The Annals of Statistics
1988, Vol. 16, No. 2, 791–807

ON ESTIMATION OF A REGRESSION MODEL WITH LONG-MEMORY STATIONARY ERRORS

BY YOSHIHIRO YAJIMA

Tokyo Institute of Technology

We consider estimation of a regression model with long-memory stationary errors. First, we estimate the regression parameters by the least-squares estimator (LSE) and, next, those describing the correlation structure of the error terms by using the residuals obtained from the LSE. Certain regularity conditions introduced to develop the asymptotic theory no longer hold in this model. In this situation we derive asymptotic properties of the preceding estimation procedure.

1. Introduction. We shall consider the regression model of the form

$$y_t = X_t'\beta + \varepsilon_t,$$

where $\{y_t\}$ is an observed sequence, $X_t = (x_{t1}, \dots, x_{tl})'$, an l-vector of nonstochastic regressors, $\{\varepsilon_t\}$, a sequence of errors and $\beta = (\beta_1, \dots, \beta_l)'$ are unknown regression parameters. $\{\varepsilon_t\}$ is usually assumed to be correlated in time series. Several authors have discussed estimation of the regression parameters β and those which describe the structure of the correlation in the errors $\{\varepsilon_t\}$. Most of them have focused on errors following a stationary ARMA process. Thus the spectral density is bounded at the frequency $\lambda = 0$ and the autocorrelation decays to 0 very rapidly.

Yet in many empirical time series, especially those of economics and geophysics, the dependence between distant observations is so strong that ARMA models are unable to express the spectral density of low frequencies adequately [Mandelbrot and Wallis (1969) and Granger and Joyeux (1980)]. Models that can represent such long-range dependence are also needed to deal with the problem of violation of assumption of independence in robust estimation [Hampel, Rousseeuw, Ronchetti and Stahel (1986)].

Hence attention has recently been paid to two parametric models that have unbounded spectral densities at $\lambda = 0$ and autocorrelations that decay to 0 more slowly than that of an ARMA model. The first model, called a fractional Gaussian noise, was introduced by Mandelbrot and Van Ness (1968) and has been used to analyze geophysical time series. It is a stationary Gaussian process with mean 0 and covariance

$$\gamma_k = E\varepsilon_t\varepsilon_{t+k} = (C/2)\{|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\},$$

where H is a parameter satisfying $\frac{1}{2} < H < 1$ and $C > 0$. Hence γ_k satisfies

$$\gamma_k \sim CH(2H-1)k^{2H-2},$$

Received March 1986; revised May 1987.
AMS 1980 *subject classifications*. Primary 62M10; secondary 62J05.
Key words and phrases. Long-memory models, regression, least-squares estimators.

as $k \to \infty$. The spectral density is given by

$$f(\lambda; H) = CF(H) f_0(\lambda; H),$$

where

$$f_0(\lambda; H) = (1 - \cos\lambda) \sum_{k=-\infty}^{\infty} |\lambda + 2k\pi|^{-1-2H}$$

and

$$F(H) = \left\{ \int_{-\infty}^{\infty} (1 - \cos x)|x|^{-1-2H}\, dx \right\}^{-1}$$

[Sinai (1976) and Geweke and Porter-Hudak (1983)]. Hence

$$f(\lambda; H) \sim (CF(H)/2)|\lambda|^{1-2H},$$

as $\lambda \to 0$. The second model was proposed by Granger and Joyeux (1980) and Hosking (1981) to analyze a time series that has an unbounded spectral density at $\lambda = 0$ but the first difference of which has a spectral density vanishing at $\lambda = 0$. Let $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$ and $\theta(z) = 1 - \sum_{j=1}^{q} \theta_j z^j$. Suppose that $\phi(z)$ and $\theta(z)$ have no zeros on or inside the unit circle and no zeros in common. For $0 < d < \frac{1}{2}$, we define the spectral density

$$f(\lambda; d, \phi, \theta) = C|\theta(e^{i\lambda})|^2 / |\phi(e^{i\lambda})(1 - e^{i\lambda})^d|^2.$$

A stationary process with mean 0 and spectral density $f(\lambda; d, \phi, \theta)$ is called a fractional ARIMA(p, d, q) process. The covariance satisfies

$$\gamma_k \sim 2\pi C|\theta(1)/\phi(1)|^2 \Gamma(1 - 2d)/\{\Gamma(d)\Gamma(1 - d)\} k^{2d-1}, \tag{1}$$

as $k \to \infty$ and the spectral density satisfies

$$f(\lambda; d, \phi, \theta) \sim C|\theta(1)/\phi(1)|^2 |\lambda|^{-2d},$$

as $\lambda \to 0$ [Hosking (1981), Theorems 1 and 2]. If we define the backward shift operator B by $B\varepsilon_t = \varepsilon_{t-1}$ and the fractional difference operator ∇^d by a binomial series,

$$\nabla^d = (1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k,$$

a fractional ARIMA(p, d, q) process $\{\varepsilon_t\}$ satisfies

$$\phi(B) \nabla^d \varepsilon_t = \theta(B) a_t,$$

where the white noise $\{a_t\}$ consists of uncorrelated random variables. Hence this model is a generalization of Box and Jenkins' ARIMA model [Box and Jenkins (1976)].

These two models are long-memory stationary models since their autocorrelations are not absolutely summable. Throughout the rest of this paper we discuss estimation of the regression model when $\{\varepsilon_t\}$ is a long-memory stationary process. In our procedure we first estimate the regression parameters β by the least-squares estimator (LSE) and, next, the parameters that describe ε_t by using the residuals obtained from the LSE.

Since the spectral density of the errors is unbounded, certain regularity conditions introduced to develop the asymptotic theory no longer hold. We shall investigate how our estimation procedure behaves asymptotically in this situation. In Section 2 we first derive a sufficient condition for the strong consistency of the LSE of β. Next, we consider a polynomial regression, $X_t = (1, t, \dots, t^{l-1})'$ and evaluate the limits of the covariance matrices of both the LSE and the best linear unbiased estimator (BLUE) in order to show how well the LSE works. Then it is shown that the LSE is not asymptotically efficient but achieves a high relative efficiency for polynomials of lower order.

In Section 3 we mainly consider an ARIMA(0, d, 0) process and discuss the asymptotic properties of the estimators for the parameters that describe ε_t. And we show that even if β are unknown, the estimators have such properties as strong consistency and asymptotic normality under some condition on X_t.

