The Annals of Statistics
1985, Vol. 13, No. 1, 369–382

ASYMPTOTIC PROPERTIES OF MINIMIZATION ESTIMATORS FOR TIME SERIES PARAMETERS

BY DANIEL MACRAE KEENAN

Brown University

1. Introduction. For a strictly stationary stochastic process in discrete time, a second-order parameter associated with the process can be viewed as a function of the spectral distribution function. Various authors (e.g. Whittle, 1953; Walker, 1964; Ibragimov, 1967; Hosoya, 1974; Taniguchi, 1979, 1981; and Hosoya and Taniguchi, 1982) have considered a certain estimator defined as a minimization solution of an integral expression. This work is concerned with estimators which are solutions of integral minimizations of which the above are special cases (analogous to the relationship of MLE to M-estimation). Asymptotic properties are shown for these estimators with the establishment of probability one bounds being the most novel contribution. The approach taken in this paper is to show that such estimators have almost sure representations as integrals of kernel functions w.r.t. the sample spectral distribution function and to invoke known results for the latter-type estimators (Keenan, 1983). An application of the results to the construction of a whole family of strongly consistent estimators of the dimension of a parameter is given.

2. Background. Let $\{X_i, -\infty < i < \infty\}$ be a strictly stationary stochastic process with mean zero. We will assume throughout that

ASSUMPTION 1.

$$\sum_{v_1, v_2, \cdots, v_{k-1}=-\infty}^{\infty} |v_j| \, |c(v_1, v_2, \cdots, v_{k-1})| < \infty \tag{2.1}$$

for $j = 1, 2, \cdots, k-1$, $k = 2, 3, \cdots$, where $c(v_1, v_2, \cdots, v_{k-1})$ is the kth order cumulant of $\{X(0), X(v_1), X(v_2), \cdots, X(v_{k-1})\}$ (see Brillinger, 1975, Section 2.6). In the case of a Gaussian process this condition is satisfied if

$$\sum_{v=-\infty}^{\infty} |v| \, |c(v)| < \infty.$$

The absolutely continuous spectral distribution function of the X_n process will be denoted by $F(\lambda)$, $\lambda \in [0, 2\pi]$. We will assume throughout that $f(\lambda) = dF(\lambda)/d\lambda > 0$, $\lambda \in [0, 2\pi]$. For a sample $\{X_1, X_2, \cdots, X_n\}$, the sample spectral distribution is defined as

$$F_n(\lambda) = \frac{2\pi}{n} \sum_s I_n\left(\frac{2\pi s}{n}\right), \quad 0 < \frac{2\pi s}{n} \le \lambda \tag{2.2}$$

where $I_n(\lambda)$ is the sample periodogram

$$I_n(\lambda) = (1/2\pi n) \left| \sum_{t=1}^{n} X_t e^{-it\lambda} \right|^2, \quad \lambda \in [0, 2\pi]. \tag{2.3}$$

Received January 1983; revised July 1984.
AMS 1980 *subject classifications.* Primary 60G10, Secondary 62M15.
Key words and phrases. Time series, minimization, probability one bounds.

Another empirical spectral distribution function is $F_n^*(\cdot)$, given by

$$F_n^*(\lambda) = \{\#\lambda_{(j)}^{(n)}\text{'s} \leq \lambda\} F_n(2\pi)/n \tag{2.4}$$

where $\lambda_{(j)}^{(n)} = \tilde{F}_n^{-1}(jF_n(2\pi)/n)$, $j = 1, 2, \cdots, n$, and $\tilde{F}_n(\cdot)$ is the piecewise linear version of $F_n(\cdot)$. The values $\{\lambda_{(j)}^{(n)}, j = 1, \cdots, n\}$ can be viewed as approximate order statistics from the spectral distribution F. The function $F_n^*(\cdot)$, like the usual empirical c.d.f. (in the i.i.d case), assigns equal weight to n points, thus making it highly compatible with linear operations. For example, minimization estimators (defined below) constructed with respect to F_n^* have the appearance of traditional M-estimators with $\lambda_{(j)}^n$, $j = 1, 2, \cdots, n$ serving as the data. By an approach analogous to that of Hosoya and Taniguchi (1982, Proposition 4.1 and Section 5) one and two-step estimators (via Newton's method) can be constructed. A third estimator of $F(\cdot)$, a continuous version of expression (2.2), is $\hat{F}_n(\cdot)$ defined as

$$\hat{F}_n(\lambda) = \int_0^\lambda I_n(\xi)\, d\xi. \tag{2.5}$$

One general perspective is to view a (second-order) characteristic of spectral distributions as a function defined on a class of spectral distributions. In this paper we will consider such functions (and their estimators) as being defined on a class of spectral distributions via an integral minimization (w.r.t. $t \in \Theta \subset \mathbb{R}^q$):

$$\int_0^{2\pi} h(\xi, t)\, d\xi + \int_0^{2\pi} m(\xi, t)\, dF(\xi)$$

where h and m are specified functions (which allow such a minimum). The most common (and important) example is where Θ is a parameterization of a family of spectral densities, $\mathscr{F} = \{f(\cdot, \theta), \theta \in \Theta \subset \mathbb{R}^q\}$, of which f, the spectral density of our process, is not necessarily an element, and h and m are functions of $f(\cdot, \theta)$ (e.g. expressions (2.11) and (2.16)); our results apply to this situation although this setup is not an assumption. That is, we may wish to consider integral expressions where h and m are not associated with a family of spectral densities; estimation of the "moments" of the spectral distribution, which are parameters associated with the expected number of zero-crossings of a process, is such an example.

