The Annals of Statistics
1992, Vol. 20, No. 3, 1329–1349

WEAK CONVERGENCE AND ADAPTIVE PEAK ESTIMATION FOR SPECTRAL DENSITIES[1]

BY HANS-GEORG MÜLLER AND KATHRYN PREWITT

University of California, Davis

Adaptive nonparametric kernel estimators for the location of a peak of the spectral density of a stationary time series are proposed and investigated. They are based on direct smoothing of the periodogram where the amount of smoothing is determined automatically in an asymptotically optimal fashion. These adaptive estimators minimize the asymptotic mean squared error. Adaptivity is derived from the weak convergence of a two-parameter stochastic process in a deviation and a bandwidth coordinate to a Gaussian limit process. Efficient global and local bandwidth choices which lead to adaptive peak estimators and practical aspects are discussed.

1. Introduction. In many instances of spectral analysis of a stationary time series, interest focuses on the location of a peak of the spectral density, the "peak frequency." One possibility is to derive estimates of the peak frequency from estimates within a suitable parametric model, either in time or frequency domains. However, parametric assumptions often are not easy to justify or require difficult choices like the determination of the order of an autoregressive model. Therefore, time-honored nonparametric methods like direct smoothing of the periodogram and then reading off the desired peak frequency have kept a prominent place in the repertoire of the time series analyst. These smoothing methods, however, depend critically on the degree of smoothing used, and this paper addresses the question of how to choose the smoothing parameter for the estimation of peaks in such a way that adaptive peak estimates are obtained. Since the peak frequency is a local concept, the desired bandwidth choice will have to be an inherently local one; what happens far away from the peak in question should not unduly influence this bandwidth choice. Adaptive peak estimation means that employing certain data-dependent bandwidths guarantees the same asymptotic distribution of the estimated peak frequency as if the optimal, but unknown bandwidths, which minimize the asymptotic mean squared error of the peak frequency, would have been chosen.

We consider kernel type estimates for spectral densities which are kernel-smoothed periodograms, in contrast to lag window smoothers. Smoothing of periodograms has been considered for a long time for the purpose of spectral

Received August 1990; revised June 1991.

[1]Research supported in part by Air Force Grant AFOSR-89-0386 and NSF Grant DMS-90-02423.

AMS 1980 *subject classifications*. Primary 62M15, 62M10; secondary 62G07, 62G20, 62E20.

Key words and phrases. Bandwidth choice, cumulants, curve estimation, efficiency, kernel estimators, stationary process, tightness, time series, variable bandwidth.

density estimation; see, for instance, Grenander and Rosenblatt (1957), Parzen (1957), Alekseev and Yaglom (1980) and Priestley (1981). For kernel estimators of spectral densities, the problem of local bandwidth choice, in particular, bandwidth choice for spectral peaks, seems not to have been addressed so far. Global bandwidth choice was considered recently in terms of adapting cross-validation for spectral densities [Hurvich (1985) and Beltrão and Bloomfield (1987)].

The asymptotic normality of the estimated peak coordinate was established for lag window type estimators by Newton and Pagano (1983). For smoothed periodograms, this will be one of the consequences (Theorem 3.2) of our main result (Theorem 2.1) on the weak convergence of a two-dimensional stochastic process. The asymptotic basis for the proposed local bandwidth choice is provided in Theorem 3.1, and two practically efficient methods are justified by Theorems 4.1 and 4.2 below.

Newton and Pagano (1983) express the opinion that nonparametric methods for peak frequency estimation are at a disadvantage since no bandwidth choice methods exist as compared to the autoregressive peak frequency approach. In the latter, a parametric autoregressive model is fitted first, and the corresponding spectral density and peak frequency are obtained for this fit [see Ensor and Newton (1988)]. Several well-established methods exist for choosing the order of an autoregressive model [compare, e.g., Hannan and Quinn (1979) or Shumway (1988)], so that this approach can be carried out in a data-dependent automatic manner. While this provides a viable and in many instances satisfying solution to the spectral peak estimation problem, there are instances where this parametric approach entails relatively larger mean squared errors as compared to an adaptive nonparametric method, as will be demonstrated by means of a simulation study in Section 5.

Section 2 contains notation, basic assumptions and the presentation of Theorem 2.1, our central result. The diverse applications of this result, notably the feasibility of adaptive estimation of frequency peaks by efficient global bandwidths (Theorem 3.2) and by efficient local or variable bandwidths (Theorem 3.3) and the asymptotic normality of peak frequencies derived from smoothed periodograms (Theorem 3.1) will be discussed in Section 3.

Section 4 presents two methods to achieve this efficiency (Theorems 4.1 and 4.2). The second of these results implies that adaptive frequency peak estimation is achieved when at each point where the spectral density is to be estimated local bandwidths are chosen which are consistent estimators of the asymptotically optimal local bandwidths for estimation of the first derivative.

Proofs and auxiliary results are compiled in Section 6 and some results on the practical performance of the proposed methods as compared to parametric modeling are discussed in Section 5.

2. Main result. We introduce the necessary notation and state the assumptions that are basic for all of the following. Let $X(t)$, $t = 0, \pm 1, \pm 2$, be a real-valued strictly stationary process with zero expectation for which all moments exist. In addition, assume that for all $r > 0$,

$$\sum_{m_1,\dots,m_{r-1}} \left[1 + |m_j|\right] \left|\operatorname{cum}(X(t+m_1),\dots,X(t+m_{r-1}),X(t))\right| < \infty,$$

$$j = 1,\dots,r-1,$$

where $\operatorname{cum}(X(t+m_1),\dots,X(t+m_{r-1}),X(t))$ is the joint cumulant of order r; see Brillinger (1981), Chapter 2. In addition, assume that given some integer $k \geq 2$, the autocovariance function $c(m) = E(X(t)X(t+m))$ satisfies

$$\sum_{m=-\infty}^{\infty} |c(m)| m^{k+1} < \infty,$$

that is, the spectral density

$$f(\lambda) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} c(m) \exp(-i\lambda m), \qquad 0 \leq \lambda \leq 2\pi,$$

satisfies

(F1) $$f \in \mathscr{C}^{k+1}([0, 2\pi]).$$

Assume that a peak of f is located at some $\theta \in (0, \pi]$ with the following properties:

(F2)
$$f(\lambda) < f(\theta) \quad \text{for all } \lambda \neq \theta,$$
$$f^{(1)}(\theta) = 0, \qquad f^{(2)}(\theta) < 0, \qquad f^{(k+1)}(\theta) \neq 0.$$

Consider frequencies $\lambda_j = (2\pi j)/n$, $j = 1,\dots,n-1$, and the periodogram values

$$I(\lambda_j) = \frac{1}{2\pi n} \left| \sum_{l=0}^{n-1} \exp(-i\lambda_j l) X(l) \right|^2.$$

We discuss the following kernel estimators of $f(\lambda)$:

(2.1) $$\hat{f}(\lambda, s) = \frac{1}{b(s)} \sum_{j=1}^{n-1} \int_{d_{j-1}}^{d_j} K\left(\frac{\lambda - u}{b(s)}\right) du\, I(\lambda_j),$$

where $d_j = (\lambda_j + \lambda_{j+1})/2$, $j = 1,\dots,n-2$, $d_0 = 0$ and $d_{n-1} = 2\pi$. Here K is a kernel function of order k, satisfying

(K1)
$$K \in \mathscr{C}^1(\mathbb{R}),$$
$K^{(1)}$ is Lipschitz continuous on $\mathbb{R}$, support $(K) = [-1, 1]$,

(K2) $$K \in \mathscr{M}_{0,k},$$

where

$$\mathscr{M}_{\nu,k} = \left\{ g \in \mathscr{C}([-1,1]): \int g(x)x^j\,dx = \begin{cases} 0, & 0 \le j < k,\ j \ne \nu, \\ (-1)^{\nu}\nu!, & j = \nu, \\ \ne 0, & j = k, \end{cases} \right\}.$$

The bandwidth sequence $b(s)$ for these estimators is defined to be

$$b(s) = sn^{-1/(2k+3)},$$

with a free parameter s. It will be seen that this is the optimal rate of decay of the bandwidth for the purpose of peak estimation. The central application of our main result (Theorem 2.1 below) is the efficient choice of the parameter s.

