The Annals of Statistics
1980, Vol. 8, No. 5, 1082–1092

AN OPTIMALITY PROPERTY OF THE LEAST-SQUARES ESTIMATE OF THE PARAMETER OF THE SPECTRUM OF A PURELY NONDETERMINISTIC TIME SERIES

BY PAUL V. KABAILA[1]

University of Newcastle, Australia

Whittle has proved that the least-squares estimator of a scalar parameter of the spectrum of a purely nondeterministic time series possesses a certain optimality property independently of the distribution of the residuals. In this paper we furnish a proof in full detail of the corresponding result for a vector parameter and also provide some examples which illustrate the application of the result.

1. Introduction. Whittle in [3] pages 428–429, proved a remarkable and important optimality property of the least-squares estimate of a scalar parameter of the spectrum of a purely nondeterministic time series. A somewhat different point of departure is taken in [5] pages 124–125, but essentially the same result is proved (for details see Remark 3.3). In [4] page 215 Whittle remarks that for a vector parameter a result similar to that for a scalar parameter is valid if the determinant of a positive semidefinite matrix is adopted as a measure of its "magnitude".

The optimality result is important from a theoretical point of view in that it exploits a somewhat remarkable analogy with the Cramér-Rao lower bound for unbiased estimators. The result is important from a practical point of view in that we do not need to assume knowledge of the distribution of the residuals.

Despite its obvious interest the result does not, however, appear to have been discussed much in the literature. In fact, it appears that neither is there available a proof in full detail of Whittle's result for a vector parameter nor are there many examples which might point-up the usefulness of the result.

Our purpose in this paper, then, is to furnish a proof in full detail for the case of a vector parameter and also to provide some examples which illustrate the application of the result. A secondary concern is to establish the result under very weak assumptions. For example, we do not need to assume strict stationarity of the process.

In Section 2 we introduce a class of consistent estimators of the parameter of the spectrum of a purely nondeterministic time-series. This class is shown to include the least-squares estimator.

In Section 3 the estimators introduced in Section 2 are shown to be asymptotically normally distributed with zero mean. A comparison of the covariance

Received April 1978; revised June 1979.

[1]Now at CSIRO, Division of Mathematics & Statistics, South Melbourne, Victoria 3205, Australia.

AMS 1970 *subject classification.* 62M10.

Key words and phrases. Time-series, least-squares, estimation efficiency.

matrices of the limit distributions shows that none of the estimators from the class introduced in Section 2 is more efficient than the least-squares estimator.

Now, if the class of estimators introduced in Section 2 were found to consist of the least-squares estimator alone, then the optimality result of Section 3 would be vacuous. In Section 4 we show how to construct a "large" set of estimators which is contained in the set introduced in Section 2. We also present, in detail, such a construction in a simple situation.

Finally, a word as to some of the conventions adopted in this paper. "$a =_\Delta b$" is meant to be read "a is defined to be the expression b". $a_n \to_{\text{a.s.}} b$, $a_n \to_{\text{prob}} b$ and $a_n \to_{\text{dist}} b$ are meant to indicate that as $n \to \infty, a_n$ converges to b almost surely, in probability and in distribution, respectively. $m^{(i)}(\lambda, \theta)$ denotes $\dfrac{\partial m(\lambda, \theta)}{\partial \theta_i}$ and so on for other derivatives. All integrals range from $-\pi$ to π and so we have chosen to omit the limits of integration. Various functions of θ and μ and/or λ will be introduced; it will be implicit that these functions are defined only for $\theta \in \Theta$ (introduced in Section 2), $\mu \in \boldsymbol{\mu}$ (introduced in Section 4) and $\lambda \in [-\pi, \pi]$.

2. A class of consistent estimators. Suppose the stochastic process $\{x_t\}$ defined on a probability space $(\Omega, \mathcal{B}, P)$ satisfies the following assumptions.

A1. $\{x_t\}$ is weakly stationary and has an absolutely continuous spectral distribution. We denote the spectral density by $f(\lambda)$.

A2. $\frac{1}{n}\sum_{t=1}^{n} x_t x_{t+m} \to_{\text{prob}} \gamma_m$ for each $m \in Z$.

Introduce Θ, a compact subset of a metric space $\mathfrak{M}$ with distance $d(\cdot, \cdot)$. Let Q denote the set $[-\pi, \pi] \times \Theta$. In the next two assumptions we introduce a class of spectral densities $\{f(\cdot, \theta) | \theta \in \Theta\}$ which is assumed to contain $f(\cdot)$.

C1. $f(\lambda, \theta) \geq 0$ for all $(\lambda, \theta) \in Q$. Suppose

$$\sigma^2(\theta) =_\Delta \exp\left\{\frac{1}{2\pi}\int \log 2\pi f(\lambda, \theta)\, d\lambda\right\} > 0$$

for each $\theta \in \Theta$ and define $m(\cdot, \theta) =_\Delta (\sigma^2(\theta)/2\pi f(\cdot, \theta))$. Suppose $m(\lambda, \theta)$ is continuous in $(\lambda, \theta) \in Q$.

C2. $f(\cdot) = f(\cdot, \theta_0)$ for a unique $\theta_0 \in \Theta$ (here two functions are considered equal if they differ at most on a set of λ-measure zero). We will denote $\sigma^2(\theta_0)$ by σ^2.

Let our aim be to estimate θ_0 on the basis of a part-realisation $x_1, \cdots, x_n$ and the prior information regarding the process $\{x_t\}$ embodied in assumptions A1, A2, C1 and C2. To this end we introduce $h(\lambda, \theta)$, a real-valued function of $(\lambda, \theta) \in Q$ which satisfies assumptions D1 and D2 defined below.

