RATE OF CONVERGENCE FOR LOGSPLINE SPECTRAL DENSITY ESTIMATION

CHARLES KOOPERBERG
University of Washington

CHARLES J. STONE
University of California at Berkeley

YOUNG K. TRUONG
University of North Carolina at Chapel Hill

Revised May 18, 1994

Abstract. The logarithm of the spectral density function for a stationary process is approximated by polynomial splines. The approximation is chosen to maximize the expected log-likelihood based on the asymptotic properties of the periodogram. Estimates of this approximation are shown to possess the usual nonparametric rate of convergence when the number of knots suitably increases to infinity.

Keywords. Maximum likelihood; polynomial splines; spectral density function; stationary time series.

1. INTRODUCTION

Let X_t denote a stationary time series with autocovariance function $\gamma(u) = \mathrm{cov}(X_t, X_{t+u})$ such that $\sum_u |\gamma(u)| < \infty$. Then the spectral density function f exists and is given by

$$f(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma(u) \exp(-iu\lambda), \qquad -\pi \le \lambda \le \pi.$$

Observe that f is nonnegative and symmetric about zero and that it can be extended to a periodic function with period 2π.

The spectral density function plays an important role in time series analysis. It is used to examine individual frequency components, variations and serial correlations of the series. See, for example, Brillinger (1981) and Priestley (1981) for a wide range of applications of this function in other fields. The spectral density function can be estimated by averaging (or smoothing) the periodogram, which is necessary since the periodogram itself is inconsistent. This type of estimate is not very flexible since it is carried out by averaging over a fixed number of periodogram ordinates. Other methods such as autoregressive spectral estimates (Priestley, 1981) and the regression method (Wahba, 1980; Wahba and Wold, 1975) have been proposed to handle this problem. The former is a parametric approach, whereas the latter is a nonparametric method using splines to smooth the periodogram ordinates. Recently, Beltrão and Bloomfield (1987) and Hurvich and Beltrão (1990) have revised the kernel method by using cross-validation to develop automatic (constant) bandwidth selection procedures, while Swanepoel and van Wyk (1986), Franke and Härdle (1992) and Politis and Romano (1992) have considered an alternative approach using bootstrap methods.

The present paper considers a nonparametric approach to the problem of estimating the spectral density function, which is related to the method of nonparametric generalized regression considered by Stone (1986, 1994). Specifically, the logarithm of the spectral density function is approximated by a polynomial spline, with maximum likelihood being used to estimate the unknown coefficients. Here the likelihood is constructed by using asymptotic properties of the periodogram ordinates. Under appropriate conditions, it is shown that the maximum likelihood estimate exists, that it is unique, and that it achieves the usual (optimal) L_2 rate of convergence when the number of knots suitably increases with the length of the observed portion of the time series. These results lend theoretical support to the related, but more adaptive, methodology developed by Kooperberg, Stone and Truong (1994).

2. STATEMENT OF RESULTS

2.1 *Preliminaries*

A stationary process $\{X_t\}$ is a linear process if

$$X_t = \sum_{j=-\infty}^{\infty} a_j Z_{t-j}, \qquad Z_j \sim_{\text{iid}} (0, \sigma^2).$$

The autocovariance function for such a linear process is given by

$$\gamma(u) = \text{cov}(X_t, X_{t+u}) = \sigma^2 \sum_j a_{j-u} a_j,$$

and its spectral density function is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=-\infty}^{\infty} a_j \exp(-ij\lambda) \right|^2, \qquad -\pi \leq \lambda \leq \pi.$$

Let $0 < \gamma \leq 1$. A function g on $[0, \pi]$ is said to satisfy a Hölder condition with exponent γ if there is a positive number c such that $|g(\lambda) - g(\lambda_0)| \leq c|\lambda - \lambda_0|^\gamma$ for $\lambda, \lambda_0 \in [0, \pi]$. Let m be a nonnegative integer and set $p = m + \gamma$. A function g on $[0, \pi]$ is said to be p-smooth if g is m times differentiable on $[0, \pi]$ and $g^{(m)}$ satisfies a Hölder condition with exponent γ. In the following condition, it is assumed that $p > 1/2$.

CONDITION 1. *$\{X_t\}$ is a stationary linear process with $\sum_j |a_j||j|^p < \infty$. Moreover, $Z_j \sim_{\text{iid}} N(0, \sigma^2)$.*

Under this condition, the process is Gaussian and the spectral density function is p-smooth. Moreover, the periodogram

$$I^{(T)}(\lambda) = (2\pi T)^{-1} \left| \sum_{t=0}^{T-1} \exp(-i\lambda t) X_t \right|^2, \quad -\pi \leq \lambda \leq \pi,$$

has the following asymptotic properties:

$$I^{(T)}(\lambda_k) = f(\lambda_k) W_k, \quad \lambda_k = \frac{2\pi k}{T} \quad \text{for} \quad k = 0, \ldots, [T/2],$$

where W_k, $k = 1, \ldots, [(T-1)/2]$, have approximately the exponential distribution with mean one, W_0 and $W_{[T/2]}$ (if T is even) have approximately the χ^2 distribution with one degree of freedom, and $W_0, W_1, \ldots, W_{[T/2]}$ are asymptotically independent; see Brillinger (1981) and Brockwell and Davis (1991).

CONDITION 2. *The spectral density function f is bounded away from zero on $[0, \pi]$.*

Under this condition, let $\varphi = \log f$ denote the logarithm of the spectral density function. It follows from Conditions 1 and 2 that φ is bounded and p-smooth.