Defining the estimator $\hat{\theta}(s)$ of the peak coordinate θ by

$$\hat{\theta}(s) = \inf\left\{\lambda \in [0, 2\pi]: \hat{f}(\lambda, s) = \sup_x \hat{f}(x, s)\right\}, \tag{2.2}$$

we find in Theorem 3.1 below that for fixed $s > 0$,

$$n^{k/(2k+3)}\big(\hat{\theta}(s) - \theta\big) \to_{\mathscr{D}} \mathscr{N}\left(\frac{-s^k f^{(k+1)}(\theta) B_k}{f^{(2)}(\theta)}, \frac{2\pi f(\theta)^2}{s^3 f^{(2)}(\theta)^2} \int K^{(1)}(v)^2\,dv \right), \tag{2.3}$$

defining $B_k = [(-1)^k/k!]\int K(v)v^k\,dv$. Therefore, the asymptotic mean squared error of $\hat{\theta}(s)$ is

$$\text{As.MSE}\big(\hat{\theta}(s)\big) = s^{2k}\left(\frac{f^{(k+1)}(\theta) B_k}{f^{(2)}(\theta)}\right)^2 + \frac{1}{s^3}\frac{2\pi f(\theta)^2}{f^{(2)}(\theta)^2}\int K^{(1)}(v)^2\,dv, \tag{2.4}$$

which is seen to be minimized by

$$s^* = \left(\frac{3\pi f(\theta)^2 \int K^{(1)}(v)^2\,dv}{k f^{(k+1)}(\theta)^2 B_k^2} \right)^{1/(2k+3)}. \tag{2.5}$$

We will show that $\text{As.MSE}(\hat{\theta}(\hat{s}))/\text{As.MSE}(\hat{\theta}(s^*)) \to 1$ as $n \to \infty$ if $\hat{s} \to_p s^*$, that is, that any such procedure is efficient (Theorem 3.2).

Let any σ_1, σ_2, τ satisfying $0 < \sigma_1 < s^* < \sigma_2 < \infty$ and $\tau > 0$ be given. Throughout this paper we set

$$\gamma_n = n^{k/(2k+3)}.$$

Define the sequence of stochastic processes

$$\zeta_n(s,t) = \gamma_n^2\left[\hat{f}(\theta + t\gamma_n^{-1}, s) - \hat{f}(\theta, s)\right], \quad (s,t) \in [\sigma_1, \sigma_2] \times [-\tau, \tau] \tag{2.6}$$

in the two variables s (the "bandwidth" variable) and t (the "deviation" variable). Obviously, $\zeta_n \in \mathscr{C}([\sigma_1, \sigma_2] \times [-\tau, \tau])$.

A related, but simpler process in only one "deviation" variable was considered in the context of estimating modes of densities from i.i.d. data by Eddy

(1980). In the following, $\Rightarrow$ denotes weak convergence of stochastic processes. Our main result is the following theorem.

THEOREM 2.1. *Under* (F1), (F2), (K1O]) *and* (K2),

$$\zeta_n(s,t) \Rightarrow \zeta(s,t) \quad \text{on } \mathscr{C}([\sigma_1,\sigma_2] \times [-\tau,\tau]), \tag{2.7}$$

where $\zeta(s,t)$ *is a two-dimensional Gaussian process characterized by the following properties*:

$$E(\zeta(s,t)) = \frac{t^2}{2} f^{(2)}(\theta) + ts^k f^{(k+1)}(\theta) B_k, \tag{2.8}$$

$$\operatorname{cov}(\zeta(s_1,t_1),\zeta(s_2,t_2)) = \frac{t_1 t_2}{(s_1 s_2)^2} 2\pi f(\theta)^2 \int K^{(1)}\left(\frac{v}{s_1}\right) K^{(1)}\left(\frac{v}{s_2}\right) dv. \tag{2.9}$$

The proof requires several auxiliary results and is deferred to Section 6. The limit process ζ can be written equivalently as

$$\begin{aligned} \zeta(s,t) &= \frac{t^2}{2} f^{(2)}(\theta) + ts^k f^{(k+1)}(\theta) B_k \\ &\quad + \frac{t}{s^2}(2\pi)^{1/2} f(\theta) \int K^{(1)}\left(\frac{v}{s}\right) dW(v), \end{aligned} \tag{2.10}$$

where W denotes the Wiener measure.

Various applications of this result are discussed in the following section.

3. Adaptive frequency peak estimation. Let (F1), (F2), (K1) and (K2) be satisfied for all of the following. Consider the mapping μ defined on $\mathscr{C}([\sigma_1,\sigma_2] \times (-\infty,\infty))$ by

$$\mu(\psi)(s) = \inf\left\{t(s) : \psi(s,t) = \sup_{r \in (-\infty,\infty)} \psi(s,r)\right\}$$

for $\psi \in \mathscr{C}([\sigma_1,\sigma_2] \times (-\infty,\infty))$. Since the limiting process ζ is a random parabola with fixed second derivative [see (2.10)], the mapping μ is measurable and continuous at ζ (2.7), considered as an element of $\mathscr{C}([\sigma_1,\sigma_2] \times (-\infty,\infty))$, equipped with Whitt's metric [compare Theorem 5 of Whitt (1970) and Eddy (1980)].

By a standard expansion argument [see, e.g., Parzen (1962)], in conjunction with Lemma 4.2 below, one can show that

$$\sup_{\sigma_1 \le s \le \sigma_2} \gamma_n(\hat{\theta}(s) - \theta) = O_p(1), \tag{3.1}$$

which implies that it is actually sufficient to consider the mapping μ on spaces $\mathscr{C}([\sigma_1,\sigma_2] \times [-\tau,\tau])$, τ finite; however, for convenience we consider in the following ζ_n, ζ as elements of $\mathscr{C}([\sigma_1, \sigma_2] \times (-\infty, \infty))$. Observe that (2.10) implies

$$\mu(\zeta)(s) = \frac{-s^k f^{(k+1)}(\theta) B_k}{f^{(2)}(\theta)} - \frac{(2\pi)^{1/2} f(\theta)}{s^2 f^{(2)}(\theta)} \int K^{(1)}\left(\frac{v}{s}\right) dW(v), \tag{3.2}$$

and according to the functional mapping theorem,

$$\mu(\zeta_n) \Rightarrow \mu(\zeta). \tag{3.3}$$

Observe that $\zeta_n(s,t)$ is maximized whenever $\theta + \hat{t}\gamma_n^{-1} = \hat{\theta}(s)$, where $\hat{\theta}(s)$ is the peak frequency estimator (2.2). This means that

$$\mu(\zeta_n)(s) = \gamma_n\big(\hat{\theta}(s) - \theta\big). \tag{3.4}$$

From (3.1)–(3.4) we get the following result.

THEOREM 3.1. *For fixed s in $[\sigma_1, \sigma_2]$,*

$$\begin{aligned} &n^{k/(2k+3)}\big(\hat{\theta}(s) - \theta\big) \\ &\quad \to_{\mathscr{D}} \mathscr{N}\left(-\frac{s^k f^{(k+1)}(\theta) B_k}{f^{(2)}(\theta)}, \frac{2\pi f(\theta)^2}{s^3 f^{(2)}(\theta)^2} \int K^{(1)}(v)^2\, dv\right). \end{aligned} \tag{3.5}$$

This result for a fixed bandwidth sequence complements a result by Newton and Pagano (1983) on the asymptotic normality of peak frequency estimators derived from lag window spectral density estimates.

From (3.5) we obtain the expression given in (2.4) for the asymptotic MSE, that is, the MSE as derived from the limiting distribution, and its minimizer is seen to be s^* (2.5). This implies that the optimal asymptotic MSE for kernel frequency peak estimators is given by

$$\begin{aligned} &\text{As.MSE}\big(\hat{\theta}(s^*)\big) \\ &\quad = \left(\left| f^{(k+1)}(\theta) B_k \right|^{6/(2k+3)} \left(2\pi f(\theta)^2 \int K^{(1)}(v)^2\, dv \right)^{2k/(2k+3)} \Big/ f^{(2)}(\theta)^2 \right) \\ &\qquad \times \left(\left(\frac{3}{2k} \right)^{2k/(2k+3)} + \left(\frac{2k}{3} \right)^{3/(2k+3)} \right). \end{aligned} \tag{3.6}$$

The goal of optimizing the estimators thus naturally leads to the following two problems:

1. Optimization of the kernel function by minimizing the leading constant of (3.6) with respect to the kernel. The optimal kernel in this sense would be the solution of the variational problem

$$\begin{aligned} &\left| \int K(v) v^k\, dv \right|^3 \left(\int K^{(1)}(v)^2\, dv \right)^k = \min \\ &\text{subject to } K \in \mathscr{M}_{0,k},\ K(-1) = K(1) = 0. \end{aligned} \tag{3.7}$$

Problems of type (3.7) are discussed in Granovsky and Müller (1991) and explicit solutions are given there under an additional side condition which restricts the number of sign changes of the kernel function. For example, if $k = 2$ in (3.7) and the kernel $K^{(1)}$ is restricted to have at most one sign change, the solution is the kernel $K(v) = (15/16)(1 - v^2)^2$ on $[-1, 1]$.

