THE SMALL SAMPLE PROPERTIES OF ESTIMATORS OF THE MOVING AVERAGE PROCESS[1]

By

James Davidson

NUMBER 109

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

THE SMALL SAMPLE PROPERTIES OF ESTIMATORS OF THE MOVING AVERAGE PROCESS[1]

By

James Davidson

NUMBER 109

May, 1977.

(1) Grateful thanks are due to Milly Casey for assistance with the computing.

Introduction

A well-known although as yet unpublished paper by K.M. Kang[1] points out some peculiarities in the behaviour of the least squares and maximum likelihood estimators of the moving average process. The investigation described in this paper was stimulated originally by some practical problems encountered in the development of a programme for the estimation of econometric models with moving average residuals, to which the Kang results appeared relevant.

The paper derives in a general context the properties of the likelihood and sum of squares functions established informally by Kang, and compares the properties of various asymptotically equivalent estimation criteria. The main issue here is the appropriate treatment of the "starting residuals" in small samples. Monte Carlo experiments were carried out for the first-order moving average model, considering both the simple time-series case and the case of linear regression residuals. The results suggest practical conclusions of some importance concerning the choice of estimation criterion in small or moderately sized samples.

---

(1) Kang {1973}.

I. General Background

The q^{th} order moving average (MA(q)) process is defined as

$$u_t = \theta(L)\,\varepsilon_t , \qquad \varepsilon_t \sim N\,I\,(0,\sigma^2) \tag{1.1}$$

where $\theta(L)$ represents the lag polynomial $\theta_o + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$ where $\theta_o \equiv 1$, and L is the lag operator such that $Lx_t = x_{t-1}$. The roots of the process, $\alpha_1, \ldots, \alpha_q$, are then defined by

$$\theta(L) = \prod_{i=1}^{q} (1 - \alpha_i L) \tag{1.2}$$

The roots may be either real, or in complex conjugate pairs.

The autocovariance function of the MA(q) process (treating the roots as parameters by substitution of (1.2) into (1.1)) is

$$\begin{aligned} E(u_t u_{t+s}) &= \sigma^2\ \omega_s\ (\alpha_1, \ldots, \alpha_q), \\ \omega_s &= \sum_{r=o}^{q-s} \theta_r\,\theta_{r+s}, \qquad s = 0, 1, \ldots, q \\ &= 0, \qquad s > q, \end{aligned} \tag{1.3}$$

and has an important property which we state as

Theorem 1

(i) When the j^{th} root is real

$$\omega_s(\ldots, \alpha_j, \ldots) = \alpha_j^2\ \omega_s(\ldots, \alpha_j^{-1}, \ldots) \tag{1.4}$$

(ii) When the j^{th} and $j+1^{th}$ roots form a complex conjugate pair,

$$\omega_s(\ldots,\alpha_j,\ \alpha_{j+1},\ \ldots) = |\alpha_j|^4\ \omega_s(\ldots,\ \alpha_j^{-1}\ ,\ \alpha_{j+1}^{-1},\ \ldots)$$

Note that complex roots cannot be inverted singly if the θ_i's are to be restricted to real values.

A proof of Theorem 1 is given in Appendix II. It's most important consequence is that when the variance of the white noise process $\{\varepsilon_t\}$ is unknown, there will exist as many as 2^q observationally equivalent representations of the MA process corresponding to the possible combinations of the roots and their inverses.

For the joint density of a realisation of the series $u_1, \ldots, u_T$ is

$$f(\underset{\sim}{u};\ \alpha_1, \ldots,\ \alpha_q,\ \sigma^2) = (2\Pi\sigma^2)^{-T/2}\ |\underset{\sim}{\Omega}|^{-\frac{1}{2}}\ \exp\{-\frac{\underset{\sim}{u}'\underset{\sim}{\Omega}^{-1}\underset{\sim}{u}}{2\sigma^2}\} \tag{1.6}$$

where $\underset{\sim}{u} = (u_1,\ldots,u_T)'$ and

$$\underset{\sim}{\Omega} = \begin{pmatrix} \omega_0 & \omega_1 & \omega_2 & \ldots & \omega_s & 0 & \ldots & 0 \\ \omega_1 & \omega_0 & \omega_1 & & & & & \vdots \\ \omega_2 & \omega_1 & \omega_0 & & & & & \\ \vdots & & & \ddots & & & & 0 \\ \omega_s & & & & & & & \omega_s \\ \vdots & & & & & & \omega_0 & \omega_1 \\ 0 & \ldots & \ldots & \ldots & 0 & \omega_s & \ldots & \omega_0 \end{pmatrix}$$

From (1.4) we obtain for real α_j the relations

$$\underset{\sim}{\Omega}^{-1}(\ldots, \alpha_j, \ldots) = \alpha_j^{-2}\ \underset{\sim}{\Omega}^{-1}(\ldots, \alpha_j^{-1}, \ldots), \tag{1.7}$$

$$|\ \underset{\sim}{\Omega}(\ldots, \alpha_j, \ldots)| = \alpha_j^{2T}\ |\ \underset{\sim}{\Omega}(\ldots, \alpha_j^{-1}, \ldots)|, \tag{1.8}$$

and it is evident, substituting in (1.6), that

$$f(\underset{\sim}{u}; \ldots, \alpha_j, \ldots, \sigma^2) = f(\underset{\sim}{u}; \ldots, \alpha_j^{-1}, \ldots, (\alpha_j^2\ \sigma^2)). \tag{1.9}$$

For a pair of complex roots, the corresponding equality is

$$f(\underset{\sim}{u}; \ldots, \alpha_j, \alpha_{j+1}, \ldots, \sigma^2) \tag{1.10}$$

$$= f(\underset{\sim}{u}; \ldots, \alpha_j^{-1}, \alpha_{j+1}^{-1}, \ldots, (|\alpha_j|^4\ \sigma^2)).$$

It follows that the likelihood function - treated conventionally as a function of the lag coefficients rather than of the roots - will possess up to 2^q equal maxima. It is usual to identify the process arbitrarily by imposing the condition that all the roots lie inside the unit circle, called the 'invertibility condition' because it ensures that the MA process has an equivalent <u>stationary</u> autoregressive representation, $\varepsilon_t = \theta^{-1}(L)\ u_t$.

All estimation procedures therefore entail a search for a maximum of the likelihood or other criterion function over the invertability region - defined as the set of points in θ-space for which all the corresponding roots lie inside the unit circle - and search routines need to incorporate some device for excluding the choice of parameter values outside the region.

II. Alternative Estimation Criteria

A sample realisation of length T of the process specified by (1.1) can be represented by

$$\underset{\sim}{u} = \underset{\sim}{M}\,\underset{\sim}{\varepsilon}_1 + \underset{\sim}{N}\,\underset{\sim}{\varepsilon}_o \qquad (2.1)$$

where $\underset{\sim}{u} = (u_1, \ldots, u_T)' \quad (T \times 1)$

$\underset{\sim}{\varepsilon}_1 = (\varepsilon_1, \ldots, \varepsilon_T)' \quad (T \times 1)$

$\underset{\sim}{\varepsilon}_o = (\varepsilon_{1-q}, \ldots, \varepsilon_o)' \quad (q \times 1)$

$$\underset{\sim}{M} = \begin{bmatrix}
1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\
\theta_1 & 1 & 0 & & & & \vdots \\
\theta_2 & \theta_1 & 1 & & & & \vdots \\
\theta_3 & \theta_2 & \theta_1 & 1 & & & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & & \vdots \\
\theta_q & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \theta_q & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & 1 & 0 \\
\vdots & & \ddots & \ddots & \theta_1 & 1 & 0 \\
0 & \cdots & 0 & \theta_q & \cdots & \theta_2 & \theta_1 & 1
\end{bmatrix} \qquad (T \times T)$$

$$\underset{\sim}{N} = \begin{bmatrix}
\theta_q & \theta_{q-1} & \cdots & \cdots & \theta_1 \\
0 & \theta_q & \ddots & & \theta_2 \\
0 & 0 & \ddots & & \vdots \\
\vdots & & \ddots & & \theta_q \\
0 & \cdots & \cdots & \cdots & 0 \\
\vdots & & & & \vdots \\
0 & \cdots & \cdots & \cdots & 0
\end{bmatrix} \qquad (T \times q)$$

From (2.1)

$$\underset{\sim}{\varepsilon}_1 = \underset{\sim}{M}^{-1}(\underset{\sim}{u} - \underset{\sim}{N}\,\underset{\sim}{\varepsilon}_o) \tag{2.2}$$

and the full $(T + q)$ - vector of independent disturbances is

$$\begin{bmatrix} \underset{\sim}{\varepsilon}_o \\ \underset{\sim}{\varepsilon}_1 \end{bmatrix} = \underset{\sim}{\varepsilon} = \underset{\sim\sim}{W}\underset{\sim}{u} + \underset{\sim}{X}\,\underset{\sim}{\varepsilon}_o \tag{2.3}$$

where

$$\underset{\sim}{W} = \begin{bmatrix} \underset{\sim}{0} \\ \underset{\sim}{M}^{-1} \end{bmatrix} \quad ((T \times q) \times T)$$

$$\underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{I}_q \\ \underset{\sim}{M}^{-1}\underset{\sim}{N} \end{bmatrix} \quad ((T + q) \times q)$$

Box and Jenkins[1] show that the exact likelihood of $\underset{\sim}{u}$ can be obtained by replacing the unobserved starting values $\underset{\sim}{\varepsilon}_o$ by their conditional expectation. Letting $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_q)$, their result can be stated as

Theorem 2 (Box and Jenkins)

$$f(\underset{\sim}{\theta}, \sigma^2; \underset{\sim}{u}) = (2\pi\sigma^2)^{-\frac{T}{2}}\, \Delta(\theta)^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}\, S(\theta)\} \tag{2.4}$$

where

$$\Delta(\underset{\sim}{\theta}) = |\underset{\sim}{X}'\underset{\sim}{X}| = |\underset{\sim}{\Omega}| \tag{2.5}$$

(where $\underset{\sim}{\Omega}$ is defined following (1.6)),

$$S(\underset{\sim}{\theta}) = \underset{\sim}{\hat{\varepsilon}}'\,\underset{\sim}{\hat{\varepsilon}} \tag{2.6}$$

$$\underset{\sim}{\hat{\varepsilon}} = \underset{\sim\sim}{W}\underset{\sim}{u} + \underset{\sim}{X}\,\underset{\sim}{\hat{\varepsilon}}_o\,, \qquad \underset{\sim}{\hat{\varepsilon}}_o = -(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'\underset{\sim\sim}{W}\underset{\sim}{u} \tag{2.7}$$

---

(1) Box and Jenkins {1970}.

Hence, since $S(\theta) = u'\Omega^{-1}u$, it follows that

$$\Omega^{-1} = W'(I - X(X'X)^{-1}X')\,W = W'Q_X\,W. \tag{2.8}$$

A proof of the theorem is given in Appendix III.

The more general case is where u is a vector of regression residuals. Let

$$u = y - Z\beta \tag{2.9}$$

where Z is a (T x k) matrix of bounded explanatory variables, y is a T-vector of dependent variables and β is a k-vector of unknown, unrestricted regression coefficients, to be regarded as arguments of the likelihood.