2.2 *Polynomial Splines*

Let $K = K_T$ be a positive integer, and let A_k, $1 \le k \le K$, denote the subintervals of $[0, \pi]$ defined by $A_k = [(k-1)\pi/K, k\pi/K)$ for $1 \le k < K$ and $A_K = [\pi(1-1/K), \pi]$. Let m and q be fixed integers such that $m \ge p$ and $m > q \ge -1$. A function g defined on $[0, \pi]$ is called a piecewise polynomial with knot sequence $\pi/K, \dots, (K-1)\pi/K$ if the restriction of g to A_k is a polynomial of degree m (or less) for $1 \le k \le K$, and is callled a spline if g is also q-times continuously differentiable on $[0, \pi]$ for $q \ge 0$. Typically, we assume that $q = m - 1$.

Let S_T denote the space of splines g on $[0, \pi]$ and let B_j, $1 \le j \le J$, denote the usual basis of S_T consisting of B-splines [see de Boor (1978)]. Then $J = (m+1)K - (q+1)(K-1)$, so $K + m \le J \le (m+1)K$. Also, $B_j \ge 0$ on $[0, \pi]$, $B_j = 0$ on the complement of an interval of length $(m+1)\pi/K$ for $1 \le j \le J$, and $\sum_j B_j = 1$ on $[0, \pi]$. Moreover, for $1 \le j \le J$, there are at most $2m + 1$ values of $j' \in \{1, \dots, J\}$ such that $B_j B_{j'}$ is not identically zero on $[0, \pi]$. The following condition is imposed so that the bias term will have the desirable rate of convergence to zero.

CONDITION 3. $J^2 = o(T^{1-\epsilon})$ *for some* $\epsilon > 0$.

Since the spectral density function f is symmetric about zero and is periodic on $[-\pi, \pi]$, we have that $f'(0) = f'''(0) = f'(\pi) = f'''(\pi) = 0$ and $\varphi'(0) = \varphi'''(0) = \varphi'(\pi) = \varphi'''(\pi) = 0$. In this paper, we use splines g on $[0, \pi]$ such that $g'(0) = g'''(0) = g'(\pi) = g'''(\pi) = 0$ to model the logarithm of the spectral density function φ. (Observe that any such spline can be extended to a spline on $\mathbb{R}$ that is periodic with period 2π, symmetric about zero, and does not have a knot at zero or π.) Equivalently, we consider $g(\cdot; \boldsymbol{\beta}) = \beta_1 B_1(\cdot) + \cdots + \beta_J B_J(\cdot)$ with $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J)^t$ constrained to lie in the subspace Ω of $\mathbb{R}^J$ given by

$$\begin{aligned}\Omega \;=\; \{&\boldsymbol{\beta} = (\beta_1, \dots, \beta_J)^T \in \mathbb{R}^J : g'(0) = g'''(0) = g'(\pi) = g'''(\pi) = 0, \\ &\qquad\qquad\qquad \text{where } g = \beta_1 B_1(\cdot) + \cdot + \beta_J B_J(\cdot)\}.\end{aligned}$$

Observe that the collection S_T° of splines $g(\cdot; \boldsymbol{\beta})$, $\boldsymbol{\beta} \in \Omega$, is a subspace of S_T. Given $\boldsymbol{\beta} \in \Omega$, the corresponding spectral density function is modeled as $f(\cdot; \boldsymbol{\beta}) = \exp g(\cdot; \boldsymbol{\beta})$.

2.3 *Maximum Likelihood Estimation*

Set $Y = f(\lambda; \boldsymbol{\beta})W$, where W has the exponential distribution with mean one when $0 < \lambda < \pi$ and it has the χ^2 distribution with one degree of freedom when $\lambda = \pi$. Omitting a term that does not depend on $\boldsymbol{\beta}$, we write the log-likelihood corresponding to the observed value y of Y as

$$\begin{aligned}\psi(y,\lambda;\boldsymbol{\beta}) &= \left(\frac{\delta_\pi(\lambda)}{2}-1\right)[g(\lambda;\boldsymbol{\beta})+y\exp(-g(\lambda;\boldsymbol{\beta}))] \\ &= \left(\frac{\delta_\pi(\lambda)}{2}-1\right)\left[\sum_{j=1}^{J}\beta_j B_j(\lambda)+y\exp\left(-\sum_{j=1}^{J}\beta_j B_j(\lambda)\right)\right]\end{aligned}$$

for $0<\lambda\le\pi$ and $y\ge 0$, where $\delta_\pi(\lambda)=1$ if $\lambda=\pi$ and $\delta_\pi(\lambda)=0$ otherwise. Thus

$$\frac{\partial}{\partial\beta_j}\psi(y,\lambda;\boldsymbol{\beta})=\left(\frac{\delta_\pi(\lambda)}{2}-1\right)B_j(\lambda)\left[1-y\exp(-g(\lambda;\boldsymbol{\beta}))\right]$$

for $1\le j\le J$, $0<\lambda\le\pi$ and $y\ge 0$. Also,

$$\frac{\partial^2}{\partial\beta_j\partial\beta_l}\psi(y,\lambda;\boldsymbol{\beta})=\left(\frac{\delta_\pi(\lambda)}{2}-1\right)yB_j(\lambda)B_l(\lambda)\exp(-g(\lambda;\boldsymbol{\beta}))$$

for $1\le j,l\le J$, $0<\lambda\le\pi$ and $y\ge 0$. It follows from the last result that $\psi(y,\lambda;\cdot)$ is a concave function on $\mathbb{R}^J$ for $y\ge 0$ and $0<\lambda\le\pi$ and that it is strictly concave for $y>0$.

