The Annals of Statistics
1978, Vol. 6, No. 6, 1369–1389

ESTIMATION OF PARAMETERS IN THE ARMA MODEL WHEN THE CHARACTERISTIC POLYNOMIAL OF THE MA OPERATOR HAS A UNIT ZERO

BY TUAN PHAM-DINH

Université de Grenoble

We consider the estimation of parameters in the time series model

$$X(t) = \sum_{j=1}^{q} a_j X(t-j) + \varepsilon(t) - \varepsilon(t-j) - \sum_{j=1}^{p} c_j\{\varepsilon(t-j) - \varepsilon(t-j-1)\}$$

where the $\varepsilon(t)$ are independently identically distributed random variables with zero mean and variance σ^2. We compute the exact log likelihood of the model, propose and justify an asymptotic approximation of it. The latter will be used to derive estimates of the parameters which are shown to be asymptotically normal and efficient.

1. Introduction. The estimation of parameters in the autoregressive moving average (ARMA) time series model

$$X(t) = \sum_{j=1}^{q} a_j X(t-j) + \varepsilon(t) - \sum_{j=1}^{p} b_j \varepsilon(t-j) .$$

with $\varepsilon(t)$ being identically and independently distributed (i.i.d.) random variables of zero mean and finite variance, has been thoroughly studied by many authors (e.g., Box and Jenkins (1970); Hannan (1970)) under the assumption that the polynomials $1 - a_1 z \cdots - a_q z^q$ and $1 - b_1 z \cdots - b_p z^p$ have no zeros in and on the unit circle. We consider here the situations where the polynomial $1 - b_1 z \cdots - b_p z^p$ has a unit zero, i.e., where

$$1 - \sum_{j=1}^{p} b_j z^j = (1-z)(1 - \sum_{j=1}^{p-1} c_j z^j) .$$

Thus we are concerned with the model

$$X(t) = \sum_{j=1}^{q} a_j X(t-j) + \varepsilon(t) - \varepsilon(t-1) - \sum_{j=1}^{p-1} c_j\{\varepsilon(t-j) - \varepsilon(t-j-1)\} \tag{1.1}$$

where the $\varepsilon(t)$ are i.i.d. random variables of zero mean and variance σ^2.

This model is characterized by the fact that the variance of $X(1) + \cdots + X(n)$ remains finite as n tends to infinity. A simple example of it is the process of events displaced by random deviations (Cox and Lewis (1966), page 204). The time interval between events is given by

$$T_i = a + \varepsilon_i - \varepsilon_{i-1} .$$

Our approach to derive estimates of the parameters of the model (1.1) is the following:

Received December 1975; revised October 1976.

AMS 1970 *subject classification.* 62F20.

Key words and phrases. Asymptotic distribution, autoregressive-moving-average model, consistency, cumulants, innovation, likelihood function, L^p norm.

(i) We compute the exact log-likelihood function of the model by evaluating the innovation

$$\zeta_\theta(t) = X(t) - E_\theta\{X(t) \mid X(1), \cdots, X(t-1)\},$$

θ denoting $(a_1, \cdots, a_q, c_1, \cdots, c_{p-1}, \sigma^2)$ so that under the assumption that the observed series is Gaussian, the log-likelihood function of the model for a sample of length N is:

$$\mathscr{L}_N(\theta) = \tfrac{1}{2} \sum_{t=1}^N \log\{2\pi\sigma_\theta^2(t)\} - \tfrac{1}{2} \sum_{t=1}^N \zeta_\theta^2(t)/\sigma_\theta^2(t)$$

where $\sigma_\theta^2(t)$ is the variance of $\zeta_\theta(t)$.

(ii) We derive an asymptotic approximation L_N to $\mathscr{L}_N$, which satisfies the condition:

$$\text{(A)} \qquad \sup_{\theta \in E} \|\mathscr{L}_N^{(i)}(\theta) - L_N^{(i)}(\theta)\| \leqq (\log N) Z_N, \qquad i = 0, 1, 2$$

where $\mathscr{L}_N^{(i)}$ and $L_N^{(i)}$, $i = 0, 1, 2$ denote respectively $\mathscr{L}_N$ and L_N, the vector of first derivatives and the matrix of second derivatives of $\mathscr{L}_N$ and of L_N, $\|\cdot\|$ denotes any vector (matrix) norm, E is any compact set and $\{Z_N\}$ is some sequence of random variables bounded in the L^1 norm.

(iii) We show that the L_N function possesses all classical properties of a log-likelihood function. By (A) the same is true for $\mathscr{L}_N$. Thus there exists a consistent maximum likelihood estimate which is asymptotically normal and efficient, and a consistent approximate maximum likelihood estimate (i.e., the one based on the maximization of L_N) with the same properties. Using (A), we can even show that the latter differs from the former by a term of order $(\log N)/N$. (See Theorem 3.)

It should be pointed out that in the special case where $p = 1$, $b_1 = 1$, i.e., when the considered model is

$$\text{(1.2)} \qquad X(t) = \sum_{j=1}^q a_j X(t-j) + \varepsilon(t) - \varepsilon(t-1),$$

then specific results can be obtained by straightforward arguments. We shall deal with these in some detail.

2. Computation of the innovation. In this paragraph, we shall omit the subscript θ and denote by $Z(t \mid s)$ the orthogonal projection (in the sense of the L^2 scalar product) of the random variable $Z(t)$ onto the linear space spanned by $X(1), \cdots, X(s)$.

We put

$$\phi(t) = \varepsilon(t) - \varepsilon(t-1)$$
$$V(t) = {}^T(V_1(t), \cdots, V_p(t)) = {}^T(\phi(t), \cdots, \phi(t+2-p), \varepsilon(t+1-p)).$$

Then the model equation (1.1) becomes

$$\text{(2.1)} \qquad X(t) = \sum_{j=1}^q a_j X(t-j) + \varepsilon(t) - {}^T b V(t-1)$$

where $b = {}^T(1 + c_1, \cdots, 1 + c_{p-1}, 1)$, since $\varepsilon(t) = V_1(t) + \cdots + V_p(t)$.

Thus, for $t > q$,

$$X(t \mid t-1) = \sum_{j=1}^{q} a_j X(t-j) - {}^T b V(t-1 \mid t-1)$$

so that $\zeta(t) = X(t) - X(t \mid t-1)$, $t > q$ is given by

$$\zeta(t) = X(t) - \sum_{j=1}^{q} a_j X(t-j) + {}^T b V(t-1 \mid t-1) \tag{2.2}$$

$$\zeta(t) = \varepsilon(t) - {}^T b\{V(t-1) - V(t-1 \mid t-1)\} . \tag{2.3}$$

On the other hand, using (1.1), we get

$$V(t) = BV(t-1) + W(t)$$

where

$$B = \begin{pmatrix} c_1 & c_2 \cdots c_{p-1} & 0 \\ 1 & 0 \cdots \;\; 0 & 0 \\ 0 & 1 \cdots \;\; \cdot & \cdot \\ \cdot & \cdot \cdots \;\; \cdot & \cdot \\ 0 & 0 \cdots \;\; 1 & 1 \end{pmatrix} ; \qquad W(t) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} [X(t) - \sum_{j=1}^{q} a_j X(t-j)] .$$

Thus $V(t \mid t) = BV(t-1 \mid t) + W(t)$ and since $V(t-1 \mid t) = V(t-1 \mid t-1) + K(t)\zeta(t)$ where $K(t)\zeta(t)$ denotes the orthogonal projection of $V(t-1)$ onto the linear space spanned by $\zeta(t)$, we have the recurrence equation

$$V(t \mid t) = BV(t-1 \mid t-1) + BK(t)\zeta(t) + W(t) .$$

By (2.2), $\zeta(t) = {}^T e W(t) + {}^T b V(t-1 \mid t-1)$ where $e = {}^T(1, \cdots, 1)$ so that

$$V(t \mid t) = B\{I + K(t)\,{}^T b\}V(t-1 \mid t-1) + \{I + BK(t)\,{}^T e\}W(t) , \quad t > q . \tag{2.4}$$

To obtain $K(t)$, we introduce the covariance matrix $Q(t)$ of $V(t) - V(t \mid t)$. By (2.3),

$$K(t) = E\{V(t-1)\zeta(t)/\sigma^2\{\zeta(t)\} = -Q(t-1)b/\sigma^2\{\zeta(t)\}$$
$$\sigma^2\{\zeta(t)\} = \sigma^2 + {}^T b Q(t-1) b = \sigma^2(t) , \quad \text{say.}$$

Now, remarking that

$$V(t) - V(t \mid t) = B\{V(t-1) - V(t-1 \mid t-1) - K(t)\zeta(t)\}$$

and that $K(t)\zeta(t)$ is also the orthogonal projection of $V(t-1) - V(t-1 \mid t-1)$ onto the linear space spanned by $\zeta(t)$, we get

$$Q(t) = B\{Q(t-1) - K(t)\sigma^2(t)\,{}^T K(t)\}\,{}^T B , \quad t > q .$$

Using the above expressions for $K(t)$ and $\sigma^2(t)$, we obtain:

$$Q(t) = B\{Q(t-1) + K(t)\,{}^T b Q(t-1)\}\,{}^T B$$
$$Q(t)\,{}^T B^{-1} b = B\{Q(t-1)b + K(t)\,{}^T b Q(t-1) b\}$$
$$= BK(t)\{-\sigma^2(t) + {}^T b Q(t-1) b\} = -BK(t)\sigma^2 ,$$

$$K(t) = -B^{-1}Q(t)\,{}^T B^{-1} b/\sigma^2 , \quad t > q . \tag{2.5}$$

Suppose that $Q(t-1)$ is invertible; then we also have

$$Q(t)\,{}^T B^{-1} Q^{-1}(t-1) = B\{I + K(t)\,{}^T b\} , \quad t > q , \tag{2.6}$$

so that using the expression (2.5) for $K(t)$,

$$Q(t)\,{}^T B^{-1}\{Q^{-1}(t-1) + b\,{}^T b/\sigma^2\} = B\,, \quad t > q\,.$$

We see that $Q(t)$ is also invertible and

$$Q^{-1}(t) = {}^T B^{-1}\{Q^{-1}(t-1) + b\,{}^T b/\sigma^2\}B^{-1}\,, \quad t > q\,.$$

