The Annals of Statistics
1992, Vol. 20, No. 1, 121–145

ON BOOTSTRAPPING KERNEL SPECTRAL ESTIMATES

BY J. FRANKE AND W. HÄRDLE

University of Kaiserslautern and Université Catholique de Louvain

An approach to bootstrapping kernel spectral density estimates is described which is based on resampling from the periodogram of the original data. We show that it is asymptotically valid under suitable conditions, and we illustrate its performance for a medium-sized time series sample with a small simulation study.

1. Introduction. During the last years, Efron's (1979) bootstrap has been recognized as a powerful tool for approximating certain characteristics, that is, variance or confidence limits, of statistics, which cannot at all or only with undue effort be calculated by analytical means. In time series analysis, due to the complicated data structure, this kind of difficulty quite often crops up, particularly if one is not willing to assume Gaussianity of the data. In spite of the need for an improved evaluation of the performance of spectrum or parameter estimates for stationary processes, the bootstrap has only recently been applied to problems from time series analysis. Most authors, like Freedman (1984), Efron and Tibshirani (1986), Swanepoel and van Wyk (1986) and Kreiss and Franke (1989), consider resampling the estimated innovations of parametric time series models, whereas Künsch (1989) discusses resampling blocks of data from a stationary process. In this paper, we discuss an intuitive approach to bootstrapping kernel spectral estimates based on resampling from the periodogram of the data, an idea which has been pursued independently in a quite different manner by Hartigan (1990). We prove a theorem asserting that our procedure works provided we take care of the bias in a particular manner. This result is related to similar observations of Romano (1988) for bootstrapping kernel probability density estimates. Some simulations illustrate that our procedure works for moderate sample sizes.

2. Kernel estimates for spectral densities. Let $X_1, \dots, X_T$ be a sample from a strictly stationary real-valued process $\{X_n, -\infty < n < \infty\}$ with mean 0, finite variance and spectral density $f(\omega)$. Let

$$I_T(\omega) = \frac{1}{T}\left|\sum_{k=1}^{T} X_k e^{ik\omega}\right|^2, \qquad -\pi \le \omega \le \pi,$$

denote the periodogram of the sample. Let N denote the largest integer less than or equal to $T/2$. Let the discrete frequencies ω_k be given by

Received June 1987; revised October 1990.
AMS 1980 *subject classifications*. Primary 62M15; secondary 62G05.
Key words and phrases. Bootstrap, time series, kernel spectral estimate, periodogram, bandwidth selection.

$2\pi k/T$, $-N \le k \le N$. We consider estimation of $f(\omega)$ by a kernel spectral estimate of the form

$$(1) \qquad \hat{f}(\omega; h) = \frac{1}{Th} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_k}{h}\right) I_T(\omega_k),$$

where the kernel $K(\theta)$ is a given symmetric, nonnegative function on the real line. We stress the dependency of $\hat{f}$ on the bandwidth h, as the performance of the estimate essentially depends on this smoothing parameter. As the functional measuring the local performance we consider the mean-square percentage error (MSPE), originally proposed by Parzen (1957),

$$\mathrm{MSPE}(\omega; h) = E\left\{\frac{\hat{f}(\omega; h) - f(\omega)}{f(\omega)}\right\}^2.$$

Here, we have taken into account that $f(\omega)$ is a scale parameter of the asymptotic distribution of $I_T(\omega)$. Under suitable assumptions on the process $\{X_n\}$ and on the kernel K,

$$(2) \quad \mathrm{MSPE}(\omega; h) = \left\{\frac{1}{2} h^2 \frac{f''(\omega)}{f(\omega)}\right\}^2 + \frac{1}{2\pi} \int_{-\infty}^{\infty} K^2(\theta)\, d\theta \, \frac{1}{Th} + o\left(\frac{1}{Th}\right),$$

and $T^{-1/5}$ is the rate at which h has to go to 0 if we want to minimize $\mathrm{MSPE}(\omega; h)$ asymptotically [compare Priestley (1981), Chapter 7.2]. In this paper, we direct our attention to this most common situation in kernel spectrum estimation.

Härdle and Bowman (1988) apply the bootstrap to kernel estimates for regression curves, and Romano (1988) discusses the related problem of bootstrapping kernel estimates for probability densities. We use the familiar device of interpreting the spectral estimation problem as an approximate multiplicative regression problem, starting from

$$(3) \qquad I_T(\omega_k) = f(\omega_k)\varepsilon_k, \qquad k = 1, \ldots, N.$$

The residuals are approximately independent and identically distributed for large T. There are several precise formulations of this vague statement which differ with respect to the—always finitely numbered—frequencies at which the periodogram is considered and with respect to the assumptions on the process $\{X_n\}$ [compare, e.g., Brillinger (1981), Chapters 4 and 5].

3. The bootstrap procedure. In this section, we apply the bootstrap approach of Härdle and Bowman (1988) to (3) by pretending that $\varepsilon_1, \ldots, \varepsilon_N$ are really i.i.d. As we want to resample from the residuals, we need an initial estimate of $f(\omega)$. For this purpose, we consider a kernel estimate $\hat{f}(\omega; h_i)$ of the form (1) with an arbitrary initial bandwidth h_i. In the resampling step, we use another kernel spectrum estimate $\hat{f}(\omega; g)$ of the form (1) to get the bootstrap approximation of the law of $\hat{f}(\omega; h)$. The bandwidth h, which we want to use in spectrum estimation, the resampling bandwidth g and the initial bandwidth h_i may all be different subject to some conditions which we shall discuss later. We now consider the following procedure for getting a bootstrap approximation for $\hat{f}(\omega; h)$.

STEP 1. We choose an initial global bandwidth $h_i > 0$ which does not depend on ω. We estimate the residuals ε_k, $k = 1, \dots, N$, of (3) as

$$\hat{\varepsilon}_k = \frac{I_T(\omega_k)}{\hat{f}(\omega_k; h_i)}, \qquad k = 1, \dots, N.$$

We rescale these empirical residuals and consider

$$\tilde{\varepsilon}_k = \frac{\hat{\varepsilon}_k}{\hat{\varepsilon}.}, \qquad k = 1, \dots, N, \quad \text{where } \hat{\varepsilon}. = \frac{1}{N} \sum_{j=1}^{N} \hat{\varepsilon}_j.$$

STEP 2. We draw independent bootstrap residuals $\varepsilon_1^*, \dots, \varepsilon_N^*$ from the empirical distribution of $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N$, that is, for all $j = 1, \dots, N$,

$$\mathrm{pr}\{\varepsilon_j^* = \tilde{\varepsilon}_k\} = \frac{1}{N}, \qquad k = 1, \dots, N.$$

Keeping (3) in mind, we define bootstrap periodogram values as

$$I_T^*(\omega_k) = I_T^*(-\omega_k) = \hat{f}(\omega_k; g)\varepsilon_k^*, \qquad k = 1, \dots, N,$$

with some resampling bandwidth g. For convenience, we set $I_T^*(0) = 0$, which corresponds to the periodogram value at 0 taken from a mean-corrected sample. Finally, we get a bootstrap spectral estimate as

$$\hat{f}^*(\omega; h, g) = \frac{1}{Th} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_k}{h}\right) I_T^*(\omega_k).$$

The rescaled empirical residual ε_j^* has mean 1 with respect to the empirical distribution of $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N$. This is asymptotically the correct value as the true residual ε_j is asymptotically distributed as an exponential variable with parameter 1. Like recentering in additive regression models [Freedman (1981)], rescaling avoids an additional bias at the resampling stage. Apart from this appealing property, we need this device also from a theoretical point of view. Without rescaling, a proof of the validity of the bootstrap procedure would require more detailed information about the asymptotic properties of $\varepsilon_1, \dots, \varepsilon_N$ than given by Chen and Hannan (1980), and, presumably, Theorem 1 would not even be true in general for resampling directly from $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N$.

Resampling from the periodogram is considered independently by Hartigan (1990). He appeals to the fact that the $I_T(\omega_j)$ asymptotically are independent exponential variables and derives resampling estimates for the variance of linear combinations of the periodogram ordinates by systematically perturbing the $I_T(\omega_j)$. However, his procedure has bias problems for non-Gaussian data.

Exploiting our knowledge about the asymptotic distribution of the ε_k, we can modify the preceding bootstrap procedure by replacing $\varepsilon_1^*, \dots, \varepsilon_N^*$ with independent exponential variables $\chi_1, \dots, \chi_N$ with parameter 1. As in Step 2, we get modified bootstrap periodogram values

$$I_T^+(\omega_k) = I_T^+(-\omega_k) = \hat{f}(\omega_k; g)\chi_k, \qquad k = 1, \dots, N,\ I_T^+(0) = 0,$$

and a modified bootstrap spectral estimate

$$\hat{f}^+(\omega; h, g) = \frac{1}{Th} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_k}{h}\right) I_T^+(\omega_k).$$

As we see in the next section, the bootstrap principle holds for $\hat{f}^+$ as well as for $\hat{f}^*$. Higher-order asymptotics and/or elaborate Monte Carlo studies would be needed to detect differences between both methods. Up to now, some scant simulation results support the intuition that $\hat{f}^*$ is to be preferred for not too large samples and, in particular, for non-Gaussian time series.