Let $X_0,\ldots,X_{T-1}$ be a realization of length T of the time series. Set

$$\lambda_k=\frac{2\pi k}{T}\quad\text{and}\quad I_k=(2\pi T)^{-1}\left|\sum_{t=0}^{T-1}\exp(-i\lambda_k t)X_t\right|^2,\quad k=1,2,\ldots,[T/2].$$

The (approximate) log-likelihood function corresponding to the periodogram and the J-parameter model for the logarithm of the spectral density function is defined by

$$\ell(\boldsymbol{\beta})=\ell(g(\cdot;\boldsymbol{\beta}))=\sum_{k=1}^{[T/2]}\psi\left(I_k,\lambda_k;\boldsymbol{\beta}\right),\quad \boldsymbol{\beta}\in\mathbb{R}^J.$$

Moreover,

$$\frac{\partial}{\partial\beta_j}\ell(\boldsymbol{\beta})=\sum_{k=1}^{[T/2]}\frac{\partial}{\partial\beta_j}\psi\left(I_k,\lambda_k;\boldsymbol{\beta}\right),\quad 1\le j\le J\text{ and }\boldsymbol{\beta}\in\mathbb{R}^J,$$

and

$$\frac{\partial^2}{\partial\beta_j\partial\beta_l}\ell(\boldsymbol{\beta})=\sum_{k=1}^{[T/2]}\frac{\partial^2}{\partial\beta_j\partial\beta_l}\psi(I_k,\lambda_k;\boldsymbol{\beta}),\quad 1\le j,l\le J\text{ and }\boldsymbol{\beta}\in\mathbb{R}^J,$$

so $\ell(\cdot)$ is a concave function and it is strictly concave if some I_k is positive. It follows that $\ell(\boldsymbol{\beta})$ is also concave for $\boldsymbol{\beta}\in\Omega$. The maximum likelihood estimate $\widehat{\boldsymbol{\beta}}\in\Omega$ is given by $\ell(\widehat{\boldsymbol{\beta}})=\max\{\ell(\boldsymbol{\beta}):\boldsymbol{\beta}\in\Omega\}$, and the corresponding maximum likelihood estimate of φ is given by $\widehat{\varphi}=g(\cdot;\widehat{\boldsymbol{\beta}})$. We refer to $\widehat{f}=\exp\widehat{\varphi}$ as the logspline estimate of the spectral density function f.

2.4 *Expected Log-Likelihood Function*

By Theorem 5.2.2 of Brillinger (1981) or Theorem 10.3.1 of Brockwell and Davis (1991), $E(I_k) = f(\lambda_k) + O(T^{-1})$, where $O(T^{-1})$ is uniform in λ_k. Thus

$$E\left(\frac{1}{[T/2]}\ell(\boldsymbol{\beta})\right) = \frac{1}{[T/2]}\sum_{k=1}^{[T/2]}\left(\frac{\delta_\pi(\lambda_k)}{2} - 1\right)[g(\lambda_k;\boldsymbol{\beta}) + f(\lambda_k)\exp(-g(\lambda_k;\boldsymbol{\beta}))] + O(T^{-1}).$$

Define the (approximate) expected log-likelihood function by

$$\Lambda(a) = \frac{1}{[T/2]}\sum_{k=1}^{[T/2]}\left(\frac{\delta_\pi(\lambda_k)}{2} - 1\right)[a(\lambda_k) + f(\lambda_k)\exp(-a(\lambda_k))],$$

where a is a function on $[0,\pi]$. Note that $\Lambda(a)$ is maximized at $a = \varphi = \log f$. Set $\Lambda(\boldsymbol{\beta}) = \Lambda(g(\cdot;\boldsymbol{\beta}))$ for $g \in S_T$. Then

$$\frac{\partial}{\partial\beta_j}\Lambda(\boldsymbol{\beta}) = \frac{1}{[T/2]}\sum_{k=1}^{[T/2]}\left(\frac{\delta_\pi(\lambda_k)}{2} - 1\right)B_j(\lambda_k)[1 - f(\lambda_k)\exp(-g(\lambda_k;\boldsymbol{\beta}))], \quad 1 \le j \le J,$$

and

$$\frac{\partial^2}{\partial\beta_j\partial\beta_l}\Lambda(\boldsymbol{\beta}) = \frac{1}{[T/2]}\sum_{k=1}^{[T/2]}\left(\frac{\delta_\pi(\lambda_k)}{2} - 1\right)B_j(\lambda_k)B_l(\lambda_k)f(\lambda_k)\exp(-g(\lambda_k;\boldsymbol{\beta})), \quad 1 \le j,l \le J.$$

Thus the expected log-likelihood function $\Lambda(\boldsymbol{\beta})$ is a strictly concave function of $\boldsymbol{\beta}$. Since $\{g(\cdot;\boldsymbol{\beta}) : \boldsymbol{\beta} \in \Omega\}$ is a subspace of S_T, $\Lambda(\boldsymbol{\beta})$ is also strictly concave for $\boldsymbol{\beta} \in \Omega$. Let $\boldsymbol{\beta}^*$ be uniquely defined by $\Lambda(\boldsymbol{\beta}^*) = \max\{\Lambda(\boldsymbol{\beta}) : \boldsymbol{\beta} \in \Omega\}$, and set $\varphi^* = g(\cdot;\boldsymbol{\beta}^*)$.