D1. $h(\lambda, \theta)$ is continuous in $(\lambda, \theta) \in Q$. Let $J(\theta) =_\Delta \int h(\lambda, \theta) f(\lambda)\, d\lambda$. Assumption D1 implies that $J(\theta)$ is continuous in θ.

D2. $J(\theta)$ is minimised at the single value of θ_0 whatever the value of θ_0 in Θ.

Now $m(\lambda,\theta)$ satisfies assumption D1 and $\int m(\lambda,\theta)f(\lambda)\,d\lambda$ is minimised at the single value of θ_0 (see e.g., [1] page 134) so that assumption D2 is also satisfied. $\tilde{\theta}_n$ defined to be a minimising value of $\tilde{S}_n(\theta) =_\Delta \int m(\lambda,\theta)I_n(\lambda)\,d\lambda$ where $I_n(\lambda) =_\Delta \frac{1}{2\pi n}|\Sigma_{t=1}^n x_t e^{i\lambda t}|^2$ is the "least-squares" (or "Gaussian" [5]) estimator of θ_0.

Two functions $h_1(\lambda,\theta)$ and $h_2(\lambda,\theta)$ will be said to be "equivalent" if there exist $c_1 > 0$, c_2 such that $h_1^+(\lambda,\theta) = c_1 h_2^+(\lambda,\theta) + c_2$ for all $(\lambda,\theta) \in Q$ where $h_1^+(\lambda,\theta) =_\Delta (h_1(\lambda,\theta) + h_1(-\lambda,\theta))/2$ and $h_2^+(\lambda,\theta) =_\Delta (h_2(\lambda,\theta) + h_2(-\lambda,\theta))/2$. This definition is natural in the present context because the subset of Θ minimising $\int h(\lambda,\theta)I_n(\lambda)\,d\lambda$ is the same for equivalent $h(\lambda,\theta)$'s.

As we shall see in Section 4 there exist functions which are not equivalent to $m(\lambda,\theta)$ but which satisfy assumptions D1 and D2. For any $h(\lambda,\theta)$ not equivalent to $m(\lambda,\theta)$ we define $\hat{\theta}_n$ to be a minimising value of $S_n(\theta) =_\Delta \int h(\lambda,\theta)I_n(\lambda)\,d\lambda$. $\hat{\theta}_n$ is yet another estimator of θ_0.

The main result of this section, Theorem 2.1, is that $\tilde{\theta}_n \to_{\text{prob}} \theta_0$ and $\hat{\theta}_n \to_{\text{prob}} \theta_0$. Thus, consistency of estimation does not provide a basis for deciding which of $\tilde{\theta}_n$ or $\hat{\theta}_n$ is a better estimator of θ_0. The proof of the main result of this section will require Lemmas 2.1, 2.2 and 2.3 which we now present.

LEMMA 2.1. *Suppose* $\{x_t\}$ *satisfies assumptions* A1, A2 *and* $h(\lambda,\theta)$ *satisfies assumptions* D1 *and* D2. *Then given* $\varepsilon > 0$, $\eta > 0$ *there exists an* $n_0 \in Z_+$ *such that for all* $n \geqslant n_0$

$$P\{\sup_{\theta\in\Theta}|S_n(\theta) - J(\theta)| > \varepsilon\} < \eta.$$

PROOF. Somewhat similar to the proof of Lemma 1 of Hannan [1]. Let $q_M(\lambda,\theta)$ be the Cesaro sum of the Fourier series of $h(\lambda,\theta)$ taken to M terms i.e.,

$$q_M(\lambda,\theta) =_\Delta \Sigma_{s=-M}^M \alpha_s(\theta)\left(1 - \frac{|s|}{M}\right)e^{is\lambda};\ \alpha_s(\theta) =_\Delta \frac{1}{2\pi}\int h(\lambda,\theta)e^{-is\lambda}\,d\lambda.$$

Also, let

$$\begin{aligned} J_1(\theta) &=_\Delta |S_n(\theta) - \int q_M(\lambda,\theta)I_n(\lambda)\,d\lambda| \\ J_2(\theta) &=_\Delta |\int q_M(\lambda,\theta)I_n(\lambda)\,d\lambda - \int q_M(\lambda,\theta)f(\lambda)\,d\lambda| \\ J_3(\theta) &=_\Delta |\int q_M(\lambda,\theta)f(\lambda)\,d\lambda - J(\theta)|. \end{aligned}$$

Our interest in these quantities is motivated by the fact that

$$|S_n(\theta) - J(\theta)| \leqslant J_1(\theta) + J_2(\theta) + J_3(\theta). \tag{2.1}$$

Now given $\varepsilon_1 > 0$ we may fix M so large that $|h(\lambda,\theta) - q_M(\lambda,\theta)| < \varepsilon_1$ uniformly in $(\lambda,\theta) \in Q$ since the Cesaro sum converges uniformly in $(\lambda,\theta) \in Q$. Hence $J_1(\theta) \leqslant \varepsilon_1 c_0$ and $J_3(\theta) \leqslant \varepsilon_1\gamma_0$. Further, a constant $K < \infty$ (which does not depend on either θ or M) can be found such that $J_2(\theta) \leqslant K\Sigma_{s=-M}^M|c_s - \gamma_s|$.

Consequently, by first fixing M sufficiently large, one can find an $n_0 \in Z_+$ such that $P\{\sup_{\theta\in\Theta}[J_1(\theta) + J_2(\theta) + J_3(\theta)] > \varepsilon\} < \eta$ for all $n \geqslant n_0$.

