The Annals of Statistics
1980, Vol. 8, No. 5, 1071–1081

THE ESTIMATION OF THE ORDER OF AN ARMA PROCESS

BY E. J. HANNAN

The Australian National University

Under general conditions strong consistency of certain estimates of the maximum lags of an autoregressive moving average process is established. A theorem on weak consistency is also proved and in certain cases where consistency does not hold the probability of over-estimation of a maximum lag is evaluated.

1. Introduction. We here consider a stationary process, $x(n)$, generated by

(1)

$$\Sigma_0^p\alpha(j)x(n-j) = \Sigma_0^q\beta(j)\varepsilon(n-j), \quad \mathcal{E}\{\varepsilon(n)\} = 0, \quad \mathcal{E}\{\varepsilon(m)\varepsilon(n)\} = \delta_{mn}\sigma^2,$$

$$\alpha(0) = \beta(0) = 1, \quad g(z) = \Sigma_0^p\alpha(j)z^j \neq 0, |z| \leqslant 1;$$

$$h(z) = \Sigma_0^q\beta(j)z^j \neq 0, |z| \leqslant 1.$$

It is also assumed that g, h have no common zero. These conditions ensure that $x(n)$ may be represented as a moving average

$$\Sigma_0^\infty\kappa(j)\varepsilon(n-j), \Sigma_0^\infty\kappa(j)z^j = k(z) = g^{-1}h,$$

where the $\kappa(j)$ decrease to zero at a geometric rate, and that the $\varepsilon(n)$ are the linear innovations. It has been assumed that $\mathcal{E}\{x(n)\} = 0$ but this is immaterial to all of the results presented below, which would continue to hold if the various statistics involved were computed from (sample) mean corrected quantities.

Our purpose is to study the estimation of the true order, which shall be called p_0, q_0. A zero subscript shall be used throughout for true quantities, e.g., $\alpha_0(j)$, g_0 and k_0. This estimation problem is considerably more complex than that for the case $q = 0$, which has been studied, for example in [7], [14]. This is because when $p > p_0, q > q_0$ the estimates $\hat{\alpha}(j), \hat{\beta}(j)$, obtained by maximising the Gaussian likelihood, do not converge in any reasonable sense because the likelihood is constant along the "line" where $k = k_0$, so that as N (the sample size) increases the sample point will search up and down that line.

Though the estimation method used will be based on the maximisation of the Gaussian likelihood the assumption of normality will not be made. It will always be assumed that

(2) $$\mathcal{E}\{\varepsilon(n)|\mathcal{F}_{n-1}\} = 0, \quad \mathcal{E}\{\varepsilon(n)^2|\mathcal{F}_{n-1}\} = \sigma^2, \quad \mathcal{E}\{\varepsilon(n)^4\} < \infty,$$

where $\mathcal{F}_n$ is the σ-algebra determined by $\varepsilon(m), m \leqslant n$. The first of these conditions, (2), seems near to minimal since without it linear prediction is suboptimal. The

Received February 1979; revised July 1979.

AMS 1970 *subject classifications.* Primary 62M10; secondary 60F15.

Key words and phrases. ARMA process, maximum lags, law of the iterated logarithm, identification, strong consistency.

second could be considerably relaxed as we indicate in Section 3 below. The third seems minor.

Two other restrictions will be imposed. *In the first place it will be assumed that* $p_0 \leqslant P, q_0 \leqslant Q$ *where* P, Q *are known a priori*. It will follow that all of the results given below will continue to hold if P, Q are allowed to increase sufficiently slowly, with N, but how slowly we do not know. The remaining restriction is that $h(z)$ *will be required to have all of its zeros bounded away from the unit circle by a positive, albeit arbitrarily small, quantity*, δ, *prescribed a priori*. Of course h_0 also must satisfy this requirement.

The estimates of p_0, q_0 here considered will be based on the estimates $\hat{\sigma}^2_{p,q}$, of σ^2, obtained by maximising the Gaussian likelihood (though Gaussian assumptions are not maintained). In fact, except for a constant, $-2N^{-1}$ times the maximised log-likelihood is $\log \hat{\sigma}^2_{p,q} + N^{-1} \log \det \hat{P}_N$ where $\sigma^2 P_N$ is the covariance matrix of the data and $\hat{P}_N$ is the estimate, which is a function only of $\hat{g}^{-1}\hat{h}$. The conditions imposed above ensure that eventually the zeros of $\hat{g}, \hat{h}$ are bounded away from the unit circle and hence it is easy to show that $N^{-1} \log \det \hat{P}_N = 0(N^{-1})$. We shall, in accordance with previous practice, therefore omit this term, though it would, even asymptotically, affect the first of the criteria introduced below (but only that one). The estimates are to be obtained by minimizing one or the other of the quantities.