2. The asymptotic properties of the LSE of the regression parameters. The strong consistency of the LSE in a regression model has been discussed in various situations of $\{\varepsilon_t\}$ such that $\{\varepsilon_t\}$ is independently, identically distributed, a martingale difference and a stationary process with a bounded spectral density. [See Anderson and Taylor (1976, 1979), Lai and Robbins (1977), Lai, Robbins and Wei (1978, 1979) and Solo (1981).]

Here we consider a stationary process $\{\varepsilon_t\}$ with a spectral density $f(\lambda)$ of the form

$$f(\lambda) = f^*(\lambda)/|1 - e^{i\lambda}|^{2d}, \qquad 0 < d < \tfrac{1}{2}, \tag{2}$$

where $f^*(\lambda)$ is a nonnegative bounded function. If $f^*(\lambda)$ is nonzero in some neighborhood of $\lambda = 0$, $f(\lambda)$ is unbounded at $\lambda = 0$ and behaves like

$$f(\lambda) = O(|\lambda|^{-2d}),$$

as $\lambda \to 0$. We derive a sufficient condition on X_t, which assures the strong consistency of the LSE by tracing the procedure developed by Solo (1981).

Let $Y_T = (y_1, \dots, y_T)'$ be the sequence of observations and $\hat{\beta}_T = (\hat{\beta}_{1,T}, \dots, \hat{\beta}_{l,T})$ be the LSE of β. Then

$$\hat{\beta}_T = V_T^{-1} \sum_{t=1}^{T} X_t y_t,$$

where

$$V_T = \sum_{t=1}^{T} X_t X_t'.$$

Assume that V_l is positive definite. $\hat{\beta}_T$ can be expressed as

$$\begin{aligned} \hat{\beta}_T &= \hat{\beta}_{T-1} + V_T^{-1} X_T e_T, \\ e_T &= y_T - X_T' \hat{\beta}_{T-1}, \end{aligned} \tag{3}$$

for $T \geq l + 1$. Here we introduce the notation used throughout this paper. α is a vector of an appropriate dimension in each context with $\|\alpha\| = 1$, where $\|\alpha\|$ implies the Euclidean norm. For any matrix A, denote the minimum eigenvalue of A by $\lambda_{\min}(A)$. K stands for general constants being independent of T and d, but is not always the same one in each context.

Now we have the following result.

THEOREM 2.1. *Let $\{\varepsilon_t\}$ be a stationary process with the spectral density $f(\lambda)$ of (2). Assume that $f^*(\lambda)$ is a nonnegative bounded function. And assume*

(i) $\mu_t/t^{2d} \to \infty$ *as* $t \to \infty$,

(ii) $\sum_{t=l+2}^{\infty} \log^2 t \cdot t^{2d-1}/\mu_{t-1} < \infty$,

where $\mu_t = \lambda_{\min}(V_t)$. *Then*

$$\lim_{T\to\infty} \hat{\beta}_T = \beta \quad a.s.$$

PROOF. Noting (3), we have only to show that

$$p - \lim_{T\to\infty} \hat{\beta}_T = \beta, \tag{4}$$

and the almost sure convergence of $\sum_{t=l+1}^{T} \bar{c}_t e_t$ as $T \to \infty$, where $\bar{c}_t = \alpha' V_t^{-1} X_t$. Define $\tilde{\gamma}_k$ by

$$\tilde{\gamma}_k = \int_{-\pi}^{\pi} e^{ik\lambda}/|1 - e^{i\lambda}|^{2d}\, d\lambda.$$

And let R_T and $\tilde{R}_T$ be the $T \times T$ matrices with (i, j)th entry γ_{i-j} and $\tilde{\gamma}_{i-j}$, respectively, and I_T be the $T \times T$ identity matrix. Then boundedness of $f^*(\lambda)$ and relation (1) imply

$$R_T < K\tilde{R}_T < KT^{2d} I_T, \tag{5}$$

where the inequality is defined in the positive definite sense. Hence

$$\operatorname{Var}\left(\alpha'\hat{\beta}_T\right) \le KT^{2d}\alpha' V_T^{-1}\alpha.$$

Then (4) is obtained by condition (i). Next, we shall derive the a.s. convergence of $\sum_{t=l+1}^{T} \bar{c}_t e_t$. Define $S_{m,n}$ by

$$S_{m,n} = \sum_{t=m+1}^{n} c_t e_t,$$

for any constants $\{c_t\}$. Using the same method as that of Solo (1981), page 690, and noting (5), we find

$$ES_{m,n}^2 < Kn^{2d} \sum_{t=m+1}^{n} c_t^2 v_t^2, \tag{6}$$

where $v_t^2 = (1 + X_t' V_{t-1}^{-1} X_t)$. Next, following the procedure of Stout (1974), Theorem 2.3.1, page 18, with relation (6) we obtain

$$E\left\{\max_{1\le i\le r} S_{m,m+i}^2\right\} \le K(\log 4r/\log 2)^2 (m + 2r)^{2d} \sum_{t=m+1}^{m+r} c_t^2 v_t^2. \tag{7}$$

Finally, if

$$\sum_{t=l+1}^{\infty} \log^2 t \cdot t^{2d} c_t^2 v_t^2 < \infty \tag{8}$$

is satisfied, $S_{l,T}$ is shown to converge a.s. as $T \to \infty$ by the method of subsequence [Stout (1974), page 20] with relation (7). Condition (ii) assures that $\bar{c}_t$ satisfies (8), completing the proof. □

REMARK 2.1. Theorem 2.1 also holds for a stationary process with $\gamma_k = O(|k|^{2d-1})$ as $k \to \infty$, as is seen from the proof.

Now we put

$$m_{ij}^T = \sum_{t=1}^{T} x_{ti} x_{tj}.$$

The following conditions are often imposed on x_{ti} to investigate the asymptotic properties of the LSE.

(C_1) The limit

$$\rho_{ij} = \lim_{T\to\infty} M_{ij}^T / (\|x_i\|_T \|x_j\|_T)$$

exists for every i, j, $1 \le i, j \le l$, where $\|x_j\|_T = (m_{ii}^T)^{1/2}$.

(C_2) $\tilde{\rho}$ is nonsingular where $\tilde{\rho} = [\rho_{ij}]$.

Then we have the following result.