By Theorem 2,

$$f(\theta, \beta, \sigma^2; y, Z) =$$

$$(2\Pi\sigma^2)^{-\frac{T}{2}}\ \Delta(\theta)^{-\frac{1}{2}}\ \exp\{-\frac{1}{2\sigma^2}\ S(\theta,\beta)\} \tag{2.10}$$

where $$S(\theta,\beta) = \hat{\varepsilon}'\hat{\varepsilon} = (y - Z\beta)'W'Q_X W(y - Z\beta) \tag{2.11}$$

Concentrating f with respect to β leads to

$$f^*(\theta, \sigma^2; y, Z) = (2\Pi\sigma^2)^{-\frac{T}{2}}\,\Delta(\theta)^{-\frac{1}{2}}\ \exp\{-\frac{1}{2\sigma^2}\,S(\theta)\} \tag{2.12}$$

where now, defining

$$\begin{aligned}\tilde{\varepsilon} &= Wy - WZ\hat{\beta} + X\hat{\varepsilon}_o \\ &= Q_X(Wy - WZ\hat{\beta}),\end{aligned} \tag{2.13}$$

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{Z}'\underset{\sim}{W}'\underset{\sim}{Q}_{\underset{\sim}{X}}\underset{\sim}{W}\underset{\sim}{Z})^{-1}\ \underset{\sim}{Z}'\underset{\sim}{W}\ \underset{\sim}{Q}_{\underset{\sim}{X}}\ \underset{\sim}{W}\ \underset{\sim}{y}\ , \tag{2.14}$$

the sum of squares function instead of (2.11) is

$$S(\underset{\sim}{\theta}) = \underset{\sim}{\varepsilon}'\ \underset{\sim}{\varepsilon} \tag{2.15}$$

$$= \underset{\sim}{y}'\ \underset{\sim}{W}'(\underset{\sim}{Q}_{\underset{\sim}{X}} - \underset{\sim}{Q}_{\underset{\sim}{X}}\ \underset{\sim}{W}\ \underset{\sim}{Z}\ (\underset{\sim}{Z}'\underset{\sim}{W}'\underset{\sim}{Q}_{\underset{\sim}{X}}\ \underset{\sim}{W}\ \underset{\sim}{Z})^{-1}\ \underset{\sim}{Z}'\underset{\sim}{W}'\underset{\sim}{Q}_{\underset{\sim}{X}})\ \underset{\sim}{W}\ \underset{\sim}{y}$$

$$= \underset{\sim}{y}'(\underset{\sim}{\Omega}^{-1} - \underset{\sim}{\Omega}^{-1}\ \underset{\sim}{Z}(\underset{\sim}{Z}'\ \underset{\sim}{\Omega}^{-1}\ \underset{\sim}{Z})^{-1}\ \underset{\sim}{Z}'\ \underset{\sim}{\Omega}^{-1})\ \underset{\sim}{y} \quad \text{by (2.8)}.$$

Estimation of $\underset{\sim}{\beta}$ can be accomplished by maximisation of (2.12) with respect to $\underset{\sim}{\theta}$, and using (2.14). For the present purpose, however, which is to investigate the properties of the likelihood itself, we shall derive results which apply equally to (2.4) and (2.12), and will not in general distinguish between them. Thus, either sum of squares can be decomposed by

$$S(\underset{\sim}{\theta}) = S^{*}(\underset{\sim}{\theta}) - S^{**}(\underset{\sim}{\theta}) \tag{2.16}$$

where when $S(\underset{\sim}{\theta})$ is defined by (2.6) and (2.7), (2.17)

$$S^{*}(\underset{\sim}{\theta}) = \underset{\sim}{u}'\underset{\sim}{W}'\underset{\sim}{W}\underset{\sim}{u}$$

and $$S^{**}(\underset{\sim}{\theta}) = \underset{\sim}{u}'\underset{\sim}{W}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\ \underset{\sim}{X}'\underset{\sim}{W}\ \underset{\sim}{u}\ . \tag{2.18}$$

If $S(\underset{\sim}{\theta})$ is defined by (2.15), on the other hand, by a standard manipulation

$$\underset{\sim}{Q}_{\underset{\sim}{X}} - \underset{\sim}{Q}_{\underset{\sim}{X}}\ \underset{\sim}{W}\ \underset{\sim}{Z}\ (\underset{\sim}{Z}'\underset{\sim}{W}'\underset{\sim}{Q}_{\underset{\sim}{X}}\ \underset{\sim}{W}\ \underset{\sim}{Z})^{-1}\ \underset{\sim}{Z}'\underset{\sim}{W}'\underset{\sim}{Q}_{\underset{\sim}{X}} \tag{2.19}$$

$$= \underset{\sim}{Q}_{\underset{\sim}{WZ}} - \underset{\sim}{Q}_{\underset{\sim}{WZ}}\ \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{Q}_{\underset{\sim}{WZ}}\ \underset{\sim}{X})^{-1}\ \underset{\sim}{X}'\ \underset{\sim}{Q}_{\underset{\sim}{WZ}}$$

where $$\underset{\sim}{Q}_{\underset{\sim}{WZ}} = \underset{\sim}{I} - \underset{\sim}{WZ}\ (\underset{\sim}{Z}'\underset{\sim}{W}'\underset{\sim}{WZ})^{-1}\ \underset{\sim}{Z}'\underset{\sim}{W}'$$

Accordingly we define

$$S^{*}(\underset{\sim}{\theta}) = \underset{\sim}{y}'\underset{\sim}{W}'\underset{\sim}{Q}_{WZ}\,\underset{\sim}{W}\,\underset{\sim}{y} \tag{2.20}$$

$$S^{**}(\underset{\sim}{\theta}) = \underset{\sim}{y}'\underset{\sim}{W}'\underset{\sim}{Q}_{WZ}\,\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{Q}_{WZ}\,\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'\,\underset{\sim}{Q}_{WZ}\,\underset{\sim}{W}\,\underset{\sim}{y} \tag{2.21}$$

$S(\underset{\sim}{\theta})$ and $S^{*}(\underset{\sim}{\theta})$ are known as the unconditional and conditional sums of squares respectively although confusingly, S is the sum of squares obtained using the conditional expectation of $\underset{\sim}{\varepsilon}_{o}$ as starting values, whereas S^{*} in effect uses the unconditional (marginal) expectation of $\underset{\sim}{\varepsilon}_{o}$, i.e., zero.

Now concentrating (2.4) (or (2.12)) with respect to σ^{2}, dropping multiplicative constants and taking the $-2/T^{th}$ power yields a convenient minimand, what Kang has dubbed the 'unlikelihood function',

$$U(\underset{\sim}{\theta}) = \Delta(\underset{\sim}{\theta})^{1/T}\,S(\underset{\sim}{\theta}) \tag{2.22}$$

Evidently we may also define the function

$$U^{*}(\underset{\sim}{\theta}) = \Delta(\underset{\sim}{\theta})^{1/T}\,S^{*}(\underset{\sim}{\theta}) \tag{2.23}$$

which might symmetrically though imprecisely be called the 'conditional unlikelihood'. In fact it is just an approximation to (2.22) with no obvious theoretical interpretation.

In the remainder of the paper we compare the merits of the four functions U, S, U^{*} and S^{*} as alternative criterion functions for estimation of the MA process, either in the simple time-series (TS) case defined by (2.1), or the linear regression residuals (LR) case defined by (2.1) and (2.9)

together.

The conditions for the asymptotic equivalence of the functions follow from the property which we state as :

Lemma 1 Defining α_d as the dominant root of $\theta(L)$, having multiplicity m, the orders of magnitude of the elements of the quadratic matrix $\underset{\sim}{X}'\underset{\sim}{P}\,\underset{\sim}{X}$ as $T\to\infty$, where $\underset{\sim}{P}$ is an arbitrary $(T+q)x(T+q)$ matrix whose elements are $O(1)$, are

(i) $O(1)$ when $|\alpha_d| < 1$

(ii) $O(T^{2(m-1)}\,|\alpha_d{}^{2T}|)$ when $|\alpha_d| > 1$

(iii) $O(T^{2(m-1)}\,f(T))$, $1 \leq f(T) \leq T^2$, when $|\alpha_d| = 1$,

and $f(T)$ depends on the structure of $\underset{\sim}{P}$.

Note that $\underset{T\to\infty}{\text{Lim}}\left(\frac{|\alpha|^{2T}}{f(T)}\right) = \infty$ for all $|\alpha| > 1$ and all f.

A proof of Lemma 1 is given in Appendix IV. For the case $\underset{\sim}{P} = \underset{\sim}{I}$, it is easily deduced from the proof that $f(T) = T$.

For a TS model, $S^{**}(\underset{\sim}{\theta}) = \hat{\underset{\sim}{\varepsilon}}_o'\,\underset{\sim}{X}'\underset{\sim}{X}\,\hat{\underset{\sim}{\varepsilon}}_o$ where the elements of $\hat{\underset{\sim}{\varepsilon}}_o$ are always $O(1)$. The order of magnitude of S^{**} is found directly from Lemma 1, setting $\underset{\sim}{P} = \underset{\sim}{I}$. For an LR model, $S^{**} = \hat{\underset{\sim}{\varepsilon}}_o\,\underset{\sim}{X}'\underset{\sim}{Q}_{WZ}\,\underset{\sim}{X}\,\hat{\underset{\sim}{\varepsilon}}_o$ where $\hat{\underset{\sim}{\varepsilon}}_o = -(\underset{\sim}{X}'\underset{\sim}{Q}_{WZ}\,\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'\underset{\sim}{Q}_{WZ}\,\underset{\sim}{W}\,\underset{\sim}{y}$. The elements of both $\hat{\underset{\sim}{\varepsilon}}_o$ and $\underset{\sim}{Q}_{WZ}$ are easily shown to be $O(1)$, and therefore setting $\underset{\sim}{P} = \underset{\sim}{Q}_{WZ}$ gives the order of magnitude of S^{**} in this case. Further,

$\Delta = |X'X|$ and $S^* \geq O(T)$, so that it follows generally from Lemma 1 and the definitions that U,S,U^* and S^* are asymptotically equivalent[1] strictly inside the invertibility region. On the boundary, it is easily found from the lemma that $\Delta^{1/T} = 1 + o(1)$, and hence the pairs, U,S and U^*, S^* are equivalent, but the 'conditional' and 'unconditional' functions are not.

Appeal to this property has been made to justify the use of S or S^* as suitable criteria, as by Box and Jenkins {1970} for example.[2] The Phillips[3] method employs S as criterion for the estimation of the LR model.

However, Kang's investigation has made it clear that the asymptotic equivalences are a poor guide to the behaviour of the functions in finite samples, particularly near the boundaries of the invertibility region. The current view (see for example Osborn {1976} , Pagan and Nicholls {1976} and Prothero and Wallis {1976}) is that the use of U (or equivalently the exact likelihood) is desirable in spite of the extra computation involved, because of problems associated with S to be examined in the next section. The consensus appears to be that except in large samples, it is desirable to calculate the starting residuals, rather than setting to zero. Åström and Bohlin {1966} employ S^*, however. As far as is known, the use of U^* has not been considered to date.

---

(1) More precisely, suitable transformations of the functions which are bounded as $T \to \infty$ are asymptotically equivalent.

(2) In fact, Box and Jenkins advocate approximating $\hat{\underset{\sim}{\varepsilon}}_o$ using the 'back-forecasting' procedure. The properties of this approximation will not be examined here.

(3) See Trivedi {1970}.

III. Properties of the Criterion Functions

Properties (1.9) and (1.10) are stated in terms of the roots (or moduli if the roots are complex) and to examine the consequences of these properties for the behaviour of the criterion functions, it is desirable to treat the roots as the independent variables. This would be straightforward if, say, the roots were always restricted to be real, since then the set of q real α_i (represented as a point in $\mathbb{R}^q$) determine a set of lag coefficient values uniquely. In her informal discussion Kang confines herself to the domain of real roots.

To generalise the problem, we would like to consider a function having as its domain the product space of m conjugate pairs of complex roots, and $q - 2m$ real roots, for each $m = 0, 1, \ldots, p$, where $p = q/2$ if q is even, or $(q-1)/2$ if q is odd. Let such a function be defined by $C_{\underset{\sim}{k}}$ (where C can stand for any of U, S, U^*, S^*), and $\underset{\sim}{k} \in K$ is an index which identifies which roots are real and which complex. 2^p elements of K are required to enumerate all possible combinations of m complex and $p-m$ real pairs, for each m. The index set K is defined precisely in Appendix I.

