The Annals of Statistics
1980, Vol. 8, No. 2, 423–435

PARAMETER ESTIMATION OF AUTOREGRESSIVE INTEGRATED PROCESSES BY LEAST SQUARES

BY HIRONAO KAWASHIMA

Keio University, Yokohama

This paper deals with the asymptotic properties of so-called autoregressive integrated moving average processes. Moreover, it is shown that least squares estimates of the parameters of a Gaussian autoregressive integrated process are consistent and also best asymptotically normal.

1. Introduction. In their well known book, Box and Jenkins (1976) give autoregressive integrated moving average processes as mathematical models for prediction and control of nonstationary time series. The integrated process can be described by a constant parameter linear stochastic difference equation in which some of the characteristic roots are exactly equal to 1 in absolute value and other moduli of the roots are less than 1. Mann and Wald (1943) first considered the properties of least squares estimates of parameters in a stochastic difference equation when its characteristic roots are all less than 1 in absolute value or the case of stationary autoregressive process. Anderson (1959), White (1958), Rubin (1950) handled the first order difference equation with a root greater than 1 in absolute value. Recently, Stigum (1974, 1976) studied the case of higher order equations in which some of the moduli of the characteristic roots are greater than 1. However, the case of integrated process is omitted from their considerations and the limit distribution of least squares estimates of parameters were not handled even in the lower order case.

This paper deals with the asymptotic behavior of the integrated processes and especially for Gaussian autoregressive integrated processes, we show that least squares estimates of the parameters are consistent and best asymptotically normal. The basic propositions essential for the derivation are the following: first the parameters of integrated processes have special structure (see Lemma 1); secondly an integrated process has an exact representation by the convolution of the stationary process obtained from taking the difference of the original process and nonsummable series (see Lemma 2); thirdly a sample covariance function of the process divided by some powers of the sample size, converges to a constant (see Theorem 1).

Two basic relations are proved in Section 2 and the asymptotic properties of the process are discussed in Section 3. In Section 4, consistency and best asymptotic normality of least squares estimates of the parameters of a Gaussian autoregressive integrated process are proved.

Received November 1977; revised June 1978.

AMS 1970 *subject classifications.* Primary 62M10; secondary 62N15.

Key words and phrases. Autoregressive integrated moving average processes, least squares estimates, ergodic, spectral representation, Hardy class, consistency, best asymptotically normal.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.*
www.jstor.org

2. Preliminaries. We consider a discrete parameter, real valued random process $x(t) \equiv x(t, \omega)$, $\omega \varepsilon \Omega(P)$, where Ω is the basic probability space with the probability measure P. Let B be a shift operator defined by $Bx(t) = x(t-1)$. The pth difference of $x(t)$ is given by

$$\nabla^1 x(t) = x(t) - x(t-1) = (1-B)x(t),$$

$$\nabla^p x(t) = \nabla^{p-1}x(t) - \nabla^{p-1}x(t-1) = (1-B)^p x(t).$$

A nonstationary discrete parameter random process $x(t)$ is called an autoregressive integrated process of order (n, p), if the pth difference of $x(t)$ is a stationary autoregressive process of order n, that is

$$(1) \qquad (1 - a_1 B - \cdots - a_n B^n)\nabla^p x(t) = e(t), \qquad t = 0, 1, 2, \cdots,$$

where $e(t)$ is a zero mean white noise process with a common variance σ^2 and the roots of $(1 - a_1 B - \cdots - a_n B^n) = 0$ all exist outside the unit circle. (Note that the characteristic polynomial of (1) is given by $B^{-n} - a_1 B^{-(n-1)} - \cdots - a_n$.) Expanding $(1-B)^p$ by the powers of B and regrouping the coefficients of B^j, $j = 0, 1, \cdots, n+p$, we see that $x(t)$ satisfies the following stochastic difference equation.

$$(2) \qquad (1 - \alpha_1 B - \cdots - \alpha_{n+p} B^{n+p})x(t) = e(t), \quad t = 0, 1, 2, \cdots.$$

Now suppose we have observed the sample path of $x(t)$, where orders (n, p) are assumed to be known, from time 0 to $N + n + p - 1$. If we define a matrix X_N, vectors $\mathbf{y}_N$, $\boldsymbol{\alpha}$, $\mathbf{e}_N$ as given below, then (2) can be written as

$$(3) \qquad \mathbf{y}_N = X_N \boldsymbol{\alpha} + \mathbf{e}_N$$

where

$$\mathbf{y}_N = \begin{bmatrix} x(n+p) \\ \vdots \\ x(n+p+N-1) \end{bmatrix}, \quad X_N = \begin{bmatrix} x(n+p-1) & ,\cdots, & x(0) \\ \vdots & & \vdots \\ x(n+p+N-2) & ,\cdots, & x(N-1) \end{bmatrix}$$

$$\equiv [\mathbf{x}_N(n+p-1), \cdots, \mathbf{x}_N(0)]$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{n+p} \end{bmatrix}, \quad \mathbf{e}_N = \begin{bmatrix} e(n+p) \\ \vdots \\ e(n+p+N-1) \end{bmatrix}.$$

On the other hand, the sequence $\nabla^p x(t), p \leqslant t \leqslant N + n + p - 1$, can be calculated from the sequence $x(t)$, $0 \leqslant t \leqslant N + n + p - 1$, and hence we rewrite (1) to the following (4) by using the matrix ∇X_N and vectors $\nabla \mathbf{y}_N$, $\mathbf{a}$.

$$(4) \qquad \nabla \mathbf{y}_N = \nabla X_N \mathbf{a} + \mathbf{e}_N$$

where

$$\nabla \mathbf{y}_N = \begin{bmatrix} \nabla^p x(n+p) \\ \vdots \\ \nabla^p x(n+p+N-1) \end{bmatrix},$$

$$\nabla X_N = \begin{bmatrix} \nabla^p x(n+p-1) & ,\cdots, & \nabla^p x(p) \\ \vdots & & \vdots \\ \nabla^p x(n+p+N-2) & ,\cdots, & \nabla^p x(p+N-1) \end{bmatrix},$$

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

As we can see from the definitions of $\boldsymbol{\alpha}$ and $\mathbf{a}$, they are related to each other and the next lemma gives the exact relationship between the two parameter vectors.