2. Of even greater practical importance is data-dependent bandwidth choice. Ideally, such a bandwidth selection procedure would have the property that the asymptotic optimal MSE (3.6) is in fact achieved. A fully data-dependent bandwidth choice with this property is *efficient*, and a peak frequency estimator employing such an efficient bandwidth choice and therefore achieving (3.6) is *adaptive*. Obviously, s^* itself is not a possible choice, since it depends on the unknowns $f(\theta)$ and $f^{(k+1)}(\theta)$.

The remainder of this section is devoted to demonstrating that all bandwidth choices $\hat{s}$ with the property $\hat{s} \to_p s^*$ are efficient. In the following section, feasibility of such bandwidth choices will be investigated.

THEOREM 3.2. *Any bandwidth sequence* $b(\hat{s}) = \hat{s}n^{-1/(2k+3)}$ *which satisfies*

$$\hat{s} \longrightarrow_p s^* \tag{3.8}$$

is efficient, and $\hat{\theta}(\hat{s})$ (2.2) *is then adaptive, that is, achieves the optimal asymptotic MSE* (3.6) *of* $\hat{\theta}(s^*)$.

PROOF. We need to show

$$\zeta_n(\hat{s}, t) \Rightarrow \zeta(s^*, t) \quad \text{on } \mathscr{C}([-\tau, \tau]), \tag{3.9}$$

since then $\mu(\zeta_n(\hat{s}, t)) \to_{\mathscr{D}} \mu(\zeta(s^*, t))$ by the functional mapping theorem [see Billingsley (1968)] and $\lim \mathscr{L}(\gamma_n(\hat{\theta}(\hat{s}) - \theta)) = \lim \mathscr{L}(\gamma_n(\hat{\theta}(s^*) - \theta))$ as $n \to \infty$, where $\mathscr{L}(X)$ is the distribution of X. But (3.9) is a direct implication of Theorem 2.1, since bivariate tightness implies that for all $\varepsilon > 0$ there exist $\eta, \delta > 0$ such that

$$P\left(\sup_{t \in [-\tau, \tau],\ |s - s^*| < \delta} |\zeta_n(s, t) - \zeta_n(s^*, t)| > \eta \right) < \varepsilon,$$

which implies that $\zeta_n(\hat{s}, t) = \zeta_n(s^*, t) + o_p(1)$, where $o_p(1)$ is uniform in t. □

Theorem 3.2 provides a class of efficient bandwidth choices which we call "global" bandwidths $b(s) = sn^{-1/(2k+3)}$; these bandwidths do not depend on x respectively t, and hence do not vary locally. An alternative is the choice of local or variable bandwidths. This seems to be a reasonable approach since the peak frequency itself is a local phenomenon. Furthermore, by using varying bandwidths, adaptivity can be achieved without initial pilot estimators for θ which are required for "global" bandwidth choices in order to obtain pilot estimators for $f(\theta)$, $f^{(k+1)}(\theta)$ according to (2.5).

Define a local bandwidth spectral density estimator

$$(3.10)\quad \hat{f}(\lambda, s(\lambda)) = \frac{1}{s(\lambda)n^{-1/(2k+3)}} \sum_{j=1}^{n-1} \int_{d_{j-1}}^{d_j} K\left(\frac{\lambda - u}{s(\lambda)n^{-1/(2k+3)}}\right) du\, I(\lambda_j),$$

$\lambda \in [0, 2\pi]$. Given estimates $\hat{s}(\lambda)$ for $s(\lambda)$, we define a corresponding peak frequency estimate by

$$(3.11)\quad \hat{\theta} = \inf\left\{\lambda\colon \hat{f}(\lambda, \hat{s}(\lambda)) = \sup_{y \in [0, 2\pi]} \hat{f}(y, \hat{s}(y))\right\}.$$

In terms of processes ζ_n, such a choice corresponds to

$$(3.12)\quad \tilde{s}(t) = \hat{s}(\theta + t\gamma_n^{-1}), \qquad t \in (-\infty, \infty),$$

where for $\theta + t\gamma_n^{-1} < 0$ resp. $> 2\pi$, we set the argument equal to 0 resp. 2π.

THEOREM 3.3. *If for all* $\tau > 0$, *with* $\tilde{s}(t)$ *as defined in* (3.12),

$$(3.13)\quad \sup_{t \in [-\tau, \tau]} |\tilde{s}(t) - s^*| = o_p(1),$$

then $\hat{\theta}$ (3.11) *is adaptive.*

PROOF. As in the proof of Theorem 3.2, we need to show that $\zeta_n(\tilde{s}(t), t) \Rightarrow \zeta(s^*, t)$. This is implied by

$$P\left(\sup_t |\zeta_n(\tilde{s}(t), t) - \zeta_n(s^*, t)| > \varepsilon\right)$$
$$\le P\left(\sup_{|s_1 - s_2| \le \delta, t} |\zeta_n(s_1, t) - \zeta_n(s_2, t)| > \varepsilon\right) + P\left(\sup_t |\tilde{s}(t) - s^*| > \delta\right).$$

□

4. Efficient bandwidth choices. A straightforward approach to achieve efficient bandwidth choice according to Theorem 3.2 is to estimate the unknowns $f(\theta)$, $f^{(k+1)}(\theta)$ in the optimal bandwidth constant s^* (2.5) by pilot estimators of f, $f^{(k+1)}$ and θ. Consider derivative estimators

$$(4.1)\quad \hat{f}^{(\nu)}(\lambda, b_\nu) = \frac{1}{b_\nu^{\nu+1}} \sum_{j=1}^{n-1} \int_{d_{j-1}}^{d_j} K_\nu\left(\frac{\lambda - u}{b_\nu}\right) du\, I(\lambda_j), \qquad 0 \le \lambda \le 2\pi,$$

for $f^{(\nu)}(\lambda)$, $0 \le \nu \le k+1$, where we require $K_\nu \in \mathscr{M}_{\nu, k+1}$. The proofs of the following auxiliary results are deferred to Section 6.

LEMMA 4.1. *Assume that* $0 \le \nu \le k+1$, $f \in \mathscr{C}^{k+1}([0, 2\pi])$, $K_\nu \in \mathscr{M}_{\nu, k+1}$ *and* $b_\nu \to 0$. *Then*

$$(4.2)\quad E\hat{f}^{(\nu)}(\lambda, b_\nu) \to f^{(\nu)}(\lambda)$$

uniformly on any compact set contained in $(0, 2\pi)$. *For* $0 \le \nu \le k$, $\lambda \in (0, 2\pi)$, *if* $nb_\nu^{2\nu+1} \to \infty$, $nb_\nu^{2k+3} \to \eta^2$ *for some* $0 \le \eta < \infty$, *then*

$$(4.3)\qquad \left(nb_\nu^{2\nu+1}\right)^{1/2}\left(\hat f^{(\nu)}(\lambda, b_\nu) - f^{(\nu)}(\lambda)\right) \to_{\mathscr{D}} \mathscr{N}\left(\eta B_{k+1} f^{(k+1)}(\lambda), 2\pi f(\lambda)^2 \int K_\nu(v)^2\,dv\right).$$

For $\nu = k+1$, $\lambda \in (0, 2\pi)$, *if* $nb_{k+1}^{2k+3} \to \infty$, *then*

$$(4.4)\qquad \left(nb_\nu^{2k+3}\right)^{1/2}\left(\hat f^{(\nu)}(\lambda, b_\nu) - E\hat f^{(\nu)}(\lambda, b_\nu)\right) \to_{\mathscr{D}} \mathscr{N}\left(0, 2\pi f(\lambda)^2 \int K_\nu(v)^2\,dv\right).$$

LEMMA 4.2. *Assume that* $0 \le \nu \le k+1$, $f \in \mathscr{C}^{k+1}([0, 2\pi])$, $K_\nu \in \mathscr{M}_{\nu, k+1}$ *is Lipschitz and* $b_\nu \to 0$, $nb_\nu^{2\nu+2} \to \infty$. *Then for any* $0 < \delta < \pi$,

$$(4.5)\qquad \sup_{\delta \le \lambda \le 2\pi - \delta} \left|\hat f^{(\nu)}(\lambda, b_\nu) - f^{(\nu)}(\lambda)\right| = o_p(1).$$

