*The Annals of Statistics*
1988, Vol. 16, No. 3, 1315–1326

# WEIGHTED LEAST SQUARES ESTIMATORS ON THE FREQUENCY DOMAIN FOR THE PARAMETERS OF A TIME SERIES[1]

BY SHEAN-TSONG CHIU

*Rice University*

A procedure for estimating the parameters of a time series is proposed. The estimate minimizes a criterion function which is the weighted sum of squares of the distances between the periodogram and the spectrum of the series. Under regularity conditions, the estimate is shown to be strongly consistent. The asymptotic distribution of the estimate is also obtained. It is shown that, for a Gaussian process, an asymptotically efficient estimate can be obtained by using an iteratively reweighted procedure.

**1. Introduction.** We consider a stationary time series $X(t)$, $t = 0, \pm 1, \ldots,$ with mean zero and autocovariance function $c(u) = E[X(t)X(t+u)]$. When the autocovariance function satisfies

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

the spectral density function of the series $X(t)$ is defined by

$$f(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} c(u)\exp(-i\lambda u).$$

This relation can be inverted to obtain the representation

$$c(u) = \int_{-\pi}^{\pi} \exp(iu\lambda) f(\lambda)\, d\lambda.$$

In this paper, we are interested in the situations in which the spectrum of the series depends on some unknown parameters. Here are some examples:

$$f(\lambda) = \frac{\Omega_0}{[1 + (\lambda/\lambda_0)]^2}, \tag{1.1}$$

$$f(\lambda) = \frac{\Omega_0}{\left[1 + (\lambda/\lambda_0)^2\right]^{3/2}}, \tag{1.2}$$

$$f(\lambda) = \beta_2^2 \exp\left(-\frac{\lambda^2\beta_1^{-2}}{2}\right), \tag{1.3}$$

---

Received October 1986; revised November 1987.

[1]This research was supported in part by ONR Contract N-00014-85-K-0100 and ARO Grant DAAG 29-85-K-0212.

*AMS* 1980 *subject classifications*. Primary 62M10; secondary 62M15, 62F10.

*Key words and phrases*. Time series, weighted least squares estimation, periodogram, spectrum, cumulants.

(1.4) $$f(\lambda) = \beta_2 \exp\left(-\frac{|\lambda|}{\beta_1}\right),$$

(1.5) $$f(\lambda) = \frac{\beta_2}{\lambda^5} \exp\left(-\frac{\beta_1}{\lambda^4}\right),$$

(1.6) $$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|1 + b_1 e^{i\lambda} + \cdots + b_q e^{iq\lambda}|^2}{|1 + a_1 e^{i\lambda} + \cdots + a_p e^{ip\lambda}|^2}.$$

Models (1.1) and (1.2) were used in Aki (1967) to model the spectra of far field body-wave displacements of earthquakes. The parameter $\lambda_0$ is the corner frequency and $\Omega_0$ is proportional to seismic moment. Seismologists are very interested in estimating the parameters $\lambda_0$ and $\Omega_0$, which are important in the study of the source properties of earthquakes [cf. Aki and Richards (1980), pages 819–825]. Slutsky (1937) used model (1.3) in studies of economic time series. Model (1.4) has been used by Lumley and Panofsky (1964) for atomospheric turbulence and by Whittle (1962) for agricultural spatial series. Pierson and Moskowitz (1964) proposed (1.5) as the spectrum of ocean waves caused by winds. The last example (1.6) is the spectrum of an autoregressive moving-average process of order $(p, q)$ [ARMA$(p, q)$]. We assume the equation $1 + a_1 z + \cdots + a_p z^p = 0$ has all its roots outside the unit circle, otherwise the series will not be stationary. However, we do not assume the equation $1 + b_1 z + \cdots + b_q z^q = 0$ has all its roots outside the unit circle. The latter condition is required in many papers. In the following discussion, we let $f(\lambda, \theta_0)$ [or $f(\lambda)$] represent the spectrum of the series $X(t)$ and we are interested in estimating the parameter $\theta_0$.

