QÜESTIIÓ, Vol 13, n° 1,2,3 pp. 13-30, 1989

BIQUADRATIC FUNCTIONS : STATIONARY AND INVERTIBILITY IN ESTIMATED TIME-SERIES MODELS.

D.S.G. POLLOCK

University of London

It is important that the estimates of the parameters of an autoregressive moving-average (ARMA) model should satisfy the conditions of stationarity and invertibility. It can be shown that the unconditional maximum-likelihood estimates are bound to fulfill these conditions regardless of the size of the sample from which they are derived; and, in some quarters, it has been argued that they should be used in preference to any other estimates when the size of the sample is small. However, the maximum-likelihood estimates are difficult to obtain; and, in practice, estimates are usually derived from a least-squares criterion. In this paper, we show that, if an appropriate form of least-squares criterion is adopted, then we can likewise guarantee that the conditions of stationarity and invertibility will be fulfilled. We also re-examine several of the alternative procedures for estimating ARMA models to see whether the criterion functions from which they are derived have the appropriate form.

Keywords: ARMA Models, Least-Squares Estimation, Stationarity and Invertibility

1. INTRODUCTION

When we use the conditional and unconditional least-squares criteria in estimating the parameters of an ARMA process, we run the risk of deriving estimates which violate the conditions of stationarity and invertibility. The danger is greatest when the poles of the autoregressive operator and the zeros of the moving-average operator are close to the boundary of the unit circle.

–D.S.G Pollock - Dept. of Economics - Queen Mary College - Mile End Road - London E1 4NS

–Article rebut el setembre de 1989.

The problematic nature of the least-squares estimators of moving-average models, which are liable to violate the condition of invertibility, was demonstrated in a widely-read but unpublished paper by Kang [10]. The problem was also analysed in the context of the MA(1) model by Osborn [14] who derived expressions for the expected values of various criterion functions. A similar analysis was conducted by Davidson [6] who used the method of Monte-Carlo experiments.

The analogous problems with the least-squares estimators of autoregressive models, which are liable to violate the condition of stationarity, have been highlighted by Wallis [16].

In view of these results, it has become clear to many practitioners that the appropriate way to estimate ARMA models is to adopt the unconditional maximum-likelihood criterion. However, the criterion is difficult to fulfil; and, in practice, estimates are derived more often from one of the least-squares criteria.

The purpose of this paper is to demonstrate that, by adopting the appropriate form of the least-squares criterion, it is possible to guarantee that the conditions of stationarity and invertibility will be fulfilled. All that is required is that the criterion function should be in the form of what we shall describe as a biquadratic function. This point is established at the beginning of the paper. In the remainder of the paper, we look at the various ways in which estimators may be derived which minimise a biquadratic criterion function. This leads us to a new proposal for an estimating system.

This paper places several of the alternative procedures for estimating ARMA models in a juxtaposition and it reveals some of their common characteristics. The synoptic comparisons which this should enable may help in overcoming the bewilderment which can arise from the fact that so many different approaches have been taken to the problem of estimation.

2. BIQUADRATIC FUNCTIONS

Definition.

Let $\alpha(z) = \alpha_0 + \alpha_1 z + \cdots + \alpha_p z^p$ and $y(z) = y_0 + y_1 z + \cdots + y_n z^n$ be the z-transforms of the sequences or vectors $\alpha = \{\alpha_0, \alpha_1, \ldots, \alpha_p\}$, and $y = \{y_0, y_1, \ldots, y_n\}$. Then the product of the z-transforms is given by

$$\begin{aligned} \alpha(z)y(z) &= \sum_t \left(\sum_l \alpha_l y_{t-l} \right) z^t \\ &= \sum_t g_t z^t. \end{aligned} \tag{1}$$

Definition.

Let $\alpha = \{\alpha_i\}$, $y = \{y_i\}$ be sequences whose convolution is the sequence $g = \{g_i\}$. Let $\alpha(z)$, $y(z)$ and $g(z) = \alpha(z)y(z)$ be the z-transforms of the sequences. Then the biquadratic function of α and y is the function

$$
\begin{aligned}
Q(\alpha, y) &= \tfrac{2}{2\pi i} \oint \alpha(z)y(z)y(z^{-1})\alpha(z^{-1})\tfrac{dz}{z} \\
&= \tfrac{1}{2\pi i} \oint g(z)g(z^{-1})\tfrac{dz}{z} \\
&= \sum_t g_t^2,
\end{aligned} \tag{2}
$$

where the contour of integration encloses the origin.

Remark.

To understand the final equality in this definition, consider writing

$$
\begin{aligned}
g(z)g(z^{-1}) &= \left(\sum_r g_r z^r\right)\left(\sum_s g_s z^{-s}\right) \\
&= \sum_r \sum_s g_r g_s z^{r-s} \\
&= \sum_t \left(\sum_r g_r g_{r-t}\right) z^t.
\end{aligned} \tag{3}
$$

Putting this back into the definition gives

$$
\begin{aligned}
Q(\alpha, y) &= \tfrac{1}{2\pi i} \oint \sum_t \left(\sum_r g_r g_{r-t}\right) z^t \tfrac{dz}{z} \\
&= \sum_t \left\{ \sum_r g_r g_{r-t} \tfrac{1}{2\pi i} \oint z^t \tfrac{dz}{z} \right\}.
\end{aligned} \tag{4}
$$