LEMMA 1. *Let $x(t)$ be an autoregressive integrated process of order (n, p). Then we have*

$$\boldsymbol{\alpha} = D\mathbf{a} - Q\mathbf{d} \tag{5}$$

where

$$D = [\mathbf{d}, P\mathbf{d}, \cdots, P^{n-1}\mathbf{d}] \quad ((n+p) \times n \text{ matrix}),$$

$$\mathbf{d} = \begin{bmatrix} d_0 \\ \vdots \\ d_p \\ 0 \\ \vdots \\ 0 \end{bmatrix} ((n+p) \times 1 \text{ vector}),$$

$$P = \begin{bmatrix} 0, \cdots\cdots\cdots\cdots, 0 \\ 1, 0, \cdots\cdots\cdots\cdots \\ 0, 1, 0, \cdots\cdots\cdots \\ 0, 0, 1, 0, \cdots\cdots\cdots \\ \cdot \quad \ddots \quad \cdot \\ \cdot \quad \ddots \quad \cdot \\ \cdot \quad \ddots \quad \cdot \\ \cdot \quad \ddots \quad \cdot \\ \cdot \quad \ddots \quad \cdot \\ 0, \cdots\cdots\cdots 0, 1, 0 \end{bmatrix} ((n+p) \times (n+p) \text{ matrix}),$$

$$Q = P',$$

and d_i is the coefficient of B^i when expanding $(1 - B)^p$, i.e., $(1 - B)^p = d_0 + d_1 B + \cdots + d_p B^p$.

PROOF. Noting the definition of $\mathbf{d}$ and the property that $P^{n+p} = \Phi$ (a null matrix) holds, we can show (5) immediately.

EXAMPLE. When $p = 2$ and $n = 2$, we have $\alpha_1 = a_1 + 2$, $\alpha_2 = -2a_1 + a_2 - 1$, $\alpha_3 = a_1 - 2a_2$, $\alpha_4 = a_2$, which is equivalent to

$$\boldsymbol{\alpha} = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} a_1 + \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \end{bmatrix} a_2 - \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Next statement can be proved directly from Lemma 1 and the definition of $\nabla^p x(t)$.

COROLLARY 1. *Let $x(t)$ be an autoregressive integrated process of order (n, p). Suppose $x(t)$ are observed from time 0 to time $N + n + p - 1$ so that we can define $\mathbf{y}_N$ by* (3) *and $\nabla \mathbf{y}_N$ by* (4) *simultaneously. Then we have*

$$\text{(6)} \qquad \nabla X_N = X_N D, \quad \nabla \mathbf{y}_N = \mathbf{y}_N + X_N Q \mathbf{d}.$$

The least squares estimate $\hat{\boldsymbol{\alpha}}_N$ of $\boldsymbol{\alpha}$ based on $\mathbf{y}_N$ and X_N is given by $\hat{\boldsymbol{\alpha}}_N = (X_N' X_N)^{-1} X_N' \mathbf{y}_N$ and the estimated error is given by $\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_N = -(X_N' X_N)^{-1} X_N' \mathbf{e}_N$. Therefore, in order to evaluate the asymptotic properties of $\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_N$, we must know the behavior of $x(t)$ when t approaches to infinity, which we shall examine in the following section.

3. Asymptotic properties of integrated processes. A random process $x(t)$ is called an autoregressive integrated moving average process of order (n, p, m), if the pth difference of $x(t)$ satisfies

$$\text{(7)} \qquad \begin{aligned} &A(B)\nabla^p x(t) = B(B)e(t), \\ &A(B) = 1 - a_1 B - \cdots - a_n B^n, \; n \geqslant 0, \; a_n \neq 0, \\ &B(B) = 1 - b_1 B - \cdots - b_m B^m, \; m \geqslant 0, \; b_m \neq 0, \end{aligned}$$

where $e(t)$ is a zero mean white noise process with a variance σ^2, the roots of $A(B) = 0$, $B(B) = 0$ all exist outside the unit circle and two polynomials have no common factor. In this section we evaluate the asymptotic properties of integrated processes of the form (7). For this purpose, we give an exact representation of $x(t)$ in a slightly general form.

LEMMA 2. *Let $z(t)$ be an arbitrary finite valued sequence defined over $0 \leqslant t < \infty$. Then, for any fixed q, $0 \leqslant q < \infty$, $z(t)$ has a unique representation given by*

$$\text{(8)} \qquad z(t) = \textstyle\sum_{j=0}^{t-q} f_{j+1}^q \nabla^q z(t-j) + \sum_{j=0}^{q-1} g_j^t \nabla^j z(j), \qquad t \geqslant q,$$

where

$$f_j^q = \frac{1}{(q-1)!} \frac{(j+q-2)!}{(j-1)!}, \qquad j \geqslant 1, \qquad q \geqslant 1,$$

$$g_j^t = f_{t+1-j}^{j+1} = \frac{t!}{j!\,(t-j)!}.$$

PROOF. First we note that, since $\nabla^r z(t) = \nabla^{r+1}z(t) + \nabla^r z(t-1)$ is valid for all $r \geqslant 0$ and any $t > r$, the following equality holds.