The domain of $C_{\underset{\sim}{k}}$ is denoted $\Phi_{\underset{\sim}{k}}$. In Appendix I the family of mappings

$$C_{\underset{\sim}{k}}(\underset{\sim\sim}{\phi_k}) : \Phi_{\underset{\sim}{k}} \to \mathbb{R} \qquad \underset{\sim}{k} \in K$$

are given rigourous definition, and the required properties are established - i.e., the sets $\Phi_{\underset{\sim}{k}}$ are open, so that $C_{\underset{\sim}{k}}$ is differentiable everywhere in $\Phi_{\underset{\sim}{k}}$, and the family $C_{\underset{\sim}{k}}$, $\underset{\sim}{k} \in K$ completely define the original function

$$C(\underset{\sim}{\theta}) : \mathbb{R}^q \to \mathbb{R}$$

in terms of the roots of $\theta(L)$. Note that a complex conjugate pair of roots is represented in polar co-ordinates so that the moduli, denoted r_j, can be treated as independent variables.

With the aid of the family of $\underset{\sim}{C}_k$ as a representation of C, we now prove first of all the general case of the two properties noted by Kang.

Theorem 3 The function U has at least $q+1$ stationary points in $\underset{\sim}{\theta}$ space, on the boundaries of the invertibility region.

Proof U is a monotonic transformation of f concentrated with respect to σ^2, and from the properties of f established in Section I, inversion of a root leaves the value of U unchanged. More precisely, defining $\underset{\sim}{U}_k$ by analogy with $\underset{\sim}{C}_k$, Lemma 4(1) ensures that one may write

$$\underset{\sim}{U}_k(\ldots, \alpha_j, \ldots) = \underset{\sim}{U}_k(\ldots, \alpha_j^{-1}, \ldots) \qquad (3.1)$$

where α_j is a real root, and

$$\underset{\sim}{U}_k(\ldots, r_j, \ldots) = \underset{\sim}{U}_k(\ldots, r_j^{-1}, \ldots) \qquad (3.2)$$

where r_j is the modulus of the j^{th} complex pair; (3.1) and (3.2) follow directly from (1.9) and (1.10) respectively.

Partially differentiating with respect to α_j (or r_j) in $\underset{\sim}{\Phi}_k$ gives

---

(1) See Appendix I.

$$\frac{\partial \underset{\sim}{U}_k(\ldots, \alpha_j, \ldots)}{\partial \alpha_j} = -\frac{1}{\alpha_j^2} \frac{\partial \underset{\sim}{U}_k(\ldots, \alpha_j^{-1}, \ldots)}{\partial \alpha_j^{-1}} \tag{3.3}$$

implying

$$\left.\frac{\partial \underset{\sim}{U}_k}{\partial \alpha_j}\right|_{\alpha_j = \pm 1} = -\left.\frac{\partial \underset{\sim}{U}_k}{\partial \alpha_j}\right|_{\alpha_j = \pm 1} = 0 \tag{3.4}$$

and similarly

$$\left.\frac{\partial \underset{\sim}{U}_k}{\partial r_j}\right|_{r_j = 1} = 0 . \tag{3.5}$$

Let $\Phi_{\underset{\sim}{k}}^*$ represent the domain of q real roots. It follows from (3.4) that there exist (at least) 2^q stationary points of the function in $\Phi_{\underset{\sim}{k}}^*$ at $\alpha_j = \pm 1$, $j = 1, \ldots, q$, although consideration of the form of the relationship (1.2) shows that these points map into $q+1$ points on the invertibility boundary in θ-space.

Theorem 4 The function S is decreasing with respect to all real roots and complex moduli as the boundary of the invertibility region is crossed from inside.

Proof From (1.9) and (3.1) (or (1.10) and (3.2)) and the fact that the ML estimator of σ^2 is S/T (compare (2.4)) it follows that for any real root α_j,

$$\underset{\sim}{S}_k(\ldots, \alpha_j^{-1}, \ldots) = \alpha_j^2 \, \underset{\sim}{S}_k(\ldots, \alpha_j, \ldots) \tag{3.6}$$

and differentiating leads to

$$\left.\frac{\partial S_{\underset{\sim}{k}}}{\partial \alpha_j}\right|_{\alpha_j = 1} = - S_{\underset{\sim}{k}}(\ldots, 1, \ldots) < 0$$

$$\left.\frac{\partial S_{\underset{\sim}{k}}}{\partial \alpha_j}\right|_{\alpha_j = -1} = S_{\underset{\sim}{k}}(\ldots, 1, \ldots) > 0 \qquad (3.7)$$

and similarly for any r_j,

$$\left.\frac{\partial S_{\underset{\sim}{k}}}{\partial r_j}\right|_{r_j = 1} = - S_{\underset{\sim}{k}}(\ldots, 1, \ldots) < 0 \qquad (3.8)$$

By Lemmas 2 and 3(1), applying (3.7) and (3.8) for each $\underset{\sim}{k} \in K$ is sufficient to prove the Theorem.

To determine what happens to the 'conditional' functions on the boundary is more difficult. While S^* may be increasing or decreasing, it is clear that U^* will 'generally' be increasing, since it differs from U by the additive term $\Delta^{1/T} S^{**} > 0$ which by Lemma 1 is $O(1)$ inside boundary, and $O(T)$ or greater outside. The relevant property is formally derived as Theorem 5.

To go beyond arguments from orders of magnitude, however, is unduly complex. A sufficient condition for U^* to be always increasing as the boundary is crossed is that

---

(1) See Appendix I.

$\frac{\partial\, X'X}{\partial\, \alpha_j}$ in the TS case, or

$\frac{\partial\, X'Q_{WZ}\, X}{\partial \alpha_j}$ in the LR case, be

positive definite; but the conditions for this to hold are not easily determined. The general result (for all T) is here proved only for the simplest case, the MA(1) model (Theorem 6).

Theorem 5 There exists a sample size T' such that for all $T > T'$, the function U^* is increasing with respect to all real roots and complex moduli as the boundary of the invertibility region is crossed from inside.

Proof By (2.16) and (2.22),

$$U_k = \Delta_k^{1/T} (S_k^* - S_k^{**}) \tag{3.9}$$

so for any real root α_j,

$$\frac{\partial U_k}{\partial \alpha_j} = \Delta_k^{1/T} \left[\frac{\partial S_k^*}{\partial \alpha_j} - \frac{\partial S_k^{**}}{\partial \alpha_j} + (S_k^* - S_k^{**})(T\Delta_k)^{-1} \frac{\partial \Delta_k}{\partial \alpha_j} \right] \tag{3.10}$$

$$= 0 \quad \text{at} \quad \alpha_j = \pm 1, \text{ by Theorem 3.}$$

Similarly from (2.11),

$$\left.\frac{\partial U_k^*}{\partial \alpha_j}\right|_{\alpha_j = \pm 1} = \Delta_k^{1/T} \left(\frac{\partial S_k^*}{\partial \alpha_j} + S_k^* (T\Delta_k)^{-1} \frac{\partial \Delta_k}{\partial \alpha_j} \right) \tag{3.11}$$

$$= \Delta_k^{1/T} \left(\frac{\partial S_k^{**}}{\partial \alpha_j} + S_k^{**} (T\Delta_k)^{-1} \frac{\partial \Delta_k}{\partial \alpha_j} \right)$$

from (3.10)

Let $\{\underset{\sim}{S}^{*}_{k(T)}\}$ denote the sequence of functions $\underset{\sim}{S}^{**}_{k}$ derived from a sample realisation of size T; and let no root have modulus greater than unity; then from Lemma 1 of Section II, there exists a T' such that for all $T > T'$,

$$\underset{\alpha_j \to 1}{\text{Lim}} - \left[\frac{\underset{\sim}{S}^{**}_{k(T)}(\ldots,\alpha_j,\ldots) - \underset{\sim}{S}^{**}_{k(T)}(\ldots,1,\ldots)}{\alpha_j - 1} \right] > 0 \qquad (3.12)$$

and

$$\underset{\alpha_j \to (-1)_+}{\text{Lim}} \left[\frac{\underset{\sim}{S}^{**}_{k(T)}(\ldots,\alpha_j,\ldots) - \underset{\sim}{S}^{**}_{k(T)}(\ldots,-1,\ldots)}{\alpha_j + 1} \right] < 0 \qquad (3.13)$$

where the limits are taken over values of α_j less than 1 or greater than -1 respectively.

From (3.10) and (2.16) we have

$$\frac{\partial \underset{\sim}{S}_k}{\partial \alpha_j} = - \underset{\sim}{S}_k (T\Delta_k)^{-1} \frac{\partial \underset{\sim}{\Delta}_k}{\partial \alpha_j} \qquad \text{at} \quad \alpha_j = \pm 1, \qquad (3.14)$$

and this implies by (3.8)

$$\left. \frac{\partial \underset{\sim}{\Delta}_k}{\partial \alpha_j} \right|_{\alpha_j = 1} = T\underset{\sim}{\Delta}_k > 0 \qquad (3.15)$$

$$\left. \frac{\partial \underset{\sim}{\Delta}_k}{\partial \alpha_j} \right|_{\alpha_j = -1} = -T\underset{\sim}{\Delta}_k < 0 \qquad (3.16)$$

Hence, substituting results (3.12) and (3.15), or (3.13) and (3.16), into the right-hand equality of (3.11), it follows that for all

$T > T'$

$$\left.\frac{\partial U^{*}_{k(T)}}{\partial \alpha_j}\right|_{\alpha_j = 1 -} > 0 \qquad (3.17)$$

$$\left.\frac{\partial U^{*}_{k(T)}}{\partial \alpha_j}\right|_{\alpha_j = (-1) +} < 0 \qquad (3.18)$$

where (3.17) and (3.18) are right-hand and left-hand derivatives respectively. In the case of a complex pair, inequalities corresponding to (3.17) and (3.18) are obtained by replacing α_j by r_j throughout, and applying the result for all j and k proves the theorem.

Theorem 6 For the first-order (MA(1)) model, the results of Theorem 5 hold for all $T > 0$.

Note, when $q = 1$, the index set K has only one member, and the single root is always real. The k-subscript is therefore discarded.

Proof The model to be considered is $u_t = \varepsilon_t - \alpha\, \varepsilon_{t-1}$; in this

case,

$$\underset{\sim}{W} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & & 0 \\ \alpha & 1 & & \\ \alpha^2 & \alpha & 1 & \\ \vdots & \vdots & & \ddots \\ & & & \alpha \quad 1 \quad 0 \\ \alpha^{T-1} & \alpha^{T-2} & \cdots & \alpha \quad 1 \end{bmatrix}, \qquad ((T+1) \times T)$$

$$\underset{\sim}{X} = (1 \ \alpha \ \alpha^2 \ \ldots \ \alpha^T)' \qquad ((T+1) \times 1),$$

$$\underset{\sim}{X}'\underset{\sim}{X} = 1 + \alpha^2 + \alpha^4 + \ldots + \alpha^{2T} \qquad (3.19)$$

$$= \Delta$$

(a scalar)

and

$$\underset{\sim}{X}'\underset{\sim}{W} = \begin{bmatrix} \alpha + \alpha^3 + \ldots + \alpha^{2T-3} + \alpha^{2T-1} \\ \alpha^2 + \alpha^4 + \ldots + \alpha^{2T-2} \\ \vdots \\ \alpha^{T-1} + \alpha^{T+1} \\ \alpha^T \end{bmatrix}' \quad (1 \times T) \qquad (3.20)$$

$S^{**} = \Delta \hat{\varepsilon}_o^2$ where $\hat{\varepsilon}_o = -\Delta^{-1} \underset{\sim}{X}'\underset{\sim}{W} \underset{\sim}{u}$, and from (3.19) and (3.20) it is easily deduced that $\hat{\varepsilon}_o(\alpha) = \hat{\varepsilon}_o(\alpha^{-1})$, (which we should expect to hold since $\hat{\varepsilon}_o$ is the conditional expectation of ε_o).