The criterion function used in this paper is based on the periodogram. The periodogram at frequency $\lambda$ of a series $X(t)$ is defined as

$$I(\lambda) = \frac{1}{2\pi T} d_X^{(T)}(\lambda) d_X^{(T)}(-\lambda),$$

where $d_X^{(T)}(\lambda)$ is the finite Fourier transform of the series $X(t)$,

$$d_X^{(T)}(\lambda) = \sum_{t=0}^{T-1} X(t) \exp(-i\lambda t).$$

The motivation of the frequency domain approach is from the observation that under some mild conditions, the periodogram ordinates of a series $X(t)$ on the Fourier frequencies $0 < \lambda_j = 2\pi j/T < \pi$ are asymptotically independently distributed according to an exponential distribution with mean $f(\lambda_j)$ [cf. Brillinger (1975)]. Therefore, an "approximate maximum likelihood estimator" can be obtained by finding the $\theta$ that maximizes the function

(1.7) $$L(\theta) = -\sum_{\lambda} \log[f(\lambda, \theta)] - \sum_{\lambda} I(\lambda)/f(\lambda, \theta),$$

where the summation is over the Fourier frequencies in $(-\pi, \pi) - \{0\}$. Unless indicated otherwise, we use this convention throughout this paper. The estimates which maximize $L(\theta)$ were studied in Bloomfield (1973), Hannan (1973), Whittle (1961), Dzhaparidze (1974), Davies (1973), Robinson (1978) and Ibragimov (1967). The estimates based on the periodogram in a finite union of intervals were considered in Rice (1979). Under various conditions, the estimates were shown to be consistent and asymptotically normal. For Gaussian processes, the estimates were also shown to be asymptotically efficient. A similar estimation procedure was proposed by Taniguchi (1981), who considered the estimates which minimize the distance between the spectrum and the smoothed periodograms.

In the following sections, we study the properties of the weighted least squares estimate which minimizes the criterion function

$$Q_T(\theta) = \sum_{\lambda} \phi(\lambda)[f(\lambda, \theta) - I(\lambda)]^2. \tag{1.8}$$

We also show that, by using an iteratively reweighted procedure [cf. Green (1984)], we can obtain an estimate which has the same asymptotic covariance matrix as that for the approximate maximum likelihood estimate.

The frequency domain approach has several advantages:

1. Since the power spectrum of a series is usually a simple function of the parameters, it is natural to work with the spectrum directly. In general, the criterion functions defined on frequency domain are much easier to handle, both numerically and theoretically, than the objective functions defined on time domain.
2. A parametric model is often used as an approximation, and the approximation is usually good only in certain frequency bands. If we are only interested in the properties of the processes in these bands, we might not want to include the information contained in the other bands to estimate the parameter. This can be done easily by excluding the periodogram in some bands [cf. Rice (1979)].

In addition, the weighted least squares estimate has some other advantages:

1. Due to the availability of various least squares algorithms, it is easy to implement the procedure [see Green (1984)].
2. Sometimes, it might be desirable to just weight some periodogram ordinates less instead of excluding them completely.
3. The approximate maximum likelihood procedure cannot deal with the cases where $f(\lambda, \theta_0)$ vanishes at some $\lambda$ (for example, moving-average processes with some roots on the unit circle, or in the case that the energy in some frequency bands is filtered out). The weighted least squares procedure does not have this difficulty.
4. Analogous to the robust procedures for independent observations [see Green (1984) and references therein], we could get an estimate resistant to the presence of peaks (corresponding to periodic components, such as seasonal effect) in the series. Therefore, it is possible to estimate the spectrum of the deseasoned series without removing the seasonal trend from the series.

The last point needs more detailed study which will appear in future research.

**2. Strong consistency.** We are interested in the series $X(t)$, $t = 0, \pm 1, \dots$, which satisfies Assumption 1.

ASSUMPTION 1($k$). $X(t)$ is a stationary series with cumulants

$$c_h(u_1, u_2, \dots, u_{h-1}) = \operatorname{cum}\{X(t+u_1), \dots, X(t+u_{h-1}), X(t)\}.$$

The cumulants satisfy

$$\sum_{u_1, \dots, u_{h-1} = -\infty}^{\infty} (1 + |u_j|)|c_h(u_1, \dots, u_{h-1})| < \infty$$

for $j = 1, \dots, h-1$ and $h = 2, 3, \dots, k$.

Under Assumption 1, $X(t)$ has $h$th order spectrum $f_h(\lambda_1, \dots, \lambda_{h-1})$ with bounded and uniformly continuous derivatives [cf. Brillinger (1975), page 27].

Since the second-order properties of the series are often expressed in terms of its spectrum, it is more convenient to check whether $\Sigma[1 + |u|]|c(u)| < \infty$ or not by inspecting the spectrum. A sufficient condition is given in Theorem 1. The proofs of this and other theorems are given in Section 5. Theorem 1 is a rather direct consequence of results from trigonometric series [cf. Zygmund (1959), page 241].