$$\nabla^r z(t) = \Sigma_{j=0}^{t-(r+1)} \nabla^{r+1}z(t-j) + \nabla^r z(r). \tag{9}$$

We shall prove the representation (8) by mathematical induction. If we substitute $r = 0$ in (9), then we obtain the representation of $z(t)$ by $\nabla^1 z(t)$ and $z(0)$, that is

$$z(t) = \Sigma_{j=0}^{t-1}\nabla^1 z(t-j) + z(0).$$

On the other hand, it is easy to show $f_j^1 = 1, j = 1, 2, \cdots, t$ and $g_0^t = 1$. Hence, (8) holds for $q = 1$. Now suppose (8) is valid for some q. Substituting (9) to (8) for $r = q$ and changing the order of the summation, we obtain

$$z(t) = \Sigma_{j=0}^{t-(q+1)} (\Sigma_{i=0}^{j} f_{i+1}^q)\nabla^{q+1}z(t-j) + \{(\Sigma_{j=0}^{t-q} f_{j+1}^q)\nabla^q z(q) + \Sigma_{j=0}^{q-1} g_j^t \nabla^j z(j)\}.$$

From elementary calculations we have $\Sigma_{j=0}^{t-q} f_{j+1}^q = \dfrac{t!}{q!\,(t-q)!} \equiv g_q^t$, $\Sigma_{i=0}^{j} f_{i+1}^q = \dfrac{(j+q)!}{q!j!} \equiv f_{j+1}^{q+1}$. This shows that (8) holds for $q + 1$.

In the case of the integrated process defined by (7), we are only interested in the representation when $q = p$ in Lemma 2, that is

$$x(t) = \Sigma_{j=0}^{t-p} f_{j+1}^p \nabla^p x(t-j) + \Sigma_{j=0}^{p-1} g_j^t \nabla^j x(j) \equiv Y_t + Z_t. \tag{10}$$

Note that Y_t is defined by the convolution of the stationary process $\nabla^p x(t)$ and the nonsummable sequence f_{j+1}^p.

If the process $x(t)$ is Gaussian, then $\nabla^p x(t)$ is a strictly stationary process with $E\nabla^p x(t) = 0$ and $E\nabla^p x(t+\tau)\nabla^p x(t) = R(\tau)$. Since the spectral density function of an autoregressive moving average process is a rational function of $e^{-i\lambda}$, $\lim_{\tau\to\infty} R(\tau) = 0$ holds and this implies that $\nabla^p x(t)$ is ergodic (see Gikhman and Skorohod (1969) page 133). Therefore, for each $\omega\varepsilon\Omega$, we can define the quantity $\delta_N(k, \omega)$ by

$$\delta_N(k, \omega) \equiv \frac{1}{N}\Sigma_{j=0}^{N-1}\nabla^p x(j+k, \omega)\nabla^p x(j, \omega) - R(k). \tag{11}$$

Next lemma shows that the asymptotic behavior of $\delta_N(k, \omega)$ is uniform in k.

LEMMA 3. *Let $\nabla^p x(t, \omega)$ be a Gaussian stationary autoregressive moving average process. Then, for any sequence $\{k_N\}$*

$$\lim_{N\to\infty}\delta_N(k_N, \omega) = 0 \qquad \text{w.p. } 1,$$

where $\delta_N(k_N, \omega)$ is given by (11).

PROOF. We shall show that the inequality

$$E\Sigma_{N=1}^{\infty}|\delta_N(k_N, \omega)|^2 < \infty$$

holds independent of the choice of the sequence $\{k_N\}$. Since $\nabla^p x(t)$ is Gaussian,

$$E|\delta_N(k_N, \omega)|^2 = \frac{1}{N^2}\Sigma_{i=0}^{N-1}\Sigma_{j=0}^{N-1}\{R(i-j)^2 + R(i-j+k_N)R(i-j-k_N)\}$$

for any k_N. In view of the relation $|R(i)| \leqslant R(0)$, it follows that

$$(12)\qquad E|\delta_N(k_N, \omega)|^2 \leqslant \frac{R(0)}{N^2}\Sigma_{i=0}^{N-1}\Sigma_{j=0}^{N-1}(|R(i-j)| + |R(i-j+k_N)|)$$

for any k_N. The assumption that $\nabla^p x(t)$ is an autoregressive moving average process implies that there exist constants C and ρ such that

$$|R(t)| \leqslant C\rho^{|t|},$$

where $0 < \rho < 1$. If $k_N \geqslant 0$, then define l by $l = i + k_N$ and if $k_N < 0$, then define l by $l = j - k_N$. In both cases the inequality (12) becomes

$$E|\delta_N(k_N, \omega)|^2 \leqslant \frac{CR(0)}{N^2}\left\{\Sigma_{i=0}^{N-1}\Sigma_{j=0}^{N-1}\rho^{|i-j|} + \Sigma_{l=|k_N|}^{N-1+|k_N|}\Sigma_{m=0}^{N-1}\rho^{|l-m|}\right\}$$

$$\leqslant \frac{2CR(0)}{N^2}\left\{\Sigma_{j=0}^{\infty}\Sigma_{i=0}^{N-1}\rho^{|i-j|}\right\} \leqslant C_1\frac{1}{N}.$$

From this inequality and the assumption that the process $\nabla^p x(t)$ is Gaussian, we can show that for any sequence $\{k_N\}$, $\lim_{N\to\infty}\delta_N(k_N, \omega) = 0$ holds w.p. 1. This can be shown by following the proof of the law of large numbers for weakly stationary processes (see for example Doob (1953), Theorem 6.2).