The parametric setup ($\mathscr{F} = \{f(\cdot, \theta), \theta \in \Theta \subset \mathbb{R}^q\}$) is the most important and the remainder of this section is (primarily) concerned with it. An example which can be kept in mind throughout, although it is not an assumption, is that of an arbitrary linear process, possibly not having an ARMA representation,

$$X_n = \textstyle\sum_{j=0}^\infty a_j(\theta) e_{n-j}, \quad \{e_i\} \sim \text{i.i.d. } (0, \sigma^2)$$

where the coefficients are functions of θ, a parameter of finite dimension q, and consequently one can view the parameter as a function of the spectral distribution. In the Gaussian case, estimation has been via maximum likelihood (MLE). Whittle (1953) showed that in this case, maximum likelihood estimation of a parameter θ (a vector) is asymptotically equivalent to minimizing the integral

$$\int_0^{2\pi} [g(\lambda, \theta)]^{-1} I_n(\lambda)\, d\lambda \qquad \text{(W)} \tag{2.6}$$

where

$$g(\lambda, \theta) = (2\pi/\sigma^2) f(\lambda, \theta) \tag{2.7}$$

with σ^2 being the innovation variance assumed to not depend on θ, and $f(\cdot, \theta) \in \mathscr{F}$. Walker (1964) generalized the results of Whittle by showing that in general the least squares estimator (not assuming Gaussian, but a linear process generated by i.i.d. noise) is asymptotically equivalent to the solution of the above minimization problem and gave conditions for asymptotic normality and consistency (weak and strong) for the estimator; the estimator is asymptotically efficient with asymptotic variance

$$\left\{\frac{1}{4\pi} \int_0^{2\pi} \frac{\partial \log g(\lambda, \theta)}{\partial(\theta_i)} \frac{\partial \log g(\lambda, \theta)}{\partial(\theta_j)}\, d\lambda\right\}^{-1}. \tag{2.8}$$

We can consider the above minimization (expression (2.6)) as a special case of the following: for a continuous function

$$\rho\colon [0, 2\pi] \times \Theta \to \mathbb{R}$$

where Θ is a subset of $\mathbb{R}^q$ and where $M_G(\cdot)$ is defined as

$$M_G(t) = \int_0^{2\pi} \rho(\xi, t)\, dG(\xi) \tag{2.9}$$

with G a spectral distribution function, define $\theta(F)$, $\theta(F_n)$, and $\theta(\hat{F}_n)$ to be values (if they exist) which minimize expression (2.9) with F, F_n, and $\hat{F}_n$, respectively, substituted for G. The procedure is analogous to that of traditional M-estimation.

Ibragimov (1967) considered a different integral expression, one not associated with the likelihood function, and provided sufficient conditions for consistency (weak and strong) for $\hat{\theta}^{(n)}$, a value which minimizes

$$\int_0^{2\pi} (-\log f_1(\lambda, \theta)) I_n(\lambda)\, d\lambda, \qquad \text{(Ib)} \tag{2.10}$$

where $f_1(\cdot, \theta)$ is $f(\cdot, \theta)$ normalized to integrate to one. Ibragimov is only assuming weak stationarity and states that more stringent conditions need to be imposed to obtain the usual asymptotic behavior. One specific result of this work will be the obtainment of the asymptotic behavior of Ibragimov's estimator. Ibragimov (1967) showed that if $f(\lambda)$ is substituted for $I_n(\lambda)$ in expressions (2.6) and (2.10), then the resulting expressions reach their minimum at $\theta = \theta(F)$. The ρ functions for Whittle and Walker and for Ibragimov are, respectively,

$$[g(\xi, t)]^{-1} \quad \text{and} \quad -\log[f_1(\xi, t)]. \tag{2.11}$$

Asymptotic results will be derived for a family of ρ functions. For a linear (in the weak sense of martingale differences with constant conditional variance) process, Kabaila (1980) (and Whittle (1962) for linear process generated by i.i.d. noise) have shown both the asymptotic normality of estimators defined for a family of ρ functions (w.r.t. $\hat{F}_n(\cdot)$) and the optimality (in the sense of efficiency) of the (Whittle) estimator defined by expression (2.6). Table I (using Corollary 3.5 below) gives the asymptotic efficiency of Ibragimov's estimator (expression (2.10)) relative to that of Whittle, Walker, etc. (expression (2.6)) in the case of an AR(1) model ($X_n = \gamma X_{n-1} + e_n$). Because of Remark 3 (below) the only parameter which need be considered is γ (σ^2 and k_4 will have no effect on the asymptotic variances in these cases).

TABLE I
Ratio of asymptotic variance for estimator defined by expression (2.6) *to that by expression* (2.10)

γ	.1	.3	.5	.7	.9	.95
	.9901	.9174	.8000	.6711	.5526	.6195

Hosoya (1979) and Taniguchi (1983) have shown second-order efficiency (in different senses, respectively) of the minimization estimators defined by expressions (2.6) and (2.12) in the case of a Gaussian linear process. It may be, however, that criteria other than efficiency and/or processes other than linear (in its weak sense) should be of consideration, in which case these more general estimators may play a role.