$$\begin{aligned} \mathrm{AIC}(p,q) &= \log \hat{\sigma}^2_{p,q} + 2(p+q)/N \\ \mathrm{BIC}(p,q) &= \log \hat{\sigma}^2_{p,q} + (p+q)\log N/N \\ \phi(p,q) &= \log \hat{\sigma}^2_{p,q} + (p+q)c \log\log N/N, \qquad c > 2. \end{aligned}$$

The first of these was introduced by Akaike [1], in connection with autoregressive model fitting and has been fairly widely used in this and other problems. BIC has been suggested by Akaike, [2], Rissanen, [12], and (in a somewhat different connection) by Schwarz, [13]. The third method was introduced in [7]. Its interest lies in the fact that the second term there decreases as fast as is possible if strong consistency for $\hat{p}, \hat{q}$ is to hold. (It may be that $c = 2$ would suffice but this would require a more delicate analysis.) Because of the sharp nature of this result, depending as it does on the law of the iterated logarithm (LIL), it is much the most difficult to establish. In the present connection that difficulty is increased by the behaviour of the $\hat{\alpha}(j), \hat{\beta}(j)$, for $p > p_0, q > q_0$, already alluded to, which makes it necessary to prove the LIL for a quantity maximised over a region of parameter values. Strong consistency has appeal because it means that from some N_0 on then $\hat{p} = p_0, \hat{q} = q_0$. Of course this of itself casts doubt on the model for one can hardly believe that one can know any parameters exactly from data. It is well known that $\hat{p}, \hat{q}$ from AIC (p,q) are not weakly consistent. That method is designed for a situation where $x(n)$ is *not* generated by one of the processes in the model set. Nevertheless it is desirable to complete the theory of ARMA model fitting by discussing the strongly consistent estimation of p_0, q_0. We shall also state a theorem on weak consistency.

The conditions of this section will be maintained throughout the paper.

2. The theorems on consistency.

THEOREM 1. *If, in addition to the conditions of Section* 1, *the* $\varepsilon(n)$ *are independent then* $\hat{p}, \hat{q}$ *obtained via* BIC (p, q) *or* $\phi(p, q)$ *are strongly consistent. If the* $\varepsilon(n)$ *are not independent but* $\mathcal{E}\{|\varepsilon(n)|^{\gamma}\} < \infty, \gamma > 4$, *then the estimates obtained via* BIC (p, q) *are strongly consistent.*

The requirement of independence, which is likely to hold only in the Gaussian case, is almost certainly not needed and the second part of the theorem almost certainly holds for $\phi(p, q)$ also. We shall discuss this briefly in Section 3.

Assuming $p_0 \leqslant P, q_0 = 0$, the limits

$$\lim_{N\to\infty} \Pr\{\hat{p} = p\}, p \leqslant P$$

are evaluated for AIC in Shibata (1976), on Gaussian assumptions for the $\varepsilon(n)$. For $p < p_0$ the limits are zero while for $p \geqslant p_0$ they depend only on $p - p_0, P - p$. We shall call the limits $\pi(p - p_0, P - p)$, this being zero for $p < p_0$. Then the following is true.

THEOREM 2. *For* AIC, *under the conditions of Section* 1

$$\lim_{N\to\infty} \Pr\{\hat{p} = p_0, \hat{q} = q\} = \pi(q - q_0, Q - q),$$

$$\lim_{N\to\infty} \Pr\{\hat{p} < p_0, \hat{q} = q\} = 0; \qquad P = p_0$$

$$\lim_{N\to\infty} \Pr\{\hat{p} = p, \hat{q} = q_0\} = \pi(p - p_0, P - p), \lim_{N\to\infty}\{\hat{p} = p, \hat{q} < q_0\} = 0;$$

$$Q = q_0.$$

This says that when $P = p_0$ (or $Q = q_0$) Shibata's evaluations continue to hold. This result seems of value only for $Q = 0$ (Shibata's case) or $P = 0$, i.e., a pure moving average. It seems very likely that these probabilities can be asymptotically evaluated in general but we proceed no further here.

THEOREM 3. *Under the conditions of Section* 1, *if the last term in* BIC *or* $\phi(p, q)$ *is replaced by* $(p + q)C_N/N$ *where* C_N *increases to infinity, then the resulting* $\hat{p}, \hat{q}$ *are weakly consistent.*

PROOF OF THEOREM 1. Let x_N be the vector of observations and Γ_N be its covariance matrix. Then $P_N = \sigma^{-2}\Gamma_N$ depends only on k. Moreover $\hat{\sigma}_{p,q}^2 = N^{-1}x_N'\hat{P}_N^{-1}x_N$ wherein α, β have been replaced by their estimated values. Then as in [4], but more simply because of the condition on h,

$$(3) \qquad \hat{\sigma}_{p,q}^2 - \int_{-\pi}^{\pi} f_0(\omega)|\hat{k}(e^{i\omega})|^{-2} d\omega \to 0, \qquad \text{a.s.}$$

Here $f_0 = \sigma^2|k_0|^2/(2\pi)$ is the true spectral density. If $p < p_0$ or $q < q_0$ then the right side of (3) is strictly greater than σ^2, uniformly in $\hat{k}$, since the contrary could hold only if there were a sequence of $\hat{k}$, for such p, q, such that $f_0|\hat{k}|^{-2}$ converged to $\sigma^2/(2\pi)$, almost everywhere, which is impossible. Thus, almost surely, eventually $\hat{p} \geqslant p_0, \hat{q} \geqslant q_0$. Henceforth we consider only that case.