To solve the above recurrence equation, putting $\mu(t) = \sigma^2\,{}^T B^t Q^{-1}(t) B^t$, we obtain

$$\begin{aligned} &\mu(t) = \mu(t-1) + {}^T B^{t-1} b\,{}^T b B^{t-1}\,, \quad t > q\,, \\ &\mu(t) = \mu(q) + \textstyle\sum_{j=q}^{t-1} {}^T B^j b\,{}^T b B^j\,, \quad t > q\,. \end{aligned} \tag{2.7}$$

Let us return to the recurrence equation (2.4) We remark that ${}^T eB = {}^T b$ so that $I + BK(t)\,{}^T e = B\{I + K(t)\,{}^T b\}B^{-1}$ and thus by (2.6), the equation (2.4) becomes

$$\begin{aligned} V(t\,|\,t) &= Q(t)\,{}^T B^{-1} Q^{-1}(t-1)\{V(t-1\,|\,t-1) + B^{-1}W(t)\}\,, \quad t > q \\ &= B^t \mu^{-1}(t)\mu(t-1)B^{1-t}\{V(t-1\,|\,t-1) + B^{-1}W(t)\}\,, \quad t > q\,. \end{aligned}$$

The solution of this recurrence equation is:

$$V(t\,|\,t) = B^t \mu^{-1}(t)\{\textstyle\sum_{j=q+1}^{t} \mu(j-1)B^{-j}W(j) + \mu(q)B^{-q}V(q\,|\,q)\}$$

for $t \geqq q$. Using (2.2), we get:

$$\begin{aligned} \zeta(t) &= {}^T e\{W(t) + BV(t-1\,|\,t-1)\} \\ &= {}^T e B^t \mu^{-1}(t-1)\{\textstyle\sum_{j=q+1}^{t} \mu(j-1)B^{-j}W(j) + \mu(q)B^{-q}V(q\,|\,q)\} \end{aligned} \tag{2.8}$$

for $t > q$. On the other hand, the variance $\sigma^2(t)$ of $\zeta(t)$ has been shown to be

$$\sigma^2(t) = \sigma^2\{1 + {}^T e B^t \mu^{-1}(t-1)\,{}^T B^t e\}\,. \tag{2.9}$$

One can compute $\zeta(t)$ by (2.7) and (2.8). However, it is preferable to rearrange (2.8) as follows: we have, by (2.7) and the fact that ${}^T eB = {}^T b$,

$$\mu(j-1) = \mu(t-1) - \textstyle\sum_{k=j}^{t-1} {}^T B^k e\,{}^T e B^k\,.$$

Thus

$$\begin{aligned} \zeta(t) = &\textstyle\sum_{j=q+1}^{t} {}^T e B^{t-j} W(j) + {}^T e B^t \mu^{-1}(t-1)\mu(q)B^{-q}V(q\,|\,q) \\ &- {}^T e B^t \mu^{-1}(t-1) \textstyle\sum_{j=q+1}^{t} \{\sum_{k=j}^{t-1} {}^T B^k e\,{}^T e B^k\}B^{-j}W(j)\,, \quad t > q\,. \end{aligned}$$

The last term of the above right-hand side is equal to

$$-{}^T e B^t \mu^{-1}(t-1) \textstyle\sum_{k=q+1}^{t-1} {}^T B^k e\,{}^T e\{\sum_{j=q+1}^{k} B^{k-j}W(j)\}\,,$$

so that if we put

$$\hat{\varepsilon}(k) = \textstyle\sum_{j=q+1}^{k} {}^T e B^{k-j} W(j)\,, \quad k > q\,, \tag{2.10}$$

then for $t > q$:

$$\zeta(t) = \hat{\varepsilon}(t) + {}^T e B^t \mu^{-1}(t-1)\{\mu(q)B^{-q}V(q\,|\,q) - \textstyle\sum_{k=q+1}^{t-1} {}^T B^k e \hat{\varepsilon}(k)\}\,. \tag{2.11}$$

The term $\hat{\varepsilon}(t)$ has a simple interpretation. Indeed let $\hat{\phi}(t)$, $t \in \mathbb{Z}$ be the solution of the recurrence equation:

$$\hat{\phi}(t) - \sum_{j=1}^{p-1} c_j \hat{\phi}(t-j) = X(t) - \sum_{j=1}^{q} a_j X(t-j), \quad t > q, \tag{2.12}$$

with the initial conditions $\hat{\phi}(t) = 0$, $t \leqq q$; then it can be easily seen that the vector $\hat{V}(t)$ with components $\hat{V}_j(t) = \hat{\phi}(t+1-j)$, $j - 1, \cdots, p-1$ and

$$\hat{V}_p(t) = \sum_{j=p}^{\infty} \hat{\phi}(t+1-j)$$

is equal to $W(t) + BW(t-1) + \cdots + B^{t-q-1}W(q+1)$. Thus

$$\hat{\varepsilon}(t) = {}^{T}e\hat{V}(t) = \sum_{j=-\infty}^{t} \hat{\phi}(j);$$

and it follows that $\hat{\varepsilon}(t)$, $t \in \mathbb{Z}$ is the solution of the recurrence equation

$$\begin{aligned} \hat{\varepsilon}(t) - \hat{\varepsilon}(t-1) - \sum_{j=1}^{p-1} c_j\{\hat{\varepsilon}(t-j) - \hat{\varepsilon}(t-j-1)\} \\ = X(t) - \sum_{j=1}^{q} a_j X(t-j), \quad t > q, \end{aligned} \tag{2.13}$$

with initial conditions $\hat{\varepsilon}(t) = 0$, $t \leqq q$.

REMARK. In the special case when $p = 1$, the expressions for $\zeta(t)$ and $\sigma^2(t)$ can be obtained by the following straightforward argument.

By the model equation (1.2), we have

$$\zeta(t) = X(t) - \sum_{j=1}^{q} a_j X(t-j) + \varepsilon(t-1 \mid t-1), \quad t > q, \tag{2.14}$$

which is a particular case of (2.2) by taking $b = 1$ and $V(t) = \varepsilon(t)$.

Now $\varepsilon(t \mid t)$ is proportional to $\zeta(t)$ since it belongs to the subspace spanned by $X(1), \cdots, X(t)$ and is orthogonal to $X(1), \cdots, X(t-1)$. Thus

$$\varepsilon(t \mid t) = [E\{\varepsilon(t)\zeta(t)\}/\sigma^2(t)]\zeta(t).$$

By (1.2) and (2.14) we have

$$\zeta(t) = \varepsilon(t) + \{\varepsilon(t-1 \mid t-1) - \varepsilon(t-1)\}, \quad t > q,$$

so that if we denote by $\lambda(t)$ the variance of $\varepsilon(t) - \varepsilon(t-1)$:

$$\begin{aligned} E\{\varepsilon(t)\zeta(t)\} &= \sigma^2, \quad t > q, \\ \sigma^2(t) &= \sigma^2 + \lambda(t-1), \quad t > q. \end{aligned}$$

We thus obtain the recurrence equation

$$\zeta(t) = X(t) - \sum_{j=1}^{q} a_j X(t-j) + \frac{\sigma^2}{\sigma^2 + \lambda(t-2)} \zeta(t-1).$$

Now

$$\begin{aligned} \lambda(t) &= \sigma^2\{\varepsilon(t) - \varepsilon(t \mid t)\} = \sigma^2 - \sigma^2\{\varepsilon(t \mid t)\} \\ &= \sigma^2 - \frac{\sigma^4}{\sigma^2 + \lambda(t-1)} = \frac{\sigma^2 \lambda(t-1)}{\sigma^2 + \lambda(t-1)}, \quad t > q. \end{aligned}$$

This gives the recurrence equation

$$\frac{1}{\lambda(t)} = \frac{1}{\sigma^2} + \frac{1}{\lambda(t-1)}, \quad t > q,$$

which has the solution

$$\lambda^{-1}(t) = (t-q)/\sigma^2 + \lambda^{-1}(q), \quad t > q,$$

so that

$$\frac{\sigma^2}{\sigma^2 + \lambda(t-2)} = \frac{\sigma^2\lambda^{-1}(t-2)}{1 + \sigma^2\lambda^{-1}(t-2)} = \frac{\sigma^2\lambda^{-1}(q) + t - 2 - q}{\sigma^2\lambda^{-1}(q) + t - 1 - q}.$$

Putting $\beta = \sigma^2\lambda^{-1}(q) - q - 1$, we finally get

$$\zeta(t) = X(t) - \sum_{j=1}^{q} a_j X(t-j) + \frac{\beta + t - 1}{\beta + t}\zeta(t-1), \quad t > q+1.$$

This recurrence equation has the solution (taking into account (2.14))

$$\zeta(t) = \sum_{k=q+1}^{t} \frac{\beta + k}{\beta + t}\{X(k) - \sum_{j=1}^{q} a_j X(k-j)\} + \frac{\beta + q + 1}{\beta + t}\varepsilon(q \mid q). \tag{2.15}$$

On the other hand

$$\begin{aligned}\sigma^2(t) &= \sigma^2 + \lambda(t-1) = \frac{\sigma^2\lambda^{-1}(t-1) + 1}{\lambda^{-1}(t-1)} \\ &= \frac{\beta + t + 1}{\beta + 1}\sigma^2, \quad t > q.\end{aligned} \tag{2.16}$$

3. Asymptotic approximation of the log-likelihood function. We shall denote by Θ the open set of $\mathbb{R}^{p+q}$ of values of $\theta = (a_1, \cdots, a_q, c_1, \cdots, c_{p-1}, \sigma^2)$ such that $\sigma^2 > 0$ and the polynomial $1 - a_1 z \cdots - a_q z^q$ and $1 - c_1 z \cdots - c_{p-1} z^{p-1}$ have no zeros in the closed unit disc of the complex plane and have no common zeros. This implies that the eigenvalues of:

$$C = \begin{pmatrix} c_1 & c_2 & \cdots & c_{p-2} & c_{p-1} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

are of modulus strictly inferior to 1.

