix of all the variables in the model.

conditions (see Hannan [1970] pp. 77-79, 219-216) and we would be able to derive the asymptotic properties of the spectral estimators for β . But for the case of model (6.1) we require the following further assumptions:

A 6.1 The autoregressive process involving the dependent variable is stable, i.e. the roots of

$$\alpha_0 + \alpha_1 Z + \alpha_2 Z^2 + \dots + \alpha_p Z^p, \quad \alpha_0 = 1$$

lie outside the unit circle.

A 6.2 The vector process $\{x_t\}$ is stationary to the fourth order with absolutely continuous spectra and

$$F_X(\lambda) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} G_X(\tau) \exp(i\tau\lambda),$$

where $F_X(\lambda)$ is the spectral density matrix of the variables in X and

$$G_X(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T x_t x_{t+\tau}' \quad , \quad \tau = 0, 1, 2, \dots$$

$$\sum_{-\infty}^{\infty} |j| \quad ||G_X(\tau)|| < \infty, \quad (6.2)$$

$$\det F_X(\lambda) \geq c > 0 \quad ;$$

$\{x_t\}$ also has fourth cumulant function that satisfies

$$\sum_{\tau_1, \tau_2, \tau_3} \sum_{-\infty}^{\infty} |K_{ljkn}^*(0, \tau_1, \tau_2, \tau_3)| < \infty, \quad (6.3)$$

and $G_X(0)$ is non-singular.

Assumptions A 6.1 and A 6.2 together with those on $\{u_t\}$ are required to ensure that the scalar process $\{y_t\}$ is stationary to the fourth order with absolutely continuous spectra and satisfies relations similar to (6.2) and (6.3), as required for the asymptotic theory presented below.

A 6.3 $\lim T/M^4 \rightarrow 0$

In the derivation of the Spec LS⁽¹⁵⁾ estimator for (6.1) we need to replace the average periodogram of a filtered (lagged) series by the average periodogram of the unfiltered series multiplied by the square of the gain of the filter and to ensure it does not produce a bias in the estimates we need A 6.3. This is the same requirement used in Hannan [1967] and Hannan [1970] section VII.7 for the derivation of the so called 'inefficient Hannan estimator' for the distributed lag relationship. The consequence is that now M must satisfy $M^2/T \rightarrow 0$ and $T/M^4 \rightarrow 0$ and therefore it must be carefully chosen.

It is possible that following work by Khatri [1965] and Wahba [1968] we could relax the above conditions assuming that the true series are Gaussian. And perhaps a further relaxation using Brillinger [1969] (see also Hannan & Thomson [1971]), is possible but we do not pursue it any further.

Under the hypothesis of a finite AR or AM process for $\{u_t\}$, the efficient estimation procedures known for these cases (see for instance, Sargan [1964] and Hannan and Nicholls [1972] respectively)⁽¹⁶⁾ also assume A6.1, but in them A6.2 can be replaced by the Grenander conditions on the exogenous variables, and A6.3 is not required at all. The point is that in those cases, in order to maximize an asymptotically valid a approximation of the log likelihood

$$c - \frac{1}{2} \log \sigma^2 - \frac{1}{2} \underline{\underline{\epsilon}}' \underline{\underline{\epsilon}} / \sigma^2$$

the advantage of the explicit parametrisation of $\{u_t\}$ is used and an estimate of $f_u(\lambda_h)$ is not required. But once a particular process is not specified an estimate of $f_u(\lambda_h)$ is required and since (6.1) contains an autoregressive relation in the endogenous variable the conditions A6.2 and A6.3 are needed for the asymptotic theory presented below.

(15) This is our terminology for the estimators proposed in this section.

(16) For an extension to the multivariate regression model with lagged endogenous variables and vector MA errors, see Nicholls [1975].

VI.1 Alternative likelihood functions.

Augmenting the X matrix with the r , y_ℓ vectors we could obtain the spectral estimators of the α 's and β as in Section IV beginning from the likelihood

$$L_u = c_u - \frac{m}{2} \sum_h \log f_u(\lambda_h) - \frac{m}{2} \sum_h I_u(\lambda_h)/f_u(\lambda_h) \quad (6.4)$$

but the Hannan theorem does not apply and it is difficult to obtain the asymptotic properties of such estimators. Therefore we derive the spectral estimators using a set of stronger assumptions that enable us to operate from an alternative form of the likelihood of w_u .

On multiplying (6.1) by $q'(\omega_j)$ we obtain:

$$\alpha_0 w_y(\omega_j) + \alpha_1 w_{y_1}(\omega_j) + \dots + \alpha_r w_{y_r}(\omega_j) = w_x'(\omega_j) \beta + w_u(\omega_j), \quad (6.5)$$

that can be expressed as

$$\alpha(\omega_j) w_y(\omega_j) = w_x'(\omega_j) \beta + w_u(\omega_j) + o(1/\sqrt{T}), \quad (6.6)$$

where

$$\alpha(\omega_j) = \alpha_0 + \alpha_1 \exp(i\omega_j) + \alpha_2 \exp(2i\omega_j) + \dots + \alpha_r \exp(ri\omega_j). \quad (6.7)$$

The residual term in (6.6) is

$$\frac{1}{\sqrt{T}} \left\{ \sum_{s=0}^{r-1} \alpha_{s+1} \exp[i\omega_j(s+1)] \left[\sum_{\ell=-s}^0 y_\ell \exp(i\omega_j \ell) - \sum_{\ell=T-s}^T y_\ell \exp(i\omega_j \ell) \right] \right\} \quad (6.8)$$

and since the expression in curly brackets only contains a finite number of terms, it is of order one, hence the whole expression is $o(1/T^{1/2})$. As a result the standard deviation of the residual term is $o(M/T^{1/2})$ (see Hannan [1970] p. 493) and therefore its contribution is negligible.

Using the results of Section III for large T, we have

$$\mathcal{L}(w_u) = c(\det D)^{-1/2} \exp \left[-\frac{1}{2} \sum_j I_u(\omega_j)/f_u(\omega_j) \right] \quad (6.9)$$

where D is a diagonal matrix with $f_u(\omega_j)$, ($j = 0, 1, \dots, T-1$), in the main diagonal. Then

$$\mathcal{L}(\underline{w}_y) = c \prod_j |\alpha(\omega_j)| (\det D)^{-1/2} \exp \left[-\frac{1}{2} \sum_j I_u(\omega_j)/f_u(\omega_j) \right], \quad (6.10)$$

where $I_u(\omega_j)$ is given by

$$I_u(\omega_j) = \bar{\alpha}(\omega_j) I_y(\omega_j) \alpha(\omega_j) - \bar{\alpha}(\omega_j) \tilde{I}'_{yX}(\omega_j) \tilde{\beta} - \tilde{\beta}' \tilde{I}_{Xy}(\omega_j) \alpha(\omega_j) + \tilde{\beta}' I_{XX}(\omega_j) \tilde{\beta} + o(1/\sqrt{T}), \quad (6.11)$$

and $\prod_j |\alpha(\omega_j)|$ is the Jacobian of the transformation. The log likelihood is

$$L(\underline{w}_y) = c_1 + \sum_j \log |\alpha(\omega_j)| - \frac{1}{2} \sum_j \log f_u(\omega_j) - \frac{1}{2} \sum_j I_u(\omega_j)/f_u(\omega_j). \quad (6.12)$$

Assuming that $f_u(\omega_j)$ is constant within bands⁽¹⁷⁾ we obtain

$$L(\underline{w}_y) = c_1 + \sum_j \log |\alpha(\omega_j)| - \frac{m}{2} \sum_h \log f_u(\lambda_h) - \frac{m}{2} \sum_h I_u(\lambda_h)/f_u(\lambda_h), \quad (6.13)$$

where $I_u(\lambda_h) = \frac{1}{m} \sum_{\omega_j \in S_h} I_u(\omega_j)$ and $I_u(\omega_j)$ as defined in (6.11).

Now, since we are assuming continuity and using assumption A6.3, we can replace (6.13) by

$$L_2 = c_2 + m \sum_h \log |\alpha(\lambda_h)| - \frac{m}{2} \sum_h \log f_u(\lambda_h) - \frac{m}{2} \sum_h \ddot{I}_u(\lambda_h)/f_u(\lambda_h), \quad (6.14)$$

$$\text{where } \ddot{I}_u(\lambda_h) = \bar{\gamma}'(\lambda_h) I_{X+X}(\lambda_h) \gamma(\lambda_h), \quad (18) \quad (6.15)$$

(17) As before we split T (number of observations) as $T = 2Mm$, where $2M$ are the number of bands between $-\pi$ and π and m - we assume is odd - the number of periodogram ordinates in each band.

(18) The difference between $I_u(\lambda_h)$ and $\ddot{I}_u(\lambda_h)$ as an estimate of $f_u(\lambda_h)$ is that in the former we apply a rectangular spectral window to the residual periodogram calculated as in (6.11), whereas in the latter we calculate a residual periodogram by applying the regression filter to the smoothed periodograms of the y and X variables.

$$\underline{\gamma}'(\lambda_h) = \left[\alpha(\lambda_h) - \underline{\beta}' \right] \quad (6.16)$$

$$X^+ = (\underline{y} \ X) \quad \text{and} \quad (6.17)$$

$I_{X^+X^+}(\lambda_h)$ denotes the average periodogram matrix for the variable in X^+ . That replacing $L(\underline{w}_y)$ by L_2 simply omits terms with negligible asymptotic contribution under the present assumptions, can be shown using an argument similar to that employed in Hannan [1970] pp. 492-495, by putting our smoothed estimator for the spectra (the average periodogram, FFT in Hannan's terminology) in terms of the sample covariances and a lag window (Hannan [1970] pp. 275-6). Finally, to reconcile L_2 with L_4 defined in (6.4) we need to show that the difference of the Jacobians is asymptotically negligible, and we show that in the Appendix for the more general case of simultaneous systems.

Now differentiating L_2 with respect to $f_u(\lambda_h)$ we have

$$\frac{\partial L_2}{\partial f_u(\lambda_h)} = - \frac{m}{2f_u(\lambda_h)} + \frac{m \ddot{I}_u(\lambda_h)}{2f_u^2(\lambda_h)}$$

and equating to zero gives

$$\hat{f}_u(\lambda_h) = \ddot{I}_u(\lambda_h) \quad . \quad (6.18)$$

Concentrating L_2 with respect to $f_u(\lambda_h)$ we obtain

$$L_3 = c_3 + m \sum_{h=0}^{2M-1} \log \alpha(-\lambda_h) - \frac{m}{2} \sum_{h=0}^{2M-1} \log \ddot{I}_u(\lambda_h) \quad , \quad (6.19)$$

where $\ddot{I}_u(\lambda_h)$ is defined as in (6.15). To obtain (6.19) we have to make use of

$$\sum_{h=0}^{2M-1} \log |\alpha(\lambda_h)| = \sum_{h=0}^{2M-1} \log \alpha(-\lambda_h) \quad .$$

VI.2 The Spec LS estimator and its asymptotic distribution.

Let us call

$$\underline{\gamma}^0 = (\alpha_0 \alpha_1 \dots \alpha_r - \beta_1 \dots - \beta_k)' = (\underline{\alpha}^0' - \underline{\beta}')' \quad , \quad \alpha_0 = 1 \quad (6.20)$$

$$\text{and } \tilde{\theta}'(\lambda_h) = [1 \exp(i\lambda_h) \exp(i2\lambda_h) \dots \exp(ir\lambda_h)], \quad (6.21)$$

$$\text{then } \alpha(\lambda_h) = \tilde{\alpha}^0 \tilde{\theta}(\lambda_h) \quad (6.22)$$

From (6.15) we have:

$$\ddot{I}_u(\lambda_h) = \alpha(-\lambda_h) \tilde{I}'_{yX^+}(\lambda_h) \tilde{\gamma}(\lambda_h) - \tilde{\beta}' I_{XX^+}(\lambda_h) \tilde{\gamma}(\lambda_h) \quad ,$$

and so

$$\begin{aligned} 1/\alpha(-\lambda_h) &= \left[1/\ddot{I}_u(\lambda_h) \right] \tilde{I}'_{yX^+}(\lambda_h) \tilde{\gamma}(\lambda_h) \\ &- \left[\alpha(-\lambda_h) \ddot{I}_u(\lambda_h) \right]^{-1} \tilde{\beta}' I_{XX^+}(\lambda_h) \tilde{\gamma}(\lambda_h) \quad . \end{aligned} \quad (6.23)$$

To obtain the first order conditions for the maximum of L_3 subject to the constraint that $\alpha_0 = 1$, we impose that condition by a Lagrange multiplier, Λ ,⁽¹⁹⁾ and form

$$g_3 = m \sum_h \log \alpha(-\lambda_h) - \frac{m}{2} \sum_h \log \ddot{I}_u(\lambda_h) + 2\Lambda(e_1' \tilde{\gamma}^0 - 1) \quad , \quad (6.24)$$

where e_1 is a $1 \times (r + 1 + k)$ vector with one in the first row and zeros everywhere else.