THEOREM 1. *Suppose $X(t)$ is stationary and has spectrum $f(\lambda)$ with a bounded and uniformly continuous derivative $f'(\lambda)$ which satisfies a Lipschitz condition of order $\alpha > 0$. Then $\Sigma[1 + |u|]|c(u)| < \infty$, where $c(u) = c_2(u)$ is the autocovariance function of the series.*

All of the examples except (1.4) given in Section 1 satisfy the conditions of Theorem 1. The spectrum of example (1.4) is not differentiable at $\lambda = 0$, and the autocovariance function is of order $u^{-2}$ [see Robinson (1978) and Whittle (1962)]; thus, Assumption 1 is not satisfied either. For situations similar to (1.4), we could filter the series to make the spectrum of the filtered series satisfy the conditions in Theorem 1. That is, we can use the filter to smooth the singular points of the spectrum. Of course, we are not able to do this when the singular points depend on unknown parameters.

We now establish two theorems which are of some independent interest and are more general than the results required to prove the asymptotic properties of the weighted least squares estimates.

THEOREM 2. *Let $\psi(\lambda)$ be a continuous function on $[-\pi, \pi]$. If $X(t)$ satisfies Assumption* 1($4k$), *then*

$$\lim_{T \to \infty} \frac{1}{T} \sum_{\lambda} \psi(\lambda) I^k(\lambda) = \frac{k!}{2\pi} \int \psi(\lambda) f^k(\lambda)\, d\lambda \quad \textit{almost surely}, \tag{2.1}$$

*for $k \geq 1$. Here $f(\lambda)$ is the (second order cumulant) spectrum of $X(t)$.*

The integral in (2.1) is from $-\pi$ to $\pi$. Unless indicated otherwise, we use this convention throughout this paper. Theorem 2 can be easily extended to the following results.

COROLLARY 1. *Suppose $\psi(\lambda)$ and $X(t)$ satisfy the conditions in Theorem* 2, *and let $\phi(\lambda) = 1_\Lambda(\lambda)\psi(\lambda)$, where $1_\Lambda(\lambda)$ is the indicator function of a set $\Lambda \subset (-\pi, \pi)$, with $\Lambda$ a finite union of intervals. Then, for $k \geq 1$,*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{\lambda} \phi(\lambda) I^k(\lambda) = \frac{k!}{2\pi} \int \phi(\lambda) f^k(\lambda)\, d\lambda \quad \textit{almost surely.}$$

COROLLARY 2. *Suppose $\phi(\lambda, \theta) = 1_\Lambda(\lambda)\psi(\lambda, \theta)$, where $\psi(\lambda, \theta)$ is uniformly continuous on $(\lambda, \theta) \in [-\pi, \pi] \times \Theta$. Also suppose $X(t)$ satisfies Assumption* 1 $(2k)$. *Then, for $k \geq 1$,*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{\lambda} \phi(\lambda, \theta) I^k(\lambda) = \frac{k!}{2\pi} \int \phi(\lambda, \theta) f^k(\lambda)\, d\lambda \quad \textit{almost surely,}$$

*and the convergence is uniform over $\Theta$.*

For the case $k = 1$, results similar to Corollaries 1 and 2 (also allowing, in effect, for omission of frequencies) are in Theorem 1 of Robinson (1976). The results of Theorem 5.10.1 of Brillinger (1975) are also of interest.

The asymptotic distributions of the variables in Corollary 1 are given in Theorem 3.

THEOREM 3. *Suppose $\phi(\lambda)$ satisfies the condition in Corollary* 1 *and $X(t)$ satisfies Assumption* $1(\infty)$. *Then, as $T \to \infty$,*

$$\sqrt{T}\left[\frac{1}{T}\sum_{\lambda}\phi(\lambda)I^k(\lambda) - \frac{k!}{2\pi}\int \phi(\lambda) f^k(\lambda)\, d\lambda\right] \to N\big(0, \sigma_k^2\big),$$

*where*

$$\begin{aligned} 2\pi\sigma_k^2 = {} & \big[(2k)! - (k!)^2\big]\left\{\int\big[\phi^2(\lambda) + \phi(\lambda)\phi(-\lambda)\big] f^{2k}(\lambda)\, d\lambda\right\} \\ & + k^2(k!)^2 \int\!\!\int \phi(\lambda_1)\phi(\lambda_2) f_4(\lambda_1, -\lambda_1, \lambda_2) \\ & \times f^{k-1}(\lambda_1) f^{k-1}(\lambda_2)\, d\lambda_1\, d\lambda_2. \end{aligned} \tag{2.2}$$

REMARK 1. For series with $f_4(\lambda_1, \lambda_2, \lambda_3) \equiv 0$, such as in the case of Gaussian processes, the second term of (2.2) vanishes, and the asymptotic distribution is the same as the one in which $I(\lambda_j)$ are true independent variables.