The proof of Lemma 2.2 will require the following well-known result relating convergence in probability and almost sure convergence.

RESULT 2.1. $x_n \rightarrow_{\text{prob}} x$ iff every subsequence of the x_j's contains a further subsequence which converges to x almost surely.

LEMMA 2.2. *Suppose that under a certain set of conditions (call these conditions* C*)*

$$a_n \rightarrow_{\text{a.s.}} a, b_n \rightarrow_{\text{a.s.}} b, \cdots, r_n \rightarrow_{\text{a.s.}} r$$

implies that $y_n \rightarrow_{\text{a.s.}} y$. *Then again under conditions* C

$$a_n \rightarrow_{\text{prob}} a, b_n \rightarrow_{\text{prob}} b, \cdots, r_n \rightarrow_{\text{prob}} r$$

implies that $y_n \rightarrow_{\text{prob}} y$.

PROOF. Result 2.1 implies that we can prove that $y_n \rightarrow_{\text{prob}} y$ by proving that any subsequence of $\{n\}$, denoted $\{n_i\}$ has a further subsequence n_r such that $y_{n_r} \rightarrow_{\text{a.s.}} y$.

Obviously $a_{n_i} \rightarrow_{\text{prob}} a$ hence by Result 2.1 $\{n_i\}$ has a subsequence, denoted $\{n_a\}$ such that $a_{n_a} \rightarrow_{\text{a.s.}} a$.

Obviously, then, $b_{n_a} \rightarrow_{\text{prob}} b$ hence by Result 2.1 $\{n_a\}$ has a subsequence, denoted $\{n_b\}$ such that $b_{n_b} \rightarrow_{\text{a.s.}} b$.

We continue in this fashion till we find a subsequence of $\{n\}$, denoted $\{n_r\}$ such that $a_{n_r} \rightarrow_{\text{a.s.}} a, b_{n_r} \rightarrow_{\text{a.s.}} b, \cdots, r_{n_r} \rightarrow_{\text{a.s.}} r$. By hypothesis this implies that $y_{n_r} \rightarrow_{\text{a.s.}} y$. Hence as already indicated $y_n \rightarrow_{\text{prob}} y$.

LEMMA 2.3. *Suppose* $b(\theta)$ *and* $a_1(\theta), a_2(\theta), \cdots$ *are continuous in* $\theta \in \Theta$ *a compact subset of a metric space* M *with distance* $d(\cdot, \cdot)$. *Suppose also that* $b(\theta)$ *is minimised at a single value of* $\theta \in \Theta$ *and let this value be denoted* θ_0. *Further suppose* $a_n(\theta) \rightarrow_{\text{a.s.}} b(\theta)$ *uniformly in* Θ. *Then if* θ_n *is any minimising value of* $a_n(\theta)$ $d(\theta_n, \theta_0) \rightarrow_{\text{a.s.}} 0$.

PROOF. The method of proof is indicated by the method of proof of Theorem 1 of Hannan [1].

We now present the main result of this section.

THEOREM 2.1. *Under assumptions* A1, A2, C1, C2, D1 *and* D2 $d(\hat{\theta}_n, \theta_0) \rightarrow_{\text{prob}} 0$ *and* $d(\tilde{\theta}_n, \theta_0) \rightarrow_{\text{prob}} 0$.

PROOF. Let z_n denote $\sup_{\theta \in \Theta} |S_n(\theta) - J(\theta)|$. Lemma 2.3 states that when $z_n \rightarrow_{\text{a.s.}} 0$ then $d(\hat{\theta}_n, \theta_0) \rightarrow_{\text{a.s.}} 0$. However, Lemma 2.1 implies that $z_n \rightarrow_{\text{prob}} 0$. Hence by Lemma 2.2 $d(\hat{\theta}_n, \theta_0) \rightarrow_{\text{prob}} 0$. Similarly $d(\tilde{\theta}_n, \theta_0) \rightarrow_{\text{prob}} 0$.

3. The main result. In Section 2 it was proved that $\tilde{\theta}_n \rightarrow_{\text{prob}} \theta_0$ and $\hat{\theta}_n \rightarrow_{\text{prob}} \theta_0$. Consequently, consistency of estimation does not provide a basis for deciding which of $\tilde{\theta}_n$ and $\hat{\theta}_n$ is a better estimator of θ_0. By introducing additional assumptions on $\{x_t\}$, $f(\lambda, \theta)$ and $h(\lambda, \theta)$ we are able to prove in this section that both $n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0)$ and $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)$ converge in distribution to normal distributions with zero means. A comparison of the covariance matrices of the limiting distributions shows that $\hat{\theta}_n$ cannot be a more efficient estimator of θ_0 than $\tilde{\theta}_n$. This is the main result of the paper.

Suppose $\{x_t\}$ satisfies the following assumptions (note that A3–A5 and B1 are introduced in [1] and A6 appears in [2]).

A3. $x_t = \Sigma_u l_u \varepsilon_{t-u}$ where $l_0 = 1$ and $l_u = 0$ for $u < 0$.

A4. $E\{\varepsilon_n | \mathfrak{F}_{n-1}\} = 0$ a.s. for all n where $\mathfrak{F}_n$ is the σ-field generated by $\{\varepsilon_n, \varepsilon_{n-1}, \cdots\}$

A5. $E\{\varepsilon_n^2 | \mathfrak{F}_{n-1}\} = \sigma^2 > 0$ a.s.