Hence, using the same argument employed to prove Theorem 3

$$\left.\frac{\partial \hat{\varepsilon}_o}{\partial \alpha}\right|_{\alpha = \pm 1} = 0 , \tag{3.21}$$

and $$\frac{\partial S^{**}}{\partial \alpha} = \frac{\partial \Delta}{\partial \alpha} \hat{\varepsilon}_o^2 \gtrless 0 \text{ as } \alpha \gtrless 0$$

from (3.15) and (3.16). The proof then follows that of Theorem 5, replacing the inequality conditional on T' with a general one.

When $\underset{\sim}{u} = \underset{\sim}{y} - \underset{\sim}{Z}\,\underset{\sim}{\beta}$ (The LR model) note that from (2.20), S^* can be written as

$$S^*(\alpha, \hat{\underset{\sim}{\beta}}(\alpha)) = (\underset{\sim}{y} - \underset{\sim}{Z}\hat{\underset{\sim}{\beta}})'\underset{\sim}{W}'\underset{\sim}{W}(\underset{\sim}{y} - \underset{\sim}{Z}\hat{\underset{\sim}{\beta}}) \tag{3.22}$$

where $$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{Z}'\underset{\sim}{W}'\underset{\sim}{W}\underset{\sim}{Z})^{-1}\,\underset{\sim}{Z}'\underset{\sim}{W}'\underset{\sim}{W}\underset{\sim}{y} ,$$ and so

$$U^*(\alpha, \hat{\underset{\sim}{\beta}}(\alpha)) = \Delta(\alpha)^{1/T}\, S^*(\alpha, \hat{\underset{\sim}{\beta}}(\alpha)) \tag{3.23}$$

$$\frac{\partial U^*(\alpha, \hat{\underset{\sim}{\beta}}(\alpha))}{\partial \alpha} = \frac{\partial U^*}{\partial \alpha} + \left[\left.\frac{\partial U^*}{\partial \underset{\sim}{\beta}}\right|_{\beta = \hat{\beta}}\right]' \frac{\partial \hat{\underset{\sim}{\beta}}}{\partial \alpha}$$

$$= \frac{\partial U^*}{\partial \alpha}$$

since $\hat{\underset{\sim}{\beta}}$ is the solution to $\frac{\partial U^*}{\partial \underset{\sim}{\beta}} = 0$. It follows that the property just established for the TS model applies equally to the LR model. This completes the proof.

IV. Monte Carlo Results

Whether the boundary turning points of U are maxima, minima or saddle points is sample-dependent, and in small samples - by which we mean samples of moderate size from an econometric point of view - there is a large probability that for a process with a root near the unit circle, U will have a minimum on the boundary. In other words, instead of a pair of minima at α_j and $1/\alpha_j$ with an intervening maximum at the boundary, the three turning points degenerate into a single minimum.

The behaviour of U and the other criteria is illustrated in Fig. I. which shows the graphs of the criterion functions[1] for the MA(1) process with fitted mean. The plotted curves are the average over 50 replications of the functions calculated from a sample of size 20 of the process $u_t = \varepsilon_t - .8\,\varepsilon_{t-1}$, ε_t being a computer generated normal process (as described below). Notice that U^* is the only function which takes an interior minimum, in this case.

Monte Carlo experiments to investigate the incidence of this kind of behaviour, and compare the performance of the different criteria were performed on the MA(1) process, which should satisfactorily exemplify the problems under consideration. The models employed were

$$u_t = \varepsilon_t - \alpha\,\varepsilon_{t-1} \qquad (4.1)$$

$$\varepsilon_t \sim NI(0,1)$$

$$u_t = \mu + \varepsilon_t - \alpha\varepsilon_{t-1} \qquad (4.2)$$

---

(1) The values actually plotted are of $\ln\left(\frac{C}{T}\right)$ where $C = S, U, S^*$ and U^*.

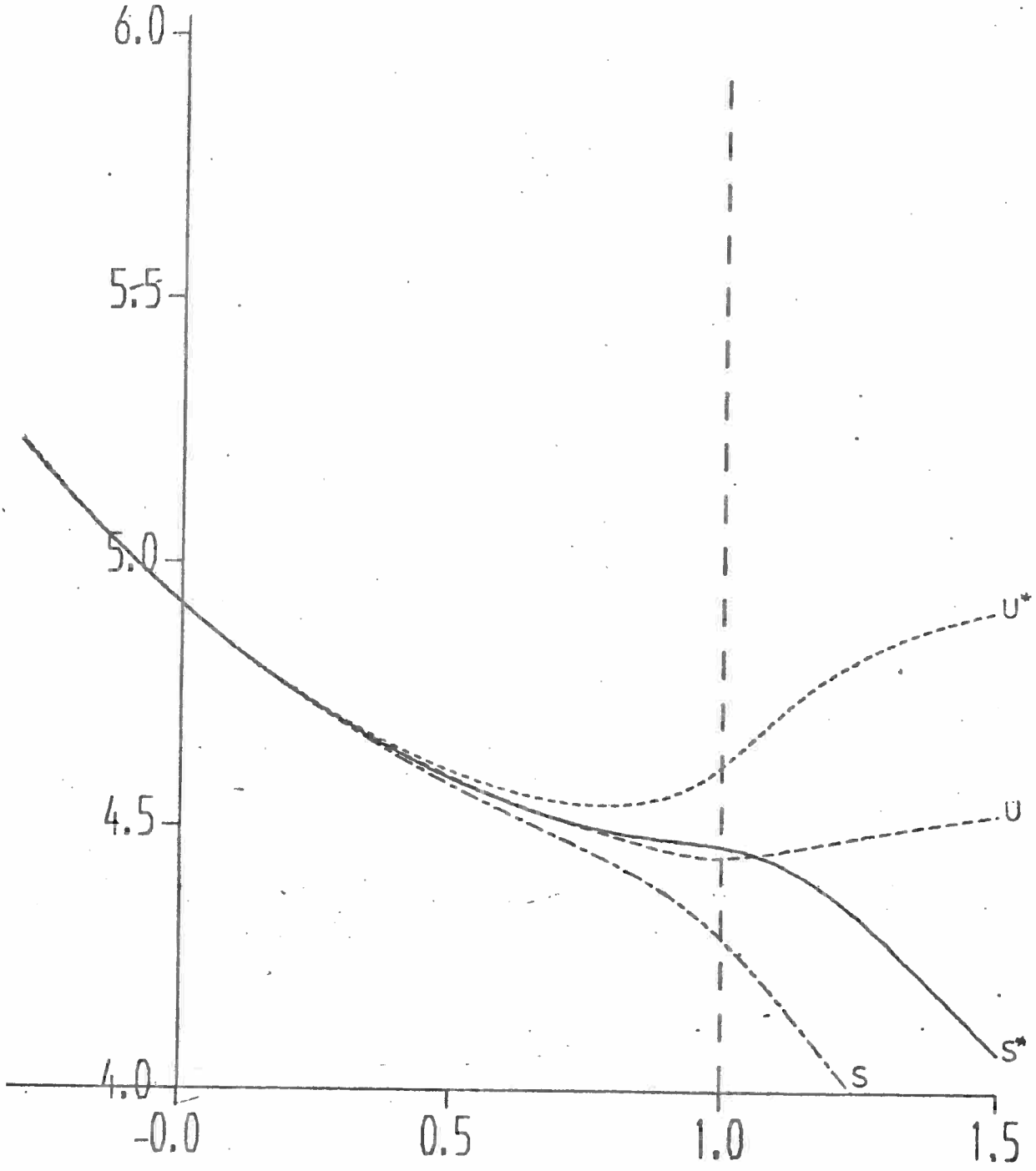

FIGURE I: Concentrated criterion functions for the model

$$u_t = \mu + \varepsilon_t - \alpha \varepsilon_{t-1}, \quad \mu = 0, \quad \alpha = .8, \quad T = 20.$$

the simplest cases of the TS and LR model respectively. The data used to generate each model was identical, with μ always being assigned the value zero in (4.2), but treated as an unknown parameter in the estimation.

The experiments were performed on the Burroughs 6700 computer at the University of Warwick. The procedure adopted was to generate 2000 standard normal pseudo-random numbers using the formula

$$Z = \left[\sum_{i=1}^{12} x_i \right] - 6 \qquad (4.3)$$

where the x_i are uniformly distributed in the interval $[0,1]$, generated using the Burroughs supplied routine RANDOM. This data set was tested for serial independence and zero mean, the computed statistics being :

Mean : - .00852

S.D. : .9837

1st order autocorrelation coefficient : .00439

(on the null hypotheses, autocorrelation coefficient and mean respectively are both normal variates with mean zero and standard deviation of $(2000)^{-\frac{1}{2}} = .02236$). This random series was then used to generate the series $\{u_t\}$, which was split into sub-series - 200 of size 10, 100 of size 20, 50 of size 40 and 25 of size 80 - to provide the samples for replications of the estimation procedures. Samples were generated in this way at 21 equally spaced values of α from -1 to 1 inclusive.

The approach at this stage was therefore extensive rather than intensive, taking advantage of the fact that the parameter space for the process is

in this case a compact interval which can be investigated uniformly. While the experiments are not on a scale of replication which allow point estimates to be treated as effectively exact, they suffice to give a good idea of the relationship between the different estimators.

To minimise the functions a univariate quadratic interpolation routine was employed, based on the algorithm of Davies, Swann and Campey.(1) Estimates were constrained to lie in the interval $[-1, 1]$, the search for the minimum being terminated whenever the function was found to decrease on moving out of the region, and the estimate set at 1 or -1 as appropriate. The true value of α was used as starting value, both to minimise search time and to reduce the possibility of missing the global minimum. The functions minimised in the estimation of the two models (4.1) and (4.2) were the univariate specialisations $S(\alpha)$, $U(\alpha)$, $S^{*}(\alpha)$ and $U^{*}(\alpha)$ of the functions defined in Section II - either conditional on $\mu = o$ or concentrated with respect to μ_{*} respectively. Note that a boundary minimum of the sum of squares functions is usually a constrained minimum, a boundary minimum of U is always unconstrained, while U^{*} always has an interior minimum.

The results of all the 672 experiments (4 criteria x 21 parameter values x 4 sample sizes x 2 models) are not given individually, since interpretation of so much data would be difficult. Instead, various methods of summarising and assessing the results are employed. Table I(a) shows, for each model and sample size, the <u>average</u> of the estimated root mean squared errors for the 21 parameter values, a figure which provides a standard of comparison for the estimators over the parameter space as a whole.

---

(1) See Box, Davies and Swann [1969], p.14.

TABLE I(a) : Mean RMSE (1 to -1)

T		LR	TS
10	S*	0.5007	0.4011
	U*	0.4662	0.3864
	S	0.5015	0.4207
	U	0.5220	0.4239
20	S*	0.3099	0.2675
	U*	0.3069	0.2609
	S	0.3345	0.2761
	U	0.3295	0.2727
40	S*	0.2027	0.1810
	U*	0.1953	0.1393
	S	0.2002	0.1720
	U	0.1933	0.1671
80	S*	0.1399	0.1358
	U*	0.1387	0.1361
	S	0.1321	0.1256
	U	0.1294	0.1250

TABLE I(b) : Mean RMSE (0.9 to -0.9)

T		LR	TS
10	S*	0.5186	0.4036
	U*	0.4654	0.3781
	S	0.5542	0.4650
	U	0.5439	0.4342
20	S*	0.3194	0.2705
	U*	0.3099	0.2584
	S	0.3697	0.3052
	U	0.3503	0.2841
40	S*	0.2063	0.1815
	U*	0.1956	0.1333
	S	0.2197	0.1887
	U	0.2057	0.1761
80	S*	0.1420	0.1365
	U*	0.1395	0.1356
	S	0.1443	0.1364
	U	0.1398	0.1331

TABLE II

Percentage of minima of U falling at -1 and 1 (LR Model)

α	T = 10		T = 20		T = 40		T = 80	
	-1.0 %	1.0 %	-1.0 %	1.0 %	-1.0 %	1.0 %	-1.0 %	1.0 %
1.0	0.0	98.0	0.0	95.0	0.0	96.0	0.0	96.0
0.8	0.0	92.5	0.0	82.0	0.0	70.0	0.0	32.0
0.6	0.5	80.0	0.0	62.0	0.0	32.0	0.0	8.0
0.4	2.0	66.0	0.0	35.0	0.0	12.0	0.0	0.0
0.2	4.5	51.5	0.0	16.0	0.0	4.0	0.0	0.0
0.0	6.5	32.0	1.0	8.0	0.0	2.0	0.0	0.0
-0.2	9.0	24.5	3.0	2.0	0.0	0.0	0.0	0.0
-0.4	16.0	11.5	6.0	1.0	0.0	0.0	0.0	0.0
-0.6	28.5	7.5	13.0	1.0	2.0	0.0	0.0	0.0
-0.8	49.0	5.0	36.0	0.0	10.0	0.0	8.0	0.0
-1.0	56.5	3.0	64.0	0.0	70.0	0.0	84.0	0.0

Note that the sum of squares functions generally have larger RMSE than the unlikelihood functions, both in the conditional (starred) and unconditional case. While on this basis of comparison no criterion appears to be uniformly superior, the conditional functions perform better in small samples, the unconditional functions better in large samples. The result is of some interest in view of the common assumption that the exact likelihood (equiv. U) should be the preferred criterion in small samples - a view which does not take account of the increased propensity of this function to "degenerate" boundary minima in small samples. The extent of this behaviour is shown in Table II which shows the percentage of replications of the LR model in which U has an upper or lower boundary minimum, for a range of parameter values and sample sizes. The comparable figures for the other model and other criteria are not given here, but on the whole S has more boundary minima (constrained), S^* has fewer, and as expected U^* none. The picture for the TS model is similar, taking account of the differing characteristics of this model.