REMARK 2. For Gaussian processes, Taniguchi (1980) studied functions of a form more general than the power functions we consider here and obtained results similar to Theorems 2 and 3.

In proving the theorem concerning the strong consistency of the weighted least squares estimates, we need assumptions about $\Theta$, $f(\lambda, \theta)$ and the weighting function $\phi(\lambda)$.

ASSUMPTION 2. $f(\lambda, \theta)$ is continuous on $(\lambda, \theta) \in [-\pi, \pi] \times \Theta$ and $\Theta$ is a compact set in $R^r$.

ASSUMPTION 3. The weighting function $\phi(\lambda) = 1_\Lambda(\lambda)\psi(\lambda)$ is symmetric, where $\psi(\lambda)$ is a positive continuous function on $[-\pi, \pi]$, and for all $\theta \neq \theta_0$, $\int \phi(\lambda)[f(\lambda, \theta) - f(\lambda, \theta_0)]^2 \, d\lambda > 0$.

Assumption 3 requires that, on the support of $\phi(\lambda)$, $f(\lambda, \theta_0)$ is different from $f(\lambda, \theta)$, $\theta \neq \theta_0$. From Corollaries 1 and 2, the strong consistency of the weighted estimate can be established.

THEOREM 4. *Under Assumptions* 1(4) *to* 3, *the estimate* $\hat{\theta}$ *which minimizes* $Q_T(\theta)$ *of* (1.8) *converges to* $\theta_0$ *almost surely.*

**3. Asymptotic distribution.** We now derive the asymptotic distribution of the weighted least squares estimate. In addition to the assumptions in the last section, we need some more assumptions.

ASSUMPTION 4. $\theta_0$ is an interior point of $\Theta$.

ASSUMPTION 5. In a neighborhood of $\theta_0$, $f(\lambda, \theta)$ is twice differentiable with respect to $\theta$ and the derivatives are continuous in $\lambda$ and $\theta$.

The asymptotic covariance matrix of the estimate depends on the matrix $A(\theta_0)$. The $jk$th element of $A(\theta)$ is

$$a_{jk}(\theta) = (2\pi)^{-1} \int \phi(\lambda) g_j(\lambda, \theta) g_k(\lambda, \theta) \, d\lambda, \tag{3.1}$$

where $g_j(\lambda, \theta) = (\partial/\partial\theta_j) f(\lambda, \theta)$.

ASSUMPTION 6. The matrix $A = A(\theta_0)$ is positive definite.

The asymptotic distribution of the weighted least squares estimate is given in Theorem 5.

THEOREM 5. *Under Assumptions* 1($\infty$) *to* 6, *let* $\hat{\theta}$ *be the estimate which minimizes* $Q_T(\theta)$ *of* (1.8). *Then* $\sqrt{T}(\hat{\theta} - \theta_0)$ *is asymptotically normal with mean zero and covariance matrix* $2A^{-1}BA^{-1} + A^{-1}DA^{-1}$. *Here* $A$ *is given in Assumption* 6. *The jkth elements of* $B$ *and* $D$ *are, respectively,*

$$b_{jk} = \frac{1}{2\pi} \int \phi^2(\lambda) f^2(\lambda, \theta_0) g_j(\lambda, \theta_0) g_k(\lambda, \theta_0) \, d\lambda$$

*and*

$$d_{jk} = \frac{1}{2\pi} \int \int \phi(\lambda)\phi(\mu) f_4(\lambda, -\lambda, \mu) g_j(\mu, \theta_0) g_k(\lambda, \theta_0)\, d\lambda\, d\mu.$$

The covariance matrix in Theorem 5 depends on the matrices $A$, $B$ and $D$. $A$ and $B$ can be estimated by $A(\hat{\theta})$ and $B(\hat{\theta})$, respectively. An estimate for the matrix $D$ is suggested by Theorem 6.

THEOREM 6. *Let $\phi(\lambda)$ satisfy Assumption* 3 *and suppose $X(t)$ satisfies Assumption* 1(8). *Then*

$$\frac{1}{T^2} \sum_{\lambda} \sum_{\mu} \phi(\lambda)\phi(\mu) I(\lambda) I(\mu) \tag{3.2}$$

*converges, in mean square, to*

$$\frac{1}{2\pi} \int \int \phi(\lambda)\phi(\mu) f_4(\lambda, -\lambda, \mu)\, d\lambda\, d\mu + \frac{2}{2\pi} \int \phi^2(\lambda) f^2(\lambda)\, d\lambda + \left[ \frac{1}{2\pi} \int \phi(\lambda) f(\lambda)\, d\lambda \right]^2. \tag{3.3}$$