A6. Suppose there exists a random variable X with $E\{X^4\} < \infty$ such that $P\{(|\varepsilon_n| > u) \leqslant cP(|X| > u)$ for some $0 < c < \infty$ and all n, all $u \geqslant 0$.

B1. $\Sigma_u u l_u^2 < \infty$.

A process satisfying assumptions A3–A5 and B1 also satisfies assumption A1. By a result of Hannan and Heyde [2] a process satisfying the above assumptions also satisfies assumption A2. Consequently, the above set of assumptions on $\{x_t\}$ is more restrictive than assumptions A1 and A2.

For the derivation of the limiting distribution of $n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0)$ we require that $f(\lambda, \theta)$ satisfy assumption C1 and assumptions C2′, C3 and C4 set out below.

C2′. $f(\cdot) = f(\cdot, \theta_0)$ for a unique $\theta_0 \in \operatorname{int}\Theta$. Here $\operatorname{int}\Theta$ denotes the interior of Θ.

C3. Suppose $\mathfrak{M}$ is $\mathbb{R}^p$ so that $\theta = [\theta_1, \cdots \theta_p]^T$. It is also supposed that there exists a neighbourhood of θ_0, denoted N_1, in which $m(\lambda, \theta)$ is a twice differentiable function of the θ_j whose second derivatives w.r.t. the θ_j are continuous in $(\lambda, \theta) \in [-\pi, \pi] \times N_1$.

When assumption D3 holds we may define $W = (w_{ij})$ where

$$w_{ij} =_\Delta \frac{1}{4\pi} \int \frac{m^{(i)}(\lambda, \theta_0) m^{(j)}(\lambda, \theta_0)}{m^2(\lambda, \theta_0)} \, d\lambda = \frac{1}{4\pi} \int \log^{(i)} m(\lambda, \theta_0) \log^{(j)} m(\lambda, \theta_0) \, d\lambda.$$

From the hypothesis that $\int m(\lambda, \theta) f(\lambda) \, d\lambda$ is minimised at θ_0 it follows that W is positive semidefinite. We now introduce the following assumption.

C4. W is nonsingular.

For the derivation of the limiting distribution of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)$ we require, in addition, that $h(\lambda, \theta)$ satisfy D1 and the following assumptions.

D2′. $J(\theta)$ is minimised at the single value θ_0 whatever the value of θ_0 in $\operatorname{int}\Theta$.

D3. $h(\lambda, \theta) \geqslant 0$ and $h(-\lambda, \theta) = h(\lambda, \theta)$ for all $(\lambda, \theta) \in Q$. $\int \log h(\lambda, \theta) \, d\lambda > -\infty$ for all $\theta \in \Theta$.

Assumption D3 is no real restriction over and above assumptions D1 and D2′. To see this suppose $h(\lambda, \theta)$ satisfies D1 and D2′ then $h^+(\lambda, \theta) =_\Delta L + \frac{1}{2}(h(-\lambda, \theta) + h(\lambda, \theta))$ has the property that the subset of Θ minimising $\int h(\lambda, \theta) I_n(\lambda) \, d\lambda$ is the same as the subset of Θ minimising $\int h^+(\lambda, \theta) I_n(\lambda) \, d\lambda$ and furthermore $h^+(\lambda, \theta)$ satisfies D1, D2′ and D3 for suitably large L.

D4. There exists some neighbourhood of θ_0 denoted N_0 in which $h(\lambda, \theta)$ is a twice differentiable function of the θ_j and whose second derivatives w.r.t. θ_j are continuous in $(\lambda, \theta) \in [-\pi, \pi] \times N_0$.

Under assumption D4 we may define $\Phi = (\phi_{ij})$ where $\phi_{ij} =_\Delta \int h^{(ij)}(\lambda, \theta_0) f(\lambda) \, d\lambda$. From the hypothesis that $\int h(\lambda, \theta) f(\lambda) \, d\lambda$ is minimised at θ_0 it follows that Φ is positive semidefinite. We now introduce the following assumption.

D5. Φ is nonsingular.

We are now ready to present the main result of the paper.

THEOREM 3.1. *Suppose assumptions* A3–A6, B1, C1, C2′, C3, C4, D1, D2′, D3, D4 *and* D5 *are satisfied. Then*

$$n^{\frac{1}{2}}\left(\tilde{\theta}_n - \theta_0\right) \rightarrow_{\text{dist}} N(0, W^{-1})$$

$$n^{\frac{1}{2}}\left(\hat{\theta}_n - \theta_0\right) \rightarrow_{\text{dist}} N(0, \Phi^{-1}\Psi\Phi^{-1})$$

where $\Psi = (\psi_{ij})$ *and* $\psi_{ij} =_\Delta 4\pi\int h^{(i)}(\lambda, \theta_0)h^{(j)}(\lambda, \theta_0)f^2(\lambda)\, d\lambda$. *Furthermore,* $\Phi^{-1}\Psi\Phi^{-1} - W^{-1}$ *is a positive semidefinite matrix. The result is usually written* $\Phi^{-1}\Psi\Phi^{-1} \geqslant W^{-1}$.

PROOF. The proof is divided into two parts.

