The Annals of Statistics
1982, Vol. 10, No. 3, 926–936

AUTOCORRELATION, AUTOREGRESSION AND AUTOREGRESSIVE APPROXIMATION

BY AN HONG-ZHI, CHEN ZHAO-GUO AND E. J. HANNAN[1]

The Australian National University

Theorems are proved relating to the rate of almost sure convergence of autocovariances, and hence autocorrelations, to their true values. These rates are uniform in the lag up to some order $P(T)$, increasing with T. The key assumption is that the process is stationary and the best linear predictor is the best predictor. In particular for an ARMA process and $P(T) = O\{(\ln T)^a\}$, $a < \infty$, the rate is $O\{(\ln \ln T/T)^{1/2}\}$. These results are used to discuss autoregressions and the use of autoregressions to approximate the structure of a more general process by increasing the order of the autoregression with T.

1. Introduction. Consider observations, $x(t)$, $t = 1, \cdots, T$, on a stationary ergodic process with zero mean and finite variance. We mention here that mean correction of the data will have no effect on the results presented below. It is always assumed that

$$x(t) = \sum_0^\infty \kappa(j)\varepsilon(t-j), \qquad \sum_0^\infty |\kappa(j)| < \infty, \qquad \kappa(0) = 1, \tag{1}$$

where the $\varepsilon(t)$ are the linear innovations, so that, (Hannan, 1970, page 142),

$$k(z) = \sum_0^\infty \kappa(j)z^j \neq 0, \qquad |z| < 1. \tag{2}$$

It is also assumed that, for some $r \geq 4$,

$$E\{\varepsilon(t) \mid \mathscr{F}_{t-1}\} = 0, \qquad E\{\varepsilon(t)^2 \mid \mathscr{F}_{t-1}\} = \sigma^2, \qquad E\{|\varepsilon(t)|^r\} < \infty. \tag{3}$$

Here $\mathscr{F}_t$ is the σ-algebra generated by $\varepsilon(s)$, $s \leq t$, and σ^2 is the prediction variance. More specifically we may require that

$$k(z) = g(z)^{-1}h(z), \qquad g(z) = \sum_0^p \alpha(j)z^j, \qquad h(z) = \sum_0^q \beta(j)z^j, \tag{4}$$

where $g(z)$ and $h(z)$ are relatively prime. Then $x(t)$ is generated by an ARMA process. In case $q = 0$, $x(t)$ is an autoregressive process. We note that one of these might be used as a model for a statistical procedure even when the model is not valid. A zero subscript will be used to distinguish true values from hypothetical values when the distinction is needed, for example $\alpha_o(j)$, $\beta_0(j)$, p_0, q_0. Since we are, in part, concerned with autoregressive approximation we shall sometimes strengthen (2) so that

$$k(z) \neq 0, \qquad |z| \leq 1, \tag{2'}$$

and consequently $|k(\exp i\omega)|$ is bounded away from zero uniformly in ω.

The theorems presented below are based on results relating to the sample autocovariances,

$$c(t) = \begin{cases} \dfrac{1}{T}\sum_1^{T-t} x(s)x(s+t), & 0 \leq t \leq T-1, \\ 0, & T \leq t. \end{cases}$$

Received May 1981; revised January 1982.

[1] The first two authors were visiting the Australian National University on leave from the Academy of Science, Beijing, China.

AMS 1979 *subject classifications.* Primary 62M10; secondary 60F15.

Key words and phrases. Stationary process, autoregression, autocovariance, law of the iterated logarithm, strong convergence, autoregressive approximation, martingale, method of subsequences.

Let $\gamma(t) = \mathscr{E}\{x(s)x(s+t)\}$. In Hannan (1974) it is shown, under very general conditions, that

$$\lim_{T\to\infty}\sup_{0\le t<\infty} | c(t) - \gamma(t) | = 0, \quad \text{a.s.}$$

We prove stronger results below. In the following, for a sequence y_T, $T = 1, 2, \cdots$, of random variables and a sequence n_T of constants we shall write $y_T = O(n_T)$ to mean that

$$\limsup_{T\to\infty} | y_T |/n_T < \infty, \quad \text{a.s.}$$

If "$<\infty$" is replaced by "$=0$" we shall write $y_T = o(n_T)$. It will also be convenient to put $Q_T = (\ln\ln T/T)^{1/2}$.

THEOREM 1. *Let $x(t)$ satisfy* (1), *and* (3), *for $r \ge 4$. Then for any $\delta > 0$ and $P(T) \le T^a$, $a = r/\{2(r-2)\}$,*

$$\max_{0\le t\le P(T)} | c(t) - \gamma(t) | = o[T^{-1/2}\{P(T)\ln T\}^{2/r}(\ln\ln T)^{2(1+\delta)/r}].$$

Of course this theorem is vacuous for $r = 4$ unless $P(T) = o[T/\{\ln T(\ln\ln T)^{(1+\delta)/2}\}]$. It is to be expected that, under reasonable conditions, the quantities $c(t) - \gamma(t)$ will, individually, follow the law of the iterated logarithm (see Hall and Heyde, 1980, pages 141 and 194). Thus one expects to be able to improve on the order given in Theorem 1, as in the following result for ARMA processes.

THEOREM 2. *Let $x(t)$ satisfy* (1), (3), (4) *for $r = 4$. Then if $P(T) = O\{(\ln T)^a\}$, for some $a < \infty$.*

$$\max_{0\le t\le P(T)} | c(t) - \gamma(t) | = O(Q_T).$$

Finally we wish to give a strengthening of the result in Hannan (1974), mentioned above.

THEOREM 3. *Let $x(t)$ satisfy* (1), *and* (3) *for $r = 4$. If*

(5)
$$\lim_{T\to\infty} T^{1/2} \sum_T^\infty | \kappa(j) | = 0$$

then

$$\sup_{0\le t<\infty} | c(t) - \gamma(t) | = O\{(\ln T/T)^{1/2}\}$$

2. Approximation by autoregressions. Consider the estimation of an autoregression when the true order, p_0, is not known. As in Akaike (1974) and Rissanen (1978), we can estimate p_0 as the $\hat{p}$ which minimizes

(6)
$$\ln\hat{\sigma}_p^2 + p\ln T/T, \qquad 0 \le p \le P(T).$$

Here $\hat{\sigma}_p^2$ is the estimate of σ^2 obtained from fitting an autoregression of order p (see Akaike, 1977; Hannan, 1970, Chapter VI). We can apply the results in Section 1 and have the following theorem.