LEMMA 4.3. *Assume that* $0 \le \nu \le k+1$, $f \in \mathscr{C}^{k+1}([0, 2\pi])$, *and* $K_\nu \in \mathscr{M}_{\nu, k+1}$ *is Lipschitz continuous. Then*

$$(4.6)\qquad b_\nu^{\nu+1}\gamma_n \to \infty$$

implies that

$$(4.7)\qquad \hat f^{(\nu)}(\theta + t\gamma_n^{-1}, b_\nu) - \hat f^{(\nu)}(\theta, b_\nu) \to_p 0 \quad \text{uniformly in } t \in [-\tau, \tau].$$

From now on, all considerations are restricted to $\lambda \in [\delta, 2\pi - \delta]$ for some small $\delta > 0$. We first consider the efficiency of global bandwidth choices, that is, adaptivity of $\hat\theta(s)$ (2.2) via (3.8). Abbreviating $\hat f^{(\nu)}(\lambda) = \hat f^{(\nu)}(\lambda, b_\nu)$, we observe

$$\left|\hat f^{(\nu)}(\hat\theta) - f^{(\nu)}(\theta)\right| \le \left|\hat f^{(\nu)}(\hat\theta) - f^{(\nu)}(\hat\theta)\right| + \left|f^{(\nu)}(\hat\theta) - f^{(\nu)}(\theta)\right|$$

and $\hat\theta \to_p \theta$ according to Theorem 3.1, as long as a global bandwidth estimator with $b(s) = sn^{-1/(2k+3)}$, s fixed, is chosen. Lemma 4.2 then implies

$$(4.8)\qquad \hat f^{(\nu)}(\hat\theta) \to_p f^{(\nu)}(\theta) \quad \text{for } 0 \le \nu \le k+1,$$

and hence we have the following result.

THEOREM 4.1. *Assume that for an arbitrary value* $s_0 \in [\sigma_1, \sigma_2]$, $\hat f(\cdot, s_0)$ (2.1) *is employed to obtain* $\hat\theta(s_0)$ (2.2) *and* $\hat f^{(k+1)}(\hat\theta(s_0), b_{k+1})$ (4.1) *to estimate* $f^{(k+1)}(\theta)$, *where* $K_\nu \in \mathscr{M}_{k+1, k+1}$ *and* b_{k+1} *satisfies* $b_{k+1} \to 0$, $nb_{k+1}^{2k+4} \to \infty$. *Then, with* s^* *as defined in* (2.5),

$$(4.9)\qquad \hat s = \left(\frac{3\pi \hat f\left(\hat\theta(s_0), s_0\right)^2 \int K^{(1)}(v)^2\,dv}{k \hat f^{(k+1)}\left(\hat\theta(s_0), b_{k+1}\right)^2 B_k^2}\right)^{1/(2k+3)} \to_p s^*,$$

that is, provides efficient global bandwidth choice for frequency peak estimation. The corresponding peak frequency estimator

$$\hat{\theta} = \inf\left\{\lambda\colon \hat{f}(\lambda, \hat{s}) = \sup_{y \in [\delta, 2\pi - \delta]} \hat{f}(y, \hat{s})\right\} \tag{4.10}$$

is adaptive.

This implies that adaptive peak estimation according to Theorem 3.2 is feasible and can be achieved by the two-step procedure provided in Theorem 4.1.

An alternative one-step method which does not require a preliminary estimate of θ can be obtained via Theorem 3.3 and Lemma 4.3. It is motivated by Lemma 4.1: Consider the case $\nu = 1$, obtain the asymptotic mean squared error from (4.3) and observe

$$B_{k+1} = \frac{(-1)^{k+1}}{(k+1)!} \int K^{(1)}(v) v^{k+1}\, dv = \frac{(-1)^k}{k!} \int K(v) v^k\, dv = B_k.$$

Then the optimal local bandwidth for the first derivative at any point λ is $b^*(\lambda) = s^*(\lambda) n^{-1/(2k+3)}$, with

$$s^*(\lambda) = \left(\frac{3\pi f(\lambda)^2 \int K^{(1)}(v)^2\, dv}{k f^{(k+1)}(\lambda)^2 B_k^2} \right)^{1/(2k+3)}, \tag{4.11}$$

so that interestingly $s^*(\theta) = s^*$ where s^* is given by (2.5). Therefore, we define

$$\hat{\theta} = \inf\left\{\lambda\colon \hat{f}(\lambda, \hat{s}(\lambda)) = \sup_{y \in [\delta, 2\pi - \delta]} \hat{f}(y, \hat{s}(y))\right\}, \tag{4.12}$$

where

$$\hat{s}(\lambda) = \left(\frac{3\pi \hat{f}(\lambda, b_0) \int K^{(1)}(v)^2\, dv}{k \hat{f}^{(k+1)}(\lambda, b_{k+1})^2 B_k^2} \right)^{1/(2k+3)}. \tag{4.13}$$

Observe that $\hat{\theta}$ (4.12) is of type (3.11). The estimator $\hat{f}(\lambda, \hat{s}(\lambda))$ can be interpreted as a spectral density estimator with locally varying bandwidths which are consistent estimates of the optimal local bandwidths for estimating the first derivative of the spectral density.

In order to demonstrate adaptivity of the peak frequency estimators $\hat{\theta}$ (4.12), one needs to establish condition (3.13) of Theorem 3.3 for $\hat{s}(\lambda)$ (4.13). But (3.13) would follow immediately from

$$\sup_{t \in [-\tau, \tau]} \left| \hat{f}^{(l)}(\theta + t\gamma_n^{-1}, b_l) - f^{(l)}(\theta) \right| \to_p 0 \quad \text{for } l = 0, k+1. \tag{4.14}$$

Now (4.14) follows from Lemma 4.3 (4.7) and Lemma 4.1 (4.2)–(4.4), if

$$(4.15)\qquad \begin{aligned} &b_0 \to 0, \quad nb_0^{2+3/k} \to \infty, \quad \limsup\left(nb_0^{2k+3}\right) < \infty, \\ &b_{k+1} \to 0, \quad nb_{k+1}^{(k+2)(2k+3)/k} \to \infty \quad \text{as } n \to \infty. \end{aligned}$$

We have shown the following result.

THEOREM 4.2. *If* (4.15) *is satisfied, the bandwidth choice* $\hat{s}(\lambda)$ (4.13) *is efficient and the peak frequency estimator* $\hat{\theta}$ (4.12) *is adaptive.*

In applications, confidence intervals for adaptive estimators $\hat{\theta}$ (4.10) or (4.12) are sometimes of interest. The asymptotic distribution of these estimators is given by Theorem 3.1 (3.5), replacing s by s^* (2.5). Uniformly consistent estimators for the unknown limiting bias and variance are then available according to Lemma 4.2, and asymptotic $100(1-\alpha)\%$ confidence intervals can be constructed by inserting such estimates.

5. Simulation examples. In the following, practical versions of the "global" bandwidth choice method of Theorem 4.1 and of the "local" bandwidth choice method of Theorem 4.2 are developed. The mean squared errors of corresponding peak location estimates are compared with those obtained by fitting autoregressive models with automatic order selection; as the order selection method, Akaike's information criterion [AIC, Akaike (1973)] is used. The mean squared errors for the various methods and examples considered are estimated from 400 Monte Carlo runs per example.