The final equality of the definition now follows from the Cauchy Integral Theorem—see Kreyszig [11], for example—which indicates that

$$
\frac{1}{2\pi i} \oint z^t \frac{dz}{z} = \begin{cases} 1, & \text{if } t = 0 \\ 0, & \text{if } t \neq 0 \end{cases} \tag{5}
$$

A familiar specialisation of this result comes from taking the perimeter of the unit circle as the contour of integration. Then, by setting $z = e^{i\omega}$ and changing the variable of integration from z to $\omega \in (-\pi, \pi]$, we get

$$(6) \qquad \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t}d\omega = \frac{1}{\pi}\int_{0}^{\pi} \cos(\omega t)d\omega = \begin{cases} 1, & \text{if } t = 0; \\ 0, & \text{if } t \neq 0. \end{cases}$$

It is clear that, however we choose to express it, the biquadratic function $Q(\alpha, y)$ is simply the square of the coefficient associated with z^0 in the product $g(z)g(z^{-1}) = \alpha(z)y(z)y(z^{-1})\alpha(z^{-1})$.

There is also a matrix representation of the biquadratic function. It is best to develop this by way of an example:

Example.

Let $\alpha(z) = \alpha_0 + \alpha_1 z$ and $y(z) = y_0 + y_1 z + y_2 z^2$. Then

$$(7) \qquad \alpha(z)y(z) = \alpha_0 y_0 + \{\alpha_0 y_1 + \alpha_1 y_0\}z + \{\alpha_0 y_2 + \alpha_1 y_1\}z^2 + \alpha_1 y_2 z^3.$$

Consider also the equation

$$(8) \qquad \begin{bmatrix} y_0 & 0 & 0 \\ y_1 & y_0 & 0 \\ y_2 & y_1 & y_0 \\ 0 & y_2 & y_1 \\ 0 & 0 & y_2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha_0 y_0 \\ \alpha_0 y_1 + \alpha_1 y_0 \\ \alpha_0 y_2 + \alpha_1 y_1 \\ \alpha_1 y_2 \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha_0 & 0 & 0 \\ \alpha_1 & \alpha_0 & 0 \\ 0 & \alpha_1 & \alpha_0 \\ 0 & 0 & \alpha_1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix}$$

which we can write as

$$(9) \qquad Y_2\alpha_2 = g = A_2 y_2.$$

This is just a matrix representation of the convolution of $\alpha_2 = \{\alpha_0, \alpha_1, 0\}$ and $y_2 = \{y_0, y_1, y_2\}$. The corresponding biquadratic function is given by

$$(10) \qquad \begin{aligned} Q(\alpha, y) &= \alpha_2' Y_2' Y_2 \alpha_2 = y_2' A_2' A_2 y_2 \\ &= g'g, \end{aligned}$$

where $Y_2'Y_2$ and $A_2'A_2$ are both Toeplitz matrices.

The subscripts on the matrices and the vectors are to give an indication of their order. The subscript on the vector y_2 is the degree of the polynomial

$y(z)$ which is one less than the length of the sequence $y = \{y_0, y_1, y_2\}$. Since they contain the same elements, there is, in fact, no need to make a notational distinction between y and y_2. The vector α_2 if formed by supplementing or "padding" the sequence α until its length becomes equal to that of y_2. These matters are formalised in the following definition and Lemma:

Definition.

Let $\alpha = \{\alpha_0, \ldots, \alpha_p\}$ be a vector of order $p+1$. Then we define $\alpha_n = \alpha_n(\alpha) = \{\alpha_{n0}, \ldots, \alpha_{nn}\}$ to be a vector of order $n+1$ which is specified by $\alpha_n = \{\alpha_0, \ldots, \alpha_n\}$ if $n \leq p$ and by $\alpha_n = \{\alpha_0, \ldots, \alpha_p, 0, \ldots, 0\}$ if $n > p$. We define A_n to be a matrix of order $(2n+1) \times (n+1)$ whose columns are the vectors $\{\alpha_{n0}, \ldots, \alpha_{nn}, 0, \ldots, 0\}, \{0, \alpha_{n0}, \ldots, \alpha_{nn}, 0, \ldots, 0\}, \ldots, \{0, \ldots, 0, \alpha_{n0}, \ldots, \alpha_{nn}\}$ We shall describe A_n as a semi-circulant matrix. The relationship between Toeplitz matrices and semi-circulant matrices is indicated by the following Lemma:

LEMMA.

Let $C = [c_{ij}]$ be a positive-definite Toeplitz matrix whose ijth elements is $c_{ij} = c_{|i-j|}$. Then the Cramér–Wald factorisation enables us to write $C = Y_n' Y_n$ where Y_n is a semi-circulant matrix.

The lemma serves to show that, if $\alpha_n = \{\alpha_0, \ldots, \alpha_n\}$ and if C is a Toeplitz matrix, then we have $\alpha_n' C \alpha_n = \alpha_n' Y_n' Y_n \alpha_n = y_n A_n' A_n y_n$ which is a biquadratic function of α_n and y_n. An algorithm for obtaining the Cramér–Wald factorisation has been described by Wilson [17]; and his implementation of it is to be found amongst the programs described by Box and Jenkins [4, Program 2]. The algorithm has also been implemented by Laurie [12], [13].

Now we are in a position to provide the theorem which will be used in demonstrating that the Yule–Walker estimates fulfil the conditions of stationarity:

THEOREM 1.