It may be questioned whether the figures in Table I(a) are the best basis for RMSE comparison, since the parameter points -1 and 1 are included in the average. Since in the smaller sample sizes a high proportion of replications produce a boundary minimum for the unconditional functions, the replication variance of these estimators for the boundary points is naturally close to zero. To assess how this fact influences the ranking of the criteria we give the average RMSE over the interior points (-.9, .9) only, in Table I(b). But in choosing a basis for comparison note that the boundaries are points of maximal bias for the conditional estimators.

Another method of assessing the experimental results is to look at the estimated biases in relation to parameter values and sample sizes. We do this in two ways. We plot the mean simulation bias against the parameter

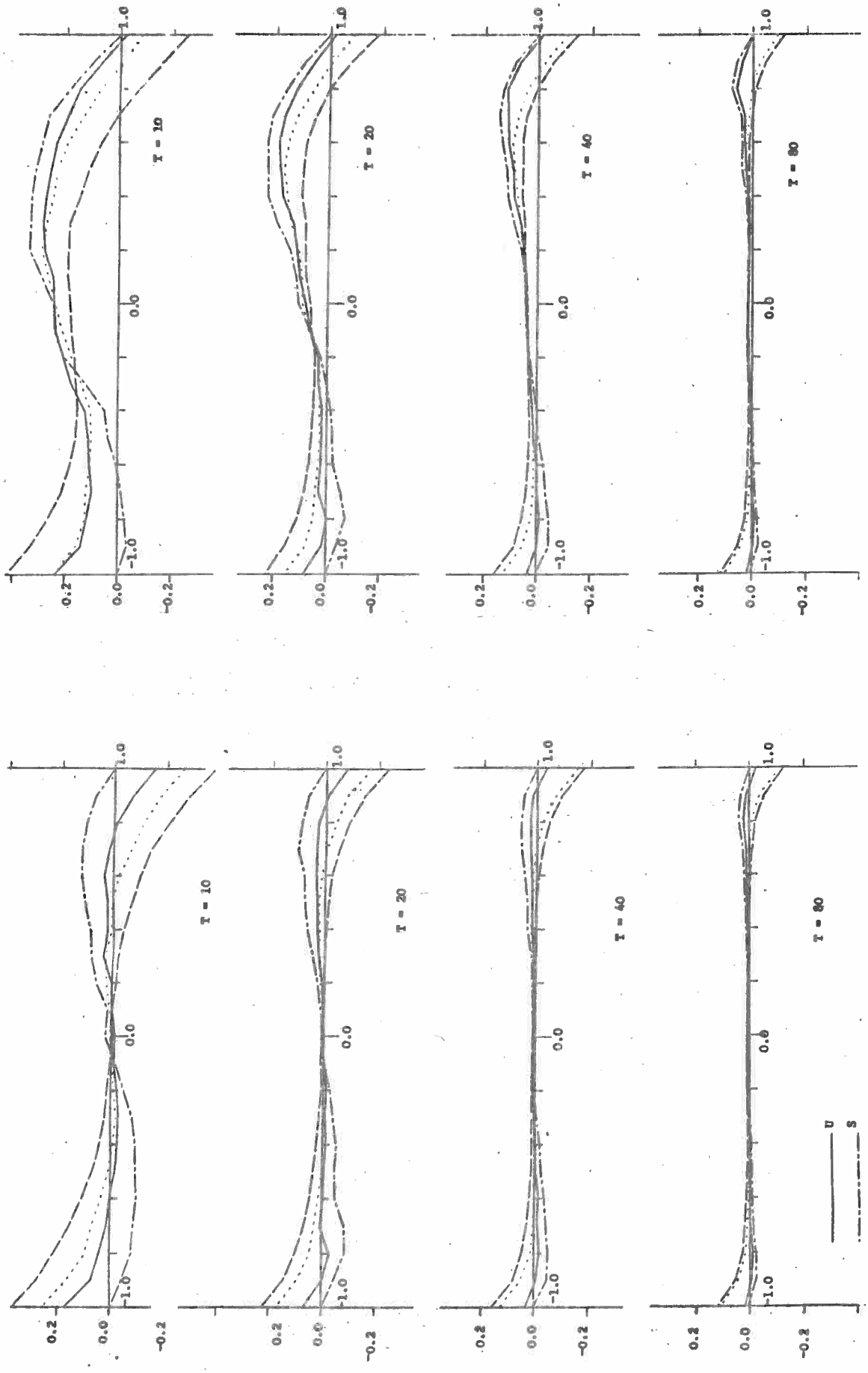

FIGURE II Bias Curves

TABLE III

Results for Bias regressions (t-ratios in brackets)

		$1/\sqrt{T}$	$\alpha/\sqrt{T}$	$\alpha^2/\sqrt{T}$	$\alpha^3/\sqrt{T}$	$\sqrt{T}$	R^2	D.W.
LR	S*	2.2112 (15.5069)	3.4926 (11.1322)	-1.4736 (-6.3890)	-5.9624 (-13.6229)	-0.0073 (-1.8029)	0.7992	1.1134
	U*	1.7970 (13.7825)	1.6994 (5.9238)	-1.0552 (-5.0034)	-5.8126 (-14.5249)	-0.0056 (-1.5214)	0.8913	1.2003
TS	S	2.4076 (15.4470)	5.1825 (15.1123)	-2.1576 (-8.5584)	-4.9070 (-10.2573)	-0.0028 (-0.6269)	0.8466	0.4321
	U	2.2832 (16.5600)	3.4411 (11.3438)	-1.2013 (-5.3869)	-4.2466 (-10.0353)	-0.0072 (-1.8363)	0.7645	0.4798
	S*	0.0788 (0.6161)	1.5747 (5.5970)	-0.2614 (-1.2637)	-4.8934 (-12.4679)	0.0051 (1.3900)	0.8259	1.4863
	U*	0.0780 (0.6358)	0.2062 (0.7636)	-0.2262 (-1.1394)	-4.7248 (-12.5404)	0.0044 (1.2535)	0.9187	1.4821
	S	0.0375 (0.4667)	3.2042 (18.1067)	-0.0531 (-0.4080)	-2.8838 (-11.6818)	0.0057 (2.5019)	0.8598	0.9719
	U	0.0087 (0.1131)	1.5350 (9.1081)	-0.0414 (-0.3345)	-2.7084 (-11.5203)	0.0051 (2.3420)	0.6522	1.0686

points, for each model and sample size, in Figures II(a) and II(b). And we estimate the "bias response functions", that is, assume that for a consistent estimator $\hat{\alpha}_T$ of α, the bias of $O(T^{-1})$ has functional form

$$E(\hat{\alpha}_T - \alpha) = \frac{1}{T} \phi(\alpha) \qquad (4.4)$$

where the function ϕ can be approximated by a polynomial in α. The latter is estimated by least squares, assuming that a cubic is sufficient to represent the bias curves plotted in Figures II(a) and II(b), and that multiplying through by $\sqrt{T}$ is a suitable transformation to ensure homoscedastic errors.(1) Eight regressions were perrormed (for each model and each criterion) on the equation

$$\sqrt{T_j}\, B_j = a_o \sqrt{T_j} + a_1 1/\sqrt{T_j} + a_2 \alpha_j/\sqrt{T_j} + a_3 \alpha_j^2/\sqrt{T_j} + a_4 \alpha_j^3/\sqrt{T_j} + U_j$$

where B_j is the mean bias obtained from simulations with T_j = 10, 20, 40, 80 and α_j = -1, -.9, -.8, ... , .9,1, j = 1, ... , 84. The results are given in Table III.

The R^2's and Durbin Watson statistics can be used as a check on the adequacy of the functional form, and the low values of some of the latter here indicate that a higher order polynomial might be desirable. But the R^2's are quite high, and for present purposes the results should be adequate. Note that the terms in $\sqrt{T}$ (the constants in the regression prior to the heteroscedasticity transformation) are generally insigificant

---

(1) See also Hendry and Trivedi {1972}, Hendry and Harrison {1974}.

which is the expected result when the estimators are consistent. (cf. (4.4)). There are various possible ways of ranking the criteria on the basis of the regressions, but we will only informally compare the coefficients, noting that inside the unit interval the coefficients of $1/\sqrt{T}$ and $\alpha/\sqrt{T}$ will play the most important role. It does appear from these results that in both models, U^* again performs best overall.

It is interesting to note the uniform positive bias in the estimates of the LR model, which the TS model does not exhibit. No immediate explanation of this result is offered, and the estimates of μ (which were recorded although are not reproduced here) show no significant bias. Incidentally, the possibility of programming error is never to be discounted in experiments of this type, but a simple check on the computation of U is provided by the property $U(\alpha) = U(\alpha^{-1})$. The computed function possesses this property to within machine accuracy, and since the other functions are all simplifications of U, the presumption of computational accuracy seems reasonable.

The conclusion to be drawn from these experiments is that no criterion is uniformly superior either in respect of sample size or parameter value, but on balance the advantage in small samples lies clearly with U^*, except in the case where the true parameter value lies sufficiently close to the boundary that a boundary minimum can be regarded as a good estimate.

V. The Distribution of the Estimates

It has been shown, for example by Pierce {1971} that the ML estimates of $\underset{\sim}{\theta}$ are asymptotically normal, and in the LR model are asymptotically independent of the estimates of $\underset{\sim}{\beta}$, provided the parameters lie strictly within the invertibility region. For a process on the boundary, however, the asymptotic distribution appears to be non-normal.(1)

All the estimators we have considered are asymptotically ML, and have the same asymptotic distribution inside the region, but it is apparent that small-sample deviations from normality may be severe, not only because the estimates are only defined over a compact region of the parameter space, but also because, with the exception of U^*, the distribution functions appear to possess discontinuities at the boundaries. This raises serious problems for hypothesis testing, and in particular for the interpretation of boundary estimates. An intensive Monte Carlo exercise which complements the extensive study of the last section is to estimate the distribution functions of the estimators by cumulative frequency functions, that is, step functions defined by (number of estimates $< X$)/N for N replications and a discrete set of points X in the domain of the function. The set {X} chosen was the integral multiples of .01 in the unit inverval, and $N = 500$. The LR model with $\alpha = .5$ and a sample size 40 was chosen as a fairly representative case. The results are plotted in Figures III(a) to III (d); the broken line in each diagram indicates the asymptotic distribution for comparison. The asymptotic variance for the MA(1) model, using the formula derived in Pierce {1971} is $(1-\alpha^2)/T$. Hence the graphed function is the normal distribution function with mean of .5 and standard deviation of .137.