Since the $\zeta(t)$, B, $V(q \mid q)$, $\mu(t)$, $\hat{\varepsilon}(t)$ and $\hat{\phi}(t)$ introduced in Section 2 depend on θ, we shall use (when necessary) the subscript θ to indicate this dependence.

We now search for an asymptotic approximation of the log-likelihood function L_N. For a sample of length $N > q$, we have

$$\begin{aligned}L_N(\theta) = &-\tfrac{1}{2}\sum_{t=q+1}^{N} \log\{2\pi\sigma_\theta^2(t)\} - \tfrac{1}{2}\sum_{t=q+1}^{N} \zeta_\theta^2(t)/\sigma_\theta^2(t) \\ &- \tfrac{1}{2}\log\det\{2\pi\Gamma_q(\theta)\} - \tfrac{1}{2}\,{}^T X_q \Gamma_q^{-1}(\theta) X_q\end{aligned} \tag{3.1}$$

where $\Gamma_q(\theta)$ is the covariance matrix of the random vector $X_q = {}^T(X(1) \cdots X(q))$ and where $\zeta_\theta(t)$ and $\sigma_\theta^2(t)$ are given by (2.8) to (2.11). By the formulas (2.8), (2.10), and (2.11), it is clear that in order to obtain an asymptotic approximation of $\mathscr{L}_N$, we should study the asymptotic behavior of ${}^T e B^t \mu^{-1}(t-1)\,{}^T B^t e$, ${}^T e B^t \mu^{-1}(t-1)$, $\hat{\varepsilon}(t)$ and

$$\sum_{k=q+1}^{t-1} {}^T e B^t \mu^{-1}(t-1)\,{}^T b^k e \hat{\varepsilon}(k).$$

The results shall be described in Lemmas 1, 2 and 3. Since the proofs of these lemmas are lengthy and technical, they are relegated to the end of the section and only the statements of the lemmas are given here.

We find it convenient to introduce the following:

DEFINITION. Let $U_N(\cdot)$ be a sequence of random functions on an open set Θ of $\mathbb{R}^k$, every sample function of which is twice continuously differentiable and $U_N^{(i)}(\cdot)$, $i = 0, 1, 2$, denote the vector whose components are the ith derivatives of $U_N(\cdot)$; we say that the sequence U_N is $O(\phi(N))$ in the L^p norm if for any compact subset E of Θ, there exists a sequence of random variables Z_N, bounded in the L^p norm such that

$$\sup_{\theta\in E} \|U_N^{(i)}(\theta)\| \leqq \phi(N)Z_N\,, \quad N \geqq 1\,, \qquad i = 0, 1, 2\,.$$

In particular, a sequence of (nonrandom) functions $U_N(\cdot)$ is said to be $O(\phi(N))$ if for every compact subset E of Θ, the sequences

$$\{\sup_{\theta\in E} \|U_N^{(i)}(\theta)\|\}/\phi(N)\,, \quad N \geqq 1 \qquad i = 0, 1, 2$$

are bounded.

We shall partition the matrix B as

$$B = \begin{pmatrix} C & 0 \\ {}^T u_{p-1} & 1 \end{pmatrix}$$

where $u_{p-1} = {}^T(0, \cdots, 1)$ and C is defined in the beginning of this section. It is easily seen that the matrix B can be block diagonalized as follows:

$$B = \begin{pmatrix} I & 0 \\ -{}^T\alpha & 1 \end{pmatrix} \begin{pmatrix} C & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I & 0 \\ {}^T\alpha & 1 \end{pmatrix}$$

where $\alpha = {}^T(I - C)^{-1}u_{p-1}$. Clearly the first and last matrix of the above right-hand side are inverse one to the other. Using this result, we get

LEMMA 1. *We have*

$${}^TeB^t\mu^{-1}(t-1) = (\beta(t), \gamma(t))\begin{pmatrix} I & -\alpha \\ 0 & 1 \end{pmatrix}, \quad t > q$$

where the sequences of functions $\beta(t)$, $t > q$ and $\gamma(t)$, $t > q$ are $O(t^{-1})$ and the sequence of functions $\gamma(t) - t^{-1}$, $t > q$ is $O(t^{-2})$. On the other hand, the sequence of functions

$${}^TeB^t\mu^{-1}(t-1)^TB^te, \quad t > q$$

is $O(t^{-1})$.

LEMMA 2. *Let $\hat{\varepsilon}_\theta(t)$ be given by* (2.13). *Then the sequence of random functions $\theta \to \hat{\varepsilon}_\theta(t)$, $t > q$ is $O(1)$ in the L^2 norm.*

LEMMA 3. *The sequence of random functions*

$$\theta \to \sum_{k=q+1}^{t-1} \{{}^TeB^t\mu^{-1}(t-1)\,{}^TB^ke - t^{-1}\}\hat{\varepsilon}(k)\,, \quad t > q$$

is $O(t^{-1})$ in the L^2 norm.

The above results suggest the approximations to $\sigma_\theta^2(t)$ and $\zeta_\theta(t)$ by σ^2 and

$$\hat{\zeta}_\theta(t) = \hat{\varepsilon}_\theta(t) - \frac{1}{t} \sum_{k=q+1}^{t=1} \hat{\varepsilon}_\theta(k) . \tag{3.2}$$

Thus we propose as an approximation to $\mathscr{L}_N$:

$$L_N(\theta) = -\frac{N-q}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=q+1}^N \hat{\zeta}_\theta^2(t) . \tag{3.3}$$

THEOREM 1. *Let $\mathscr{L}_N$ and L_N be given by* (3.1) *and* (3.3). *Then the sequence of random functions $\mathscr{L}_N(\cdot) - L_N(\cdot)$, $N > q$ is $O(\log N)$ in the L^2 norm.*

PROOF. The difference $\mathscr{L}_N(\cdot) - L_N(\cdot)$ is

$$-\tfrac{1}{2} \log \det \{2\pi\Gamma_q(\theta)\} - \tfrac{1}{2}\,{}^TX_q\Gamma_q^{-1}(\theta)X_q - \tfrac{1}{2} \sum_{t=q+1}^N [\log\{\sigma_\theta^2(t)/\sigma^2\} + \eta_\theta(t)]$$

where

$$\eta_\theta(t) = \zeta_\theta^2(t)/\sigma_\theta^2(t) - \hat{\zeta}_\theta^2(t)/\sigma^2 .$$

Now, the function $\Gamma_q(\cdot)$ being twice differentiable, it is clear that the sequence

$$\delta_N(\cdot) = \tfrac{1}{2} \log \det \{2\pi\Gamma_q(\cdot)\} + \tfrac{1}{2}\,{}^TX_q\Gamma_q^{-1}(\cdot)X_q , \quad N > q$$

is $O(1)$ in the L^1 norm. On the other hand, using (2.9) and the last result of Lemma 1, it is not difficult to show that the sequence of functions $\theta \to \log\{\sigma_\theta^2(t)/\sigma^2\}$, $t > q$, $\theta \to \sigma_\theta^{-2}(t)$, $t > q$ and $\theta \to \sigma_\theta^{-2}(t) - \sigma^{-2}$, $t > q$ are respectively $O(t^{-1})$, $O(1)$ and $O(t^{-1})$. Also the results of Lemmas 1 and 3 and (2.11) show that the sequence of random functions

$$\theta \to \delta\zeta_\theta(t) = \zeta_\theta(t) - \hat{\zeta}_\theta(t) , \quad t > q$$

is $O(t^{-1})$ in the L^2 norm and those of Lemma 2 and (3.2) show that the sequence of random functions $\theta \to \hat{\zeta}_\theta(t)$, $t > q$ is $O(1)$ in the L^2 norm. Now

$$\eta_\theta(t) = [\{\delta\zeta_\theta(t)\}^2 + 2\delta\zeta_\theta(t)\hat{\zeta}_\theta(t)]\sigma_\theta^{-2}(t) + \hat{\zeta}_\theta^2(t)\{\sigma_\theta^{-2}(t) - \sigma^{-2}\}$$

so that, using the above results, one can easily see that the sequence of functions $\theta \to \eta_\theta(t)$, $t > q$ is $O(t^{-1})$ in the L^1 norm (this comes from the fact that the derivatives of the function $\theta \to \eta_\theta(t)$ are sums of products of the derivatives of $\theta \to \delta\zeta_\theta(t)$ or of $\theta \to \sigma_\theta^{-2}(t) - \sigma^{-2}$ with those of $\theta \to \zeta_\theta(t)$ or $\theta \to \sigma_\theta^{-2}(t)$ or $\theta \to \delta\zeta_\theta(t)$).

Thus the derivatives up to second order with respect to θ of

$$\delta_N'(\theta) = \sum_{t=q+1}^N [\log\{\sigma_\theta^2(t)/\sigma^2 + \eta_\theta(t)]$$

can be bounded in absolute value on any compact E of Θ by $Z_{q+1}/(q+1) + \cdots + Z_N/N$ where Z_t, $t > q$ is some sequence of random variables bounded in the L^1 norm. This gives the theorem.

We now turn to the proofs of Lemmas 1, 2 and 3.