4. The bootstrap principle holds. The basic idea of bootstrapping, as applied to the spectral estimation context, is to infer properties of the distribution of the estimate $\hat{f}(\omega; h)$ from the conditional distribution of its bootstrap approximation $\hat{f}^*(\omega; h, g)$, given the original data. To prove the theoretical validity of this bootstrap principle, we follow Bickel and Freedman (1981) and consider the Mallows distance between the pivotal quantity $\sqrt{Th}\{\hat{f}(\omega; h) - f(\omega)\}/f(\omega)$ and its bootstrap approximation $\sqrt{Th}\{\hat{f}^*(\omega; h, g) - \hat{f}(\omega; g)\}/\hat{f}(\omega; g)$. Here, the Mallows distance between distributions F and G is defined as

$$d_2(F, G) = \inf\left\{E(X - Y)^2\right\}^{1/2},$$

where the infimum is taken over all pairs of random variables X and Y having marginal distributions F and G, respectively. We adopt the convention that where random variables appear as arguments of d_2 these represent the corresponding distributions. In particular, bootstrap quantities represent their conditional distribution given the original data $X_1, \dots, X_T$.

For our main result, we need the process generating the data to show sufficiently weak dependence between observations taken at time points far apart. To make this statement precise, we restrict our attention to linear processes, and we assume that the coefficients of the infinite moving average representation decrease sufficiently fast. Furthermore, we consider only the most common situation in kernel spectrum estimation by assuming that the spectral density $f(\omega)$ which we want to estimate is twice continuously differentiable and by choosing a kernel K for which $T^{-1/5}$ is the optimal rate of decrease for the bandwidth h if one is interested in a small mean-square percentage error. This fact is guaranteed by condition (C4) [compare, e.g., Priestley (1981), page 511].

If we want the bootstrap principle to hold in the simple form described in Section 3, we have to make the crucial assumption that the resampling bandwidth g, which we use for defining the bootstrap spectral estimate, converges to 0 a bit slower than $T^{-1/5}$. The reference estimate $\hat{f}(\omega; g)$, therefore, is a bit smoother than an optimal estimate of $f(\omega)$. However, this should not worry us as we do not use $\hat{f}(\omega; g)$ for estimating $f(\omega)$ but only for inferring information about the distribution of $\hat{f}(\omega; h)$, which itself is a kernel estimate with bandwidth h decreasing to 0 with optimal rate $T^{-1/5}$.

We make use of the following notational convention: $h \sim a_T$ if and only if there are constants c, c' such that $0 < c \le h/a_T \le c' < \infty$ for all T large enough.

THEOREM 1. *Let $\{X_n, -\infty < n < \infty\}$ be a real-valued linear process:*

$$X_n = \sum_{k=-\infty}^{\infty} b_k \xi_{n-k}, \qquad -\infty < n < \infty,$$

where ξ_n, $-\infty < n < \infty$, are independent identically distributed random variables satisfying

(C1) $E\xi_n = 0$, $E\xi_n^2 = 1$, $E|\xi_n|^5 < \infty$, *the characteristic function $q(u)$ of ξ_j satisfies* $\sup\{|q(u)|; |u| \ge \delta\} < 1$ *for all* $\delta > 0$.

Assume that the spectral density f of $\{X_n\}$ is nonvanishing and twice continuously differentiable on $[-\pi, \pi]$, and

$$\text{(C2)} \qquad \sum_{k=-\infty}^{\infty} |kb_k| < \infty.$$

Let K be a symmetric, nonnegative kernel on $(-\infty, \infty)$ satisfying

$$\text{(C3)} \qquad \frac{1}{2\pi}\int_{-\infty}^{\infty} K(\theta)\, d\theta = 1, \qquad \frac{1}{2\pi}\int_{-\infty}^{\infty} \theta^2 K(\theta)\, d\theta = 1,$$

where K has compact support $[-\kappa, \kappa]$ and K is uniformly Lipschitz with constant L_K.

Let $k(u)$ denote the Fourier transform of $K(\theta)$, and assume that it is locally quadratic around 0:

$$\text{(C4)} \qquad \lim_{u \to 0} \frac{k(0) - k(u)}{u^2} \quad \textit{exists, is finite and not } 0.$$

For $T \to \infty$, let the bandwidth h of the estimate of interest, the initial bandwidth h_i and the resampling bandwidth g satisfy

$h \sim T^{-1/5}$, $h_i \to 0$ *such that* $(Th_i^4)^{-1} = O(1)$, $g \to 0$ *such that* $h/g \to 0$.

Then, using the preceding definitions, the bootstrap principle holds:

$$\text{(i)} \quad d_2\left[\sqrt{Th}\,\frac{\hat{f}(\omega; h) - f(\omega)}{f(\omega)}; \sqrt{Th}\,\frac{\hat{f}^*(\omega; h, g) - \hat{f}(\omega; g)}{\hat{f}(\omega; g)}\right] \to 0$$

in probability,

$$\text{(ii)} \quad d_2\left[\sqrt{Th}\,\frac{\hat{f}(\omega; h) - f(\omega)}{f(\omega)}; \sqrt{Th}\,\frac{\hat{f}^+(\omega; h, g) - \hat{f}(\omega; g)}{\hat{f}(\omega; g)}\right] \to 0$$

in probability.

The proof of the theorem, which is deferred to the Appendix, crucially depends on a theorem of Chen and Hannan (1980) which states the almost sure uniform convergence of the empirical distribution function F_N of the sample $I_T(\omega_j)/f(\omega_j)$, $j = 1, \dots, N$, to the distribution function $1 - \exp(-x)$ of the exponential distribution with parameter 1. To make use of this theorem, we need finiteness of the fifth moment and the condition on the characteristic function in (C1). Theorem 1 is presumably correct, assuming $E\xi_n^4 < \infty$ only, because, for our purposes, the weaker convergence in probability of $\sup|F_N(x) - (1 - e^{-x})|$ suffices. We do not try to prove this assertion, as (C1) does not appear excessively restrictive.

To cope with the bias part of the Mallows distance in (i) and (ii), the kernel K must decrease sufficiently fast to 0. For simplicity, we even assume that its support is compact. Some of the kernels which are frequently used in applied spectral analysis satisfy this assumption, for example, the Bartlett–Priestley window [compare Priestley (1981), Chapters 6.2 and 7.5], and restricting attention to kernels with compact support gives rise to considerable simplification of already quite technical proofs.

In the literature, rescaled kernel estimates of the form

$$\tilde{f}(\omega; h) = \frac{\hat{f}(\omega; h)}{S_T(\omega)}, \qquad S_T(\omega) = \frac{1}{Th} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_j}{h}\right)$$

sometimes are considered. As we shall show in the appendix, $\frac{1}{4} \le S_T(\omega) \le 2$ for all ω, if T is large enough. Therefore, the results of this paper hold for $\tilde{f}(\omega; h)$ too if they are appropriately rephrased.

As already mentioned, Härdle and Bowman (1988) propose a similar procedure for bootstrapping kernel regression estimates. In contrast to our Theorem 1, they consider resampling regression function estimates with bandwidth $g \sim T^{-1/5}$ only. In this case, the bootstrap principle does not hold in the straightforward form of Theorem 1 as the bias of the bootstrap approximation does not approach the bias of the kernel estimate fast enough. However, it is possible to handle this difficulty by essentially bootstrapping only the variance part of the bootstrap approximation and by introducing the bias part by means of an explicit estimate of $f''(\omega)$, remembering the asymptotic relation (2). The same idea works in the spectral estimation context too, and we formulate the result as Theorem 2. We do not give the proof, as its larger part is identical and the rest is quite similar to the proof of Theorem 1. Details can be found in a technical report [Franke (1987)].

THEOREM 2. *Let $\{X_n, -\infty < n < \infty\}$ be a real-valued linear process satisfying the assumptions of Theorem 1. Let the kernel K satisfy the assumptions of Theorem 1, too. Let $\hat{f}''(\omega)$ be a weakly consistent estimate of $f''(\omega)$. Let*

$$\hat{f}_c(\omega; h, g) = E^*\hat{f}^*(\omega; h, g) = \frac{1}{Th} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_j}{h}\right) \hat{f}(\omega_j; g)$$

be the conditional expectation of $\hat{f}^*(\omega; h, g)$ and of $\hat{f}^+(\omega; h, g)$ given the *original data. Then, for* $T \to \infty$, $h \sim T^{-1/5}$, $g \sim T^{-1/5}$, $h_i \to 0$ *such that* $(Th_i^4)^{-1} = O(1)$, *we have*

(i)
$$d_2\left[\sqrt{Th}\,\frac{\hat{f}(\omega;h) - f(\omega)}{f(\omega)};\right.$$
$$\left.\sqrt{Th}\,\frac{\hat{f}^*(\omega;h,g) - \hat{f}_c(\omega;h,g) + (h^2/2)\hat{f}''(\omega)}{\hat{f}(\omega;g)}\right] \to 0$$

in probability,

(ii)
$$d_2\left[\sqrt{Th}\,\frac{\hat{f}(\omega;h) - f(\omega)}{f(\omega)};\right.$$
$$\left.\sqrt{Th}\,\frac{\hat{f}^+(\omega;h,g) - \hat{f}_c(\omega;h,g) + (h^2/2)\hat{f}''(\omega)}{\hat{f}(\omega;g)}\right] \to 0$$

in probability.

Here and in the following, E^* denotes the expectation with respect to the empirical distribution of $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N$.