---

(1) The distribution in this case has recently been investigated for the MA(1) process by Sargan {1977}.

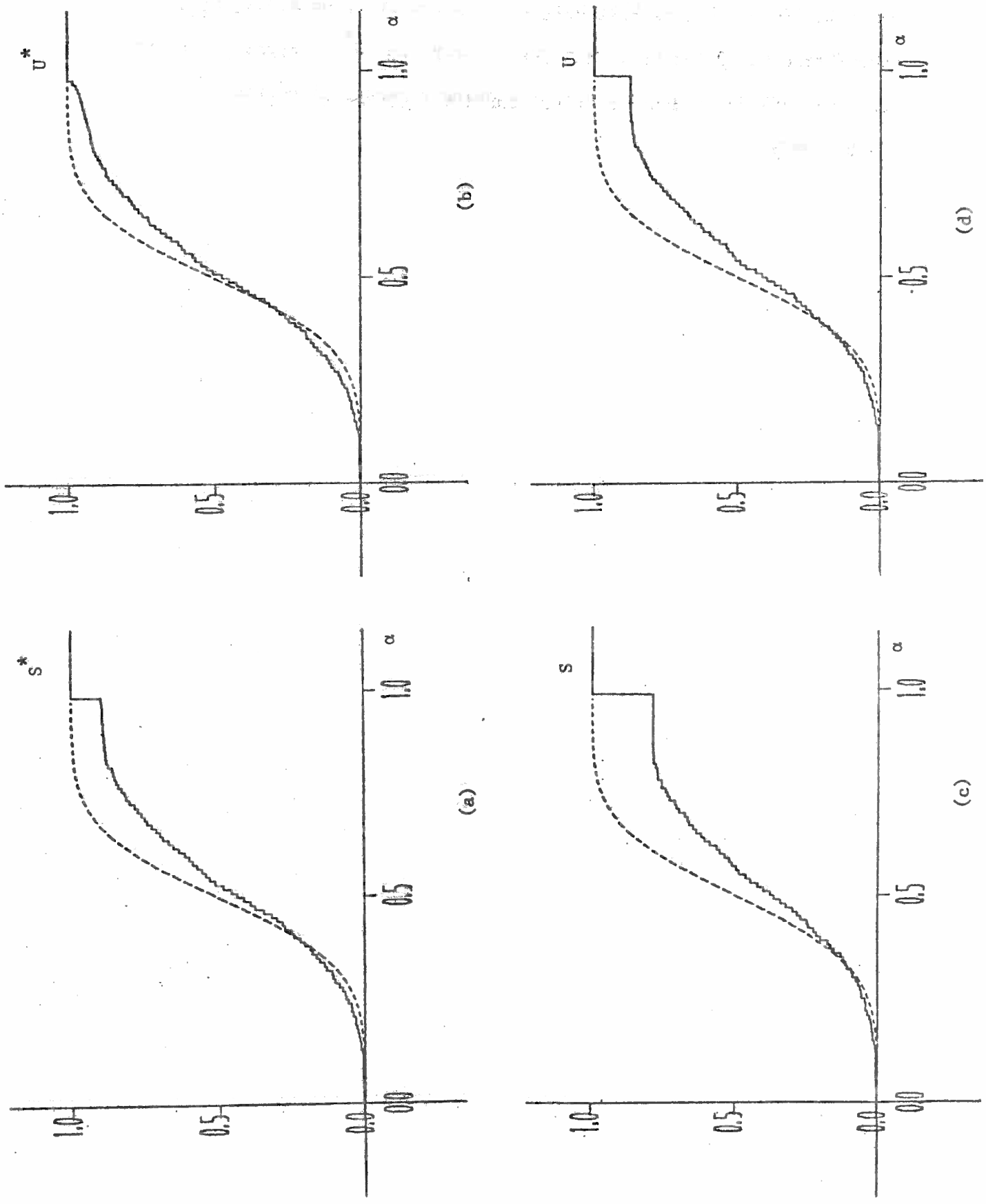

FIGURE III Cumulative Frequency Curves for the model

$u_t = \mu + \varepsilon_t - \alpha \varepsilon_{t-1}$, $\mu = 0$, $\alpha = .5$, $T = 40$

None of the empirical distributions are a close fit to the asymptotic distribution - and note that this is in a decent-sized sample from the econometricians point of view - but clearly the U^* criterion performs best, all the other distrib utions showing a marked discontinuity at the boundary.

VI. Conclusions

Several important conclusions appear to emerge from the foregoing. Firstly, that exact ML (the U criterion) is not an ideal estimator in small samples. Estimates fall on the boundary of the invertibility region with rather high probability, presenting difficulties with interpretation and inference. An example of these difficulties is found in the recent study of Prothero and Wallis (1976), who consider a number of explanations for the frequency with which they obtain boundary estimates for a fourth-order moving average, in the context of exact ML estimation of quarterly Box-Jenkins models (P.479-480). They consider mis-specification as a possible cause, as well as (referring to the Kang paper) the possibility of degenerate boundary maxima of the likelihood; but apparently believe the latter explanation to be implausible in a sample of size 40.(1) While we have not studied the ARMA model here, there is no reason to suppose that our Monte Carlo results are a misleading guide to that case. Reference to Table 2 shows that in a sample of size 40, exact ML estimates of an MA(1) process with true parameter of .8 fell on the boundary seven times out of ten. The Prothero and Wallis estimates, having rather fewer degrees of freedom than this, are not improbable even if the true parameter values are well inside the invertibility region.

The second conclusion is that U^* is the best criterion of those studied. We have noted that the estimator has marginally superior bias and mean squares error characteristics in small samples, although one should not ignore the fact that it is fairly strongly biased away from boundary values. In view of the previous section, we may also surmise that its small sample distribution diverges less drastically from the asymptotic normal approximation than those for the other criteria, and inference using these approximations is correspondingly more reliable.

---

(1) They suggest, I think mistakenly, that estimates of a fourth-order process from forty observations are ".... in effect based on 10 observations."

To these advantages we may add two others; that it is less expensive to calculate than the exact ML estimator, and also that there is no problem, in the first order case and probably generally, with imposing the invertibility constraints during estimation. Whatever the estimator used, it is desirable to impose the constraints, for even when the function possesses an interior minimum, a temporary step outside the invertibility region by the search routine can lead to failure due to rounding errors (because of the powers of order T of the roots involved).

Osborne {1976}, for example, suggests an ingenious transformation to achieve this,(1) but her type of method has the difficulty that whenever the constraint becomes binding, the derivatives of the transformation tend to zero,(2) resulting in a singular Hessian, and failure of a gradient minimisation routine wherever the minimum lies on the boundary. Modified transformations to overcome this are possible, but not wholly satisfactory. A criterion which guarantees an interior minimum is clearly preferred, except only in the case where the true process is on the boundary. The importance of this case should not be under-estimated, but it appears that with the estimators here considered, one must choose between observing a process on the boundary when the true process is not, with a large probability, or never observing a process on the boundary at all. There appears to be a fundamental problem of identifiability for which no simple solution is apparent.

---

(1) Her procedure (with modified notation) is to employ the β_i and γ_i defined by (A.1.1) of Appendix I below as independent variables, and impose the invertibility constraints $|\gamma_i| < 1$, $|\beta_i| < 1+\gamma_i$ by defining $\gamma_i = d_i/(1+|d_i|)$ and $\beta_i = g_i(1+\gamma_i)/(1+|g_i|)$, so that the constraints are satisfied for $-\infty < d_i, g_i < \infty$. The function is then minimised unrestrictedly with respect to the d_i and g_i.

(2) For example, with the change of variable described in the last footnote,

$$\lim_{d_i \to \infty} \left(\frac{\partial \gamma_i}{\partial d_i} \right) = 0.$$

A final conclusion, from the Monte Carlo results, is that if for reasons of computational convenience a sum-of-squares criterion is to be used, then S^* is preferable to S, the latter having the worst small sample characteristics of all the functions studied. On the presumption that in large samples the difference between the criteria becomes negligible, there never appears to be a justification for estimating the starting residuals, in spite of the apparent theoretical desirability of this procedure.

APPENDIX I

Representation of the criterion functions with the roots of the lag polynomial as independent variables.

Observe that $\theta(L) = \prod_{i=1}^{q} (1 - \alpha_i L)$ can be partially solved as

$$\theta(L) = \prod_{i=1}^{q/2} (1 + \beta_i L + \gamma_i L^2) \qquad (A.1.1)$$

when q is even, or

$$\theta(L) = \prod_{i=1}^{(q-1)/2} (1 + \beta_i L + \gamma_i L^2)(1 - \alpha_q L) \qquad (A.1.2)$$

if q is odd, where β_i, γ_i and α_q are always real. (Note: henceforth it will be assumed that q is odd, to avoid tedious duplication of expressions. The modifications in the even case are mostly obvious, and will be noted where necessary.)

The vector-valued function implied by (A.1.1) or (A.1.2),

$$\underset{\sim}{\theta} = \underset{\sim}{h}(\underset{\sim}{t}) \quad : \quad \mathbb{R}^q \to \mathbb{R}^q \qquad (A.1.3)$$

where $\underset{\sim}{t} = (\beta_1, \gamma_1, \beta_2, \gamma_2, \cdots, \beta_{(q-1)/2}, \gamma_{(q-1)/2}, \alpha_q)$

is continuous although not 1-1 in general, and there is no difficulty in making this change of variable and writing

$$C = C(\underset{\sim}{h}(\underset{\sim}{t})) .$$

Considering the pair (β_i, γ_i) - we drop the subscripts whenever the meaning is clear - the original pairs of roots are identical with the roots of the quadratic

$$\lambda^2 + \beta\lambda + \gamma = 0.$$

Complex conjugate pairs of roots will be represented using polar co-ordinates, $\lambda_1 = re^{i\omega}$, $\lambda_2 = re^{-i\omega}$. It will be found convenient to represent a complex conjugate pair as a point in the open subset of $\mathbb{R}^2$,

$$S = \{(r,\omega) : r > 0, 0 < \omega < \pi\}$$

instead of in the more conventional Argand plane.

We then define

$$\underset{\sim}{g}(\lambda_1, \lambda_2) : \mathbb{R}^2 \to T \qquad \text{(A.1.4)}$$

where $T = \{(\beta, \gamma) : -\infty < \beta < \infty, \gamma < \tfrac{1}{4}\beta^2 \}$

$$\underset{\sim}{g} = (g_1, g_2) = (-(\lambda_1 + \lambda_2), \lambda_1 \lambda_2),$$

and $$\underset{\sim}{g}'(r, \omega) : S \to \mathbb{R}^2 - T \qquad \text{(A.1.5)}$$

where $\underset{\sim}{g}' = (g_1', g_2') = (-2r\cos\omega, r^2)$.

i.e. (A.1.4) and (A.1.5) define respectively mappings from pairs of real and complex roots to the corresponding subsets of the real plane T and $\mathbb{R}^2 - T$.

Now letting $p = (q - 1)/2$ (or $q/2$ when q is even), consider the mapping from m complex and $p - m$ real pairs of roots to the corresponding pairs (β_i, γ_i) - in addition to the identity mapping from an odd real root to itself, whenever q is odd. The necessary notation is provided by defining an index set K whose members are ordered $(p + 1)$-tuples denoted $\underset{\sim}{k} = (k_1, k_2, \ldots, k_p, m)$. The first p elements are some ordering of the integers from one to p, (corresponding to the indices in (A.1.2)) such that the first m are the indices of a complex pair, and the last $p-m$ the indices of a real pair of roots; the ordering is otherwise arbitrary, so to avoid indeterminacy assume the natural ordering within each sub-vector. m runs over integral values from 0 to p.

There are $\binom{p}{m}$ possible combinations of m complex and $p-m$ real pairs, and the members of K enumerate all possibilities for each m, $m = 0, 1, \ldots, p$, giving a total of $\sum_m \binom{p}{m} = 2^p$ members.