PROOF OF LEMMA 1. Using the diagonalized form of B, we get from (2.7)

$$\mu(t-1) = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \tilde{\mu}(t-1) \begin{pmatrix} I & 0 \\ {}^T\alpha & 1 \end{pmatrix}$$

where

$$\tilde{\mu}(t-1) = \begin{pmatrix} I & -\alpha \\ 0 & 1 \end{pmatrix} \mu(q) \begin{pmatrix} I & 0 \\ -{}^T\alpha & 1 \end{pmatrix} + \sum_{k=q+1}^{t-1} \begin{pmatrix} {}^TC^k & 0 \\ 0 & 1 \end{pmatrix} \Lambda \begin{pmatrix} C^k & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} I & \alpha \\ 0 & 1 \end{pmatrix} e\,{}^Te \begin{pmatrix} I & 0 \\ -{}^T\alpha & 1 \end{pmatrix}.$$

Let ρ_E denote the least upper bound of the moduli of the eigenvalues of C_θ for $\theta \in E$; then $\|C_\theta^k\| \leqq K\rho_E^k$, $\theta \in E$ where K is some constant. Since $\rho_E < 1$, it follows that if we put

$$\tilde{\mu}(t-1) = \begin{pmatrix} \tilde{\mu}_1(t-1) & \tilde{\mu}_{12}(t-1) \\ \tilde{\mu}_{21}(t-1) & \tilde{\mu}_2(t-1) \end{pmatrix}$$

then the sequences $\tilde{\mu}_1(t-1)$, $t > q$, $\tilde{\mu}_{12}(t-1)$, $t > q$ and $\tilde{\mu}_{21}(t-1)$, $t > q$ are $O(1)$. On the other hand, $\tilde{\mu}_2(t-1) = t - q - 1 + \mu_2(q)$, $\mu_2(q)$ being the element on the last row and the last column of $\mu(q)$. Now let

$$\tilde{\mu}^{-1}(t-1) = \begin{pmatrix} \{\tilde{\mu}^{-1}(t-1)\}_1 & \{\tilde{\mu}^{-1}(t-1)\}_{12} \\ \{\tilde{\mu}^{-1}(t-1)\}_{21} & \{\tilde{\mu}^{-1}(t-1)\}_2 \end{pmatrix}.$$

Then it can be verified that (the variable $t-1$ being omitted)

$$\{\tilde{\mu}^{-1}\}_1 = \{\tilde{\mu}_1 - \tilde{\mu}_{12}\tilde{\mu}_2^{-1}\tilde{\mu}_{21}\}^{-1}$$
$$\{\tilde{\mu}^{-1}\}_{12} = {}^T\{\tilde{\mu}^{-1}\}_{21} = -\{\tilde{\mu}^{-1}\}_1\tilde{\mu}_{12}\tilde{\mu}_2^{-1}$$
$$\{\tilde{\mu}^{-1}\}_2 = \tilde{\mu}_2^{-1} + \tilde{\mu}_2^{-1}\tilde{\mu}_{21}\{\tilde{\mu}^{-1}\}_1\tilde{\mu}_{12}\tilde{\mu}_2^{-1}.$$

Thus the sequence $\{\tilde{\mu}^{-1}(t-1)\}_1$, $t > q$ is $O(1)$, the sequences $\{\tilde{\mu}^{-1}(t-1)\}_{12}$, $t > q$ and $\{\tilde{\mu}^{-1}(t-1)\}_{21}$, $t > q$ are $O(t^{-1})$ and the sequence $\{\tilde{\mu}^{-1}(t-1)\}_2 - t^{-1}$, $t > q$ is $O(t^{-2})$.

Now, using the diagonalized form of B

$${}^TeB^t\mu^{-1}(t-1) = {}^Te \begin{pmatrix} I & 0 \\ -{}^T\alpha & 1 \end{pmatrix} \begin{pmatrix} C^t & 0 \\ 0 & 1 \end{pmatrix} \tilde{\mu}^{-1}(t-1) \begin{pmatrix} I & -\alpha \\ 0 & 1 \end{pmatrix}.$$

Thus ${}^TeB^t\mu^{-1}(t-1)$ is of the form indicated in Lemma 1 with

$$\beta(t) = {}^Te \begin{pmatrix} C^t \\ -{}^T\alpha C^t \end{pmatrix} \{\tilde{\mu}^{-1}(t-1)\}_1 + \{\tilde{\mu}^{-1}(t-1)\}_{21}$$

$$\gamma(t) = {}^Te \begin{pmatrix} C^t \\ -{}^T\alpha C^t \end{pmatrix} \{\tilde{\mu}^{-1}(t-1)\}_{12} + \{\tilde{\mu}^{-1}(t-1)\}_2.$$

Using the fact that $\|C^t\| \leqq K\rho_E^t$, $\theta \in E$, with $\rho_E < 1$ (E being a compact subset of Θ and K being some constant), one can deduce that the sequences $\beta(t)$, $t > q$ and $\gamma(t)$, $t > q$ are $O(t^{-1})$ and the sequence $\gamma(t) - t^{-1}$, $t > q$ is $O(t^{-2})$.

Finally, we have

$$ {}^{T}eB^{t}\mu^{-1}(t-1)\,{}^{T}B^{t}e = (\beta(t), \gamma(t))\begin{pmatrix} {}^{T}C^{t} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} I & -\alpha \\ 0 & 1 \end{pmatrix} e $$

and thus the sequence ${}^{T}eB^{t}\mu^{-1}(t-1)\,{}^{T}B^{t}e$, $t > q$ is easily seen to be $O(t^{-1})$.

To show Lemma 2, we need the following result which is also useful later:

LEMMA 4. *Let $Y(t)$, $t \in \mathbb{Z}$, be the process*

$$ Y(t) = \sum_{j=1}^{q} a_j^{*} Y(t-j) + \varepsilon(t) - \sum_{j=1}^{p-1} c_j^{*}\varepsilon(t-j) $$

where a_j^{*}, c_j^{*} are the true values of a_j, c_j; then $X(t) = Y(t) - Y(t-1)$.

PROOF. Let ω_j, $j \geqq 0$ be the coefficients of the Taylor development of $(1 - c_1^{*}z \cdots - c_{p-1}^{*}z^{p-1})(1 - a_1^{*}z \cdots - a_q^{*}z^{q})^{-1}$; then

$$ X(t) = \sum_{j=0}^{\infty} \omega_j\{\varepsilon(t-j) - \varepsilon(t-j-1)\} $$
$$ Y(t) = \sum_{j=0}^{\infty} \omega_j \varepsilon(t-j)\,, $$

for which the result follows:

PROOF OF LEMMA 2. We have seen that

$$ \hat{\varepsilon}_\theta(t) = \sum_{k=-\infty}^{t} \hat{\phi}_\theta(k) $$

where $\hat{\phi}(t)$, $t > q$ is the solution of the recurrence equation (2.12). Thus putting $V'(t) = {}^{T}(\hat{\phi}(t), \cdots, \hat{\phi}(t+2-p))$, we see that $V'(t)$ satisfies

$$ V'(t) = CV'(t-1) + u_1\{X(t) - \sum_{j=1}^{q} a_j X(t-j)\}\,, \quad t > q\,; \qquad V'(q) = 0 $$

where $u_1 = {}^{T}(1, 0, \cdots, 0)$. This gives

$$ V'(t) = \sum_{k=0}^{t-q-1} C_\theta^{k} u_1\{X(t-k) - \sum_{j=1}^{q} a_j X(t-k-j)\}\,, \quad t > q\,. $$

Therefore $\hat{\varepsilon}_\theta(t)$ is equal to

$$ {}^{T}u_1 \sum_{k=q+1}^{t} [\sum_{l=0}^{k-q-1} C_\theta^{l} u_1\{X(k-l) - \sum_{j=1}^{q} a_j X(k-l-j)\}] $$
$$ = {}^{T}u_1 \sum_{l=0}^{t-q-1} C_\theta^{l} u_1[\sum_{k=q+1}^{t-l}\{X(k) - \sum_{j=1}^{q} a_j X(k-j)\}]\,. $$

Using Lemma 4, we get

$$ \hat{\varepsilon}_\theta(t) = {}^{T}u_1 \sum_{l=0}^{t-q-1} C_\theta^{l} u_1\{Y(t-l) - Y(q) + \sum_{j=1}^{q} a_j\{Y(t-l-j) - Y(q-j)\}\} $$
$$ = \sum_{j=0}^{q} [\sum_{l=0}^{t-q-1} g_{jl}(\theta)\{Y(t-l-j) - Y(q-j)\}]\,, \quad \text{say}. $$

Now since $\|C_\theta^{l}\| \leqq K\rho_E$, $\theta \in E$ with $\rho_E < 1$, E being a compact subset of Θ and K being some constant (see the proof of Lemma 1), we have

$$ \sup_{\theta \in E} \|g_{jl}(\theta)\| \leqq K'\rho_E^{\,l}\,, \quad K' = \text{constant}, $$

and the same result for the derivatives of g_{jl} up to second order. Since $\sum_0^\infty \rho_E^{\,l} < +\infty$ and the L^2 norms of $Y(t-l-j) - Y(q-j)$, $0 \leqq l < t-q$, $t > q$ are bounded, the sequence $\theta \to \hat{\varepsilon}_\theta(t)$ can be seen to be $O(1)$ in the L^2 norm.

PROOF OF LEMMA 3. By Lemma 1 and the diagonalized form of B:

$$\begin{aligned} {}^T eB^t\mu^{-1}(t-1)\,{}^T B^k e &= (\beta(t), \gamma(t)) \begin{pmatrix} {}^T C^k & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I & -\alpha \\ 0 & 1 \end{pmatrix} e \\ &= (\beta(t)\,{}^T C^k, \{-\beta(t)\,{}^T C^k\alpha + \gamma(t)\})e \\ &= \gamma(t) + \delta(k, t)\,, \quad \text{say.} \end{aligned}$$

Thus the sequence considered in the lemma is

$$\textstyle\sum_{k=q+1}^{t} \{\gamma(t) - t^{-1} + \delta(t, k)\}\hat{\varepsilon}(k)\,.$$

Now, since $\|C_\theta{}^k\| \leqq K\rho_E{}^k$, $\theta \in E$, with $\rho_E < 1$ (E being a compact subset of Θ and K being some constant), the derivatives of $\theta \to \delta(t, k)$ up to second order can be seen to be bounded in absolute value on E by $Kk^2\rho_E{}^k/t$. Thus let $g_\theta(k, t) = \gamma(t) - t^{-1} + \delta(t, k)$ and $g_\theta{}^{(i)}(k, t)$ be the vector with components the ith derivatives of $g_\theta(k, t)$ with respect to θ; then by the above result and the fact that the sequence $\gamma(t) - t^{-1}$ is $O(t^{-2})$, for any compact set E of Θ, we have

$$\textstyle\sum_{k=q+1}^{t-1} \sup_{\theta\in E} \|g_\theta{}^{(i)}(k, t)\| = K/t\,, \quad t > q\,, \qquad i = 0, 1, 2$$

where K is some constant.