Now

$$\frac{\partial g_3}{\partial \tilde{\gamma}^0} = m \sum_h \left[1/\alpha(-\lambda_h) \right] \frac{\partial \alpha(-\lambda_h)}{\partial \tilde{\gamma}^0} - \frac{m}{2} \sum_h \left[1/\ddot{I}_u(\lambda_h) \right] \frac{\partial \ddot{I}_u(\lambda_h)}{\partial \tilde{\gamma}^0} + 2\Lambda e_1 \quad . \quad (6.25)$$

The first term in the RHS of (6.25) can be expressed using (6.23) as

$$\begin{aligned} m \sum_h \left[1/\ddot{I}_u(\lambda_h) \right] \frac{\partial \alpha(-\lambda_h)}{\partial \tilde{\gamma}^0} \tilde{I}'_{yX^+}(\lambda_h) \tilde{\gamma}(\lambda_h) \\ - m \sum_h \left[1/\ddot{I}_u(\lambda_h) \right] \frac{\partial \alpha(-\lambda_h)}{\partial \tilde{\gamma}^0} \left[1/\alpha(-\lambda_h) \right] \tilde{\beta}' I_{XX^+}(\lambda_h) \tilde{\gamma}(\lambda_h), \end{aligned} \quad (6.26)$$

(19) Note that Λ is a scalar. We use a capital letter for the multiplier to avoid confusion with λ_h .

and the second as

$$\begin{aligned}
 & - \frac{m}{2} \sum_h [1/\bar{I}_u(\lambda_h)] \frac{\partial \underline{\gamma}'(-\lambda_h)}{\partial \underline{\gamma}^o} \frac{\partial \bar{I}_u(\lambda_h)}{\partial \underline{\gamma}(-\lambda_h)} \\
 & = - m \sum_h [1/\bar{I}_u(\lambda_h)] \frac{\partial \underline{\gamma}'(-\lambda_h)}{\partial \underline{\gamma}^o} I_{X^+X^+}(\lambda_h) \underline{\gamma}(\lambda_h) \\
 & = - m \sum_h [1/\bar{I}_u(\lambda_h)] \left[\frac{\partial \alpha(-\lambda_h)}{\partial \underline{\gamma}^o} I_{YX^+}(\lambda_h) \underline{\gamma}(\lambda_h) \right. \\
 & \quad \left. - \frac{\partial \underline{\beta}'}{\partial \underline{\gamma}^o} I_{XX^+}(\lambda_h) \underline{\gamma}(\lambda_h) \right] . \tag{6.27}
 \end{aligned}$$

Therefore, doing the appropriate cancellations we get

$$\begin{aligned}
 \frac{\partial g_3}{\partial \underline{\gamma}^o} & = - m \sum_h [1/\bar{I}_u(\lambda_h)] \frac{\partial \alpha(-\lambda_h)}{\partial \underline{\gamma}^o} [1/\alpha(-\lambda_h)] \underline{\beta}' I_{XX^+}(\lambda_h) \underline{\gamma}(\lambda_h) \\
 & \quad + m \sum_h [1/\bar{I}_u(\lambda_h)] \frac{\partial \underline{\beta}'}{\partial \underline{\gamma}^o} I_{XX^+}(\lambda_h) \underline{\gamma}(\lambda_h) + 2\Lambda \underline{e}_1 . \tag{6.28}
 \end{aligned}$$

Using

$$\frac{\partial \alpha(-\lambda_h)}{\partial \underline{\gamma}^o} = \left[\frac{\partial \alpha}{\partial \underline{\gamma}^o} \right] \underline{\varrho}(-\lambda_h) \tag{6.29}$$

and

$$P^*(\lambda_h) = \begin{bmatrix} [1/\hat{\alpha}(-\lambda_h)] \underline{\varrho}(-\lambda_h) \hat{\underline{\beta}}' \\ I_k \end{bmatrix} , \tag{6.30}$$

$$\frac{\partial g_3}{\partial \underline{\gamma}^o} = - m \sum_h [1/\bar{I}_u(\lambda_h)] P^*(\lambda_h) I_{XX^+}(\lambda_h) \underline{\gamma}(\lambda_h) + 2\Lambda \underline{e}_1 \tag{6.30a}^{(20)}$$

(20) Note that $\frac{\partial \underline{\beta}'}{\partial \underline{\gamma}^o} = \begin{bmatrix} 0 \\ -I_k \end{bmatrix}$.

Define

$$X^\dagger = (y_1 \dots y_r X) = (Y X) . \quad (6.31)$$

Arguing as we did after (6.17) we have that

$$I_{XX}^+(\lambda_h) \tilde{y}(\lambda_h) = I_{XY}(\lambda_h) \tilde{\theta}'(\lambda_h) \alpha^0 - I_{XX}(\lambda_h) \tilde{\beta} \quad (6.32)$$

can, in the limit, be simplified. To do that note that

$$I_{XY}(\lambda_h) = W_X^*(\lambda_h) \tilde{w}_y(\lambda_h) . \quad (6.33)$$

Using (6.6) and relying on the continuity assumption, in the limit $\tilde{w}_y(\lambda_h)$ can be replaced by

$$\frac{1}{\alpha(\lambda_h)} \left[W_X(\lambda_h) \tilde{\beta} + \tilde{w}_u(\lambda_h) \right] . \quad (6.34)$$

Using (6.34) and neglecting $\frac{1}{\alpha(\lambda_h)} W_X^*(\lambda_h) \tilde{w}_u(\lambda_h)$, $I_{XY}(\lambda_h)$ can be expressed as

$$I_{XY}(\lambda_h) = [1/\alpha(\lambda_h)] I_{XX}(\lambda_h) \tilde{\beta} , \quad (6.35)$$

and making use of it

$$I_{XX}^+(\lambda_h) \tilde{y}(\lambda_h) = I_{XX}(\lambda_h) P(\lambda_h) \tilde{y}^0 . \quad (6.36)$$

Then

$$\frac{\partial g_3}{\partial \tilde{y}^0} = - m \tilde{J} \tilde{y}^0 + 2\Lambda \tilde{e}_1 , \quad (6.37)$$

where

$$\tilde{J} = \sum_h [1/\tilde{I}_u(\lambda_h)] P^*(\lambda_h) I_{XX}(\lambda_h) P(\lambda_h) , \quad (6.38)$$

and equating (6.37) to zero we get

$$\tilde{y}^0 = \frac{2}{m} \tilde{J}^{-1} \Lambda \tilde{e}_1 . \quad (6.39)$$

Substituting (6.39) into the first order condition for Λ and solving for Λ after equating to zero we have

$$\Lambda = \frac{m}{2} (\tilde{e}_1' \tilde{J}^{-1} \tilde{e}_1)^{-1} \quad (6.40)$$

and then

$$\underline{\gamma}^0 = (\underline{e}_1' J^{-1} \underline{e}_1)^{-1} J^{-1} \underline{e}_1 \quad . \quad (6.41)$$

Splitting J as

$$J = \begin{bmatrix} \underline{a} & \underline{p}' \\ \underline{p} & P \end{bmatrix} \quad , \quad (21) \quad (6.42)$$

from (6.41) we have that

$$\hat{\underline{\gamma}} = - P^{-1} \underline{p} \quad , \quad (6.43)$$

where

$$\hat{\underline{\gamma}} = (\hat{\alpha}_1 \dots \hat{\alpha}_r - \hat{\beta}_1 \dots - \hat{\beta}_k)' = (\underline{\alpha}' - \underline{\beta}')' \quad . (6.44)$$

Then

$$\hat{\underline{\gamma}} = - \left[\begin{array}{c} \Sigma_h^{-1} \ddot{I}_u^{-1}(\lambda_h) |\hat{\alpha}(\lambda_h)|^2 \bar{\theta}_1 \hat{\beta}' I_{XX}(\lambda_h) \hat{\beta} \theta_1' | \Sigma_h \left[\ddot{I}_u(\lambda_h) \hat{\alpha}(-\lambda_h) \right]^{-1} \bar{\theta}_1 \hat{\beta}' I_{XX}(\lambda_h) \\ \Sigma_h [1/\ddot{I}_u(\lambda_h)] [1/\hat{\alpha}(\lambda_h)] I_{XX}(\lambda_h) \hat{\beta} \theta_1' | \Sigma_h [1/\ddot{I}_u(\lambda_h)] I_{XX}(\lambda_h) \\ \Sigma_h [1/\ddot{I}_u(\lambda_h)] [1/\hat{\alpha}(-\lambda_h)] \bar{\theta}_1 \hat{\beta}' I_{XY}(\lambda_h) \\ \Sigma_h [1/\ddot{I}_u(\lambda_h)] I_{XY}(\lambda_h) \end{array} \right]^{-1} \\ = - \left[\Sigma_h [1/\ddot{I}_u(\lambda_h)] P_1^*(\lambda_h) I_{XX}(\lambda_h) P_1(\lambda_h) \right]^{-1} \left[\Sigma_h [1/\ddot{I}_u(\lambda_h)] P_1^*(\lambda_h) I_{XY}(\lambda_h) \right] \quad , \quad (6.45)$$

where $P_1^*(\lambda_h)$ is defined as $P^*(\lambda_h)$ but using $\bar{\theta}_1 \hat{\beta}' = \theta_1(-\lambda_h) \hat{\beta}'$ instead of $\theta_1(-\lambda_h) \hat{\beta}'$ and

$$\bar{\theta}_1 = \theta_1(-\lambda_h) = \left[\exp(-i\lambda_h) \exp(-i2\lambda_h) \dots \exp(-ir\lambda_h) \right]' \quad . \quad (6.46)$$

To obtain (6.45) from (6.43) we have to rewrite \underline{p} using (6.35).

(21) The P matrix here is different to the one defined in (4.45).

If we express P as

$$P = \Sigma \left[1/\ddot{I}_u(\lambda_h) \right] P_1^*(\lambda_h) W_X^*(\lambda_h) W_X(\lambda_h) P_1(\lambda_h) \quad (6.47)$$

and if as argued just before (6.34) we use

$$\frac{1}{\alpha(-\lambda_h)} \beta' W_X^*(\lambda_h) = \bar{w}'_y(\lambda_h) - \frac{1}{\alpha(-\lambda_h)} w'_u(\lambda_h), \quad (6.48)$$

and if we neglect the term $(1/\alpha(-\lambda_h))W_X^*(\lambda_h)w_u(\lambda_h)$, (6.45) can be written as

$$\hat{Y} = \left[\begin{array}{c} \Sigma \ddot{I}_u^{-1}(\lambda_h) \bar{\theta}'_1 [I_Y(\lambda_h) - \ddot{I}_u(\lambda_h)/|\alpha(\lambda_h)|^2] \theta'_1 \quad \Sigma \ddot{I}_u^{-1}(\lambda_h) \bar{\theta}'_1 I'_{yX}(\lambda_h) \\ \Sigma (1/\ddot{I}_u(\lambda_h)) I_{XY}(\lambda_h) \theta'_1 \quad \Sigma (1/\ddot{I}_u(\lambda_h)) I_{XX}(\lambda_h) \end{array} \right]^{-1} \left[\begin{array}{c} \Sigma (1/\ddot{I}_u(\lambda_h)) \bar{\theta}'_1 [I_Y(\lambda_h) - \ddot{I}_u(\lambda_h)/|\alpha(\lambda_h)|^2] \\ \Sigma (1/\ddot{I}_u(\lambda_h)) I_{XY}(\lambda_h) \end{array} \right]. \quad (6.49)$$

If we note that matrix $\bar{\theta}'_1 \theta'_1$ applied to $f_y(\lambda)$, gives us the spectra matrix of y_1, \dots, y_r and that $\bar{\theta}'_1$ applied to $f'_{yX}(\lambda)$ gives us the cross-spectra matrix between $Y = (y_1, \dots, y_r)$ and X , we can see that what we are doing in (6.49) is to remove from $I_{YY}(\lambda_h)$ and $I_{Yy}(\lambda_h)$ the effect due to noise, i.e. the intensity of the regressors other than the purely exogenous ones has been corrected by subtracting the part that is correlated with the noise of the equation. For our particular model, provided it is well specified, the sensible correction to do is to replace $I_{YY}(\lambda_h)$ and $I_{Yy}(\lambda_h)$ by

$$\begin{aligned} (1/|\alpha(\lambda_h)|^2) \bar{\theta}'_1 \hat{\beta}' I_{XX}(\lambda_h) \hat{\beta} \theta'_1 \quad \text{and} \\ (1/|\alpha(\lambda_h)|^2) \bar{\theta}'_1 \hat{\beta}' I_{XX}(\lambda_h) \hat{\beta} \end{aligned} \quad (6.50)$$

respectively, and we can refer to this as the endogenization of the lagged endogenous variables.

To obtain the asymptotic properties of $\hat{\gamma}$ note that from (6.30a)

$$\hat{\gamma} = - \begin{bmatrix} \sum_h (1/\ddot{I}_u(\lambda_h))(1/\hat{\alpha}(-\lambda_h)) \ddot{\theta}_1 \hat{\beta}' I_{XX^\dagger}(\lambda_h) \\ \sum_h (1/\ddot{I}_u(\lambda_h)) I_{XX^\dagger}(\lambda_h) \end{bmatrix}^{-1} \begin{bmatrix} \sum_h (1/\ddot{I}_u(\lambda_h))(1/\hat{\alpha}(-\lambda_h)) \ddot{\theta}_1 \hat{\beta}' I_{XY}(\lambda_h) \\ \sum_h (1/\ddot{I}_u(\lambda_h)) I_{XY}(\lambda_h) \end{bmatrix} = - P^{\dagger-1} \underline{p}, \quad (6.50a)$$

and using $I_{XY}(\lambda_h) = - I_{XX^\dagger}(\lambda_h)\underline{\gamma} + I_{Xu}(\lambda_h)$, (6.50a) can be written as

$$\begin{aligned} \sqrt{T} (\hat{\gamma} - \underline{\gamma}) &= - \left[\frac{1}{2M} P^\dagger \right]^{-1} \frac{\sqrt{T}}{2M} \left[\sum_h (1/\ddot{I}_u(\lambda_h)) P_1^*(\lambda_h) I_{Xu}(\lambda_h) \right] \\ &= - \left\{ \frac{1}{2M} P^\dagger \right\}^{-1} \left\{ \sqrt{T}/2M \right\} \ddot{\underline{p}}_u. \end{aligned} \quad (6.51)$$

Under the present assumptions, and arguing along the lines of Hannan [1970] pp. 492-95, we have that asymptotically we can replace $\sqrt{T}/2M \ddot{\underline{p}}_u$ by $\sqrt{T}/2M \underline{p}_u$, where \underline{p}_u is defined as $\ddot{\underline{p}}_u$ but replacing $\ddot{I}_u(\lambda_h)$ by $I_u(\lambda_h)$. Then by the Hannan theorem $\sqrt{T}/2M \underline{p}_u$ is asymptotically normally distributed with zero mean and variance covariance matrix given by

$$V = \int_{-\pi}^{\pi} \frac{1}{2\pi} \dot{P}_1^*(\lambda) F_X(\lambda) \dot{P}_1(\lambda) / f_u(\lambda) d\lambda$$

$$= \left[\begin{array}{l} \int_{-\pi}^{\pi} \frac{1}{2\pi} \frac{1}{f_u(\lambda) \cdot |\alpha(\lambda)|^2} \bar{\theta}' \beta' F_X(\lambda) \beta \, d\lambda \\ \int_{-\pi}^{\pi} \frac{1}{2\pi} \frac{1}{f_u(\lambda) \cdot \alpha(-\lambda)} \bar{\theta}' \beta' F_X(\lambda) \, d\lambda \end{array} \right] \\ = \left[\begin{array}{l} \int_{-\pi}^{\pi} \frac{1}{2\pi} \frac{1}{f_u(\lambda) \cdot \alpha(\lambda)} F_X^*(\lambda) \beta \theta' \, d\lambda \\ \int_{-\pi}^{\pi} \frac{1}{2\pi} \frac{1}{f_u(\lambda)} F_X(\lambda) \, d\lambda \end{array} \right] \quad (6.52)$$

where $\dot{P}_1(\lambda)$ is defined as $P_1(\lambda_h)$ but using the true values of $\underline{\alpha}$ and $\underline{\beta}$ and $f_u(\lambda)$ instead of $\hat{I}_u(\lambda_h)$. Also

$$p \lim_{T \rightarrow \infty} \frac{1}{2M} P = p \lim_{T \rightarrow \infty} \frac{1}{2M} P^{\dagger} \rightarrow V$$

therefore $\sqrt{T} (\hat{\gamma} - \gamma)$ is asymptotically normally distributed with zero mean and variance-covariance matrix given by V^{-1} . The matrix V can be consistently estimated by P .