Part a. We note that the following relationships hold

$$\int m^{(i)}(\lambda, \theta_0)f(\lambda)\, d\lambda = 0 \tag{3.1}$$

$$\int h^{(i)}(\lambda, \theta_0)f(\lambda)\, d\lambda = 0. \tag{3.2}$$

Now assumption D2′ implies that whatever the value of θ_0 in int $\Theta \int h^{(j)}(\lambda, \theta_0)f(\lambda)\, d\lambda = 0$. Hence

$$\int \frac{h^{(j)}(\lambda, \theta_0)}{m(\lambda, \theta_0)}\, d\lambda = 0 \qquad \text{whatever the value of } \theta_0 \text{ in int } \Theta.$$

Differentiating w.r.t. θ_0 we see that

$$\int h^{(j)}(\lambda, \theta_0)\frac{m^{(i)}(\lambda, \theta_0)}{m^2(\lambda, \theta_0)}\, d\lambda = \int \frac{h^{(ij)}(\lambda, \theta_0)}{m(\lambda, \theta_0)}\, d\lambda.$$

Hence

$$4\pi\int h^{(j)}(\lambda, \theta_0)m^{(i)}(\lambda, \theta_0)f^2(\lambda)\, d\lambda = 2\sigma^2\phi_{ij}. \tag{3.3}$$

Also,

$$4\pi\int m^{(j)}(\lambda, \theta_0)m^{(i)}(\lambda, \theta_0)f^2(\lambda)\, d\lambda = 4\sigma^4 w_{ij}. \tag{3.4}$$

Let us now introduce the quantities

$$V_n^i =_\Delta n^{\frac{1}{2}}S_n^{(i)}(\theta_0) = n^{\frac{1}{2}}\int h^{(i)}(\lambda, \theta_0)I_n(\lambda)\, d\lambda$$

$$T_n^i =_\Delta n^{\frac{1}{2}}\tilde{S}_n^{(i)}(\theta_0) = n^{\frac{1}{2}}\int m^{(i)}(\lambda, \theta_0)I_n(\lambda)\, d\lambda.$$

Let us define

$$A =_\Delta \left[\begin{array}{c|c} \Psi & 2\sigma^2\Phi \\ \hline 2\sigma^2\Phi & 4\sigma^4 W \end{array}\right].$$

We will prove that the limiting distribution of $\alpha =_\Delta [V_n^1, \cdots V_n^p, T_n^1 \cdots T_n^p]^T$ is $N(0, A)$.

Now let us define $I(\lambda, \varepsilon) =_\Delta \frac{1}{2\pi n}|\Sigma_{t=1}^n \varepsilon_t e^{it\lambda}|^2$. As proved by Hannan [1] we may replace consideration of the expressions T_n^i and V_n^i by

$$n^{\frac{1}{2}}\int I(\lambda,\varepsilon)f(\lambda)m^{(i)}(\lambda,\theta_0)\,d\lambda \qquad \text{and} \qquad n^{\frac{1}{2}}\int I(\lambda,\varepsilon)f(\lambda)h^{(i)}(\lambda,\theta_0)\,d\lambda$$

respectively. But equations (3.1) and (3.2) imply that

$$n^{\frac{1}{2}}\int I(\lambda,\varepsilon)f(\lambda)m^{(i)}(\lambda,\theta_0)\,d\lambda = n^{\frac{1}{2}}\int\left\{I(\lambda,\varepsilon) - \frac{1}{2\pi n}\Sigma_{t=1}^n\varepsilon_t^2\right\}f(\lambda)m^{(i)}(\lambda,\theta_0)\,d\lambda$$

and

$$n^{\frac{1}{2}}\int I(\lambda,\varepsilon)f(\lambda)h^{(i)}(\lambda,\theta_0)\,d\lambda = n^{\frac{1}{2}}\int\left\{I(\lambda,\varepsilon) - \frac{1}{2\pi n}\Sigma_{t=1}^n\varepsilon_t^2\right\}f(\lambda)h^{(i)}(\lambda,\theta_0)\,d\lambda.$$

As proved by Hannan [1] this type of expression can be reduced to the consideration of an expression which is asymptotically normal by Hannan and Heyde [2]. The precise value of the limiting covariance matrix follows from equations (3.3), (3.4) and the definition of Ψ.

Part b. By Theorem 2.1 $\tilde{\theta}_n \to_{\text{prob}} \theta_0$ and $\hat{\theta}_n \to_{\text{prob}} \theta_0$. Hence we may derive the limiting distribution on the assumption that $\tilde{\theta}_n \in N_1$ and $\hat{\theta}_n \in N_0$. By the mean-value theorem

$$\Sigma_{i=1}^p\left\{-\tilde{S}_n^{(ij)}\left(\bar{\theta}_{nj}\right)\right\}\left\{n^{\frac{1}{2}}\left(\tilde{\theta}_{n,i} - \theta_{0,i}\right)\right\} = T_n^j \tag{3.5}$$

$$\Sigma_{i=1}^p\left\{-S_n^{(ij)}\left(\theta_{nj}^*\right)\right\}\left\{n^{\frac{1}{2}}\left(\hat{\theta}_{n,i} - \theta_{0,i}\right)\right\} = V_n^j \tag{3.6}$$

where $\bar{\theta}_{nj} = \lambda_{1j}\tilde{\theta}_n + (1-\lambda_{1j})\theta_0$, $\theta_{nj}^* = \lambda_{2j}\hat{\theta}_n + (1-\lambda_{2j})\theta_0$, $0 < \lambda_{ij} < 1$ and $0 < \lambda_{2j} < 1$.