Lemma 4. *Let $x(t)$ be an autoregressive integrated moving average process of order (n, p, m) and suppose $\Sigma_{j=0}^{p-1}E|x(j)|^2 < \infty$ holds. Then, for any fixed $l \geqslant p$, we have*

$$\Sigma_{k=l}^{l+N-1}|\Sigma_{j=0}^{p-1}g_j^k\nabla^j x(j)|^2 \leqslant C(\omega)N^{2p-1}, \qquad \text{w.p. 1},$$

where $C(\omega)$ is a constant which depends only on $\omega\varepsilon\Omega$.

Proof. Define C_{p-1} by $C_{p-1} = 2\max\{|d_0|, \cdots, |d_{p-1}|\}$. Then we have

$$\Sigma_{k=l}^{l+N-1}\left|\Sigma_{j=0}^{p-1}g_j^k\nabla^j x(j)\right|^2 \leqslant C_{p-1}^2\Sigma_{k=l}^{l+N-1}\Sigma_{i=0}^{p-1}\Sigma_{j=0}^{p-1}g_i^k g_j^k|x(i)|\,|x(j)|$$

$$\leqslant C_{p-1}^2D(\omega)^2\Sigma_{k=l}^{l+N-1}\left(\Sigma_{j=0}^{p-1}g_j^k\right)^2,$$

where $D(\omega) = \max_{0\leqslant j\leqslant p-1}\{|x(j, \omega)|^2\}$. From $\Sigma_{j=0}^{p-1}E|x(j)|^2 < \infty$, $D(\omega)^2$ is finite for each $\omega\varepsilon\Omega$, w.p. 1. On the other hand, we see that

$$\Sigma_{j=0}^{p-1}g_j^k = \Sigma_{j=0}^{p-1}\frac{k!}{j!\,(k-j)!} \leqslant p\frac{k!}{(k-(p-1))!}$$

holds for any $k \geqslant 1 \geqslant p$. Thus, we obtain

$$\Sigma_{k=l}^{l+N-1}\left|\Sigma_{j=0}^{p-1}g_j^k\nabla^j x(j)\right|^2 \leqslant p^2C_{p-1}^2D(\omega)^2\Sigma_{k=l}^{l+N-1}\{k(k-1)\cdots(k-p+2)\}^2$$

$$\leqslant C(\omega)N^{2(p-1)+1}.$$

This proves the lemma.

Now we are ready to prove

THEOREM 1. *Let $x(t)$ be a Gaussian autoregressive integrated moving average process of order (n, p, m). If $\Sigma_{j=0}^{p-1}E|x(j)|^2 < \infty$, then for any τ we have*

(13) $\lim_{N\to\infty}\mathbf{x}_N'(l+\tau)\mathbf{x}_N(l)N^{-(2p+1)} = \int_{-\pi}^{\pi}e^{i\lambda\tau}|H(e^{i\lambda})|^2 f(\lambda)d\lambda \equiv \hat{R}(\tau)$, w.p. 1,

where

(i) *$f(\lambda)$ is the spectral density function of $\nabla^p x(t)$ and*

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|\frac{B(e^{-i\lambda})}{A(e^{-i\lambda})}\right|^2,$$

(ii) *$x_N(l)$ is a $N \times 1$ vector defined by $\mathbf{x}_N'(l) = [x(l), x(l+1), \cdots, x(l+N-1)]$,*

(iii) *$H(e^{i\lambda})$ is defined by $H(e^{i\lambda}) = \text{l.i.m.}_{N\to\infty}H_N(e^{i\lambda})$ in $L^2_{(-\pi,\pi)}$, where $H_N(e^{i\lambda}) = \Sigma_{j=0}^{\infty}h_N(j)e^{i\lambda j}$ and*

$$h_N(j) = \begin{cases} f_{j+1}^p N^{-p}, & 0 \leqslant j \leqslant N-1, \\ 0, & j \geqslant N, j < 0. \end{cases} \tag{14}$$

PROOF. Since we have $\mathbf{x}_N'(l+\tau)\mathbf{x}_N(l) = \Sigma_{k=l}^{l+N-1}(Y_{k+\tau} + Z_{k+\tau})(Y_k + Z_k)$, where Y_k and Z_k are defined in (10), we shall first evaluate the asymptotic behavior of $\Sigma_{k=l}^{l+N-1}Y_{k+\tau}Y_k$. If we define the values of $\nabla^p x(j)$ by $\nabla^p x(j) = 0$ for $j < p$, then we have

$$\Sigma_{k=l}^{l+N-1}Y_{k+\tau}Y_k = \Sigma_{k=0}^{N-1}\Sigma_{j=0}^{N_1-1}f_{j+1}^p\nabla^p x(k+l+\tau-j)\Sigma_{i=0}^{N_2-1}f_{i+1}^p\nabla^p x(k+l-i),$$

where $N_1 = N + l + \tau - p$, $N_2 = N + l - p$. By interchanging the order of the summation and using the variable $\delta_N(k, \omega)$ defined by (11), we obtain

$$\Sigma_{k=l}^{l+N-1}Y_{k+\tau}Y_k = \Sigma_{j=0}^{N_1-1}f_{j+1}^p\Sigma_{i=0}^{N_2-1}f_{i+1}^p N[R(\tau-j+i) + \delta_N(\tau-j+i, \omega)]$$
$$\equiv A_1 + A_2.$$

In view of the spectral representation of $R(\tau)$, we can rewrite the term A_1 as

$$A_1 = N\int_{-\pi}^{\pi}e^{i\lambda\tau}\Sigma_{j=0}^{N_1-1}f_{j+1}^p\overline{e^{i\lambda j}}\Sigma_{k=0}^{N_2-1}f_{k+1}^p e^{i\lambda k}f(\lambda)d\lambda.$$