Consider the biquadratic function $Q(\alpha, y) = \alpha_n' Y_n' Y_n \alpha_n$ where $\alpha_n = \alpha_n(\alpha)$ with $\alpha = \{1, \alpha_1, \ldots, \alpha_p\}$. If α is chosen so as to minimise the value of Q, then all the roots of equation $\alpha(z) = 1 + \alpha_1 z + \cdots + \alpha_p z^p = 0$ will lie outside the unit circle.

Proof.

In place of $Q(\alpha, y)$, let us consider the equivalent function

$$S = \frac{\alpha_n' Y_n' Y_n \alpha_n}{n+1} = \alpha_p' C_p \alpha_p, \tag{11}$$

where $C_p = [c_{ij}]$ is the principal minor of order p of the matrix $Y_n'Y_n/(n+1)$ and where $\alpha_p = \{1, \alpha_1, \ldots, \alpha_p\}$. If we factorise $\alpha(z)$ as $\alpha(z) = q(z)\phi(z)$ where $q(z) = 1 + q_1 z + q_2 z^2$ is quadratic, then, for a given value of $\phi(z)$, we can write the function as

$$S(q) = [1 \quad q_1 \quad q_2] \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix} \begin{bmatrix} 1 \\ q_1 \\ q_2 \end{bmatrix}, \tag{12}$$

where $\gamma_k = \sum_i \sum_j \phi_i \phi_j c_{|i-j+k|}$. At the point of the minimum, we find that

$$\begin{aligned} q_1 &= \tfrac{\gamma_1\gamma_2 - \gamma_0\gamma_1}{\gamma_0^2 - \gamma_1^2}, \\ q_2 &= \tfrac{\gamma_1^2 - \gamma_0\gamma_2}{\gamma_0^2 - \gamma_1^2} \quad \text{and} \\ S &= \gamma_0 + q_1\gamma_1 + q_2\gamma_2. \end{aligned} \tag{13}$$

In terms of these values, we can can express $\gamma_0, \gamma_1, \gamma_2$ as

$$\begin{aligned} \gamma_0 &= \tfrac{S(1+q_2)}{d} \\ \gamma_1 &= \tfrac{-Sq_1}{d} \quad \text{and} \\ \gamma_2 &= \tfrac{S\{q_1^2 - q_2(1+q_2)\}}{d}, \quad \text{where} \\ d &= (1-q_2)(1+q_2+q_1)(1+q_2-q_1). \end{aligned} \tag{14}$$

Now the matrix of (12) is positive definite by virtue of its construction. In particular, the principal minor of order 2 must be positive definite which is equivalent to the conditions that $\gamma_0 > 0$ and $\gamma_0^2 - \gamma_1^2 > 0$. Given that $S > 0$, these imply that

$$\begin{aligned} &(1+q_2)^2 > q_1^2 \quad \text{and} \\ &\tfrac{1+q_2}{1-q_2} > 0 \quad \text{or, equivalently,} \quad 1 - q_2^2 > 0. \end{aligned} \tag{15}$$

The latter conditions are necessary and sufficient to ensure that the roots of $q(z) = 0$ lie outside the unit circle. They are equivalent to the conditions listed under (3.2.18) by Box and Jenkins [4, p. 58].

We can repeat this analysis for every other quadratic factor of the polynomial $\alpha(z)$ in order to show that all of the complex roots must lie outside the unit circle when S is minimised. It is also easy to show that the real roots must lie outside the unit circle.

3. AUTOREGRESSIVE ESTIMATION

The Yule–Walker Equations

Consider a stationary autoregressive process of order p which is represented by the equation

$$\alpha(L)y(t) = \varepsilon(t), \tag{16}$$

where $\alpha(L) = \alpha_0 + \alpha_1 L + \cdots + \alpha_p L^p$, which has $\alpha_0 = 1$, is a polynomial of degree p in the lag operator L, and where $\varepsilon(t)$ is a white-noise process with $E\{\varepsilon(t)\} = 0$ and $V\{\varepsilon(t)\} = \sigma_\varepsilon^2$. The process is stationary if and only if it has an alternative representation as an infinite-order moving-average process:

$$\begin{aligned} y(t) &= \alpha^{-1}(L)\varepsilon(t) \\ &= \theta(L)\varepsilon(t), \end{aligned} \tag{17}$$

where $\theta(L) = \{\theta_0 + \theta_1 L + \cdots\}$ is of infinite degree with $\sum |\theta_i| < \infty$. For stationarity it is necessary and sufficient that the roots of the equation $\alpha(z) = 0$ lie outside the unit circle.

Given a sample of T mean-adjusted observations in $y = \{y_0, y_1, \ldots, y_{T-1}\}$, the Yule–Walker estimates of the elements of $\alpha = \{1, \alpha_1, \ldots, \alpha_p\}$ may be obtained by minimising the function

$$\begin{aligned} S(\hat{\alpha}, y) &= \tfrac{1}{2\pi i T} \oint \hat{\alpha}(z)y(z)y(z^{-1})\hat{\alpha}(z^{-1})\tfrac{dz}{z} \\ &= \oint \hat{\alpha}(z)\hat{\alpha}(z^{-1})I(z)\tfrac{dz}{iz} \\ &= \hat{\alpha}' C_p \hat{\alpha} \end{aligned} \tag{18}$$

in respect of $\hat{\alpha}$. Here the Toeplitz matrix C_p, which is the usual estimate of the covariance matrix, is a minor of order $p+1$ of $C(y) = Y'_{T-1}Y_{T-1}/T$. When $z = e^{i\omega}$, the function

$$I(z) = \frac{y(z)y(z^{-1})}{2\pi T} \tag{19}$$

becomes the periodogram. The empirical autocovariance of lag τ may also be expressed, in terms of the periodogram, as

$$\begin{aligned} c_\tau &= \oint z^\tau I(z)\tfrac{dz}{iz} \\ &= \int_{-\pi}^{\pi} e^{i\omega\tau} I(e^{i\omega}) d\omega. \end{aligned} \tag{20}$$

Remark.