**4. Asymptotically efficient estimate.** If we assume $f(\lambda, \theta_0) > 0$ for all $\lambda$ and let the weighting function $\phi(\lambda) = f^{-2}(\lambda, \theta_0)$, we obtain the same covariance matrix as the one in Theorem 5 of Robinson (1978), which is the asymptotic covariance matrix of the estimate which maximizes the approximate likelihood function $L(\theta)$ of (1.7). Furthermore, for Gaussian processes, the matrix $D$ vanishes, and the covariance matrix is $2A^{-1}$, which is the asymptotic covariance matrix of the maximum likelihood estimate [see Hannan (1973)]. Therefore, the $\hat{\theta}$ which minimizes $Q_T(\theta)$ of (1.8) with $\phi(\lambda) = f^{-2}(\lambda, \theta_0)$ is asymptotically efficient. However, $\theta_0$ is unknown in advance, and so the preceding $\hat{\theta}$ is not an estimate.

Though $\theta_0$ is unknown, a $\sqrt{T}$ consistent estimate $\tilde{\theta}$ could be obtained by using the suggested procedure with some arbitrary weighting. We can approximate $f^{-2}(\lambda, \theta_0)$ by $f^{-2}(\lambda, \tilde{\theta})$ and expect to obtain an asymptotically efficient estimate by using the weighting function $\phi(\lambda) = f^{-2}(\lambda, \tilde{\theta})$.

We redefine the sum of squares,

$$Q_T(\theta, \eta) = \sum_{\lambda} \psi(\lambda) [f(\lambda, \theta) - I(\lambda)]^2 / f^2(\lambda, \eta).$$

Theorem 7 describes the asymptotic properties of the sequence of estimates $\hat{\theta}_T = \hat{\theta}_T(\eta_T)$ which minimize $Q_T(\theta, \eta_T)$, where $\eta_T$ is a convergent sequence of random variables in $\Theta$.

THEOREM 7. *Let $\phi(\lambda) = \psi(\lambda)/f^2(\lambda, \eta_0)$; also suppose $\eta_T$ converges to $\eta_0$ almost surely, and $\sqrt{T}(\eta_T - \eta_0)$ converges in law. Then, under Assumptions 1–6, $\hat{\theta}_T$ is strongly consistent and $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is asymptotically normal with mean zero and covariance matrix as indicated in Theorem* 5.

If we assume $f(\lambda, \theta_0) > 0$ for all $\lambda$ and let $\hat{\theta}_1$ be the unweighted least squares estimate, then Theorem 7 implies that the estimate $\hat{\theta}_2$ obtained by using the weighting function $\phi(\lambda) = f^{-2}(\lambda, \hat{\theta}_1)$ has the same asymptotic distribution as that for the "estimate" with $\phi(\lambda) = f^{-2}(\lambda, \theta_0)$. Therefore, for Gaussian processes, the estimate $\hat{\theta}_2$ is asymptotically efficient. In practice, the iteration is repeated a few times or until the estimates converge.

Finally, we remark that the iterative procedure previously discussed is similar to the procedure in Green (1984). Green pointed out that if the iterative procedure converges to a point $\hat{\theta}$, then $\hat{\theta}$ is a solution of the likelihood equation.

## 5. Proofs.

We now give proofs of the results.

PROOF OF THEOREM 1. From integration by parts, we have

$$(5.1)\quad \begin{aligned} iuc(u) &= iu\int_{-\pi}^{\pi} f(\lambda)\exp(iu\lambda)\,d\lambda \\ &= f(\pi)[\exp(iu\pi) - \exp(-iu\pi)] - \int_{-n}^{\pi} f'(\lambda)\exp(iu\lambda)\,d\lambda. \end{aligned}$$

Since the first term in (5.1) is equal to zero, then

$$\sum_{u=-\infty}^{\infty} |uc(u)| = \sum_{u=-\infty}^{\infty} \left| \int_{-\pi}^{\pi} f'(\lambda)\exp(iu\lambda)\,d\lambda \right| < \infty$$

from Zygmund [(1959), Theorem (3.6), page 241]. The theorem is finished by noting that $c(0)$ is bounded. □

PROOF OF THEOREM 2. Let

$$\psi_L(\lambda) = \sum_{u=-L}^{L} q(u)\exp(iu\lambda)$$

be the Cesaro sum of the Fourier series of $\psi(\lambda)$ and let $S_T = \sum_\lambda \psi_L(\lambda) I^k(\lambda)$. From Theorem 4.3.2 of Brillinger (1975), we find that the expected value of $I^k(\lambda)$ is equal to $k!\,f^k(\lambda) + o(1)$, so

$$(5.2)\quad \lim_{T\to\infty} \frac{1}{T} E\left[\sum_\lambda \psi_L(\lambda) I^k(\lambda)\right] = \frac{k!}{2\pi}\int \psi_L(\lambda) f^k(\lambda)\,d\lambda.$$