As a consistent estimate for $f''(\omega)$, we can choose, for example, a kernel estimate of the simple form

$$\hat{f}''(\omega;h_2) = \frac{1}{Th_2^3}\sum_{k=-N}^{N} W\left(\frac{\omega - \omega_k}{h_2}\right) I_T(\omega_k), \tag{4}$$

where W is a kernel of order (2, 4) as defined by Gasser, Müller, Köhler, Molinari and Prader (1984).

5. Simulations. In this section, a small simulation study illustrates the performance of our bootstrap approach for a medium sample size $T = 256$. We consider data from an autoregressive process of order 5:

$$X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + \varepsilon_t,$$

where the ε_t, $-\infty < t < \infty$, are independent standard normal variables. The process parameters have been chosen such that the spectral density has a specified shape: one major peak, one minor peak, and local minima between the peaks, at 0 and at π. We consider estimating the spectral density at the discrete frequencies $\omega_k = 2\pi k/256$ for $k = 42$, 84 (approximately at the two peaks), for $k = 30$, 54 (at the left and right slope of the major peak) and for $k = 67$ (approximately at the trough between both peaks). For those ω_k, we consider the density and skewness of the law of the asymptotic pivot $\sqrt{Th}\{\hat{f}(\omega_k;h) - f(\omega_k)\}/f(\omega_k)$, or, to be precise, a kernel probability density estimate $p_{k,h}$ with Gaussian kernel and bandwidth $b = 0.4$, chosen by a cross-validatory argument, and the sample skewness $s_{k,h}$, both based on 500 simulated data sets. For the spectral estimate, we used the parabolic Bartlett–Priestley kernel [Priestley (1981), Chapters 6.2 and 7.5], scaled such that condition (C3) of Theorem 1 is satisfied. Inspection of various spectrum estimates showed that a good global bandwidth selection lies somewhere between 0.10 and 0.15.

We compare five approximations of $p_{k,h}$ and $s_{k,h}$, three of them derived from the bootstrap principle and the other two from asymptotic normality. All are based on *one* particular sample, $X_1, \ldots, X_{256}$. To get something like a representative data set, we chose that one out of nine independent samples for which the average mean-square percentage error of $\hat{f}(\omega; 0.1)$ assumed its median value. The three bootstrap approximations are provided by the conditional laws of $\sqrt{Th}\{\hat{f}^*(\omega; h, g) - \hat{f}(\omega; g)\}/\hat{f}(\omega; g)$ for bootstrap bandwidths $g = 0.2, 0.3, 0.4$ and initial bandwidths $h_i = g$ in all three cases. Based on 500 resamples, we calculated kernel probability density estimates $p^*_{k,h,g}$ with, again, Gaussian kernel and bandwidth $b = 0.4$ as bootstrap approximations of $p_{k,h}$, and sample skewnesses $s^*_{k,h,g}$ as approximations of $s_{k,h}$.

Using asymptotic normality of $\hat{f}(\omega; h)$, as in Proposition A2, and the asymptotic bias expansion, contained in (2), we know that $\sqrt{Th}\{\hat{f}(\omega_k; h) - f(\omega_k)\}/f(\omega_k)$ is also approximately normally distributed with mean $\mu_{k,h}$ and variance σ^2 given by

$$\mu_{k,h} = 0.5\sqrt{Th^5}\, f''(\omega_k)/f(\omega_k), \qquad \sigma^2 = \int K^2(\theta)\, d\theta/(2\pi).$$

To cope with the additional smoothing introduced by kernel probability density estimation, we have to compare $p_{k,h}$ with the normal density $\varphi_{k,h}$ with mean $\mu_{k,h}$, but with larger variance $\sigma^2 + b^2$. The normal approximation to the skewness $s_{k,h}$ is, of course, 0.

To get $\varphi_{k,h}$, we have to know f and f''. As a realistic competitor for the bootstrap, we therefore consider $\hat{\varphi}_{k,h}$, a plug-in normal approximation with mean

$$\hat{\mu}_{k,h} = 0.5\sqrt{Th^5}\, \hat{f}''(\omega; h_2)/\hat{f}(\omega; h_1)$$

and variance $\sigma^2 + b^2$; $\hat{f}(\omega; h_1)$ denotes again a spectral estimate, given by (1) with Bartlett–Priestley kernel K and bandwidth $h_1 = 0.15$; $\hat{f}''(\omega; h_2)$ denotes a kernel estimate of $f''(\omega)$, as in (4), where the kernel W has the same support as K and, there, equals $\{c_1 \cos^4(c_2 u)\}''$ with suitable constants c_1, c_2. The bandwidths h_1, h_2 are chosen to give a visually good correspondence between the true functions and their estimates.

Figures 1 and 2 show plots of $p_{k,h}$ and its approximations for $k = 42$ (peak) and $k = 30$ (slope) and bandwidths $h = 0.05$ and $h = 0.10$, respectively. Among all these selections of ω_k and h which we have considered, Figure 1c is typical for the majority of those situations: Visually, the bootstrap provides a better fit to the true density than its competitor, the plug-in normal approximation. In a few cases, for which Figure 2 is an example, the bootstrap approximation is not better than the plug-in normal approximation, but it never was considerably worse. A bit surprising was the observation that the bootstrap densities $p^*_{k,h,g}$ did not depend as much on the chosen bootstrap bandwidth g as we originally expected, as can be seen from Figures 1b and 2.

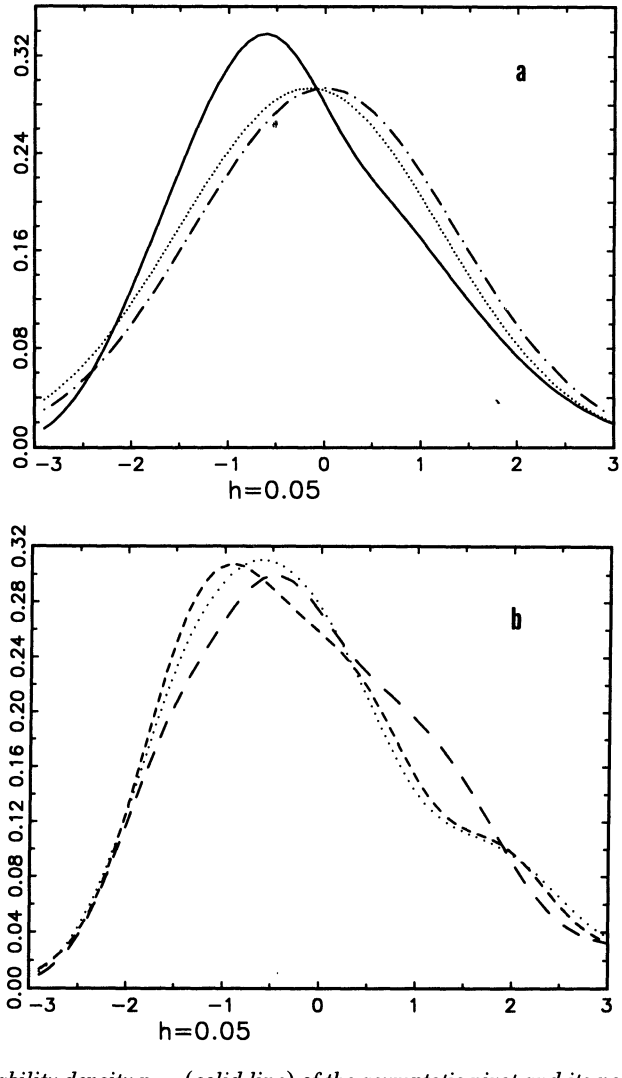

FIG. 1. *(a) Probability density $p_{k,h}$ (solid line) of the asymptotic pivot and its normal approximations $\varphi_{k,h}$ (dotted line) and $\hat{\varphi}_{k,h}$ (dots and dashes) for $k = 42$ and $h = 0.05$. (b) Bootstrap approximation $p^*_{k;h,g}$ for $k = 42$ and $h = 0.05$ and $g = 0.2$ (long dashes), $g = 0.3$ (short dashes) and $g = 0.4$ (dots).*

In some cases, only the heavily oversmoothed reference spectral estimate ($g = 0.4$) deviated considerably from the $p^*_{k,h,g}$ for smaller $g = 0.2$ and 0.3.

Table 1 compares the skewness $s_{k,h}$ of the asymptotic pivot and its bootstrap approximations $s^*_{k,h,g}$ for $g = 0.2$, 0.3 and 0.4. The bootstrap manages to reproduce the skewness of the distribution, which we want to approximate, quite well.

We also have repeated the simulation study with innovations ε_t drawn from a centered and scaled χ^2_4 distribution. Qualitatively, the results are the same

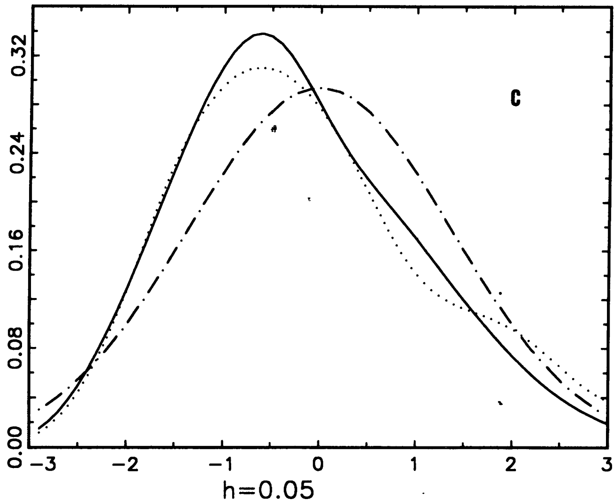

FIG. 1. *(c) Probability density* $p_{k,h}$ *(solid line), its plug-in normal approximation* $\hat{\varphi}_{k,h}$ *(dots and dashes) and a bootstrap approximation* $p^*_{k,h,0.4}$ *(dots) for* $k = 42$ *and* $h = 0.05$.