COROLLARY 2.1. *Under* (C_1) *and* (C_2), *if*

$$0 < \liminf_{t\to\infty} \|x_i\|_t^2 / t^{\S}, \qquad i = 1, 2, \dots, l,$$

for some $\S > 2d$, *then*

$$\lim_{T\to\infty} \hat{\beta}_T = \beta \quad a.s.$$

PROOF. Let

$$D_t = \operatorname{diag}(\|x_1\|_t, \|x_2\|_t, \dots, \|x_l\|_t)$$

and

$$G_t = D_t^{-1} V_t D_t^{-1}.$$

Then for any α,

$$\alpha' V_t \alpha \ge \lambda_{\min}(G_t) \alpha' D_t^2 \alpha \ge \lambda_{\min}(G_t) \min_{1\le i\le l} \|x_i\|_t^2,$$

which means

$$\mu_t \ge \lambda_{\min}(G_t) \min_{1\le i\le l} \|x_i\|_t^2.$$

Hence

$$\liminf_{t\to\infty} \mu_t / t^{\S} \ge \lambda_{\min}(\tilde{\rho}) \liminf_{t\to\infty} \min_{1\le i\le l} \|x_i\|_t^2 / t^{\S} > 0.$$

Therefore, μ_t satisfies conditions (i) and (ii) of Theorem 2.1. □

For example, $x_{ti} = \cos \nu_i t$, $\sin \nu_i t$ or t^{i-1}, $1 \le i \le l$, satisfies the assumptions of Corollary 2.1. [See Anderson (1971), pages 581–582.] In the special case of a stationary process with $\gamma_k = O(|k|^{-\delta})$, $\delta > 0$, and $l = 1$, $x_{t1} \equiv 1$, the strong consistency of the LSE, that is, the sample mean, is proved by Doob (1953), Theorem X.6.2, page 492.

Now we turn to the next problem. We pay attention to the polynomial regression, $X_t = (x_{t1}, x_{t2}, \dots, x_{tl})' = (1, t, \dots, t^{l-1})'$, namely,

$$y_t = \beta_1 + \beta_2 t + \cdots + \beta_l t^{l-1} + \varepsilon_t. \tag{9}$$

And we assume that $f^*(\lambda)$ in (2) is a positive continuous function, a stronger condition than $f^*(\lambda)$ is a nonnegative bounded function. Grenander (1954) proved the asymptotic efficiency of the LSE relative to the BLUE under some conditions. One of them is that the spectral density of the errors is positive and continuous but our spectral density $f(\lambda)$ of ε_t diverges to ∞ at $\lambda = 0$ because of the term, $1/|1 - e^{i\lambda}|^{2d}$. Hence we shall evaluate the asymptotic covariance matrices of the LSE and the BLUE and see how the LSE works in our model. First, we have the result on the LSE.

Theorem 2.2. *Let $\{y_t\}$ satisfy (9) and $\{\varepsilon_t\}$ be a stationary process with the spectral density $f(\lambda)$ of (2). Assume that $f^*(\lambda)$ is a positive continuous function. Then*

$$\lim_{T\to\infty} D_T E(\hat\beta_T - \beta)(\hat\beta_T - \beta)' D_T / T^{2d} = 2\pi f^*(0) M^{-1} H(d) M^{-1},$$

where the (i, j)th entries of M and $H(d)$ are

$$m_{ij} = \{(2i-1)(2j-1)\}^{1/2}/(i+j-1)$$

and

$$h_{ij}(d) = \left[\{(2i-1)(2j-1)\}^{1/2}\Gamma(1-2d)/\{\Gamma(d)\Gamma(1-d)\}\right] \times \int_0^1\int_0^1 x^{i-1} y^{j-1} |x-y|^{2d-1}\,dx\,dy.$$

Proof. We can assume that $f^*(\lambda) = (1/2\pi)|\theta(e^{i\lambda})|^2$, that is, $\{\varepsilon_t\}$ is a fractional ARIMA$(0, d, q)$ process since we can apply the same method as that of Theorem 2 of Grenander and Rosenblatt (1954) to a general $f^*(\lambda)$. $D_T E(\hat\beta_T - \beta)(\hat\beta_T - \beta)' D_T / T^{2d}$ is expressed as

$$\left(D_T^{-1} V_T D_T^{-1}\right)^{-1}\left\{D_T^{-1}\tilde X_T' R_T \tilde X_T D_T^{-1}/T^{2d}\right\}\left(D_T^{-1} V_T D_T^{-1}\right)^{-1},$$

where $\tilde X_T$ is the $T \times l$ matrix with (i, j)th entry x_{ij}. Then

$$\lim_{T\to\infty} D_T^{-1} V_T D_T^{-1} = M.$$

Next, noting (1), we have

$$\lim_{T\to\infty} D_T^{-1}\tilde{X}_T' R_T \tilde{X}_T D_T^{-1}/T^{2d} = |\theta(1)|^2 H(d). \qquad \square$$

Next, we shall consider the BLUE. This is an extension of the corollary of Theorem 5.1 of Adenstedt (1974), who considered the case $l = 1$, i.e., estimation of the mean.

THEOREM 2.3. *Let $\tilde{\beta}_T = (\tilde{\beta}_{1,T}, \tilde{\beta}_{2,T}, \dots, \tilde{\beta}_{l,T})'$ be the BLUE. Then*

$$\lim_{T\to\infty} D_T E(\tilde{\beta}_T - \beta)(\tilde{\beta}_T - \beta)' D_T/T^{2d} = 2\pi f^*(0) W^{-1}(d),$$

where $W(d) = (w_{ij}(d))$ and

$$w_{ij}(d) = \frac{\left[\Gamma(i-d)\Gamma(j-d)\{(2i-1)(2j-1)\}^{1/2}\right]}{\{\Gamma(i-2d)\Gamma(j-2d)(i+j-1-2d)\}}.$$

PROOF. Similar to Theorem 2.2, we can assume that $f^*(\lambda) = (1/2\pi)|\theta(e^{i\lambda})|^2$. Furthermore, we can assume that $x_{t1} = 1$ and $x_{tj} = \prod_{n=1}^{j-1}(t-n)$, $2 \le j \le l$, since the limit of $D_T E(\tilde{\beta}_T - \beta)(\tilde{\beta}_T - \beta)' D_T/T^{2d}$ is the same as that of $x_{tj} = t^{j-1}$, $1 \le j \le l$. Define $\sum_{j=1}^T \tilde{\tau}_j y_j$ by

$$\sum_{j=1}^{T} \tilde{\tau}_j y_j = \sum_{i=1}^{l} \chi_i \tilde{\beta}_{i,T},$$

for any constant χ_i. And put $\tilde{P}_T(z) = \sum_{j=1}^T \tilde{\tau}_j z^{j-1}$, $P_T(z) = \sum_{j=1}^T \tau_j z^{j-1}$, $\tilde{S}_T(z) = \tilde{P}_T(z)\theta(z)$ and $S_T(z) = P_T(z)\theta(z)$. Then $\tilde{S}_T(z)$ minimizes

$$(1/2\pi)\int_{-\pi}^{\pi} |S_T(e^{i\lambda})|^2/|1 - e^{i\lambda}|^{2d}\, d\lambda, \tag{10}$$

subject to the restraints

$$S_T^{(k-1)}(1) = \sum_{n=0}^{k-1} \binom{k-1}{n} \chi_{n+1} \theta^{(k-1-n)}(1), \qquad 1 \le k \le l. \tag{11}$$