Let θ be the vector of coefficients in the polynomial $(gh_0 - hg_0)$, which is of degree $r = \max(p + q_0, q + p_0)$ and has zero constant term. Then, putting α, β for the vectors comprised of the $\alpha(j), \beta(j)$,

$$\theta = A\begin{pmatrix} \alpha - \alpha_0 \\ \beta - \beta_0 \end{pmatrix}$$

where A is a function only of α_0, β_0. The matrix A is of rank r. To see this observe that $Ax = 0$ implies that if the components of x are used to define transfer functions g_x, h_x then $g_x h_0 - h_x g_0 = 0$. But then $g_x = g_1 g_0, h_x = g_1 h_0$ where g_1 is a polynomial with $g_1(0) = 1$. But the degree of g_1 cannot exceed $s = \min(p - p_0, q - q_0)$ so that the space of vectors x satisfying $Ax = 0$ is s dimensional at most. Since A is $r \times (r + s)$, because $r + s = p + q$, then A is of rank r. Now we may find s further linearly independent, linear combinations of α, β, linearly independent of the elements of θ. Call these ψ. We may then use θ, ψ as new parameters. At $\theta = 0$ we have $gh_0 = hg_0$ i.e., $k = g^{-1}h = k_0 = g_0^{-1}h_0$. Because h_0, g_0 are prime polynomials it must also be true that, at $\theta = 0, h = h_0 g_1, g = g_0 g_1$, where g_1, of degree s, is a function of ψ. Of course g_1 must have all zeros bounded away from the unit circle because this is true of h.

Now it is shown in [4] that, in any case $\hat{k} \to k_0$, a.s., uniformly on every closed subset of the open unit disc, so that $\hat{\theta} \to 0$, a.s. Nothing of this kind can be said about $\hat{\psi}$. It will now be convenient to write $\hat{\sigma}^2_{p,q}(\hat{\theta}, \hat{\psi})$ for the optimised quantity. Because $\theta' = (\alpha' - \alpha_0', \beta' - \beta_0')A'$ then it is clear that $\theta = 0$ is an interior point of the region over which θ varies, since (α_0, β_0) is an interior point of the region over which (α, β) varies. Since $\hat{\theta} \to 0$ then $\hat{\theta}$ is an interior point, also, for sufficiently large N. In our developments below we therefore assume that $\hat{\theta}$ is interior.

$$\hat{\sigma}^2_{p,q}(\hat{\theta}, \hat{\psi}) = \hat{\sigma}^2 + \hat{\theta}' \frac{\partial \hat{\sigma}^2(0, \hat{\psi})}{\partial \theta} + \tfrac{1}{2}\hat{\theta}' \frac{\partial^2 \hat{\sigma}^2(\bar{\theta}, \hat{\psi})}{\partial \theta \partial \theta'} \hat{\theta}. \tag{4}$$

Here $\hat{\sigma}^2 = \hat{\sigma}^2_{p,q}(0, \hat{\psi}) = \hat{\sigma}^2_{p_0 q_0}(0)$, the latter following from the fact that $\hat{P}_N$ is now being evaluated at k_0. It will be sufficient to show that

$$\lim_{N\to\infty}\left\{\left(\hat{\sigma}^2_{p,q} - \hat{\sigma}^2\right)\sigma^{-2} + (p - p_0 + q - q_0)c \log\log N/N\right\} > 0, \qquad \text{a.s.,}$$
$$p > p_0, q > q_0.$$

Indeed if this is so the same will certainly be true when $\hat{\sigma}^2$ is replaced by $\hat{\sigma}^2_{p_0 q_0}$ because, to $0(N^{-1}), \hat{\sigma}^2 - \hat{\sigma}^2_{p_0, q_0} \geqslant 0$ because $\hat{\sigma}^2_{p_0 q_0}$ has been obtained after minimisation of $-2N^{-1}$ by the log likelihood (and the omitted term is $0(N^{-1})$). Since it will follow that $\log(\hat{\sigma}^2_{p,q}/\hat{\sigma}^2_{p_0, q_0}) = (\hat{\sigma}^2_{pq} - \hat{\sigma}^2_{p_0 q_0})/\hat{\sigma}^2_{p_0 q_0} + o(N^{-1})$ and $\hat{\sigma}^2_{p_0 q_0} \to \sigma^2$, a.s., we need consider only $(\hat{\sigma}^2_{p,q} - \hat{\sigma}^2)$. Since $\bar{\theta}$ is nearer to 0 than $\hat{\theta}$ then $\partial^2 \hat{\sigma}^2(\bar{\theta}, \hat{\psi})/\partial \theta \partial \theta' - \partial^2 \hat{\sigma}^2(0, \hat{\psi})/\partial \theta \partial \theta'$ converges to zero.