Whittle and Walker considered a linear process for which the innovation variance σ^2 was assumed not to depend on the parameter (vector) θ. Hosoya (1974) pointed out that this is not always the case and proposed instead the minimization of

$$\int_0^{2\pi} \{\log f(\xi, \theta) + [f(\xi, \theta)]^{-1} I_n(\xi)\}\, d\xi \qquad \text{(H)} \tag{2.12}$$

rather than expression (2.6). Hosoya (1974) and Taniguchi (1979) derived the asymptotic distribution of this (quasi-maximum likelihood) estimator under regularity conditions. Hosoya and Taniguchi (1982) consider minimization of expression (2.12) for the situation where a family of parametric spectral densities $\{f(\xi, \theta), \theta \in \Theta \subset \mathbb{R}^q\}$ are fitted for a linear process with spectral density $\{f(\xi), \xi \in [0, 2\pi]\}$, not necessarily in the parametric family.

For ease of interpretation, we will also refer to expressions (2.6), (2.10), and (2.12) as (W), (Ib), and (H), respectively. We can include expression (H) in the general results by considering ρ functions of the form

$$\rho\colon [0, 2\pi] \times \Theta \times \mathbb{R}^+ \to \mathbb{R} \tag{2.13}$$

with

$$\rho(\xi, t, y) = h(\xi, t) y^{-1} + m(\xi, t) \tag{2.14}$$

and $M_G(t)$ in (2.9) is replaced by

$$M_G(t, g) = \int_0^{2\pi} \rho(\xi, t, g(\xi))\, dG(\xi) \tag{2.15}$$

where g is the derivative (generalized) of G. Expression (H) can be written with ρ given by

$$\rho(\xi, t, f(\xi)) = \log[f(\xi, t)][f(\xi)]^{-1} + [f(\xi, t)]^{-1}. \tag{2.16}$$

An equivalent (and informative although not as notationally convenient) representation of the integral expression (i.e. (2.15)) to be minimized (w.r.t. t) is:

$$a(\theta) + \left[\int_0^{2\pi} m(\lambda, t)\, dF_n(\lambda)\right]$$

(where $a(\theta) = \int_0^{2\pi} h(\xi, \theta]\, d\xi$) and minimization of the bracketed term, although typically w.r.t. $d\hat{F}_n(\cdot)$, having been extensively studied (e.g. expression (2.6)).

Therefore, for a whole family of ρ functions (again, not necessarily associated with a family of spectral densities, $\mathscr{F}$) we will define estimators (of a parameter θ) as minimization solutions of the integral of ρ w.r.t. $F_n(\cdot)$ and show various asymptotic results for these estimators.

Expressions (W), (Ib), and (H) are integrals with respect to the continuous estimator

$$\hat{F}_n(\lambda) = \int_0^{\lambda} I_n(\xi)\, d\xi \tag{2.17}$$

rather than w.r.t. $F_n(\cdot)$, given by expression (2.2). Under the assumptions on the process given above, the difference between the integral of a function of bounded variation with respect to $F_n(\cdot)$ and $\hat{F}_n(\cdot)$, respectively, is $O_p(n^{-1})$ (by Theorem 5.10.2, Brillinger, 1975, and Markov's Inequality). This is not a rate fast enough to pass properties via strong approximation from one to the other; probability one bounds will be established for estimators defined w.r.t. $F_n(\cdot)$ and $F_n^*(\cdot)$.

3. Strong consistency and asymptotic normality.

ASSUMPTION 2. Θ is a subset of $\mathbb{R}^q$, ρ is a function of the form of (2.14) such that $h(\xi, t)$ and $m(\xi, t)$ are continuous functions which are of bounded variation (in ξ) with $\sup_{t\in\Theta} \| m(\cdot, t) \|_v < \infty$, $\sup_{t\in\Theta} \| h(\cdot, t) \|_v < \infty$, and $\sup_{t\in\Theta} | m(2\pi, t) |_v < \infty$ where $\| \cdot \|_v$ is the total variation norm. It is assumed that

$$M_F(t, f) = \int_0^{2\pi} \rho(\xi, t, f(\xi))\, dF(\xi)$$

has a unique minimum at t_0 (to be referred to as $\theta(F)) \in \operatorname{int} \Theta$ with ρ continuous (in t) in a nbhd of t_0, N_{t_0}, such that for any $\delta > 0$ sufficiently small there exists a nbhd of t_0 such that off that neighborhood, $| M_F(t, f) - M_F(t_0, f) | > \delta$. We are not assuming that the true spectral density, f, is in a family parameterized by Θ.

Let $\|\cdot\|_\infty$ be the sup norm

$$\|H\|_\infty = \sup_{0\le\lambda\le 2\pi}|H(\lambda)|$$

where $H(\cdot)$ is a function on $[0, 2\pi]$.

Lemma 3.2 below shows that under the assumptions stated above for the X_n process, the minimimization estimator is strongly consistent. The following lemma is a slight restatement (i.e., discrete analogue) of a result due to Brillinger (1969, Corollary, Theorem 4.2).

LEMMA 3.1. *If* $\{X_n\}_{n=-\infty}^{\infty}$ *is a strictly stationary process satisfying assumption* (1), *then*

$$\|F_n - F\|_\infty \to 0 \quad w.p.1.$$

PROOF. Same proof as in Brillinger (1969) (Corollary to Theorem 4.2) except that reference to Theorem 4.2 is replaced by Theorem 5.10.1, Brillinger (1975).