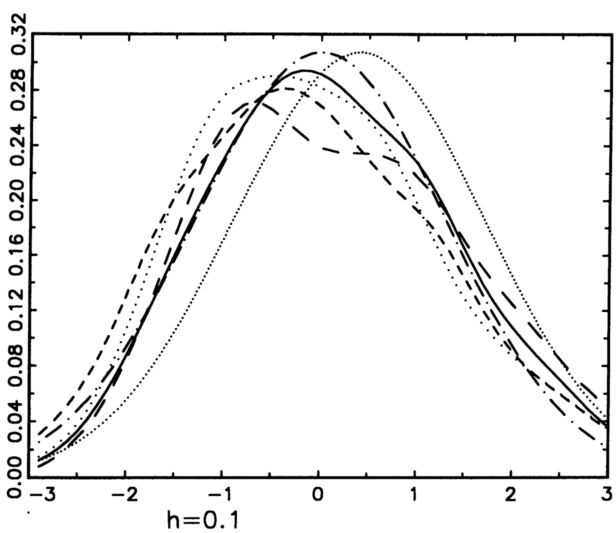

FIG. 2. *Probability density* $p_{k,h}$ *(solid line) of the asymptotic pivot, its normal approximations* $\varphi_{k,h}$ *(narrowly spaced dots) and* $\hat{\varphi}_{k,h}$ *(dots and dashes) and its bootstrap approximation* $p^*_{k,h,g}$ *for* $k = 30$ *and* $h = 0.10$ *and for* $g = 0.2$ *(long dashes),* $g = 0.3$ *(short dashes) and* $g = 0.4$ *(widely spaced dots).*

TABLE 1
Skewness of asymptotic pivot for various bandwidths h and its bootstrap approximations for several values of g taken from one representative sample

		k				
		30	42	54	67	84
$h = 0.05$	$s_{k,h}$	0.800	0.606	0.866	1.055	0.674
	$s^*_{k,h,0.2}$	0.852	0.795	0.963	0.610	0.352
	$s^*_{k,h,0.3}$	0.930	0.759	0.726	0.721	0.702
	$s^*_{k,h,0.4}$	0.719	0.895	0.665	0.869	0.978
$h = 0.10$	$s_{k,h}$	0.426	0.613	0.560	0.542	0.287
	$s^*_{k,h,0.2}$	0.557	0.642	0.630	0.511	0.491
	$s^*_{k,h,0.3}$	0.417	0.546	0.532	0.583	0.372
	$s^*_{k,h,0.4}$	0.330	0.701	0.625	0.494	0.428
$h = 0.15$	$s_{k,h}$	0.414	0.489	0.467	0.554	0.330
	$s^*_{k,h,0.2}$	0.426	0.507	0.309	0.351	0.337
	$s^*_{k,h,0.3}$	0.454	0.566	0.323	0.539	0.499
	$s^*_{k,h,0.4}$	0.495	0.425	0.457	0.399	0.470

as in the Gaussian case, that is, the bootstrap outperforms the plug-in normal approximation in approximating the probability density and the skewness of the law of interest.

6. Confidence intervals and bandwidth selection. Once we know that the bootstrap principle holds for spectral density estimation we can apply it in the usual manner to get estimates for statistical quantities of interest. For the sake of illustration, we have a look at the problem of getting a confidence interval for $f(\omega)$ and of selecting a local bandwidth $h = h(\omega)$ of the kernel estimate $\hat{f}(\omega; h)$ at a given frequency ω. In this entire section, we implicitly assume that the conditions of Theorem 1 are satisfied.

Let c_α be characterized by

$$\mathrm{pr}\left[\sqrt{Th}\,\frac{\hat{f}(\omega; h) - f(\omega)}{f(\omega)} \le c_\alpha\right] = \alpha,$$

that is, $\{1 + c_\alpha(Th)^{-1/2}\}\hat{f}(\omega; h)$ is the upper bound of a $(1 - 2\alpha)$-confidence interval for $f(\omega)$. A bootstrap approximation for the generally unknown quantity c_α is given as c^*_α, defined by

$$\mathrm{pr}^*\left[\sqrt{Th}\,\frac{\hat{f}^*(\omega; h, g) - \hat{f}(\omega; g)}{\hat{f}(\omega; g)} \le c^*_\alpha\right] = \alpha,$$

where the bootstrap distribution pr^* corresponds to drawing the bootstrap residuals $\varepsilon^*_1, \dots, \varepsilon^*_N$ from the empirical distribution of the rescaled residuals as described in Section 3. From Theorem 1, we know that $c^*_\alpha \to c_\alpha$ in probability if $T \to \infty$. Explicit calculation of c^*_α will be quite difficult, and, therefore, we propose to estimate it by the familiar Monte Carlo algorithm, as described, for example, by Efron and Tibshirani (1986), which usually is associated with bootstrap procedures. Analogously, we can get a bootstrap approximation for the lower bound of a confidence interval for $f(\omega)$.

A major problem with kernel spectral estimates is the choice of bandwidth h. Until quite recently, the literature contains only rough guidelines for choosing h which often depend on some prior information on the shape of f. An extensive discussion of this problem has been given by Priestley [(1981), Chapter 7]. In a recent paper, Beltrão and Bloomfield (1987) have investigated the problem of selecting a global bandwidth which minimizes the average mean-square percentage error,

$$\mathrm{AMSPE}(h) = \frac{1}{N} \sum_{j=1}^{N} \mathrm{MSPE}(\omega_j; h),$$

where $\mathrm{MSPE}(\omega; h)$ is defined as in Section 2. They have proposed a cross-validatory choice of bandwidth, and they have shown that their procedure produces a bandwidth which approximately minimizes $\mathrm{AMSPE}(h)$.

We consider the problem of selecting a good local bandwidth $h = h(\omega)$ which minimizes approximately the mean-square percentage error $\mathrm{MSPE}(\omega; h)$ for a given frequency ω. Following Rice (1984) who considered bandwidth choice for the related nonparametric regression estimates, we restrict the minimization to an interval $B_T = [aT^{-1/5}, bT^{-1/5}]$ of bandwidths which shrinks to 0 at the optimal rate. Here, $0 < a < b < \infty$ are suitable constants. Let h_0, depending on the sample size, be defined by

$$\mathrm{MSPE}(\omega; h_0) = \min_{h \in B_T} \mathrm{MSPE}(\omega; h).$$

As we shall discuss in proving Theorem 3,

$$(5) \qquad T^{1/5} h_0 \to z_\infty = \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} K^2(\lambda)\, d\lambda \left\{ \frac{f(\omega)}{f''(\omega)} \right\}^2 \right]^{1/5}, \quad \text{for } T \to \infty,$$

provided $f''(\omega) \neq 0$ and a, b are chosen such that $a < z_\infty < b$. Notice that $z_\infty T^{-1/5}$ minimizes the dominating part of the asymptotic formula (2) for $\mathrm{MSPE}(\omega; h)$ considered as a function of h. As $\mathrm{MSPE}(\omega; h)$ depends on the unknown spectral density f, we cannot calculate h_0. Therefore, we propose to estimate $\mathrm{MSPE}(\omega; h)$ by its bootstrap approximation,

$$\mathrm{MSPE}^*(\omega; h) = E^* \left\{ \frac{\hat{f}^*(\omega; h, g) - \hat{f}(\omega; g)}{\hat{f}(\omega; g)} \right\}^2,$$

and then to choose the bandwidth h_0^* which minimizes $\mathrm{MSPE}^*(\omega; h)$,

$$\mathrm{MSPE}^*(\omega; h_0^*) = \min_{h \in B_T} \mathrm{MSPE}^*(\omega; h).$$

The calculation of h_0^* can be accomplished easily, as $\mathrm{MSPE}^*(\omega; h)$ can be given explicitly. We do not have to resort to Monte Carlo methods in this case. A straightforward calculation, using the independence of the bootstrap residuals

ε_j^* and $E^*\varepsilon_j^* = 1$, shows

$$\hat{f}^2(\omega; g)\text{MSPE}^*(\omega; h) = \frac{\text{var}^*(\varepsilon_1^*)}{T^2h^2} \sum_{j=-N}^{N} K^2\left(\frac{\omega - \omega_j}{h}\right)\hat{f}^2(\omega_j; g) \tag{6}$$
$$+ \left\{\frac{1}{Th} \sum_{j=-N}^{N} K\left(\frac{\omega - \omega_j}{h}\right)\hat{f}(\omega_j; g) - \hat{f}(\omega; g)\right\}^2,$$

where, by Proposition A1 of the Appendix,

$$\text{var}^*(\varepsilon_1^*) = E^*(\varepsilon_1^* - 1)^2 = \frac{1}{N}\sum_{k=1}^{N} \tilde{\varepsilon}_k^2 - 1 \to 1 \quad \text{in probability.}$$

Restricting minimization to a finite subset of B_T which is allowed to increase with the sample size at a certain rate, Härdle and Bowman (1988) have shown that the analogous bootstrap selection of the bandwidth of a kernel regression estimate is asymptotically optimal in the sense that the ratio of the minimum of the bootstrap error estimate and the minimum of the true error converges to 1 in probability. Using the explicit formula (6), we are able to prove the same result without restrictions to B_T and, furthermore, to prove consistency of h_0^* in the sense that $T^{1/5}(h_0^* - h_0) \to 0$ in probability.