It may be shown that

$$\frac{\partial^2 \hat{\sigma}^2(0, \hat{\psi})}{\partial \theta \partial \theta'} = 2\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sigma^2}{|g_0 h_0 \hat{g}_1|^2} e^{i(j-k)\omega}\, d\omega\right]_{j,k=1,\cdots,r} \{1 + o(1)\} \tag{5}$$

where the quantity that is $o(1)$ is of that order a.s. and uniformly in $\hat{\psi}$, while $\hat{g}_1$ is g_1

evaluated at $\hat{\psi}$. Indeed from [4] it follows that

$$\frac{\partial^2 \hat{\sigma}^2(0,\hat{\psi})}{\partial\theta\partial\theta'} = \int_{-\pi}^{\pi} f_0(\omega) \frac{\partial^2 |k|^2}{\partial\theta\partial\theta'}\bigg|_{0,\hat{\psi}} d\omega\{1 + o(1)\}.$$

Remembering that, at $\theta = 0, k = k_0$ and that $f_0 = \sigma^2|k_0|^2/(2\pi)$ the result (5) follows by elementary, albeit tedious, manipulations.

On the other hand, expanding about $\hat{\theta}$,

$$\frac{\partial\hat{\sigma}^2_{pq}(0,\hat{\psi})}{\partial\theta} = -\frac{\partial^2\hat{\sigma}^2_{pq}(\bar{\theta},\hat{\psi})}{\partial\theta\partial\theta'}\hat{\theta} + 0(N^{-1}), \tag{6}$$

where the last term is of the indicated order almost surely and uniformly in $\hat{\psi}$. Put $\hat{b}$ for the left side of (6) and $\hat{B}$ for the right side of (5), ignoring the factors 2 and $\{1 + o(1)\}$. Then from (4), (5), (6)

$$\hat{\sigma}^2_{p,q}(\hat{\theta},\hat{\psi}) - \hat{\sigma}^2 = -\tfrac{1}{4}\hat{b}'\hat{B}^{-1}\hat{b}\cdot\{1 + o(1)\} + 0(N^{-1}). \tag{7}$$

We shall hold p, q fixed for the moment and hence have omitted these from the notation on the right. Of course $\hat{b}, \hat{B}$ are functions only of $\hat{\psi}$ and not of $\hat{\theta}$. Now

$$\tfrac{1}{2}\hat{b} = \frac{1}{2\pi}\int_{-\pi}^{\pi} I_\varepsilon(\omega)\frac{1}{g_0 h_0 \hat{g}_1}\{e^{ij\omega}\}_{j=1,\dots,r} d\omega + o(N^{-\frac{1}{2}}) \tag{8}$$

where again the last term is of that order a.s. and uniformly in $\hat{\psi}$ and

$$I_\varepsilon(\omega) = \frac{1}{N}|\Sigma_1^N \varepsilon(n)e^{in\omega}|^2.$$

The equation (8) may be established by first showing that

$$\tfrac{1}{2}\hat{b} = \tfrac{1}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} I_x(\omega)\frac{\partial}{\partial\theta}|k|^{-2}|_{0,\hat{\psi}} d\omega + o(N^{-\frac{1}{2}}), I_x(\omega) = \frac{1}{N}|\Sigma_1^N x(n)e^{in\omega}|^2.$$

Again the proof is given, essentially in [4]. Since $\frac{1}{2}\partial|k|^{-2}/\partial\theta$, evaluated at $0, \hat{\psi}$ has, as jth component, the real part of $|k_0|^{-2}(g_0 h_0 \hat{g}_1)^{-1}e^{ij\omega}$, (8) is obtained by showing that $I_x(\omega)|k_0|^{-2}$ may be replaced by $I_\varepsilon(\omega)$. This may be done by using the formulae in [5] page 246, which shows that

$$\frac{1}{N^{\frac{1}{2}}}\Sigma_1^N x(n)e^{in\omega} = \left\{\frac{1}{N^{\frac{1}{2}}}\Sigma_1^N\varepsilon(n)e^{in\omega}\right\}k_0(e^{i\omega}) + \frac{1}{N^{\frac{1}{2}}}\Sigma_0^\infty \kappa_0(j)e^{ij\omega}R_{jN}(\omega)$$

$$\begin{aligned} R_{jN}(\omega) &= \Sigma_{-j+1}^0 \varepsilon(n)e^{in\omega} - \Sigma_{N-j+1}^N \varepsilon(n)e^{in\omega}, && 0 \leqslant j \leqslant N \\ &= \Sigma_{-N+1}^0 \varepsilon(n)e^{in\omega} - \Sigma_1^N \varepsilon(n)e^{in\omega}, && N \leqslant j. \end{aligned}$$