2.5 *Rate of Convergence*

Given a square-integrable function g on $[0,\pi]$, set $\|g\| = \left(\int_0^\pi |g(\lambda)|^2 d\lambda\right)^{1/2}$. The proof of the following result is given in Section 3.

THEOREM. *Suppose Conditions* 1–3 *hold. Then the maximum likelihood estimate* $\widehat{\varphi}$ *of* φ *exists, and it is unique except on an event whose probability tends to zero as* $T \to \infty$. *Moreover,*

$$\|\widehat{\varphi} - \varphi\| = O_P\left(\sqrt{J/T} + J^{-p}\right).$$

Given positive numbers a_T and b_T for $T \ge 1$, let $a_T \sim b_T$ mean that a_T/b_T is bounded away from zero and infinity. The next result is an immediate consequence of the theorem; here the rate of convergence coincides with that achievable by averaging the periodogram ordinates; see Brillinger (1981, p.251).

COROLLARY. *Suppose Conditions* 1 *and* 2 *hold and that* $J \sim T^{1/(2p+1)}$. *Then*

$$\|\widehat{\varphi} - \varphi\| = O_P(T^{-p/(2p+1)}).$$

3. PROOFS

The proof of the theorem is broken into three parts. The first part considers the bias term, the second proves the existence, uniqueness and consistency of the maximum likelihood estimate, and the third part deals with the variance term.

Throughout this section, it is assumed that Conditions 1–3 hold. Given a function a on $(0, \pi]$, set $\|a\|_\infty = \sup_{0<\lambda\leq\pi} |a(\lambda)|$ and $\|a\|_T^2 = ([T/2])^{-1} \sum_{k=1}^{[T/2]} [a(\lambda_k)]^2$.

3.1 *The Bias*

In order to bound the bias term $\varphi^* - \varphi$, we state a result that, roughly speaking, bounds the second derivative of the log-likelihood function in terms of the discrete L_2 norm $\|\cdot\|_T$. The proof of this result is similar to the proofs of Lemma 5 in Stone (1986) and Lemma 4.2 in Stone (1994).

LEMMA 1. *Let M_0 be a positive constant. Then there are positive constants M_1 and M_2 such that $-M_1\|a - \varphi\|_T^2 \leq \Lambda(a) - \Lambda(\varphi) \leq -M_2\|a - \varphi\|_T^2$ for all functions a on $(0, \pi]$ such that $\|a\|_\infty \leq M_0$.*

To apply Lemma 1, it is necessary to check the boundedness condition based on the L_∞ norm. To this end, we can use Lemma 7 of Stone (1986), which says that there is a positive constant M_3 such that

$$\|g\|_\infty \leq M_3 J^{1/2} \|g\|, \qquad g \in S_T. \tag{1}$$

This result will also be needed in the proof of the existence of the maximum likelihood estimate (see Section 3.2) and in the computation of the variance term (see Section 3.3).

In order to use the L_2 norm to bound the bias term, we need to establish that this norm and its discrete version are equivalent when restricted to S_T. By elementary properties of polynomials and a standard compactness argument, there are positive constants M_4 and M_5 such that if $T \geq M_4 J$, then

$$M_5^{-1}\|g\|^2 \leq \|g\|_T^2 \leq M_5\|g\|^2, \quad g \in S_T. \tag{2}$$

The next result contains the L_2 and L_∞ bounds on the bias term. The proof of this result is similar to the proofs of Lemma 8 of Stone (1986) and Lemma 4.4 of Stone (1994). The L_∞ bound will be used in the proof of the existence of $\hat{\varphi}$ and in bounding the variance term $\hat{\varphi} - \varphi^*$.

LEMMA 2. $\|\varphi^* - \varphi\|^2 = O(J^{-2p})$ *and* $\|\varphi^* - \varphi\|_\infty = O(J^{1/2-p})$.

PROOF. By Condition 3 and Theorem 8.15 of Schumaker (1981), there are periodic polynomial splines g_T° for $T \geq 1$ defined on $[-\pi, \pi]$ with the knot sequence $\{k\pi/K : k =$

$\pm 1, \ldots, \pm(K-1)\}$ and a positive constant c_1 such that $\sup_{\lambda \in [-\pi,\pi]} |g_T^\circ(\lambda) - \varphi(\lambda)| \leq c_1 J^{-p}$. Set $g_T(\lambda) = [g_T^\circ(-\lambda) + g_T^\circ(\lambda)]/2$ for $\lambda \in [0, \pi]$. Then $g_T \in S_T^\circ$. Moreover, $\|g_T - \varphi\|_\infty \leq c_1 J^{-p}$ and hence $\|g_T - \varphi\|_T^2 \leq c_1^2 J^{-2p}$. By Lemma 1, there is a positive constant c_2 such that

$$\Lambda(g_T) - \Lambda(\varphi) \geq -c_2 J^{-2p}. \tag{3}$$

Let b be a positive constant. Choose $g \in S_T^\circ$ with $\|g - \varphi\|^2 = bJ^{-2p}$. Then

$$\|g - g_T\|^2 \leq 2\Big(\|g - \varphi\|^2 + \|\varphi - g_T\|^2\Big) \leq 2(b + c_1^2)J^{-2p}.$$