The statistical consistency of the Yule–Walker estimates can be seen as a consequence of the convergence in mean square of the function $S(\hat{\alpha}, y)$ to the function

$$\begin{aligned}\bar{S}(\hat{\alpha}) &= \oint \hat{\alpha}(z)\hat{\alpha}(z^{-1})f(z)\frac{dz}{iz} \\ &= \hat{\alpha}'\Gamma_p\hat{\alpha},\end{aligned} \tag{21}$$

where

$$f(z) = \frac{\sigma_\varepsilon^2}{2\pi\alpha(z)\alpha(z^{-1})}. \tag{22}$$

The latter is the spectral density function of the AR process when $z = e^{i\omega}$, whilst Γ_p is the true variance-covariance matrix of order p pertaining to the process.

The convergence of S to $\bar{S}$ follows in consequence of a spectral result which is proved, for example, by Priestley [6, Theorem 6.2.4]. It also follows as a consequence of the convergence of C_p to Γ_p. Given that the leading coefficients of both $\hat{\alpha}(z)$ and $\alpha(z)$ are unity, the asymptotic function $\bar{S}$ attains its minimum value of σ_ε^2 when $\hat{\alpha}(z) = \alpha(z)$. A fundamental theorem—which is proved by Amemiya [1, Theorem 4.1.1] and by Domowitz and White [7, Theorem 2.2] amongst others—serves to show that if S tends in probability, uniformly, to the function $\bar{S}$, then the values which minimise S will tend to those which minimise $\bar{S}$. Thus the Yule–Walker estimates are statistically consistent.

The function $\bar{S}(\hat{\alpha})$ gives the variance $V\{e(t)\}$ of the residual sequence $e(t)$ which comes from fitting the AR model $\hat{\alpha}(L)y(t) = e(t)$ to the infinite data sequence $y(t)$. The hallmark of the biquadratic criterion functions is that they closely mimic the form of this asymptotic function.

Some of the other methods which are commonly used in estimating autoregressive models are subject to the hazard that, if the roots of the polynomial $\alpha(z) = 0$ are close to the boundary of the unit circle, then roots of the estimated polynomial may fall inside the unit circle which implies that the estimated model is unstable or nonstationary. However the Yule–Walker estimates are not subject to this hazard:

THEOREM 2.

The Yule–Walker estimates always correspond to a stationary autoregressive process.

Proof.

This follows directly from Theorem 1 by virtue of the biquadratic nature of the criterion function $S(\hat{\alpha}, y)$ of (18).

4. MOVING-AVERAGE ESTIMATION

The Approximation Method

Consider a moving-average process of order q which is represented by the equation

$$y(t) = \mu(L)\varepsilon(t), \tag{23}$$

where $\mu(L) = \mu_0 + \mu_1 L + \cdots + \mu_q L^q$, which has $\mu_0 = 1$, is a polynomial of degree q in the lag operator L. The process is invertible if and only if it has an alternative representation as an infinite-order autoregressive process:

$$\begin{aligned} \varepsilon(t) &= \mu^{-1}(L)y(t) \\ &= \psi(L)y(t), \end{aligned} \tag{24}$$

where $\psi(L) = \{\psi_0 + \psi_1 L + \cdots\}$ is of infinite degree with $\sum |\psi_i| < \infty$. For invertibility, it is necessary and sufficient that the roots of the equation $\mu(z) = 0$ lie outside the unit circle.

Given a sample of T mean-adjusted observations in $y = \{y_0, y_1, \ldots, y_{T-1}\}$, we can estimate the elements of $\mu = \{1, \mu_1, \ldots, \mu_q\}$ according to a principle which is similar to the one which generates the Yule–Walker estimates of an autoregressive process. The principle gives rise to what we shall describe, for reasons which will become clear shortly, as the Approximation Method.

The estimates of the Approximation Method are the values which minimise

$$\begin{aligned} S(\hat{\mu}, y) &= \oint \frac{I(z)}{\hat{\mu}(z)\hat{\mu}(z^{-1})} \frac{dz}{iz} \\ &= \oint \hat{\psi}(z) I(z) \hat{\psi}(z^{-1}) \frac{dz}{iz} \\ &= \frac{1}{T} \hat{\psi}'_\infty Y'_\infty Y_\infty \hat{\psi}_\infty. \end{aligned} \tag{25}$$

Here $\hat{\psi}_\infty$ is a vector of infinite order containing the coefficients of the expansion of $\hat{\psi}(z)$ whilst Y_∞ is an infinite-order semi-circulant matrix constructed from the elements of the finite data sequence $\{y_0, y_1 \ldots, y_{T-1}\}$. The matrix $Y'_\infty Y_\infty / T$ is a Toeplitz matrix, likewise of infinite order, whose τth subdiagonal

and supradiagonal contain repeated instances of the τth empirical autocovariance which is specified by

$$(26)\qquad \begin{aligned} c_\tau &= \frac{1}{T}\sum_{t=\tau}^{T-1} y_{t-\tau}y_t \quad \text{if} \quad |\tau| \leq T-1, \\ c_\tau &= 0 \quad \text{if} \quad |\tau| > T-1. \end{aligned}$$