THEOREM 4. *If $x(t)$ is generated by an autoregression and* (3) *holds for $r = 4$ then, for $P(T) \uparrow \infty$ and $P(T) = O(\ln T)$ in* (6), $\hat{p} \to p_0$ a.s.

This theorem also holds true if $P(T) = O(\ln\ln T)$ and $\ln T$ in (6) is replaced by $c\ln\ln T$ for sufficiently large constant $c > 0$.

Having fitted an autoregression of order p, whether or not the true process is an autoregression, the spectrum $f(\omega) = \sigma^2 | \sum \kappa(j)\exp ij\omega |^2/(2\pi)$ may be estimated by

$$\hat{f}_p(\omega) = \frac{\hat{\sigma}_p^2}{2\pi} | \textstyle\sum_0^p \hat{\alpha}_p(j)e^{ij\omega} |^{-2},$$

where the $\hat{\alpha}_p(j)$ and $\hat{\sigma}_p^2$ are obtained from the Yule-Walker procedure (Hannan, 1970). If p depends on T, i.e. $p = p(T)$, we shall write $\hat{\sigma}_T^2$, $\hat{\alpha}_T(j)$ $\hat{f}_T(\omega)$, for brevity. In Berk (1974) it is shown that $\hat{f}_T(\omega)$ converges in probability to $f(\omega)$ when $p(T) = o(T^{1/3})$, the $\varepsilon(t)$ are independent with zero mean and variance σ^2 and (1), (2′), (5) hold. In the following theorem we establish almost sure convergence.

THEOREM 5. *If $x(t)$ satisfies* (1), (2′), (3) *for $r = 4$ and* (5) *then for $p(T) \uparrow \infty$, $p(T) = o\{(T/\ln T)^{1/2}\}$*

$$\sup_\omega |\hat{f}_T(\omega) - f(\omega)| = o(1). \tag{7}$$

If $p(T) = O[T^{1/3}/\{\ln T(\ln\ln T)^{1+\delta}\}]$, for some $\delta > 0$, and (1), (2′), (3) *for $r = 4$, then again* (7) *holds.*

Let the expression

$$\int_{-\pi}^{\pi} f(\omega)\,|\textstyle\sum_0^{p(T)} a(j)e^{ij\omega}|^2\,d\omega, \qquad a(0) = 1$$

be minimized at $a(j) = \alpha_T(j)$, $\alpha_T(0) = 1$, and let σ_T^2 be the minimum value. We finally state a theorem about the uniform convergence of the $\hat{\alpha}_T(j)$.

THEOREM 6. *If $x(t)$ satisfies* (1), (2′), (3) *for $r = 4$, and* (4) *and let $p(T) = O\{(\ln T)^a\}$, $1 < a < \infty$ or $p(T) = [d \ln T]$ for d sufficiently large,* then

$$\sup_{1\le j\le p(T)} |\alpha_T(j) - \hat{\alpha}_T(j)| = O(Q_T). \tag{8}$$

3. Some further observations. In this section we consider only Theorems 4 and 6 and thus maintain (1), (2′), (3) and (4). Then

$$\{\textstyle\sum_0^\infty \kappa(j)e^{ij\omega}\}^{-1} = \sum_0^\infty a(j)e^{ij\omega}, \qquad \sum_0^\infty j|a(j)| < \infty.$$

It is shown in Baxter (1962) that, under the conditions of Theorems 6, $\sigma_T^2 - \sigma^2$ decreases to zero at a geometric rate and, in general,

$$\textstyle\sum_0^{p(T)} |\sigma_T^2\alpha_T(j) - \sigma^2 a(j)| \le c \sum_{p(T)+1}^\infty |a(j)|, \tag{9}$$

so that $\max|\alpha_T(j) - a(j)|$ also converges geometrically to zero. (The rate of geometric convergence is determined by the zero of $h(z)$ nearest to the unit circle.) As a consequence, if $p(T) = d \ln T$ for d sufficiently large, $\alpha_T(j)$ may be replaced by $a(j)$ in (8). Of course this is also true if $p(T) = c(\ln T)^a$, $1 < a < \infty$. These considerations are of some importance in connection with the recursive methods for the estimation of (4), including p_0, q_0 in the estimation, discussed in Hannan and Rissanen (1982).

As observed in Hannan and Heyde (1972, page 2059) the first part of (3) is equivalent to the assertion that the best linear predictor is the best predictor, in the least squares sense. This is a natural condition on $x(t)$, since the final purpose of estimation is often linear prediction. On the other hand, it is difficult to find reasonable conditions on $x(t)$ that would make the second part of (3) hold, other than the requirement that $x(t)$ be Gaussian. For that reason it is of interest to see how far that part of (3) can be modified. The condition is fairly essential for Theorems 1, 2, 3 since the $c(t) - \gamma(t)$ involve

$$\frac{1}{T}\textstyle\sum_1^T \{\varepsilon(t)^2 - \sigma^2\}.$$

The Yule-Walker equations of estimation for the $\hat{\alpha}_T(j)$ are

$$\textstyle\sum_0^{p(T)} \hat{\alpha}_T(j)c(j-k) = 0, \qquad k = 1, \cdots, p(T); \qquad \hat{\alpha}_T(0) = 1.$$

Also

$$\begin{aligned}\textstyle\sum_0^{p(T)} a(j)c(j-k) &= T^{-1}\{\textstyle\sum_0^k a(j)\sum_1^{T-k+j} x(t)x(t+k-j)\\ &\quad+\textstyle\sum_{k+1}^{p(T)} a(j)\sum_1^{T-j+k} x(t)x(t+j-k)\}\\ &= T^{-1}[\textstyle\sum_0^k a(j)\{\sum_{k+1}^T x(t-j)x(t-k)+\sum_{T+1}^{T+j} x(t-j)x(t-k)\}\\ &\quad+\textstyle\sum_{k+1}^{p(T)} a(j)\{\sum_{k+1}^T x(t-j)x(t-k)+\sum_{T+1}^{T+k} x(t-j)x(t-k)\\ &\quad-\textstyle\sum_{k+1}^{j} x(t-j)x(t-k)\}]\\ &= T^{-1}[\textstyle\sum_0^k a(j)\{\sum_{k+1}^T x(t-j)x(t-k)+jo(T^{1/2})\}\\ &\quad+\textstyle\sum_{k+1}^{p(T)} a(j)\{\sum_{k+1}^T x(t-j)x(t-k)+ko(T^{1/2})+jO(p(T)^{1/2})\}].\end{aligned}$$