By (1), for J sufficiently large,

$$\|g\|_\infty \leq \|g - g_T\|_\infty + \|g_T - \varphi\|_\infty + \|\varphi\|_\infty \leq 1 + \|\varphi\|_\infty,$$

since $p > 1/2$. Moreover, $\|g - g_T\|^2 \geq (b/2)J^{-2p}$ for b sufficiently large, since $\|g - g_T\| \geq \|g - \varphi\| - \|g_T - \varphi\|$. It follows from (2) and the inequality $\|g - g_T\|_T^2 \leq 2(\|g - \varphi\|_T^2 + \|\varphi - g_T\|_T^2)$ that, for b sufficiently large, there is a positive constant c_3 such that $\|g - \varphi\|_T^2 \geq c_3 bJ^{-2p}$. Thus by Lemma 1, there is a positive constant c_4 such that, for b and J sufficiently large,

$$\Lambda(g) - \Lambda(\varphi) \leq -c_4 bJ^{-2p} \qquad \text{for all } g \in S_T^\circ \text{ with } \|g - \varphi\|^2 = bJ^{-2p}. \tag{4}$$

Let b be chosen such that $b > \max(c_1^2, c_2/c_4)$ and otherwise sufficiently large. It follows from (3) and (4) that for J sufficiently large,

$$\Lambda(g) < \Lambda(g_T) \qquad \text{for all } g \in S_T^\circ \text{ with } \|g - \varphi\|^2 = bJ^{-2p}.$$

Therefore, $\Lambda(\cdot)$ has a local maximum on $\|g - \varphi\|^2 < bJ^{-2p}$ and hence, by the concavity of $\Lambda(\cdot)$, $\|\varphi^* - \varphi\|^2 < bJ^{-2p}$ for J sufficiently large. Consequently, $\|\varphi^* - g_T\|^2 = O(J^{-2p})$, so we conclude from (1) that $\|\varphi^* - g_T\|_\infty = O(J^{1/2-p})$ and hence that $\|\varphi^* - \varphi\|_\infty = O(J^{1/2-p})$. □

3.2 *The Existence, Uniqueness and Consistency of The Maximum Likelihood Estimate*

The proof of the existence involves several approximations. First we approximate the expected log-likelihood function by the log-likelihood for independent observations. Specifically, suppose Condition 3 holds. Let $\tau_T, T \geq 1$, be positive numbers such that $J\tau_T^2 = O(1)$ and $\log T/\sqrt{T} = o(\tau_T^2)$. Set

$$\tilde{\ell}(g) = \sum_{k=1}^{[T/2]} \left(\frac{\delta_\pi(\lambda_k)}{2} - 1\right) [g(\lambda_k; \boldsymbol{\beta}) + |A(\lambda_k)|^2 I_Z(\lambda_k) \exp(-g(\lambda_k; \boldsymbol{\beta}))],$$

where $A(\lambda) = \sum_j a_j \exp(-ij\lambda)$ and $I_Z(\cdot)$ is the periodogram of Z_t defined by

$$I_Z(\lambda) = I_Z^{(T)}(\lambda) = (2\pi T)^{-1} \left|\sum_{t=0}^{T-1} \exp(-i\lambda t) Z_t\right|^2, \quad \lambda \in [-\pi, \pi].$$

It follows from Condition 1 that the random variables $I_Z(\lambda_k)$ for $1 \le k \le [T/2]$, are independent, $I_Z(\lambda_k)$ has the exponential distribution with mean $\sigma^2/(2\pi)$ for $1 \le k < [T/2]$, and $I_Z(\lambda_k)/E[I_Z(\lambda_k)]$ has the χ^2 distribution with one degree of freedom if T is even and $k = T/2$ (see Theorem 5.2.6 of Brillinger (1981)). Hence $[T/2]^{-1}E[\tilde{\ell}(g)] = \Lambda(g)$ for $g \in S_T$, and it follows by applying the Markov inequality to a suitable moment generating function that the result below is valid (see the proofs of Lemma 10 of Stone (1986) and Lemma 4.6 of Stone (1994)).

LEMMA 3. *Let b and ε be positive constants. There is a positive constant M_6 such that, for T sufficiently large,*

$$P\left(\left|\frac{\tilde{\ell}(g) - \tilde{\ell}(\varphi^*)}{[T/2]} - [\Lambda(g) - \Lambda(\varphi^*)]\right| \ge \varepsilon\tau_T^2\right) \le 2\exp(-M_6 T\tau_T^2)$$

for all $g \in S_T$ with $\|g - \varphi^*\| \le b\tau_T$.

The next result describes the approximation of the observable log-likelihood function $\ell(\cdot)$ by $\tilde{\ell}(\cdot)$.

LEMMA 4. *Let ε and M_7 be positive constants. Then, except on an event with probability tending to zero as $T \to \infty$,*

$$\left|\frac{\tilde{\ell}(g) - \ell(g)}{[T/2]}\right| \le \varepsilon\tau_T^2$$

for all $g \in S_T$ with $\|g\|_\infty \le M_7$.