The result now follows. For example

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\overline{\frac{1}{N^{\frac{1}{2}}}\Sigma_1^N\varepsilon(n)e^{in\omega}}\right\}\left\{\frac{1}{N^{\frac{1}{2}}}\Sigma_0^\infty\kappa_0(j)e^{ij\omega}R_{jN}(\omega)\right\}k_0^{-1}(g_0 g_0 \hat{g}_1)^{-1}d\omega$$

$$= \Sigma_0^\infty d_k \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\overline{\frac{1}{N^{\frac{1}{2}}}\Sigma_1^N\varepsilon(n)e^{in\omega}}\right\}\left\{\frac{1}{N^{\frac{1}{2}}}\Sigma_0^\infty\kappa_0(j)e^{ij\omega}R_{j\omega}(\omega)\right\}e^{ik\omega}d\omega$$

when the d_k are the Fourier coefficients of $k_0^{-1}(g_0h_0\hat{g}_1)^{-1}$

$$= \Sigma_0^\infty\Sigma_0^\infty d_\kappa\kappa_0(j)\frac{1}{N}\frac{1}{2\pi}\int_{-\pi}^{\pi}\{\Sigma_1^N\varepsilon(n)e^{in\omega}\}\{R_{jN}(\omega)e^{ij\omega}\}e^{ik\omega}\,d\omega.$$

The integrated expression contains $\min(j,N)$ terms each of which is $o(N^{\frac{1}{2}})$, uniformly in j, k, because of the fourth moment condition. Since $\Sigma\Sigma|d_k\kappa_0(j)|\cdot\min(j,N)$ converges the assertion follows. All other terms constituting the term that is $o(N^{-\frac{1}{2}})$ may be treated in much the same way.

Now

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}I_\varepsilon(\omega)\frac{1}{g_0h_0\hat{g}_1}e^{ij\omega}\,d\omega = \Sigma_0^{N-j-1}\gamma_k(\hat{\psi})c(k+j)$$

(9) $$c(k) = \frac{1}{N}\Sigma_{k+1}^N\varepsilon(n)\varepsilon(n-k),$$

where $\gamma_j(\hat{\psi})$ are the coefficients in the expression of $(g_0h_0\hat{g}_1)^{-1}$ and thus converge to zero at a geometric rate uniformly in $\hat{\psi}$. The expression (9) may also be written as

(10) $$\frac{1}{N}\Sigma_1^N\varepsilon(n)\hat{\zeta}(n-j) + 0(N^{-1}),\ \hat{\zeta}(n) = \Sigma_0^\infty\gamma_k(\hat{\psi})\varepsilon(n-k).$$

Indeed the error is

$$\frac{1}{N}\Sigma_1^j\varepsilon(n)\hat{\zeta}(n-j) + \frac{1}{N}\Sigma_{j+1}^N\varepsilon(n)\Sigma_{n-j}^\infty\gamma_k(\hat{\psi})\varepsilon(n-k-j)$$

$$= \frac{1}{N}\Sigma_1^j\varepsilon(n)\hat{\zeta}(n-j) + \frac{1}{N}\Sigma_{j+1}^N\varepsilon(n)\Sigma_{m=0}^\infty\gamma_{n-j+m}(\hat{\psi})\varepsilon(-m).$$

Since $|\gamma_k(\hat{\psi})| \leqslant \rho^k, |\rho| < 1$ the second term is dominated by

$$N^{-1}\Sigma_{j+1}^N|\varepsilon(n)\rho^n|\Sigma_0^\infty\rho^m|\varepsilon(-m)|\rho^{-j}$$

which is clearly $0(N^{-1})$ a.s.

For any fixed ψ it follows that

(11) $$\limsup\{-\hat{\sigma}^2(\hat{\theta},\psi) + \hat{\sigma}^2\}\frac{N}{2\sigma^2\log\log N} = 1 \quad \text{a.s.}$$