Now $\tilde{S}_n^{(ij)}(\bar{\theta}_{nj}) - \tilde{S}_n^{(ij)}(\theta_0) \to_{\text{prob}} 0$ and $S_n^{(ij)}(\theta_{nj}^*) - S_n^{(ij)}(\theta_0) \to_{\text{prob}} 0$. By an argument similar to that used in the proof of Lemma 2.1 it can be seen that $\tilde{S}_n^{(ij)}(\theta_0) \to_{\text{prob}} w_{ij}$ and $S_n^{(ij)}(\theta_0) \to_{\text{prob}} \phi_{ij}$. Let us define

$$\beta =_\Delta n^{\frac{1}{2}}\left[\begin{array}{c}\hat{\theta}_n - \theta_0 \\ \hline \tilde{\theta}_n - \theta_0\end{array}\right] \qquad \text{and} \quad B =_\Delta \left[\begin{array}{c:c}\Phi^{-1}\Psi\Phi^{-1} & 2\sigma^2 I \\ \hdashline 2\sigma^2 I & 4\sigma^4 W^{-1}\end{array}\right].$$

From the limiting distribution for α found in Part a and from equations (3.5) and (3.6) it follows that $\beta \to_{\text{dist}} N(0, B)$. B is clearly positive semidefinite as it is a covariance matrix. Hence

$$\left[\begin{array}{c:c} I & \frac{1}{2\sigma^2}I\end{array}\right]\left[\begin{array}{c:c}\Phi^{-1}\Psi\Phi^{-1} & 2\sigma^2 I \\ \hdashline 2\sigma^2 I & 4\sigma^4 W^{-1}\end{array}\right]\left[\begin{array}{c} I \\ \hdashline \frac{1}{2\sigma^2}I\end{array}\right] \geqslant 0.$$

Therefore $\Phi^{-1}\Psi\Phi^{-1} \geqslant W^{-1}$.

REMARK 3.1. $\Phi^{-1}\Psi\Phi^{-1} \geqslant W^{-1}$ is a strong result. It implies, for instance, that each diagonal element of $\Phi^{-1}\Psi\Phi^{-1}$ is larger than the corresponding element of W^{-1}. It also implies that $\det \Phi^{-1}\Psi\Phi^{-1} \geqslant \det W^{-1}$ which, as mentioned in Section 1, is the result stated by Whittle in [4] page 215.

REMARK 3.2. It is evident that $\tilde{\theta}_n$ is a better estimator of θ_0 than $\hat{\theta}_n$ when assumption D5 is not satisfied i.e., when Φ is singular.

REMARK 3.3. In its basic approach the proof given above is much the same as that given in [3] pages 428–429. We outline, now, another interpretation of Theorem 3.1. In [5] pages 124–125, this other interpretation is used as the basis for an alternative proof of the result given in [3].

If we let $x = [x_1, \cdots, x_n]^T$ then we see that $S_n(\theta)$ is approximately x^TCx where

$$C =_\Delta \frac{1}{2\pi n}\begin{bmatrix} \alpha_0(\theta) & \alpha_1(\theta) & \cdot\ \ \cdot & \alpha_{n-1}(\theta) \\ \alpha_{-1}(\theta) & \ddots & \ddots & \\ \vdots & \ddots & \ddots & \\ \alpha_{-n+1}(\theta) & & \ddots & \end{bmatrix}$$

and we remind the reader that $\alpha_s(\theta) =_\Delta \frac{1}{2\pi}\int h(\lambda, \theta)e^{-is\lambda}\,d\lambda$. Thus Theorem 3.1 may be interpreted as indicating the optimality of the least-squares estimator from within the class of consistent estimators of θ_0 which minimise certain types of quadratic forms in the data.

4. Some examples. Suppose assumptions A3–A6, B1, C1, C2′, C3 and C4 are satisfied. It is of interest to ask how "large" is the class of estimators of θ_0 generated by minimising $\int h(\lambda, \theta)I_n(\lambda)\,d\lambda$ where $h(\lambda, \theta)$ satisfies assumptions D1, D2′, D3, D4 and D5. As already indicated in Section 3 the least-squares estimator belongs to this class of estimators of θ_0. We may generate (in the obvious way) a "large" class of nonequivalent $h(\lambda, \theta)$'s which satisfy assumptions D1, D2′, D3, D4 and D5 based on two observations:

First observation. Suppose $h_1(\lambda, \theta)$ and $h_2(\lambda, \theta)$ satisfy assumptions D1, D2′, D3, D4 and D5. Also suppose that $h_3(\lambda, \theta)$ and $h_4(\lambda, \theta)$ satisfy D1, D3 and D4 and that $\int h_3(\lambda, \theta)f(\lambda)\,d\lambda$ and $\int h_4(\lambda, \theta)f(\lambda)\,d\lambda$ are minimised on sets containing θ_0 whatever the value θ_0 is assumed to take in $\operatorname{int}\Theta$. Then $h_5(\lambda, \theta) =_\Delta h_1(\lambda, \theta) + \alpha h_2(\lambda, \theta)$ satisfies assumptions D1, D2′, D3, D4 and D5 for $\alpha > 0$. $h_6(\lambda, \theta) =_\Delta h_3(\lambda, \theta) + \beta h_4(\lambda, \theta)$ satisfies assumptions D1, D3, D4 and $\int h_6(\lambda, \theta)f(\lambda)$ is minimised on a set containing θ_0 whatever the value of $\theta_0 \in \operatorname{int}\Theta$ for $\beta > 0$. $h_7(\lambda, \theta) =_\Delta h_1(\lambda, \theta) + \sigma h_3(\lambda, \theta)$ satisfies assumptions D1, D2′, D3, D4 and D5 for $\sigma > 0$.