PROOF. According to Theorem 10.3.1 of Brockwell and Davis (1991),

$$I_k = I^{(T)}(\lambda_k) = |A(\lambda_k)|^2 I_Z(\lambda_k) + R_T(\lambda_k),$$

where

$$R_T(\lambda) = A(\lambda)J_Z(\lambda)Y_T(-\lambda) + A(-\lambda)J_Z(-\lambda)Y_T(\lambda) + |Y_T(\lambda)|^2$$

with

$$J_Z(\lambda) = (2\pi T)^{-1/2}\sum_{t=0}^{T-1}\exp(-i\lambda t)Z_t, \quad \lambda \in [-\pi, \pi],$$

$$Y_T(\lambda) = (2\pi T)^{-1/2}\sum_{j=-\infty}^{\infty} a_j \exp(-i\lambda j)U_{Tj}(\lambda)$$

and

$$U_{Tj}(\lambda) = \sum_{t=-j}^{T-1-j}\exp(-i\lambda t)Z_t - \sum_{t=0}^{T-1}\exp(-i\lambda t)Z_t.$$

It follows from Condition 1 that $\max_k |A(\lambda_k)| \le \sum |a_j| < \infty$ and that $J_Z(\lambda)$ and $Y_T(\lambda)$ have normal distributions with mean zero and variances $O(1)$ and $\sigma_T^2 = O(1/T)$, respectively. Since

$$\int_y^\infty \exp(-x^2/2)dx \le \frac{1}{y}\exp(-y^2/2), \quad y > 0,$$

we see that

$$\max_k |J_Z(\lambda_k)| = O_P(\sqrt{\log T}) \quad \text{and} \quad \max_k |Y_T(\lambda_k)| = O_P\Big(\sqrt{\frac{\log T}{T}}\Big)$$

and hence that

$$\max_k |R_T(\lambda_k)| = O_P\left(\frac{\log T}{\sqrt{T}}\right) = o_P(\tau_T^2). \tag{5}$$

The desired result now follows from the observation that

$$\left|\frac{\tilde{\ell}(g) - \ell(g)}{[T/2]}\right| = \left|\frac{1}{[T/2]}\sum_k R_T(\lambda_k)\exp(-g(\lambda_k;\boldsymbol{\beta}))\right| = O\left(\max_k |R_T(\lambda_k)|\right)$$

for all $g \in S_T$ with $\|g\|_\infty \leq M_7$. □

The next result gives the variability of the log-likelihood function over a small neighborhood.

Lemma 5. *Given positive constants ε and M_8, there is a positive constant M_9 such that, except on an event with probability tending to zero as $T \to \infty$,*

$$\left|\frac{\ell(g_1) - \ell(g_2)}{[T/2]}\right| \leq \varepsilon \tau_T^2$$

for all $g_1, g_2 \in S_T$ with $\|g_1\|_\infty \leq M_8$, $\|g_2\|_\infty \leq M_8$ and $\|g_1 - g_2\|_\infty \leq M_9\tau_T^2$.

Proof. Write $g_1 = g(\cdot;\boldsymbol{\beta}_1)$ and $g_2 = g(\cdot;\boldsymbol{\beta}_2)$. It follows from Theorem 10.3.1 of Brockwell and Davis (1991) and (5) that

$$\begin{aligned}\left|\frac{\ell(g_1) - \ell(g_2)}{[T/2]}\right| &\leq \frac{1}{[T/2]}\sum_{k=1}^{[T/2]} |g(\lambda_k;\boldsymbol{\beta}_2) - g(\lambda_k;\boldsymbol{\beta}_1)| \\ &+ \frac{1}{[T/2]}\sum_{k=1}^{[T/2]} |A(\lambda_k)|^2 I_Z(\lambda_k)|\exp(-g(\lambda_k;\boldsymbol{\beta}_2)) - \exp(-g(\lambda_k;\boldsymbol{\beta}_1))| + o_P(\tau_T^2).\end{aligned}$$

The desired result now follows from Condition 1 and the fact that $I_Z(\lambda_k)$, $k = 1, \ldots, [T/2]$, are independent and have exponential type distributions with mean $\sigma^2/(2\pi)$. □

We now discuss the approximation of certain sets by smaller sets. In fact, according to Lemma 12 of Stone (1986), for positive constants b and c, there is a positive constant M_{10} such that, for T sufficiently large, the set $\{g : \|g - \varphi^*\| \leq b\tau_T\}$ can be covered by $O(\exp(M_{10} J \log T))$ subsets each having diameter at most $c\tau_T^2$. Here the diameter of a subset G of S_T is defined as $\sup\{\|g_1 - g_2\|_\infty : g_1, g_2 \in G\}$.

It follows from the above covering result, Lemma 1 with φ replaced by φ^*, and Lemmas 3–5 and (2) that, for a given positive constant b, except on an event whose probability tends to zero as $T \to \infty$, $\ell(g) < \ell(\varphi^*)$ for all g such that $\|g - \varphi^*\| = b\tau_T$. Consequently, by the strict concavity of $\ell(g)$ as a function of g, the maximum likelihood estimate $\widehat{\varphi}$ exists and is unique except on an event whose probability tends to zero as $T \to \infty$. Moreover, $\|\widehat{\varphi} - \varphi^*\| = o_P(\tau_T)$. Thus we conclude from (1) that $\|\widehat{\varphi} - \varphi^*\|_\infty = o_P(J^{1/2}\tau_T) = o_P(1)$ and hence from Lemma 2 that $\|\widehat{\varphi} - \varphi\| = o_P(1)$. This completes the proof of the existence, uniqueness and consistency of the maximum likelihood estimate $\widehat{\varphi}$.