Now the derivatives of

$$\textstyle\sum_{k=q+1}^{t-1} g_\theta(k, t)\hat{\varepsilon}_\theta(t)\,,$$

with respect to θ up to second order, are sums of products of the derivatives with respect to θ up to second order of $\hat{\varepsilon}_\theta(t)$ and of $g_\theta(k, t)$, and thus by Lemma 2 and the above result are bounded in the L^2 norm. This is the result of the lemma.

REMARK. In the special case when $p = 1$, the vector e and the matrix B both reduce to the scalar 1, $Q(t) = \lambda(t)$, $\mu(t) = \sigma^2\lambda^{-1}(t) = \beta + t + 1$ (see the remark at the end of Section 2), so that the results of Lemma 1 are obvious, and $\hat{\varepsilon}_\theta(t)$ is just

$$\textstyle\sum_{k=q+1}^{t} \{X(k) - \sum_{j=1}^{q} a_j X(k-j)\} = \sum_{k=q+1}^{t} W(k)\,,$$

the $W(k)$ now being scalars, so that using the simplified form of Lemma 4 with all $c_j{}^* = 0$, one gets almost immediately:

$$\hat{\varepsilon}_\theta(t) = Y(t) - Y(q) - \textstyle\sum_{j=1}^{q} a_j\{Y(t-j) - Y(q-j)\} \tag{3.4}$$

from which the results of Lemma 2 are obvious. Also the proof of Lemma 3 is considerably simplified since the considered sequence is reduced to

$$-\frac{\beta}{t(\beta + t)} \textstyle\sum_{j=q+1}^{t-1} \hat{\varepsilon}_\theta(j)\,, \quad t > q\,.$$

4. Asymptotic properties of the log-likelihood function. We now show that the approximate (and thus the exact) log-likelihood function possesses all the classical properties of a log-likelihood function, namely:

THEOREM 2. *Let $L_N{}^{(i)}(\cdot)$, $i = 1, 2$ denote the vector of first derivatives of L_N*

and the matrix of second derivatives of L_N, respectively, and θ^* the true value of θ. Then as $N \to \infty$

(i) *$N^{-1}L_N^{(1)}(\theta^*)$ converges in probability to 0;*

(ii) *$N^{-1}L_N^{(2)}(\theta^*)$ converges in probability to*

$$-J = -\begin{pmatrix} (\sigma^*)^{-2}\Gamma & 0 \\ 0 & \frac{1}{2}(\sigma^*)^{-4} \end{pmatrix}$$

where Γ is the matrix with general element

$$= \frac{(\sigma^*)^2}{2\pi} \int_{-\pi}^{\pi} \left\{\frac{\partial}{\partial\theta_\alpha} \psi_\theta(e^{i\lambda}) \frac{\partial}{\partial\theta_\beta} \psi_\theta(e^{-i\lambda})\right\}_{\theta=\theta^*} |\psi_{\theta^*}(e^{i\tau})|^{-2}\, d\lambda\,,$$

where $\psi_\theta(z) = (1 - c_1 z \cdots - c_{p-1}z^{p-1})^{-1}(1 - a_1 z \cdots - a_q z^q)$;

(iii) *if the $\varepsilon(t)$ possesses a finite fourth cumulant κ, then $N^{-\frac{1}{2}}L_N^{(1)}(\theta^*)$ converges in distribution to a Gaussian random vector with zero mean and covariance matrix*

$$\begin{pmatrix} (\sigma^*)^{-2}\Gamma & 0 \\ 0 & \frac{1}{2}(\sigma^*)^{-4} + \frac{1}{4}\kappa(\sigma^*)^{-8} \end{pmatrix}.$$

PROOF. We remark that $\hat{\zeta}_\theta(t)$ depends only on $(a_1, \cdots, a_q, c_1, \cdots, c_{p-1})$. Let $\hat{\zeta}_\theta^{(i)}(t)$, $i = 1, 2$ denote respectively the vector of first derivatives and the matrix of second derivatives of $\hat{\zeta}_\theta(t)$ with respect to these parameters and

$$\gamma_N(\theta) = \frac{1}{N-q} \sum_{t=q+1}^N \hat{\zeta}_\theta(t)\hat{\zeta}_\theta^{(1)}(t)\,, \tag{4.1}$$

$$\gamma_N'(\theta) = \frac{1}{N-q} \sum_{t=q+1}^N \hat{\zeta}_\theta(t)\hat{\zeta}_\theta^{(2)}(t)\,, \tag{4.2}$$

$$\sigma_N^2(\theta) = \frac{1}{N-q} \sum_{t=q+1}^N \hat{\zeta}_\theta^2(t)\,, \tag{4.3}$$

$$\Gamma_N(\theta) = \frac{1}{N-q} \sum_{t=q+1}^N \hat{\zeta}_\theta^{(1)}(t)\, {}^T\hat{\zeta}_\theta^{(1)}(t)\,. \tag{4.4}$$

Then a direct computation shows that

$$L_N^{(1)}(\theta) = -(N-q)\begin{pmatrix} \gamma_N(\theta)/\sigma^2 \\ \{\sigma^2 - \sigma_N^2(\theta)\}/(2\sigma^4) \end{pmatrix}, \tag{4.5}$$

$$L_N^{(2)}(\theta) = -(N-q)\begin{pmatrix} \{\Gamma_N(\theta) + \gamma_N'(\theta)\}/\sigma^2 & -\gamma_N(\theta)/\sigma^4 \\ -{}^T\gamma_N(\theta)/\sigma^4 & \{\sigma_N^2(\theta) - \sigma^2/2\}/\sigma^6 \end{pmatrix}. \tag{4.6}$$

The theorem then follows from Lemmas 5 and 6.

LEMMA 5. *As $N \to \infty$, $\gamma_N(\theta^*)$, $\gamma_N'(\theta^*)$, $\sigma_N^2(\theta^*)$ and $\Gamma_N^{-1}(\theta^*)$ converge in probability respectively to $0, 0, (\sigma^*)^2$ and Γ.*

LEMMA 6. *If the $\varepsilon(t)$ possess a finite fourth cumulant, κ, then as $N \to \infty$, $N^{\frac{1}{2}}\, {}^T({}^T\gamma_N(\theta^*), \sigma_N^2(\theta^*) - (\sigma^*)^2)$ converges in distribution to a Gaussian random vector* *of zero mean and covariance matrix*

$$\begin{pmatrix} \Gamma(\sigma^*)^2 & 0 \\ 0 & 2(\sigma^*)^4 + \kappa \end{pmatrix}.$$

The proofs of Lemmas 5 and 6 involve the approximations to $\hat{\zeta}_{\theta^*}(t)$ and $\hat{\zeta}_{\theta^*}^{(i)}(t)$, $i = 1, 2$ by $\varepsilon(t)$ and some stationary processes $\bar{\varepsilon}^{(i)}(t)$, $i = 1, 2$ as described in

LEMMA 7. *Let $Y(t)$ be defined as in Lemma 4 and*

$$\bar{\varepsilon}_\theta(t) = \sum_{j=0}^{\infty} \pi_j(\theta) Y(t-j)$$

where $\pi_j(\theta)$, $j \geqq 0$ are the coefficients of the Taylor expansion of $(1 - c_1 z \cdots - c_{p-1} z^{p-1})^{-1}(1 - a_1 z \cdots - a_q z^q)$; then

$$\hat{\zeta}_\theta(t) = \bar{\varepsilon}_\theta(t) - \frac{1}{t} \sum_{k=q+1}^{t-1} \bar{\varepsilon}_\theta(k) + \eta_\theta(t)$$

where the sequence of random functions $\theta \to \eta_\theta(t)$, $t > q$ is $O(t^{-1})$ in the L^2 norm.

PROOF. It is well known that $\bar{\varepsilon}_\theta(t)$ satisfies the equation

$$\bar{\varepsilon}_\theta(t) - \sum_{j=1}^{p-1} c_j \bar{\varepsilon}_\theta(t-j) = Y(t) - \sum_{j=1}^{q} a_j Y(t-j).$$

This is the same equation as (2.12) except that $\hat{\phi}_\theta(t)$, $X(t)$ are replaced by $\bar{\varepsilon}_\theta(t)$, $Y(t)$. Using an argument similar to that in the proof of Lemma 2, we get

$$\bar{\varepsilon}_\theta(t) = {}^T u_1 \sum_{k=0}^{\infty} C_\theta{}^k u_1 \{Y(t-k) - \sum_{j=1}^{q} a_j Y(t-k-j)\},$$
$$\hat{\phi}_\theta(t) = {}^T u_1 \sum_{k=0}^{t-q-1} C_\theta{}^k u_1 \{X(t-k) - \sum_{j=1}^{q} a_j X(t-k-j)\}.$$

Thus

$$\begin{aligned} \bar{\varepsilon}_\theta(t) - \bar{\varepsilon}_\theta(t-1) - \hat{\phi}_\theta(t) &= \sum_{k=t-q}^{\infty} {}^T u_1 C_\theta{}^k u_1 \{X(t-k) - \sum_{j=1}^{q} a_j X(t-k-j)\} \\ &= \sum_{l=0}^{\infty} \{X(q-l), \cdots, X(-l)\} g(t-q-l), \end{aligned}$$

where $g_\theta(k)$ is the vector with components ${}^T u_1 C_\theta{}^k u_1$, $-{}^T u_1 C_\theta{}^k u_1 a_1, \cdots, -{}^T u_1 C_\theta{}^k u_1 a_q$. Consequently

$$\begin{aligned} \hat{\zeta}_\theta(t) &= \hat{\varepsilon}_\theta(t) - \frac{1}{t} \sum_{k=q+1}^{t-1} \hat{\varepsilon}_\theta(k) = \sum_{k=q+1}^{t} \frac{k}{t} \hat{\phi}_\theta(k) \\ &= \bar{\varepsilon}_\theta(t) - \frac{1}{t} \sum_{k=q+1}^{t-1} \bar{\varepsilon}_\theta(k) - \frac{q+1}{t} \bar{\varepsilon}_\theta(q) \\ &\quad - \sum_{l=0}^{\infty} \{X(q-l), \cdots, X(-l)\} \left\{ \sum_{k=q+t}^{t} \frac{k}{t} g(k-q-l) \right\}. \end{aligned}$$

Now $\|C^k\| \leqq K\rho_E{}^k$, $\theta \in E$ with $\rho_E < 1$, where E is any compact set of Θ and K is some constant, it can be verified that the derivatives up to second order of

$$\theta \to \sum_{k=q+1}^{t} \frac{k}{t} g(k-q-l)$$

are bounded in norm in E by $K' \rho_E{}^l / t$ (K' being some constant). We deduce that the last term of the above right-hand side is $O(t^{-1})$ in the L^2 norm.