Second observation. We may generate functions $h(\lambda, \theta)$ which satisfy D1, D3, D4 and $\int h(\lambda, \theta)f(\lambda)\,d\lambda$ being minimised on a set containing θ_0 whatever the value of θ_0 in $\operatorname{int}\Theta$ as follows. Consider the class of estimators of x_t of the form $\hat{x}_t = -\sum_{u;u\neq 0}h_u(\mu)x_{t-u}$ where $\hat{x}_t$ denotes the estimated value of x_t and $\mu \in \boldsymbol{\mu}$, a compact subset of $\mathbb{R}^q$. Suppose that for each $\mu \in \boldsymbol{\mu}$ $\sum_{u,u\neq 0}h_u(\mu)z^u$ is analytic in an annulus containing the unit circle. The mean-square error of estimation is $E\{(x_t - \hat{x}_t)^2\} = \int h(\lambda, \mu)f(\lambda)\,d\lambda$. Suppose, also, that for each $\theta \in \operatorname{int}\Theta$ $\int h(\lambda, \mu)f(\lambda, \theta)\,d\lambda$ ($h(\lambda, \mu) =_\Delta |1 + \sum_{u,u\neq 0}h_u(\mu)e^{iu\lambda}|^2$) is minimised at a unique $\mu \in \operatorname{int}\boldsymbol{\mu}$ and denote this value by $\mu(\theta)$. From assumptions C1, C2′ and C3 it follows that in N_1 $f(\lambda, \theta)$ is a twice differentiable function of the θ_j whose second derivatives w.r.t. the θ_j are continuous in $(\lambda, \theta) \in [-\pi, \pi] \times N_1$. Assume there exists a neighbourhood of $\mu(\theta_0)$ denoted N_2, in which $h(\lambda, \mu)$ is a twice differentiable function of the μ_j and whose third derivatives w.r.t. θ_j are continuous in $(\lambda, \mu) \in [-\pi, \pi] \times N_2$.

If $f_i(\theta, \mu) =_\Delta \int h^{(i)}(\lambda, \mu) f(\lambda, \theta)\, d\lambda$ then clearly $\mu(\theta_0)$ is a solution of $f_i(\theta_0, \mu) = 0$ for $i = 1, \cdots, q$. Let us now introduce the assumption that the Jacobian $(\partial(f_1, \cdots f_q)/\partial(\mu_1, \cdots \mu_q))$ is nonzero at $(\theta_0, \mu(\theta_0))$. An application of the implicit function theorem indicates that there is neighbourhood of θ_0, denoted N_3, such that for θ in this neighbourhood $\mu(\theta)$ is a twice differentiable function of the θ_j whose second derivatives w.r.t the θ_j are continuous in $(\lambda, \theta) \in [-\pi, \pi] \times N_3$. Hence $h(\lambda, \mu(\theta))$ has the desired properties. Let it be remarked that an especially simple class of estimators suitable for the construction just described is given by $\hat{x}_t = -\sum_{j=1}^{q} \mu_j x_{t-\tau(j)}$ where $\tau(j) \neq 0$ for $j = 1, \cdots q$ and $\boldsymbol{\mu}$ is defined to be the set of those μ's minimising $\int h(\lambda, \mu) f(\lambda, \theta)\, d\lambda$ as θ varies over Θ.

Based on the second observation we present, in Subsections 4.1 and 4.2, an infinite class of nonequivalent $h(\lambda, \theta)$'s which satisfy assumptions D1, D2′, D3, D4 and D5 for $\{x_t\}$ known to be given by a first order autoregression.

4.1 *An infinite class of nonequivalent* h(λ, θ)'s. Suppose $\{x_t\}$ is given by $x_t = \sum_u l_u \varepsilon_{t-u}$ where $l_u = \theta_0^u$ for $u \geqslant 0$, $l_u = 0$ for $u < 0$ and where for some $0 < \xi < \delta < 1$ known to us, $\xi < |\theta_0| \leqslant \delta$. It is clear that $\sum_u u l_u^2 < \infty$ so that assumptions A4 and B1 are satisfied. Suppose, further, that $\{\varepsilon_t\}$ satisfies assumptions A4, A5 and A6.

If we define $f(\lambda, \theta) = (\sigma^2/2\pi)(1/|1 + \theta e^{i\lambda}|^2)$ and $\Theta = [-\delta, -\xi] \cup [\xi, \delta]$ it is clear that assumptions C1, C2′, C3 and C4 are satisfied.

Now let k be a fixed, positive, odd integer. Consider the class of predictors of the form $\hat{x}_t = \mu x_{t-k}$ where $\boldsymbol{\mu} =_\Delta [-\delta^k, \delta^k]$. It is not difficult to show that for each $\theta \in \Theta$ the value of μ which minimises $\int h(\lambda, \mu) f(\lambda, \theta)\, d\lambda$ is $\mu(\theta) = \theta^k$.

$h(\lambda, \mu(\theta))$ is seen to satisfy assumptions D1, D2′, D3, D4 and D5. By Theorem 3.1 the limiting distribution of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)$ is normal with mean zero and variance $J_k =_\Delta (1/k^2)\theta^{2(1-k)} a b^{-2}$ where $a =_\Delta 4\pi \int (\partial h(\lambda, \mu(\theta_0))/\partial\mu) f(\lambda)\, d\lambda$ and $b =_\Delta \int (\partial^2 h(\lambda, \mu(\theta_0))/\partial\mu^2) f(\lambda)\, d\lambda$. By a lengthy calculation we can find explicit expressions for the Fourier coefficients of $f(\lambda)$ and $f^2(\lambda)$. It can then be shown that:

$$(4.1) \quad \frac{J_k}{J_1} = \frac{\theta_0^2(1 + \theta_0^2)}{(1 - \theta_0^2)^2} \frac{1}{k^2 \theta_0^{2k}} - \frac{\theta_0^2}{1 - \theta_0^2}\left(\frac{2k+1}{k^2}\right) - 2\left(\frac{\theta_0^2}{1 - \theta_0^2}\right)^2 \frac{1}{k^2}.$$