In the proof of Theorem 3, we will show that the variance of $T^{-1}S_T$ is of order $T^{-1}$. This implies that $N^{-2}S_{N^2}$ converges almost surely to the right-hand side of (5.2). Also, it can be shown by straight calculation that $E\{\max_{N^2 < M \le (N+1)^2}[N^{-2}S_M - N^{-2}S_{N^2}]^2\}$ is of order $N^{1-2k}$ for $k \ge 2$ and is of order $N^{-2}$ for $k = 1$. Applying Chebyshev's inequality and the Borel–Cantelli lemma gives

$$\lim_{T\to\infty} \frac{S_T}{T} = \frac{k!}{2\pi}\int \psi_L(\lambda) f^k(\lambda)\,d\lambda$$

almost surely. The theorem now follows from the uniform convergence of $\psi_L(\lambda)$

to $\psi(\lambda)$ [cf. Edwards (1979), page 87] and the almost sure convergence of $\Sigma_\lambda I^k(\lambda)/T$. □

PROOF OF COROLLARY 1. The corollary follows immediately from the fact that for any $\delta > 0$, we can find a continuous function $\tilde{\phi}(\lambda)$ on $[-\pi, \pi]$, such that the set $W = \{\lambda\colon \phi(\lambda) = \tilde{\phi}(\lambda)\}$ is contained in a finite union of intervals with a total length less than $\delta$. □

PROOF OF COROLLARY 2. From the argument in Corollary 1, we see that it is sufficient to prove the result for continuous $\psi(\lambda, \theta)$. Let

$$\psi_L(\lambda, \theta) = \sum_{u=-L}^{L} q(u, \theta)\exp(iu\lambda)$$

be the Cesaro sum of the Fourier series of $\psi(\lambda, \theta)$. Since

$$|q(u, \theta)| = \left|\int \{1 - |u|/(L+1)\}\psi(\lambda, \theta)\exp(i\lambda u)\, d\lambda\right| \le 2\pi \sup_{\lambda, \theta} |\psi(\lambda, \theta)|,$$

$q(u, \theta)$ is uniformly bounded. From this and the fact that $\psi_L(\lambda, \theta)$ is a linear combination of the $2L + 1$ functions

$$\left(1 - \frac{|u|}{L+1}\right)\exp(i\lambda u), \qquad u = -L, \dots, L,$$

we establish the uniform convergence of $\Sigma_\lambda \psi_L(\lambda, \theta) I^k(\lambda)/T$. Corollary 2 follows from the uniform convergence of $\psi_L(\lambda, \theta)$ to $\psi(\lambda, \theta)$, which can be shown by slightly modifying the proof of 3.2.2 of Edwards (1979). □

PROOF OF THEOREM 3. The variance of $T^{-1/2}\Sigma\phi(\lambda)I^k(\lambda)$ is equal to

$$\begin{aligned} &T^{-1}\sum_{\lambda_1}\sum_{\lambda_2}\phi(\lambda_1)\phi(\lambda_2)\mathrm{cum}\big(I^k(\lambda_1), I^k(\lambda_2)\big) \\ &\quad = T^{-2k-1}(2\pi)^{-2k}\sum_{\nu}\sum_{\lambda_1}\sum_{\lambda_2}\phi(\lambda_1)\phi(\lambda_2) \\ &\qquad \times \mathrm{cum}\big\{d_X^{(T)}(\omega_{ij});\, ij \in \nu_1\big\} \cdots \mathrm{cum}\big\{d_X^{(T)}(\omega_{ij});\, ij \in \nu_p\big\}, \end{aligned} \tag{5.3}$$

where $\omega_{ij} = (-1)^j\lambda_i$ and the summation in $\nu$ is over all indecomposable partitions of the table [cf. Brillinger (1975)]

$$\begin{matrix} (1,1) \cdots (1, 2k) \\ (2,1) \cdots (2, 2k). \end{matrix}$$

Applying Theorem 4.3.2 of Brillinger (1975) yields that (5.3) is equal to

$$\begin{aligned} &2\pi k^2(k!)^2 T^{-2}\sum_{\lambda_1}\sum_{\lambda_2}\phi(\lambda_1)\phi(\lambda_2) f(\lambda_1, -\lambda_1, \lambda_2) f^{k-1}(\lambda_1) f^{k-+1}(\lambda_2) \\ &\quad + \left[(2k)! - (k!)^2\right]T^{-1}\left\{\sum_{\lambda}\left[\phi^2(\lambda) + \phi(\lambda)\phi(-\lambda)\right] f^{2k}(\lambda)\right\} + o(1), \end{aligned}$$

which converges to the $\sigma_k^2$ of (2.2) in Theorem 3.