PROOF OF LEMMA 5. We shall denote by $\bar\varepsilon^{(i)}(t)$, $i = 1, 2$ the vector of first derivatives and the matrix of second derivatives of the function $\theta \to \bar\varepsilon_\theta(t)$, at $\theta = \theta^*$. Now, since $\bar\varepsilon_{\theta^*}(t) = \varepsilon(t)$, by Lemma 7, we have

$$\hat\zeta_{\theta^*}(t) - \varepsilon(t) = -\frac{1}{t} \sum_{k=q+1}^{t-1} \varepsilon(k) + O(t^{-1}) = \delta(t) + O(t^{-1}), \quad \text{say;} \tag{4.7}$$

$$\begin{aligned}\hat\zeta_{\theta^*}^{(i)}(t) - \bar\varepsilon^{(i)}(t) &= -\frac{1}{t} \sum_{k=q+1}^{t-1} \bar\varepsilon^{(i)}(k) + O(t^{-1}) \\ &= \delta^{(i)}(t) + O(t^{-1}), \quad \text{say} \quad (i = 1, 2),\end{aligned} \tag{4.8}$$

where $O(t^{-1})$ denotes a sequence of random variables such that $tO(t^{-1})$ is bounded in the L^2 norm.

Now the variance of $\delta(t)$ is $(t - q - 1)(\sigma^*)^2/t^2$, that of $\delta_j^{(1)}(t)$ (resp. $\delta_{ij}^{(2)}(t)$) is

$$\frac{1}{t^2} \sum_{k=q+1}^{t-1} \sum_{l=q+1}^{t-1} R(k - l) = \frac{1}{t} \left\{ \sum_{|s|<t-q} \left(1 - \frac{q + 1 + |s|}{t}\right) R(s) \right\}$$

where $R(\cdot)$ denotes the covariance function of the stationary process $\bar\varepsilon_j^{(1)}(t)$ (resp. $\bar\varepsilon_{ij}^{(2)}(t)$). Since $\sum_{-\infty}^{\infty} R(s) < +\infty$, the variances of $\delta(t)$, $\delta_j^{(1)}(t)$ and $\delta_{ij}^{(2)}(t)$ can be bounded by K/t, K being some constant. Thus

$$\frac{1}{N} \sum_{t=q+1}^{N} \hat\zeta_{\theta^*}^2(t) = \frac{1}{N} \sum_{t=q+1}^{N} \varepsilon^2(t) + \frac{1}{N} \sum_{t=q+1}^{N} \eta(t), \tag{4.9}$$

$$\begin{aligned}
\frac{1}{N} \sum_{t=q+1}^{N} \hat\zeta_{\theta^*}(t)\hat\zeta_{\theta^*,j}^{(1)}(t) &= \frac{1}{N} \sum_{t=q+1}^{N} \varepsilon(t)\bar\varepsilon_j^{(1)}(t) + \frac{1}{N} \sum_{t=q+1}^{N} \eta_j'(t), \\
\frac{1}{N} \sum_{t=q+1}^{N} \hat\zeta_{\theta^*,j}^{(1)}(t)\hat\zeta_{\theta^*,k}^{(1)}(t) &= \frac{1}{N} \sum_{t=q+1}^{N} \bar\varepsilon_j^{(1)}(t)\bar\varepsilon_k^{(1)}(t) + \frac{1}{N} \sum_{t=q+1}^{N} \eta_{jk}''(t), \\
\frac{1}{N} \sum_{t=q+1}^{N} \hat\zeta_{\theta^*}(t)\hat\zeta_{\theta^*,jk}^{(2)}(t) &= \frac{1}{N} \sum_{t=q+1}^{N} \varepsilon(t)\bar\varepsilon_{jk}^{(2)}(t) + \frac{1}{N} \sum_{t=q+1}^{N} \eta_{jk}'''(t),
\end{aligned} \tag{4.10}$$

where the L^1 norms of $\eta(t)$, $\eta_j'(t)$, $\eta_{jk}''(t)$, and $\eta_{jk}'''(t)$ tend to 0 as $t \to \infty$. Using the result

$$\lim \frac{1}{N}(x_1 + \cdots + x_N) = 0$$

if $x_N \to 0$, as $N \to \infty$, we see that the last term of the above right-hand sides converges in L^1 to 0 as $N \to \infty$. On the other hand, the first terms converge respectively almost surely to $(\sigma^*)^2$, 0, Γ_{jk} and 0, the matrix $\Gamma = \{\Gamma_{jk}\}$ being the covariance matrix of $\bar\varepsilon^{(1)}(0)$. Now, the stationary process $\bar\varepsilon^{(1)}(t)$ admits the spectral representation

$$\bar\varepsilon^{(1)}(t) = e^{i\lambda t}\psi_{\theta^*}^{(1)}(e^{i\lambda})\, dZ_Y(\lambda)$$

where $\psi_\theta^{(1)}(z)$ is the vector of derivatives with respect to θ of $(1 - c_1 z \cdots - c_{p-1}z^{p-1})^{-1}(1 - a_1 z \cdots - a_q z^q)$ and $Z_Y(\lambda)$ is the process of orthogonal increments appearing in the spectral representation of $Y(t)$. Thus Γ is equal to the expression given in Theorem 2. This completes the proof of the theorem.

PROOF OF LEMMA 6. From (4.9) and (4.10) we clearly have

$$N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \hat{\zeta}_{\theta^*}(t)\hat{\zeta}_{\theta^*,j}^{(1)}(t) = N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \varepsilon(t)\bar{\varepsilon}^{(1)}(t) + N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \eta_j'(t)$$
$$N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \{\hat{\zeta}_{\theta^*}^2(t) - (\sigma^*)^2\} = N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \{\varepsilon^2(t) - (\sigma^*)^2\} + N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \eta(t)\,,$$

where $\eta(t)$, $\eta_j'(t)$ are seen to be

$$\eta(t) = \delta^2(t) + 2\delta(t)\varepsilon(t) + O(t^{-1})$$
$$\eta_j'(t) = \delta(t)\delta_j^{(1)}(t) + \delta(t)\bar{\varepsilon}_j^{(1)}(t) + \delta_j^{(1)}(t)\varepsilon(t) + O(t^{-1})\,,$$

$\delta(t)$ and $\delta^{(1)}(t)$ being defined by (4.7) and (4.8) and $O(t^{-1})$ being a sequence of random variables such that $tO(t^{-1})$ tends to O in the L^1 norm as $t \to \infty$.

Let $Z(t)$ be the random vector with components $Z_j(t) = \varepsilon(t)\bar{\varepsilon}_j^{(1)}(t)$, $j = 1, \cdots, q$, $Z_{q+1}(t) = (\sigma^*) - \varepsilon^2(t)$, then $Z(t)$ is a stationary ergodic process with the martingale property: $E\{Z(t) \mid Z(s), s < t\} = 0$. By Billingsley's (1968, page 206) theorem, the sum $N^{-\frac{1}{2}}\{Z(q+1) + \cdots + Z(N)\}$ converges in distribution as $N \to \infty$ to a Gaussian vector with zero mean and covariance matrix

$$E\{Z(t)^T Z(t) = \begin{pmatrix} (\sigma^*)^2\Gamma & 0 \\ 0 & 2(\sigma^*)^4 + \kappa \end{pmatrix}.$$

Thus we obtain the result of the lemma if we show that

$$N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \eta_j(t) \qquad \text{and} \qquad N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \eta(t)$$

tend to O in probability as $N \to \infty$.

Now we have seen in the proof of Lemma 5 that the variances of $\delta(t)$ and of $\delta_j^{(1)}(t)$ are bounded by K/t. Thus the L^1 norms of:

$$N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \delta^2(t) \qquad \text{and} \qquad N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \delta(t)\delta_j^{(1)}(t)$$

are bounded by $KN^{-\frac{1}{2}} \log (N/q)$, which tends to 0 as $N \to \infty$. Thus we need only to show that the sums

$$N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \delta(t)\varepsilon(t)\,, \qquad N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \delta(t)\bar{\varepsilon}_j^{(1)}(t) \qquad \text{and}$$
$$N^{-\frac{1}{2}} \sum_{t=q+1}^{N} \delta_j^{(1)}(t)\varepsilon(t)$$

tend to 0 in probability as $N \to \infty$. To do that we shall compute their variances. Let ε_N, $\varepsilon_{N,j}$ and $\varepsilon'_{N,j}$ denote respectively the above sums. Then we have

$$\sigma^2(\varepsilon_{N,j}) = \frac{1}{N} \sum_{t=q+1}^{N} \sum_{s=q+1}^{N} (\mathrm{Cov}\,\{\delta(t), \bar{\varepsilon}_j^{(1)}(s)\}\, \mathrm{Cov}\,\{\delta(s), \bar{\varepsilon}_j^{(1)}(t)\}$$
$$+ \mathrm{Cov}\,\{\delta(t), \delta(s)\}R(t-s) + \mathrm{Cum}\,\{\delta(s), \bar{\varepsilon}_j^{(1)}(s), \delta(t), \bar{\varepsilon}_j^{(1)}(t)\})$$

where $R(\cdot)$ denotes the covariance function of the stationary process $\bar{\varepsilon}_j^{(1)}(t)$.

The above result follows from the formula

$$\mathrm{Cov}\,(X_1X_2, X_3X_4) = \mathrm{Cov}\,(X_1, X_3)\,\mathrm{Cov}\,(X_2, X_4) + \mathrm{Cov}\,(X_1, X_4)\,\mathrm{Cov}\,(X_2, X_3)$$
$$+ \mathrm{Cum}\,(X_1, X_2, X_3, X_4)$$

for any zero mean random variables X_1, X_2, X_3, X_4.