3.3 *The Variance*

The next task is to bound the variance term $\widehat{\varphi} - \varphi^*$ by establishing an upper bound to $\|\widehat{\varphi} - \varphi^*\|^2$. Let $\mathbf{S}(\boldsymbol{\beta})$ denote the score at $\boldsymbol{\beta}$; that is, the J-dimensional column vector with entries $\partial\ell(\boldsymbol{\beta})/\partial\beta_j$. Let $\mathbf{H}(\boldsymbol{\beta})$ denote the Hessian at $\boldsymbol{\beta}$; that is, the $J \times J$ matrix with entries $\partial^2\ell(\boldsymbol{\beta})/\partial\beta_j\partial\beta_l$. Then

$$\int_0^1 \frac{d}{du}\mathbf{S}(\boldsymbol{\beta}^* + u(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*))du = \mathbf{S}(\widehat{\boldsymbol{\beta}}) - \mathbf{S}(\boldsymbol{\beta}^*).$$

This can further be written as $\mathbf{D}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = \mathbf{S}(\widehat{\boldsymbol{\beta}}) - \mathbf{S}(\boldsymbol{\beta}^*)$, where $\mathbf{D}$ is the $J \times J$ matrix given by

$$\mathbf{D} = \int_0^1 \mathbf{H}(\boldsymbol{\beta}^* + u(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*))du.$$

Since $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^t\mathbf{S}(\widehat{\boldsymbol{\beta}}) = 0$ and $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^t E[\mathbf{S}(\boldsymbol{\beta}^*)] = 0$, we conclude that

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^t\mathbf{D}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = -(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^t\{\mathbf{S}(\boldsymbol{\beta}^*) - E[\mathbf{S}(\boldsymbol{\beta}^*)]\}. \tag{6}$$

We claim that

$$|\mathbf{S}(\boldsymbol{\beta}^*) - E[\mathbf{S}(\boldsymbol{\beta}^*)]|^2 = O_P(T) \tag{7}$$

and that there is a positive constant c_1 such that

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^t\mathbf{D}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \le -c_1 TJ^{-1}|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*|^2 \tag{8}$$

except on an event whose probability tends to zero with T. Since

$$|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^t\{\mathbf{S}(\boldsymbol{\beta}^*) - E[\mathbf{S}(\boldsymbol{\beta}^*)]\}| \le |\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*||\mathbf{S}(\boldsymbol{\beta}^*) - E[\mathbf{S}(\boldsymbol{\beta}^*)]|,$$

it follows from (6)–(8) that $|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*|^2 = O_P(J^2/T)$ and hence that

$$\|\widehat{\varphi} - \varphi^*\|^2 = O_P(J/T). \tag{9}$$

The theorem follows from (9) and Lemma 2.

Proof of (7). Now,

$$E\left[\left\{\frac{\partial\ell(\boldsymbol{\beta}^*)}{\partial\beta_j} - E\left(\frac{\partial\ell(\boldsymbol{\beta}^*)}{\partial\beta_j}\right)\right\}^2\right] = \operatorname{var}\left[\sum_k \left(\frac{\delta_\pi(\lambda_k)}{2} - 1\right) B_j(\lambda_k)[1 - I_k\exp(-g(\lambda_k;\boldsymbol{\beta}^*))]\right].$$

Thus, by Theorem 10.3.2 (ii) of Brockwell and Davis (1991) and a property of splines ($\sum_j B_j = 1$),

$$E[|\mathbf{S}(\boldsymbol{\beta}^*) - E[\mathbf{S}(\boldsymbol{\beta}^*)]|^2] = \sum_j \mathrm{var}\left[\sum_k \left(\frac{\delta_\pi(\lambda_k)}{2} - 1\right) B_j(\lambda_k)[1 - I_k \exp(-g(\lambda_k; \boldsymbol{\beta}^*))]\right] = O(T).$$

Therefore (7) holds.

The proof of (8) depends on the following result.

LEMMA 6. *There is a positive constant M_{11} such that, except on an event whose probability tends to zero as $T \to \infty$,*

$$\frac{1}{T}\sum_k I_k[g(\lambda_k; \boldsymbol{\theta})]^2 \geq M_{11} J^{-1} |\boldsymbol{\theta}|^2, \quad \boldsymbol{\theta} \in \mathbb{R}^J.$$

PROOF. By Theorem 10.3.1 of Brockwell and Davis (1991),

$$\frac{1}{T}\sum I_k[g(\lambda_k; \boldsymbol{\theta})]^2 = \frac{1}{T}\sum_k I_Z(\lambda_k)|A(\lambda_k)|^2[g(\lambda_k; \boldsymbol{\theta})]^2 + \frac{1}{T}\sum R_T(\lambda_k)[g(\lambda_k; \boldsymbol{\theta})]^2, \tag{10}$$

where $|A(\cdot)|^2$ is bounded away from zero by Condition 2. According to (2) and (5),

$$\begin{aligned}\frac{1}{T}\sum R_T(\lambda_k)[g(\lambda_k; \boldsymbol{\theta})]^2 &\leq \max_{\lambda_k} R(\lambda_k) M_5 \|g(\cdot; \boldsymbol{\theta})\|^2 \\ &= O_P\left(\frac{\log T}{\sqrt{T}} J^{-1}|\boldsymbol{\theta}|^2\right) = o_P\left(J^{-1}|\boldsymbol{\theta}|^2\right). \end{aligned} \tag{11}$$

We claim that there is a positive constant c_1 such that, except on an event whose probability tends to zero as $T \to \infty$,

$$\frac{1}{T}\sum_k I_Z(\lambda_k)[g(\lambda_k; \boldsymbol{\theta})]^2 \geq c_1 J^{-1}|\boldsymbol{\theta}|^2, \quad \boldsymbol{\theta} \in \mathbb{R}^J. \tag{12}$$

The desired conclusion follows from (10)–(12).