These follow from the fact that $x^2(t)/t^{1/2}$ converges a.s. to zero because $x(t)$ has a finite fourth moment. Thus

$$\textstyle\sum_0^{p(T)} a(j)c(j-k) = T^{-1}\sum_0^{p(T)} a(j)\sum_{k+1}^T x(t-j)x(t-k) + o(T^{-1/2}).$$

Also for Theorem 4 $a(j) = 0$, $j > p_0$, while under the conditions of Theorem 6

$$\textstyle\sum_{p(T)+1}^{\infty} a(j)x(t-j) \le \rho^{p(T)}\sum_0^{\infty} d(j)\,|\,x(t-j)\,|, \qquad d(j) > 0, \quad 0 < \rho < 1, \tag{10}$$

where the $d(j)$ also converge to zero at a geometric rate, as explained at the beginning of this section, so that for $p(T) \ge d \ln T$ and d sufficiently large (10) is $o(T^{-1/2})$. Thus, then, uniformly in $k = 1, \cdots, p(T)$,

$$\textstyle\sum_0^{p(T)} a(j)c(j-k) = T^{-1}\sum_{k+1}^T \varepsilon(t)x(t-k) + o(T^{-1/2})$$

and hence

$$\textstyle\sum_0^{p(T)} \{\hat{a}_T(j) - a(j)\}c(j-k) = -T^{-1}\sum_{k+1}^T \varepsilon(t)x(t-k) + o(T^{-1/2}),\ k = 1, \cdots, p(T). \tag{11}$$

Thus, in Theorems 4 and 6 we are reduced to the consideration of (11). In the first term on the right side of (11), $\varepsilon(t)^2$ does not occur and for this reason we may show, under the conditions of those two theorems, but without the middle part of (3), that

$$\sup_{1\le k\le p(T)}\,|\,T^{-1}\textstyle\sum_{k+1}^T \varepsilon(t)x(t-k)\,| = O(Q_T).$$

The proof is virtually the same as that given for Theorem 2, in the next section.

Let us, in Theorems 4 and 6, replace the middle part of (3) by

$$\frac{1}{T}\textstyle\sum_1^T \{\varepsilon(t)^2 - \sigma^2\} = o\{p(T)^{-2}\}. \tag{12}$$

We shall discuss this later. Then as in the proof of Theorem 2 we may show that

$$\max_{1\le t\le p(T)}\,|\,c(t) - \gamma(t)\,| = o\{p(T)^{-2}\}. \tag{13}$$

Let $\hat{\Gamma}_T$ be the matrix with $c(j-k)$ in the (j, k)th place, $j, k = 1, \cdots, p(T)$, and let Γ_T be similarly composed of the $\gamma(j-k)$. If x is a vector of unit length with $p(T)$ components and (2′) holds then

$$x'\Gamma_T x = \int f(\omega)\,|\textstyle\sum x_j e^{ij\omega}|^2\, d\omega \ge \inf f(\omega) > 0$$

so that the smallest eigenvalue of Γ_T is bounded away from zero. The sum of the moduli of the elements in any row of Γ_T^{-1} is bounded uniformly in the row index. If u_T is the vector composed of the quantities of the right side of (11) then $\Gamma_T^{-1}u_T$ has elements that are, uniformly, $O(Q_T)$. It follows that

$$\sum\sum_{j,k=1}^{p(T)} \{c(j-k) - \gamma(j-k)\}^2 = O\{p(T)^2 p(T)^{-4}\} = o(1).$$

Since this is the trace of $(\hat{\Gamma}_T - \Gamma_T)^2$ the largest eigenvalue of $(\hat{\Gamma}_T - \Gamma_T) = o(1)$ and hence the smallest eigenvalue of $\hat{\Gamma}_T = \Gamma_T + (\hat{\Gamma}_T - \Gamma_T)$ is also, a.s., bounded away from zero. It follows that the elements of $\hat{\Gamma}_T^{-1}$ are a.s., uniformly bounded. Now

$$\hat{\Gamma}_T^{-1} u_T = \{I_T + \hat{\Gamma}_T^{-1}(\Gamma_t - \hat{\Gamma}_T)\}\Gamma_T^{-1} u_T$$

where I_T is the $p(T)$-rowed unit matrix. The sum of the moduli of the elements along any row of $\hat{\Gamma}_T^{-1}(\Gamma_T - \hat{\Gamma}_T)$ is $O\{p(T)^2 o(p(T)^{-2})\} = o(1)$, because of (13) and the uniform boundedness of the elements of $\hat{\Gamma}_T^{-1}$. Thus Theorems 4 and 6 hold with the middle part of (3) replaced by (12), if in Theorem 6 $p(T) \geq d \ln T$, for suitably large d. For example we might take $p(T) = (\ln T)^a$, $a > 1$. The condition (12) seems mild. For example if $x(t)$ is regular (Ibragimov and Linnik, 1971, page 301) then $\{\varepsilon(t)^2 - \sigma^2\}$ has an absolutely continuous spectrum. If the spectral density is bounded then Menchoff's inequality holds (see Hannan, 1978) and by the methods of Moricz (1976) it may be shown that the left side of (12) is $O\{(\ln T)^a/T^{1/2}\}$, $a > 3/2$. Thus Theorems 4 and 6 hold under rather general conditions. There will be analogous relaxations in relation to Theorem 5 but the considerations are more complex since (10) does not hold.

4. Proofs of theorems.

Throughout this section we take $\sigma^2 = 1$ for simplicity.