The criterion function in (25) is biquadratic. Therefore we may write $S(\hat{\mu}, y)$ as

$$(27)\qquad \begin{aligned} \tfrac{1}{T}\hat{\psi}'_\infty Y'_\infty Y_\infty \hat{\psi}_\infty &= \tfrac{1}{T} y'[\hat{\Psi}'_\infty \hat{\Psi}_\infty]_{T-1} y \\ &= \tfrac{1}{T} y' \Delta(\mu) y, \end{aligned}$$

where $[\hat{\Psi}'_\infty \hat{\Psi}_\infty]_{T-1} = \Delta(\hat{\mu})$ stands for the principal minor of $\hat{\Psi}'_\infty \hat{\Psi}_\infty$ of order T. The effect of commuting the elements in equation (27) is to reduce the vectors and the matrix to a finite order.

THEOREM 3.

The estimates of the Approximation Method fulfil the condition of invertibility.

Proof.

By setting $\hat{\psi}_i = 0$ for $i > 1$ in (25), we can obtain the inequality $\min S(\hat{\mu}, y) \leq c_0$ which shows that there is a finite-valued minimum. But $S(\hat{\mu}, y)$ is bounded if and only if $\sum |\hat{\psi}_i| < \infty$. Therefore $\hat{\mu}^{-1}(z) = \hat{\psi}(z)$ converges for $|z| \leq 1$; and so the roots of $\hat{\mu}(z) = 0$ must lie outside the unit circle. Therefore the estimates fulfil the condition of invertibility.

Remark.

The matrix $[\Psi'_\infty \Psi_\infty]_{T-1} = \Delta(\mu)$ is identical to the dispersion matrix $\Delta(\mu)$ of a vector $y = \{y_0, y_1, \ldots, y_{T-1}\}$ generated by a qth-order autoregressive process specified by $\mu(L)y(t) = \varepsilon(t)$ with $V\{\varepsilon(t)\} = 1$.

When $\varepsilon(t)$ is a Gaussian process, the likelihood function for the vector $y = \{y_0, y_1, \ldots, y_{T-1}\}$ generated by the MA model under (23) is given by

$$(28)\qquad N(y) = (2\pi\sigma_\varepsilon^2)^{-T/2} |\Omega(\mu)|^{-1/2} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} y'\Omega^{-1}(\mu)y\right\},$$

where $\sigma_\varepsilon^2 \Omega(\mu)$ stands for the dispersion matrix of the vector. The generalised least-squares estimates the MA parameters—which were called the unconditional least-squares estimates by Box and Jenkins [4]—are obtained by minimising the quadratic function $y'\Omega^{-1}(\mu)y$ which is to be found in the exponent

of the likelihood function. The biquadratic function $y'[\Psi'_\infty\Psi_\infty]_{T-1}y = y'\Delta(\mu)y$ is similar to the function $y'\Omega^{-1}(\mu)y$; and $\Delta(\mu)$ may be regarded as an approximation to $\Omega^{-1}(\mu)$.

Sharman [15] has shown that the product $\Omega(\mu)\Delta(\mu)$ differs from the identity matrix I_T by a matrix consisting of zeros everywhere except in the first q rows and the last q rows. Balestra [2] has provided an expression for the matrix $\Omega^{-1}(\mu)$ for the case of an MA(1) process which has the form of $\Omega^{-1}(\mu) = \Delta(\mu) + W$ where W is a matrix whose elements decline in size as we move away from the borders. In view of these results, the differences between $\Omega^{-1}(\mu)$ and $\Delta(\mu)$ appear minor. Nevertheless, they are crucial; and the unconditional least-squares estimates often violate the condition of invertibility. In fact, the second edition of the book by Box and Jenkins [4] contains an added appendix (Appendix A7.6) which bears witness to the difficulties which were encountered in practice with the method of estimation which they had recommended.

We may note that the biquadratic criterion function under (18), from which the Yule–Walker estimates of the parameters of an AR process are obtained, can be written as $S(\alpha, y) = y'A'_{T-1}A_{T-1}y/T = y'\Omega(\alpha)y/T$. Here $\Omega(\alpha)$ is the dispersion matrix of a vector $y = \{y_0, y_1, \ldots, y_{T-1}\}$ generated by an pth-order moving-average process specified by $y(t) = \alpha(L)\varepsilon(t)$ with $V\{\varepsilon(t)\} = 1$. It can also be construed as an approximation to the matrix $\Delta^{-1}(\alpha)$. The Yule–Walker Method and the Approximation Method are, in a sense, mirror images of each other.

IMPLEMENTING THE APPROXIMATION METHOD.

In evaluating the function $S(\mu, y) = T^{-1}y'\Delta(\mu)y$, it is usually impractical to form the $T \times T$ Toeplitz matrix $\Delta = \delta_{|i-j|}$ and to store it in its entirety for the reason that it is too large. However, there are at least two ways of evaluating the function which are practical.

The first way depends upon forming the T elements $\delta_0, \ldots, \delta_{T-1}$ which are to found on successive diagonals of Δ. These elements are the autocovariances of the synthetic AR process $\mu(L)y(t) = \varepsilon(t)$. Once they have been calculated, we can proceed to evaluate the expression

$$\begin{aligned} S(\mu, y) &= T^{-1}\sum_t\sum_s y_t y_s \delta_{|t-s|} \\ &= T^{-1}\sum_\tau\sum_t y_t y_{t-\tau}\delta_\tau; \quad \tau = t - s \\ &= c_0 + 2\sum_{t=1}^{T-1} c_\tau\delta_\tau, \end{aligned} \tag{29}$$

wherein c_τ is the empirical autocovariance specified in (26). It is possible to minimise the function $S(\mu, y)$ by using a numerical algorithm which demands

nothing more than a facility for evaluating the function for arbitrary values of μ. Nevertheless, Godolphin [9] has devised a specialised method for minimising the function which makes use of some close approximations to the derivatives $d\delta_\tau/d\mu_j$. However, his iterative procedure has only a linear rate of convergence.