VI.3 The spectral estimator and the maximum likelihood estimators.

If the $\{u_t\}$ process has a finite MA representation as

$$u_t = \sum_0^q a_j \varepsilon_{t-j} \quad , \quad a_0 = 1 \quad (6.53)$$

then $f_u(\lambda) = \frac{\sigma^2}{2\pi} |g|^2 \quad , \quad (6.54)$

where

$$g = \sum_0^q a_j \exp(ij\lambda) \quad . \quad (22)$$

If we substitute $f_u(\lambda)$ in V by its value in (6.54) we get the V matrix for the finite MA case and it can be compared with the corresponding matrix in Hannan and Nicholls [1972], where

(22) To keep notation close to that of Hannan-Nicholls [1972] we have to omit the argument λ in g .

they propose an estimator for the parameters of a model of the type (6.1) with errors as (6.53). It can be seen that our V corresponds to their

$$\begin{bmatrix} \Delta_{11} & -\Delta_{12} \\ -\Delta_{21} & \Delta_{22} \end{bmatrix} = \Delta^0, \text{ say.}$$

The variance that they obtain for $\sqrt{T}(\tilde{\gamma} - \gamma)$, ($\tilde{\gamma}$ is our notation for the estimate that they propose for γ) is

$$\left\{ \Delta^0 + \begin{bmatrix} \Psi - \Omega\Phi^{-1}\Omega & 0 \\ 0 & 0 \end{bmatrix} \right\}^{-1}, \quad (6.55)$$

where

$$\Psi = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{|\alpha(\lambda)|^2} \bar{\theta}_1 \theta_1' d\lambda, \quad (6.56)$$

$$\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g\alpha(-\lambda)} \bar{\theta}_1 \theta_1' d\lambda, \quad (6.57)$$

$$\Phi = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{|g|^2} \bar{\theta}_1 \theta_1' d\lambda. \quad (6.58)$$

It can be shown that $[\Psi - \Omega\Phi^{-1}\Omega]$ is a positive definite matrix, therefore $\tilde{\gamma}$ will be efficient with respect to $\hat{\gamma}$. It can be proved, following Akaike [1973], that the Hannan-Nicholls procedure "is equivalent to a three stage realization of one step of the Newton-Raphson procedure for the numerical maximization of the likelihood function, using the gradient and approximate Hessian" developed in Akaike [1973]. Thus in the case of MA errors $\hat{\gamma}$ is not efficient with respect to the maximum likelihood estimator. And what is worse, since $[\Psi - \Omega\Phi^{-1}\Omega]$ is positive definite, even if we overstate the order of the MA in the calculation of $\tilde{\gamma}$ this estimator, although no longer having the efficiency of the maximum likelihood estimator, still will be efficient with respect to $\hat{\gamma}$. Of course

if the order of the MA is understated or if $\{u_t\}$ does not follow a MA process $\tilde{\gamma}$ will not be a consistent estimator, when $\hat{\gamma}$ will be anyway.

The matrix $[\Psi - \Omega\Phi^{-1}\Omega]$ is the zero matrix if $g = c.\alpha(\lambda)$, where c is a constant. In this case $\hat{\gamma}$ has the efficiency of the maximum likelihood estimator. If the order of the MA process is r we can test (see Hannan and Nicholls [1972], theorem 2) the hypothesis $\underline{\alpha} = \underline{a} = (a_0, a_1, \dots, a_r)$, by an asymptotic chi-square test. If the hypothesis is not rejected it will be then simpler to estimate $\underline{\gamma}$ by $\hat{\underline{\gamma}}$ than by $\tilde{\underline{\gamma}}$. We can also test (6.53) against (6.2) by means of a likelihood ratio test, the likelihoods being those corresponding to $\hat{\underline{\gamma}}$ and $\tilde{\underline{\gamma}}$, taking as degrees of freedom the difference in number of parameters used in estimating $\sum_h f_u(\lambda_h)$ under both hypothesis. It must be warned that the likelihoods are not necessarily nested for small samples and in any case it may be necessary to correct for degrees of freedom.

The loss in efficiency with respect to the case that the spectrum of the $\{u_t\}$ process is known can be seen from the fact that then in (6.49) we do not need to subtract $I_u(\lambda_h)/|\alpha(\lambda_h)|^2$ from $I_y(\lambda_h)$ and the asymptotic variance matrix can be obtained from the asymptotic variance matrix of

$$\frac{\sqrt{T}}{2M} \sum_h \left[\frac{1}{f_u(\lambda_h)} \right] I_{X^t u}(\lambda_h) .$$

Therefore, in the case of known spectrum the matrix Ψ must be added to the north-west corner of V , and we call \dot{V} the resulting matrix. The matrix \dot{V}^{-1} is the asymptotic variance matrix obtained when the errors follow a MA or AR process with known parameters. For the MA case this is already clear for the exposition above. For the p^{th} AR case there is no difficulty in proving that \dot{V} takes the form

$$\frac{1}{\sigma^2} \left[G_{X^t X^t} - \sum_{j=1}^p \rho_j \left\{ G_{X^t X_j^t} + G_{X_j^t X^t} \right\} + \sum_{\ell=1}^p \sum_{j=1}^p \rho_\ell \rho_j G_{X_\ell^t X_j^t} \right] ,$$

where G stands for the covariance matrix of the variables appearing as subscripts and the ρ_j 's are the parameters of the AR process.

We can now compare, from the point of view of efficiency, the spectral estimator $\hat{\gamma}$ with the time domain estimator of γ , $\tilde{\gamma}$ say, when the errors are AR and we need to estimate the parameters of the AR process. We do that for $k = r = 1$ and for an AR process of first order. Then the corresponding V matrix for $\sqrt{T}(\hat{\gamma} - \gamma)$ is

$$\left[\hat{V} - \begin{pmatrix} \int_{-\pi}^{\pi} \frac{1}{2\pi(1 + \alpha^2 + 2\alpha \cos \lambda)} d\lambda & 0 \\ 0 & 0 \end{pmatrix} \right]^{-1} \quad (6.59)$$

and the corresponding one for $\sqrt{T}(\tilde{\gamma} - \gamma)$ is, (see Maddala [1971])

$$\left[\hat{V} - \begin{pmatrix} \frac{1-\rho^2}{(1-\rho\alpha)^2} & 0 \\ 0 & 0 \end{pmatrix} \right]^{-1} \quad (6.60)$$

where ρ is the parameter of the $\{u_t\}$ AR process. Now

$$d_1 = \int_{-\pi}^{\pi} \frac{d\lambda}{2\pi(1+\alpha^2+2\alpha \cos \lambda)} = \frac{1}{1-\alpha^2} \quad , \quad (6.61)$$

using the fact that by the stationarity condition $|\alpha| < 1$.

Also

$$d_2 = \frac{1 - \rho^2}{(1-\rho\alpha)^2} = \frac{1}{1-\alpha^2} - \frac{(\alpha-\rho)^2}{(1-\rho\alpha)^2(1-\alpha^2)} \quad ,$$

therefore if $\rho \neq \alpha$, $d_1 > d_2$ and $\tilde{\gamma}$ will be efficient respect with $\hat{\gamma}$.

The model (6.1) for $r = k = 1$ takes the form

$$y_t = -\alpha y_{t-1} + \beta x_t + u_t \quad (6.62)$$

or
$$y_t = \beta \sum_{\ell=0}^{\infty} -\alpha^{\ell} x_{t-\ell} + v_t \quad , \quad (6.63)$$

where $v_t = \frac{1}{1+\alpha L} u_t$ and L is the lag operator. The model (6.63) is known as the geometric lag model and its estimation by spectral methods has been considered by Hannan [1965]. The basic regression equation considered is (6.63), defined with u_t instead of v_t and consequently the error term corresponding to the transformed model (6.62) is given by $u_t^+ = u_t + \alpha u_{t-1}$. In fact, since a general formulation for the specification of the error term is adopted, it is of no consequence for the estimation of α and β whether u_t or u_t^+ is taken as the basic random error. Thus, if we multiply by β the second equation in (4.3) in Hannan [1965] p. 216, the resulting formulae correspond to the ones in (6.45) here, where then $\hat{I}_u(\lambda_h)$ is an estimate of $f_u(\lambda_h)$. One obvious difference between the Hannan [1965] approach and the one followed in this section is the different sort of estimates used for the spectra of the variables. In passing, note that on multiplying the first order condition for α , by β , the estimates for $\underline{\gamma}$ are defined in terms of P^{-1} which is a consistent estimate of their asymptotic variance.

Starting from a different motivation, what we have done in this section is a generalization of the Hannan [1965] estimators. This generalization has been previously mentioned in Hannan and Nicholls [1972], section 3, where they also show the effects of assuming a general stationary structure (GSS) for the errors when they follow a MA process. The effort of going through the details of the derivation of $\hat{\underline{\gamma}}$ has been proved quite useful to understand the relation of the estimation methods assuming a general stationary structure (GSS) or a AR or MA process for $\{u_t\}$ when lagged endogenous variables are and are not present. Thus, and in contrast with the results in section IV, we have that with the presence of lagged endogenous variables in the model the assumption of a GSS for the errors when they are AR or MA results in, except for some special cases, a loss in efficiency with respect to the maximum likelihood estimators. The reason can be seen as follows.

The log likelihood of (6.1) can be expressed in a general form as

$$L^0 = c^0 - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_u(\lambda) d\lambda - \frac{1}{2\pi} \int_{-\pi}^{\pi} I_u(\lambda) f_u^{-1}(\lambda) d\lambda, \quad (6.64)$$

but if u_t is AR or MA we should add the restriction

$$r(\gamma) = \int_{-\pi}^{\pi} \log f_u(\lambda) d\lambda - \int_{-\pi}^{\pi} g(\underline{a}, \lambda) d\lambda = 0, \quad (6.65)$$

where $g(\underline{a}, \lambda)$ is an appropriate function of the true parameters defining the AR or MA process for $\{u_t\}$. Now the restrictions imposed by (6.65) will decrease the lower band for the variance matrix of the estimator of γ by (see Rothenberg [1973] Chapter 2) the positive definite matrix

$$v^{-1} q(\underline{a}, \lambda) [q'(\underline{a}, \lambda) v^{-1} q(\underline{a}, \lambda)]^{-1} q'(\underline{a}, \lambda) v^{-1} \quad (6.66)$$

where

$$q(\underline{a}, \lambda) = \left[\frac{\partial r(\gamma)}{\partial \gamma} \right] = c \int_{-\pi}^{\pi} f_u^{-1}(\lambda) f_{X^+u}(\lambda) d\lambda. \quad (6.67)$$

Now, if the model does not contain lagged endogenous variables X^+ is X and $q(\underline{a}, \lambda) = 0$ and nothing is gained asymptotically by imposing (6.65). On the other hand, if the model contains lagged endogenous variables $q(\underline{a}, \lambda)$ does not vanish and we reduce the asymptotic variance by considering the appropriate error structure.

As shown in Appendix the estimators obtained from the maximization of L_2 (i.e. $\hat{\gamma}$) are asymptotically equivalent to those derived from L_4 (i.e. following the approach of the section IV) denoted $\hat{\tilde{\gamma}}$ and defined by (6.49) without subtracting $\frac{I_u(\lambda_h)}{|\alpha(\lambda)^2|}$ from $I_y(\lambda_h)$. So for computational purposes it will be easier to compute $\hat{\tilde{\gamma}}$ than $\hat{\gamma}$. The fact is that if we do not impose the restriction (6.65) the endogenization of the lagged endogenous variables does not carry any loss in efficiency but - as we saw comparing $\hat{\tilde{\gamma}}$ with $\tilde{\gamma}$ and $\tilde{\tilde{\gamma}}$ - this is not the case if we correctly introduce the restriction.

The equations for $\hat{\gamma}$ are a general expression for the maximum likelihood estimators of γ when the $\{u_t\}$ is AR or MA, by replacing $\ddot{I}_u(\lambda_h)$ with the formula of $\hat{f}_u(\lambda_h)$ in terms of the parameters of the error process. Of course in those cases the mentioned general expression must be complemented with the first order conditions corresponding to the parameters of the error process and the iteration on both sets of equations is required to obtain the maximum likelihood estimators.

In the computation of the Spec LS estimators of (6.1) we generally have two alternatives, one of which is to compute $\hat{\gamma}$, that apart from the factor $(\ddot{I}_u(\lambda_h))^{-1}$ is linear. However, a linearization of $\hat{\gamma}$ will result in a loss in efficiency. This is because a consistent estimator of $f_u(\lambda_h)$ is not enough, due to the fact that some signals (the lagged endogenous variables) are correlated with the noise and they have not been replaced by adequate instruments. The other alternative is to compute $\tilde{\gamma}$ for which a consistent estimator of $f_u(\lambda_h)$ seems enough, because the signals have been replaced by instruments that are uncorrelated with the errors. The problem here is that $\tilde{\gamma}$ is non-linear on account of $(\ddot{I}_u(\lambda_h))^{-1}$ and because of $(|\alpha(\lambda_h)|^2)^{-1}$. Therefore if the $\hat{\alpha}$ we obtain is different from that used to compute $|\alpha(\lambda_h)|^2$ by non-trivial amounts it will pay to iterate by estimating $\ddot{I}_u(\lambda_h)$ and $|\alpha(\lambda_h)|^2$ using $\hat{\alpha}$. In the case of both $\hat{\gamma}$ and $\tilde{\gamma}$, and initial estimate for $\ddot{I}_u(\lambda_h)$ and for $\ddot{I}_u(\lambda_h)$ and $|\alpha(\lambda_h)|^2$ respectively, could be based on the results of an estimation of γ using a high order autoregressive process for the disturbances.