This follows from the observation, first, that the first term in (10) is, for fixed ψ, N^{-1} by a martingale with stationary, square integrable, ergodic martingale differences. Hence, [8], it obeys the law of the iterated logarithm. The matrix $\hat{B}$ is, for ψ fixed, just the covariance matrix of the expressions in the first term of (10), neglecting a factor σ^2N^{-1}. Thus (11) is of the form

$$\Sigma_1^r z_j(N)^2, z_j(N) = \Sigma_1^N\{y_j(n)\}/\{2N\log\log N)^{\frac{1}{2}}\},\quad j = 1,\cdots,r$$

where the $y_j(n)$ constitute a vector of stationary, ergodic, square integrable martingale differences, with unit covariance matrix. Then $\limsup z_j(N)^2 = 1, j = 1,\cdots,r$. Consider then r-vectors $\alpha(u), u = 1,\cdots,M$, with $\Sigma\alpha_j(u)^2 \equiv 1$, chosen so that they are uniformly spread over the unit sphere. We may choose M so that any

r-vector, z, has an angle with some $\alpha(u)$ whose cosine is not less than $(1-\varepsilon)$, no matter how small ε may be. Then, if $\theta_N(u)$ is the angle between the vector with components $z_j(N)$ and $\alpha(u)$, and $\tilde{u}$ makes $\{\cos\theta_N(u)\}^2$ smallest,

$$\limsup \Sigma_1^r z_j(N)^2 = \limsup \frac{\{\Sigma\alpha_j,(\tilde{u})z_j(N)\}^2}{\{\cos\theta_N(\tilde{u})\}^2} \leqslant \limsup \frac{\max_u\{\Sigma\alpha_j(u)z_j(N)\}^2}{(1-\varepsilon)^2}$$
$$= \frac{1}{(1-\varepsilon)^2}$$

since evidently

$$\limsup \max_u\{\Sigma\alpha_j(u)z_j(N)\}^2 = 1$$

because $\Sigma\alpha_j(u)y_j(n)$ has the same properties as $y_n(n)$ and there are only finitely many $\alpha(u)$. Thus $\limsup \Sigma z_j(N)^2 = 1$, a.s.

We shall now show that (11) *also holds when* ψ *is allowed to vary*. We shall do this by showing that

$$\{\Sigma_0^{N-j-1}\gamma_k(\psi)c(k+j)\}\left\{\frac{N}{2\log\log N}\right\}^{\frac{1}{2}}$$

is continuous in ψ, uniformly in N (i.e., is equicontinuous). Since $\hat{B}$ clearly has this property the same will then be true of $\chi_N(\hat{\psi}) = [N/\{2\log\log N\}]\hat{b}'\hat{B}^{-1}\hat{b}$. Now choosing $\psi(l), l = 1,\cdots,L$, arbitrarily dense in the compact space within which $\hat{\psi}$ lies and observing that there is an $\hat{l}$ so that $|\hat{\psi}-\psi(\hat{l})| < \delta$, where δ may be made arbitrarily small by choosing L large, it will follow that

$$\max\{\chi_N(\hat{\psi}) - \chi_N(\psi(\hat{l})\} \leqslant c\delta.$$

Since

$$\limsup \max_l \chi_N\{\psi(l)\} = 1.$$

The theorem will then hold.

To show equicontinuity, we first show that there is a finite constant d so that, putting $d(N) = d\log\log N$ then

(12) $$\lim_{N\to\infty}\max_\psi|\Sigma_{d(N)}^{N-j+1}\{\gamma_k(\psi)c(k+j)\}\{N/(2\log\log N)\}^{\frac{1}{2}}| = 0, \qquad \text{a.s.}$$

For this it is sufficient to show, as $M\to\infty$, and using c here and below for a finite constant, not always the same one, that

(13) $$\mathcal{E}\{\max_\psi \max_{c<N\leqslant M}|\Sigma_{d(N)}^{N-j-1}\{\gamma_k(\psi)c(k+j)\}\{N/(2\log\log N)\}^{\frac{1}{2}}|\}$$

may be made arbitrarily small by choosing c large. However $|\gamma_k(\psi)| \leqslant \rho^k$, $\rho < 1$, uniformly in ψ so that for d sufficiently large, there is a $\rho_0, \rho < \rho_0 < 1$, for which

$$\rho_0^k(\log N)^{-\delta/2} > \rho^k, \delta > 1, k > d(N).$$

Thus (13) is bounded by

(14) $$\mathcal{E}\left[\max_{c<N\leqslant M}\Sigma_{d(N)}^{N-j-1}\{\rho_0^k N^{-\frac{1}{2}}(\log N)^{-\delta/2}|\Sigma_{k+j+1}^N\varepsilon(n)\varepsilon(n-j-k)|\}\right].$$

Choosing c so that $d(N) > A, N > c$ and using the easily established inequality

$$\max_{1 \leqslant k \leqslant n} \frac{1}{b_n} |\Sigma_1^k x_j| \leqslant 2 \max_{1 \leqslant k \leqslant n} \left| \Sigma_1^k \frac{x_j}{b_j} \right|,$$

$$0 < b_1 \leqslant \cdots \leqslant b_n,$$

we bound (14) by

$$2\,\Sigma_A^\infty \rho_0^k \mathcal{E}\left[\max_{c \leqslant N \leqslant M} \left| \Sigma_{j+k+1}^N \frac{\varepsilon(n)\varepsilon(n-k-j)}{\{n(\log n)^\delta\}^{\frac{1}{2}}} \right| \right].$$