Now define for each $\underset{\sim}{k} \in K$ a set

$$\Phi_{\underset{\sim}{k}} = \left(\prod_{i=1}^{m} S_{k_i}\right) \times \mathbb{R}^{2(p-m)+1} \subseteq \mathbb{R}^q$$

with typical elements

$$\phi_{\underset{\sim}{k}} = (r_{k_1}, \omega_{k_1}, \ldots, r_{k_m}, \omega_{k_m}, \lambda_{1,k_{m+1}}, \lambda_{2,k_{m+1}}, \ldots,$$

$$\lambda_{1,k_p}, \lambda_{2,k_p}, \alpha_q),$$

where the subscript $\underset{\sim}{k}$ is to be interpreted as a vector-valued index, the member $\underset{\sim}{k} \in K$, such that the subscripts of the elements of $\phi_{\underset{\sim}{k}}$ correspond to the elements of $\underset{\sim}{k}$. Note that the last $2(p-m) + 1$ co-ordinates of $\phi_{\underset{\sim}{k}} \in \Phi_{\underset{\sim}{k}}$ are identical with the real α_i of (1.2).

Similarly define

$$T^*_{\underset{\sim}{k}} = \{ t_{\underset{\sim}{k}} = (\beta_1, \gamma_1, \cdots, \beta_p, \gamma_p, \alpha_q) \mid -\infty < \beta_j < \infty ;$$

$$\gamma_j < \tfrac{1}{4}\beta_j^2 \text{ iff } j = k_i \text{ and } i \geq m, \forall j; -\infty < \alpha_q < \infty\}$$

(A.1.7)

$$= \left(\overset{p}{\underset{j=1}{X}} T'_j\right) X \; R \subseteq R^2$$

where $$T'_j = \begin{cases} T & \text{if } j = k_i, \; i > m \\ R^2 - T & \text{if } j = k_i, \; i \leq m \end{cases}$$

Then by a natural extension of (A.1.4) and (A.1.5) we can define a family of continuous mappings.

$$t_{\underset{\sim}{k}} = G_{\underset{\sim}{k}}(\phi_{\underset{\sim}{k}}) : \Phi_{\underset{\sim}{k}} \to T^*_{\underset{\sim}{k}}, \quad \underset{\sim}{k} \in K \qquad (A.1.8)$$

By a similar extension from (A.1.3), we also define

$$\theta_{\underset{\sim}{k}} = \underset{\sim}{h}(t_{\underset{\sim}{k}}) : T^*_k \to \Theta_k \subseteq R^q \qquad (A.1.9)$$

Important properties of $G_{\underset{\sim}{k}}$ and $\underset{\sim}{h}$ are summarised by the following lemmas:

Lemma 2 The family $T^*_{\underset{\sim}{k}}$, $\underset{\sim}{k} \in K$, forms a partition of R^q.

Proof of Lemma 2:

Assume q is odd - the modifications in the even case are trivial. We are required to prove (i) any $\underset{\sim}{t} \in \mathbb{R}^q$ is in at least one set $T^*_{\underset{\sim}{k}}$, $\underset{\sim}{k} \in K$; (ii) any $\underset{\sim}{t} \in \mathbb{R}^q$ can be in at most one set $T^*_{\underset{\sim}{k}}$, $\underset{\sim}{k} \in K$.

Consider $\underset{\sim}{t} = (\beta_1, \gamma_1, \ldots, \beta_p, \gamma_p, \alpha_q) \in R^q$.

For each pair (β_j, γ_j) either (a) $\gamma_j > \frac{1}{4}\beta_j^2$ or (b) $\gamma_j \leq \frac{1}{4}\beta_j^2$. Arrange the indices j into a p-tuple, those for which (a) holds first followed by those for which (b) holds. Together with the number m of pairs for which (a) holds, this defines an element $\underset{\sim}{k}$ of K, and so by definition $\underset{\sim}{t} \in T^*_{\underset{\sim}{k}}$ for $\underset{\sim}{k} \in K$. This proves (i).

Now suppose that $\exists\, \underset{\sim}{t}$ s.t. $\underset{\sim}{t} \in T^*_{\underset{\sim}{k}'}$ and $\underset{\sim}{t} \in T^*_{\underset{\sim}{k}''}$ where $\underset{\sim}{k}' \neq \underset{\sim}{k}''$. Form the $(p+1)$-tuple $\underset{\sim}{k}$ from $\underset{\sim}{t}$ in the manner just described. If $\underset{\sim}{k} \neq \underset{\sim}{k}'$ then $\underset{\sim}{t} \notin T^*_{\underset{\sim}{k}'}$ and hence $\underset{\sim}{k} = \underset{\sim}{k}'$. Similarly $\underset{\sim}{k} = \underset{\sim}{k}''$, contradicting the assertion that $\underset{\sim}{k}' \neq \underset{\sim}{k}''$. This proves (ii), and hence the lemma.

Corollary. Every boundary point of $T^*_{\underset{\sim}{k}}$ is the limit point of a sequence in $T^*_{\underset{\sim}{k}'}$ for some $\underset{\sim}{k}' \neq \underset{\sim}{k}$.

Lemma 3: The image sets $\theta_{\underset{\sim}{k}}$ of $T^*_{\underset{\sim}{k}}$ under $\underset{\sim}{h}$ form a covering for $\mathbb{R}^q$.

Proof of Lemma 3:

We have to prove that $\underset{\sim}{\theta} \in \mathbb{R}^q$ is in at least one set $\underset{\sim}{\Theta}_k = \underset{\sim}{h}(\underset{\sim}{T}_k^*)$, $\underset{\sim}{k} \in K$.

The function $\underset{\sim}{h}$ is defined by (A.1.3) to be a mapping from $\mathbb{R}^q$ to $\mathbb{R}^q$, and from its functional form derived from (A1.1) or (A1.2), it is evident that

$$\mathbb{R}^q = \underset{\sim}{h}(\mathbb{R}^q) \qquad \text{(A.1.10)}$$

i.e. any point $\underset{\sim}{\theta} \in \mathbb{R}^q$ can be associated with one or more points $\underset{\sim}{t} \in \mathbb{R}^q$ by $\underset{\sim}{h}$. Unless the lemma is true,

$$\exists\ \underset{\sim}{\theta}_k \text{ s.t. } \underset{\sim}{\theta}_k \neq \underset{\sim}{h}(t_k), \qquad \forall\ \underset{\sim}{t}_k \in \underset{\sim}{T}_k^*,\ \underset{\sim}{k} \in K \qquad \text{(A.1.11)}$$

But by Lemma 2, every $\underset{\sim}{t} \in R^q$ is identical with some $\underset{\sim}{t}_k \in \underset{\sim}{T}_k^*$, $\underset{\sim}{k} \in K$, so (A.1.10) is false unless $\exists\ \underset{\sim}{\theta}_k = \underset{\sim}{\theta} \neq \underset{\sim}{h}(\underset{\sim}{t})$, $t \in \mathbb{R}^q$, which contradicts (A.1.11).

Lemma 4: If a point is in $\underset{\sim}{\Phi}_k$, then the point obtained by replacing either a real root r_j or a complex modulus α_j by its' inverse is also in $\underset{\sim}{\Phi}_k$.

The proof of Lemma 4 follows immediately from the definitions.

Using the preceding results it becomes straightforward to define a collection of composite mappings

$$\underset{\sim}{C}_k = \underset{\sim}{Coh} \circ \underset{\sim}{G}_k \qquad \underset{\sim}{\Phi}_k \to \mathbb{R}, \quad \underset{\sim}{k} \in K.$$

Lemmas 2 and 3 ensure that the collection jointly represents the function C over the whole of the original parameter space.

Appendix II

Proof of Theorem 1

(i) When α_j is a real root, observe that (1.2) can be written in a partially factored form as

$$\theta(L) = \theta^*(L)(1 - \alpha_j L) \qquad (A.2.1)$$

$$= (1 + \theta_1^* L + \ldots + \theta_{q-1}^* L^{q-1})$$

$$- \alpha_j(L + \theta_1^* L^2 + \ldots + \theta_{q-1}^* L^q)$$

$$= \sum_{r=o}^{q} (\theta_r^* - \alpha_j \theta_{r-1}^*) L^r$$

where $\theta_{-1}^* = \theta_q^* = 0$ and $\theta_o^* = 0$, so that

$$\theta_r = \theta_r^* - \alpha_j \theta_{r-1}^*, \quad r = 1, \ldots, q. \qquad (A.2.2)$$

The s^{th}-order autocovariance $\sigma^2\omega_s$ is defined by

$$\omega_s = \sum_{r=o}^{q-s} \theta_r \theta_{r+s} \qquad s = o, 1, \ldots, q \qquad (A.2.3)$$

$$= \sum_{r=o}^{q-s} \left[(1 + \alpha_j^2) \theta_r^* \theta_{r+s}^* - \alpha_j(\theta_{r-1}^* \theta_{r+s}^* + \theta_r^* \theta_{r+s-1}^*)\right]$$

substituting from (A.2.2) and rearranging, using the fact that terms containing θ_{-1}^* and θ_q^* are zero. The proof of (i) is immediate on substituting $1/\alpha_j$ for α_j in (A.2.3).

(ii) When α_j is complex, and $\alpha_{j+1} = \bar{\alpha}_j$, write the roots in polar co-ordinates as

$$\alpha_j = |\alpha_j| e^{i\omega_j} \quad \text{and} \quad \bar{\alpha}_j = |\alpha_j| e^{-i\omega_j} .$$

Define $\beta = -2|\alpha_j| \cos \omega_j$ and $\gamma = |\alpha_j|^2$

A suitable partial factorisation is

$$\theta(L) = \theta^+(L)(1 - \alpha_j L)(1 - \bar{\alpha}_j L)$$

$$= \theta^+(L)(1 + \beta L + \gamma L^2) \qquad \text{(A.2.4)}$$

$$= \sum_{r=o}^{q} [\theta_r^+ + \beta\theta_{r-1}^+ + \gamma \theta_{r-2}^+] L^r$$

where $\theta_{-2}^+ = \theta_{-1}^+ = \theta_{q-1}^+ = \theta_q^+ = 0$ and $\theta_o^+ = 1$, and

$$\theta_r = \theta_r^+ + \beta\theta_{r-1}^+ + \gamma\theta_{r-2}^+ . \qquad \text{(A.2.5)}$$

Substituting into the definition of ω_s as before gives

$$\omega_s = \sum_{r=o}^{q-s} [(1 + \beta^2 + \gamma^2)\, \theta_r^+ \theta_{r+s}^+ +$$

$$\beta(1 + \gamma)(\theta_{r-1}^+ \theta_{r+s}^+ + \theta_r^+ \theta_{r+s-1}^+) \qquad \text{(A.2.6)}$$

$$+ (\theta_r^+ \theta_{r+s-2}^+ + \theta_{r-2}^+ \theta_{r+s}^+)]$$

where $1 + \beta^2 + \gamma^2 = 1 + |\alpha_j|^2 \cos^2 \omega_j + |\alpha_j|^4$

$$\beta(1 + \gamma) = -(|\alpha_j| + |\alpha_j|^3) \cos \omega_j$$

$$\gamma = |\alpha_j|^2 .$$

Since $\frac{1}{\alpha_j} = \frac{1}{|\alpha_j|} e^{-i\omega}$ and $\frac{1}{\bar{\alpha}_j} = \frac{1}{|\alpha_j|} e^{i\omega}$, inversion of the roots in pairs is equivalent to inversion of the modulus of the pair. Hence, the proof of (ii) is immediate on substituting $|\frac{1}{\alpha_j}|$ for $|\alpha_j|$ in (A.2.6).