We have mentioned above that it will be simpler to compute $\hat{\gamma}$ than $\tilde{\gamma}$. Now taking advantage of the fact that the imaginary parts of the elements of P and p cancel out, when summing over the whole range or frequencies, the equation for $\hat{\gamma}$ can be further simplified. To do that, let us call

$$\theta_1 = \theta_c + i\theta_s \quad , \quad (6.68)$$

where θ_c is formed with the cosine part of the exponential and θ_s with sine part; and also ϕ_c is a $r \times r$ matrix with

$\cos [(k-\ell)\lambda_h]$ in the k, ℓ position;

$$\tilde{I}_{XY}(\lambda_h) = A_{XY}(\lambda_h) - i B_{XY}(\lambda_h) \quad (23) \quad (6.69)$$

and

$$\tilde{I}_{XX}(\lambda_h) = A_{XX}(\lambda_h) - i B_{XX}(\lambda_h) .$$

Thus the equation for $\hat{\tilde{Y}}$ can be written as follows

$$\hat{\tilde{Y}} = - \begin{bmatrix} \sum_{h=0}^M \delta_h \ddot{I}_u^{-1}(\lambda_h) I_y(\lambda_h) \phi_c & \sum_{h=0}^M \delta_h \ddot{I}_u^{-1}(\lambda_h) [\theta_c A'_{yX}(\lambda_h) + \theta_s B'_{yX}(\lambda_h)] \\ \sum_{h=0}^M \delta_h \ddot{I}_u^{-1}(\lambda_h) [A_{XY}(\lambda_h) \theta'_c + B_{XY}(\lambda_h) \theta'_s] & \sum_{h=0}^M \delta_h \ddot{I}_u^{-1}(\lambda_h) A_{XX}(\lambda_h) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{h=0}^M \delta_h (1/\ddot{I}_u(\lambda_h)) I_y(\lambda_h) \theta_c \\ \sum_{h=0}^M \delta_h (1/\ddot{I}_u(\lambda_h)) A_{XY}(\lambda_h) \end{bmatrix} , \quad (6.70)$$

where $\delta_h = 1$ except for $h = 0$ and M , that $\delta_h = \frac{1}{2}$.

(23) Note the $A_{XY}(\lambda_h)$ and $B_{XY}(\lambda_h)$ are vectors.

VII The Specfiml Method as applied to models with lagged endogenous variables.

In this section we derive the Specfiml estimates for a simultaneous equations model that contains r lagged values of the dependent variables, treated as predetermined, in addition to a set of k exogenous variables. Following the terminology used in Section IV this model can be expressed as:

$$B_0 Y'_1 + B_1 Y'_1 + \dots + B_r Y'_r + \Gamma Z' = U' \quad (7.1)$$

or using

$$A = (B_0 \ B_1 \ \dots \ B_r \ \Gamma) = (B_0 \ \Gamma^\dagger) \quad (7.2)$$

and

$$X = (Y \ Y_1 \ \dots \ Y_r \ Z) = (Y \ Y^\dagger \ Z) = (Y \ Z^\dagger) \quad (7.3)$$

as

$$AX' = B_0 Y' + \Gamma^\dagger Z^{\dagger'} = U' \quad (7.4)$$

For the vector process $\{u_t\}$ we use the assumptions adopted in Section IV, while for the exogenous variables $\{z_t\}$ we require condition A6.2. We also need condition A6.3 and that the model is identified. For the problem of identification we refer to Sargan [1972] and Hannan [1971]. A further requirement is that the autoregressive process involving the dependent variables is stable, i.e. the roots of

$$\det [B_0 + B_1 \xi + \dots + B_r \xi^r] = 0$$

lie outside of the unit circle.

VII.1 Alternative likelihood functions and first order conditions.

Arguing as in section IV we can write the log likelihood function as

$$\begin{aligned} \frac{L(Y', V)}{T} = L = c - \frac{m}{2T} \sum_{\lambda_h} \log \det F_U(\lambda_h) - \frac{m}{2T} \sum_{\lambda_h} \text{tr}\{F_U^{-1}(\lambda_h) A I_{XX}(\lambda_h) A'\} \\ + \log |\det B_0| \quad (7.5) \end{aligned}$$

and using

$$\hat{F}_U(\lambda_h) = I_U(\lambda_h) = AI_{XX}(\lambda_h)A' \quad (7.6)$$

the concentrated likelihood can be expressed as

$$L_c = c_1 + \log |\det B_0| - \frac{m}{2T} \sum_{\lambda_h} \log \det I_U(\lambda_h). \quad (7.7)$$

The expressions (7.5), (7.6) and (7.7) correspond to (4.17), (4.18) and (4.19), but now A and X are defined by (7.2) and (7.3)⁽²⁴⁾.

Now writing

$$\Pi = - B_0 \Gamma^\dagger \quad (7.8)$$

$$P' = (\Pi' \ ; \ I) \quad (7.8a)$$

we can deduce first order conditions for the Specfml estimators of the form:

$$\left[- \frac{1}{2M} \sum_{\lambda_h} \hat{F}_U^{-1}(\lambda_h) \hat{A}I_{XZ^\dagger}(\lambda_h) \hat{P}' \right]^{v_u} = 0. \quad (7.9)$$

If $r = 0$ (i.e. there are no lagged endogenous variables), we saw, that this is quite a convenient form for deriving the asymptotic distribution and the asymptotic variance matrix can be estimated by

$$\left[\frac{1}{2M} \sum_{\lambda_h} \hat{P}I_{Z^\dagger Z^\dagger}(-\lambda_h) \hat{P}' \otimes I_U^{-1}(\lambda_h) \right]^{u^{-1}} \quad (7.10)$$

which leads to the suggested spectral three stage least squares (Spec 3SLS) estimator given by

$$\left[\frac{1}{2M} \sum_{\lambda_h} \tilde{F}_U^{-1}(\lambda_h) \tilde{A}I_{XZ^\dagger}(\lambda_h) \tilde{P}' \right]^{v_u} = 0 \quad (7.11)$$

(24) In this section we follow the notation adopted in Section IV, except that in the latter case A, X, Π , P, etc. possessed simpler definitions, owing to the absence of lagged endogenous variables.

where \tilde{A} is the Spec 3SLS estimator based on consistent estimators of $F_U(\lambda_h)$ and P .

However if $r \neq 0$ (i.e. there are lagged endogenous variables), we have that:

(a) (7.9) does not provide a good basis for discussing asymptotic theory,

(b) (7.10) does not provide a consistent estimate of the asymptotic variance matrix and

(c) (7.11) does not give an \tilde{A} asymptotically equivalent to \hat{A} .

To do this we use an alternative form of the likelihood function.

Applying an unitary transformation to (7.1) by postmultiplying it by Q^* we get

$$B\tilde{w}_Y^* + \Gamma^\dagger \tilde{w}_{Z^\dagger}^* = \tilde{w}_U^*$$

and for a particular ω_j we have

$$B\tilde{w}_Y(\omega_j) + \Gamma^\dagger \tilde{w}_{Z^\dagger}(\omega_j) = \tilde{w}_U(\omega_j) . \quad (7.12)$$

Writing

$$B(\omega_j) = B_0 + B_1 \exp(-i\omega_j) + \dots + B_r \exp(-ir\omega_j) \quad (7.13)$$

it follows that

$$B(\omega_j) \tilde{w}_Y(\omega_j) + \Gamma^\dagger \tilde{w}_{Z^\dagger}(\omega_j) = \tilde{w}_U(\omega_j) + o(1/\sqrt{T}) , \quad (7.14)$$

and the contribution of the residual term (see Hannan [1970] p. 493) is negligible.

Now assuming (a) continuity and (b) that the spectra of the errors are constant within bands and using A6.3, L can be replaced by

$$L_1 = c - \frac{m}{2T} \sum_{\lambda_h} \log \det F_U(\lambda_h) - \frac{m}{2T} \sum_{\lambda_h} \text{tr} \{F_U^{-1}(\lambda_h) \ddot{I}_U(\lambda_h)\} + \frac{m}{T} \sum_{\lambda_h} \log |\det B(\lambda_h)| , \quad (7.15)$$

where

$$\ddot{I}_U(\lambda_h) = A(\lambda_h) I_{XX}(\lambda_h) A^*(\lambda_h) \quad , \quad (7.16)$$

$$\chi = (Y \ Z) \quad (7.17)$$

and

$$A(\lambda_h) = [B(\lambda_h) \ ; \ \Gamma] \quad (7.18)$$

To reconcile L_1 with L we proceed by an argument similar to one after (6.17).

Concentrating as before, but now using,

$$\hat{F}_U(\lambda_h) = \ddot{I}_U(\lambda_h) \quad (7.19)$$

we have

$$L_e = c_e + \frac{1}{2M} \sum_{\lambda_h} \log \det B(\lambda_h) - \frac{1}{4M} \sum_{\lambda_h} \log \det \ddot{I}_U(\lambda_h) \quad . \quad (7.20)$$

Now if we call

$$\phi = [\phi_j] = (\text{Vec } A)^u \quad , \quad (7.21)$$

we can differentiate L_e with respect to any elements of ϕ , and we get

$$\begin{aligned} \frac{\partial L_e}{\partial \phi_j} &= \frac{1}{2M} \sum_{\lambda_h} \text{tr} \left[B^{-1}(\lambda_h) \frac{\partial B'(\lambda_h)}{\partial \phi_j} \right] \\ &\quad - \frac{1}{2M} \sum_{\lambda_h} \text{tr} \left[\ddot{I}_U^{-1}(\lambda_h) A(\lambda_h) I_{XX}(\lambda_h) \frac{\partial A^*(\lambda_h)}{\partial \phi_j} \right] \quad . \quad (7.22) \end{aligned}$$

Now if we write

$$\phi(\lambda_h) = (I_p \ ; \ e^{i\lambda_h} I_p \ ; \ e^{2i\lambda_h} I_p \ ; \ \dots \ ; \ e^{ri\lambda_h} I_p) \quad (7.23)$$

so that

$$(B_0 \ B_1 \ \dots \ B_r) \ \phi^*(\lambda_h) = B(\lambda_h) \quad , \quad (7.24)$$

we have

$$\ddot{\Gamma}_U(\lambda_h) = A(\lambda_h)I_{\chi Y}(\lambda_h)B'(-\lambda_h) + A(\lambda_h)I_{\chi Z}(\lambda_h)\Gamma', \quad (7.25)$$

so that

$$\hat{B}'^{-1}(-\lambda_h) = \ddot{\Gamma}_U^{-1}(\lambda_h)\hat{A}(\lambda_h)I_{\chi Y}(\lambda_h) + \ddot{\Gamma}_U^{-1}(\lambda_h)\hat{A}(\lambda_h)I_{\chi Z}(\lambda_h)\hat{\Gamma}'\hat{B}'^{-1}(-\lambda_h) \quad (7.26)$$

and

$$\begin{aligned} \sum_{\lambda_h} \text{tr} \left[\hat{B}'^{-1}(-\lambda_h) \frac{\partial B'(-\lambda_h)}{\partial \phi_p} \right] &= \sum_{\lambda_h} \text{tr} \left[\ddot{\Gamma}_U^{-1}(\lambda_h)\hat{A}(\lambda_h)I_{\chi Y}(\lambda_h) \frac{\partial B'(-\lambda_h)}{\partial \phi_p} \right] \\ &+ \sum_{\lambda_h} \text{tr} \left[\ddot{\Gamma}_U^{-1}(\lambda_h)\hat{A}(\lambda_h)I_{\chi Z}(\lambda_h)\hat{\Gamma}'\hat{B}'^{-1}(-\lambda_h) \frac{\partial B'(-\lambda_h)}{\partial \phi_p} \right]. \end{aligned} \quad (7.27)$$

Therefore

$$\begin{aligned} \frac{\partial L_e}{\partial \phi_p} &= \frac{1}{2M} \sum_{\lambda_h} \text{tr} \left[\ddot{\Gamma}_U^{-1}(\lambda_h)\hat{A}(\lambda_h)I_{\chi Z}(\lambda_h)\hat{\Gamma}'\hat{B}'^{-1}(-\lambda_h) \frac{\partial B'(-\lambda_h)}{\partial \phi_p} \right] \\ &- \frac{1}{2M} \sum_{\lambda_h} \text{tr} \left[\ddot{\Gamma}_U^{-1}(\lambda_h) \hat{A}(\lambda_h) I_{\chi Z} \frac{\partial \hat{\Gamma}'}{\partial \phi_p} \right]. \end{aligned} \quad (7.28)$$

Then noting that

$$\frac{\partial B'(-\lambda_h)}{\partial \phi_p} = \Phi(\lambda_h) \frac{\partial B^+}{\partial \phi_p}, \quad (7.29)$$

where

$$B^+ = (B_0 \ B_1 \ \dots \ B_r), \quad (7.30)$$

we define

$$P'(\lambda_h) = (-\Gamma'B'^{-1}(-\lambda_h) \Phi(\lambda_h) : I) \quad (25) \quad (7.31)$$

and we can write

$$\frac{\partial L_e}{\partial \phi_p} = - \frac{1}{2M} \sum_{\lambda_h} \text{tr} \left[\ddot{\Gamma}_U^{-1}(\lambda_h)\hat{A}(\lambda_h)I_{\chi Z}(\lambda_h)P'(\lambda_h) \frac{\partial A'}{\partial \phi_p} \right]. \quad (7.32)$$

(25) This is a generalization of the $P(\lambda_h)$ defined in (6.30).

If the A matrix has only the usual-standardizing restrictions then the first order conditions can be written as

$$-\frac{1}{2M} \left\{ \sum_{\lambda_h} \ddot{I}_U^{-1}(\lambda_h) \hat{A}(\lambda_h) I_{XZ}(\lambda_h) P'(\lambda_h) \right\}^{V_U} = 0 \quad (7.33)$$

Now (7.33) is a convenient form for deriving an asymptotic error variance matrix and suggesting a suitable 3SLS approximation. Indeed we can immediately write down a Spec 3SLS estimator of the form

$$\left\{ \sum_{\lambda_h} \ddot{I}_U^{-1}(\lambda_h) \tilde{A}(\lambda_h) I_{XZ}(\lambda_h) \tilde{P}'(\lambda_h) \right\}^{V_U} = 0 \quad , \quad (7.34)$$

or since

$$A(\lambda_h) W_X^*(\lambda_h) = A W_X^*(\lambda_h) + R_0 \quad , \quad (7.35)$$

where R_0 is asymptotically negligible, as

$$\left\{ \sum_{\lambda_h} \ddot{I}_U^{-1}(\lambda_h) \tilde{A} I_{XZ}(\lambda_h) \tilde{P}'(\lambda_h) \right\}^{V_U} = 0 \quad , \quad (7.36)$$

where $\tilde{I}_U^{-1}(\lambda_h)$ and $\tilde{P}'(\lambda_h)$ could be based on estimating A by using a high order vector autoregressive models and using them to generate corresponding estimates of $I_U(\lambda_h)$ and $P(\lambda_h)$. If the \tilde{A} we obtain are different from those used to compute $\tilde{I}_U^{-1}(\lambda_h)$ and $\tilde{P}'(\lambda_h)$ by non-trivial amounts then it will pay to iterate by estimating $\tilde{I}_U^{-1}(\lambda_h)$ and $\tilde{P}'(\lambda_h)$ using \tilde{A} .