Let us examine this expression a little. For $k > 1$ and $\theta_0 = 0$ J_k/J_1 clearly diverges to $+\infty$. For $\theta_0 \neq 0$ J_k/J_1 is of the form $a_1(\theta_0) f_1(k, \theta_0) - a_2(\theta_0) f_2(k) - a_3(\theta_0) f_3(k)$ where $a_1(\theta_0) > 0$, $a_2(\theta_0) > 0$, $a_3(\theta_0) > 0$, $f_1(k, \theta_0) > 0$, $f_2(k) > 0$ and $f_3(k) > 0$. Obviously $f_2(k + 2) < f_2(k)$ and $f_3(k + 2) < f_3(k)$. Furthermore, $f_1(k + 2, \theta_0) > f_1(k, \theta_0)$ for $|\theta_0| < \left(\frac{k}{k+2}\right)^{\frac{1}{2}}$. Hence for fixed θ_0 such that $\theta_0 \neq 0$ and $|\theta_0| < \frac{1}{3^{\frac{1}{2}}}$: $1 < \frac{J_3}{J_1} < \frac{J_5}{J_1} < \frac{J_7}{J_1} < \cdots$. In fact, for any fixed $0 < \delta < 1$ and for k_0

sufficiently large $\frac{J_{k_0}}{J_1} < \frac{J_{k_0+2}}{J_1} < \frac{J_{k_0+4}}{J_1} < \cdots$. Theorem 3.1 indicates that for any θ_0 such that $|\theta_0| < 1$ that $J_k(\theta_0)/J_1(\theta_0) \geqslant 1$ and numerical evaluations of J_k/J_1 verify the truth of this.

4.2 *An additional* $h(\lambda, \theta)$. Suppose $\{x_t\}$ is given by $x_t = \Sigma_u l_u \varepsilon_{t-u}$ where $l_u = \theta_0^u$ for $u \geqslant 0$, $l_u = 0$ for $u < 0$ and where for some $0 < \delta < 1$ known to us, $|\theta_0| \leqslant \delta$. It is clear that $\Sigma_u u l_u^2 < \infty$ so that assumptions A3 and B1 are satisfied. Suppose, further, that $\{\varepsilon_t\}$ satisfies assumptions A4, A5 and A6.

If we define $f(\lambda, \theta) = (\theta^2/2\pi)(1/|1 + \theta e^{i\lambda}|^2)$ and $\Theta = [-\delta, \delta]$ it is clear that assumptions C1, C2′, C3 and C4 are satisfied.

Consider the class of interpolators of the form $\hat{x}_t = \mu(x_{t-1} + x_{t+1})$ where $\boldsymbol{\mu} =_\Delta [-\delta/(L + \delta^2), \delta/(1 + \delta^2)]$. Thus the mean-square error of interpolation is $E\{(x_t - \hat{x}_t)^2\} = \int h(\lambda, \mu) f(\lambda)\, d\lambda$. For each $\theta \in \operatorname{int}\Theta$ $\int h(\lambda, \mu) f(\lambda, \mu)\, d\lambda$ is minimised at $\mu(\theta) = \dfrac{\theta}{1 + \theta^2}$.

$h(\lambda, \mu(\theta))$ satisfies assumptions D1, D2′, D3, D4 and D5. By Theorem 3.1 it is clear that the limiting distribution of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)$ is normal with mean zero and variance

$$I_1 =_\Delta \left[\frac{(1 + \theta_0^2)^2}{1 - \theta_0^2} \right]^2 c d^{-2}$$

where

$$c =_\Delta 4\pi \int \left[\frac{h(\lambda, \mu(\theta_0))}{\partial \mu} \right]^2 f(\lambda)\, d\lambda \text{ and } d =_\Delta \int \frac{\partial^2 h(\lambda, \mu(\theta_0))}{\partial \mu^2} f(\lambda)\, d\lambda.$$

A calculation of the Fourier coefficients of $f(\lambda)$ and $f^2(\lambda)$ shows that $I_1 = 1$. Now J_1 as defined in the previous section is $1 - \theta_0^2$. Hence it is obvious that $J_1 < I_1$ except for $\theta_0 = 0$ where $J_1 = I_1$. This is in accord with Theorem 3.1.

Acknowledgments. The author wishes to express his sincere thanks to Professor G. C. Goodwin for his encouragement and advice. The author also wishes to thank Professor E. J. Hannan for his comments on a preliminary version of the paper. Thanks also to a referee whose comments led to a substantial improvement in the presentation of the paper.

REFERENCES

[1] Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Probability* **10** 130–145.

[2] Hannan, E. J. and Heyde, C. C. (1972). On limit theorems for quadratic forms of discrete time series. *Ann. Math. Statist.* **43** 2058–2068.

[3] Whittle, P. (1953). Estimation and information in stationary time-series. *Ark. Mat.* **2** 423–434.

[4] Whittle, P. (1954). Some recent contributions to the theory of stationary processes. Appendix 2. In *A Study in the Analysis of Stationary Time-Series*. 2nd Ed. (H. Wold, ed.). Almquist and Wiksell, Stockholm.

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

Division of Mathematics and Statistics
CSIRO
P.O. Box 310
South Melbourne, Victoria 3205
Australia