Moreover, since $\Sigma_{j=0}^{N-1}f_{j+1}^p = \frac{1}{p!}\frac{(N+p-1)!}{(N-1)!}$, we have $\|h_N\|_1 = \Sigma_{j=0}^{\infty}h_N(j) = \frac{1}{p!} + \frac{K}{N}$, where K is a constant. From the definition of $h_N(j)$ we can show immediately that $\lim_{N\to\infty}(\Sigma_{j=0}^{\infty}|h_{N+\nu}(j) - h_N(j)|^2)^{\frac{1}{2}} = 0$ holds for any $\nu > 0$. Therefore, there exists a function $H(e^{i\lambda})$ in $L^2_{(-\pi,\pi)}$ such that $\lim_{N\to\infty}\frac{1}{2\pi}\int_{-\pi}^{\pi}|H_N(e^{i\lambda}) - H(e^{i\lambda})|^2 d\lambda = 0$ and this implies $\lim_{N\to\infty}\int_{-\pi}^{\pi}|H_N(e^{i\lambda}) - H(e^{i\lambda})|^2 f(\lambda)d\lambda = 0$. Consequently, we have

$$\lim_{N\to\infty}A_1N^{-(2p+1)} = \int_{-\pi}^{\pi}e^{i\lambda\tau}|H(e^{i\lambda})|^2 f(\lambda)d\lambda \equiv \hat{R}(\tau), \qquad \text{w.p. 1}.$$

On the other hand, let $\{k_N^*\}$ be a sequence defined by

$$|\delta_N(k_N^*(\omega), \omega)| = \max_{0 \leqslant i \leqslant N_2 - 1,\, 0 \leqslant j \leqslant N_1 - 1}$$

$$\times \left| \frac{1}{N} \Sigma_{k=0}^{N-1} \nabla^p x(N + l + \tau - j - k) \nabla^p x(N + l - i - k) - R(\tau - j + i) \right|, \quad N = 1, 2, 3, \cdots, \tag{15}$$

for each fixed $\omega \varepsilon \Omega_1$, $\Omega_1 \subset \Omega$, $P(\Omega_1) = 1$. Since $\nabla^p x(j) = 0$ for $j < p$, we can rewrite (15) as

$$|\delta_N(k_N^*(\omega), \omega)| = \left| \frac{N^*}{N} \frac{1}{N^*} \Sigma_{k=0}^{N^*-1} \nabla^p x(N + k_N^* - k) \nabla^p x(N - k) - R(k_N^*) \right|,$$

where $N^* = \min\{N + l + \tau - j^* - p, N + l - i^* - p\}$ and $k_N^*(\omega) = \tau - j^* + i^*$. The result of Lemma 3 tells us that for any positive number ε, we can find a constant $M = M(\varepsilon, \omega)$ such that the inequality $|\delta_N(k_N, \omega)| < \varepsilon$, $(N > M)$, holds whenever $\omega \varepsilon \Omega_2$, $\Omega_2 \subset \Omega$, $P(\Omega_2) = 1$. Therefore, if $N^* > M$, then we obtain $|\delta_N(k_N^*(\omega), \omega)| < \varepsilon$ and if $N^* \leqslant M$, then we get $|\delta_N(k_N^*(\omega), \omega)| \leqslant MN^{-1}\max_{1 \leqslant n \leqslant M} |\delta_n(k_N^*(\omega), \omega)|$. Moreover, by following the procedures of Lemma 3 we can easily show that

$$\Sigma_{k=1}^{\infty} E\{\Sigma_{n=1}^{M} |\delta_n(k, \omega)| k^{-(1+2\beta)/2}\}^2 < \infty$$

holds for any fixed β, $0 < \beta < 1/2$. Thus, for any $\omega \varepsilon \Omega_3$, $\Omega_3 \subset \Omega$, $P(\Omega_3) = 1$, and $k_N^*(\omega)$, $|k_N^*(\omega)| < N + \tau + l - p$, we have $MN^{-1}\max_{1 \leqslant n \leqslant M} |\delta_n(k_N^*(\omega), \omega)| \leqslant C_M(\omega) N^{\beta - (\frac{1}{2})}$, where $C_M(\omega)$ is a finite constant which depends only on ω and M. Noting that the set Ω_2 can be determined independent of the choice of the sequence $\{k_N\}$, we can define a sequence $\{k_N^*(\omega^*)\}$ for each $\omega^* \varepsilon \cap_{i=1}^3 \Omega_i$, $P(\cap_{i=1}^3 \Omega_i) = 1$, and with this sequence we have

$$|\delta_N(k_N^*(\omega^*), \omega^*)| \leqslant C_1(\omega^*)(\varepsilon + N^{\beta - (\frac{1}{2})}), 0 < \beta < 1/2,$$

for any $N > M(\varepsilon, \omega^*)$. Therefore, we obtain

$$A_2 N^{-(2p+1)} \leqslant \Sigma_{j=0}^{\infty} |h_{N_1}(j)| \Sigma_{i=0}^{\infty} |h_{N_2}(i)| |\delta_N(k_N^*(\omega^*), \omega^*)| \leqslant C_2(\omega^*)(\varepsilon + N^{\beta - (\frac{1}{2})}).$$

This shows that $\lim_{N \to \infty} A_2 N^{-(2p+1)} = 0$, w.p. 1. Thus, we have

$$\lim_{N \to \infty} (\Sigma_{k=l}^{l+N-1} Y_{k+\tau} Y_k) N^{-(2p+1)} = \int_{-\pi}^{\pi} e^{i\lambda\tau} |H(e^{i\lambda})|^2 f(\lambda) d\lambda \equiv \hat{R}(\tau), \text{ w.p. } 1.$$

The result of Lemma 4 and this equality imply that

$$\lim_{N \to \infty} (\Sigma_{k=l}^{l+N-1} Y_{k+\tau} Z_k) N^{-(2p+1)} = 0, \qquad \text{w.p. } 1.$$

Hence we have $\lim_{N \to \infty} \mathbf{x}_N'(l + \tau)\mathbf{x}(l) N^{-(2p+1)} = \hat{R}(\tau)$, w.p. 1.