Appendix III

Proof of Theorem 2:

By assumption the joint density of the observed realisation $\underset{\sim}{u}$ and the pre-sample disturbances $\underset{\sim}{\varepsilon}_o$ is

$$f(\underset{\sim}{u}, \underset{\sim}{\varepsilon}_o ; \underset{\sim}{\theta}, \sigma^2) = (2 \Pi \sigma^2)^{-(T+q)/2} \exp \{ - \frac{\underset{\sim}{\varepsilon}' \underset{\sim}{\varepsilon}}{2 \sigma^2} \} \quad (A.3.1)$$

where $\underset{\sim}{\varepsilon}$ is the $(T+q)$ -vector of independent random variates defined by (2.3).

(A.3.1) is related to the joint density of $\underset{\sim}{u}$ by

$$f(\underset{\sim}{u}, \underset{\sim}{\varepsilon}_o ; \underset{\sim}{\theta}, \sigma^2) = f(\underset{\sim}{u}; \underset{\sim}{\theta}, \sigma^2) \cdot f(\underset{\sim}{\varepsilon}_o | \underset{\sim}{u} ; \underset{\sim}{\theta}, \sigma^2) \quad (A.3.2)$$

It is therefore required to evaluate $f(\underset{\sim}{\varepsilon}_o | \underset{\sim}{u}; \underset{\sim}{\theta}, \sigma^2)$. First premultiply (2.3) by $(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}'$ and rearrange to give

$$\underset{\sim}{\varepsilon}_o = (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}' (\underset{\sim}{\varepsilon} - \underset{\sim}{W}\, \underset{\sim}{u}) \quad (A.3.3)$$

$\underset{\sim}{\varepsilon}_o$ is normally distributed with conditional mean vector

$$E(\underset{\sim}{\varepsilon}_o | \underset{\sim}{u}) = -(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{W}\, \underset{\sim}{u}$$

$$= \hat{\underset{\sim}{\varepsilon}}_o \quad \text{by definition (2.7)}$$

and covariance matrix

$$E\left(\underset{\sim}{\varepsilon}_o - \hat{\underset{\sim}{\varepsilon}}_o\right)\left(\underset{\sim}{\varepsilon}_o - \hat{\underset{\sim}{\varepsilon}}_o\right)' \,|\, \underset{\sim}{u}) = \sigma^2 (X'X)^{-1}$$

Hence

$$f\left(\underset{\sim}{\varepsilon}_o \,|\, \underset{\sim}{u};\ \underset{\sim}{\theta},\ \sigma^2\right) = (2\Pi\sigma^2)^{-q/2} \,|X'X\,|^{\frac{1}{2}} \exp\left\{-\frac{(\underset{\sim}{\varepsilon}_o - \hat{\underset{\sim}{\varepsilon}}_o)'X'X(\underset{\sim}{\varepsilon}_o - \hat{\underset{\sim}{\varepsilon}}_o)}{2\sigma^2}\right\}. \tag{A.3.4}$$

From (2.3) and (2.7), $\underset{\sim}{\varepsilon} = \hat{\underset{\sim}{\varepsilon}} + X\,(\underset{\sim}{\varepsilon}_o - \hat{\underset{\sim}{\varepsilon}}_o)$

so that

$$\underset{\sim}{\varepsilon}'\,\underset{\sim}{\varepsilon} = \hat{\underset{\sim}{\varepsilon}}'\,\hat{\underset{\sim}{\varepsilon}} + (\underset{\sim}{\varepsilon}_o - \hat{\underset{\sim}{\varepsilon}}_o)'X'X(\underset{\sim}{\varepsilon}_o - \hat{\underset{\sim}{\varepsilon}}_o) \quad \text{since} \tag{A.3.5}$$

$$\hat{\underset{\sim}{\varepsilon}}'\,X = \underset{\sim}{0}.$$

Hence combining (A.3.1), (A.3.2), (A.3.4) and (A.3.5),

$$f(\underset{\sim}{u};\ \underset{\sim}{\theta}, \sigma^2) = (2\,\Pi\,\sigma^2)^{-T/2} \,|X'X|^{-\frac{1}{2}} \exp\left\{-\frac{\hat{\underset{\sim}{\varepsilon}}\,\hat{\underset{\sim}{\varepsilon}}}{2\sigma^2}\right\} \tag{A.3.6}$$

which is equivalent to (2.4).

Appendix IV

Proof of Lemma 1 :

It may be deduced from the definitions that the elements of $\underset{\sim}{X}$ are defined by the recursive formula

$$x_{jt} = -\theta_1 x_{jt-1} - \dots - \theta_q x_{jt-q} , \qquad (A.4.1.)$$

$$t = 1 + q , \dots , T + q, \quad j = 1, \dots , q,$$

the starting values $x_{j1} , \dots , x_{jq}$ being provided by the j^{th} column of the q-order identity matrix. Assuming all the roots to be distinct, the solution of the difference equation (A.4.1) is

$$x_{jt} = \sum_{i=1}^{q} A^{ij} \alpha_i^t \qquad (A.4.2)$$

where the coefficients A^{ij} are the $(i,j)^{th}$ elements of $\underset{\sim}{A}^{-1}$, and

$$\underset{\sim}{A} = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots\dots & \alpha_q \\ \alpha_1^2 & \alpha_2^2 & & \vdots \\ \alpha_1^3 & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \alpha_1^q & \alpha_2^q & \dots\dots & \alpha_q^q \end{bmatrix} \quad (q \times q)$$

Hence, the $(j,\ell)^{th}$ element of $\underset{\sim}{X'PX}$ is

$$\sum_{t=1}^{T+q} \sum_{s=1}^{T+q} P_{ts} x_{jt} x_{\ell s} \qquad (A.4.3)$$

$$= \sum_{i=1}^{q} \sum_{k=1}^{q} A^{ij} A^{k\ell} \sum_{t=1}^{T+q} \sum_{s=1}^{T+q} P_{ts} \alpha_i^t \alpha_k^s \quad \text{from (A.4.2)}$$

The elements of $\underset{\sim}{A}^{-1}$ and by assumption $\underset{\sim}{P}$ are $O(1)$ as $T \to \infty$. Hence to determine the order of magnitude of (A.4.3) as $T \to \infty$ it is sufficient to consider the order of the double power series

$$\sum_{t=1}^{T+q} \sum_{s=1}^{T+q} P_{ts} \alpha_d^{t+s} \qquad \text{(A.4.4)}$$

where α_d is the dominant root, and P_{ts}, $t,s = 1, \ldots, T+q$ is a sequence of bounded variables.

Now consider the case where a root, say α_1, has multiplicity m. (There is no need to consider the case of more than one multiple root). The solution to (A.4.1) becomes

$$x_{jt} = A^{1j} \alpha_1^t + A^{2j} t \alpha_1^t + \ldots + A^{mj} t^{m-1} \alpha_1^t + \sum_{i=m+1}^{q} A^{ij} \alpha_i^t \qquad \text{(A.4.5.)}$$

where

$$\underset{\sim}{A} = \begin{pmatrix} \alpha_1 & \alpha_1 \cdots \alpha_1 & \alpha_{m+1} \cdots \alpha_q \\ \alpha_1^2 & 2\alpha_1^2 \cdots 2^{m-1}\alpha_1^2 & \alpha_{m+1}^2 \cdots \alpha_q^2 \\ \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\ \alpha_1^q & q\alpha_1^q \cdots q^{m-1}\alpha_1^q & \alpha_{m+1}^q \cdots \alpha_q^q \end{pmatrix}$$

From consideration of the analogue of (A.4.3) corresponding to this case, it is evident that to determine the order of magnitude of $\sum_t \sum_s P_{ts} x_{jt} x_{\ell s}$ as $T \to \infty$, when α_d has multiplicity m, it is sufficient to consider the order of

$$\sum_{t=1}^{T+q} t^{m-1} \alpha_d^t \sum_{s=1}^{T+q} P_{ts} s^{m-1} \alpha_d^s \qquad (A.4.6)$$

This reduces to (A.4.4) when $m = 1$, so we consider (A.4.5) as the general case.

Using standard results on the convergence of power series[1], (A.4.6) is convergent if α_d is inside the unit circle, and divergent if it is outside. This proves part (i) of the Lemma.

We assume $|\alpha_d| > 1$, and consider

$$\frac{1}{T^{2(m-1)}|\alpha_d^{2T}|} \sum_{t=1}^{T} t^{m-1} |\alpha_d^t| \sum_{s=1}^{T} P_{ts} s^{m-1} |\alpha_d^s| \qquad (A.4.7)$$

$$= \sum_{t=1}^{T} \left(\frac{t}{T}\right)^{m-1} |\alpha_d^{t-T}| \sum_{s=1}^{T} P_{ts} \left(\frac{s}{T}\right)^{m-1} |\alpha_d^{s-T}|$$

$$= \sum_{t=0}^{T-1} \left(1 - \frac{t}{T}\right)^{m-1} |\alpha_d^{-t}| \sum_{s=0}^{T-1} P_{T-t,T-s} \left(1-\frac{s}{T}\right)^{m-1} |\alpha_d^{-s}|$$

This series is convergent, which is sufficient to prove (ii).

If $|\alpha_d| = 1$, consider

$$\frac{1}{T^{2(m-1)}} \sum_{t=1}^{T} t^{m-1} \sum_{s=1}^{T} P_{ts} s^{m-1} \qquad (A.4.8)$$

---

(1) See Apostol {1974}, Chapters 8 and 9.

$$= \sum_{t=o}^{T-1} \left(1 - \frac{t}{T}\right)^{m-1} \sum_{s=o}^{T-1} P_{T-t,T-s} \left(1 - \frac{s}{T}\right)^{m-1} .$$

(A.4.8)

/has the same order of magnitude as $\sum_{t,s} P_{ts}$, and $\sum_{t,s} P_{ts} = O(f(T))$,

where $1 \leqslant f(T) \leqslant T^2$.

This proves (iii).

References:

Apostol, T.M. {1974} Mathematical Analysis (2nd Edn). Addison-Wesley.

Åström, K.J. and Bohlin, T. {1966}. "Numerical identification of linear dynamic systems from normal operating records", in P.H. Hammond (ed.), Theory of Self Adaptive Control Systems, New York, Plenum Press.

Box, M.J., Davies, D. and Swann, W.H. {1969} Non-Linear Optimisation Techniques, Mathematical and Statistical Techniques for Industry, I.C.I., Monograph No. 5, Oliver and Boyd.

Box, G.E.P. and Jenkins, G.M. {1970} . Time Series Analysis, Forecasting and Control. Holden-Day.

Hendry, D.F. and Harrison, R.W. {1974}. "Monte-Carlo Methodology and the Small Sample Behaviour of Ordinary and Two-Stage Least Squares". Journal of Econometrics, Vol. 2, No. 2.

Hendry, D.F. and Trivedi, P.K. {1972} "Maximum Likelihood Estimation of Difference Esquations with Moving Average Errors: A Simulation Study", Review of Economic Studies, Vol. 39 (2), No. 118, 117-146.

Kang, K.M. {1973} "A Comparison of Least Squares and Maximum Likelihood Estimation for Moving Average Processes". Unpublished paper, Australian Bureau of Census and Statistics, Canberra.

Osborn, D.R. {1976} "Maximum Likelihood Estimation of Moving Average Processes". Annals of Economic and Social Measurement, 5, p. 75-87.

Pagan, A.R., and Nicholls, D.F. {1976} "Exact Maximum Likelihood Estimation of Regression Models with Finite Order Moving Average Errors." Review of Economic Studies, Vol. 43 (3), No. 135, p. 383-387.

Pierce, D.A. {1971} "Least Squares estimation in the regression model with autoregressive-moving average errors." Biometrika, 58, 2, p. 299-312.

Prothero, D.L. and Wallis, K.F. {1976} "Modelling Macroeconomic Time Series". Journal of the Royal Statistical Society, Series A, 139, Part 4, p. 468-486.

Sargan, J.D. {1977} "Estimating the First-Order Moving Average when the True Value of the Coefficient is One". London School of Economics Working Paper.

Trivedi, P.K. {1970} "Inventory Behaviour in U.K. Manufacturing, 1956-67". Review of Economic Studies, Vol. 37(4), No. 112, p. 517-536.