THEOREM 3. *If the conditions of Theorem* 1 *are satisfied and if, additionally,* $f''(\omega) \neq 0$ *and* $0 < a < z_\infty < b < \infty$, *then, for* h_0, h_0^* *defined as before,*

(i) $T^{1/5}(h_0^* - h_0) \to 0$ *in probability for* $T \to \infty$,

(ii) $\dfrac{\text{MSPE}^*(\omega; h_0^*)}{\text{MSPE}(\omega; h_0)} \to 1$ *in probability for* $T \to \infty$.

The proof of the theorem is again postponed to the Appendix.

7. Concluding remarks. We have shown that a rather straightforward approach to bootstrapping kernel spectrum estimates works. Our procedure is quite similar to the bootstrap for both parametric and nonparametric regression with fixed design. Some care has to be taken if the bootstrap principle is to hold. Either one has to restrict the bootstrap essentially to $\hat{f}(\omega; h) - E\hat{f}(\omega; h)$, estimating the bias $E\hat{f}(\omega; h) - f(\omega)$ explicitly as in Theorem 2, or one has to choose a preliminary estimate $\hat{f}(\omega; g)$ which is asymptotically smoother than an optimal kernel spectrum estimate. If h/g does not converge to 0, then the assertion of Theorem 1 does not hold. As can be seen from a careful look at the proof, the critical quantity is

$$\sqrt{Th}\left(E^*\hat{f}^*(\omega; h, g) - E\hat{f}^*(\omega; h, g)\right), \tag{7}$$

which dominates the left-hand side of (A7) of the Appendix, and which converges to 0 in probability if $h/g \to 0$. Now, remark that

$$E^*\hat{f}^*(\omega; h, g) = \frac{1}{Tg} \sum_{j=-N}^{N} \left\{ \frac{1}{Th} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_j}{h}\right) K\left(\frac{\omega_j - \omega_k}{g}\right) \right\} I_T(\omega_j).$$

By the compactness of the support of K and by the asymptotic properties of the periodogram, exhibited in (A6), $E^*\hat{f}^*(\omega; h, g)$ behaves asymptotically like the mean of Tg independent random variables with uniformly bounded variance. Therefore, (7) converges to 0 only if the scaling factor $(Th)^{1/2}$ converges to ∞ slower than $(Tg)^{1/2}$. The necessity to oversmooth in resampling kernel type function estimates is not a particular feature of spectrum estimation. Similar results have been found by Romano (1988) for probability density estimates and by Härdle and Bowman (1988) for regression function estimates.

Finally, let us remark that our results do not strongly depend on the particular assumptions on the stationary process. Essentially, we need asymptotic normality of $\hat{f}(\omega; h)$, as stated in Proposition A2 of the Appendix, and the empirical distribution function of the $I_T(\omega_j)/f(\omega_j)$, $j = 1, \dots, N$, must converge uniformly to $1 - e^{-x}$ in probability.

APPENDIX

Some auxiliary results and proofs of the theorems. For real numbers a_T and random variables Z_T, we write $Z_T = o_p(a_T)$ for $T \to \infty$ [$Z_T = O_p(a_T)$ for $T \to \infty$] if $Z_T/a_T \to 0$ in probability [$Z_T/b_T \to 0$ in probability for all sequences b_T such that $a_T = o(b_T)$]. For analyzing the bias of the kernel spectrum estimate, we repeatedly consider

$$S_T(\omega) = \frac{1}{Th} \sum_{j=-N}^{N} K\left(\frac{\omega - \omega_j}{h}\right).$$

If the kernel K satisfies the assumptions of Theorem A1, we have

$$\text{(A1)} \qquad \begin{aligned} |S_T(\omega) - 1| &= \left| S_T(\omega) - \frac{1}{2\pi} \int_{\omega-\pi}^{\omega+\pi} \frac{1}{h} K\left(\frac{\theta}{h}\right) d\theta \right| \\ &\le \frac{2\pi L_K}{Th}, \quad \text{if } |\omega| \le \pi - \kappa h, \end{aligned}$$

where $[-\kappa, \kappa]$ contains the support of K, and, for any bounded function ψ,

$$\text{(A2)} \qquad \left| \frac{1}{Th} \sum_{j=-N}^{N} K^m\left(\frac{\omega - \omega_j}{h}\right) \psi(\omega_j) \right| \le c_m^* \sup_\theta |\psi(\theta)|, \qquad m \ge 1,$$

with a suitable constant c_m^*, because only about $2\kappa Th$ summands do not vanish.

THEOREM A1. *Let* $\{X_n, -\infty < n < \infty\}$ *be a linear process,*

$$X_n = \sum_{k=-\infty}^{\infty} b_k \xi_{n-k}, \qquad -\infty < n < \infty,$$

satisfying assumptions (C1) *and* (C2) *of Theorem* 1. *Let the spectral density* f *of* $\{X_n\}$ *be nonvanishing and satisfying a uniform Lipschitz condition. Let* K *be a symmetric, nonnegative kernel satisfying assumption* (C3) *of Theorem* 1. *Let* $I_T(\omega)$ *and* $\hat{f}(\omega; h)$ *denote the periodogram and the kernel spectrum estimate based on* $X_1, \dots, X_T$, *as in Section* 2.

If, for $T \to \infty$, *we have* $h \to 0$, $(Th^4)^{-1} = O(1)$, *then*

$$\sup_{-\pi \le \omega \le \pi} \left| \hat{f}(\omega; h) - f(\omega) S_T(\omega) \right| = O_p(h^{-1}T^{-1/3}) + O_p(h).$$

PROOF. The theorem is related to Theorem 2.1 of Woodroofe and van Ness (1967) who, under assumptions on K which are too restrictive for our purposes, give an exact rate for the convergence in probability of $\sup|\hat{f}(\omega; h) - E\hat{f}(\omega; h)|/f(\omega)$. Referring to the similarity of arguments, we only sketch the proof of our theorem. Let J_T, $\hat{\varphi}$ denote the periodogram and spectral estimate of $\xi_1, \dots, \xi_T$:

$$J_T(\omega) = \frac{1}{T}\left|\sum_{k=1}^{T} \xi_k e^{ik\omega}\right|^2, \qquad \hat{\varphi}(\omega; h) = \frac{1}{Th}\sum_{j=-N}^{N} K\left(\frac{\omega - \omega_j}{h}\right) J_T(\omega_j).$$

Because f is bounded, it suffices to show that the assertion of the theorem holds for the independent ξ_j and that the supremum of $|\hat{f}(\omega; h) - \hat{\varphi}(\omega; h) f(\omega)|$ is of the order $O_p(h)$, for

$$\begin{aligned}&\left| \hat{f}(\omega; h) - f(\omega) S_T(\omega) \right| \\ &\quad \le \left| \hat{f}(\omega; h) - \hat{\varphi}(\omega; h) f(\omega) \right| + \left| \hat{\varphi}(\omega; h) - S_T(\omega) \right| f(\omega).\end{aligned}$$

(i) We split $\sup|\hat{\varphi}(\omega; h) - S_T(\omega)|$ into two parts and show that both of them converge in probability to 0 with the desired speed. Let $a_T = hT^{-1/3}$, $m_T = [a_T^{-1}]$, and $\theta_k = \pi k / m_T$ for $-m_T \le k \le m_T$:

$$\begin{aligned}&\sup_{\omega} \left| \hat{\varphi}(\omega; h) - S_T(\omega) \right| \\ &\quad = \sup_{|k| \le m_T} \sup_{|\omega - \theta_k| \le \pi a_T} \left| \hat{\varphi}(\omega; h) - S_T(\omega) \right| \\ &\quad \le \sup_{|k| \le m_T} \left| \hat{\varphi}(\theta_k; h) - S_T(\theta_k) \right| \\ &\qquad + \sup_{|\omega - \theta| \le \pi a_T} \left| \hat{\varphi}(\theta; h) - S_T(\theta) - \hat{\varphi}(\omega; h) + S_T(\omega) \right|.\end{aligned}$$

Both terms on the right-hand side are of order $O_p(h^{-1}T^{-1/3})$. For the second term, using Lipschitz continuity of K, we have

$$hT^{1/3} \sup_{|\omega-\delta|\le \pi a_T} |\hat{\varphi}(\theta;h) - S_T(\theta) - \hat{\varphi}(\omega;h) + S_T(\omega)|$$

$$\le \pi L_K \left\{ \frac{1}{Th} \sum_{j=-N}^{N} J_T(\omega_j) + 1 \right\} \to 2\pi L_K \quad \text{a.s.},$$

for $T \to \infty$, by results of Chen and Hannan (1980) on the empirical distribution of the $J_T(\omega_j)$, $j = 1, \dots, N$. For the first term, we have by Chebyshev's inequality, for all $\delta > 0$,

$$\mathrm{pr}\left\{ hT^{1/3} \sup_{|k|\le m_T} |\hat{\varphi}(\theta_k;h) - S_T(\theta_k)| \ge \delta \right\}$$

$$\le \sum_{|k|\le m_T} \frac{h^2T^{2/3}}{\delta^2} E\{\hat{\varphi}(\theta_k;h) - S_T(\theta_k)\}^2$$

$$\le (2m_T + 1)\frac{hT^{1/3}}{\delta^2},$$

and the right-hand side is bounded for $T \to \infty$. We have used

$$E\{\hat{\varphi}(\omega;h) - S_T(\omega)\}^2$$

$$\le \frac{c}{Th} \quad \text{for all } \omega \in [-\pi, \pi] \text{ and suitable constant } c > 0,$$

which follows from independence of the ξ_j and then from using (A2).