VII.2 The asymptotic distribution of the Specfiml estimators.

To obtain the asymptotic distribution of \hat{A} we have

$$A I_{XZ}(\lambda_h) = I_{UZ}(\lambda_h)$$

so that

$$\frac{1}{2M} \left[\sum_{\lambda_h} \hat{P}(\lambda_h) I_{ZX}(-\lambda_h) \otimes \ddot{I}_U^{-1}(\lambda_h) \right]^u \sqrt{T} (\hat{\phi} - \phi)$$

$$= \frac{\sqrt{m}}{\sqrt{2M}} \left[\begin{array}{c} \Sigma \\ \lambda_h \end{array} \hat{P}(\lambda_h) \otimes \ddot{I}_U^{-1}(\lambda_h) \text{Vec } I_{UZ}(\lambda_h) \right]^u . \quad (7.37)$$

Note also that using (7.14) and the results in Hannan [1970] chapter seven, we have

$$\begin{aligned} & \underset{T}{P} \lim_{T \rightarrow \infty} \frac{1}{2M} \left[\begin{array}{c} \Sigma \\ \lambda_h \end{array} \hat{P}(\lambda_h) I_{ZX}(-\lambda_h) \otimes \ddot{I}_U^{-1}(\lambda_h) \right]^u \\ &= \frac{1}{2\pi} \left[\int_{\pi}^{\pi} P(\lambda) F_{ZZ}(-\lambda) P^*(\lambda) \otimes F_U^{-1}(\lambda) d\lambda \right]^u . \quad (7.38) \end{aligned}$$

Now, under the present assumptions and arguing along the lines of Hannan [1970] p. 492-495 we deduce that asymptotically the RHS of (7.37) can be replaced by

$$- \frac{\sqrt{m}}{\sqrt{2M}} \left[\begin{array}{c} \Sigma \\ \lambda_h \end{array} \hat{P}(\lambda_h) \otimes I_U^{-1}(\lambda_h) \text{Vec } I_{UZ}(\lambda_h) \right]^u , \quad (7.39)$$

which by the Hannan theorem (see Hannan [1970] p. 442 and 487-492) has an asymptotic multivariate normal distribution with variance-covariance matrix given by the RHS of (7.38). Therefore $\sqrt{T}(\hat{\phi} - \phi)$ has an asymptotic multivariate normal distribution with zero means and variance-covariance matrix given by the inverse of the RHS of (7.38) which can be consistently estimated by

$$\left[\frac{1}{2M} \int_{\pi}^{\pi} \hat{P}(\lambda_h) I_{ZZ}(-\lambda_h) \hat{P}^*(\lambda_h) \otimes \ddot{I}_U^{-1}(\lambda_h) d\lambda \right]^u^{-1} , \quad (7.40)$$

i.e. minus the inverse of the Hessian corresponding to L_e .

The result can be interpreted by considering a stationary stochastic process defining the random part (η_t) of y_t from

$$B_0 \eta_t + B_1 \eta_{t-1} + \dots + B_r \eta_{t-r} = u_t .$$

If we denote the systematic part of y_t by y_{st} so that

$$y_{st} = y_t - \eta_t \quad ,$$

we have that

$$B_0 y_{st} + B_1 y_{s(t-1)} + \dots + B_r y_{s(t-r)} + \Gamma z_t = 0 \quad .$$

Now writing

$$x'_{st} = (y'_{st}, y'_{s(t-1)}, \dots, y'_{s(t-r)}, z'_t)$$

and

$$Y'_s = (y_{s1}, y_{s2}, \dots, y_{sT})$$

$$X'_s = (x_{s1}, x_{s2}, \dots, x_{sT}) \quad ,$$

we have that

$$F_{Y_s Y_s}(\lambda) = B^{-1}(\lambda) \Gamma F_{ZZ}(\lambda) \Gamma' B^{-1*}(\lambda)$$

$$F_{Y_s Z}(\lambda) = B^{-1}(\lambda) \Gamma F_{ZZ}(\lambda) \quad .$$

Thus

$$F_{X_s X_s}(\lambda) = \begin{bmatrix} \Phi^*(\lambda) F_{Y_s Y_s}(\lambda) \Phi(\lambda) & \Phi^*(\lambda) F_{Y_s Z}(\lambda) \\ F_{ZY_s}(\lambda) \Phi(\lambda) & F_{ZZ}(\lambda) \end{bmatrix}$$

$$= \begin{bmatrix} -\Phi^*(\lambda) B^{-1}(\lambda) \Gamma \\ I \end{bmatrix} F_{ZZ}(\lambda) \begin{bmatrix} -\Gamma' B^{-1}(-\lambda) \Phi(\lambda) \\ \vdots \\ I \end{bmatrix}$$

$$= P(-\lambda) F_{ZZ}(\lambda) P'(\lambda) \quad .$$

We can replace (7.38) by

$$\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} (F_{X_s X_s}(-\lambda) \otimes F_U^{-1}(\lambda)) d\lambda \right]^u \quad (7.40a)$$

If $r = 0$ (i.e. there are no lagged endogenous variables) the asymptotic error variance collapses to the one derived in Hannan and Terrell [1973]. In section IV we have shown that in such a case there is no loss in efficiency by assuming a GSS for the errors when they in fact follow an AR process.

Note that (7.40) also gives an estimate of the asymptotic matrix of the errors in the estimators obtained by maximizing (7.5) which leads us to the problems of computation and so note that if the spectral estimators were to be computed from the first order conditions, it will be simpler to use (7.9) rather than (7.33) and that asymptotically there is no difference⁽²⁶⁾.

Note that the estimates obtained by maximizing L_e have the same limitations as those obtained by maximizing L , i.e. we must have $m > p(r + 1) + k$. Moreover if we look at the similar conditions for the Spec 3SLS defined by (7.35), we require also $\tilde{I}_U^{-1}(\lambda_h)$ to be non-singular, which requires that $p < m$. We obviously require that the estimate of A , \tilde{A} , on which the estimate of $\tilde{P}(\lambda_h)$ is based should lead to a non-singular $\tilde{B}(\lambda_h)$ for all h , and also that the matrix

$$\begin{bmatrix} \Sigma & \tilde{P}(\lambda_h) & I_{ZX}(-\lambda_h) & \otimes & \tilde{I}_U^{-1}(\lambda_h) \end{bmatrix}^u$$

should be non-singular. It seems that all we need is that the sample should be sufficiently large to enable \tilde{A} to be well defined.

It is interesting to compare the asymptotic properties of the Spectiml estimators with the asymptotic properties of the maximum likelihood estimators when the errors follow a vector moving average process. There is no difficulty in proving that the results obtained, in a similar comparison, in Section VI generalized here, for the simultaneous model. In particular:

(26) This is true provided the autoregressive process involving the dependent variables is stable.

- I The Specfiml estimators have the efficiency of the maximum likelihood estimators if the parameters of the error moving average process are equal to the parameters of the autoregressive process of the endogenous variables;
- II if they are not equal the Specfiml estimators result in a loss in efficiency compared to the maximum likelihood estimators;
- III provided that the disturbance process is MA, an estimation of A - for instance along the lines of Nicholls [1975] - overstating the order of the MA process will be efficient with respect to the Specfiml estimator. Of course if the error process is of higher order or is not of a MA form, the Nicholls [1975] procedure does not lead to consistent estimators of A unlike the Specfiml procedure which does anyway.

Similar conclusions are obtained when comparing the Specfiml estimator with the maximum likelihood (ML) estimators if the errors follow an AR process. In order to prove them, the next section is devoted to the derivation of the asymptotic variance matrix (AVM) of the ML estimators in the AR case.

VIII The Asymptotic variance matrix of the Structural Estimators
when the Errors Follow an AR process.

The model considered here is the one set up in the previous section but now the errors are generated by a vector autoregressive model of the form

$$\underline{u}_t - R_1 \underline{u}_{t-1} \dots - R_r \underline{u}_{t-r} = \underline{\varepsilon}_t .$$

We find it convenient to work with the reduced form, which we write

$$\underline{y}_t - \Pi \underline{z}_t^\dagger = \underline{v}_t , \quad (8.1)$$

where Π and \underline{z}_t^\dagger are as defined in (7.8) and (7.3) respectively and

$$\underline{v}_t - S_1 \underline{v}_{t-1} \dots - S_r \underline{v}_{t-r} = \underline{\varepsilon}_t^0 \quad (8.2)$$

$$\underline{v}_t = B_0^{-1} \underline{u}_t$$

$$\underline{\varepsilon}_t^0 = B^{-1} \underline{\varepsilon}_t . \quad (8.3)$$

Then we have

$$\text{var}(\underline{\varepsilon}_t) = G$$

and

$$\text{var}(\underline{\varepsilon}_t^0) = G^0 .$$

To find the AVM of $\sqrt{T}(\hat{A} - A)$, \hat{A} is the time domain autoregressive ML estimator, we consider the first and second derivatives of the likelihood function. We can write the concentrated likelihood function as

$$L_r = - \frac{1}{2} T \log \det (E^0, E^0) ,$$

where E^0 is a matrix formed with the column vectors $\underline{\varepsilon}_t^0$ and we assume that the $\underline{\varepsilon}_t^0$ have been expressed as functions of some set of parameters ϕ , eliminating the S_ϕ , by using the equations

$$\sum_{k=0}^{\tilde{r}} V_j^k V_k S_k' = 0 \quad j = 1, \dots, \tilde{r} \quad , \quad (8.4)$$

where V_k is the matrix whose rows are v_{t-k}^t , $t = 1, \dots, T$ and $S_0 = -I$. Then $\varepsilon_{\tilde{t}}^0$ is determined from $v_{\tilde{t}-s}$ using (8.2) and (8.4) and then is expressed as a function of A from (8.1) and (7.8). We assume that A depends upon the n parameters of the vector ϕ .

Now since equations (8.4) are equivalent to $\frac{\partial L_r}{\partial S_s} = 0$, we can write the first derivatives of L_r with respect to ϕ_n , ignoring the dependence of S_s on ϕ .

Thus

$$\begin{aligned} \frac{\partial L_r}{\partial \phi_n} &= -T \operatorname{tr} \left[(E^0, E^0)^{-1} \left(E^0, \frac{\partial E^0}{\partial \phi_n} \right) \right] \\ &= -T \operatorname{tr} \left[\hat{G}^0^{-1} \left(E^0, \frac{\partial E^0}{\partial \phi_n} \right) \right] = 0 \quad . \quad (8.5) \end{aligned}$$

And for $\frac{\partial E^0}{\partial \phi_n}$, again ignoring the dependence of S_s on ϕ_n , and differentiating (8.2)

$$\frac{\partial E^0}{\partial \phi_n} = - \sum_{s=0}^{\tilde{r}} \frac{\partial v_s}{\partial \phi_n} S_s' = \sum_{s=0}^{\tilde{r}} Z_s^+ \left(\frac{\partial \Pi}{\partial \phi_n} \right)' S_s' \quad .$$

Now from (7.8)

$$\frac{\partial \Pi}{\partial \phi_n} = -B_0^{-1} \frac{\partial B_0}{\partial \phi_n} \Pi - B_0^{-1} \frac{\partial \Gamma^+}{\partial \phi_n} = -B_0^{-1} \left(\frac{\partial A}{\partial \phi_n} \right) P \quad ,$$

where P was defined in (7.8a).

For the Hessian we have

$$\frac{\partial^2 L_r}{\partial \phi_n \partial \phi_q} = -T \operatorname{tr} \left[\frac{\partial}{\partial \phi_q} (\hat{G}^0^{-1}) \left(E^0, \frac{\partial E^0}{\partial \phi_n} \right) \right]$$

$$- T \operatorname{tr} \left[\hat{G}^{0-1} \left(E^0, \frac{\partial^2 E^0}{\partial \phi_n \partial \phi_q} \right) \right] - T \operatorname{tr} \left[\hat{G}^{0-1} \left(\frac{\partial E^0}{\partial \phi_q} \right) \left(\frac{\partial E^0}{\partial \phi_n} \right) \right].$$

Note that the first two terms are $O(\sqrt{T})$ since E^0 in each occurs multiplied by Z'_s when explicit expression for $\partial E^0 / \partial \phi_n$ and $\partial^2 E^0 / \partial \phi_n \partial \phi_q$ are substituted, and so can be neglected compared with the last term which is $O(T)$. In this term we must take account of the dependence in $\frac{\partial E^0}{\partial \phi_q}$ of S_s on ϕ_q .