The second way of evaluating the function $S(\mu, y)$ depends upon forming the leading elements of the sequence $e(t) = \mu^{-1}(L)y(t)$ from $\mu = \{1, \mu_0, \ldots, \mu_q\}$ and $y = \{y_0, \ldots, y_{T-1}\}$. The sequence $e(t) = \{e_0, \ldots, e_{T-1}, \ldots\}$ is, of course, infinite; but, given that the roots of $\mu(z) = 0$ are outside the unit circle, we shall find that it converges rapidly to zero after the element e_{T-1}, which is where the influence of y ceases. Thus $TS(\mu, y)$ can be approximated to a high degree of accuracy by an extended sum of squares $\sum_{t=0}^{T+n} e_t^2$.

The function $S(\mu, y) = y'[\Psi'_\infty \Psi_\infty]_{T-1} y$ may also be compared with the function $\bar{e}'\bar{e} = y'M'^{-1}M^{-1}y$ wherein $M = M(\mu)$ is a lower triangular Toeplitz matrix of order T whose leading column is the vector $\{\mu_0, \ldots, \mu_q, 0, \ldots, 0\}$. The matrix M^{-1} is simply the the principal minor of order T of the infinite order semi-circulant matrix Ψ_∞. Therefore the product $M'^{-1}M^{-1}$, which is not exactly a Toeplitz matrix, can be viewed as an approximation to $\Delta(\mu) = [\Psi'_\infty \Psi_\infty]_{T-1}$. The value of μ which minimises the function $y'M'^{-1}M^{-1}y$ corresponds to what Box and Jenkins have described as the conditional least-squares estimates of the moving-average process. Whilst such estimates are capable of violating the condition of invertibility, Osborn's [14] analysis of the expected value of the criterion function suggests that they should do so less frequently than the unconditional least-squares estimates. This supposition seems to be confirmed in practice.

An attractive feature of the criterion function $\bar{e}'\bar{e} = y'M'^{-1}M^{-1}y$ of the unconditional least-squares estimation is the relative ease with which it may be minimised via the Gauss–Newton iterative procedure. In applying this procedure, we can can make use of the analytic form of the derivative $d\bar{e}/d\mu$ which is readily avilable.

The formal algebra associated with the unconditional least-squares estimates is not affected when the tail of the data vector $\{y_0, y_1, \ldots, y_{T-1}\}$ is padded by zeros. However, as the extent of the padding increases, the value of function $y'M'^{-1}M^{-1}y$ converges rapidly upon that of $TS(\mu, y) = y'\Delta(\mu)y$. The upshot is that the estimates of the approximation method can be obtained, in practice, from a procedure which is intended for calculating the unconditional least-squares estimates.

The asymptotic form of the criterion function in (25) is given by

$$\bar{S}(\hat{\mu}) = \oint \frac{f(z)}{\hat{\mu}(z)\hat{\mu}(z^{-1})}\frac{dz}{iz}, \tag{30}$$

where

$$f(z) = \frac{\sigma_\varepsilon^2}{2\pi}\mu(z)\mu(z^{-1}) \tag{31}$$

is the spectral density function of the MA process depicted by equation (23) when $z = e^{i\omega}$. The function $\bar{S}(\hat{\mu})$ is minimised when $\hat{\mu}(z) = \mu(z)$, in which case we have $\bar{S}(\mu) = \sigma_\varepsilon^2$.

Consider the "inverse" function

$$\bar{V}(\hat{\mu}) = \oint \frac{\hat{\mu}(z)\hat{\mu}(z^{-1})}{f(\omega)}\frac{dz}{iz}. \tag{32}$$

It can be seen that this also attains its minimum value when $\hat{\mu}(z) = \mu(z)$. This result suggests that we might estimate $\mu(z)$ by finding the value which minimises a finite-sample version of the function in the form of

$$\begin{aligned} V(\hat{\mu}, y) &= \oint \frac{\hat{\mu}(z)\hat{\mu}(z^{-1})}{\hat{f}(z)}\frac{dz}{iz} \\ &= \hat{\mu}' R_q \hat{\mu}, \end{aligned} \tag{33}$$

wherein $\hat{f}(z)$ is a consistent estimate of the spectral density function $f(z)$. Here R_q is a Toeplitz matrix whose elements are given by

$$\begin{aligned} r_\tau &= \frac{1}{(2\pi)^2}\oint \frac{z^\tau}{\hat{f}(z)}\frac{dz}{iz} \\ &= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi} \frac{e^{i\omega\tau}}{\hat{f}(e^{i\omega\tau})}d\omega. \end{aligned} \tag{34}$$

These are the empirical counterparts of what Cleveland [5] has described as the inverse covariances. The expression above can be compared with a similar expression for the empirical autocovariances given under (20). Theorem 1 serves to show the following:

THEOREM 4.

The estimates of the moving-average parameters obtained by the Inverse Method fulfil the conditions of invertibility.