Choosing $k = 2$, define $\hat{R}(b_0) = \Sigma_j \hat{f}(\tau_j, b_0)^2 / \Sigma_j \hat{f}^{(3)}(\tau_j, b_0)^2$ and an initial global bandwidth by

$$\hat{b} = \left(\frac{3\pi}{2} \frac{\int K^{(1)}(v)^2\, dv}{B_2^2} \frac{1}{n} \right)^{1/7} \hat{R}(b_0)^{1/7},$$

where $0 = \tau_1 < \tau_2 < \cdots < \tau_m = 2\pi$ are equidistant on $[0, 2\pi]$ for some large m, which was chosen to be $m = 300$ and b_0 is a small starting bandwidth. Then, in obvious notation, the "global" adaptive bandwidth for peak estimation is defined as

$$(5.1)\quad \hat{b}_{\text{glob}} = \hat{b}\left[\left[\left(\hat{f}(\hat{\theta}(\hat{b}), \hat{b})^2 \Big/ \left(\hat{f}^{(3)}(\hat{\theta}(\hat{b}), \hat{b})^2 \hat{R}(b_0)\right)\right)^{1/7} \wedge 1.4\right] \vee 0.6\right],$$

whereas "local" adaptive bandwidths for estimation at points τ_j, $1 \le j \le m$, are defined as

$$(5.2)\quad \hat{b}_{\text{loc}}(\tau_j) = \hat{b}\left[\left[\left(\hat{f}(\tau_j, \hat{b})^2 \Big/ \left(\hat{f}^{(3)}(\tau_j, \hat{b})^2 \hat{R}(b_0)\right)\right)^{1/7} \wedge 1.4\right] \vee 0.6\right].$$

The truncation serves to stabilize the otherwise high variability of global and local bandwidth choices (4.9) and (4.13). The estimated local bandwidths

$\hat{b}_{loc}(\cdot)$ are subjected to an additional smoothing step with a simple five-point smoother, with weights $(2/15, 3/15, 5/15, 3/15, 2/15)$. Kernels used were $K_0(x) = (3/4)(1 - x^2)$ to estimate $\hat{f}$ and kernel $K_3(x) = (315/32)(18x - 60x^3 + 42x^5)$ to estimate $\hat{f}^{(3)}$, both restricted to $[-1, 1]$.

The following models were considered:

I $$X(t) = X(t-1) - 0.9X(t-2) + \varepsilon(t),$$

a stationary autoregressive process of order 2 with characteristic roots near the unit circle and the values $\theta = 1.0148$, $f(\theta) = 22.0369$ and $f^{(3)}(\theta) = -3.1405 \times 10^4$.

II $$X(t) = X(t-1) - 0.6X(t-2) + \varepsilon(t),$$

another stationary autoregressive process of order 2 with characteristic roots a little further away from the unit circle and values $\theta = 0.8411$, $f(\theta) = 1.7052$ and $f^{(3)}(\theta) = -1.3088 \times 10^2$.

III $$X(t) = 1.387X(t-1) - 0.471X(t-2) - 0.127X(t-3) + \varepsilon(t),$$

a stationary autoregressive process of order 3 with values $\theta = 0.4112$, $f(\theta) = 9.7645$ and $f^{(3)}(\theta) = 5.0156 \times 10^3$.

IV $$\begin{aligned} X(t) &= 1.905X(t-1) - 1.981X(t-2) + 1.349X(t-3) \\ &\quad - 0.791X(t-4) + 0.352X(t-5) - 0.09X(t-6) + \varepsilon(t), \end{aligned}$$

a stationary autoregressive process of order 6 with values $\theta = 0.6827$, $f(\theta) = 5.9194$ and $f^{(3)}(\theta) = -6.8074 \times 10^2$.

V $$\begin{aligned} X(t) &= 1.84X(t-1) - 1.44X(t-2) + 1.12X(t-3) - 1.07X(t-4) \\ &\quad + 0.88X(t-5) - 0.87X(t-6) + 0.87X(t-7) \\ &\quad - 0.73X(t-8) + 0.6X(t-9) - 0.63X(t-10) \\ &\quad + 0.63X(t-11) - 0.53X(t-12) + 0.4X(t-13) \\ &\quad - 0.36X(t-14) + 0.22X(t-15) + \varepsilon(t), \end{aligned}$$

a stationary autoregressive process of order 15 with values $\theta = 0.2834$, $f(\theta) = 6.909 \times 10^2$ and $f^{(3)}(\theta) = -5.7252 \times 10^8$.

VI $$X(t) = 1.3X(t-1) - 0.7X(t-2) + \varepsilon(t) - 0.7\varepsilon(t-1),$$

an autoregressive moving average process of orders 2 and 1, ARMA(2, 1), with values $\theta = 0.6966$, $f(\theta) = 1.7869$ and $f^{(3)}(\theta) = -3.2158 \times 10^2$.

For all processes, $\varepsilon(t)$, $t = 0, 1, \dots,$ were simulated as independent $\mathscr{N}(0, 1)$ pseudo-random variables. For each Monte Carlo run, $n = 100$ data points were generated for each of these processes, and an autoregressive process was fitted with order selection according to the AIC, considering orders up to 30. Then the location of the maximum of the corresponding spectral density was computed, denoted as the AIC estimate. Peak estimates based on methods (5.1)—GLOB estimate, and (5.2)—LOC estimate, were obtained by choosing $b_0 = 0.2$, which corresponds to approximately six points in the initial smoothing window. As long as extreme values are avoided, the results do not depend

TABLE 1

Mean squared errors averaged from 400 Monte Carlo runs, for AIC-based autoregressive (AIC), global bandwidth (5.1)-based (GLOB) and local bandwidth (5.2)-based (LOC) spectral density peak location estimates for models I–VI with $n = 100$. Small numbers denote powers of 10 by which to multiply

Model		AIC	GLOB	LOC
I	AR(2)	2.455^{-2}	2.108^{-3}	2.172^{-3}
II	AR(2)	5.700^{-2}	3.993^{-2}	4.043^{-2}
III	AR(3)	2.101^{-2}	3.104^{-2}	3.186^{-2}
IV	AR(6)	6.481^{-2}	6.126^{-2}	6.017^{-2}
V	AR(15)	8.923^{-3}	7.136^{-3}	7.177^{-3}
VI	ARMA(2, 1)	1.130^{-1}	1.134^{-2}	1.115^{-2}

critically on the choice of b_0. We found that the nonparametric procedures required considerably less computing time than the AIC-based autoregressive fits.

The results are shown in Table 1. It is evident, that the nonparametric adaptive methods work well when compared to autoregressive modeling, in particular for less well behaved models. The global method (5.1) appears to be slightly advantageous.

6. Auxiliary results and proofs. A sequence of lemmas is given which leads to the proof of Theorem 2.1. It is assumed that (F1), (F2), (K1) and (K2) are satisfied. We first investigate the moment structure of processes ζ_n.

LEMMA 6.1.

$$E\zeta_n(s,t) = \frac{t^2}{2} f^{(2)}(\theta) + s^k t f^{(k+1)}(\theta) B_k + o(1),$$

where the o-term is uniform in $(s,t) \in [\sigma_1, \sigma_2] \times [-\tau, \tau]$.

PROOF. According to Brillinger (1981), Theorem 5.2.4,

$$\begin{aligned} E\zeta_n(s,t) &= \frac{\gamma_n^2}{b(s)} \sum_{j=1}^{n-1} \left[\int_{d_{j-1}}^{d_j} \left\{ K\left(\frac{\theta + t\gamma_n^{-1} - u}{b(s)} \right) - K\left(\frac{\theta - u}{b(s)} \right) \right\} du \right. \\ &\qquad \left. \times \left\{ f(\lambda_j) + O(n^{-1}) \right\} \right] \\ &= A_n + B_n. \end{aligned}$$

For the remainder term, due to the Lipschitz continuity of K, $|B_n| = O(n^{-(k+2)/(2k+3)})$ uniformly in s, t, owing to the fact that K has compact support $[-1,1]$, which implies

$$\sum_{j=1}^{n-1}\int_{d_{j-1}}^{d_j} 1_{\{|\theta-v|\le b(s)\}\cup\{|\theta-v+t\gamma_n^{-1}|\le b(s)\}}\,dv = O(n^{-1/(2k+3)}) \tag{6.1}$$

uniformly in s,t. Furthermore, using the Lipschitz continuity of f on $[0,2\pi]$, by an analogous argument,

$$\left|A_n - \frac{\gamma_n^2}{b(s)}\int_0^{2\pi}\left\{K\left(\frac{\theta+t\gamma_n^{-1}-u}{b(s)}\right)-K\left(\frac{\theta-u}{b(s)}\right)\right\}f(u)\,du\right|$$
$$= O(n^{-(k+2)/(2k+3)})$$

uniformly in s,t. Therefore, by substitution and Taylor expansion,

$$\begin{aligned} E\zeta_n(s,t) &= \gamma_n^2\int_{-1}^{1}K(v)\bigg[f^{(1)}(\theta-vb(s))(t\gamma_n^{-1}) \\ &\qquad + \frac{1}{2}f^{(2)}(\theta-vb(s))(t\gamma_n^{-1})^2 + O\big((t\gamma_n^{-1})^3\big)\bigg]\,dv + o(1) \\ &= C_n + D_n + E_n + o(1), \end{aligned}$$

where the o-term is uniform in s,t. Now $E_n = O(t^3\gamma_n^{-1}) = o(1)$, $D_n = (t^2/2)f^{(2)}(\theta) + o(1)$, where the o-terms are uniform in s,t. The result follows by another Taylor expansion of $f^{(1)}$ around θ in C_n, using the moment conditions on the kernel function and observing $f^{(1)}(\theta) = 0$. □