Condition (11) is equivalent to that $\sum_{j=1}^T \tau_j y_j$ is an unbiased estimator of $\sum_{i=1}^l \chi_i \beta_i$. Now let $\eta_\nu(z)$, $\nu = 0, 1, 2, \dots$, be the orthonormal polynomials obtained from $1, z, z^2, \dots$, by the Gram–Schmidt procedure, where the inner product of two polynomials $g(z)$, $h(z)$ is defined by

$$(g, h) = (1/2\pi)\int_{-\pi}^{\pi} g(e^{i\lambda})\overline{h(e^{i\lambda})}/|1 - e^{i\lambda}|^{2d}\, d\lambda.$$

And let c_k be the right-hand side term of (11). Then the minimum of (10) is

$$\tilde{c}' W_T^{-1}(d)\tilde{c},$$

where $\tilde{c} = (c_1, c_2, \dots, c_l)'$ and $W_T(d)$ is the $l \times l$ matrix with (i, j)th entry $\sum_{\nu=0}^{T+g} \eta_\nu^{(i-1)}(1)\eta_\nu^{(j-1)}(1)$ [Grenander and Rosenblatt (1954), Theorem 1]. Put $\chi_i =$

$\|x_i\|_T\omega_i/T^d$. Then, using Lemma 2.1 stated in the Appendix,

$$\lim_{T\to\infty} \operatorname{Var}\left(\sum_{i=1}^{l} \omega_i\|x_i\|_T\tilde{\beta}_{i,T}/T^d\right)$$
$$= \lim_{T\to\infty} T^d\tilde{c}'D_T^{-1}\left(T^{2d}D_T^{-1}W_T(d)D_T^{-1}\right)^{-1}D_T^{-1}\tilde{c}T^d$$
$$= \theta^2(1)\tilde{\omega}'W^{-1}(d)\tilde{\omega},$$

where $\tilde{\omega} = (\omega_1, \omega_2, \dots, \omega_l)'$. Hence the proof is complete. □

Hence we see the variances of both estimators become larger by T^{2d} than those of a stationary process with a positive continuous spectral density (cf. Grenander (1954)].

We define the relative efficiency of the LSE by

$$e(d) = \lim_{T\to\infty} \det\left[E(\tilde{\beta}_T - \beta)(\tilde{\beta}_T - \beta)'\right]\Big/\det\left[E(\hat{\beta}_T - \beta)(\hat{\beta}_T - \beta)'\right].$$

Then Theorems 2.2 and 2.3 give, for example,

$$e(d) = \frac{(1+2d)\Gamma(1+d)\Gamma(2-2d)}{\Gamma(1-d)}, \qquad l = 1,$$
$$= \frac{36\Gamma^2(2-d)}{\left\{(1+2d)^2(3+2d)(3-2d)\Gamma^2(1+d)\Gamma^2(3-2d)\right\}}, \qquad l = 2.$$

The actual values of $e(d)$ are listed in Table 1.

The value for $d = 0.5$ is defined by $\lim_{d\to 0.5} e(d)$. It is seen from Table 1 that the LSE is not asymptotically efficient and its performance becomes poorer for $l = 2$ than $l = 1$ but achieves a high relative efficiency.

REMARK 2.2. The spectral density $f(\lambda; d, \phi, \theta)$ of a fractional ARIMA process clearly satisfies the assumptions of Theorems 2.1–2.3. The spectral density $f(\lambda; H)$ of a fractional Gaussian noise also satisfies these assumptions since it can be written in the form $f(\lambda; H) = f^*(\lambda; H)/|1 - e^{i\lambda}|^{2H-1}$ with a positive continuous function $f^*(\lambda; H)$ [cf. Geweke and Porter-Hudak (1983), Theorem 1].

3. Estimation of the parameters of the errors. In this section for simplicity let $\{\varepsilon_t\}$ be a fractional ARIMA(0, d, 0) process. Then its spectral density

TABLE 1
The relative efficiency $e(d)$ of the LSE

$l \setminus d$	0.0	0.1	0.2	0.3	0.4	0.5
1	1.000	0.995	0.987	0.982	0.985	1.000
2	1.000	0.986	0.956	0.925	0.901	0.889($=\frac{8}{9}$)

$f(\lambda; d)$ is written as $f(\lambda; d) = \sigma_a^2/(2\pi|1 - e^{i\lambda}|^{2d})$, where σ_a^2 is the innovation variance of the white noise, that is, $\sigma_a^2 = Ea_t^2$. And we assume that $\sigma_a^2 \in (0, \infty)$ and $d \in D = [\delta, 1/2 - \delta]$, $0 < \delta < \frac{1}{4}$. We shall construct an estimator for (d, σ_a^2) by substituting $\hat{\beta}_T$ for the unknown parameters β. Let $\tilde{d}_T$ be the d that minimizes

$$U_T(d, \hat{\beta}_T) = (1/2\pi)\int_{-\pi}^{\pi}\left|\sum_{t=1}^{T}(y_t - X_t'\hat{\beta}_T)e^{it\lambda}\right|^2 / g(\lambda; d)\, d\lambda,$$

where $g(\lambda; d) = |1 - e^{i\lambda}|^{-2d}$. Define $\tilde{\sigma}_{a,T}^2 = U_T(\tilde{d}_T, \hat{\beta}_T)/T$.

In the case that β is known, despite the lack of certain regularity conditions introduced by Walker (1964) and Hannan (1973), we have derived the strong consistency, the limiting distribution and the rate of convergence of this estimator [cf. Yajima (1985)].

Now we shall show that the same asymptotic properties still hold under some assumptions on X_t even if β is unknown. Hereafter, let d_0 and $\sigma_{a,0}^2$ be the true values of d and σ_a^2, respectively. We introduce the following conditions.