PROOF OF THEOREM 1.

$$c(t) - \gamma(t) = T^{-1} \sum_0^\infty\sum \kappa(j)\kappa(k) \sum_{s=1}^{T-t} \{\varepsilon(s-j)\varepsilon(s+t-k) - \delta_{j,k-t}\} \tag{14}$$
$$- \frac{t}{T} \sum_0^\infty\sum \kappa(j)\kappa(k)\delta_{j,k-t}.$$

The last term is $t\gamma(t)/T$ and is thus $o(P(T)/T)$. Since $P(T) \leq T^a$, $a = r/\{2(r-2)\}$ then $P(T)/T \leq P(T)^{2/r}/T^{1/2}$ and this last term may be neglected. Thus we need consider only the first term in the right side. Put

$$S_\tau(j, k, t) = \sum_{s=1}^{\tau} \{\varepsilon(s-j)\varepsilon(s+t-k) - \delta_{j,k-t}\}.$$

This is a square integrable martingale relative to the σ-algebras $\mathscr{F}_{\tau+m}$, $m = \max(-j, t-k)$. Since the martingale differences have moments of order $r/2$ then first by Doob's inequality (Hall and Heyde, 1980, page 15) and then Burkholder's inequality, (op cit., page 23) we have

$$\mathscr{E}\{\max_{\tau\leq T} | S(j, k, t) |^{r/2}\} \leq cT^{r/4}. \tag{15}$$

Here and below we use c for a finite, positive constant, not always the same one. Put

$$m(t, \tau) = | \textstyle\sum_j \sum_k \kappa(j)\kappa(k) S_{\tau-t}(j, k, t) |.$$

Then from (15)

$$\mathscr{E}\{\max_{\tau<T} | m(t, \tau) |^{r/2}\} \leq cT^{r/4}$$

so that

$$P\{\max_{\tau<T}\max_{0\leq t\leq P(T)} | m(t, \tau) | \geq \psi(T)\} \leq cP(T)T^{r/4}\psi(T)^{-r/2}. \tag{16}$$

Thus if $\psi(T) = T^{1/2}\{P(T)\ln T\}^{2/r}(\ln\ln T)^{2(1+\delta)/r}$ then the right side of (16) is $c/\{\ln T(\ln\ln T)^{1+\delta}\}$. Now by precisely the same argument as in Moricz (1976, page 309, below (4.8)), for example, we may show that

$$\max_{0\leq t\leq P(T)} \frac{1}{T} | m(t, T) | = o\{T^{-1}\psi(T)\}$$

which proves the theorem.

PROOF OF THEOREM 2. The proof is long and complex but we can find no simpler proof. The neglected term in (14) is again negligible. We may truncate the sum in the first term on the right side in (14) at $j, k \le d \ln T$ for some sufficiently large d. For example

$$P\{\max_t | T^{-1} \sum_{d \ln T}^{\infty} \sum_{k=0}^{\infty} \kappa(j)\kappa(k) S_{T-t}(j, k, t) | \ge \varepsilon Q_T\}$$
$$\le \sum_t (\varepsilon Q_T)^{-2} E\{| T^{-1} \sum_{d \ln T}^{\infty} \sum_{k=0}^{\infty} \kappa(j)\kappa(k) S_{T-t}(j, k, t) |^2\}$$
$$\le c T^{-1} P(T) (\varepsilon Q_T)^{-2} \{\sum_{d \ln T}^{\infty} | \kappa(j) |\}^2 \{\sum_0^{\infty} | \kappa(k) |\}^2 = O(T^{-a}), \qquad a > 1,$$

since $|\kappa(j)|$ decreases to zero at a geometric rate. The required result now follows from the Borel-Cantelli lemma. Thus it is sufficient to consider, for a suitable $d < \infty$,

$$c_\varepsilon(t) = T^{-1} \sum_1^{T-t} \varepsilon(s)\varepsilon(s+t), \qquad 1 \le t \le d \cdot \ln T$$

We have omitted $c_\epsilon(0) - 1$. The $\varepsilon(t)^2 - 1$ are martingale differences and the law of the iterated logarithm holds for this; see Stout (1974a). The proof now follows the proof of the law of the iterated logarithm as given in Stout (1974b, pages 299–302). However first we must truncate the $\varepsilon(t)$ sequence and we have found it easiest to use the method of truncation in Hartman and Wintner (1941), rather than that in Stout (1974a). We choose $\lambda(t)$ as the function, given as $\lambda(r)$, on page 172 of Hartman and Wintner (1941). In the definition of $\lambda(r)$ the first two absolute moments of a distribution function, $\tau(x)$, occur and in our application these are to be the first two moments of the distribution of $\varepsilon(t)^2$. The function $\lambda(t)$ is of the form

$$\lambda(t) = Q_t^{-1} \eta(t), \qquad \lambda(t) > t^{1/3},$$

where $\eta(t)$ decreases monotonically to zero but $\lambda(t)$ increases monotonically to ∞. It is not difficult to see that, for $s, t \ge 3$,

$$\lambda(s+t)^{1/2} \le \lambda(s)^{1/2} + \lambda(t)^{1/2}. \tag{17}$$

This follows from the same result for Q_t^{-1} since $\eta(t)$ decreases monotonically. Now put, for $t \ge 3$,

$$g(t) = \begin{cases} \varepsilon(t), & |\varepsilon(t)| < \lambda(t)^{1/2}, \\ 0, & \text{otherwise}. \end{cases}$$

It is shown in Hartman and Wintner (1941, page 174) that, putting $\phi(t) = (t \ln \ln t)^{1/2}$, $e(t) = \varepsilon(t) - g(t)$,

$$\sum_3^{\infty} \mathscr{E}\{e(t)^2\}/\phi(t) < \infty. \tag{18}$$

Now we show that $Tc_\varepsilon(t)$ can be replaced by the martingale,

$$\sum_3^{T-t} [g(s)g(s+t) - \mathscr{E}\{g(s)g(s+t) \mid \mathscr{F}_{s+t-1}\}].$$

The effect of truncation is essentially dominated in modulus by

$$\phi(T)^{-1} \sum_3^T | e(s)g(s+t) + g(s)e(s+t) + e(s)s(s+t) |.$$

The contribution from the last term under the modulus sign is dominated in modulus by

$$\phi(T)^{-1} \sum_3^{T+\ln T} e(s)^2$$

which converges a.s. to zero by (18), using Kronecker's lemma. The contribution from the first term is dominated in modulus by

$$\phi(T)^{-1} \max_t \sum_3^T | e(s) | \lambda(s+t)^{1/2} \le \phi(T)^{-1} \max_t \sum_3^T | e(s) | \{\lambda(s)^{1/2} + \lambda(t)^{1/2}\}.$$