Using Doob's inequality ([11], page 68, Proposition IV-2-8) we obtain the bound

$$4\Sigma_A^\infty \rho_0^k \left\{ \Sigma_{j+k+1}^M \frac{\sigma^4}{n(\log n)^\delta} \right\}^{\frac{1}{2}} < c\Sigma_A^\infty \rho_0^k.$$

The result (12) now follows. The proof of equicontinuity of $\chi_N(\psi)$ now may be established by showing that

$$\text{(15)} \qquad \limsup_{N\to\infty} \max_{k \leqslant d(N)} \{k^{-1}|c(k)\{N/(2\log\log N)\}^{\frac{1}{2}}|\} \leqslant c.$$

For this purpose, *and for the first time* we use the independence of the $\varepsilon(n)$. Then $Nc(k)$ can be decomposed as $(k+1)$ sums, $S_N(k,j), j = 0, \cdots, k$, of independent random variables where

$$S_N(k,j) = \Sigma_m \varepsilon\{(k+1)m + j\}\varepsilon\{(k+1)m + j - k\}.$$

Now almost precisely as in [9], Section 8.16, the quantities $\varepsilon(n)\varepsilon(n-k), k = 1, \cdots, d(N), n = 1, 2, \cdots$ may be truncated so that the truncated quantities are $o\{(n\log\log n)^{-\frac{1}{2}}\}$, uniformly in k, and so that the contribution of the truncation errors to $T_N(k,j) = S_N(k,j)/\{2N\log\log N\}^{\frac{1}{2}}, k \leqslant d(N), 1 \leqslant j \leqslant k+1$, will converge almost surely to zero, uniformly in j, k. Thus, assuming such a truncation having been effected on the summands in $S_N(k,j)$, we must establish the convergence, for $c > 1, \delta > 0$, of

$$\text{(16)} \qquad \Sigma_{u=1}^\infty P\{\max_{1 \leqslant N \leqslant c^u} \max_k \max_j T_N(k,j) > 1 + \delta\}.$$

For then, by the Borel-Cantelli lemma $T_N(k,j)$ has limit superior unity, uniformly in k,j and since the limit inferior of -1 will be established in the same way by considering $-S_N(k,j)$ while $c(k)\{N/(2\log\log N)\}^{\frac{1}{2}}$ is composed of $(k+1)$ such bounded quantities then (15) and, the first part of the theorem, will result. However (16) is bounded by

$$\Sigma_{u=1}^\infty \Sigma_j \Sigma_k P\{\max_{1 \leqslant N \leqslant c^u} T_N(k,j) > 1 + \delta\},$$

and remembering that the $\varepsilon(n)\varepsilon(n-k)$ have the same distribution for all $n, k > 1$, then using the classical result in [9], pages 376–379, we see that this is bounded by

$$c\Sigma_{u=1}^\infty \frac{(\log u)^2}{u^{1+\varepsilon}} < \infty.$$

The second part of Theorem 1 is established by following the proof just given, down to the point where

$$\max_\psi\left[\Sigma_1^{d(N)}\gamma_k(\psi)c(k+j)\left(\frac{N}{\log N}\right)^{\frac{1}{2}}\right]$$

must be shown to converge almost surely to zero. It is now sufficient to show that

$$\max_{1\leqslant k\leqslant d(N)}\left\{c(k+j)\left(\frac{N}{\log N}\right)^{\frac{1}{2}}\right\} \tag{17}$$

converges almost surely to zero. The proof may now be completed using extensions of Menshov's inequality and the method of subsequences as in [10]. Indeed $\varepsilon(n)\varepsilon(n-j)$ has a moment of order $a>2$, by assumption and consequently as is shown in the reference cited,

$$\mathcal{E}\{\max_{j+1\leqslant N\leqslant M}|Nc(k+j)|^a\}\leqslant CN^{a/2}. \tag{18}$$

Hence, by Markov's inequality,

$$P\{\max_{1\leqslant N\leqslant M}\max_{k\leqslant d(N)}|c(k+j)|\geqslant\lambda(M)\}\leqslant\frac{C\log\log M}{\log M(\log\log M)^{1+\delta}}\leqslant\frac{C}{\log M}$$

if $\lambda(n)=n^{\frac{1}{2}}(\log n)^{1/a}(\log\log N)^{(1+\delta)/a}$. The proof may now be completed as in the last reference cited by showing that $N^{-\frac{1}{2}}(\log N)^{-1/a}\max_{1\leqslant k\leqslant d(N)}|c(k+j)|\to 0$, a.s., which evidently completes the result since $a>2$.