(C_3) $\{\varepsilon_t\}$ is an ergodic process.

(C_4) $E\{a_t|\mathscr{F}_{t-1}\}$, $E\{a_t^2|\mathscr{F}_{t-1}\}$, $E\{a_t^3|\mathscr{F}_{t-1}\}$ and $E\{a_t^4|\mathscr{F}_{t-1}\}$ are all constants a.s., where $\mathscr{F}_{t-1}$ is the sub σ-field generated by $\{a_s|s \le t-1\}$ and $E\{\cdot\,|\,\cdot\}$ is the conditional expectation.

Then we have the strong consistency of $(\tilde{d}_T, \tilde{\sigma}_{a,T}^2)$. The key to proving the result is to derive the rate of a.s. convergence of $\hat{\beta}_T$.

Theorem 3.1. *Under* (C_1)–(C_3), *if there are* $0 < K_1, K_2 < \infty$ *and* $\delta_i \ge 0$ *such that*

$$K_1 n^{\delta_i} \le \|x_i\|_n \le K_2 n^{\delta_i}, \qquad 1 \le i \le l,$$

then

$$\lim_{T\to\infty} \tilde{d}_T = d_0 \quad a.s. \quad \text{and} \quad \lim_{T\to\infty} \tilde{\sigma}_{a,T}^2 = \sigma_{a,0}^2 \quad a.s.$$

Proof. First, we show that

$$\lim_{T\to\infty}\{U_T(d, \hat{\beta}_T) - U_T(d, \beta)\}/T = 0 \quad \text{a.s.},$$

and the convergence is uniform in d for $d \in D$. In fact,

$$\begin{aligned}\{U_T(d, \hat{\beta}_T) &- U_T(d, \beta)\}/T \\ &= -(\pi T)^{-1}(\hat{\beta}_T - \beta)'D_T\int_{-\pi}^{\pi}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_t D_T^{-1}X_s e^{i(t-s)\lambda}\right\}\Big/ g(\lambda; d)\, d\lambda \\ &\quad + (2\pi T)^{-1}\int_{-\pi}^{\pi}\left|\sum_{t=1}^{T}(\hat{\beta}_T - \beta)'X_t e^{it\lambda}\right|^2\Big/ g(\lambda; d)\, d\lambda.\end{aligned}$$

Lemmas 3.1 and 3.2 in the Appendix ensure that the first term converges to 0 a.s.

as $T \to \infty$ and the convergence is uniform in d for $d \in D$. The second term is dominated by

$$K(\hat{\beta}_T - \beta)' D_T G_T D_T (\hat{\beta}_T - \beta)/T,$$

which is shown to converge to 0 a.s. as $T \to \infty$ by Lemma 3.1. Then the result is obtained by applying the same argument as in Hannan (1973) to $U_T(d, \beta)/T$. □

Next, we consider the case of (9), the polynomial regression, and derive the limiting distribution or the rate of convergence of $(d_T, \sigma^2_{a,T})$.

THEOREM 3.2. *Let* $\{y_t\}$ *satisfy* (9). *Assume the same conditions as in Theorem* 3.1.

(i) *If* $0 < d_0 < \frac{1}{4}$ *and* $\{a_t\}$ *satisfies* (C_4), $T^{1/2}(\tilde{d}_T - d_0, \tilde{\sigma}^2_{a,T} - \sigma^2_{a,0})$ *is asymptotically normal with zero mean vector and covariance matrix,*

$$A = \begin{pmatrix} 6/\pi^2 & 0 \\ 0 & 2\sigma^4_{a,0} + \kappa_{4,0} \end{pmatrix},$$

where $\kappa_{4,0} = Ea_t^4 - 3\sigma^4_{a,0}$.

(ii) *If* $d_0 = \frac{1}{4}$, *then*

$$\left(\tilde{d}_T - d_0, \tilde{\sigma}^2_{a,T} - \sigma^2_{a,0}\right) = O_p\left((\log T/T)^{1/2}\right).$$

(iii) *If* $\frac{1}{4} < d_0 < \frac{1}{2}$, *then*

$$\left(\tilde{d}_7 - d_0, \tilde{\sigma}^2_{a,T} - \sigma^2_{a,0}\right) = o_p\left(1/T^{1-2d_0}\right).$$

PROOF. We shall prove only the properties of $(\tilde{d}_T - d_0)$ since the results on $(\tilde{\sigma}^2_{a,T} - \sigma^2_{a,0})$ are obtained in the same way. We have

$$0 = U_T^{(1)}(d_0, \hat{\beta}_T) + U_T^{(2)}(d_T^*, \hat{\beta}_T)(\tilde{d}_T - d_0),$$

with $|d_T^* - d_0| \le |\tilde{d}_T - d_0|$, where $U_T^{(i)}(d, \beta)$, $i = 1, 2$ is the ith partial derivative with respect to d. Similar to Theorem 3.1, it is enough to show

$$\lim_{T \to \infty} \left\{U_T^{(2)}(d, \hat{\beta}_T) - U_T^{(2)}(d, \beta)\right\}/T = 0 \quad \text{a.s.,} \tag{12}$$

and the convergence is uniform in d for $d \in D$ and

$$p - \lim_{T \to \infty} \left\{U_T^{(1)}(d_0, \hat{\beta}_T) - U_T^{(1)}(d_0, \beta)\right\}/T^{1/2} = 0 \tag{13}$$

[cf. Yajima (1985)]. Assertion (12) can be proved in the same way as in Theorem 3.1. Put $h(\lambda; d) = g^{-1}(\lambda; d)$ and let $h^{(i)}(\lambda; d)$ be its ith partial derivative with respect to d. Then

$$\begin{aligned} &\left\{U_T^{(1)}(d_0, \hat{\beta}_T) - U_T^{(1)}(d_0, \beta)\right\}/T^{1/2} \\ &\quad = \pi^{-1}\left\{(\hat{\beta}_T - \beta)' D_T / T^{d_0}\right\} Z_{1,T} \\ &\qquad + (2\pi)^{-1}\left\{(\hat{\beta}_T - \beta)' D_T / T^{d_0}\right\} Z_{2,T} \left\{D_T(\hat{\beta}_T - \beta)/T^{d_0}\right\}, \end{aligned} \tag{14}$$

where

$$Z_{1,T} = T^{d_0-1/2}\int_{-\pi}^{\pi}\left\{\sum_{s=1}^{T}\sum_{t=1}^{T} D_T^{-1}X_s\varepsilon_t e^{i(t-s)\lambda}\right\}h^{(1)}(\lambda; d_0)\,d\lambda,$$

and

$$Z_{2,T} = T^{2d_0-1/2}D_T^{-1}\int_{-\pi}^{\pi}\left\{\sum_{s=1}^{T} X_s e^{is\lambda}\right\}\left\{\sum_{t=1}^{T} X_T' e^{-it\lambda}\right\}h^{(1)}(\lambda; d_0)\,d\lambda\, D_T^{-1}.$$