Differentiating (8.4) we have

$$\frac{\partial \Pi}{\partial \phi_n} \sum_{k=0}^{\tilde{r}} (Z_j^\dagger V_k) S'_k + \sum_{k=0}^{\tilde{r}} (V_j^\dagger Z_k^\dagger) \left(\frac{\partial \Pi}{\partial \phi_n} \right)' S'_k - \sum_{k=1}^{\tilde{r}} (V_j^\dagger V_k) \frac{\partial S'_k}{\partial \phi_n} = 0.$$

Now

$$\sum_{k=0}^{\tilde{r}} (Z_j^\dagger V_k) S'_k = Z_j^\dagger E^0 = O(\sqrt{T}),$$

and so the first term is negligible. So now write

$$\hat{G}_{V^+} = \frac{1}{T} \begin{pmatrix} V_r^\dagger V_{\tilde{r}} & V_r^\dagger V_{(\tilde{r}-1)} \cdots V_r^\dagger V_1 \\ \cdots & \cdots \cdots \cdots \\ \cdots & \cdots \cdots \cdots \\ V_1^\dagger V_{\tilde{r}} & V_1^\dagger V_1 \end{pmatrix}, \quad K_n = \frac{1}{T} \begin{pmatrix} \sum_{k=0}^{\tilde{r}} V_r^\dagger Z_k^\dagger \left(\frac{\partial \Pi}{\partial \phi_n} \right)' S'_k \\ \cdots \cdots \cdots \\ \sum_{k=0}^{\tilde{r}} V_1^\dagger Z_k^\dagger \left(\frac{\partial \Pi}{\partial \phi_n} \right)' S'_k \end{pmatrix}.$$

Then we have

$$\begin{pmatrix} \frac{\partial S'_r}{\partial \phi_n} \\ \cdots \\ \cdots \\ \frac{\partial S'_1}{\partial \phi_n} \end{pmatrix} = \hat{G}_{V^+}^{-1} K_n$$

and

$$\frac{dE^{0'}}{d\phi_q} = \frac{\partial E^{0'}}{\partial \phi_q} - \sum_{k=1}^{\tilde{r}} \left(\frac{\partial S_k}{\partial \phi_q} \right) V_k' = \frac{\partial E^{0'}}{\partial \phi_q} - K_q' \hat{G}_{V^+}^{-1} \begin{bmatrix} V_r' \\ \vdots \\ V_1' \end{bmatrix},$$

so that

$$\begin{aligned} \frac{1}{T} \frac{\partial^2 L}{\partial \phi_n \partial \phi_q} &= - \operatorname{tr} \left[(E^{0'}, E^0)^{-1} \begin{bmatrix} \frac{\partial E^{0'}}{\partial \phi_q} \\ \frac{\partial E^0}{\partial \phi_n} \end{bmatrix} \right] \\ &+ \operatorname{tr} \left[(E^{0'}, E^0)^{-1} K_q' \hat{G}_{V^+}^{-1} \begin{bmatrix} V_r' \\ \vdots \\ V_1' \end{bmatrix} \sum_{s=0}^{\tilde{r}} Z_s^+ \left(\frac{\partial \Pi}{\partial \phi_n} \right)' S_s' \right]. \\ &= - \operatorname{tr} \left\{ \left[\frac{E^{0'}, E^0}{T} \right]^{-1} \sum_k \sum_s \left[S_k \left(\frac{\partial \Pi}{\partial \phi_q} \right) \left(\frac{Z_k^+, Z_s^+}{T} \right) \left(\frac{\partial \Pi}{\partial \phi_n} \right)' S_s' \right] \right\} \\ &+ \operatorname{tr} \left[\left[\frac{E^{0'}, E^0}{T} \right]^{-1} K_q' \hat{G}_{V^+}^{-1} K_n \right]. \end{aligned} \quad (8.5a)$$

Now write

$$\Xi_\phi = - \left[p \lim \left(\frac{\partial^2 L_r}{\partial \phi_n \partial \phi_q} / T \right) \right] = \left[\xi_\phi^{nq} \right] = \Omega_\phi^{-1},$$

and

$$\begin{aligned} G_{VV}(\tau) &= p \lim \frac{V_{k+\tau}' V_k}{T} \\ G_{VZ^+}(\tau) &= p \lim \frac{V_{k+\tau}' Z_k^+}{T} \\ G_{Z^+Z^+}(\tau) &= p \lim \frac{Z_{k+\tau}^+ Z_k^+}{T}. \end{aligned}$$

Then we have

$$\begin{aligned} \xi_{\phi}^{nq} = & \text{tr} \left\{ G^{o-1} \sum_{k=0}^{\tilde{r}} \sum_{s=0}^{\tilde{r}} \left[S_k \left(\frac{\partial \Pi}{\partial \phi_q} \right) G_{Z^+Z^+(k-s)} \left(\frac{\partial \Pi}{\partial \phi_n} \right)' S'_s \right] \right\} \\ & - \text{tr} \left[G^{o-1} \tilde{K}'_q G_{V^+(\tilde{r})}^{-1} \tilde{K}_n \right], \end{aligned} \quad (8.6)$$

where

$$\tilde{K}'_q = \underset{T \rightarrow \infty}{p \lim} K_q = \begin{pmatrix} \sum_k G_{VZ^+(\tilde{r}-k)} \left(\frac{\partial \Pi}{\partial \phi_q} \right)' S'_k \\ \dots \\ \sum_k G_{VZ^+(1-k)} \left(\frac{\partial \Pi}{\partial \phi_q} \right)' S'_k \end{pmatrix},$$

$$G_{V^+(\tilde{r})}^{-1} = \underset{T \rightarrow \infty}{p \lim} \hat{G}_{V^+}^{-1} = \begin{pmatrix} G_{VV}(0) & G_{VV}(1) & \dots & G_{VV}(\tilde{r}-1) \\ G_{VV}(-1) & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ G_{VV}(-\tilde{r}+1) & \dots & \dots & G_{VV}(0) \end{pmatrix}^{-1}.$$

Now we consider the case where the order of the AR process is r^+ , but we mistakenly assume in estimation that the order is $\tilde{r} > r^+$, and in particular we consider the limit of ξ_{ϕ}^{nq} as $\tilde{r} \rightarrow \infty$ with $S_s, s \leq r^+$ remaining constant and $S_s = 0, s > r^+$.

First we note that the increasing \tilde{r} in general increases the AVM.

Since the first term in ξ_{ϕ}^{nq} is dependent of \tilde{r} , given $S_s = 0, s > r^+$, we consider only the second term and if we have that minus the second term is increasing in \tilde{r} , then the AVM (Ω_{ϕ}) will also be increasing.

More formally write $\Omega_{\phi}^{\tilde{r}}$ for the AVM corresponding to autoregressive order \tilde{r} , then if and only if

$$\Omega_{\phi}^{\tilde{r}+1} \geq \Omega_{\phi}^{\tilde{r}}$$

we have

$$(\Omega_{\phi}^{\tilde{r}+1})^{-1} \leq (\Omega_{\phi}^{\tilde{r}})^{-1} \quad ,$$

and then for an arbitrary constant vector $\underline{\alpha}$ we have

$$\underline{\alpha}' \left[(\Omega_{\phi}^{\tilde{r}+1})^{-1} \right] \underline{\alpha} \leq \underline{\alpha}' \left[(\Omega_{\phi}^{\tilde{r}})^{-1} \right] \underline{\alpha} \quad .$$

Now write

$$\Delta = \sum_{p=1}^n \alpha_p \frac{\partial \Pi}{\partial \phi_p}$$

and

$$\Delta_{\tilde{r}} = \begin{bmatrix} \tilde{r} \\ \sum_{k=0} G_{VZ}^+(\tilde{r}-k) \Delta' S'_k \\ \tilde{r} \\ \sum_{k=0} G_{VZ}^+(\tilde{r}-k-1) \Delta' S'_k \\ \vdots \\ \tilde{r} \\ \sum_{k=0} G_{VZ}^+(1-k) \Delta' S'_k \end{bmatrix} \quad .$$

Then

$$\text{tr} \left[G^{o-1} \sum_p \sum_q \alpha_p \tilde{K}'_p G^{-1}_{V^+} \tilde{K}_q \alpha_q \right] = \text{tr} \left[G^{o-1} \Delta'_{\tilde{r}} G^{-1}_{V^+(\tilde{r})} \Delta_{\tilde{r}} \right] \quad ,$$

and if this increases with \tilde{r} , then so does the AVM, but now it is clear that

$$\Delta'_{(\tilde{r}+1)} G^{-1}_{V^+(\tilde{r}+1)} \Delta_{(\tilde{r}+1)} \geq \Delta'_{\tilde{r}} G^{-1}_{V^+(\tilde{r})} \Delta_{\tilde{r}}$$

by using the partitioned inverse of $G_{V^+(\tilde{r}+1)}$. The condition for strict inequality is rather complex, so that we now concentrate on the limit as $\tilde{r} \rightarrow \infty$.

The first term in (8.6) is relatively simple to express in spectral terms. We have that this term is, (note that $S_k = 0$

if $k > r^+$)

$$\begin{aligned}
 & \text{tr} \left\{ G^{0-1} \begin{matrix} r^+ & r^+ \\ \Sigma & \Sigma \\ k=0 & s=0 \end{matrix} \left[S_k \left(\frac{\partial \Pi}{\partial \phi_q} \right) G_{Z^+Z^+(k-s)} \left(\frac{\partial \Pi}{\partial \phi_n} \right)' S'_s \right] \right\} \\
 &= \int_{-\pi}^{\pi} \text{tr} \left[G^{0-1} \begin{matrix} \Sigma & \Sigma \\ k & s \end{matrix} \exp[i(s-k)\lambda] S_k \left(\frac{\partial \Pi}{\partial \phi_q} \right) F_{Z^+Z^+(\lambda)} \left(\frac{\partial \Pi}{\partial \phi_n} \right)' S'_s \right] d\lambda \\
 &= \int_{-\pi}^{\pi} \text{tr} \left\{ \begin{bmatrix} r^+ \\ \Sigma \\ s=0 \end{bmatrix} S'_s \exp(is\lambda) \right\} G^{0-1} \begin{bmatrix} r^+ \\ \Sigma \\ k=0 \end{bmatrix} S_k \exp(-ik\lambda) \\
 & \quad \left(\frac{\partial \Pi}{\partial \phi_q} \right) F_{Z^+Z^+(\lambda)} \left(\frac{\partial \Pi}{\partial \phi_n} \right)' \right\} d\lambda
 \end{aligned}$$

where $F_{Z^+Z^+(\lambda)}$ is the matrix spectral density corresponding to $G_{Z^+Z^+(\tau)}$, and now writing $F_{VV}(\lambda)$ as the spectral density matrix corresponding to $G_{VV}(\tau)$, we have

$$(2\pi) F_{VV}(\lambda) = \begin{bmatrix} r^+ \\ \Sigma \\ k=0 \end{bmatrix} S_k \exp(-ik\lambda) \right]^{-1} G^0 \begin{bmatrix} r^+ \\ \Sigma \\ s=0 \end{bmatrix} S'_s \exp(is\lambda) \right]^{-1}$$

so that the expression above can be written

$$\frac{1}{2\pi} \left(\text{vec} \frac{\partial \Pi}{\partial \phi_n} \right)' \left[\int_{-\pi}^{\pi} \left(F_{Z^+Z^+(\lambda)} \otimes F_{VV}^{-1}(-\lambda) \right) d\lambda \right] \text{vec} \left(\frac{\partial \Pi}{\partial \phi_q} \right). \quad (8.7)$$

This, of course, equal to the usual formula for the AVM when $r = 0$ (i.e. no lagged endogenous variables).

Now considering the second term of (8.6), as $\tilde{r} \rightarrow \infty$, it is necessary to evaluate $G_{V^+}^{-1}$, and we note that this a Toeplitz matrix, so that we can consider a $(\tilde{r} \times \tilde{r})$ unitary transformation

matrix of the form

$$Q_{\tilde{r}} = \left[\frac{1}{\sqrt{\tilde{r}}} \left(\exp(2\pi ijk/\tilde{r}) \right) \right]_{j,k} \quad j,k = 0, \dots, \tilde{r}-1.$$

Then as $\tilde{r} \rightarrow \infty$

$$(Q_{\tilde{r}}^* \otimes I_p) G_{V^+} (Q_{\tilde{r}} \otimes I_p) = H_{\tilde{r}} \quad ,$$

tends to a block diagonal matrix with diagonal blocks equal to $2\pi F_{VY}(2\pi ij/\tilde{r})$.

Now to simplify \tilde{K}_q we write

$$G_{VZ^+}(s-k) = (G_{VY}(s-k-1) : G_{VY}(s-k-2) \dots G_{VY}(s-k-r) : 0)$$

and then approximate

$$G_{VY}(s-k-s^*) = \frac{2\pi}{\tilde{r}} \sum_{j=0}^{\tilde{r}-1} \exp\left\{2\pi i(k+s^*-s)j/\tilde{r}\right\} F_{VY}(2\pi ij/\tilde{r}) + o(1/\tilde{r}) \quad .$$

Thus \tilde{K}_q can be approximated by

$$\begin{aligned} & (2\pi/\tilde{r}) \sum_{j=0}^{\tilde{r}-1} \sum_{k=0}^{\tilde{r}} \exp(2\pi ikj/\tilde{r}) \left[\left[\begin{array}{c} 1 \\ \exp(2\pi ij/\tilde{r}) \\ \vdots \\ \exp(2\pi ij(\tilde{r}-1)/\tilde{r}) \end{array} \right] \left[\exp(2\pi ij/\tilde{r}) \dots \right. \right. \\ & \left. \left. \dots \exp(2\pi ijr/\tilde{r}) \right] \otimes F_{VY}(2\pi ij/\tilde{r}) \right] \left[\begin{array}{c} \partial \Pi_1 \\ \partial \Phi_q \end{array} \right]' S_k' \quad , \end{aligned}$$

where $\Pi = (\Pi_1 : \Pi_2)$ and $\Pi_2 = (-B_0^{-1} \Gamma)$.

Then

$$G_{V^+}^{-1} \tilde{K}_q = (2\pi/\tilde{r}) \sum_{j=0}^{\tilde{r}-1} \sum_{k=0}^{\tilde{r}} \exp(2\pi ijk/\tilde{r}) \left[Q_{\tilde{r}} \otimes I_p \right] H_{\tilde{r}}^{-1} \left[Q_{\tilde{r}}^* \otimes I_p \right]$$

$$\left[\left[\begin{array}{c} 1 \\ \vdots \\ \exp(2\pi i j (\tilde{r}-1)/\tilde{r}) \end{array} \right] \left[\exp(2\pi i j / \tilde{r}) \dots \exp(2\pi i j r / \tilde{r}) \right] \right]$$

$$\otimes F_{VY}(2\pi i j / r) \left[\begin{array}{c} \left(\frac{\partial \Pi_1}{\partial \phi_q} \right)' \\ S'_k \end{array} \right] + o(1/\tilde{r})$$

and note that

$$Q_{\tilde{r}}^* \left[\begin{array}{c} 1 \\ \vdots \\ \exp(2\pi i j (\tilde{r}-1)/\tilde{r}) \end{array} \right] = \underline{e}_j \sqrt{\tilde{r}} ,$$

where \underline{e}_j has all its elements zero, except for a one in the j^{th} position.