Proof of Theorem 2. When $q=0$ the limits

$$\lim_{N\to\infty}P\{\hat{p}=p\},p_0\leqslant p\leqslant P$$

were evaluated in [12]. Of course the limits are zero for $p<p_0$. It was shown above that, when $P=p_0 (Q=q_0$ is the same) then again we need consider only $P\{\hat{q}=q,\hat{p}=p_0\},q_0\leqslant q\leqslant Q$. However now $r=q+p_0,s=0$ and it follows from (5), (7), (8) that $(\hat{\sigma}^2_{p_0,q}-\hat{\sigma}^2)$ is of exactly the same form, to $o(N^{-\frac{1}{2}})$, as if an autoregression of order $r=p_0+q$ is being fitted when the truth is an autoregression of order p_0+q_0. The term $(\hat{\sigma}^2_{p_0,q_0}-\hat{\sigma}^2)$ can be expressed in the same way, with r now equal to (p_0+q_0), again to $o(N^{-\frac{1}{2}})$. Thus to this order of magnitude the limits

$$\lim_{N\to\infty}P\{\hat{q}=q,\hat{p}=p_0\},q_0\leqslant q\leqslant Q$$

are precisely the same as those obtained by Shibata, in [14], for a true autoregression of order p_0+q_0 when the fitted order is p_0+q. The result then follows. (Though Shibata's treatment is for the Gaussian case this relates only to the asymptotic distribution of the partial autocorrelations, which is the same under the present circumstances. See [6].)

Proof for theorem. We have now only to show that

$$\max_\psi\{|\Sigma_k\gamma_k(\psi)S_N(k)/\{NC_N\}^{\frac{1}{2}}|,S_N(k)=Nc(k+j),$$

converges in probability to zero. However this is no greater than

$$\Sigma_k \rho^k |S_N(k)| / \{NC_N\}^{\frac{1}{2}}$$

whose root mean square is dominated by

$$\Sigma_k \rho^k / C_N^{\frac{1}{2}} \to 0.$$

3. Some comments. The proof of the first part of Theorem 1 could be completed without the requirement of independence for the $\varepsilon(n)$ but with the higher moment condition of the second part of the theorem if the Skorohod representation theorem was available for a triangular array of martingale differences, the proof following the lines of [8]. The martingale difference array would be composed of $X_{n,k} = \varepsilon(n)\varepsilon(n-k), n \geqslant k+1$, where $k = 1, 2, \cdots$ indexes the rows of the array. However that construction has not so far been completed.

We have maintained the second part of (2) throughout the paper for uniformity. Of course it is necessarily satisfied if the $\varepsilon(n)$ are i.i.d. and so is not used in the first part of Theorem 1. However it would be necessary for that result, if the proof were to be completed without the use of independence along the lines of the previous paragraph, because without the second part of (2) the vector $\frac{1}{2}\hat{b}$, for $\hat{\psi}$ fixed, does not have $N^{-1}\hat{B}$ as covariance matrix. In the same way Theorem 2 would not hold. *However Theorem* 3 *and the second part of Theorem* 1 *are not affected by the elimination of the second part of* (2). This is obvious for Theorem 3. For Theorem 1 the proof down to (17) is not altered. The fact that (18) holds follows from Doob's inequality and the inequality in [3], Theorem 9. The remainder of the proof is independent of the second part of (2).

REFERENCES

[1] Akaike, H. (1969). Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.* **21** 243–247.

[2] Akaike, H. (1977). On entropy maximisation principle. In *Applications of Statistics* (P. R. Krishnaiah, ed.) 27–41. North Holland, Amsterdam.

[3] Burkholder, D. L. (1966). Martingale transforms. *Ann. Math. Statist.* **37** 1494–1504.

[4] Dunsmuir, W. T. M. and Hannan, E. J. (1976). Vector linear time series models. *Adv. Appl. Prob.* **8** 339–364.

[5] Hannan, E. J. (1970). *Multiple Time Series*. Wiley, New York.

[6] Hannan, E. J. and Heyde, C. C. (1972). On limit theorems for quadratic functions of discrete time series. *Ann. Math. Statist.* **43** 2058–2066.

[7] Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *J. Roy. Statist. Soc. Ser. B.* **41** 190–195.

[8] Heyde, C. C. and Scott, D. J. (1978). Invariance principles for the law of the iterated logarithm for martingales and processes with stationary increments. *Ann. Probability* **1** 428–436.

[9] Moran, P. A. P. (1968). *An Introduction to Probability Theory*. Oxford.

[10] Moricz, F. (1976). Moment inequalities and the strong law of large numbers. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35** 298–314.

[11] Neveu, J. (1975). *Discrete Parameter Martingales*. North Holland, Amsterdam.

[12] Rissanen, J. (1978). Modeling by shortest data description. *Automatica* **14** 465–471.

[13] SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.

[14] SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63** 117–126.

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
MATHEMATICAL SCIENCES BUILDING
P.O. BOX 4
CANBERRA, A.C.T. 2600
AUSTRALIA