LEMMA 3.2. *If* $\{X_i, -\infty < i < \infty\}$ *is a strictly stationary process with mean zero satisfying Assumption* 1 *and* ρ *is a function satisfying Assumption* 2, *then there exists a solution* T_n *(i.e.* $M_{F_n}(t, I_n)$ *has a minimum at* T_n*) such that* $(\theta(F_n) =)\ T_n \to t_0\ (= \theta(F))$ *with probability one.*

PROOF. For $t \in \Theta$

$$\begin{aligned} &|M_{F_n}(t, I_n) - M_F(t, f)| \\ &\quad = \left| \int_0^{2\pi} \rho(\xi, t, I_n(\xi))\, dF_n(\xi) - \int_0^{2\pi} \rho(\xi, t, f(\xi))\, dF(\xi) \right| \\ &\quad \le \left| \int_0^{2\pi} m(\xi, t)\, d[F_n(\xi) - F(\xi)] \right| \\ &\qquad + \left| \left[\frac{2\pi}{n} \textstyle\sum_{s=1}^n h\left(\frac{2\pi s}{n}, t\right)\right] - \int_0^{2\pi} h(\xi, t)\, dF \right| \end{aligned}$$

and since h and m are continuous functions of bounded variation (in ξ) and $F_n(\cdot) - F(\cdot)$ is a right continuous, bounded function, we have by Lemma 7.2.2B, page 254, Serfling (1980),

$$\begin{aligned} &|M_{F_n}(t, I_n) - M_F(t, f)| \\ &\qquad \le h(2\pi, t)\,|F_n(2\pi) - F(2\pi)| + \|m(\cdot, t)\|_v\, \|F_n - F\|_\infty + o(1). \end{aligned}$$

By Lemma 3.1 and the assumptions of the Lemma, we have

$$\sup_{t\in\Theta}|M_{F_n}(t, I_n) - M_F(t, f)| \to_{n\to\infty} 0 \quad \text{w.p.1.} \tag{3.1}$$

and the remainder of the proof is analogous to Theorem 1, Taniguchi (1979).

REMARK 1. If Θ is compact and ρ is continuous in ξ and t along with bounded variation in ξ and a unique minimum exists at t_0 then Assumption 2 (and consequently the assumptions of Lemma 3.2) is satisfied. However, often a compactness assumption is not appropriate, but the bounded variation conditions of Lemma 3.2 are satisfied (e.g. expression (2.6) in the AR(1) case). Hannan (1973) has given conditions for the existence of the asymptotic MLE when the parameter space is only required to be in a bounded set.

REMARK 2. The preceding Lemma and the flavor of Theorem 3.4 are motivated by Theorems 1 and 2, Taniguchi (1979) in which convergence in probability of time series analogues of minimum Hellinger distance estimators is considered.

COROLLARY 3.3. *In the statement of Lemma* 3.2, F_n^* *and* $\hat{F}_n$, *given by expressions* (2.4) *and* (2.5), *respectively, may be substituted for* F_n.

PROOF. For F_n^* the proof is the same except that in conjunction with Lemma 3.1 above, Lemma 3.4 Keenan (1982) is applied. For $\hat{F}_n$ the proof is the same except that reference to Lemma 3.1 is replaced by Corollary, Theorem 4.2, Brillinger (1969). □

THEOREM 3.4. *If* $\{X_i, -\infty < i < \infty\}$ *is a strictly stationary process with mean zero satisfying assumption* 1 *and* ρ *satisfies assumption* 2 *such that*

(i) $\partial m(\xi, t)/\partial t$ *and* $\partial^2 m(\xi, t)/\partial t \partial t'$ *are of bounded variation* (*component-wise*) *in* ξ *for all* t;

(ii) *there exists matrix-valued functions* g_1 *and* g_2 *with*

$$\left| \left(\frac{\partial^3 m(\xi, t)}{\partial \xi \partial t \partial t'} \right)_{j,\ell} \right| \le (g_1(\xi))_{j,\ell},$$
$$\left| \left(\frac{\partial^2 h(\xi, t)}{\partial t \partial t'} \right)_{j,\ell} \right| \le (g_2(\xi))_{j,\ell}, \quad 1 \le j, \quad \ell \le k \tag{3.2}$$

for t *in a neighborhood of* $\theta(F)$ *where*

$$\int_0^{2\pi} g_i(\xi)\, d\xi < \infty \text{ (element-wise)}, \quad i = 1, 2;$$

and

(iii) *the matrix* B *is nonsingular where* B *is given by*

$$B = -\left\{ \int_0^{2\pi} \frac{\partial^2 h(\xi, t)}{\partial t \partial t'}\Big|_{t=\theta(F)} d\xi + \int_0^{2\pi} \frac{\partial^2 m(\xi, t)}{\partial t \partial t'}\Big|_{t=\theta(F)} dF(\xi) \right\} \tag{3.3}$$

then

$$n^{1/2}[\theta(F_n) - \theta(F)] \\ = n^{1/2}A_nB^{-1}\left\{O(n^{-1}) + \int_0^{2\pi} \frac{\partial m(\xi, t)}{\partial t}\Big|_{t=\theta(F)} d[F_n(\xi) - F(\xi)]\right\} \tag{3.4}$$

where the components of the matrix $A_n \to 1$ *w.p.1. and the* $O(n^{-1})$ *term does not depend upon the realization.*

PROOF. The proof is analogous to Theorem 2 of Taniguchi (1979) with Lemma 3.2, above, giving the almost sure convergence.