Now we shall evaluate the right-hand terms of (14). First, using (5), (C_1) and (C_2),

$$D_T(\hat{\beta}_T - \beta)/T^{d_0} = O_p(1). \tag{15}$$

Next, we evaluate $Z_{1,T}$ for each case of d_0. Define

$$\S_{T,t}(j) = (2\pi)^{-1}\int_{-\pi}^{\pi}\left(\sum_{s=1}^{T} x_{sj}e^{-is\lambda}/\|x_j\|_T\right)h^{(1)}(\lambda; d_0)e^{it\lambda}\,d\lambda,$$

$$H_{T,j}(\lambda) = (1/2)\left|\sum_{s=1}^{T} x_{sj}e^{-is\lambda}\right|^2/\|x_j\|_T^2.$$

Let $Z_{1,T}(j)$ be the jth component of $Z_{1,T}$. Then, using (5) and Parseval's formula,

$$\begin{aligned}\operatorname{Var}\big(Z_{1,T}(j)\big) &\le KT^{4d_0-1}\sum_{t=1}^{T}\S_{T,t}^2(j)\\ &\le KT^{4d_0-1}\pi^{-1}\int_{-\pi}^{\pi}H_{T,j}(\lambda)\left\{h^{(1)}(\lambda; d_0)\right\}^2 d\lambda\\ &= KT^{4d_0-1}Z_{3,T}(j),\quad \text{say.}\end{aligned}$$

Since $Z_{3,T}(j)$ is bounded,

$$\lim_{T\to\infty} T^{4d_0-1}Z_{3,T}(j) = 0, \tag{16}$$

for $d_0 < \frac{1}{4}$. Next $H_{T,j}(\lambda)$ has the property

$$\pi^{-1}\int_{-\pi}^{\pi}H_{T,j}(\lambda)\,d\lambda = 1,\qquad H_{T,j}(\lambda)\ge 0,\qquad -\pi\le\lambda\le\pi,$$

$$\lim_{T\to\infty}\ \sup_{\xi\le\lambda\le\pi} H_{T,j}(\lambda) = 0,\qquad 1\le j\le l,$$

for any $\xi > 0$, which is the same as that of the Fejér kernel. Hence

$$\lim_{T\to\infty} Z_{3,T}(j) = \left\{h^{(1)}(0; d_0)\right\}^2 = 0,$$

for $d_0 = \frac{1}{4}$. Finally, we consider the case of $\frac{1}{4} < d_0 < \frac{1}{2}$. Noting that $\{h^{(1)}(\lambda; d_0)\}^2$

satisfies a Lipschitz condition of order 1 and that

$$H_{T,j}(\lambda) \le T/2, \qquad -\pi \le \lambda \le \pi,$$

$$H_{T,j}(\lambda) \le K/(T\lambda^2), \qquad 0 < \lambda \le \pi,$$

we have

$$Z_{3,T}(j) = O(\log T/T)$$

[Zygmund (1959), Chapter 3, Theorem 3.15]. Thus relation (16) also holds for $\frac{1}{4} < d_0 < \frac{1}{2}$. Hence

$$p - \lim_{T\to\infty} Z_{1,T} = 0, \tag{17}$$

for $0 < d_0 < \frac{1}{2}$. Now we consider $Z_{2,T}$. Let $Z_{2,T}(j,k)$ be the (j,k)th entry of $Z_{2,T}$ and define $\tilde{Z}_{2,T}(j)$ by

$$\tilde{Z}_{2,T}(j) = 2T^{2d_0-1/2}\int_{-\pi}^{\pi} H_{T,j}(\lambda)|h^{(1)}(\lambda; d_0)|\,d\lambda.$$

Then

$$|Z_{2,T}(j,k)|^2 = \tilde{Z}_{2,T}(j)\tilde{Z}_{2,T}(k) \tag{18}$$

and

$$\tilde{Z}_{2,T}(j) = O(1/T^{1/2}) \tag{19}$$

[Zygmund (1959), Chapter 3, Theorem 3.15]. Noting (15)–(19), we complete the proof of (13). □

REMARK 3.1. Theorems 3.1 and 3.2 can be easily extended to the estimation of $(d, \sigma_a^2, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ of a general fractional ARIMA(p, d, q) model with $0 < d < \frac{1}{2}$. And as is seen from its proof, if $0 < d_0 < \frac{1}{4}$, Theorem 3.2 still holds for $\{X_t\}$ satisfying the assumptions of Theorem 3.1.

REMARK 3.2. Fox and Taqqu (1986) recently considered the model,

$$y_t = \beta_1 + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a stationary Gaussian process and its spectral density $f(\lambda;\theta)$ satisfies $f(\lambda;\theta) \sim |\lambda|^{-\alpha(\theta)}L_\theta(\lambda)$ as $\lambda \to 0$ with $0 < \alpha(\theta) < 1$ and $L_\theta(\lambda)$ varies slowly at $\lambda = 0$ and θ is a vector of unknown parameters. $f(\lambda;\theta)$ includes the spectral densities of a fractional Gaussian noise and a fractional ARIMA process. They proved the strong consistency and the asymptotic normality of the same estimator for θ as ours. Hence, combining their results with our consideration, we can show that Theorems 3.1 and 3.2 also hold for a stationary process with spectral density $f(\lambda;\theta)$. Furthermore, if we assume that $\{\varepsilon_t\}$ is Gaussian, a stronger condition than (C_3) and (C_4), the estimator is asymptotically normally distributed.