Remark. It is notable that we need to use a consistent estimator $\hat{f}(z)$ of the spectral density function in order to obtain consistent estimates of the inverse autocovariances. By contrast, the sequence of ordinary empirical autocovariances, which represent consistent estimates of their population counterparts, may be obtained, according to equation (20), from the periodogram $I(z)$ which is an inconsistent estimator of the spectrum. It might be hoped that, if $\hat{f}(z)$ were replaced by $I(z)$, the function $V(\mu)$ would still converge to $\bar{V}(\mu)$, thereby preserving the consistency of the estimates of the MA parameters. However, the results of Bhansali [3] suggest that this is not the case. The consistent estimator $\hat{f}(z)$ would normally be obtained by smoothing the periodogram, and this is liable to add significantly to the burden of computation.

Durbin's Method.

It is interesting to note that the method of MA estimation proposed by Durbin [8] is, in fact, a variant of the inverse method. Consider the asymptotic form of the inverse criterion function under (32) which can be combined with the definition under (31) to give

$$\begin{aligned}\bar{V}(\hat{\mu}) &= \frac{2\pi}{\sigma_\varepsilon^2}\oint \frac{\hat{\mu}(z)\hat{\mu}(z^{-1})}{\mu(z)\mu(z^{-1})}\frac{dz}{iz} \\ &= \frac{2\pi}{\sigma_\varepsilon^2}\oint \hat{\mu}(z)\psi(z)\psi(z^{-1})\hat{\mu}(z^{-1})\frac{dz}{iz} \\ &= \left(\frac{2\pi}{\sigma_\varepsilon^2}\right)^2 \hat{\mu}'\Delta_q(\mu)\hat{\mu}.\end{aligned} \tag{35}$$

In effect, Durbin's proposal was to approximate the asymptotic form of the inverse criterion function by replacing $\psi(z) = \mu^{-1}(z)$ by a finite polynomial $\phi(z)$ obtained by fitting a autoregressive model of a high order to the data generated by the MA process. In fact, Durbin's method is the result of using the method of autoregressive spectral estimation to derive the function $\hat{f}(z)$ which is to be found within the criterion function $V(\mu, y)$ of (33).

5. AUTOREGRESSIVE MOVING-AVERAGE ESTIMATION

Consider an autoregressive moving-average process of order (p, q) which is represented by the equation

$$\alpha(L)y(t) = \mu(L)\varepsilon(t). \tag{36}$$

We assume that the model is both stationary and invertible by virtue of the condition that the roots of the equations $\alpha(z) = 0$ and $\mu(z) = 0$ lie outside the unit circle.

Given our sample of mean-adjusted observations, we might derive estimates of the autoregressive and moving-average parameters by finding the values which minimise the function

$$S(\hat{\alpha}, \hat{\mu}, y) = \frac{1}{2\pi} \oint I(z) \frac{\hat{\alpha}(z)\hat{\alpha}(z^{-1})}{\hat{\mu}(z)\hat{\mu}(z^{-1})} \frac{dz}{iz}. \tag{37}$$

Such estimates would fulfil the conditions of stationarity and invertibility. There are a variety of alternative matrix expressions which we can use for the function $S(\alpha, \mu, y)$. However, if we wish to avoid expressions involving matrices or vectors of infinite order, then we are constrained to write

$$\begin{aligned} S(\alpha, \mu, y) &= \alpha_r' Y_r' \Delta_r(\mu) Y_r \alpha_r \\ &= y_r' A_r' \Delta_r(\mu) A_r y_r. \end{aligned} \tag{38}$$

Here $Y_r\alpha_r = A_r y_r$ is the matrix expression for the convolution of $\alpha = \{1, \alpha_1, \ldots, \alpha_p\}$ and $y = \{y_0, y_1, \ldots, y_{T-1}\}$ which depends upon the semicirculant matrices $Y_r(y)$ and $A_r(\alpha)$ of order $(2T-1) \times T$, whilst $\Delta_r(\mu)$ is the variance–covariance matrix of order $r = 2T - 1$ of the process $\mu(L)y(t) = \varepsilon(t)$. These expressions for $S(\alpha, \mu, y)$ may be compared with similar expressions based on the lower triangular Toeplitz matrices $A(\alpha)$, $M(\mu)$ and $Y(y)$ of order T:

$$\bar{e}'\bar{e} = \alpha' Y' M'^{-1} M^{-1} Y \alpha = y' A' M'^{-1} M^{-1} A y. \tag{39}$$

The values of α and μ which minimises the function $\bar{e}'\bar{e}$ corresponds to what Box and Jenkins have described as the conditional least-squares estimates of the autoregressive moving-average process. To find these values, we may use a Gauss–Newton iterative procedure based on the analytic expressions for the derivatives $d\bar{e}/d\alpha$ and $d\bar{e}/d\mu$. By padding the tail of the vector y with zeros, we can obtain values for α and μ which closely approximate the ones which minimise the function $S(\alpha, \mu, y)$.

THE MIXED METHOD.

For an alternative way of obtaining estimates which are guaranteed to satisfy the conditions of stationarity and invertibility, we might resort to a Mixed Method with combines the Yule–Walker Method of estimating the autoregressive parameters with the Inverse Method of estimating the moving-average parameters.