COROLLARY 3.5. *Under the conditions of Theorem* 3.4 *we have* $n^{1/2}[\theta(F_n) - \theta(F)]$ *is asymptotically normal with mean zero and variance* Ω *given by*

$$\Omega = 2\pi(B')^{-1}\left\{\int_0^{2\pi}\left[\frac{\partial m}{\partial t}(2\pi - \alpha, \theta(F))\frac{\partial m}{\partial t'}(\alpha, \theta(F)) + \frac{\partial m}{\partial t}(\alpha, \theta(F))\frac{\partial m}{\partial t'}(\alpha, \theta(F))\right]f_{XX}^2(\alpha)\,d\alpha + \int_0^{2\pi}\int_0^{2\pi}\frac{\partial m}{\partial t}(\alpha, \theta(F))\frac{\partial m}{\partial t'}(\beta, \theta(F))f_{XXXX}(\alpha, \beta, -\alpha)\,d\alpha\,d\beta\right\}B^{-1} \tag{3.5}$$

where f_{XX} *and* f_{XXXX} *are the* 2*nd and* 4*th order cumulant spectra, respectively.*

PROOF. Apply Theorem 5.10.1, Brillinger (1975), to expression (3.4). □

NOTE. Corollary 3.5 will still hold if in expression (3.2) $\partial^3 m(\xi, t)/\partial\xi\partial t\partial t'$ is replaced by $\partial^2 m(\xi, t)/\partial t\partial t'$ and the assumption is made that $\partial^2 m(\xi, t)/\partial t\partial t'$ is continuous in ξ and t.

COROLLARY 3.6. *In the statements of Theorem* 3.4 *and Corollary* 3.5, F_n^* *and* $\hat{F}_n$ *may be substituted for* F_n.

PROOF. In the proof of Theorem 3.4, reference to Lemma 3.2 is replaced by Corollary 3.3. In the proof of Corollary 3.5, reference to Theorem 10.5.1, Brillinger (1975) is replaced by Theorems 10.5.1 and 10.5.2 Brillinger (1975) in the case of $\hat{F}_n$ and by the same argument as in Keenan (1982) (Theorem 3.5, in particular expression (3.11)) in the case of F_n^*. □

REMARK 3. For a linear process (i.e., a linear filtering of i.i.d. noise with variance σ^2 and 4th order joint cumulant k_4), the 4th order cumulant spectra in expression (3.9) factors as

$$f_{XXXX}(\alpha, \beta, -\alpha) = (k_4/(2\pi\sigma^4))f_{XX}(\alpha)f_{XX}(\beta)$$

and thus for minimization estimators defined for ρ such that $\rho(\xi, t) = m(\xi, t)$

(i.e. $h(\xi, t) \equiv 0$), as in expression (2.6) (Whittle) and 2.10 (Ibragimov), the latter integral in expression (3.5) is zero (under the conditions of Theorem 3.4), reflecting the well known fact (see Kabaila, 1980) that expression (2.6) (Whittle) gives the most efficient estimator within this class.

This work is concerned primarily with estimators defined w.r.t. $F_n(\cdot)$ and $F_n^*(\cdot)$. Theorem 3.4 gives more than just the asymptotic normality of our estimators; it states that, (w.p.1) along each realization, the difference between our estimators and the true value can be represented by integrals w.r.t. $[F_n(\cdot) - F(\cdot)]$. The asymptotic normality could have been established by using results of Kabaila (1980) and Theorems 10.5.1 and 10.5.2, Brillinger (1975). Under a weak form of linearity (and stationarity), Kabaila (1980) proves asymptotic normality for estimators which minimize integrals defined w.r.t. $\hat{F}_n(\cdot)$. However, for numerical computation, one ordinarily uses estimators defined w.r.t. to $F_n(\cdot)$. In the next section, probability one bounds are established for estimators defined as the minimization solution of integrals w.r.t. $F_n(\cdot)$ and $F_n^*(\cdot)$. The method uses Theorem 3.4 above and probability one bounds for $\| F_n - F \|_\infty$ and $\| F_n^* - F \|_\infty$ established in Keenan (1982). This same bound has not been proven for $\| \hat{F}_n - F \|_\infty$ (see also Malevich, 1965) and it does not appear that it can by this method be established for estimators defined w.r.t. $\hat{F}_n(\cdot)$ (see comments at the end of Section 2).

4. Probability one bounds. For the following theorem an additional assumption will be made concerning the kth order cumulants of the X_n process. The assumption is assumption 7.7.2 of Brillinger (1975), page 264. We will assume that C_k is finite for all $k \in N$ where C_k is defined as

$$C_k = \sum_{v_1, v_2, \cdots, v_{k-1}} | c(v_1, v_2, \cdots, v_{k-1}) |$$

where $c(v_1, v_2, \cdots, v_{k-1})$ is the kth order cumulant of $(X(0), X(v_1), \cdots, X(v_{k-1}))$. We will also assume that

ASSUMPTION 3.

$$\sum_{L=1}^{\infty} (\sum_\nu C_{n_1} C_{n_2} \cdots C_{n_p})(Z^L/L!) < \infty \tag{4.1}$$

for Z in a neighborhood of zero, where the inner summation is over all indecomposable partitions $\nu = (\nu_1, \cdots, \nu_p)$ of the table

$$\begin{matrix} 1 & 2 \\ 3 & 4 \\ 2L-1 & 2L \end{matrix}$$

with ν_j having $n_j > 1$ elements, $j = 1, \cdots, p$. In the case of a Gaussian process, this condition is satisfied if

$$\sum_{v=-\infty}^{\infty} | c(v) | < \infty.$$

THEOREM 4.1. *If, in addition to the conditions of Theorem 3.4, assumption*

(3) *is satisfied, then*

$$\text{(4.2)} \qquad \limsup_{n\to\infty} \frac{n^{1/2}\,|\,\theta_i(F_n) - \theta_i(F)\,|}{(2\Omega_{ii}\log n)^{1/2}}, \quad \le 1 \quad w.p.1, \quad i = 1, \cdots, q$$

where Ω_{ii} *is the* (i, i)th *element of the matrix* Ω *given by expression* (3.5) *and the subscript* i *indicates the* ith *component of the* q*-dim. vector. Expression* (4.2) *similarly holds with* F_n *replaced by* F_n^*.