Note that we can then consider $\tilde{K}'_n(Q_{\tilde{r}} \otimes I_p) = \tilde{K}^*(Q_{\tilde{r}} \otimes I_r)$ in the same way and that the product is only non-zero, if the j on the left hand side is the same as the j on the right hand side. It follows if we write $\tilde{\omega}_j = 2\pi i j / \tilde{r}$

$$\tilde{K}'_n G_{V+}^{-1} \tilde{K}_q \approx (2\pi/\tilde{r}) \sum_{j=0}^{\tilde{r}-1} \sum_{k=0}^{\tilde{r}} \sum_{s=0}^{\tilde{r}} \exp(i(\tilde{\omega}_k - \tilde{\omega}_s)S_s) \frac{\partial \Pi_1}{\partial \phi_n}$$

$$\left[\left[\begin{array}{c} \exp(-\tilde{\omega}_j) \\ \vdots \\ \exp(-r\tilde{\omega}_j) \end{array} \right] \left[\exp(\tilde{\omega}_j) \dots \exp(r\tilde{\omega}_j) \right] \otimes F_{YV}(\tilde{\omega}_j) F_{VV}^{-1}(\tilde{\omega}_j) F_{VY}(\tilde{\omega}_j) \right]$$

$$\left(\frac{\partial \Pi_1}{\partial \phi_q} \right)' S'_k .$$

Thus the limit as $\tilde{r} \rightarrow \infty$ is

$$\int_0^{2\pi} \left[\begin{array}{c} r^+ \\ \Sigma \\ s=0 \end{array} S_s \exp(-is\omega) \right] \left[\frac{\partial \Pi_1}{\partial \phi_n} \right] \left\{ \left[\begin{array}{c} \exp(-i\omega) \\ \vdots \\ \exp(-ri\omega) \end{array} \right] \left[\exp(i\omega) \dots \exp(ri\omega) \right] \right\} \otimes$$

$$F_{YV}(\omega) F_{VV}^{-1}(\omega) F_{VY}(\omega) \left\{ \left[\frac{\partial \Pi_1}{\partial \phi_n} \right] \left[\begin{array}{c} r^+ \\ \Sigma \\ k=0 \end{array} S'_k \exp(ki\omega) \right] \right\} d\omega . \quad (8.8)$$

Thus $\lim_{\tilde{r} \rightarrow \infty} \text{tr} (G^{0-1} \tilde{K}'_q G^{-1} \tilde{K}_n) =$

$$\int_{-\pi}^{\pi} \text{tr} \left[\begin{array}{c} r^+ \\ \Sigma \\ k=0 \end{array} S'_k \exp(ki\omega) \right] G^{0-1} \left[\begin{array}{c} r^+ \\ \Sigma \\ s=0 \end{array} S_s \exp(-si\omega) \right] \left[\frac{\partial \Pi}{\partial \phi_n} \right]$$

$$F_{Z+V}(\omega) F_{VV}^{-1}(\omega) F_{VZ+}(\omega) \left[\frac{\partial \Pi}{\partial \phi_q} \right]' d\omega ,$$

where we use

$$F_{VZ+}(\omega) = \left[\left[\exp(i\omega) \dots \exp(ri\omega) \right] \otimes F_{VY}(\omega) \cdot : 0 \right] ,$$

and using the vec notation

$$\left[\text{vec} \left[\frac{\partial \Pi}{\partial \phi_q} \right] \right]' \int_{-\pi}^{\pi} \left[F_{Z+V}(\lambda) F_{VV}^{-1}(\lambda) F_{VZ+}(\lambda) \right] \otimes F_{VV}(-\lambda) d\lambda \left[\text{vec} \left[\frac{\partial \Pi}{\partial \phi_n} \right] \right] .$$

(8.9)

Thus compactly we can write

$$\xi_{\phi}^{nq} = \left[\text{vec} \left[\frac{\partial \Pi}{\partial \phi_n} \right] \right]' \int_{-\pi}^{\pi} \left[F_{Z+Z+}(\lambda) - F_{Z+V}(\lambda) F_{VV}^{-1}(\lambda) F_{VZ+}(\lambda) \right] \otimes$$

$$F_{VV}^{-1}(-\lambda) d\lambda \left[\text{vec} \left[\frac{\partial \Pi}{\partial \phi_q} \right] \right]$$

and that

$$\Omega_{\phi} = \left\{ \left[\text{vec} \frac{\partial \Pi}{\partial \phi} \right]' \int_{-\pi}^{\pi} \left[F_{Z^+Z^+}(\lambda) - F_{Z^+V}(\lambda) F_{VV}^{-1}(\lambda) F_{VZ^+}(\lambda) \right] \otimes F_{VV}^{-1}(-\lambda) d\lambda \text{vec} \left(\frac{\partial \Pi}{\partial \phi} \right) \right\}^{-1},$$

where $\text{vec} \frac{\partial \Pi}{\partial \phi}$ is the $(pk^+ \times n)$ matrix whose j^{th} row is $\frac{\partial}{\partial \phi_j}(\text{vec } \Pi)'$, and k^+ , the number of predetermined variables, equals $(pr + k)$.

Now $F_{ZV}(\lambda) = 0$, and the elements of $F_{Z^+Z^+}(\lambda) - F_{VZ^+}^*(\lambda) F_{VV}^{-1}(\lambda) F_{VZ^+}(\lambda)$ only differ from $F_{Z^+Z^+}(\lambda)$ when both row and column correspond to lagged endogenous variables.

Now considering the submatrix which corresponds to the lagged endogenous variables we see that this is equal to

$$\int_{-\pi}^{\pi} \left\{ \left[\begin{array}{c} \exp(-i\lambda) \\ \vdots \\ \exp(-ri\lambda) \end{array} \right] \left[\exp(i\lambda) \dots \exp(ri\lambda) \right] \right\} \otimes \left[F_{YY}(\lambda) - F_{YV}(\lambda) F_{VV}^{-1}(\lambda) F_{VY}(\lambda) \right] \otimes F_{VV}^{-1}(-\lambda) d\lambda.$$

Also from

$$y_t - \Pi_{01} y_{t-1} - \dots - \Pi_{0r} y_{t-r} - \Pi_2 z_t = y_t,$$

(note that $\Pi_1 = (\Pi_{01} \dots \Pi_{0r})$),

we have if

$$\Pi(\lambda) = I - \sum_{s=1}^r \Pi_{0s} \exp(-is\lambda), \quad (8.10)$$

$$F_{YV}(\lambda) = [\Pi(\lambda)]^{-1} F_{VY}(\lambda)$$

and

$$F_{YY}(\lambda) = (\Pi(\lambda))^{-1} F_{VV}(\lambda) (\Pi^*(\lambda))^{-1} + (\Pi(\lambda))^{-1} \Pi_2 F_{ZZ}(\lambda) \Pi_2' (\Pi^*(\lambda))^{-1}.$$

Then we have that

$$F_{YY}(\lambda) - F_{YV}(\lambda) F_{VV}^{-1}(\lambda) F_{VY}(\lambda) = (\Pi(\lambda))^{-1} \Pi_2 F_{ZZ}(\lambda) \Pi_2' (\Pi^*(\lambda))^{-1}.$$

This can, of course, be written

$$F_{YZ}(\lambda) F_{ZZ}^{-1}(\lambda) F_{ZY}(\lambda) .$$

It follows easily that the whole asymptotic variance matrix can now be written

$$\begin{aligned} \Omega_\phi = & \left\{ \left[\text{vec } \frac{\partial \Pi}{\partial \underline{\phi}} \right]' \left[\int_{-\pi}^{\pi} \left\{ [F_{Z^+Z}(\lambda) F_{ZZ}^{-1}(\lambda) F_{ZZ^+}(\lambda)] \otimes F_{VV}^{-1}(-\lambda) \right\} d\lambda \right] \right. \\ & \left. \left[\text{vec } \frac{\partial \Pi}{\partial \underline{\phi}} \right] \right\}^{-1} . \end{aligned} \quad (8.11)$$

This is perfectly general for the case when the reduced form matrix is a general function of a set of parameters. Now consider the standard simultaneous equations case where

$$\frac{\partial \Pi}{\partial \phi_n} = - B_0^{-1} \left[\frac{\partial A}{\partial \phi_n} \right] P .$$

Then

$$\begin{aligned} \xi_\phi^{nq} = & \text{tr} \int_{-\pi}^{\pi} F_{VV}^{-1}(\lambda) \left[\frac{\partial \Pi}{\partial \phi_q} \right] F_{Z^+Z}(\lambda) F_{ZZ}^{-1}(\lambda) F_{ZZ^+}(\lambda) \left[\frac{\partial \Pi}{\partial \phi_n} \right]' d\lambda \\ = & \int_{-\pi}^{\pi} B_0'^{-1} F_{VV}^{-1}(\lambda) B_0^{-1} \left[\frac{\partial A}{\partial \phi_q} \right] P F_{Z^+Z}(\lambda) F_{ZZ}^{-1}(\lambda) F_{ZZ^+}(\lambda) P' \left[\frac{\partial A}{\partial \phi_n} \right]' d\lambda . \end{aligned}$$

Now

$$P F_{Z^+Z}(\lambda) = \begin{bmatrix} \Pi F_{Z^+Z}(\lambda) \\ F_{Z^+Z}(\lambda) \end{bmatrix} ,$$

and $F_{YZ}(\lambda) = \Pi F_{Z^+Z}(\lambda) + F_{VZ}(\lambda)$, and $F_{VZ}(\lambda) = 0$.

Thus

$$P F_{Z^+Z}(\lambda) = F_{XZ}(\lambda) ,$$

so that

$$\Omega_{\phi} = \left\{ \left[\text{vec} \frac{\partial A}{\partial \phi} \right]' \int_{-\pi}^{\pi} \left[F_{XZ}(\lambda) F_{ZZ}^{-1}(\lambda) F_{ZX}(\lambda) \otimes F_{UU}^{-1}(-\lambda) \right] d\lambda \right. \\ \left. \left[\text{vec} \frac{\partial A}{\partial \phi} \right] \right\}^{-1},$$

where $\left[\text{vec} \frac{\partial A}{\partial \phi} \right]$ is $(p^2(r+1+k) \times n)$ matrix defined in an analogous way as $\text{vec} \frac{\partial \Pi}{\partial \phi}$ above.

If the restrictions are linear, or zero-one restrictions then we can simple replace $\text{vec} \left[\frac{\partial A}{\partial \phi} \right]$ by S, which, in the latter case, is a selection matrix.

Finally we can give an explicit expression for $F_{XZ}(\lambda)$, by using $\Phi(\lambda)$ as defined in (7.23).

Thus we have

$$F_{XZ}(\lambda) = \begin{bmatrix} \Phi^*(\lambda) F_{YZ}(\lambda) \\ F_{ZZ}(\lambda) \end{bmatrix}.$$

Also from comparing (8.10) on (7.13)

$$(\Pi(\lambda))^{-1} \Pi_2 = - (B(\lambda))^{-1} \Gamma$$

so that

$$F_{YZ}(\lambda) = (\Pi(\lambda))^{-1} \Pi_2 F_{ZZ}(\lambda) = - (B(\lambda))^{-1} \Gamma F_{ZZ}(\lambda),$$

so that

$$F_{XZ}(\lambda) F_{ZZ}^{-1}(\lambda) F_{ZX}(\lambda) \\ = \begin{bmatrix} \Phi^*(\lambda) B(\lambda)^{-1} \Gamma F_{ZZ}(\lambda) \Gamma' (B^*(\lambda))^{-1} \Phi(\lambda) & -\Phi^*(\lambda) B(\lambda)^{-1} \Gamma F_{ZZ}(\lambda) \\ -F_{ZZ}(\lambda) \Gamma' (B^*(\lambda))^{-1} \Phi(\lambda) & F_{ZZ}(\lambda) \end{bmatrix}$$

$$= P(-\lambda) F_{ZZ}(\lambda) P^*(-\lambda)$$

where $P(\lambda)$ was defined in (7.31).

Now it is clear that the resulting AVM is the one derived in the previous section for the Specfiml estimators.

IX Final Comment

The last section makes it clear how much we depend, when we generalize the stochastic process generating the errors on our structural equations, on the assumption that there are a set of exogenous variables, completely incoherent with the errors on the equations which we are estimating. This is worrying, because there is a serious problem of inadequate sample size in the estimation of these spectral estimators, so that only relatively small models can be estimated, as discussed in Section V. In highly aggregate models, the aggregative variables which may be treated as exogenous should usually more properly be considered endogenous, and the resulting inconsistencies may be serious. Ignoring the inadequate sample size problem, the set of variables, which it may be really appropriate to consider exogenous, may behave like a set of trends or dummy variables, which lead to estimation of the model coefficients with large standard errors. Of course the same will be true, if general ARMA models are used in the time domain with high order moving average and/or autoregressive equations. But in this context, it seems a natural procedure to reduce the order of the stochastic processes by using the usual type of significance test, and in doing so, we end up with an acceptable explanation of the data, in which the error generation stochastic processes have been simplified, in effect allocating as much as possible of the explanation of the behaviour of the observed data to the structural equation model. The econometrician would agree that this may be a rather arbitrary procedure, and would judge the value of the result largely on a priori judgement of the economic acceptability of the model. A comparison with the Specfml estimates of the model would certainly throw light on the limitations of the ARMA pretesting estimates.

In the case of variables which are generated by a stochastic model outside the economic model which is being estimated, a time domain approach may be advantageous, if it is assumed that the errors on the two models are incoherent, and that there are rather long lags in the effect of economic variables on the variables generated by the outside model. In this case there

will be no harm in the time domain in treating the outside variables as exogenous, provided the internal errors stochastic model is correctly specified. Of course, if the combined model is sufficiently long, the treatment of the outside variables as exogenous in spectral methods would lead to only small inconsistencies. An alternative treatment when using spectral methods would be to develop methods of estimating an incomplete set of equations. Thus equations of the form (7.36) can be used dropping the requirement that there are as many equations as current endogenous variables, and using for Z any set of suitable instrumental variables (i.e. variables which are endogenous from the point of view of the combined model generating both inside and outside variables).

$P'(\lambda_n)$ can be replaced by $I_{ZZ}(\lambda_h)^{-1} I_{ZX}(\lambda_h)$, and the asymptotic error variance matrix estimated as

$$\left[\frac{1}{2M} \sum_h \hat{I}_U(\lambda_h) \otimes I_{XZ}(-\lambda_h) I_{ZZ}(-\lambda_h)^{-1} I_{ZX}(-\lambda_h) \right] u^{-1}.$$

As a special case we can consider similarly the estimation of a single equation.

Appendix. Two asymptotically equivalent expressions for the likelihood function.

If we write the simultaneous equations model with lagged endogenous variables and stationary errors as

$$B_0 y_t + B_1 y_{t-1} + \dots + B_r y_{t-r} + \Gamma z_t = u_t, \quad t = 1, \dots, T \quad (B.1)$$

or using

$$A = (B_0 \dots B_r \Gamma) = (B_0 \Gamma^T) \quad \text{and} \quad (B.2)$$

$$x_t' = (y_t', y_{t-1}', \dots, y_{t-r}', z_t') = (y_t', z_t'^T), \quad (B.3)$$

as

$$A x_t = u_t,$$

we have that log likelihood, under the assumption that the spectra of the errors is constant within frequency bands, can be written as

$$T \log |\det B_0| - \frac{m}{2} \sum_h \log \det F_U(\lambda_h) - \frac{m}{2} \sum_h \text{tr} F_U^{-1}(\lambda_h) A I_{XX}(\lambda_h) A'$$

and if using

$$F_U(\lambda_h) = A I_{XX}(\lambda_h) A'$$

we concentrated the likelihood, we get

$$L_c = T \log |\det B_0| - \frac{m}{2} \sum_h \log \det A I_{XX}(\lambda_h) A'. \quad (B.5)$$

Now an alternative expression for the concentrated likelihood is:

$$L_e = m \sum_h \log |\det B(\lambda_h)| - \frac{m}{2} \sum_h \log \det A I_{XX}(\lambda_h) A', \quad (B.6)$$

where

$$B(\lambda_h) = B_0 + B_1 \exp(i\lambda_h) + \dots + B_r \exp(ir\lambda_h). \quad (B.7)$$

To reconcile L_c and L_e we assume that $\{y_t\}$, $\{z_t\}$ and $\{u_t\}$ are stationary processes, and we proceed as follows.