Now define a $(k+1) \times (k+1)$ matrix $\hat{\mathbb{R}}(k)$ by

$$\hat{\mathbb{R}}(k) = \begin{bmatrix} \hat{R}(0), & \hat{R}(1), & \cdots\cdots, & \hat{R}(k-1), & \hat{R}(k) \\ \vdots & & & & \vdots \\ \hat{R}(k), & \hat{R}(k-1), & \cdots\cdots, & \hat{R}(1), & \hat{R}(0) \end{bmatrix}.$$

Then, we obtain

THEOREM 2. *For any fixed $k \geqslant 0$, we have*

$$\det \hat{\mathbb{R}}(k) \neq 0.$$

PROOF. If $\det \hat{\mathbb{R}}(k) = 0$ holds, then there exists a nontrivial $(k+1) \times 1$ vector $\mathbf{u}' = [u_0, u_1, \cdots, u_k]$ such that $\hat{\mathbb{R}}(k)\mathbf{u} = \mathbf{0}$. Without restricting the generality, we may assume that $u_0 = 1$. Since $\hat{\mathbb{R}}(k)$ is a symmetric matrix, $\hat{\mathbb{R}}(k)\mathbf{u} = \mathbf{0}$ implies that

$$(16) \qquad \hat{R}(t) = u_1\hat{R}(t-1) + u_2\hat{R}(t-2) + \cdots + u_k\hat{R}(t-k)$$

holds for any $t \geqslant 0$ with the initial values $\hat{R}(k), \hat{R}(k-1), \cdots, \hat{R}(0)$. From the existence and the uniqueness theorems of the linear difference equation (16) and from the spectral representation (13) of $\hat{R}(\tau)$, we see that $\hat{R}(\tau)$ is the covariance function of some autoregressive process of order k (see Doob (1953) page 505). Hence, we have

$$|H(e^{i\lambda})|^2 f(\lambda) = \frac{\sigma_1^2}{2\pi}\left|\frac{1}{U(e^{-i\lambda})}\right|^2, \qquad \text{a.e. } \lambda,$$

where $U(e^{-i\lambda}) = 1 - u_1e^{-i\lambda} - \cdots - u_ke^{-i\lambda k}$ and σ_1 is a nonzero constant. Since $\hat{R}(\tau)$ is the covariance function of some stationary process, the zero points of $U(1/z)$, $z = re^{i\lambda}$, must lie inside the unit disk. Noting that $f(\lambda) \neq 0$ for $-\pi < \lambda \leqslant \pi$, we obtain

$$|H(e^{i\lambda})|^2 = \frac{\sigma_1^2}{\sigma^2}\left|\frac{A(e^{-i\lambda})}{B(e^{-i\lambda})U(e^{-i\lambda})}\right|^2 = \frac{\sigma_1^2}{\sigma^2}\left|\frac{A(e^{-i\lambda})}{V(e^{-i\lambda})}\right|^2$$

$$= \frac{\sigma_1^2}{\sigma^2}\left|\frac{\hat{A}(e^{i\lambda})}{\hat{V}(e^{i\lambda})}\right|^2 \equiv |\hat{H}(e^{i\lambda})|^2,$$

where $V(e^{-i\lambda}) = 1 - v_1e^{-i\lambda} - \cdots - v_{m+k}e^{-i\lambda(m+k)}$, $\hat{A}(e^{i\lambda}) = \overline{A(e^{-i\lambda})}$, $\hat{V}(e^{i\lambda}) = \overline{V(e^{-i\lambda})}$. Without restricting the generality, we may assume that the two polynomials $\hat{A}(z)$, $\hat{V}(z)$ have no common factor.

Now since $H_N(e^{i\lambda})$ belongs to $L^2_{(-\pi,\pi)}$ and $h_N(j) = 0$, $j < 0$ holds for any N, $H_N(z)$ belongs to the Hardy class H^2 in $|z| < 1$ for any N (see Duren (1970) page 38). From the completeness of the space H^2, $H(z)$ also belongs to H^2 in $|z| < 1$. In view of the canonical factorization theorem of H^2, we obtain

$$H(z) = C(z)S_1(z)S_2(z)$$

where

(i) $C(z)$ is the outerfunction of $H(z)$ given by

$$C(z) = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{i\lambda}+z}{e^{i\lambda}-z}\log|H(e^{i\lambda})|d\lambda\right];$$

(ii) $S_1(z)$ is the Blaschke product of $H(z)$ given by

$$S_1(z) = z^l\Pi_{k=1}^{\infty}\frac{|z_k|}{z_k}\frac{z_k - z}{1-\overline{z_k}z},$$

where l is a nonnegative integer and $\{z_k\}$ are the zero points of $H(z)$ in $|z| < 1$ and $z_k \neq 0$ for all k;

(iii) $S_2(z)$ is the singular inner function of $H(z)$ given by

$$S_2(z) = \exp\left[-\int_{-\pi}^{\pi}\frac{e^{i\lambda}+z}{e^{i\lambda}-z}d\mu(t)\right],$$

where $\mu(t)$ is a bounded nondecreasing singular function ($\mu'(t) = 0$, a.e.).