APPENDIX

LEMMA 2.1.

$$\lim_{T\to\infty} T^{2d} D_T^{-1} W_T(d) D_T^{-1} = W(d).$$

PROOF. We have

$$\eta_\nu(z) = -\sum_{j=0}^{\nu} \eta_{\nu,\nu-j} z^j / \sigma_\nu(d),$$

where

$$\eta_{k,j} = -\binom{k}{j} \Gamma(j-d)\Gamma(k-d-j+1)/\{\Gamma(-d)\Gamma(k-d+1)\}$$

and

$$\sigma_\nu^2(d) = \Gamma(\nu+1)\Gamma(\nu+1-2d)/\Gamma^2(\nu+1-d) \tag{A.1}$$

[Hosking (1981), page 168, and Yajima (1985), Lemma 3.2]. Hence

$$\eta_\nu^{(i-1)}(1) = \sum_{k=0}^{i-1} \S_k L_{(i-1)-k}(\nu+1, d)/\sigma_\nu(d),$$

where $\{\S_k\}$ are the constants with $\S_0 = -1$ and

$$L_m(\nu, d) = \sum_{j=0}^{\nu-1} \eta_{\nu-1,\nu-1-j}(j+1)^m.$$

Then, using (A.1) and the relation

$$(n+1)^{t-1} \le \Gamma(n+t)/\Gamma(n+1) \le n^{t-1}, \tag{A.2}$$

for $0 \le t \le 1$, $n = 1, 2, 3, \dots$, and applying Lemma A.1 in the following to $L_m(\nu, d)$, we have the result. □

LEMMA A.1.

$$L_m(\nu, d) = -\frac{\Gamma(1-2d)\Gamma(m+1-d)}{\Gamma(1-d)\Gamma(m+1-2d)} \nu^m c_{\nu-1}(d) + o(\nu^{m-d}),$$

where

$$c_t(d) = 1 - \sum_{j=1}^{t} \eta_{t,j}.$$

PROOF. We only sketch the outline. Since

$$\eta_{t,j} = \eta_{t-1,j} - \eta_{t,t}\eta_{t-1,t-j}, \qquad j = 1, \dots, t-1, \tag{A.3}$$

$$c_t(d) = (1 - \eta_{t,t}) c_{t-1}(d). \tag{A.4}$$

Hence

$$c_t(d) = \frac{\Gamma(1-d)\Gamma(t+1-2d)}{\Gamma(1-2d)\Gamma(t+1-d)}, \qquad t \geq 1. \tag{A.5}$$

Next, let

$$M_m(\nu, d) = \sum_{k=0}^{\nu} \eta_{\nu-1,k} k^m,$$

$$\psi_m(\nu, d) = L_m(\nu, d) + M_m(\nu, d),$$

$$\zeta_m(\nu, d) = L_m(\nu, d) - M_m(\nu, d).$$

Then, using (A.3),

$$\psi_m(\nu+1, d) = (1-\eta_{\nu,\nu})\psi_m(\nu, d) + \sum_{n=0}^{m-1}\binom{m}{n}\{(1-\eta_{\nu,\nu})\psi_n(\nu,d) + (1+\eta_{\nu,\nu})\zeta_n(\nu,d)\}/2,$$

$$\zeta_m(\nu+1, d) = (1+\eta_{\nu,\nu})\zeta_m(\nu, d) + \sum_{n=0}^{m-1}\binom{m}{n}\{(1-\eta_{\nu,\nu})\psi_n(\nu,d) + (1+\eta_{\nu,\nu})\zeta_n(\nu,d)\}/2.$$

Solving these difference equations with the help of (A.2), (A.4) and (A.5), we have

$$\psi_m(\nu, d) = -\frac{m!\Gamma(1-2d)\Gamma(m-d)}{\Gamma(1-d)\Gamma(m-2d)}\binom{\nu}{m}c_{\nu-1}(d) + o(\nu^{m-d}),$$

$$\zeta_m(\nu, d) = -\frac{m!\Gamma(1-2d)\Gamma(m-d)}{\Gamma(1-d)\Gamma(m+1-2d)}\binom{\nu-1}{m-1}(\nu-2d)c_{\nu-1}(d) + o(\nu^{m-d}).$$

Finally, we evaluate the term of the highest order, completing the proof. □

LEMMA 3.1. *Under* (C_1) *and* (C_2), *if* $\|x_i\|_n$ *satisfies the same condition as that of Theorem* 3.1, *then*

$$\lim_{T\to\infty} (D_T/T^{1/2})(\beta_T - \beta) = 0 \quad a.s.$$

PROOF. Let

$$S_n = (\|x_j\|_n/n^{1/2})(\hat{\beta}_{j,n} - \beta_j).$$

Then (C_1), (C_2) and relation (5) imply

$$ES_n^2 \leq K/n^{1-2d_0}.$$

Choosing ν and $n(m)$ so that $(1-2d_0)\nu > 1$ and $n(m)$ is the smallest integer that satisfies $n(m) \geq m^\nu$, we have

$$\sum_{m=1}^{\infty} ES_{n(m)}^2 \leq K \sum_{m=1}^{\infty} m^{-\nu(1-2d_0)} < \infty.$$

Hence

$$\lim_{m\to\infty} S_{n(m)} = 0 \quad \text{a.s.} \tag{A.6}$$

Next, we show that $\max_{n(m)\le n<n(m+1)}|S_n - S_{n(m)}|$ converges to 0 a.s. as $m \to \infty$. We have

$$\begin{aligned}
&\max_{n(m)\le n<n(m+1)} |S_n - S_{n(m)}| \\
&= \max_{n(m)\le n<n(m+1)} \left|S_n - \|x_j\|_n\left(\hat{\beta}_{j,n(m)} - \beta_j\right)/n^{1/2}\right| \\
&\quad + \max_{n(m)\le n<n(m+1)} \left|\|x_j\|_n\left(\hat{\beta}_{j,n(m)} - \beta_j\right)/n^{1/2} - S_{n(m)}\right| \\
&= Z_m(1) + Z_m(2), \quad \text{say.}
\end{aligned}$$

Let $V_i^{-1}X_i(j)$ be the jth component of $V_i^{-1}X_i$. Then, noting $V_i^{-1} = D_i^{-1}G_i^{-1}D_i^{-1}$ and the assumption, we have, with the help of (3) and (7),

$$\begin{aligned}
E\{Z_m(1)\}^2 &= E\left\{\max_n \left|\sum_{j=n(m)+1}^{n} \|x_j\|_n V_i^{-1}X_i(j)e_i/n^{1/2}\right|^2\right\} \\
&\le K\{\log 4(n(m+1) - n(m))/\log 2\}^2\{2n(m+1) - n(m)\}^{2d_0} \\
&\quad \times \left\{\|x_j\|^2_{n(m+1)}/n(m)\right\} \sum_{i=n(m)+1}^{n(m+1)} \left\{V_i^{-1}X_i(j)\right\}^2 v_i^2 \\
&\le K(\log m)^2/m^{\nu(1-2d_0)}.
\end{aligned}$$