Write

$$B(L) = B_0 + B_1 L + \dots + B_r L^r \quad (B.8)$$

and note that

$$\det B(L) = \det B_0 \prod (1 - \mu_\ell L)$$

where μ_ℓ is the inverse of the ℓ^{th} root of $B(L)$ and ℓ goes from 1 to rp and p is the number of equations in the system.

Note that if

$$B(e^{i\lambda}) = B_1(\lambda) + i B_2(\lambda)$$

then

$$\det \left[B(e^{i\lambda}) \right] \det \left[B(e^{-i\lambda}) \right] = \det \begin{bmatrix} B_1(\lambda) & -B_2(\lambda) \\ B_2(\lambda) & B_1(\lambda) \end{bmatrix}.$$

Then the Jacobian contribution to the log likelihood function divided by T can be written

$$\frac{1}{2M} \sum_{h=-(M-1)}^M \log \left| B \left[e^{i\lambda_h} \right] \right| = \left[\sum_{\ell=1}^{rp} \frac{1}{2M} \sum_{h=-(M-1)}^M \log \left[1 - \mu_\ell e^{i\lambda_h} \right] \right] + \log \det B_0, \quad (B.9)$$

and the difference between L_e and L_c is given by the term in curly brackets in (B.9). Now we consider each term of the form

$$\frac{1}{2M} \sum_{h=-(M-1)}^M \log \left[1 - \mu_\ell e^{i\lambda_h} \right] = \frac{1}{2M} \sum_{h=0}^{2M-1} \log \left[1 - \mu_\ell e^{i\lambda_h} \right].$$

We have that if we write $v^* = \exp(i\pi/M)$, then v^* is the $(1/2M)$ complex root of unity, and if we consider the equation

$$x^{2M} - 1 = 0,$$

the corresponding roots can be written v^{*h} , $h=0, \dots, 2M-1$. Then

we can write

$$x^{2M} - 1 = \prod_{h=1}^M (x - v^{*h}) .$$

It follows that

$$\begin{aligned} \sum_{h=0}^{2M-1} \left[1 - \mu_{\ell} e^{i\lambda_h} \right] &= \prod_{h=0}^{2M-1} (1 - \mu_{\ell} v^{*h}) \\ &= \mu_{\ell}^{2M} \prod_{h=0}^{2M-1} (1/\mu_{\ell} - v^{*h}) = \mu_{\ell}^{2M} [(1/\mu_{\ell})^{2M} - 1] = (1 - \mu_{\ell}^{2M}) . \end{aligned}$$

Thus

$$\frac{1}{2M} \sum_{h=0}^{2M-1} \log \left[1 - \mu_{\ell} e^{i\lambda_h} \right] = \frac{1}{2M} \log (1 - \mu_{\ell}^{2M}) .$$

Now if $|\mu_{\ell}| < c < 1$, then $|\mu_{\ell}^{2M}|$ is negligible as $M \rightarrow \infty$, indeed if $M = o(T^{1/a})$, a some fixed integer, then $\mu_{\ell}^{2M} = o(T^{-r})$ for all positive r and then

$$\frac{1}{2M} \log (1 - \mu_{\ell}^{2M}) = o(T^{-r}) \text{ as } T \rightarrow \infty ,$$

thus all the terms in the RHS of (B.9) but the $\log \det B_0$ are negligible.

Also considering

$$\frac{\partial}{\partial \phi_k} \left[\frac{1}{2M} \sum_h \log (B(\lambda_h)) \right] ,$$

where ϕ_k is any parameter on which $B(\lambda_h)$ depends, we find ourselves considering

$$- \sum_{\ell} \left[\frac{\mu_{\ell}^{2M-1}}{1 - \mu_{\ell}^{2M}} \right] \left[\frac{\partial \mu_{\ell}}{\partial \phi_k} \right] . \tag{B.10}$$

Now if ϕ_k is such that $B(\lambda_h)$ has elements which have bounded

and continuous first derivatives, and we restrict the range of variation of ϕ_k , such that $|\mu_\ell| < c < 1$ for all ℓ , then it follows that (B.10) is $o(T^{-r})$ for all r as T and $M \rightarrow \infty$.

Similarly, the second derivative can be written

$$-\sum_{\ell} \left(\frac{\mu_{\ell}^{2M-1}}{1-\mu_{\ell}^{2M}} \right) \left(\frac{\partial^2 \mu_{\ell}}{\partial \phi_k \partial \phi_j} \right)$$

$$-\sum_{\ell} \left(\frac{(2M-1)\mu_{\ell}^{2M-2} + 2\mu_{\ell}^{4M-2}}{(1-\mu_{\ell}^{2M})^2} \right) \frac{\partial \mu_{\ell}}{\partial \phi_k} \frac{\partial \mu_{\ell}}{\partial \phi_j}$$

and on the same conditions as before this is $o(T^{-r})$ for all r , and so definitely negligible.

Therefore both likelihood and their first and second derivatives are pairwise equivalent for large T and M .

Bibliography

- Akaike, H. (1973), 'Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models', Biometrika, 60, pp. 255-265.
- Amemiya, T. (1973a), 'Generalized Least Squares with Estimated Autocovariance Matrix', Econometrica, 41, pp. 723-732.
- Amemiya, T. (1973b), 'Regression Analysis when the Variance of the Dependent Variable is Proportional to the Square of its Expectation', Journal of the American Statistical Association, 68, pp. 928-934.
- Anderson, T.W. (1971), The Statistical Analysis of Time Series, John Wiley.
- Bock, M.E., T.A. Yancey and G.G. Judge (1973), 'The Statistical Consequences of Preliminary Test Estimators in Regression', Journal of the American Statistical Association, 68, pp. 106-116.
- Box, G.E.P. and G.M. Jenkins (1971), Time Series Analysis Forecasting and Control, Holden-Day, second printing.
- Brillinger, D.R. (1969), 'Asymptotic Properties of Spectral Estimates of Second-Order', Biometrika, 56, pp. 375-90.
- Brundy, J.M. and D.W. Jorgenson (1971), 'Efficient Estimation of Simultaneous Equations by Instrumental Variables', Review of Economics and Statistics, 53, pp. 207-224.
- Dhrymes, P.J. (1970), Econometrics, Statistical Foundations and Applications, Harper and Row.
- Dhrymes, P.J. (1973), 'Small Sample and Asymptotic Relations Between Maximum Likelihood and Three-stage Least Squares Estimators', Econometrica, 41, pp. 357-364.
- Durbin, J. (1963), 'Maximum Likelihood Estimation of the Parameters of a System of Simultaneous Regression Equation', Paper presented at the Copenhagen Meeting of the Econometric Society.
- Engel, R.F. and R. Gardner (1976), 'Some Finite Sample Properties of Spectral Estimators of a Linear Regression', Econometrica, 44, pp. 149-166, January.
- Goldberger, A.S. (1964), Econometric Theory, John Wiley.
- Goodman, N.R. (1963), 'Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution', The Annals of Mathematical Statistics, 34, pp. 152-177.
- Granger, C.W.J. (1966), 'The Typical Spectral Shape of an Economic Variable', Econometrica, 34, pp. 150-161.

- Hannan, E.J. (1963), 'Regression for Time Series', in Proceedings of the Symposium on Time Series Analysis, M. Rosenblatt (ed.) John Wiley, pp. 17-37.
- Hannan, E.J. (1965), 'The Estimation of Relationships involving Distributed Lags', Econometrica, 33, pp. 206-224.
- Hannan, E.J. (1967), 'The Estimation of a Lagged Regression Relation', Biometrika, 54, pp. 409-418.
- Hannan, E.J. (1970), Multiple Time Series, John Wiley.
- Hannan, E.J. (1971), 'The Identification Problem for Multiple Equation Systems with Moving Average Errors', Econometrica, V. 39, pp. 751-765.
- Hannan, E.J. and D.F. Nicholls (1972), 'The Estimation of Mixed Regression, Autoregression, Moving Average, and Distributed Lag Models', Econometrica, 40, 529-548.
- Hannan, E.J. and R.D. Terrell (1973), 'Multiple Equation Systems with Stationary Errors', Econometrica, 41, pp. 299-230.
- Hannan, E.J. and P.J. Thomson (1971), 'Spectral Inference over Narrow Bands' J. Appl. Prob., 8, pp. 157-169.
- Hendry, D.F. (1970), 'The Estimation of Economic Models with Autoregressive Errors', unpublished Ph. D. thesis, London University.
- Hendry, D.F. (1971), 'Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process', International Economic Review, 12, pp. 257-272.
- Hendry, D.F. (1976), 'The Structure of Simultaneous Equations Estimators', Journal of Econometrics, 4, pp. 51-88, February.
- Judge, G.G., T.A. Yancey and M.E. Bock (1973), 'Properties of Estimators after Preliminary Tests of Significance when Stochastic Restrictions are Used in Regression', Journal of Econometrics, 1, pp. 29-48.
- Kendall, M.G. and A. Stuart (1967), The Advanced Theory in Statistics, vol. 2, Inference and Relationship. C. Griffin & Co., second edition.
- Khatri, C.G. (1965) 'Classical Statistical Analysis based on a Certain Multivariate Complex Gaussian Distribution', Ann. Math. Statist., 36, pp. 98-114.
- Koopmans, T.C. and W.C. Hood (1953), 'The Estimation of Simultaneous Linear Economic Relationships', Chapter VI in Studies in Econometric Method, W.C. Hood and T.C. Koopman (eds.), John Wiley, pp. 112-199.

- Macduffee, C.C. (1956), The Theory of Matrices, Chelsea Publishing Co.
- Maddala, G.S. (1971), 'Generalized Least Squares with an Estimated Variance Covariance Matrix', Econometrica, 39, pp. 23-24.
- Nerlove, M. (1972), 'On the Structure of Serial Dependence in Some U.S. Price Series' in The Econometrics of Price Determination Conference. O. Eckstein (ed.) Board of Governors of the Federal Reserve System, Washington.
- Nicholls, D.F. (1975), 'The Efficient Estimation of Vector Linear Time Series Models', unpublished paper, Australian Nat. University and Michigan State University.
- Nicholls, D.F., A.R. Pagan and R.D. Terrel (1975), 'The Estimation and Use of Models with Moving Average Disturbance Terms: A survey', International Economic Review, 16, pp. 113-134, February.
- Orcutt, G.H. (1948), 'A Study of the Autoregressive Nature of Time Series Used for Tinbergen's Model of the Economic System of the United States, 1919-1932', Journal of the Royal Statistical Society, Ser. B, 10, pp. 1-45.
- Powell, M.J.D. (1964), 'An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives', Computer Journal, 7, pp. 303-307.
- Rothenberg, T.J. (1973), Efficient Estimation with a Priori Information, Cowles Foundation Monograph, 23, Yale University Press.
- Rothenberg, T.J. and C.T. Leenders (1964), 'Efficient Estimation of Simultaneous Equations Systems', Econometrica, 32, pp. 57-76.
- Rozanov, Yu. A. (1967), Stationary Random Processes. Holden-Day.
- Sargan, J.D. (1961), 'The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals', Econometrica, 29, pp. 414-426.
- Sargan, J.D. (1964), 'Wages and Prices in the United Kingdom: A Study in Econometric Methodology', Colston Papers, 16, Butterworths Scientific Publications.
- Sargan, J.D. (1972), 'The Identification and Estimation of Sets of Simultaneous Stochastic Equations', unpublished paper, London School of Economics.
- Sargan, J.D. (1975), 'Asymptotic Theory and Large Models', International Economic Review, V. 16, n°1, pp. 75-91.

- Wahba, G. (1968), 'On the Distribution of Some Statistics Useful in the Analysis of Jointly Stationary Time Series' , The Annals of Mathematical Statistics, 39, pp. 1849-1862
- Wallis, K.F. (1966), 'Some Econometric Problems in the Analysis of Inventory Cycles', Cowles Commission Discussion Paper, 209, Yale University.
- Watson, G.S. and E.J. Hannan (1956), 'Serial Correlation in Regression Analysis II', Biometrika, 43, pp. 436-48.
- Whittle, P. (1953), 'The Analysis of Multiple Stationary Time Series', Journal of the Royal Statistical Society, Series B, 15, pp. 125-139.
- Zellner, A. and H. Theil (1962), 'Three-Stage Least-Squares Simultaneous Estimation of Simultaneous Equations', Econometrica, 30, pp. 54-78.

Abstract

This monograph is concerned with the estimation theory of econometric models with stationary errors. The theory is based on the likelihood function using a spectral transformation of the covariance matrix. This study considers firstly, a simultaneous equations model in which all the predetermined variables are exogenous, and shows the generality of the spectral approach. In particular it shows that from the likelihood function, first order conditions and Hessian in spectral form it is possible to derive the corresponding expressions used in different time domain procedures. In fact the spectral estimator proposed encompasses most estimators of the simultaneous model under different hypothesis on the process generating the disturbances. Secondly it shows the constraints in the computation of the spectral estimators from small samples. Then it studies models with lagged endogenous variables. Various alternative methods of computing maximum likelihood estimates are considered and shown to be asymptotically equivalent. The spectral maximum likelihood estimators are compared to estimators obtained when the disturbance process is a pure autoregressive (AR) or a pure moving average (MA) and it is shown that failure to impose the AR or MA restriction (the results for the MA case were previously shown by Hannan and Nicholls, *Econometrica* 1972, p. 529-548) results in a loss of asymptotic efficiency when lagged endogenous variables are included in the model. This differs from the case in which there are no lagged endogenous variables, where failure to impose the AR or AM restrictions does not lead to a loss in asymptotic efficiency. Finally this study shows that as the order of the AR error vector process, assumed in autoregressive full information maximum likelihood estimation, tends to infinity the asymptotic variance matrix (AVM) for that case is the AVM derived for the spectral estimators.