LOCAL CROSS VALIDATION FOR SPECTRUM BANDWIDTH CHOICE

By Carlos Velasco

Universidad Carlos III de Madrid

First version received January 1998

Abstract. We investigate an automatic method of determining a local bandwidth for non-parametric kernel spectral density estimates at a single frequency. This procedure is a modification of a cross-validation technique for global bandwidth choices, avoiding the computation of any pilot estimate based on initial bandwidths or on approximate parametric models. Only local conditions on the spectral density around the frequency of interest are assumed. We illustrate with a Monte Carlo study the performance in finite samples of the bandwidth estimates proposed.

Keywords. Bandwidth selection; nonparametric spectral estimation; cross-validation; time series; periodogram.

1. INTRODUCTION

Smoothed estimation of the spectral density of stationary time series, like many non parametric methods of inference, relies on the choice of a bandwidth or lag number depending on the sample size. The properties of the estimates depend crucially on the value of this number. Asymptotic theory prescribes a rate for the lag number M with respect to the sample size N as this tends to infinity, but gives no practical guidance for the choice of M in finite samples. Different techniques have been proposed in the literature to that end. The usual criterion is the minimization of some estimate of the asymptotic mean square error of the estimator. This can be implemented by plug in or cross validation methods. Also, global and local choices are possible, depending on whether we are interested in the behaviour of the spectral density for the whole range of frequencies or at a specific point or small interval.

The plug in method consists of substituting the unknowns of the leading term in the asymptotic expression for the mean square error by consistent estimates, generally non parametric, but parametric estimates based on approximate models can also be used. Cross validation procedures avoid the use of initial estimates and approximate the mean square error indirectly. They are based on estimates that do not use the information contained in the sample about the function of interest at the particular point at which the estimate is being evaluated (at each Fourier frequency in the case of spectral estimation).

Beltrao and Bloomfield (1987) (BB hereafter) considered bandwidth choice for discrete periodogram average type spectral estimates. They justified a

method based on a cross validated form of Whittle's frequency domain approximation to the likelihood function of a stationary Gaussian process (see also Hurvich, 1985). Robinson (1991) extended their results under more general conditions for a wider class of models, including spectral estimation for the construction of efficient regression estimates, and proved the consistency of the estimate of M. This cross validated method selects a global bandwidth for all the range of frequencies $[-\pi, \pi]$ or for a fixed subset of it. Here we propose a modified version of cross validation to justify a local bandwidth choice for a single frequency, following some ideas suggested by Robinson (1991, p. 1346), related also with the work of Hurvich and Beltrao (1994) in a different context. For this single frequency choice, we only use local smoothness properties of the spectral density of the time series around this frequency, allowing for a broader range of dependence models. This local adaptation could also lead to efficiency gains when estimation of the spectrum for the whole range of frequencies $[-\pi, \pi]$ is in mind.

The method we analyse here can be seen as the cross validation alternative to Bühlmann's (1996) iterative local plug in procedure for lag window spectral estimates, proposed by Brockmann *et al.* (1993) in the context of kernel regression estimators (see also Herrmann, 1997), or to the related proposal of Newey and West (1994) for covariance matrix estimation. Local adaptation is also studied by Fan and Kreutzberger (1998), who used the variable bandwidth selector of Fan and Gijbels (1995) for a local polynomial maximum likelihood fit of the spectral density, and by Lepskii *et al.* (1997) for kernel estimates of a regression model with inhomogeneous smoothness.

Alternatively, Wahba (1980) considered automatic smoothing methods for the log periodogram and Pawitan and O'Sullivan (1994) and Kooperberg *et al.* (1995) extended Wahba's (1980) idea using a penalized Whittle likelihood. These methods, like Efromovich's (1998) efficient data driven kernel procedure for uniformly smooth spectral densities, make a global choice of the smoothing parameter, while in many cases we are interested in obtaining optimal local bandwidths for the original scale.

The next section is devoted to the assumptions used in the paper, together with a brief introduction to the main cross validation concepts for non parametric spectrum estimation and a detailed analysis of the mean square error for the spectral estimate at a fixed frequency under local smoothness assumptions. Section 3 introduces the local cross validation criterion and the main result of the paper. Then in Section 4 we carry out a Monte Carlo analysis of the finite sample behaviour of the techniques proposed. All the proofs and some technical lemmas required are given in the Appendix.