Now $\phi(T)^{-1} \sum | e(s) | \lambda(s)^{1/2}$ will converge to zero, by Kronecker's lemma, if

$$\sum_3^{\infty} \frac{| e(s) | \lambda(s)^{1/2}}{\phi(s)} < \infty, \quad \text{a.s.} \tag{19}$$

This holds true, because of (18), since $\mathscr{E}|e(s)| > \lambda(s)^{1/2}$. On the other hand,

$$\max_t \phi(T)^{-1}\lambda(t)^{1/2} \sum |e(s)| \le \lambda(T)^{1/2}\phi(T)^{-1} \sum |e(s)|,$$

again converges to zero a.s. by Kronecker's lemma because of (19). Finally

$$\phi(T)^{-1}\max_t \textstyle\sum_1^T |e(s+t)g(s)| \le \phi(T)^{-1}\max_t \sum_1^T |e(s+t)|\lambda(s+t)^{1/2},$$

which again converges to zero by the same argument. Then the effect of truncation is negligible. Moreover, since $\mathscr{E}\{\varepsilon(s)\varepsilon(s+t) \mid \mathscr{F}_{s+t-1}\} = 0$, $t > 0$,

$$\begin{aligned}-\phi(T)^{-1} \textstyle\sum_3^{T-t} &\mathscr{E}\{g(s)g(s+t) \mid \mathscr{F}_{s+t-1}\} \\ &= \phi(T)^{-1} \textstyle\sum_3^{T-t} \mathscr{E}\{e(s)g(s+t) + g(s)e(s+t) + e(s)e(s+t) \mid \mathscr{F}_{s+t-1}\},\end{aligned}$$

and this converges to zero, uniformly in t, by the same arguments. Therefore the effect of mean correction is negligible. Before proving the desired result, we need the evaluation that, uniformly in t, and almost surely,

$$\begin{aligned}s_t^2(T) &= \textstyle\sum_3^{T-t} \operatorname{var}(g(s)g(s+t) \mid \mathscr{F}_{s+t-1}) \\ &= \textstyle\sum_3^{T-t} \mathscr{E}\{g(s)^2 g(s+t)^2 \mid \mathscr{F}_{s+t-1}\} - \sum_3^{T-t} [\mathscr{E}\{g(s)g(s+t) \mid \mathscr{F}_{s+t-1}\}]^2 \qquad (20) \\ &= T\{1 + o(1)\}.\end{aligned}$$

However

$$\begin{aligned}[\mathscr{E}\{g(s)g(s+t) \mid \mathscr{F}_{s+t-1}\}]^2 &= g(s)^2[\mathscr{E}\{e(s+t) \mid \mathscr{F}_{s+t-1}\}]^2 \\ &\le \lambda(s+t)[\mathscr{E}\{e(s+t) \mid \mathscr{F}_{s+t-1}\}]^2 \le Q_{s+t}^{-1}\mathscr{E}\{e(s+t)^2 \mid \mathscr{F}_{s+t-1}\},\end{aligned}$$

so that the second term in (20) is $o(T)$, again using (18). Also since $e(s)g(s) \equiv 0$

$$\mathscr{E}\{g^2(s)g^2(s+t) \mid \mathscr{F}_{s+t-1}\} = \{\varepsilon(s)^2 - e(s)^2\}\,[1 - \mathscr{E}\{e(s+t)^2 \mid \mathscr{F}_{s+t-1}\}]$$

and thus

$$\textstyle\sum_3^{T-t} \mathscr{E}\{g(s)^2 g(s+t)^2 \mid \mathscr{F}_{s+t-1}\} = T\{1 + o(1)\},$$

by the same kind of argument as before, and (20) is established. Now we prove the desired result. Put

$$S_t(T) = \textstyle\sum_{u=t+1}^T Y_t(u) \qquad (21)$$

where $Y_t(u) = g(u)g(u-t) - \mathscr{E}\{g(u)g(u-t) \mid \mathscr{F}_{u-1}\}$. Then it suffices to show that

$$P\{\max_t S_t(T) > b\phi(T) \text{ infinitely often}\} = 0, \quad \text{for some} \quad b < \infty. \qquad (22)$$

Let T_k be the smallest integer T such that $\max_t s_t^2(T+1) \ge p^{2k}$, and $S_t^{(k)}$ be $S_t(T)$ for $T \le T_k$ and $S_t(T_k)$ for $T > T_k$. Then $p^{2k-1} < T_k < p^{2k+1}$ a.s. for large enough T, as is seen from (20). Therefore it suffices to show that

$$\textstyle\sum_k P\{\sup_{T\ge 3}\max_t S_t^{(k)}(T) > b\phi(T_k)\}.$$

Put $(\mathrm{p}^{2k+1}) = p(k)$. The above is implied by

$$\textstyle\sum_k \sum_{t=1}^{p(k)} P\{\sup_{T\ge p(k)} S_t^{(k)}(T) > b\phi(T_k)\} < \infty$$

and, as shown in Stout (1974b, pages 301–305), the left hand side is dominated by

$$\textstyle\sum_k p(k)\exp(\tfrac14 b^2 \ln\ln p^{2k}),$$

which converges for b sufficiently large. Replacing $S_t(T)$ by $-S_t(T)$ and repeating the proof establishes the result.

Proof of Theorem 3. The proof is basically of the same form as for Theorem 2 and can be given in outline except for the proof that we may truncate the sum over j, k in (14).