Given a value for $\hat{\mu}$, we should obtain one for α by minimising the function

$$S(\hat{\alpha}|\hat{\mu}) = \oint \hat{\alpha}(z)\hat{\alpha}(z^{-1})\frac{\hat{f}(z)}{\hat{\mu}(z)\hat{\mu}(z^{-1})}\frac{dz}{iz} = \hat{\alpha}'G_p\hat{\alpha}, \tag{40}$$

where the elements of the Toeplitz matrix G_p are given by

$$g_\tau = \oint z^\tau \frac{\hat{f}(z)}{\hat{\mu}(z)\hat{\mu}(z^{-1})}\frac{dz}{overiz}. \tag{41}$$

Given a value for $\hat{\alpha}$, we should obtain one for $\hat{\mu}$ by minimising the function

$$L(\hat{\mu}|\hat{\alpha}) = \oint \frac{\hat{\mu}(z)\hat{\mu}(z^{-1})}{\hat{\alpha}(z)\hat{f}(z)\hat{\alpha}(z^{-1})}\frac{dz}{iz} = \hat{\mu}'H_q\hat{\mu}, \tag{42}$$

where the elements of the Toeplitz matrix H_q are given by

$$h_\tau = \frac{1}{(2\pi)^2}\oint \frac{z^\tau}{\hat{\alpha}(z)\hat{f}(z)\hat{\alpha}(z^{-1})}\frac{dz}{iz}. \tag{43}$$

These two minimisations represent successive steps in an iterative algorithm which will generate a convergent sequence of values for $\hat{\alpha}$ and $\hat{\mu}$. The recursion starts with an initial estimate of $\hat{\alpha}$ provided by the equations

$$\begin{bmatrix} c_q & c_{q-1} & \cdots & c_{q-p+1} \\ c_{q+1} & c_q & \cdots & c_{q-p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q+p-1} & c_{q+p-2} & \cdots & c_q \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{bmatrix} = \begin{bmatrix} c_{q+1} \\ c_{q+2} \\ \vdots \\ c_{q+p} \end{bmatrix}$$

Remark.

In order to implement the algorithm, we must decide precisely how the elements of G_p and H_q are to be calculated. An obvious way is to reexpress equations (41) and (43) in terms of the discrete Fourier transform and to use the fast Fourier algorithm to perform the computations. An alternative way is to borrow the method which Durbin [8] has used for estimating MA models which we have described above. We can calculate the elements of G_p by forming the covariances of the filtered process $\hat{\mu}^{-1}(L)y(t)$. We can calculate the

elements of H_q by fitting an autoregressive model of high order to the filtered process $\hat{\alpha}(L)y(t)$.

An algorithm for the Mixed Method, which has been programmed by the author, appears to converge at roughly the same speed as an unsophistocated version of the Gauss–Newton algorithm for computing the unconditional least-squares estimates. In the more sophistocated versions of the Gauss–Newton algorithm, the rate of convergence is accelerated by selecting an optimal step-length for each iteration. The alternative step lengths are evaluated in terms of the associated reductions in the value of the criterion function. In the absense of an explicit criterion function for the Mixed Method, it difficult to know how to accelerate its rate of convergence in such a manner.

It can be shown that the estimates generated by the Mixed Method are asymptotically equivalent to the maximum-likelihood estimates.

6. REFERENCES

[1] **Amemiya, T.** (1985). "Advanced Econometrics". Basil Blackwell, Cambridge.

[2] **Balestra, P.** (1980). "A note on the exact transformation associated with the first-order moving-average process". Journal of Econometrics, 14, 381–394.

[3] **Bhansali, R. J.** "Autoregressive and window estimates of the inverse autocorrelation function". Biometrika, 67, 551–556.

[4] **Box G. E. P.** and **Jenkins G.M.** (1970). "Time-Series Analysis : Forecasting and Control". Holden-Day, San Francisco.

[5] **Cleveland, W. S.** (1972). "The inverse autocorrelations of a time series and their applications". Technometrics, 14, 277–293.

[6] **Davidson, E. H.** (1981). "Problems with the estimation of moving-average processes". Journal of Econometrics, 16, 295–310.

[7] **Domowitz, I.** and **White H.** (1982). "Misspecified models with dependent observations". Journal of Econometrics, 20, 35–58.

[8] **Durbin, J.** (1959). "Efficient estimation of parameters in moving-average models". Biometrika, 46, 306–316.

[9] **Godolphin, E. J.** (1977). "A direct representation for the maximum likelihood estimator of a Gaussian moving average process". Biometrika, 64, 375–84.

[10] **Kang, K. M.** (1975). "A comparison of the least-squares and maximum-likelihood estimation for moving-average processes". Bureau of Census

and Statistics, Canberra.

[11] **Kreyszig, E.** (1988). "Advanced Engineering Mathematics : Sixth Edition". John Wiley and Sons, New York.

[12] **Laurie, D. P.** (1980). "Efficient implementation of Wilson's algorithm for factorising a self-reciprocal polynomial". BIT, 20, 257–259.

[13] **Laurie, D. P.** (1982). "Algorithm AS 175 : Cramér–Wold Factorisation". Applied Statistics , 86–93.

[14] **Osborn, D. R.** (1982). "On the criteria functions used for the estimation of moving-average processes". Journal of the American Statistical Association, 77, 388–392.

[15] **Sharman, P.** (1975). "An approximate inverse for the covariance matrix of moving-average and autoregressive processes". Annals of Statistics, 3, 532–8.

[16] **Wallis, K. E.** (1976) "Multiple time-series and the final form of Econometric models". Econometrica, 45, 1481–1497.

[17] **Wilson, G. T.** (1969). "Factorisation of the covariance generating function of a pure moving-average process". SIAM J. Statist. Comput. Simul, 8, 301–309.