In verifying (12), we can assume that $q = -1$. It suffices to show that there is a positive constant c_2 such that, except on an event whose probability tends to zero as $T \to \infty$,

$$\frac{1}{T}\sum_{k:\lambda_k \in A_v} I_Z(\lambda_k)[g_v(\lambda_k; \boldsymbol{\theta}_v)]^2 \geq c_2 J^{-1} \qquad \text{for } 1 \leq v \leq K,\ \boldsymbol{\theta}_v \in \mathbb{R}^{m+1} \text{ and } |\boldsymbol{\theta}_v|^2 = 1, \tag{13}$$

where $A_v = [(v-1)\pi/K, v\pi/K)$ for $1 \leq v < K$, $A_K = [\pi(1-1/K), \pi]$ (see Section 2.2), and $g_v(\cdot; \boldsymbol{\theta}_v)$ is the restriction of $g(\cdot; \boldsymbol{\theta})$ to A_v. By applying a simple compactness argument to the set $\{\boldsymbol{\theta}_v \in \mathbb{R}^{m+1}; |\boldsymbol{\theta}_v|^2 = 1\}$ and using the distributional properties of $I_Z(\lambda_k)$, we see that

(13) holds. (A detailed proof of (12) can be found in Kooperberg, Stone and Truong, 1993.) □

PROOF OF (8). Now

$$\boldsymbol{\theta}^t \mathbf{H}(\boldsymbol{\beta})\boldsymbol{\theta} \leq -\frac{1}{2}\sum_k [g(\lambda_k;\boldsymbol{\theta})]^2 I_k \exp(-g(\lambda_k;\boldsymbol{\beta})).$$

By Condition 1, Lemma 2 and the result $\|\widehat{\varphi}-\varphi^*\|_\infty = o_P(1)$ in Section 3.2, there is a positive constant c_1 such that $\|\varphi^*\|_\infty \leq c_1$ and, for T sufficiently large, $\|\widehat{\varphi}\|_\infty \leq c_1$ except on an event whose probability tends to zero as $T \to \infty$. Consequently, there is a positive constant c_2 such that, except on an event whose probability tends to zero as $T \to \infty$,

$$\frac{1}{T}\boldsymbol{\theta}^t \mathbf{D}\boldsymbol{\theta} \leq -c_2 \frac{1}{T}\sum_k I_k [g(\lambda_k;\boldsymbol{\theta})]^2, \quad \boldsymbol{\theta} \in \mathbb{R}^J. \tag{14}$$

Equation (8) follows from (14) and Lemma 6 applied to $\boldsymbol{\theta} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$. This completes the proof of the theorem.

ACKNOWLEDGEMENTS

Charles J. Stone was supported in part by National Science Foundation Grant DMS-920427. Young K. Truong was supported in part by a Research Council Grant from the University of North Carolina at Chapel Hill.

REFERENCES

BELTRÃO, K. I. and BLOOMFIELD, P. (1987) Determining the bandwidth of a kernel spectrum estimate. *J. of Time Ser. Anal.* 8, 21–38.

BRILLINGER, D. R. (1981) *Time Series, Data Analysis and Theory.* San Francisco: Holden-Day.

BROCKWELL, P. J. and DAVIS, R. A. (1991) *Time Series: Theory and Methods.* Second edition, New York: Springer.

DE BOOR, C. (1978) *A Practical Guide to Splines.* New York: Springer.

FRANKE, J. and HÄRDLE, W. (1992) On bootstrapping kernel spectral estimates. *Ann. Statist.* 20, 121–45.

HURVICH, C. M. and BELTRÃO, K. I. (1990) Cross-validatory choice of a spectrum estimate and its connections with AIC. *J. of Time Ser. Anal.* 11, 121–137.

KOOPERBERG, C., STONE, C. and TRUONG, Y. K. (1994) Logspline estimation of a possibly mixed spectral distribution. Manuscript.

KOOPERBERG, C., STONE, C. and TRUONG, Y. K. (1993) Rate of convergence for logspline spectral density estimation. Technical Report No. 396, Department of Statistics, University of California, Berkeley.

POLITIS, D. N. and ROMANO, J. P. (1992) A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.* 20, 1985–2007.

PRIESTLEY, M. B. (1981) *Spectral Analysis and Time Series.* London: Academic Press.

SCHUMAKER, L. L. (1981) *Spline Functions: Basic Theory.* New York: Wiley.

STONE, C. J. (1986) The dimensionality reduction principle for generalized additive models. *Ann. Statist.* 14, 590–606.

— (1994) The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.* 22, 118–184.

SWANEPOEL, J. W. and VAN WYK, J. W. J. (1986) The bootstrap applied to spectral density function estimation. *Biometrika* 73, 135–42.

WAHBA, G. (1980) Automatic smoothing of the log periodogram. *J. Amer. Statist. Assoc.* 75, 122–32.

— and WOLD, S. (1975) Periodic splines for spectral density estimation: The use of cross-validation for determining the degree of smoothing. *Comm. Statist.* 4, 125–41.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

SCHOOL OF PUBLIC HEALTH
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-7400