PROOF. In Keenan (1982), Lemma 4.1, it is shown that, under conditions presently satisfied, $n^{1/2}\|F_n - F\|$ is $O((\log n)^{1/2})$, w.p.1, and that integration w.r.t $[F_n(\cdot) - F(\cdot)]$ of a function of bounded variation maintains that same rate. By Theorem 3.4, since the components of A_n converge to one w.p.1, it suffices to establish

$$\text{(4.3)} \qquad \limsup_{n\to\infty} \left(\frac{n}{2\Omega_{ii}\log n}\right)^{1/2} \left| \int_0^{2\pi} \frac{\partial m'}{\partial t}(\xi, \theta(F))b^{(i)}\, d[F_n(\xi) - F(\xi)\right| \le 1 \quad \text{w.p.1.}$$

where $b^{(i)}$ is the ith column of B^{-1}, B given by expression (3.3). By the same proof as in Theorem 4.4, Keenan (1982), it follows that expression (4.3) holds if $m(\xi, \theta(F))$ is a function of bounded variation in ξ (Ω_{ii} substituted for Λ, $(\partial m'(\xi, \theta(F)))/(\partial t)$ for $h_1(\xi)$ everything remains the same and the proof is done componentwise). As in Keenan (1982, Corollary 4.5), the result follows for F_n replaced by F_n^*. □

Klimko and Nelson (1978) and An, Chen, and Hannan (1982) have also considered probability bounds for estimators of time series parameters. Generally speaking, when the parameter depends on only a finite amount of the process (e.g., autocovariance of a fixed lag), law of the iterated logarithm results are typically available (see Phillipp, 1967, or the above two references). However, if this assumption is not made on the parameter, i.e. it can depend on the entire process, then Theorem 4.1 may still give a general law of the "uniterated" logarithm for estimators of such parameters. Spectral parameters typically are of this form.

5. An application to the estimation of the dimension of a parameter. Various authors have proposed methods for estimating the dimension of a parameter. In the case of autoregressive processes (of finite order), Akaike (1974) and Hannan and Quinn (1979) have proposed, respectively, choosing the value of K which minimizes

$$\text{(5.1)} \qquad \ln \hat{\sigma}_K^2 + 2K/n$$

$$\text{(5.2)} \qquad \ln \hat{\sigma}_K^2 + 2Kc \ln \ln n/n$$

where $c(>1)$ is an arbitrary constant, $\hat{\sigma}_K^2$ is the residual variance from fitting an AR(K) model for $K \le q$ (given), and n is the sample size. The ln ln n is a result of an application of a law of the iterated logarithm result, which was possible because of the autoregressive assumption (or more generally because of the determination of the parameter by a given finite dimensional joint distribution in the time domain (e.g., Hannan, 1980, extends expressions (5.2) and (5.3) to ARMA models)). Schwarz (1978) has proposed a general method which in the autoregressive time series setting reduces to expression (5.2) with ln ln n replaced by ln n:

$$\ln \hat{\sigma}_K^2 + 2Kc \ln n/n. \tag{5.3}$$

Below, we propose an estimator which does not depend on the parameter being determined by some finite dimensional joint distribution in the time domain. We will exploit the "uniterated" logarithm result of Theorem 4.1, in a manner motivated by Hannan and Quinn (1979).

Let $\Theta \subset \mathbb{R}^q$ be a compact subset with $\Theta^{(j)}$, $j = 1, 2, \cdots, q$, being an increasing sequence of compact subsets ($\Theta^{(j)} \subset \Theta^{(j+1)}$) with $\Theta^{(q)} = \Theta$. The dimension of a subset $A \subset \mathbb{R}^q$ is defined as the dimension of the affine hull (in $\mathbb{R}^q$) of A. Typically, we will have dim $(\Theta^{(j)})$ equal to j. For example, if Θ is the parameter space of an AR(q), then $\Theta^{(j)}$ is that of an AR(j), $1 \leq j \leq q$.

$$\Theta^{(j)} = \{\mathbf{a} = (a_1, \cdots, a_q) \mid \mathbf{a} \in \Theta, a_1 \neq 0, a_j \neq 0, a_{j+1} = \cdots = a_q = 0\}.$$

ASSUMPTION 4. The function ρ is given by expression (2.14) where $h(\xi, t)$ and $m(\xi, t)$ are continuous functions (in ξ and t) of bounded variation (in ξ) on $[0, 2\pi] \times \Theta$ and

$$M_F(t, f) = \int_0^{2\pi} \rho(\xi, t, f)\, dF(\xi)$$

has a unique minimum on Θ, denoted by $\Theta(F)$, which is an interior point of Θ. Denote by $\theta^{(j)}(F)$, a value in $\Theta^{(j)}$ such that

$$\mathrm{Min}_{t\in\Theta^{(j)}} M_F(t, f) = M_F(\theta^{(j)}(F), f).$$

DEFINITION. The dimension of $\theta(F)$ relative to $\{\Theta^{(j)}\}_{j=1}^q$ is defined as

$$K_0 = \dim(\theta(F)) = \mathrm{Min}\{1 \leq K \leq q \mid \theta(F) \in \Theta^{(K)}\}.$$