By slightly generalizing the proof of Theorem 5.10.1 of Brillinger (1975), it can be shown that

$$\operatorname{cum}\left\{T^{-1/2}\sum_{\lambda_1}\phi(\lambda)I^k(\lambda),\ldots,T^{-1/2}\sum_{\lambda_h}\phi(\lambda)I^k(\lambda)\right\}$$

converges to zero for $h \geq 3$. Since, for Gaussian random variables, the cumulants of order higher than 2 are zero, we obtain the asymptotic normality and finish the proof. □

**Proof of Theorem 4.** From Corollaries 1 and 2, uniformly on $\Theta$,

$$\begin{aligned}\lim_{T\to\infty}\frac{Q_T(\theta)}{T} &= \frac{1}{2\pi}\int\phi(\lambda)\left[f^2(\lambda,\theta) - 2f(\lambda,\theta)f(\lambda,\theta_0) + 2f^2(\lambda,\theta_0)\right]d\lambda\\ &= \frac{1}{2\pi}\int\phi(\lambda)\left[f(\lambda,\theta) - f(\lambda,\theta_0)\right]^2 d\lambda + \frac{1}{2\pi}\int\phi(\lambda)f^2(\lambda,\theta_0)\,d\lambda\\ &= Q(\theta).\end{aligned}$$

Under Assumption 3, the minimum of $Q(\theta)$ is attained only at $\theta = \theta_0$. Let $\hat{\theta}_T$ be a sequence of estimates which minimizes $Q_T(\theta)$. Suppose that there exists a subsequence $\hat{\theta}_{T_n}$ converging to $\theta' \neq \theta_0$. Then

$$Q(\theta') = \lim_{T\to\infty} Q_T\left(\hat{\theta}_{T_n}\right)/T \leq \lim_{T\to\infty} Q_T(\theta_0)/T = Q(\theta_0),$$

which is a contradiction, and the proof is finished. □

**Proof of Theorem 5.** This follows immediately from Theorem 3 and Corollary 1 by using the classical argument in Jennrich (1969). □

**Proof of Theorem 6.** The expected value of (3.2) is equal to

$$(2\pi T)^{-2}T^{-2}\sum_{\lambda}\sum_{\mu}\phi(\lambda)\phi(\mu)E\left[d_X^{(T)}(\lambda)d_X^{(T)}(-\lambda)d_X^{(T)}(\mu)\right]. \tag{5.4}$$

Since

$$\begin{aligned}E&\left[d_X^{(T)}(\lambda)d_X^{(T)}(-\lambda)d^{(T)D}X(\mu)d_X^{(T)}(-\mu)\right]\\ &= \operatorname{cum}\left\{d_X^{(T)}(\lambda), d_X^{(T)}(-\lambda), d_X^{(T)}(\mu), d_X^{(T)}(-\mu)\right\}\\ &\quad+\operatorname{cum}\left\{d_X^{(T)}(\lambda), d_X^{(T)}(-\lambda)\right\}\operatorname{cum}\left\{d_X^{(T)}(\mu)d_X^{(T)}(-\mu)\right\}\\ &\quad+\operatorname{cum}\left\{d_X^{(T)}(\lambda), d_X^{(T)}(\mu)\right\}\operatorname{cum}\left\{d_X^{(T)}(-\lambda), d_X^{(T)}(-\mu)\right\}\\ &\quad+\operatorname{cum}\left\{d_X^{(T)}(\lambda), d_X^{(T)}(-\mu)\right\}\operatorname{cum}\left\{d_X^{(T)}(-\lambda), d_X^{(T)}(\mu)\right\},\end{aligned}$$

(5.4) is equal to

$$\begin{aligned}&2\pi T^{-2}\sum_{\lambda}\sum_{\mu}\phi(\lambda)\phi(\mu)f_4(\lambda,-\lambda,\mu) + T^{-1}\sum_{\lambda}\phi^2(\lambda)f^2(\lambda)\\ &\quad+T^{-1}\sum_{\lambda}\phi(\lambda)\phi(-\lambda)f^2(\lambda) + \left[T^{-1}\sum_{\lambda}\phi(\lambda)f(\lambda)\right]^2 + o(1),\end{aligned}$$

which converges to (3.3) in Theorem 5. Similar to the proof of Theorem 3, we can show that the variance of (3.2) is of order $T^{-1}$ and finish the proof. □