Hence

$$\sum_{m=1}^{\infty} E\{Z_m(1)\}^2 < \infty,$$

which implies $\lim_{m\to\infty} Z_m(1) = 0$ a.s. On the other hand,

$$|Z_m(2)| \le \left(\|x_i\|_{n(m+1)}/\|x_i\|_{n(m)} + 1\right)|S_{n(m)}|.$$

Hence it follows from the assumption and (A.6) that $Z_m(2)$ converges to 0 a.s. as $T \to \infty$. Thus the proof is complete. □

LEMMA 3.2. *Under* (C_1)–(C_3), *if* $\|x_i\|_n$ *satisfies the same conditions as those of Theorem* 3.1, *then*

$$\lim_{T\to\infty} \left(1/T^{1/2}\right)\int_{-\pi}^{\pi}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T} \varepsilon_t D^{-1}X_s e^{i(t-s)\lambda}\right\}\Big/ g(\lambda; d)\, d\lambda = 0 \quad a.s.,$$

and the convergence is uniform in d *for* $d \in D$.

PROOF. Let

$$Z_T = (1/T^{1/2})\int_{-\pi}^{\pi}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_t x_{sj} e^{i(t-s)\lambda}/\|x_j\|_T\right\}\Big/ g(\lambda; d)\, d\lambda$$

and

$$\S_n(d) = (1/2\pi)\int_{-\pi}^{\pi} e^{-in\lambda}/g(\lambda; d)\, d\lambda.$$

Also let

$$\hat{g}_M^{-1}(\lambda; d) = \sum_{n=-M}^{M}(1 - |n|/M)\S_n(d)e^{in\lambda}.$$

Since $g^{-1}(\lambda; d)$ is continuous with respect to $(\lambda; d)$, we can choose M so that $|g^{-1}(\lambda; d) - \hat{g}_M^{-1}(\lambda; d)| < \zeta$ uniformly on $-\pi \le \lambda \le \pi$, $d \in D$. Then Z_T is bounded by

$$\zeta\left(\sum_{t=1}^{T}\varepsilon_t^2/T\right)^{1/2} + \left|2\pi\sum_{n=-M}^{M}\S_n(d)(1 - |n|/M)r_T(n)\right|,$$

where

$$\begin{aligned} r_T(n) &= \sum_{t=1}^{T-n}\varepsilon_t x_{t+n,j}/\left(\|x_j\|_T T^{1/2}\right), \qquad n \ge 0,\\ &= \sum_{t=1}^{T+n}\varepsilon_{t-n} x_{tj}/\left(\|x_j\|_T T^{1/2}\right), \qquad n < 0 \end{aligned}$$

[cf. Hannan (1971), page 774]. Hence it suffices to show that $\lim_{T\to\infty} r_T(n) = 0$ a.s. for any n. Define $S_{m,n}$ by

$$S_{m,n} = \sum_{t=m+1}^{n} c_t\varepsilon_t,$$

for any constants $\{c_t\}$. Then, similar to (6) and (7), we have

$$ES_{m,n}^2 \le K(n-m)^{2d_0}\sum_{t=m+1}^{n} c_t^2,$$

and, hence,

$$E\left\{\max_{1\le i\le r} S_{m,m+i}^2\right\} \le K(\log 4r/\log 2)^2(2r)^{2d_0}\sum_{t=m+1}^{m+r} c_t^2.$$

Then we can show that $\lim_{T\to\infty} r_T(n) = 0$ a.s. in the same way as in Lemma 3.1. □

Acknowledgments. The author is grateful to Mr. Stewart G. Hartley for improving the presentation of the paper. The author would like to thank the referees for their very helpful comments.

REFERENCES

ADENSTEDT, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Ann. Statist.* **2** 1095–1107.

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

ANDERSON, T. W. and TAYLOR, J. B. (1976). Strong consistency of least squares estimates in normal linear regression. *Ann. Statist.* **4** 788–790.

ANDERSON, T. W. and TAYLOR, J. B. (1979). Strong consistency of least squares estimates in dynamic models. *Ann. Statist.* **7** 484–489.

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis Forecasting and Control*, 2nd ed. Holden-Day, San Francisco.

DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

FOX, R. and TAQQU, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532.

GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long-memory time series models. *J. Time Ser. Anal.* **4** 221–238.

GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.* **1** 15–29.

GRENANDER, U. (1954). On estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Statist.* **25** 252–272.

GRENANDER, U. and ROSENBLATT, M. (1954). An extension of a theorem of G. Szegö and its application to the study of stochastic processes. *Trans. Amer. Math. Soc.* **76** 112–126.

HAMPEL, F. R., ROUSSEEUW, P. J., RONCHETTI, E. M. and STAHEL, W. A. (1986). *Robust Statistics*. Wiley, New York.

HANNAN, E. J. (1971). Non-linear time series regression. *J. Appl. Probab.* **8** 767–780.

HANNAN, E. J. (1973). The asymptotic theory of linear-time series models. *J. Appl. Probab.* **10** 130–145.

HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176.

LAI, T. L. and ROBBINS, H. (1977). Strong consistency of least squares estimates in regression models. *Proc. Nat. Acad. Sci. U.S.A.* **74** 2667–2669.

LAI, T. L., ROBBINS, H. and WEI, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci. U.S.A.* **75** 3034–3036.

LAI, T. L., ROBBINS, H. and WEI, C. Z. (1979). Strong consistency of least squares estimates in multiple regression. II. *J. Multivariate Anal.* **9** 343–361.

MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noise and applications. *SIAM Rev.* **10** 422–437.

MANDELBROT, B. B. and WALLIS, J. R. (1969). Some long-run properties of geophysical records. *Water Resources Res.* **5** 321–340.

SINAI, YA. G. (1976). Self-similar probability distributions. *Theory Probab. Appl.* **21** 64–80.

SOLO, V. (1981). Strong consistency of least squares estimators in regression with correlated disturbances. *Ann. Statist.* **9** 689–693.

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

WALKER, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Austral. Math. Soc.* **4** 363–384.

YAJIMA, Y. (1985). On estimation of long-memory time series models. *Austral. J. Statist.* **27** 303–320.

ZYGMUND, A. (1959). *Trignometric Series*. Cambridge Univ. Press, London.

DEPARTMENT OF INFORMATION SCIENCES
TOKYO INSTITUTE OF TECHNOLOGY
OH-OKAYAMA, MEGURO-KU
TOKYO 152
JAPAN