THEOREM 5.1. *If the conditions of Theorem* 4.1 *in addition to Assumption* 4 *and the condition that*

$$M_F(\theta^{(j)}(F), f) < M_F(\theta^{(j-1)}(F), f), \quad 2 \leq j \leq K_0.$$

are satisfied, then for $\hat{K}$ *corresponding to the*

$$\mathrm{Min}_{1\leq K\leq q}\{M_{F_n}(\theta^{(K)}(F_n), I_n) + 2KC(n^{-1}\log n)\}, \tag{5.4}$$

$\hat{K}$ *converges almost surely to* $K_0 = \dim(\theta(F))$ *where* C *and* $\theta^{(K)}(F_n)$ *are given by expressions* (5.6) *and* (5.5), *respectively.*

PROOF. By continuity and compactness, we have that there exists a (not necessarily unique for $j < K_0$) $\theta^{(j)}(F_n) \in \Theta^{(j)}$, $j = 1, 2, \cdots, q$, such that

$$\text{Min}_{t\in\Theta^{(j)}}[M_{F_n}(t, I_n)] = M_{F_n}(\theta^{(j)}(F_n), I_n). \tag{5.5}$$

Since $\Theta^{(j)} \subset \Theta^{(j+1)}$, we have for $1 \le K \le q$

$$\begin{aligned} M_{F_n}(\theta^{(K)}(F_n), I_n) &+ 2KCn^{-1}\log n \\ &= M_{F_n}(\theta^{(1)}(F_n), I_n) + 2Cn^{-1}\log n \\ &\quad + \textstyle\sum_{j=2}^{K} \{-[M_{F_n}(\theta^{(j-1)}(F_n), I_n) - M_{F_n}(\theta^{(j)}(F_n), I_n)] + 2Cn^{-1}\log n\} \end{aligned}$$

with $[M_{F_n}(\theta^{(j-1)}(F_n), I_n) - M_{F_n}(\theta^{(j)}(F_n), I_n)]$ being nonnegative. If $(j-1) \ge K_0$, then by Taylor's theorem and the fact that $\theta(F)$ is a relative interior point of $\Theta^{(\ell)}$, $\ell = K_0, K_0 + 1, \cdots, q$, we have (for n sufficiently large)

$$\begin{aligned} M_{F_n}(\theta^{(j-1)}(F_n), I_n) &- M_{F_n}(\theta^{(j)}(F_n), I_n) \\ &= \frac{\partial M'_{F_n}}{\partial\theta}(\theta, I_n)_{|\theta=\theta^{(j)}_{(F)}}(\theta^{(j-1)}(F_n) - \theta^{(j)}(F_n)) \\ &\quad + \frac{1}{2}(\theta^{(j-1)}(F_n) - \theta^{(j)}(F_n))' \frac{\partial^2 M_{F_n}(\theta, I_n)}{\partial\theta\partial\theta'}_{\theta=\tilde\theta}(\theta^{(j-1)}(F_n) - \theta^{(j)}(F_n)) \end{aligned}$$

where $\tilde\theta$ is an interior point on the line segment between $\theta^{(j-1)}(F_n)$ and $\theta^{(j)}(F_n)$. By the assumptions of the theorem,

$$\frac{\partial M}{\partial\theta}{}_{F_n}(\theta, I_n)_{|\theta=\theta^{(j)}_{(F)}} = 0$$

for n sufficiently large; applying Theorem 4.1 to $[\theta^{(j-1)}(F_n) - \theta(F)]$ and $[\theta^{(j)}(F_n) - \theta(F)]$, we have for $(j-1) \ge K_0$

$$0 \le \lim\sup_{n\to\infty}(n/2C\log n)[M_{F_n}(\theta^{(j-1)}(F_n), I_n) - M_{F_n}(\theta^{(j)}(F_n), I_n)] \le 1 \quad \text{w.p.1}$$

where

$$C = 2(\Omega_{11}^{1/2}, \cdots, \Omega_{qq}^{1/2})\left[\left|\frac{\partial^2 M_F(\theta(F), f)}{\partial\theta\partial\theta'}\right|\right](\Omega_{11}^{1/2}, \cdots, \Omega_{qq}^{1/2})'. \tag{5.6}$$

and $[|\ \ |]$ represents the matrix of absolute values. For $(j-1) < K_0$, with probability one

$$\begin{aligned} M_{F_n}(\theta^{(j-1)}(F_n), I_n) &- M_{F_n}(\theta^{(j)}(F_n), I_n) \\ &\to_{n\to\infty} M_F(\theta^{(j-1)}(F), f) - M_F(\theta^{(j)}(F), f) > 0 \end{aligned}$$

and by expression (3.1), this convergence does not depend on any uniqueness requirement of $\theta^{(K)}(F_n)$ and $\theta^{(K)}(F)$ for $K < K_0$. Therefore, the asymptotic minimum is reached at K_0, with probability one. □

In the case of Θ being the parameter space for a Gaussian autoregressive process of order $K_0 \le q$, and maximum likelihood estimation is used, then for ρ

given by expression (2.6), expression (5.4) reduces to

(5.7) $$\hat{\sigma}_K^2 + 2KC \ln n/n$$

in comparison to expressions (5.1), (5.2), and (5.3), where $\hat{\sigma}_K^2$ is replaced by $\ln \hat{\sigma}_K^2$. The reason is that expression (2.6) was chosen for minimization rather than expression (2.12); the latter is asymptotically equivalent to $(-2 \log \text{likelihood}/n)$.