(ii) From (A2) and part (i), we know that $\hat{\varphi}(\omega;h)$ is $O_p(h)$ uniformly in ω. Using this result and Lipschitz continuity of f, we can show that

$$\hat{f}(\omega;h) - \hat{\varphi}(\omega;h) f(\omega) = \frac{1}{Th} \sum_{j=-N}^{N} K\left(\frac{\omega - \omega_j}{h}\right)\{I_T(\omega_j) - J_T(\omega_j) f(\omega)\}$$

is $O_p(h)$ uniformly in ω. For this purpose, we use the approximation of the discrete Fourier transform of the X_k by the discrete Fourier transform of the ξ_k as given by Hannan [(1970), page 246]. □

Theorem A1 and (A1) immediately imply Corollary A1, from which, together with (A2) and the compactness of the support of K, Corollary A2 follows.

COROLLARY A1. *Under the assumptions of Theorem* A1

$$\sup_{|\omega|\le \pi - \kappa h} |\hat{f}(\omega;h) - f(\omega)| = O_p(h^{-1}T^{-1/3}) + O_p(h).$$

COROLLARY A2. *Let the assumptions of Theorem* A1 *be satisfied. If for* $T \to \infty$, $g \to 0$ *and* $(Tg^4)^{-1} = O(1)$, *we have, for* $\omega \neq \pm\pi$,

$$\frac{1}{Th} \sum_{j=-N}^{N} K^2\left(\frac{\omega - \omega_j}{h}\right)\left\{\hat{f}(\omega_j, g) - f(\omega_j)\right\}^2 = O_p(g^{-2}T^{-2/3}) + O_p(g^2).$$

Relation (5) of Chen and Hannan (1980) on the empirical distribution function of the $I_T(\omega_j)/f(\omega_j)$, $j = 1, \dots, N$, implies for $N = [T/2] \to \infty$

$$\frac{1}{N} \sum_{j=1}^{N} \frac{I_T(\omega_j)}{f(\omega_j)} \to 1, \qquad \frac{1}{N} \sum_{j=1}^{N} \left[\frac{I_T(\omega_j)}{f(\omega_j)}\right]^2 \to 2 \quad \text{a.s.}$$

under the assumptions of Theorem 1. (A1), (A2) and Theorem A1 allow us to replace $f(\omega)$ by its estimate $\hat{f}(\omega; h)$ if we settle for convergence in probability.

PROPOSITION A1. *Under the assumptions of Theorem* 1, *we have, for* $N = [T/2] \to \infty$ *and* $h \to 0$ *such that* $(Th^4)^{-1} = O(1)$,

$$\frac{1}{N} \sum_{j=1}^{N} \frac{I_T(\omega_j)}{\hat{f}(\omega_j; h)} \to 1, \qquad \frac{1}{N} \sum_{j=1}^{N} \left\{\frac{I_T(\omega_j)}{\hat{f}(\omega_j; h)}\right\}^2 \to 2,$$

$$\frac{1}{N} \sum_{j=1}^{N} \left\{\frac{I_T(\omega_j)}{f(\omega_j)} - \frac{I_T(\omega_j)}{\hat{f}(\omega_j; h)}\right\}^2 \to 0 \quad \textit{in probability}.$$

PROPOSITION A2. *Let* $\{X_n, -\infty < n < \infty\}$ *be a linear process,*

$$X_n = \sum_{k=-\infty}^{\infty} b_k \xi_{n-k}, \qquad -\infty < n < \infty,$$

satisfying the assumptions of Theorem 1, *and let* $\hat{f}(\omega; h)$ *denote a kernel spectral estimate with a nonnegative symmetric kernel K satisfying assumption* (C3) *of Theorem* 1. *If, for* $T \to \infty$, *we have* $h \to 0$ *and* $Th^2 \to \infty$, *then, for* $|\omega| < \pi$:

(i) $Th \operatorname{var}(\hat{f}(\omega; h)) \to \sigma^2 = f^2(\omega) \frac{1}{2\pi} \int_{-\infty}^{\infty} K^2(\theta)\, d\theta$, *for* $T \to \infty$,

(ii) $\sqrt{Th}\left\{\hat{f}(\omega; h) - E\hat{f}(\omega; h)\right\} \to Z$ *in distribution*,

where Z is a Gaussian random variable with mean 0 *and variance* σ^2.

PROOF. Using the compactness of the support of K and the asymptotic properties of the periodogram $I_T(\omega_j)$, $j = 1, \dots, N = [T/2]$, as given in Theorem 6.2.3 of Priestley (1981), we have, for a suitable constant C,

$$Th \operatorname{var}\{\hat{f}(\omega; h)\} \le \frac{1}{Th} \sum_{j=-N}^{N} K^2\left(\frac{\omega - \omega_j}{h}\right) f^2(\omega_j) + CS_T^2(\omega)h \quad \text{for } \omega \ge \kappa h.$$

As, by (A2), $S_T(\omega)$ is bounded, (i) follows.

Part (ii) can be shown by the same methods used in the proof of Theorem V.11 of Hannan (1970), which states a stronger result for a slightly different, but asymptotically equivalent spectral estimate. □

LEMMA A1. (i) *Let K satisfy the assumptions of Theorem* A1, *and, for* $|\omega| < \pi - \kappa h$, *let p be twice continuously differentiable on* $[\omega - \kappa h, \omega + \kappa h]$. *Then, for* $T \to \infty$, $h \to 0$ *such that* $Th \to \infty$,

$$\left| \frac{1}{Th} \sum_{j=-N}^{N} K\left(\frac{\omega - \omega_j}{h}\right) p(\omega_j) - p(\omega) - \frac{h^2}{2} p''(\omega) \right|$$
$$\leq \frac{c}{Th} \left\{ \sup_\theta |p(\theta)| + h \sup_\theta |p'(\theta)| \right\}$$
$$+ \frac{h^2}{2} \sup_\theta |p''(\theta) - p''(\omega)|,$$

where c is a suitable constant and the suprema are taken over the interval $[\omega - \kappa h, \omega + \kappa h]$.

(ii) *Let the assumptions of Theorem* 1 *be satisfied. Then, for* $T \to \infty$, $h \to 0$ *such that* $(Th^4)^{-1} = O(1)$, *the bias of* $\hat{f}(\omega; h)$ *satisfies*

$$E\hat{f}(\omega; h) - f(\omega) = \frac{h^2}{2} f''(\omega) + o(h^2) + O\left(\frac{\log T}{T}\right)$$

uniformly in $|\omega| \leq \pi - \kappa h$.

PROOF. (i) The compactness of the support of K, its Lipschitz continuity and the differentiability of p imply, uniformly in $|\omega| < \pi - \kappa h$,

$$\left| \frac{1}{Th} \sum_{j=-N}^{N} K\left(\frac{\omega - \omega_j}{h}\right) p(\omega_j) - \frac{1}{2\pi} \int_{-\infty}^{\infty} K(\theta) p(\omega + \theta h)\, d\theta \right|$$
$$\leq \frac{c}{Th} \left\{ \sup_\theta |p(\theta)| + h \sup_\theta |p'(\theta)| \right\}.$$

The assertion follows from the Taylor expansion of $p(\omega + \theta h)$, using that $K(\theta)/(2\pi)$ and $\theta^2 K(\theta)/(2\pi)$ integrate to 1 and $\theta K(\theta)$ integrates to 0.

(ii) Replacing p by f in (i) and noting that, under our assumptions,

$$EI_T(\omega_j) = f(\omega_j) + O\left[\frac{\log T}{T}\right], \qquad j = 1, \ldots, N, \tag{A3}$$

uniformly in j [Priestley (1981), page 418], the second assertion follows. □

PROOF OF THEOREM 1. (a) To prove (i), we use Lemma 8.8 of Bickel and Freedman (1981) and split the squared Mallows metric into a variance part and a squared bias part,

$$V_T^2 = d_2^2\left[\sqrt{Th}\,\frac{\hat{f}(\omega;h) - E\hat{f}(\omega;h)}{f(\omega)}, \sqrt{Th}\,\frac{\hat{f}^*(\omega;h,g) - E^*\hat{f}^*(\omega;h,g)}{\hat{f}(\omega;g)}\right]$$

$$B_T^2 = Th|b_T(\omega) - b_T^*(\omega)|^2,$$

where

$$b_T(\omega) = \frac{\{E\hat{f}(\omega;h) - f(\omega)\}}{f(\omega)} \quad \text{and} \quad b_T^*(\omega) = \frac{\{E^*\hat{f}^*(\omega;h,g) - \hat{f}(\omega;g)\}}{\hat{f}(\omega;g)}.$$

Throughout the proof, we use the abbreviations $I_{T,j} = I_T(\omega_j)$, $I_{T,j}^* = I_T^*(\omega_j)$ and

$$\alpha_j(\omega;h) = \frac{1}{Th}K\left(\frac{\omega - \omega_j}{h}\right),$$

$$\gamma_j(\omega;h,g) = \sum_{j=-N}^{N} \alpha_k(\omega,h)\alpha_j(\omega_k;g) - \alpha_j(\omega;g).$$