PROOF OF THEOREM 7. Since $\phi(\lambda) = \psi(\lambda)/f^2(\lambda, \eta_0)$ satisfies Assumption 3, $f(\lambda, \eta_0)$ is bounded away from zero on $\Lambda$. The strong consistency of $\hat{\theta}_T$ follows from the uniformly and almost sure convergence of $1/f^2(\lambda, \eta_T)$. To prove the theorem, we only need to show that the vectors $(\partial/\partial\theta)T^{-1/2}Q(\theta_0, \eta_0)$ and $(\partial/\partial\theta)T^{-1/2}Q(\theta_0, \eta_T)$ have the same asymptotic distribution.

Expanding $(\partial/\partial\theta_j)Q_T(\theta_0, \eta_T)$ about $\eta_0$ yields

$$\begin{aligned}\frac{\partial}{\partial\theta_j}&T^{-1/2}Q_T(\theta_0, \eta_T)\\ &= \frac{\partial}{\partial\theta_j}T^{-1/2}Q_T(\theta_0, \eta_0)\\ &\quad + \sum_{k=1}^{r}\sqrt{T}(\eta_{kT} - \eta_{k0})\left\{\frac{4}{T}\sum_{\lambda}\psi(\lambda)[I(\lambda) - f(\lambda, \theta_0)]\right.\\ &\qquad\qquad \left.\times g_j(\lambda, \theta_0)g_k(\lambda, \eta^*)/f^3(\lambda, \eta^*)\right\}\end{aligned} \tag{5.5}$$

for some $\eta^*$ which lies between $\eta_T$ and $\eta_0$. By applying Corollary 2, it can be seen that the term inside the braces in (5.5) converges to zero almost surely, and the proof is finished. □

**Acknowledgments.** The valuable comments and references from the referees are gratefully appreciated. Several errors in the earlier version of the manuscript, including a mistake in Theorem 3, were pointed out by the referees.

## REFERENCES

AKI, K. (1967). Scaling law of seismic spectrum. *J. Geophys. Res.* **72** 1217–1231.

AKI, K. and RICHARDS, P. G. (1980). *Quantitative Seismology Theory and Methods.* Freeman, San Francisco.

BLOOMFIELD, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60** 217–226.

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

DAVIES, R. B. (1973). Asymptotic inference in stationary Gaussian series. *Adv. in Appl. Probab.* **5** 469–497.

DZHAPARIDZE, K. O. (1974). A new method for estimating spectral parameters of a stationary regular time series. *Theory Probab. Appl.* **19** 122–132.

EDWARDS, R. E. (1979). *Fourier Series: A Modern Introduction* **1**, 2nd ed. Springer, New York.

GREEN, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives (with discussion). *J. Roy. Statist. Soc. Ser. B* **46** 149–192.

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

IBRAGIMOV, I. A. (1967). On maximum likelihood estimation of parameters of the spectral density of stationary time series. *Theory Probab. Appl.* **12** 115–119.

JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

LUMLEY, J. L. and PANOFSKY, H. A. (1964). *The Structure of Atmospheric Turbulence*. Wiley-Interscience, New York.

PIERSON, W. J. and MOSKOWITZ, L. (1964). A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. *J. Geophys. Res.* **69** 5181–5190.

RICE, J. (1979). On the estimation of the parameters of a power spectrum. *J. Multivariate Anal.* **9** 378–392.

ROBINSON, P. M. (1976). Fourier estimation of continuous time models. In *Statistical Inference in Continuous Time Economic Models* (A. R. Bergstrom, ed.) 215–266. North-Holland, Amsterdam.

ROBINSON, P. M. (1978). Alternative models for stationary stochastic processes. *Stochastic Process. Appl.* **8** 141–152.

SLUTSKY, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica* **5** 105–146.

TANIGUCHI, M. (1980). On estimation of the integrals of certain functions of spectral density. *J. Appl. Probab.* **17** 73–83.

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

WHITTLE, P. (1961). Gaussian estimation in stationary time series. *Bull. Inst. Internat. Statist.* **39** 105–130.

WHITTLE, P. (1962). Topographic correlations, power-law covariance functions, and diffusion. *Biometrika* **49** 305–314.

ZYGMUND, A. (1959). *Trigonometric Series* **1**. Cambridge Univ. Press, Cambridge.

DEPARTMENT OF STATISTICS
RICE UNIVERSITY
P.O. BOX 1892
HOUSTON, TEXAS 77251-1892