If we write $S_1(z)S_2(z) = z^lP(z)$, then $P(z)$ is analytic in $|z| < 1$ and satisfies $P(0) \neq 0$. This is because the sequence $\{z_k\}$ satisfies $z_k \neq 0$ for all k and $S_2(z)$ does not vanish inside the unit disk. Thus, $P(z)$ possesses the Taylor expansion $P(z) = \Sigma_{j=0}^{\infty}p_jz^j$ with $p_0 \neq 0$. Noting that $C(z)$ is also analytic, we have

$$H(z) \equiv \Sigma_{j=0}^{\infty}h(j)z^j = z^lC(z)P(z) = z^l\Sigma_{j=0}^{\infty}c_jz^j\Sigma_{j=0}^{\infty}p_jz^j \equiv z^l\Sigma_{j=0}^{\infty}r_jz^j,$$

and if we compare the lth power of z, then we obtain $h(l) = r_0 = c_0p_0$. Since $C(z)$ is an outerfunction and $|H(e^{i\lambda})| = |\hat{H}(e^{i\lambda})|$ holds a.e. λ, we obtain

$$c_0 = C(0) = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{\sigma_1}{\sigma}\left|\frac{\hat{A}(e^{i\lambda})}{\hat{V}(e^{i\lambda})}\right|d\lambda.\right.$$

Noting that the zeros of $\hat{A}(z)$ and $\hat{V}(z)$ all exist outside the unit disk and by Jensen's formula (see Ahlfors (1966) page 206) we obtain

$$c_0 = \exp\left[\log\frac{\sigma_1}{\sigma}\left|\frac{\hat{A}(0)}{\hat{V}(0)}\right|\right] = \frac{\sigma_1}{\sigma} \neq 0. \tag{17}$$

On the other hand, since we have $h(j) = \frac{1}{2\pi}\int_{-\pi}^{\pi}H(e^{i\lambda})e^{-i\lambda j}d\lambda$,

$$(j) - h_N|(j)| \leqslant \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}|H(e^{i\lambda}) - H_N(e^{i\lambda})|^2d\lambda^{\frac{1}{2}}\right.$$

holds by Parseval's relation. Thus, we see that $h(j) = \lim_{N\to\infty}h_N(j)$ holds for any j and in particular for $j = l$, we have

$$h(l) = \lim_{N\to\infty}\frac{f_{l+1}^p}{N^p} = \lim_{N\to\infty}N^{-p}\frac{1}{(p-1)!}\frac{(l+p-1)!}{l!} = 0, \tag{18}$$

from (14). Consequently, from the relations (17) and (18), we see that

$$0 = h(l) = p_0c_0 \neq 0$$

holds and this leads to a contradiction. Hence, we see that there is no nonzero vector $\mathbf{u}$ such that $\hat{\mathbb{R}}(k)\mathbf{u} = \mathbf{0}$ for any fixed $k \geq 0$.

4. Asymptotic properties of least squares estimates. In view of Lemma 1 and Corollary 1, we shall show in the next theorem that $\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_N$ can be approximated by the corresponding estimates calculated from $\nabla^p x(t)$.

THEOREM 3. *Let $x(t)$ be a Gaussian autoregressive integrated process of order (n, p). If $\Sigma_{j=0}^{p-1} E|x(j)|^2 < \infty$, then we have*

$$\lim_{N\to\infty} \|N^{\frac{1}{2}}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_N) - N^{\frac{1}{2}} D(\mathbf{a} - \hat{\mathbf{a}}_N)\| = 0, \qquad \text{w.p. } 1,$$

where $\hat{\mathbf{a}}_N = (\nabla X_N' \nabla X_N)^{-1} \nabla X_N' \nabla y_N$.

PROOF. From the relation (5), we obtain

$$\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_N = D(\mathbf{a} - \hat{\mathbf{a}}_N) - (\hat{\boldsymbol{\alpha}}_N + Q\mathbf{d} - D\hat{\mathbf{a}}_N)$$

by adding and subtracting the term $D\hat{\mathbf{a}}_N$. Noting the relation (6), we get

$$\begin{aligned}(\hat{\boldsymbol{\alpha}}_N + Q\mathbf{d} - D\hat{\mathbf{a}}_N) &= (X_N' X_N)^{-1} X_N' \nabla \mathbf{y}_N - D(\nabla X_N' \nabla X_N)^{-1} \nabla X_N' \nabla \mathbf{y}_N \\ &= (X_N' X_N)^{-1} X_N' I_N \nabla y_N,\end{aligned}$$

where $I_N = (E_N - \nabla X_N(\nabla X_N' \nabla X_N)^{-1} \nabla X_N')$. If we denote the Euclidean norm of a vector $\mathbf{x}$ by $\|\mathbf{x}\|$ and define the matrix norm of A by $\|A\| = (trAA')^{\frac{1}{2}}$, then we have

$$\begin{aligned}\|\hat{\alpha}_N + Qd - D\hat{a}_N\| &\leq \left\| \left(\frac{X_N' X_N}{N^{(2p+1)/2}} \right)^{-1} X_N' \right\| \|(I_N \nabla \mathbf{y}_N) N^{-(2p+1)/2}\| \\ &\leq \left\{ \operatorname{tr}\left(\frac{X_N' X_N}{N^{2p+1}} \right)^{-1} \right\}^{\frac{1}{2}} \|(I_N \nabla \mathbf{y}_N) N^{-\frac{1}{2}}\| N^{-p}.\end{aligned}$$

Since I_N is a $N \times N$ idempotent matrix and $\nabla^p x(t)$ is an ergodic stationary process, we have

$$\lim_{N\to\infty} \|(I_N \nabla \mathbf{y}_N) N^{-\frac{1}{2}}\| \leq \lim_{N\to\infty} \left(\frac{1}{N} \Sigma_{j=n+p}^{N+n+p-1} |\nabla^p x(j)|^2 \right)^{\frac{1}{2}} = R(0)^{\frac{1}{2}}.$$

Thus, we can show the existence of a finite constant $C(\omega)$ which only depends on ω, $\omega \varepsilon \Omega$, such that

$$\|N^{\frac{1}{2}}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_N) - N^{\frac{1}{2}} D(\mathbf{a} - \hat{\mathbf{a}}_N)\| \leq C(\omega) \operatorname{tr}\{\hat{\mathbb{R}}(n + p - 1)^{-1}\}^{\frac{1}{2}} R(0)^{\frac{1}{2}} N^{-p+\frac{1}{2}}.$$

This shows that the statement of the theorem holds.