Using the estimator proposed by Ibragimov (expression 2.10), expression (5.4) reduces to

$$\mathrm{Min}_{1\le k\le q}\left[-\frac{2\pi}{n}\sum_{s=1}^{n}\log f_1\left(\frac{2\pi s}{n}, \theta^{(k)}(F_n)\right)I_n\left(\frac{2\pi s}{n}\right)\right] + 2KC \ln n/n.$$

If $\{X_i \mid -\infty < i < \infty\}$ is a linear process of infinite order

(5.8) $$X_n = \sum_{j=0}^{\infty} a_j(\theta)e_{n-j}, \quad \{e_i\} \sim \mathrm{iid}(0, \sigma^2), \quad a_0 \equiv 1$$

where the coefficients are functions of $\theta \in \Theta^{(q)}$, then expression (5.4) gives a whole class of strongly consistent estimators (one for each ρ function) for the true dimension of θ.

REMARK 4. In practice, C is unknown, but because of Lemma 3.2, there exists a strongly consistent estimator, $\hat{C}_n$ (i.e. substitute $\theta^{(q)}(F_n)$ for $\theta(F)$ in expressions (3.3), (3.5), and (5.6)), of C and, therefore, if we replace C by $\hat{C}_n + \delta$, for some fixed, known constant $\delta > 0$, Theorem 5.1 is still valid. To illustrate the calculation of C, if (the simplest case) $q = 1$, the ρ function is that of Hosoya (expression (2.16)), and the process is given by expression (5.8), then

$$C = 8\pi \int_0^{2\pi}\left[\frac{\partial}{\partial\theta} g(\xi, \theta_0)^{-1}\right]^2 g(\xi, \theta_0)\, d\xi \Bigg/ \left|\int_0^{2\pi}\left[\frac{\partial^2}{\partial\theta^2} g(\xi, \theta_0)\right] g(\xi, \theta_0)\, d\xi\right|$$

where

$$g(\xi, \theta) = \left|\sum_{j=0}^{\infty} a_j(\theta)e^{ij\xi}\right|^2.$$

Note that in this case C is scale invariant.

Within this family of dimension estimators, it may be possible to define measures of optimality and to choose ρ functions which are optimal (possibly more so than maximum likelihood) in estimating the dimension of the parameter. For example, in a rather specialized instance, Taniguchi (1980) has shown that order selection of Akaike's information criterion gives asymptotic minimum mean square error of prediction.

6. Summary. An application of Theorem 4.1 gives probability one bounds for the estimators of Whittle and Walker (expression (2.6)), Hosoya (expression (2.12)), and Ibragimov (expression (2.10)). An application of Theorem 3.4 gives conditions for asymptotic normality of Ibragimov's estimator (expression (2.11)). Conditions for almost sure convergence (Lemma 3.2) of the above estimators are given which don't require the parameter set to be compact or contained within a bounded set. In conclusion, probability one bounds, asymptotic distributions and almost sure convergence have been established for a class of estimators defined via integral minimization.

REFERENCES

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* **AC-19** 716–723.

AN, H-Z., CHEN, Z-G. and HANNAN, E. J. (1982). Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.* **10** 926–936.

BRILLINGER, D. R. (1969). Asymptotic properties of spectral estimates of second order. *Biometrika* **56** 375–390.

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory.* Holt, New York.

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

HANNAN, E. J. (1980). The estimation of the order of an ARMA process. *Ann. Statist.* **8** 1071–1081.

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

HOSOYA, Y. (1974). Estimation problems on stationary time series models. Ph.D. Thesis, Yale University.

HOSOYA, Y. (1979). High-order efficiency in the estimation of linear processes. *Ann. Statist.* **7** 516–530.

HOSOYA, Y. and TANIGUCHI, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Statist.* **10** 132–153.

IBRAGIMOV, I. A. (1967). On maximum likelihood estimation of parameters of the spectral density of stationary time series. *Theory Probab. Appl.* **120** 115–119.

KABAILA, P. (1980). An optimality property of the least-squares estimate of the parameter of the spectrum of a purely nondeterministic time series. *Ann. Statist.* **8** 1082–1092.

KEENAN, D. M. (1983). Limiting behavior of functionals of the sample spectral distribution. *Ann. Statist.* **11** 1206–1217.

KLIMKO, L. and NELSON, P. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6** 629–642.

MALEVICH, T. L. (1965). Some properties of the estimators of the spectrum of a stationary process. *Theory Probab. Appl.* **10** 457–465.

PHILLIPP, W. (1967). Das Gesetz vom iterierten logarithmus fur stark mishende stationare prozesse. *Z. Wahrsch. verw. Gebiete* **8** 204–209.

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

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

TANIGUCHI, M. (1979). On estimation of parameters of Gaussian stationary processes. *J. Appl. Probab.* **16** 575–591.

TANIGUCHI, M. (1980). On selection of the order of the spectral density model for a stationary process. *Ann. Inst. Statist. Math.* **32** 401–419.

TANIGUCHI, M. (1981). An estimation procedure of parameters of a certain spectral density model. *J. Roy. Statist. Soc. B* **43** 34–40.

TANIGUCHI, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Ann. Statist.* **11** 157–169.

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.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Math. Astr. Fys.* **2** 423–434.

WHITTLE, P. (1962). Gaussian estimation in stationary time series. *Bull. Internat. Statist. Inst.* **39** 105–129.

DIVISION OF APPLIED MATHEMATICS
BROWN UNIVERSITY
PROVIDENCE, RHODE ISLAND 02912