Since $t^{1/2}\gamma(t) \to 0$ because of (5), $t\gamma(t)T^{-1} = o(T^{1/2})$, $t \le T$ and $\gamma(t) = o(T^{-1/2})$, $t \ge T$. Thus we need consider only $c(t) - \gamma(t)$, $t < T$ and may neglect the last term in (14). Put

$$S_{j,k,t}(T) = \textstyle\sum_1^{T-t} \{\varepsilon(s-j)\varepsilon(s+t-k) - \delta_{j,k-t}\}.$$

Now we shall show that, uniformly in t,

$$\sum_0^{\infty}\sum \kappa(j)\kappa(k)S_{j,k,t}(T) - \sum_0^{T\ln T}\sum \kappa(j)\kappa(k)S_{j,k,t}(T)$$

is $o\{(\ln T/T)^{1/2}\}$. This is made up of three terms, of which one is

$$a_t(T) = \textstyle\sum_{T\ln T+1}^{\infty} \sum_{k=0}^{\infty} \kappa(j)\kappa(k)S_{j,k,t}(T),$$

another is the same with lower limits of summation for j, k reversed and a third has both lower limits of summation at $(T \ln T + 1)$. The proof is essentially the same for all three. Uniformly in t we have

$$\mathscr{E}[\{\textstyle\sum_{T\ln T+1}^{\infty} \sum_0^{\infty} \kappa(j)\kappa(k)S_{j,k,t}(T)\}^2] \le \{\sum_{T\ln T+1}^{\infty} |\kappa(j)|\}^2\{\sum_0^{\infty}|\kappa(j)|\}^2 O(T)$$

which is $o\{(\ln T)^{-1}\}$. Since T values of t are involved

$$P\{\max_t |a_t(T)| \ge \varepsilon(T\ln T)^{1/2}\} = o\{(\ln T)^{-2}\}.$$

Put $T_\ell = 2^\ell$ and $T(\ell) = T_\ell \ln T_\ell$. Then $\max_t |a_t(T_\ell)|/T(\ell)^{1/2}$ is, a.s., eventually less than ε and thus $\max_t|a_t(T_\ell)|/T(\ell)^{1/2}$ converges a.s. to zero.

For $T_\ell < T \le T_{\ell+1}$, putting $S_{j,k,t}(T_\ell) = 0$, $t \ge T_\ell$, we have

$$\text{(23)} \qquad |a_t(T) - a_t(T_\ell)| \le |\textstyle\sum_{T(\ell)+1}^{T\ln T} \sum_0^{\infty} \kappa(j)\kappa(k)S_{j,k,t}(T_\ell)| + |\sum_{T\ln T+1}^{\infty}\sum_0^{\infty} \kappa(j)\kappa(k)\{S_{j,k,t}(T) - S_{j,k,t}(T_\ell)\}|.$$

Now, uniformly in t,

$$\mathscr{E}[\max_{T_\ell<T\le T_{\ell+1}} |\textstyle\sum_{T(\ell)+1}^{T\ln T}\sum_0^{\infty} \kappa(j)\kappa(k)\{S_{j,k,t}(T_\ell)|^2\}] \le \{\sum_{T(\ell)+1}^{T(\ell+1)} |\kappa(j)|\}^2\{\sum_0^{\infty}\kappa(j)\}^2 O(T_\ell) = o\{(\ln T_\ell)^{-1}\}.$$

On the other hand, uniformly in t,

$$\begin{aligned}\mathscr{E}\{\max_{T_\ell<T\le T_{\ell+1}} &|\textstyle\sum_{T\ln T+1}^{\infty}\sum_0^{\infty} \kappa(j)\kappa(k)\{S_{j,k,t}(T) - S_{j,k,t}(T_\ell)\}|^2\}\\ &\le [\textstyle\sum_{T(\ell)+1}^{\infty}|\kappa(j)|\sum_0^{\infty}|\kappa(k)(\mathscr{E}[\max_{T_\ell<T\le T_{\ell+1}}\{S_{j,k,t}(T) - S_{j,k,t}(T_\ell)\}^2])^{1/2}]^2\\ &\le \{\textstyle\sum_{T(\ell)+1}^{\infty}|\kappa(j)|\}^2\{\sum_0^{\infty}|\kappa(k)|\}^2 O(T_{\ell+1} - T_\ell) = o\{(\ln T_\ell)^{-1}\}\end{aligned}$$

by Doob's inequality applied to the martingales

$$\textstyle\sum_{T_\ell+1-t}^{T-t} \{\varepsilon(s-j)\varepsilon(s+t-k) - \delta_{j,k-t}\}, \qquad t < T_\ell;$$

$$\textstyle\sum_1^{T-t} \{\varepsilon(s-j)\varepsilon(s+t-k) - \delta_{j,k-t}\}, \qquad T_\ell < t \le T_{\ell+1}.$$

Thus from (23), putting $a_t(T_\ell) = 0$, $t \ge T_\ell$, we obtain, uniformly in t,

$$\mathscr{E}\{\max_{T_\ell<T\le T_{\ell+1}} |a_t(T) - a_t(T_\ell)|^2\} = O\{(\ln T_\ell)^{-1}\}$$

so that

$$\max_{T_\ell<T\le T_{\ell+1}}\max_t |\{a_t(T) - a_t(T_\ell)\}/T(\ell)^{1/2}| \to 0, \quad \text{a.s.}$$

Thus we need to consider only

$$\sum_0^{T\ln T}\sum \kappa(j)\kappa(k)S_{j,k,t}(T), \qquad t < T.$$

Now we truncate the $\varepsilon(s)$ at $\lambda(s)^{1/2}$ where $\lambda(s) = (s/\ln s)^{1/2}\eta(s)$. All of the properties of $\lambda(s)$ used in the proof of Theorem 2 are retained by a suitable choice of $\eta(s)$ and the proofs are not different. Again effects of truncation are negligible. Thus we introduce $g(u)g(u - a_{j,k,t})$ where $a_{j,k,t} = |t + j - k|$ and

$$Y_{j,k,t}(u) = g(u)g(u - a_{j,k,t}) - \mathscr{E}\{g(u)g(u - a_{j,k,t}) \mid \mathscr{F}_{u-1}\}$$

and consider

$$\sum_T P\{\max_{j,k,t} \sum Y_{j,k,t}(u) > c(T \ln T)^{1/2}\}. \tag{24}$$

Now put

$$S^2_{j,k,t}(T) = \sum_1^{T-t} \mathscr{E}\{Y_{j,k,t}(u)^2 \mid \mathscr{F}_{u-1}\}$$

which is easily seen to be less than or equal to $T\{1 + o(1)\}$. Then (24) is dominated by

$$\sum_T \sum_{j,k,t} P[\exp\{(\ln T/T)^{1/2} \sum Y_{j,k,t}(u) - ½(\ln T/T)(1 + \dot{\eta}(T))S^2_{j,k,t}(T)\} > \exp c_1 \ln T].$$

Here $c_1 = c - \{1 + \eta(T)\}s^2_{j,k,t}(T)/(2T)$. There are $O\{(T \ln T)^2\}$ values of j, k, t since $a_{j,k,t}$ depends only on j and $t - k$. Thus from Lemma 5.4.1 and Corollary (5.4.1) of Stout (1974b), we bound (24) by

$$c_2 \sum_T (T \ln T)^2 e^{-c_1 \ln T}$$

which is certainly finite for c_1, and hence c, large enough. Again replacing $Y_{j,k,t}(u)$ by $-Y_{j,k,t}(u)$ the result follows.