We are now ready to prove

THEOREM 4. *Let $x(t)$ be a Gaussian autoregressive integrated process of order (n, p). If $\Sigma_{j=0}^{p-1} E|x(j)|^2 < \infty$, then we have*

$$\lim_{N\to\infty} \hat{\boldsymbol{\alpha}}_N = \boldsymbol{\alpha}, \qquad \text{w.p. } 1,$$

that is, $\hat{\boldsymbol{\alpha}}_N$ is a strongly consistent estimate of $\boldsymbol{\alpha}$.

PROOF. From Theorem 3, we have $\|(\alpha - \hat{\alpha}_N) - D(\mathbf{a} - \hat{\mathbf{a}}_N)\| \leqslant K(\omega)N^{-p}$, w.p. 1, where $K(\omega)$ is a finite constant which only depends on ω, $\omega\varepsilon\Omega$. On the other hand the estimated error of $\hat{\mathbf{a}}_N = (\nabla X_N' \nabla X_N)^{-1} \nabla X_N' \nabla y_N$ is given by

$$(\mathbf{a} - \hat{\mathbf{a}}_N) = -(\nabla X_N' \nabla X_N)^{-1} \nabla X_N' \mathbf{e}_N = -\left(\frac{\nabla X_N' \nabla X_N}{N}\right)^{-1} \frac{\nabla X_N' \mathbf{e}_N}{N}.$$

Since $\nabla^p x(t)$ is ergodic, we see that $\lim_{N\to\infty}\|\mathbf{a} - \hat{\mathbf{a}}_N\| = 0$, w.p. 1. These two facts show that

$$\lim_{N\to\infty}\|\hat{\alpha}_N - \alpha\| = 0, \qquad \text{w.p. } 1.$$

As a direct consequence of Theorem 3, we have

THEOREM 5. *Let $x(t)$ be a Gaussian autoregressive integrated process of order (n, p). If $\sum_{j=0}^{p-1} E|x(j)|^2 < \infty$, then the least squares estimate $\hat{\alpha}_N$ of α is a best asymptotically normal estimate, namely*

$$N^{\frac{1}{2}}(\alpha - \hat{\alpha}_N) \to \Phi(0, \sigma^2 D \Gamma_n^{-1} D') \text{ in law as } N \to \infty.$$

where

$$\Gamma_n = \begin{bmatrix} R(0), & R(1), & \cdots\cdots, & R(n-1) \\ R(1), & R(0), & \cdots\cdots, & R(n-2) \\ \vdots & & & \vdots \\ R(n-2), & R(n-3), & \cdots\cdots, & R(1) \\ R(n-1), & R(n-2), & \cdots\cdots, & R(0) \end{bmatrix}$$

and $R(\tau) = E\nabla^p x(t+\tau)\nabla^p x(t)$.

PROOF. From Theorem 4, $N^{\frac{1}{2}}(\alpha - \hat{\alpha}_N)$ converges to $N^{\frac{1}{2}}D(\mathbf{a} - \hat{\mathbf{a}}_N)$ w.p. 1. On the other hand $N^{\frac{1}{2}}(\mathbf{a} - \hat{\mathbf{a}}_N)$ converges in law to a zero mean normal random variable with covariance matrix $\sigma^2\Gamma_n^{-1}$ (see Whittle (1952)). Therefore, $N^{\frac{1}{2}}(\alpha - \hat{\alpha}_N)$ converges in law to a zero mean normal random variable with covariance matrix $\sigma^2 D\Gamma_n^{-1}D'$. Since $\hat{\mathbf{a}}_N$ is an asymptotically minimal variance estimate of the parameter $\mathbf{a}$ and in view of the relation (5), we see that $\hat{\alpha}_N$ is a best asymptotically normal estimate of α.

REFERENCES

[1] AHLFORS, L. V. (1966). *Complex Analysis*, 2nd Ed. McGraw-Hill, New York.

[2] ANDERSON, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30** 676–687.

[3] BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis-Forecasting and Control*, Revised Edition. Holden-Day, San Francisco.

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

[5] Duren, P. L. (1970). *Theory of H^p Spaces*. Academic Press, New York.
[6] Gikhman, I. I. and Skorohod, A. V. (1969). *Introduction to the Theory of Random Processes*. Saunders Company, Philadelphia.
[7] Mann, H. B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica*. **11** 173–220.
[8] Rubin, H. (1950). Consistency of maximum-likelihood estimates in the explosive case. In *Statistical Inference in Dynamic Economic Models* (Koopmans, T. C., ed.), Wiley, New York.
[9] Stigum, B. P. (1974). Asymptotic properties of dynamic stochastic parameter estimates (III). *J. Multivariate Anal*. **4** 351–381.
[10] Stigum, B. P. (1976). Least squares and stochastic difference equations. *J. Econometrics*. **4** 349–370.
[11] White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist*. **29** 1188–1197.
[12] Whittle, P. (1952). Estimation and information in stationary time series. *Arkiv. Für Mathematik* **2** 423–434.

Department of Administration Engineering
Keio University
3-14-1 Hiyoshi Kohoku-ku
Yokohama 223
Japan