Define now processes $\phi_n(s,t) = \zeta_n(s,t) - E\zeta_n(s,t)$,

$$\begin{aligned} \phi_n(s,t) &= \frac{\gamma_n^2}{b(s)}\sum_{j=1}^{n-1}\int_{d_{j-1}}^{d_j}\left\{K\left(\frac{\theta+t\gamma_n^{-1}-u}{b(s)}\right)-K\left(\frac{\theta-u}{b(s)}\right)\right\}du \\ &\qquad \times\{I(\lambda_j)-EI(\lambda_j)\}. \end{aligned} \tag{6.2}$$

LEMMA 6.2.

$$\begin{aligned} &\operatorname{cov}(\zeta_n(s_1,t_1),\zeta_n(s_2,t_2)) \\ &\quad = \frac{t_1t_2}{(s_1s_2)^2}2\pi f^2(\theta)\int K^{(1)}\left(\frac{v}{s_1}\right)K^{(1)}\left(\frac{v}{s_2}\right)dv + o(1) \end{aligned} \tag{6.3}$$

uniformly in s_1, t_1, s_2, t_2.

PROOF. Defining

$$W_j(s,t) = \frac{\gamma_n^2}{b(s)}\int_{d_{j-1}}^{d_j}\left\{K\left(\frac{\theta+t\gamma_n^{-1}-u}{b(s)}\right)-K\left(\frac{\theta-u}{b(s)}\right)\right\}du, \tag{6.4}$$

we find

$$|W_j(s,t)| = O(n^{-(k+1)/(2k+3)}) \quad \text{uniformly in } j,s,t. \tag{6.5}$$

With (6.1) this implies

$$\sum_{j=1}^{n-1} \left| W_j(s,t) \right| = O(n^{(k+1)/(2k+3)}),$$

$$\begin{aligned} \operatorname{cov}(\zeta_n(s_1,t_1),\zeta_n(s_2,t_2)) &= \operatorname{cov}(\phi_n(s_1,t_1),\phi_n(s_2,t_2)) \\ &= \sum_{j=1}^{n-1} W_j(s_1,t_1)W_j(s_2,t_2)\{f^2(\lambda_j) + O(n^{-1})\} \\ &\quad + \sum_{\substack{l,k=1 \\ l\neq k}}^{n-1} W_l(s_1,t_1)W_k(s_2,t_2)O(n^{-1}) \\ &= A_n + B_n, \end{aligned} \tag{6.6}$$

according to Brillinger (1981), Theorem 5.2.4 where $\lambda_j = 2\pi j/n$. By (6.5), we find $B_n = O(n^{-1/(2k+3)})$ uniformly in s_1, t_1, s_2, t_2, and

$$A_n = \sum_{j=1}^{n-1} W_j(s_1,t_1)W_j(s_2,t_2)\, f^2(\lambda_j) + O(n^{-1})$$

uniformly. Observing

$$W_j(s,t) = \frac{t\gamma_n}{b(s)^2}\int_{d_{j-1}}^{d_j} K^{(1)}\left(\frac{\theta-u}{b(s)}\right) du + O(n^{-2k/(2k+3)}),$$

we obtain for the leading term of A_n,

$$\begin{aligned} &\sum_{j=1}^{n-1} W_j(s_1,t_1)W_j(s_2,t_2)\, f^2(\lambda_j) \\ &= \frac{t_1t_2\gamma_n^2}{b(s_1)^2 b(s_2)^2} \sum_{j=1}^{n-1} \int_{d_{j-1}}^{d_j} K^{(1)}\left(\frac{\theta-u}{b(s_1)}\right) du \\ &\qquad \times \int_{d_{j-1}}^{d_j} K^{(1)}\left(\frac{\theta-u}{b(s_2)}\right) du\, f^2(\lambda_j) + O(n^{-(k-1)/(2k+3)}), \end{aligned} \tag{6.7}$$

whence the result follows by a simple integral approximation and transformation. □

Turning now to the asymptotic normality of processes ϕ_n at a point, we employ the method of cumulants, following Leonov and Shiryaev (1959), compare Brillinger (1981).

LEMMA 6.3. *For fixed* $(s_1,t_1),\dots,(s_\kappa,t_\kappa)$, $\kappa \geq 1$,

$$(\phi_n(s_1,t_1),\dots,\phi_n(s_\kappa,t_\kappa))^T \to_{\mathscr{D}} \mathscr{N}_\kappa(0,C), \tag{6.8}$$

where $C = (c_{ij})$, $1 \le i,\ j \le \kappa$ *with*

$$c_{ij} = \frac{t_i t_j}{(s_i s_j)^2} 2\pi f^2(\theta) \int K^{(1)}\left(\frac{v}{s_i}\right) K^{(1)}\left(\frac{v}{s_j}\right) dv.$$

PROOF. For any given constants γ_i, $1 \le i \le \kappa$, define $Z_n = \sum_{i=1}^{\kappa} \gamma_i \phi_n(s_i, t_i)$. It is sufficient to show that all cumulants of order $r \ge 3$ vanish asymptotically. Define $W_j(s,t)$ as in (6.4) and

$$J(s,t) = \left\{ j \in \{1, \dots, n-1\} : 1_{(\{|\theta - v| \le b(s)\} \cup \{|\theta - v + t\gamma_n^{-1}| \le b(s)\} \cap [d_{j-1}, d_j])} \ne 0 \right\}.$$

Observe that for the cardinality $|J|$ of $J = \bigcup_{\mu=1}^{\kappa} J(s_\mu, t_\mu)$, $|J| = O(nb)$, or $|J| = O(n^{2(k+1)/(2k+3)})$. Then we obtain for the rth-order cumulant of Z_n, $r \ge 0$:

$$\begin{aligned} \operatorname{cum}(Z_n, \dots, Z_n) &= \sum_{(j_1, \dots, j_r) \in J^r}^{n-1} \prod_{i=1}^{r} \sum_{\mu=1}^{\kappa} \gamma_\mu W_{j_i}(s_\mu, t_\mu) \operatorname{cum}\big(I(\lambda_{j_1}), \dots, I(\lambda_{j_r})\big) \\ &= O(n^{-r(k+1)/(2k+3)}) \sum_{(j_1, \dots, j_r) \in J^r} \operatorname{cum}\big(I(\lambda_{j_1}), \dots, I(\lambda_{j_r})\big), \end{aligned} \tag{6.9}$$

according to (6.5). Define

$$d(\lambda) = \sum_{t=0}^{n-1} X(t) e^{-i\lambda t}, \qquad \Delta(\lambda) = \sum_{t=0}^{n-1} e^{-(2\pi\lambda/n)t}.$$

Following Leonov and Shiryaev (1959) [see Theorem 2.3.2 of Brillinger (1981)],

$$\begin{aligned} &\operatorname{cum}\big(I(\lambda_{j_1}), \dots, I(\lambda_{j_r})\big) \\ &\quad = \frac{1}{(2\pi n)^r} \operatorname{cum}\big(d(\lambda_{j_1})\, d(-\lambda_{j_1}), \dots, d(\lambda_{j_r}) d(-\lambda_{j_r})\big) \\ &\quad = O(n^{-r}) \sum_{\rho} C_{\rho_1} \cdots C_{\rho_p}, \end{aligned} \tag{6.10}$$

where the last summation is over all indecomposable partitions $\rho = (\rho_1, \dots, \rho_p)$ of the table consisting of the two columns $(-j_1, \dots, -j_r)$ and $(j_1, \dots, j_r)$ and $\rho_l = [r_{li},\ i = 1, \dots, m_l\}$, $C_{\rho_l} = \operatorname{cum}(d(\lambda_{r_{l1}}), \dots, d(\lambda_{r_{lm_l}}))$, as in the proof of Theorem 7.4.4 of Brillinger (1981). According to Theorem 4.3.2 of Brillinger (1981), $C_{\rho_l} = O(1)(\Delta(\sum_{i=1}^{m_l} \lambda_{r_{li}}) + 1)$ and

$$\Delta\left(\sum_{i=1}^{m_l} \lambda_{r_{li}}\right) = \begin{cases} n, & \text{if } \sum_{i=1}^{m_l} r_{li} = \kappa n,\ \kappa = 0, \pm 1, 2, \dots, \\ 0, & \text{otherwise.} \end{cases} \tag{6.11}$$