Now,

$$\operatorname{Cov}\{\delta(t), \bar{\varepsilon}_j^{(1)}\bar{\varepsilon}_j^{(1)}(t)\} = \frac{1}{t}\sum_{k=q+1}^{t-1} E\{\varepsilon(k)\bar{\varepsilon}_j^{(1)}(s)\}\,.$$

By using the fact that $\bar{\varepsilon}_j^{(1)}(s)$ can be expressed in the form

$$\bar{\varepsilon}_j^{(1)}(s) = \sum_{k=0}^{\infty} d_k\varepsilon(s-k)$$

with $\sum_0^\infty |d_k| < +\infty$, we have

$$\sum_k |E\{\varepsilon(k)\bar{\varepsilon}_j^{(1)}(s)\}| \leqq K\,, \qquad \forall\, s \quad (K = \text{constant}).$$

Thus $|\operatorname{Cov}\{\delta(t), \bar{\varepsilon}_j^{(1)}(s)\}| \leqq K/t$ and consequently

$$\begin{aligned}&|\textstyle\sum_{t=q+1}^N \sum_{s=q+1}^N \operatorname{Cov}\{\delta(t), \bar{\varepsilon}_j^{(1)}(s)\}\operatorname{Cov}\{\delta(s), \bar{\varepsilon}_j^{(1)}(t)\}|\\ &\qquad \leqq \sum_{t=q+1}^N \sum_{s=q+1}^N \frac{K^2}{st} = (\textstyle\sum_{t=q+1}^N K/t)^2 \leqq \left(K\log\dfrac{N}{q}\right)^2.\end{aligned}$$

On the other hand, the covariance between $\delta(t)$ and $\delta(s)$ is:

$$(\sigma^*)^2\frac{1}{st}\{\min(s, t) - q - 1\}$$

and is thus bounded in absolute value by $K/\max(s, t)$, K being some constant. It follows that

$$\begin{aligned}&|\textstyle\sum_{t=q+1}^N \sum_{s=q+1}^N \operatorname{Cov}\{\delta(t), \delta(s)\}R(t-s)|\\ &\qquad \leqq 2\sum_{t=q+1}^N \frac{K}{t}\left\{\textstyle\sum_{s=q+1}^t |R(t-s)| \leqq K'\log\left(\dfrac{N}{q}\right)\right\}\end{aligned}$$

where K' is some other constant.

Finally, the cumulant between $\delta(s)$, $\delta(t)$, $\bar{\varepsilon}_j^{(1)}(s)$, and $\bar{\varepsilon}_j^{(1)}(t)$ is

$$\frac{1}{st}\sum_{l=q+1}^{\min(s,t)-1}\operatorname{Cum}\{\varepsilon(l), \varepsilon(l), \bar{\varepsilon}_j^{(1)}(s), \bar{\varepsilon}_j^{(1)}(t)\}\,.$$

By using the fact that $\bar{\varepsilon}^{(1)}(t) = \sum_0^\infty d_k\varepsilon(t-k)$, $\sum_0^\infty d_k^2 < +\infty$, we get:

$$|\textstyle\sum_p \operatorname{Cum}\{\varepsilon(l), \bar{\varepsilon}_j^{(1)}(t), \varepsilon(l), \bar{\varepsilon}_j^{(1)}(s)\}| \leqq \sum K|d_{t-l}d_{s-l}| < +\infty$$

where K is some constant. Thus

$$|\operatorname{Cum}\{\delta(t), \bar{\varepsilon}_j^{(1)}(t), \delta(s), \bar{\varepsilon}_j^{(1)}(s)\}| \leqq \frac{K'}{st}\,, \quad K' = \text{constant}\,,$$

and consequently

$$\sum_{t=q+1}^N \sum_{s=q+1}^N \operatorname{Cum}\{\delta(t), \bar{\varepsilon}_j^{(1)}(t), \delta(s), \bar{\varepsilon}_j^{(1)}(s)\} \leqq \left(K\log\frac{N}{q}\right)^2.$$

The results above show that $\sigma^2(\varepsilon_{N,j}) \to 0$ as $N \to \infty$. A similar argument shows that $\sigma^2(\varepsilon_N) \to 0$ as $N \to \infty$. It remains to consider

$$\sigma^2(\delta'_{N,j}) = \sigma^2\{N^{-\frac{1}{2}}\textstyle\sum_{t=q+1}^N \delta_j^{(1)}(t)\varepsilon(t)\} = \dfrac{(\sigma^*)^2}{N}\textstyle\sum_{t=q+1}^N \sigma^2\{\delta_j^{(1)}(t)\}$$

since the random variable $\delta_j^{(1)}(t)$ is independent of $\varepsilon(t)$. Now the variance of $\delta_j^{(1)}(t)$ is bounded by K/t where K is a constant. It follows that the variance of $\varepsilon'_{N,j}$ is bounded by $K(\sigma^*)^2 \log (N/q)/N$ which tends to 0 as $N \to \infty$.

This completes the proof of the lemma.

REMARK. In the special case when $p = 1$, the proof for Theorem 2 can be somewhat shortened. Indeed $\hat{\zeta}_\theta(t)$ in this case can be seen to be a linear function of $a_1, \cdots, a_q$. We have, by (3.2) and (3.4),

$$\hat{\zeta}_\theta(t) = U_0(t) - \sum_{j=1}^q a_j U_j(t) , \tag{4.11}$$

where

$$U_j(t) = Y(t-j) - \frac{1}{t} \sum_{k=q+1}^{t-1} Y(k-j) - \frac{q+1}{t} Y(q-j) , \qquad j = 0, \cdots, q , \tag{4.12}$$

$Y(t)$ being defined in Lemma 4. Since $\tilde{\varepsilon}_\theta(t) - Y(t) - a_1 Y(t-1) \cdots - a_q Y(t-q)$ in the present case, the result of Lemma 7 is trivial. Also, $\hat{\zeta}_\theta^{(2)}(t)$ and $\tilde{\varepsilon}_\theta^{(2)}(t)$ are now null so that we need not consider them in the proof of Lemma 5.

Finally, the matrix Γ of Theorem 2 in this case reduces to

$$\begin{pmatrix} R(0) & R(1) & \cdots & R(q-1) \\ R(1) & R(0) & \cdots & R(q-2) \\ \cdot & \cdot & \cdots & \cdot \\ R(q-1) & R(q-2) & \cdots & R(0) \end{pmatrix}$$

where $R(\cdot)$ is the covariance function of the stationary process $Y(t)$.

5. Estimation of the parameters. We estimate θ by maximizing the approximate log-likelihood function L_N, or more frequently by solving the equation $L_N^{(1)}(\theta) = 0$. Using (4.5) and remarking that $\gamma_N(\theta)$ and $\sigma_N^2(\theta)$ depend only on $a = {}^T(a_1, \cdots, a_q)$ and $c = {}^T(c_1, \cdots, c_{p-1})$, the resulting estimate $\hat{\theta}_N = (\hat{a}_N, \hat{c}_N, \hat{\sigma}_N^2)$ is given by

$$\gamma_N(\hat{a}_N, \hat{c}_N) = 0$$
$$\hat{\sigma}_N^2 = \sigma_N^2(\hat{a}_N, \hat{c}_N)$$

where we have written $\gamma_N(a, c)$ and $\sigma_N^2(a, c)$ in place of $\gamma_N(\theta)$ and $\sigma_N^2(\theta)$.

One should remark that in the special case when $p = 1$, the equation $L_N^{(1)}(\theta) = 0$ has a explicit solution. Indeed, using (4.11), a direct computation shows that the estimate $\hat{\theta}_N = (\hat{a}_N, \hat{\sigma}_N^2)$ which maximizes L_N is the solution of:

$$\sum_{k=1}^q \{\sum_{t=q+1}^N U_j(t) U_k(t)\} \hat{a}_k = \sum_{t-q+1}^N U_j(t) U_0(t) , \tag{5.1}$$

$$\hat{\sigma}^2 = \frac{1}{N-q} \sum_{t=q+1}^N \hat{\zeta}_{\hat{\theta}}^2(t) . \tag{5.2}$$

Using the fact that

$$\sum_{k=1}^q \{\sum_{t=q+1}^N U_j(t) U_k(t)\}(\hat{a}_k - a_k^*) = \sum_{t=q+1}^N \hat{\zeta}_{\theta^*}(t) U_j(t)$$

$$\begin{aligned}\hat\sigma^2 &= \frac{1}{N-q}\,[\textstyle\sum_{t=q+1}^{N} \hat\zeta_{\hat\theta}(t)\{\hat\zeta_{\theta^*}(t) - \sum_{k=1}^{q}(\hat a_k - a_k^*)U_k(t)\}] \\ &= \frac{1}{N-q}\textstyle\sum_{t=q+1}^{N}\hat\zeta_{\hat\theta}(t)\hat\zeta_{\theta^*}(t) \\ &= \frac{1}{N-q}\,[\textstyle\sum_{t=q+1}^{N}\hat\zeta^2_{\theta^*}(t) - \sum_{j=1}^{q}(\hat a_j - a_j^*)\{\sum_{t=q+1}^{N}U_j(t)\hat\zeta_{\theta^*}(t)\}]\,,\end{aligned}$$

and the fact that $\zeta^{(1)}(t) = -^T(U_1(t), \cdots, U_q(t))$, the results of Lemmas 5 and 6 give:

Suppose that Γ is invertible, then the estimate $\hat\theta_N = (\hat a_N, \hat\sigma_N{}^2)$ is consistent and $N^{\frac12}(\hat\theta_N - \theta^*)$ converges in distribution as $N \to \infty$ to a Gaussian vector with zero mean and covariance matrix

$$\begin{pmatrix}(\sigma^*)^2\Gamma^{-1} & 0 \\ 0 & 2(\sigma^*)^4 + \kappa\end{pmatrix}.$$

However, in the general case, the existence of a consistent approximate maximum likelihood estimate (i.e., satisfying the equation $L_N{}^{(1)}(\theta) = 0$) is not immediate. To show the existence of such an estimate, we need this result:

LEMMA 8. *There exists a function $g: (0, \infty) \to [0, \infty]$ with $g(r) \downarrow 0$ as $r \downarrow 0$ and a sequence of random variables M_N, $N > q$ bounded in the L^1 norm such that*

$$\forall\,\theta\,,\quad \|\theta - \theta^*\| \leqq r:\ N^{-1}\|L_N{}^{(2)}(\theta) - L_N{}^{(2)}(\theta^*)\| \leqq g(r)M_N\,.$$

PROOF. Using (4.6) and (4.1)—(4.4), and remarking that if the sequences $U(t)$, $t > q$ and $V(t)$, $t > q$ of random variables are bounded in the L^2 norm then the sequence

$$\frac{1}{N}\textstyle\sum_{t=q+1}^{N} U(t)V(t)\,,\quad N > q$$

is bounded in the L^1 norm, we obtain the result of the lemma if

(i) the sequence $\hat\zeta^{(i)}_{\theta^*}(t)$, $t > q$, $i = 0, 1, 2$ is bounded in the L^2 norm;

(ii) there exists a function $g: (0, \infty) \to [0, \infty]$ with $g(r) \downarrow 0$ as $r \downarrow 0$ and a sequence of random variables $U(t)$, $t > q$, bounded in the L^2 norm such that

$$\forall\,\theta\,,\quad \|\theta - \theta^*\| \leqq r:\ \|\hat\zeta_\theta{}^{(i)}(t) - \hat\zeta^{(i)}_{\theta^*}(t)\| \leqq g(r)U(t) \qquad i = 0, 1, 2\,.$$

Point (i) is a direct consequence of Lemma 2 and (3.2); as for point (ii), it can be proved using the mean value theorem if we have shown that for some $r_0 > 0$, the sequence

$$\sup_{\theta:\|\theta-\theta^*\|\leqq r_0}\|\hat\zeta_\theta{}^{(3)}(t)\|\,,\quad t > q$$

is bounded in the L^2 norm. That this is true can be shown by an argument similar to that in the proof of Lemma 2.

The above result and Theorem 2 permit us to show, by a standard argument (see, e.g., Aitchison and Silvey (1958)) *the existence of a consistent estimate $\hat\theta_N$,* *function of* $X(1), \cdots, X(N)$ *only and satisfying* $L_N^{(1)}(\hat{\theta}_N) = 0$ *with probability tending to one. The sequence* $\hat{\theta}_N$ *is unique in the sense that for any other sequence* $\hat{\theta}_N'$ *with the same properties,* $P\{\hat{\theta}_N = \hat{\theta}_N'\}$ *tends to one as* $N \to \infty$. By Theorem 1, the same results hold for the function $\mathscr{L}_N$. Thus, there exists also a consistent maximum likelihood estimate $\tilde{\theta}_N$. A standard argument using Theorem 2 also shows that $\hat{\theta}_N$ as well as $\tilde{\theta}_N$ are asymptotically normal and efficient. Moreover:

THEOREM 3. *Let* $\hat{\theta}_N$ *and* $\tilde{\theta}_N$ *be defined as above, then the sequence* $N(\log N)^{-1}(\hat{\theta}_N - \tilde{\theta}_N)$ *is bounded in probability.*

Here a sequence U_N is said to be bounded in probability if

$$\lim_{a\to\infty} \limsup_{N\to\infty} P\{\|U_N\| > a\} = 0 .$$

PROOF. By the mean value theorem

$$\frac{\partial}{\partial\theta_\alpha} L_N(\tilde{\theta}_N) = \sum_\beta \frac{\partial^2}{\partial\theta_\alpha \partial\theta_\beta} L_N(\theta_N')(\tilde{\theta}_{N,\beta} - \hat{\theta}_{N,\beta}) ,$$

where θ_N' is some point on the segment joining $\hat{\theta}_N$ and $\tilde{\theta}_N$ and thus tends to θ^* as $N \to \infty$. Using the results of Theorem 2 and Lemma 8, we get

$$N^{-1}L_N^{(1)}(\tilde{\theta}_N) = -J_N(\tilde{\theta}_N - \hat{\theta}_N)$$

where $-J_N$ tends in probability to the matrix $-J$ of Theorem 2.

Let E be a compact neighborhood of θ^* and Z_N be the sequence of random variables of Theorem 1 such that

$$\sup_{\theta\in E} \|\mathscr{L}_N^{(1)}(\theta) - L_N^{(1)}(\theta)\| \leqq (\log N)Z_N .$$

If $\tilde{\theta}_N \in E$ and $\mathscr{L}_N^{(1)}(\tilde{\theta}_N) = 0$ then

$$N(\log N)^{-1}\|\tilde{\theta}_N - \hat{\theta}_N\| \leqq \|J_N^{-1}\|Z_N = R_N , \quad \text{say.}$$

Thus

$$P\{N(\log N)^{-1}\|\tilde{\theta}_N - \hat{\theta}_N\| > a\} \leqq P\{\mathscr{L}_N^{(1)}(\tilde{\theta}_N) \neq 0\} + P\{\tilde{\theta}_N \notin E\} + P\{R_N > a\} .$$

The sequence R_N can be shown to be bounded in probability since the sequences $\|K_N^{-1}\|$ and Z_N are. On the other hand, $P\{\mathscr{L}_N^{(1)}(\tilde{\theta}_N) = 0\}$ tends to one and $\tilde{\theta}_N \to \theta^*$ as $N \to \infty$, so that

$$\limsup_{N\to\infty} P\{N(\log N)^{-1}\|\tilde{\theta}_N - \hat{\theta}_N\| > a\} \leqq \limsup_{N\to\infty} P\{R_N > a\} .$$

The above right-hand side tends to 0 as $a \to \infty$, which gives the result.

In conclusion, one should remark that the main results are Theorems 1 and 2 and Lemma 8. They ensure that we can estimate parameters by standard methods as in the case of an ordinary ARMA process, provided that we take the $\hat{\zeta}_\theta(t)$ defined by (3.2) as the residuals. In particular, we can obtain estimates of θ by nonlinear regression (see Box and Jenkins (1970)) as follows:

Let $\theta_N^{(n)} = (a_N^{(n)}, c_N^{(n)}, (\sigma_N^2)^{(n)})$ by the n-step estimate of θ; then

$$\begin{pmatrix} a_N^{(n+1)} \\ c_N^{(n+1)} \end{pmatrix} = \begin{pmatrix} a_N^{(n)} \\ c_N^{(n)} \end{pmatrix} - \{\Gamma_N(a_N^{(n)}, c_N^{(n)})\}^{-1}\gamma_N(a_N^{(n)}, c_N^{(n)})$$

where Γ_N and γ_N are given in (4.4) and (4.1). If the estimate $\theta_N^{(0)}$ is $N^{\frac{1}{2}}$-consistent, in the sense that $N^{\frac{1}{2}}(\theta_N^{(0)} - \theta^*)$ is bounded in probability, then it can be shown that $N^{\frac{1}{2}}(\theta_N^{(1)} - \hat{\theta}_N)$ tends to 0 in probability as $N \to \infty$ (see Pham-dinh (1975)).

Remark. In the special case when $p = 1$, the above iteration is unnecessary since the approximate maximum likelihood estimate $\hat{\theta}_N$ of θ can be computed directly by (5.1) and (5.2). Now the matrix Γ_N

$$\Gamma_{N,jk} = \frac{1}{N} \sum_{t=q+1}^{N} U_j(t)U_k(t) , \qquad j, k = 0, \cdots, q ,$$

which is involved in the computation of $\hat{\theta}_N$, is not a stationary matrix (i.e., $\Gamma_{N,jk}$ is not of the form $R_N(j-k)$). However, $\Gamma_{N,jk}$ is a consistent estimate of $R(j-k)$ so that it is more convenient to replace it by

$$R_N(j-k) = \frac{1}{N} \sum_{t=1}^{N} U_N(t-j)U_N(t-k)$$

where

$$U_N(t) = \sum_{k=1}^{t} \frac{k}{t} X(k) \qquad \text{if} \quad 1 \leqq t \leqq N, = 0 \quad \text{otherwise.}$$

Thus for $j < t < N + j$:

$$U_N(t-j) = \sum_{k=1}^{t-j} \frac{k}{t-j} X(k) = \sum_{k=j+1}^{t} \frac{k-j}{t-j} X(k-j) .$$

By replacing $X(t)$ by $Y(t) - Y(t-j)$ (see Lemma 4), one can show that the L^2 norms of $U_j(t) - U_N(t-j)$, $j = 0, 1, \cdots, q$ are bounded for all $t = q + 1, \cdots, N$ by $K/(t-q)$ where K is some constant. It can then be shown that:

$$NE|\Gamma_{N,jk} - R_N(j-k)| \leqq K \log\left(\frac{N}{q}\right) , \qquad j, k = 0, \cdots, q$$

where K is some constant.

Thus, if we define L_N' by

$$L_N' = -\frac{N}{2} \log(2\pi\sigma^2)$$
$$- \frac{N}{2\sigma^2} \{\sum_{j=1}^{q} \sum_{k=1}^{q} a_j a_k R_N(j-k) - 2 \sum_{j=1}^{q} a_j R_N(j) + R_N(0)\} ,$$

then it is not difficult to see that the sequence $L_N'(\cdot) - L_N(\cdot)$, $N > q$ is $O(\log N)$. Following the same argument as that of Theorem 3, if $\hat{\theta}_N'$ denotes the estimate which maximizes L_N' then $N(\log N)^{-1}(\hat{\theta}_N - \hat{\theta}_N')$ is bounded in probability. The use of $R_N(j-k)$ in place of $\Gamma_{N,jk}$ thus produces an equally good estimate and has the advantage of simplifying the numerical computation.

REFERENCES

Aitchison, J. and Silvey, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29** 813–828.

BILLINGSLEY, P. (1968). *Convergence of Probability Measure*. Wiley, New York.
BOX, G. P. E. and JENKINS, G. M. (1970). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.
COX, D. R. and LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Methuen, London.
HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.
PHAM-DINH, T. (1975). Estimation et test dans les modèles paramétriques de processus stationnaires. Thesis, Université de Grenoble.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720