(b) We first prove that $V_T \to 0$ in probability. For this purpose, let χ_j, $|j| \geq 1$, be independent, exponentially distributed variables with parameter 1, and let $\chi_0 \equiv 0$. We remark that $I_T(\omega)/f(\omega)$ converges to χ_1 in distribution. We define

$$f^0(\omega;h) = \sum_{j=-N}^{N} \alpha_j(\omega;h) f(\omega_j)\chi_j, \qquad D^0 = \sqrt{Th}\{f^0(\omega;h) - Ef^0(\omega;h)\},$$

$$D = \sqrt{Th}\{\hat{f}(\omega;h) - E\hat{f}(\omega;h)\},$$

$$D^* = \sqrt{Th}\{f^*(\omega;h,g) - E^*\hat{f}^*(\omega;h,g)\}.$$

We use that d_2 is a metric, and we get

$$V_T \leq \frac{d_2(D,D^0)}{f(\omega)} + d_2\left(\frac{D^0}{f(\omega)}, \frac{D^0}{\hat{f}(\omega;g)}\right) + \frac{d_2(D^0,D^*)}{\hat{f}(\omega;g)}.$$

To prove $d_2(D,D^0) \to 0$ in probability, consider a zero-mean Gaussian variable Z with variance σ^2 given in Proposition A2. By this proposition, D converges to Z in distribution, and $ED^2 \to EZ^2$. Exactly as in proving the first part of Proposition A2, $E(D^0)^2 \to EZ^2$ follows. Using boundedness of f and the regularity conditions on K, it is easy to show that D^0 satisfies Liapounov's condition [Shiryayev (1984), page 331] and, therefore, converges to Z in distribution, too. Now

$$d_2(D,D^0) \leq d_2(D,Z) + d_2(Z,D^0) \to 0,$$

where the convergence holds by Lemma 8.3 of Bickel and Freedman (1981).

Theorem 8.1 of Major (1978) provides an explicit formula for the Mallows metric of real-valued random variables which implies

$$d_2^2\left[\frac{D^0}{f(\omega)}, \frac{D^0}{\hat{f}(\omega;g)}\right] = \left[\frac{1}{f(\omega)} - \frac{1}{\hat{f}(\omega;g)}\right]^2 E(D^0)^2 \to 0 \quad \text{in probability},$$

as, for example, by Theorem A1 and (A1), $\hat{f}(\omega; g) \to f(\omega) > 0$ in probability, and as, by (A2) and the boundedness of f, $E(D^0)^2$ is bounded.

(c) To finish the proof that $V_T \to 0$ in probability, it suffices to show that $d_2(D^0, D^*) \to 0$ in probability, as $\hat{f}(\omega; g) \to f(\omega) > 0$. As D^0, D^* are sums of independent random variables (conditional on the original data), we have, by a slight modification of Lemma 8.7 of Bickel and Freedman (1981),

$$\text{(A4)} \qquad d_2^2(D^0, D^*) \le Th \sum_{j=-N}^{N} \alpha_j^2(\omega; h) d_2^2\left[f(\omega_j)\{\chi_j - 1\}, I_{T,j}^* - E^* I_{T,j}^*\right].$$

As the distributions of χ_j, ε_j^* do not depend on j, we have, using the definition of $I_{T,j}^*$,

$$\begin{aligned} &d_2^2\left[f(\omega_j)\{\chi_j - 1\}, I_{T,j}^* - E^* I_{T,j}^*\right] \\ &\quad \le 2 d_2^2\left[f(\omega_j)\{\chi_j - 1\}, \hat{f}(\omega_j;g)\{\chi_j - 1\}\right] + 2\hat{f}^2(\omega_j;g) d_2^2(\chi_j - 1, \varepsilon_j^* - 1) \\ &\quad = 2\left| f(\omega_j) - \hat{f}(\omega_j;g)\right|^2 E(\chi_1 - 1)^2 + 2\hat{f}^2(\omega_j;g) d_2^2(\chi_1, \varepsilon_1^*). \end{aligned}$$

Therefore, using Corollary A2 and (A2), we conclude that the right-hand side of (A4) converges to 0 in probability if $d_2(\chi_1, \varepsilon_1^*) \to 0$ in probability. To prove the latter convergence, we use

$$d_2(\chi_1, \varepsilon_1^*) \le d_2(\chi_1, \varepsilon_1^0) + d_2(\varepsilon_1^0, \hat{\varepsilon}_1^*) + d_2(\hat{\varepsilon}_1^*, \varepsilon_1^*),$$

where the distributions of ε_1^0 and $\hat{\varepsilon}_1^*$ are the empirical distributions of the true residuals $\varepsilon_1, \dots, \varepsilon_N$ and of the unscaled empirical residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N$, respectively.

Theorem 1 and relation (5) of Chen and Hannan (1980) imply that the distribution function of ε_1^0 converges to the distribution function of χ_1 uniformly a.s. for $N \to \infty$ and that

$$E^0(\varepsilon_1^0)^2 = \frac{1}{N}\sum_{j=1}^{N} \varepsilon_j^2 = \frac{1}{N}\sum_{j=1}^{N}\left[\frac{I_T(\omega_j)}{f(\omega_j)}\right]^2 \to E\chi_1^2 \quad \text{a.s.}$$

Therefore, $d_2(\chi_1, \varepsilon_1^0) \to 0$ a.s. by Lemma 8.3 of Bickel and Freedman (1981).

To get an upper bound for $d_2(\varepsilon_1^0, \hat{\varepsilon}_1^*)$, we choose the joint distribution of $(\varepsilon_1^0, \hat{\varepsilon}_1^*)$ such that it assumes the value $(\varepsilon_j, \hat{\varepsilon}_j)$ with probability $1/N$, $j = 1, \dots, N$. Then, by Proposition A1,

$$d_2^2(\varepsilon_1^0, \hat{\varepsilon}_1^*) \le \frac{1}{N} \sum_{k=1}^{N} (\varepsilon_k - \hat{\varepsilon}_k)^2$$

$$= \frac{1}{N} \sum_{k=1}^{N} \left[\frac{1}{f(\omega_k)} - \frac{1}{\hat{f}(\omega_k; h_i)} \right]^2 I_{T,k}^2 \to 0 \quad \text{in probability.}$$

By exactly the same argument, we also get

$$d_2^2(\hat{\varepsilon}_1^*, \varepsilon_1^*) \le \frac{1}{N} \sum_{k=1}^{N} (\hat{\varepsilon}_k - \tilde{\varepsilon}_k)^2$$

$$= \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\varepsilon}_k^2 \right] \left[1 - \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\varepsilon}_k \right]^{-1} \right]^2 \to 0 \quad \text{in probability,}$$

by Proposition A1, using $\hat{\varepsilon}_k = I_T(\omega_k)/\hat{f}(\omega_k; h_i)$.

(d) We now start to discuss the bias part B_T. First, we remark that we may neglect the denominators of $b_T(\omega)$ and $b_T^*(\omega)$, as $\hat{f}(\omega; g) \to f(\omega) > 0$ in probability, and as, by Lemma A1 and Proposition A2,

$$\sqrt{Th} \left\{ 1 - \frac{f(\omega)}{\hat{f}(\omega; g)} \right\} b_T(\omega)$$

$$\text{(A5)} \qquad = \sqrt{Th} \frac{\hat{f}(\omega; g) - E\hat{f}(\omega; g) + E\hat{f}(\omega; g) - f(\omega)}{f(\omega)\hat{f}(\omega; g)} \left[E\hat{f}(\omega; h) - f(\omega) \right]$$

$$= O_p(g^2).$$

By Theorem 6.2.3 of Priestley (1981) we have, with $\Gamma_T(j, k)$ uniformly bounded in j, k, T and with $\delta_{j,k}^* = 1$ for $j = \pm k$ and $\delta_{j,k}^* = 0$ otherwise,

$$\text{(A6)} \qquad \operatorname{cov}(I_{T,j}, I_{T,k}) = \delta_{j,k}^* f^2(\omega_j) + \frac{1}{T}\Gamma_T(j, k) \quad \text{for all } 1 \le |j|, |k| \le N.$$

Using this relation, (A3), Lemma A2 and the compactness of the support of K, a straightforward calculation shows

$$\text{(A7)} \qquad ThE \left\{ \sum_{j=-N}^{N} \gamma_j(\omega; h, g) \left[I_{T,j} - f(\omega_j) \right] \right\}^2 \le \frac{c^* h^3}{g^3} \to 0$$

for suitable $c^* > 0$. As $E^* I_{T,j}^* = \hat{f}(\omega_j; g)$, we have

$$b_T(\omega) - b_T^*(\omega) = \frac{1}{f(\omega)} \{ E\hat{f}(\omega; h) - f(\omega) \} - \frac{1}{\hat{f}(\omega; g)} \sum_{j=-N}^{N} \gamma_j(\omega; h, g) I_{T,j}.$$

Therefore, using (A5), (A7) and $\hat{f}(\omega; g) \to f(\omega) > 0$ in probability, we finally get $B_T = \sqrt{Th}\,(b_T(\omega) - b_T^*(\omega)) \to 0$ in probability by proving $\sqrt{Th}\,a_T(\omega) \to 0$ with

$$a_T(\omega) = \sum_{j=-N}^{N} \alpha_j(\omega; h) f(\omega_j) - \sum_{j=-N}^{N} \gamma_j(\omega; h, g) f(\omega_j).$$