Therefore, in obvious notation,

$$\prod_{l=1}^{p} C_{\rho_l} = \sum_{s=0}^{p} n^q O(1) 1_{\{\prod \Delta(\cdots) = n^q\}}$$

and, observing that (6.11) imposes at least $q-1$ linear constraints on the partitions contributing to n^q,

$$\sum_{(j_1,\dots,j_r)\in J^r} \operatorname{cum}\big(I(\lambda_{j_1}),\dots,I(\lambda_{j_r})\big) = O(n^{-r})\sum_{\rho}\sum_{q=0}^{p} |J|^{r-(q-1)} n^q. \tag{6.12}$$

Observing that we may assume $p \le r$, we obtain

$$\begin{aligned}&\sum_{(j_1,\dots,j_r)\in J^r} \operatorname{cum}\big(I(\lambda_{j_1}),\dots,I(\lambda_{j_r})\big)\\ &\quad = O\Big(n \max_{0\le p\le r} b^{r-p+1}\Big) = O(nb) = O\big(n^{2(k+1)/(2k+3)}\big).\end{aligned} \tag{6.13}$$

This together with (6.5) implies $\operatorname{cum}(Z_n,\dots,Z_n) = O(n^{(2-r)(k+1)/(2k+3)}) \to 0$ for $r \ge 3$, that is, asymptotic normality. □

Next we show tightness. In order to apply a criterion of Bickel and Wichura (1971), we require bounds on fourth moments of differences.

LEMMA 6.4. *There exists a constant c such that*

$$E(\phi_n(s,t_2) - \phi_n(s,t_1))^4 \le c(t_2 - t_1)^4 \quad \textit{uniformly in } s \in [\sigma_1, \sigma_2]. \tag{6.14}$$

PROOF. Define

$$\begin{aligned}&w_j(s,t_1,t_2)\\ &\quad = \frac{\gamma_n^2}{b(s)}\int_{d_{j-1}}^{d_j}\left[K\left(\frac{\theta + t_2\gamma_n^{-1} - u}{b(s)}\right) - K\left(\frac{\theta + t_1\gamma_n^{-1} - u}{b(s)}\right)\right]du\end{aligned} \tag{6.15}$$

and observe that by Lipschitz continuity of K,

$$|w_j(s,t_1,t_2)| = O(n^{-(k+1)/(2k+3)})|t_2 - t_1| \quad \text{uniformly in } s. \tag{6.16}$$

Let M_r be the rth-order cumulant of

$$\phi_n(s,t_2) - \phi_n(s,t_1) = \sum_{j=1}^{n-1} w_j(s,t_1,t_2)\big(I(\lambda_j) - EI(\lambda_j)\big). \tag{6.17}$$

For $r = 1$ we find $M_1 = 0$; for $r \ge 2$,

$$|M_r| = O(n^{(2-r)(k+1)/(2k+3)})|t_2 - t_1|^r, \tag{6.18}$$

according to (6.13) and (6.16). Therefore, $|M_2| = O(1)|t_2 - t_1|^2$, $|M_4| = O(n^{(2-r)(k+1)/(2k+3)})|t_2 - t_1|^4$, where the O-terms are uniform in s. Observing

$$\begin{aligned}E(\phi_n(s,t_2) - \phi_n(s,t_1))^4 &= M_4 + 4M_3M_1 + 3M_2^2 + 6M_2M_1^2 + M_1^4\\ &\le |M_4| + 3M_2^2,\end{aligned}$$

the result follows. □

LEMMA 6.5.

$$E(\phi_n(s_2,t) - \phi_n(s_1,t))^4 \le c(s_2 - s_1)^4 \quad \textit{uniformly in } t \in [-\tau,\tau]. \tag{6.19}$$

PROOF. Define

$$\tilde{w}_j(s_1,s_2,t) = W_j(s_2,t) - W_j(s_1,t), \tag{6.20}$$

where W_j is defined in (6.4). Proceeding analogously as in the proof of Lemma 6.4, we first establish a bound on $\tilde{w}_j$. All O-terms are uniform in s_1, s_2, t:

$$\begin{aligned}
&|\tilde{w}_j(s_1,s_2,t)|\\
&\quad\le n^{(2k+1)/(2k+3)}\Bigg(\int_{d_{j-1}}^{d_j} |s_2^{-1} - s_1^{-1}|\left|K\left(\frac{\theta + t\gamma_n^{-1} - u}{b(s_2)}\right) - K\left(\frac{\theta - u}{b(s_2)}\right)\right| du\\
&\qquad + s_1^{-1}\int_{d_{j-1}}^{d_j}\left|K\left(\frac{\theta + t\gamma_n^{-1} - u}{b(s_2)}\right) - K\left(\frac{\theta - u}{b(s_2)}\right)\right.\\
&\qquad\qquad \left. - \left(K\left(\frac{\theta + t\gamma_n^{-1} - u}{b(s_1)}\right) - K\left(\frac{\theta - u}{b(s_1)}\right)\right)\right| du\Bigg)\\
&\quad = |s_2 - s_1| O(n^{-(k+1)/(2k+3)})\\
&\qquad + O(n^{-2/(2k+3)}) \sup_{u\in[0,2\pi],\, t\in[-\tau,\tau]} |H_n(s_2,t,u) - H_n(s_1,t,u)|,
\end{aligned} \tag{6.21}$$

where

$$H_n(s,t,u) = K\left(\frac{\theta + t\gamma_n^{-1} - u}{b(s)}\right) - K\left(\frac{\theta - u}{b(s)}\right).$$

Define $I_1(s,t,u) = 1_{\{|\theta - u| \le b(s)\} \cup \{|\theta + t\gamma_n^{-1} - u| \le b(s)\}}$ and $I_2(s,u) = 1_{\{|\theta - u| \le 2b(s)\}}$. Observe that for sufficiently large n, $I_1(s,t,u) \le I_2(s,u)$, uniformly in t. For n large enough,

$$\begin{aligned}
&\left|\frac{d}{ds}H_n(s,t,u)\right|\\
&\quad = \Bigg|\left\{\frac{d}{ds}\left(\frac{\theta + t\gamma_n^{-1} - u}{b(s)}\right)\right\} I_1(s,t,u) K^{(1)}\left(\frac{\theta + t\gamma_n^{-1} - u}{b(s)}\right)\\
&\qquad - \left\{\frac{d}{ds}\left(\frac{\theta - u}{b(s)}\right)\right\} I_1(s,0,u) K^{(1)}\left(\frac{\theta - u}{b(s)}\right)\Bigg|\\
&\quad = \left|K^{(1)}\left(\frac{\theta + t\gamma_n^{-1} - u}{b(s)}\right)\right|\left|\frac{d}{ds}\left(\frac{\theta + t\gamma_n^{-1} - u}{b(s)}\right) - \frac{d}{ds}\left(\frac{\theta - u}{b(s)}\right)\right| I_2(s,u)\\
&\qquad + \left|\frac{d}{ds}\left(\frac{\theta - u}{b(s)}\right)\right|\left|K^{(1)}\left(\frac{\theta + t\gamma_n^{-1} - u}{b(s)}\right) - K^{(1)}\left(\frac{\theta - u}{b(s)}\right)\right| I_2(s,u)\\
&\quad = O(n^{-(k-1)/(2k+3)})
\end{aligned}$$

uniformly in s, t, owing to the Lipschitz continuity of $K^{(1)}$ and to

$$\left\{\frac{d}{ds}\left(\frac{\theta - u}{b(s)}\right)\right\} I_2(s, u) = O(1).$$

Therefore,

$$|\tilde{w}_j(s_1, s_2, t)| \le |s_2 - s_1| O(n^{-(k+1)/(2k+3)}). \tag{6.22}$$

As before, we find for the rth-order cumulants $\tilde{M}_r$ of $\phi_n(s_1, t) - \phi_n(s_2, t)$:

$$\tilde{M}_1 = 0$$

and

$$\begin{aligned}\tilde{M}_r &= \sum_{(j_1, \dots, j_r) \in J^r}^{n-1} \prod_{i=1}^{r} \tilde{w}_{j_i}(s_1, s_2, t) \operatorname{cum}\big(I(\lambda_{j_1}), \dots, I(\lambda_{j_r})\big)\\ &= |s_2 - s_1|^r O(n^{-(r-2)(k+1)/(2k+3)})\end{aligned}$$

for $r \ge 2$ by (6.13) and (6.21). Thus

$$E(\phi_n(s_2, t) - \phi_n(s_1, t))^4 \le |\tilde{M}_4| + 3\tilde{M}_2^2 = |s_2 - s_1|^4 O(1). \tag{6.23}$$ □