Proof of Theorem 4. As was shown in Hannan and Quinn (1979), (7) is not minimised at $p < p_0$, symptotically. Thus we need to show that, for all T sufficiently large,

$$\min_p\{\ln \hat{\sigma}_p^2 - \ln \hat{\sigma}_{p_0}^2 + (p - p_0)\ln T/T\} > 0$$

where $p_0 < p \le c \ln T$, $c > 0$. Now we introduce a matrix $\hat{\Gamma}_p = \{c(j - k), j, k = 1, \cdots, p\}$ and the vector $\hat{\gamma}_p = \{c(j), j = 1, \cdots, p\}$. Then the vector $\hat{\alpha}_p = \{\hat{\alpha}_p(j), j = 1, \cdots, p\}$ of autoregressive coefficient estimates, is a solution of the equation $\hat{\Gamma}_p\hat{\alpha}_p = -\hat{\gamma}_p$ and as is well-known (see Hannan, 1970)

$$\hat{\sigma}_p^2/\hat{\sigma}_{p_0}^2 = \prod_{j=p_0+1}^{p} \{1 - \hat{\alpha}_j^2(j)\}.$$

It is evidently sufficient to show that

$$\max_{p_0 < j \le c \ln T} |\hat{\alpha}_j(j)| = o\{(\ln T/T)^{1/2}\}.$$

Let Γ_p and γ_p be the matrix and vector made by replacing the $c(j)$ in $\hat{\Gamma}_p$ and $\hat{\gamma}_p$ by $\gamma(j)$. Let $\Gamma_p\alpha_p = -\gamma_p$. The smallest eigenvalue of Γ_p is bounded away from zero uniformly by p since (2′) necessarily holds when $h(z) \equiv 1$. (See Section 3 below (13)). Also the largest eigenvalue in modulus of $(\hat{\Gamma}_p - \Gamma_p)$ is bounded, uniformly in p, by

$$\max_p\left[\sum_{j,k=1}^{p}\sum \{c(j - k) - \gamma(j - k)\}^2\right]^{1/2} = O\left(\frac{\ln T(\ln \ln T)^{1/2}}{T^{1/2}}\right),$$

from Theorem 2. Thus the smallest eigenvalue of $\hat{\Gamma}_p = \Gamma_p + (\hat{\Gamma}_p - \Gamma_p)$ is bounded away from zero uniformly in $p \le c \ln T$. (Again see Section 3 below (13)). Now

$$\begin{aligned}(\hat{\alpha}_p - \alpha_p) &= -\Gamma_p^{-1}\{(\hat{\gamma}_p - \gamma_p) + (\hat{\Gamma}_p - \Gamma_p)(\hat{\alpha}_p - \alpha_p) + (\hat{\Gamma}_p - \Gamma_p)\alpha_p\},\\ \{I_p + \Gamma_p^{-1}(\hat{\Gamma}_p - \Gamma_p)\}(\hat{\alpha}_p - \alpha_p) &= -\Gamma_p^{-1}\{(\hat{\gamma}_p - \gamma_p) + (\hat{\Gamma}_p - \Gamma_p)\alpha_p\}.\end{aligned} \tag{25}$$

Γ_p^{-1}, being the inverse of the covariance matrix of an autoregressive process, has all elements null outside of the first p_0 diagonals above and below the main diagonal (Hannan, 1970, page 351). Let $\mu_p(i, j)$ be the typical element of Γ_p^{-1}. It is evidently bounded, uniformly in p. Now

$$\max_{p_0<p\le P(T)}\max_{1\le j\le p}\Big|\sum_{k,\ell}\sum \mu_p(j,k)\{c(k-\ell)-\gamma(k-\ell)\}\{\hat\alpha_p(\ell)-\alpha_p(\ell)\}\Big|$$

$$\le \max_p\max_j|\hat\alpha_p(j)-\alpha_p(j)|\,\{\textstyle\sum_k|\mu_p(j,k)|\sum_\ell \max_\ell|c(k-\ell)-\gamma(k-\ell)|\}$$

$$\le \max_p\max_j|\hat\alpha_p(j)-\alpha_p(j)|\,O\{Q_TP(T)\}$$

since $|\mu_p(j, k)|$ is uniformly bounded and null save for $2p_0$ values of k, at most, $\max_\ell|c(k-\ell)-\gamma(k-\ell)|=O(Q_T)$, by Theorem 2, and there are $P(T)$ values of ℓ to sum over. Thus

$$\{I_p+\Gamma_p^{-1}(\hat\Gamma_p-\Gamma_p)\}(\hat\alpha_p-\alpha_p)=\{1+o(1)\}(\hat\alpha_p-\alpha_p).$$

Moreover

$$\max_p\max_j\Big|\textstyle\sum_1^p \mu_p(j,k)\{c(k)-\gamma(k)\}$$
$$-\textstyle\sum_{k=1}^p\sum_{\ell=1}^p \mu_p(j,k)\{c(k-\ell)-\gamma(k-\ell)\}\alpha_p(\ell)\Big|=O(Q_T)$$

by the same kind of argument, since $\alpha_p(\ell)$ is null except for p_0 values and using Theorem 2. Thus from (25), since $\alpha_p(p)=0$, $p>p_0$,

$$\max_{p_0<p<P(T)}|\hat\alpha_p(p)|=\max_{p_0<p<P(T)}|\hat\alpha_p(p)-\alpha_p(p)|=O(Q_T),$$

which established Theorem 4.