For this purpose, we split $a_T(\omega)$ into three parts and show that the first and third parts are of order $O(1/(Tg))$ and the second is of order $o(h^2)$:

$$\begin{aligned} a_T(\omega) = {} & \sum_{j=-N}^{N} \alpha_j(\omega; g) f(\omega_j) - \frac{1}{2\pi}\int_{-\infty}^{\infty} K(\theta) f(\omega + \theta g)\, d\theta \\ & + p(\omega; g) - \sum_{j=-N}^{N} \alpha_j(\omega; h) p(\omega_j; g) \\ & + \sum_{j=-N}^{N} \alpha_j(\omega; h) \left\{ \frac{1}{2\pi}\int_{-\infty}^{\infty} K(\theta) f(\omega_j + \theta g)\, d\theta \right. \\ & \left. \qquad - \sum_{j=-N}^{N} \alpha_k(\omega_j; g) f(\omega_k) \right\}, \end{aligned} \tag{A8}$$

where

$$p(\omega; g) = \frac{1}{2\pi}\int_{-\infty}^{\infty} K(\theta) f(\omega + \theta g)\, d\theta - f(\omega).$$

As in the proof of Lemma A1, the compactness of the support of K and the Lipschitz continuity of K and f imply that the first part of (A8) is bounded by a constant multiple of $1/(Tg)$ for T large enough. This upper bound is uniform in $|\omega| \le \pi - \kappa g$. Therefore, the third line of (A8) is asymptotically of order $1/(Tg)$ too, because only summands with $|\omega - \omega_j| \le \kappa h$ do not vanish.

Because f is twice continuously differentiable and K is bounded, $p(\omega; g)$ is twice continuously differentiable on $[-\pi + \kappa g,\ \pi - \kappa g]$. Using Lebesgue's theorem on dominated convergence, we conclude that $p(\omega; g)$, $p'(\omega; g)$ and $p''(\omega; g)$ converge to 0 uniformly on $[\omega - \delta,\ \omega + \delta]$ for all $\delta < \pi - |\omega|$. Applying the first part of Lemma A1, we get that the second line of (A8) is asymptotically $o(h^2)$.

(e) The proof of (ii) follows exactly the same lines as the proof of (i), but is easier. In particular, defining D^+ as D^* with $\hat{f}^+$ replacing $\hat{f}^*$, (A4) would be replaced by

$$d_2^2(D^0, D^+) \le Th \sum_{j=-N}^{N} \alpha_j^2(\omega; h) d_2^2\big[f(\omega_j)\{\chi_j - 1\},\ \hat{f}(\omega_j; g)\{\chi_j - 1\}\big] \to 0$$

in probability by Corollary A2, and the rest of part (c) of the proof is not necessary.

LEMMA A2. *Let K, h, g satisfy the assumptions of Theorem 1. Then $\gamma_j(\omega; h, g) = O(h/(Tg^2))$ uniformly in $|\omega| \le \pi - \kappa h$. In particular, $\gamma_j(\omega; h, g) = 0$ if $|\omega - \omega_j| > (h+g)\kappa$.*

PROOF. By definition of γ_j,

$$\begin{aligned}|\gamma_j(\omega; h, g)| \le & \frac{1}{T^2 hg} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_k}{h}\right) \left| K\left(\frac{\omega_j - \omega_k}{g}\right) - K\left(\frac{\omega_j - \omega}{g}\right) \right| \\ & + \frac{1}{Tg} K\left(\frac{\omega_j - \omega}{g}\right) |S_T(\omega) - 1|.\end{aligned}$$

The first term on the right-hand side is of order $h/(Tg^2)$ by Lipschitz continuity of K and (A2). The second term is of order $1/(T^2 gh)$ by (A1). The compactness of the support of K implies $\gamma_j(\omega; h, g) = 0$ for $|\omega - \omega_j| > (h+g)\kappa$. □

PROOF OF THEOREM 3. The proof is a combination of arguments given by Rice (1984) and of results which we have already obtained in the course of proving Theorem 1. We, therefore, only give a sketchy outline of the arguments. We use the notation

$$l(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} K^2(\theta)\, d\theta \frac{1}{z} + \left[\frac{z^2 f''(\omega)}{2 f(\omega)}\right]^2$$

which is the asymptotically dominating part of $T^{4/5}$ MSPE$(\omega; h)$ for $h = zT^{-1/5}$ [compare (2) of Section 2]. Using (A7) and a Taylor expansion argument as in the proof of Lemma A1,

$$\text{(A9)} \qquad \sup_{a \le z \le b} \left| T^{4/5} \, \text{MSPE}(\omega; zT^{-1/5}) - l(z) \right| \to 0 \quad \text{for } T \to \infty$$

for arbitrary $0 < a < b < \infty$. A calculation of derivatives shows that $l(z)$ is strictly convex and infinitely often differentiable on $(0, \infty)$, and that it has z_∞ of (5) as a unique minimum, provided $f''(\omega) \ne 0$. These properties and (A9) imply $T^{1/5} h_0 \to z_\infty$ for $T \to \infty$, provided $a < z_\infty < b$. As the next step, we prove

$$\text{(A10)} \qquad \sup_{h \in B_T} T^{4/5} |\text{MSPE}^*(\omega; h) - \text{MSPE}(\omega; h)| \to 0 \quad \text{in probability for } T \to \infty.$$

This convergence is shown separately for the variance part and for the bias part of the mean-square percentage error, noticing also that we can forget about the denominators as $\hat{f}(\omega; g) \to f(\omega) > 0$ in probability. The convergence of the variance part of (A10) follows rather easily from Corollary A1, using (A2) and the asymptotic properties (A7) of the periodogram. To prove that the difference of the bootstrap bias and the bias of $\hat{f}(\omega; h)$ itself converges to 0 faster than $T^{-2/5}$, one has to repeat the arguments of part (d) of the proof of Theorem 1, remarking that all of them hold uniformly in $h \in B_T$.

Now (A9), (A10), (5) and the regularity of $l(z)$ imply $T^{1/5}(h_0^* - h_0) \to 0$ in probability, using exactly the same arguments as by Rice (1984) in the proof of his Corollary 2.2. By the first part of Theorem 3, we immediately conclude the second part of Theorem 3 because $l(z)$ is continuous and because, by (A9) and (A10), $\mathrm{MSPE}^*(\omega; h)$ and $\widehat{\mathrm{MSPE}}(\omega; h)$ can both be approximated by $T^{-4/5}l(hT^{1/5})$ uniformly in $h \in B_T$. □

Acknowledgments. This research has been supported by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 123 Stochastische Mathematische Modelle. A substantial part of this work has been done while the first author, as a visiting fellow, enjoyed the hospitality of the Department of Statistics, IAS, Australian National University, Canberra. We thank Don Poskitt and Enno Mammen for some stimulating remarks on earlier versions of this paper.

REFERENCES

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

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

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

CHEN, Z.-G. and HANNAN, E. J. (1980). The distribution of periodogram ordinates. *J. Time Ser. Anal.* **1** 73–82.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

EFRON, B. and TIBSHIRANI, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Statist. Sci.* **1** 54–77.

FRANKE, J. (1987). On the choice of local bandwidth for kernel spectral estimates using the bootstrap. Preprint 409, SFB 123 Stochastische Mathematische Modelle, Univ. Heidelberg.

FREEDMAN, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

FREEDMAN, D. A. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.* **12** 827–842.

GASSER, T., MÜLLER, H. G., KÖHLER, W., MOLINARI, L. and PRADER, A. (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.* **12** 210–229.

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.

HÄRDLE, W. and BOWMAN, A. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83** 100–110.

HARTIGAN, J. A. (1990). Perturbed periodogram estimates of variance. *Internat. Statist. Rev.* **58** 1–7.

KREISS, J. P. and FRANKE, J. (1989). Bootstrapping stationary ARMA-models. *J. Time Ser. Anal.* To appear

KÜNSCH, H.-R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241.

MAJOR, P. (1978). On the invariance principle for sums of independent, identically distributed random variables. *J. Multivariate Anal.* **8** 487–501.

PARZEN, E. (1957). On choosing an estimate of the spectral density function of a stationary time series. *Ann. Math. Statist.* **28** 921–932.

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series* **1**. Academic, New York.

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.

ROMANO, J. P. (1988). Bootstrapping the mode. *Ann. Inst. Statist. Math.* **40** 565–586.

SHIBATA, R. (1981). An optimal autoregressive spectral estimate. *Ann. Statist.* **9** 300–306.

SHIRYAYEV, A. N. (1984). *Probability*. Springer, Berlin.

SWANEPOEL, J. W. H. and VAN WYK, J. W. Y. (1986). The bootstrap applied to power spectral density function estimation. *Biometrika* **73** 135–141.

WOODROOFE, M. B. and VAN NESS, J. W. (1967). The maximum deviation of sample spectral densities. *Ann. Math. Statist.* **38** 1559–1569.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF KAISERSLAUTERN
P.O. BOX 3049
D-6750 KAISERSLAUTERN
GERMANY

C.O.R.E.
UNIVERSITÉ CATHOLIQUE DE LOUVAIN
34, VOIE DU ROMAN PAYS
B-1348 LOUVAIN-LA-NEUVE
BELGIUM