PROOF OF THEOREM 2.1. According to the corollary of Bickel and Wichura (1971), page 1664, we need to show convergence of the finite-dimensional distributions of processes ϕ_n to the corresponding distributions of the limit process ϕ and tightness of processes ϕ_n. The former is implied by Lemmas 6.1–6.3, and for the latter we employ the moment tightness condition (3) of Bickel and Wichura (1971). For this, let $B = [s_1, s_2] \times [t_1, t_2]$ and $D = |\tilde{s}_1, \tilde{s}_2| \times [\tilde{t}_1, \tilde{t}_2]$. The increment of processes ζ_n around a block B is $\phi_n(B) = \phi_n(s_2, t_2) - \phi_n(s_1, t_2) - \phi_n(s_2, t_1) + \phi_n(s_1, t_1)$. The tightness condition amounts to $E[(\phi_n(B))^2(\phi_n(D))^2] \le c\lambda(B)\lambda(D)$ for some constant $c > 0$, λ being the Lebesgue measure. According to the Cauchy–Schwarz inequality, it suffices to show $E(\zeta_n(B))^4 \le c|s_2 - s_1|^2|t_2 - t_1|^2$, which follows from Lemmas 6.4 and 6.5. □

PROOF OF LEMMA 4.1. For the bias part, observe that analogous to the result on the expectation of the process for any $\lambda \in (0, 2\pi)$, $E(\hat{f}^{(\nu)}(\lambda) - f^{(\nu)}(\lambda)) = b^{k+1-\nu} B_k f^{(k+1)}(\lambda)(1 + o(1))$, which together with the assumption on b implies the special form of the asymptotic bias for $0 \le \nu \le k$. The o-term is uniform on compact sets. For $\nu = k + 1$, the Taylor expansion is carried out up to the $(k + 1)$st derivative.

In analogy to the proof of the covariance structure of processes ζ_n we find, defining

$$\tilde{W}_j(\lambda, b) = \frac{1}{b^{\nu+1}} \int_{d_{j-1}}^{d_j} K_\nu\left(\frac{\lambda - u}{b}\right) du,$$

$$\operatorname{var}(\hat{f}^{(\nu)}(\lambda, b)) = \frac{2\pi f^2(\lambda)}{nb^{2\nu+1}} \int_{-1}^{1} K_\nu(v)^2 \, dv (1 + o(1)),$$

using Brillinger (1981), Theorem 5.2.4, and integral approximations (6.7). Let $\tilde{J} = \{j \in \{1, \dots, n-1\}: |t - d_{j-1}| \le b \text{ or } |t - d_j| \le b\}$ and consider the rth-order cumulant of $\tilde{Z}_n = (nb^{2\nu+1})^{1/2}(\hat{f}^{(\nu)}(\lambda) - f^{(\nu)}(\lambda))$ for $r \ge 2$:

$$\begin{aligned}
\operatorname{cum}_r(\tilde{Z}_n) &= (nb^{2\nu+1})^{r/2} \\
&\quad \times \sum_{(j_1,\dots,j_r)\in \tilde{J}_r}^{n-1} \operatorname{cum}\Big(\tilde{W}_{j_1}(\lambda, b) I(\lambda_{j_1}), \dots, \tilde{W}_{j_r}(\lambda, b) I(\lambda_{j_r})\Big) \\
&\le (nb^{2\nu+1})^{r/2} \Big(\max_{1\le j_l \le n-1} \big|\tilde{W}_{j_l}(\lambda, b)\big| \Big)^r \\
&\quad \times \sum_{(j_1,\dots,j_r)\in \tilde{J}^r} \operatorname{cum}\big(I(\lambda_{j_1}), \dots, I(\lambda_{j_r})\big).
\end{aligned}$$

Since $\max_{1\le j_l \le n-1} |\tilde{W}_{j_l}(\lambda, b)| = O(1/nb^{\nu+1})$, (6.13) implies

$$\operatorname{cum}_r(\tilde{Z}_n) = O\big((nb)^{1-r/2}\big) \to 0 \tag{6.24}$$

for $r \ge 3$, whence the result follows. □

PROOF OF LEMMA 4.2. Let $\rho_n \sim nb^{-(\nu+2)}$. Partition the interval $[\delta, 2\pi - \delta]$ into ρ_n intervals I_l with center z_l and width $2(\pi - \delta)/\rho_n$, $l = 1, \dots, \rho_n$. Observe

$$\begin{aligned}
&P\Big(\sup_{\lambda \in [\delta, 2\pi-\delta]} \big|\hat{f}^{(\nu)}(\lambda) - f^{(\nu)}(\lambda)\big| > \varepsilon \Big) \\
&\quad \le P\Big(\sup_l \sup_{\lambda\in I_l} \big|\hat{f}^{(\nu)}(\lambda) - \hat{f}^{(\nu)}(z_l)\big| > \frac{\varepsilon}{4} \Big) \\
&\qquad + P\Big(\sup_l \big|\hat{f}^{(\nu)}(z_l) - E\hat{f}^{(\nu)}(z_l)\big| > \frac{\varepsilon}{4} \Big) \\
&\qquad + P\Big(\sup_l \big|E\hat{f}^{(\nu)}(z_l) - f^{(\nu)}(z_l)\big| > \frac{\varepsilon}{4} \Big) \\
&\qquad + P\Big(\sup_l \sup_{\lambda\in I_l} \big|f^{(\nu)}(z_l) - f^{(\nu)}(\lambda)\big| > \frac{\varepsilon}{4} \Big).
\end{aligned}$$

Now, using the Lipschitz continuity of K_ν, $\tilde{J}$ as defined in the proof of Lemma 4.1, (6.24) and Lemma 4.1, one shows that each of the terms on the r.h.s. is indeed $o_p(1)$. □

PROOF OF LEMMA 4.3. For large enough n,

$$\begin{aligned}
\sup_t \big|\hat{f}^{(\nu)}(\theta + t\gamma_n^{-1}) - \hat{f}^{(\nu)}(\theta)\big| &= O\bigg(\frac{1}{b^{r+1} n^{k/(2k+3)}} \bigg) \sum_{j=1}^{n} \frac{1}{b} \int_{d_{j-1}}^{d_j} 1_{\{|(\theta - x)/b| \le 2\}}\, dx \\
&= o_p(1).
\end{aligned}$$

□

REFERENCES

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (B. N. Petrov and F. Csaki, eds.) 267–281. Akademiai Kiado, Budapest.

ALEKSEEV, V. G. and YAGLOM, A. M. (1980). Nonparametric and parametric spectrum estimation methods for stationary time series. In *Time Series* (O. D. Anderson, ed.) 401–422. North-Holland, New York.

BELTRÃO, K. I. and BLOOMFIELD, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *J. Time Ser. Anal.* **8** 21–28.

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

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

EDDY, W. F. (1980). Optimal kernel estimators of the mode. *Ann. Statist.* **8** 870–882.

ENSOR, K. B. and NEWTON, H. J. (1988). The effect of order estimation on estimating the peak frequency of an autoregressive spectral density. *Biometrika* **75** 587–589.

GRANOVSKY, B. and MÜLLER, H. G. (1991). Optimizing kernel methods for the nonparametric estimation of functions and characteristic points: A unifying variational principle. *Internat. Statist. Rev.* **59** 373–388.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

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

HURVICH, C. M. (1985). Data-driven choice of a spectrum estimate: Extending the applicability of cross-validation methods. *J. Amer. Statist. Assoc.* **80** 933–940.

LEONOV, V. P. and SHIRYAEV, A. N. (1959). On a method of calculation of semi-invariants. *Theory Probab. Appl.* **4** 319–329.

NEWTON, H. J. and PAGANO, M. (1983). A method for determining periods in time series. *J. Amer. Statist. Assoc.* **78** 152–157.

PARZEN, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* **28** 329–348.

PARZEN, E. (1962). On the estimation of a probability density and mode. *Ann. Math. Statist.* **33** 1065–1076.

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series*. Academic, New York.

SHUMWAY, R. H. (1988). *Applied Statistical Time Series Analysis*. Prentice-Hall, Englewood Cliffs, N.J.

WHITT, W. (1970). Weak convergence of probability measures on the function space $\mathscr{C}[0, \infty)$. *Ann. Math. Statist.* **41** 939–944.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616-8705