Proof of Theorem 5. We recall the definitions above Theorem 6 and at the beginning of Section 3. As before we write Γ_T, $\hat\Gamma_T$, γ_T, $\hat\gamma_T$, α_T, $\hat\alpha_T$ instead of Γ_p, etc. at $p=p(T)$. By (9) and the fact that $\sigma_T^2\to\sigma^2=1$ it follows that $|\alpha_T(j)-a(j)|$ converges uniformly to zero for $0\le j\le p(T)$. Hence

$$\Big|\textstyle\sum_0^{p(T)}\alpha_T(j)e^{ij\omega}\Big|^2-\Big|\sum_0^\infty a(j)e^{ij\omega}\Big|^2\to 0$$

and since (2′) holds

$$\lim_{T\to\infty}\sup_\omega\Big|\sigma_T^2\Big|\textstyle\sum_0^{p(T)}\alpha_T(j)e^{ij\omega}\Big|^{-2}-2\pi f(\omega)\Big|=0.$$

Also

$$\textstyle\sum_0^{p(T)}|\alpha_T(j)|\le c<\infty.$$

Putting

$$f_T(\omega)=(\sigma_T^2/2\pi)\Big|\textstyle\sum_0^{p(T)}\alpha_T(j)e^{ij\omega}\Big|^{-2}$$

it is thus only necessary to show that $\sup|f_T(\omega)-\hat f_T(\omega)|=o(1)$. Thus from (25) we consider

$$\Gamma_T\{\hat\alpha_T-\alpha_T\}=-(\hat\Gamma_T-\Gamma_T)(\hat\alpha_T-\alpha_T)-(\hat\gamma_T-\gamma_T)-(\hat\Gamma_T-\Gamma_T)\alpha_T.$$

For the first part of Theorem 5 we have, using Theorem 3

$$\|(\hat\Gamma_T-\Gamma_T)(\alpha_T-\alpha_T)\|^2=\textstyle\sum_{k=1}^{p(T)}\big[\sum_{j=1}^{p(T)}\{c(j-k)-\gamma(j-k)\}\{\hat\alpha_T(j)-\alpha_T(j)\}\big]^2$$
$$\le\sum_{j,k}\sum\{c(j-k)-\gamma(j-k)\}^2\textstyle\sum_1^{p(T)}\{\hat\alpha_T(j)-\alpha_T(j)\}^2$$
$$=o(1)\textstyle\sum_1^{p(T)}\{\hat\alpha_T(j)-\alpha_T(j)\}^2.$$

Similarly

$$\|\hat\gamma_T-\gamma_T\|^2=\textstyle\sum_1^{p(T)}\{c(j)-\gamma(j)\}^2=o\{(\ln T/T)^{1/2}\},$$
$$\|(\hat\Gamma_T-\Gamma_T)\alpha_T\|^2=o\{(\ln T/T)^{1/2}\}.$$

Hence $\{1 + o(1)\} \| \hat{\alpha}_T - \alpha_T \|^2 = o\{(\ln T/T)^{1/2}\}$ and

$$\sup_\omega | \sum \{\hat{\alpha}_T(j) - \alpha(j)\} e^{ij\omega} | = o\{P(T)(\ln T/T)^{1/2}\} = o(1).$$

Finally

$$| \hat{\sigma}_T^2 - \sigma_T^2 | = | c(0) - \gamma(0) - \hat{\alpha}'_T \hat{\gamma}_T + \alpha'_T \gamma_T | = o(1)$$

so that the first part of the theorem is established. The second part of Theorem 5 is proved in the same way using Theorem 1.

PROOF OF THEOREM 6. This theorem is proved in almost the same way as were Theorems 4 and 5. Thus we use (24) for $p = p(T) = O(\ln T)$. Γ_T again has its smallest eigenvalue bounded away from zero because of (2′) and $(\hat{\gamma}_T - \gamma_T)$ is $O(Q_T)$. The elements of α_T decrease at a geometric rate so that $(\hat{\Gamma}_T - \Gamma_T)\alpha_T = O(Q_T)$. Since, using Γ_T for Γ_p, $p = p(T)$, we also have $\{I_T + \Gamma_T^{-1}(\hat{\Gamma}_T - \Gamma_T)\} = I_T\{1 + o(1)\}$ then $(\hat{\alpha}_T - \alpha_T) = O(Q_T)$.

REFERENCES

AKAIKE, H. (1977). On entropy maximisation principle. In *Applications of Statistics*, Ed. P. R. Krishnaiah, 27–41. North-Holland, Amsterdam.

BAXTER, G. (1962). An asymptotic result for the finite predictor. *Math. Scand.* **10** 137–144.

BERK, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489-502.

HALL, P. and HEYDE, C. C. (1980). *Martingale Theory and its Application.* Academic, New York.

HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

HANNAN, E. J. (1974). The uniform convergence of autocovariances. *Ann. Statist.* **2** 803–806.

HANNAN, E. J. (1978). Rates of convergence for time series regression. *Adv. Appl. Probability* **10** 740–743.

HANNAN, E. J. and HEYDE, C. C. (1972). On limit theorems for quadratic functions of discrete time series. *Ann. Math. Statist.* **43** 2058–2066.

HANNAN, E. J. and QUINN, B. G. (1979). The determination of the order of an autoregression. *J. Roy. Statist. Soc. B.* **41** 190–195.

HANNAN, E. J. and RISSANEN, J. (1982). Recursive estimation of ARMA order. *Biometrika,* **69** 81–94.

HARTMAN, P. and WINTNER, A. (1941). On the law of the iterated logarithm. *Amer. J. Math.* **63** 169–176.

IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

MORICZ, F. (1976). Moment inequalities and the strong law of large numbers. *Z. Wahrsch. verw. Gebiete* **35** 298–314.

RISSANEN, J. (1978). Modeling by shortest data description. *Automatica* **14** 465–471.

STOUT, W. F. (1974a). The Hartman-Wintner law of the iterated logarithm for martingales. *Ann. Math. Statist.* **41** 2158–2160.

STOUT, W. F. (1974b). *Almost Sure Convergence.* Academic, New York.

AN HONG-ZHI
INSTITUTE OF APPLIED MATHEMATICS
CHINESE ACADEMY OF SCIENCES
BEIJING, CHINA

CHEN ZHAO-GUO
INSTITUTE OF APPLIED MATHEMATICS
CHINESE ACADEMY OF SCIENCES
BEIJING, CHINA

E. J. HANNAN
THE RESEARCH SCHOOL OF SOCIAL SCIENCES
THE AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 2600, AUSTRALIA