2. ASSUMPTIONS AND DEFINITIONS

In this section we introduce some assumptions and definitions, together with some intuitions about cross validation and BB's results. Given the observed data

X_t, $t = 1, 2, \ldots, N$, the periodogram at the frequency $\lambda_j = 2\pi j/N$, j integer, is equal to

$$I(\lambda_j) \stackrel{\text{def}}{=} \frac{1}{2\pi N} \left| \sum_{t\ 1}^{N} X_t \exp(it\lambda_j) \right|^2 .$$

The averaged periodogram spectral estimate with lag number $M = M_N = h_N^{-1}$, where h_N is the bandwidth of the estimate in BB's notation, and kernel or spectral window K (this function was denoted W in BB, but we use this notation later for another analogous function) is

$$\hat{f}_M(\lambda_j) \stackrel{\text{def}}{=} \sigma_M^{-1} \sum_k K(M\lambda_k) I(\lambda_j \quad \lambda_k)$$

where the summation runs for all values of k in the support of K, but not including indices k with $\lambda_k \quad \lambda_j = 2\pi r$ for $r \in \mathbb{N}$ (to account for mean correction), and σ_M gives the exact sum of the weights used,

$$\sigma_M \stackrel{\text{def}}{=} \sum_k K(M\lambda_k).$$

Here σ_M^{-1} could be replaced by the approximate value $2\pi M/N$, using the fact that K integrates to 1, but this simplifes some arguments. We stress the dependence of $\hat{f}_M$ on M in the notation, since this is the *parameter* of interest.

We now give BB's regularity conditions to compare later with our *local* Assumptions 1 4 below. BB (cf. their Theorem 3.1) considered a zero mean stationary Gaussian process $\{X_t\}$ with autocovariance function $\gamma(r) = E(X_0 X_r)$ satisfying

$$\sum_1^{\infty} r|\gamma(r)| < \infty \tag{1}$$

and spectral density $f(\lambda) = (2\pi)^{-1} \sum_{-\infty}^{\infty} \gamma(r) \exp\{ir\lambda\}$ everywhere positive. The kernel function K for non parametric estimation was non negative, even, bounded, with

$$\int_{-\infty}^{\infty} K(x)\, dx = 1 \qquad w_2 \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x^2 K(x)\, dx < \infty. \tag{2}$$

They also wrote $K(x) = \int w(y) \exp(ixy)\, dy$, where w is of compact support. Finally, the bandwidth h_N used by BB satisfied $h_N^{-1} = \mathrm{O}(N^{\rho})$, for some $\rho < 2/5$ and $h_N = \mathrm{o}(1)$.

The 'leave two out version' of the estimator $\hat{f}_M$ (we leave only two frequencies out if K is actually compactly supported inside $[\ \pi, \pi]$, as we assume later on, or if we have defined its periodic version in that interval) is

$$\hat{f}^{\mathrm{j}}_M(\lambda_j) \stackrel{\text{def}}{=} \sigma_{j,M}^{-1} \sum_k{}' K(M\lambda_k) I(\lambda_j \quad \lambda_k) \tag{3}$$

where $\sum_k'$ runs for the same values as before except in the set of indices $k \in \{2j, 2j \pm N, \ldots\}$ (by the definition of $\sum_k$ this means $k \notin \{0, \pm N, \ldots\} \cup \{2j, \ldots\}$). Also, the normalizing number $\sigma_{j,M}$ is equal to

$$\sigma_{j,M} \stackrel{\text{def}}{=} \sum_k{}' K(M\lambda_k).$$

Introduce the pseudo log likelihood type criterion

$$L(f) \stackrel{\text{def}}{=} \sum_{j\ 1}^{N-1} \left\{ \log f(\lambda_j) + \frac{I(\lambda_j)}{f(\lambda_j)} \right\} \tag{4}$$

which is Whittle's approximation for the likelihood of a Gaussian sequence in the frequency domain. BB showed under the previous conditions that

$$L(\hat{f}_M) \quad L(f) = \frac{N}{2}\text{IMSE}(M)$$

plus a term of smaller order in probability, where $\text{IMSE}(M)$ is the discrete approximation to the relative integrated mean squared error of $\hat{f}_M$, weighted by f^{-1}:

$$\text{IMSE}(M) \stackrel{\text{def}}{=} N^{-1} \sum_{j\ 1}^{N-1} E\left\{ \frac{\hat{f}_M(\lambda_j) \quad f(\lambda_j)}{f(\lambda_j)} \right\}^2.$$

Then minimizing $L(\hat{f}_M)$ and $\text{IMSE}(M)$ should be approximately equivalent, and this is the basis for the estimation of the M that minimizes the $\text{IMSE}(M)$ of $\hat{f}_M(\lambda)$ for in $[\ \pi, \pi]$.

If we are interested in non parametric spectral estimation at a single frequency (of special interest is the zero frequency; see Bühlmann's (1996) examples, together with covariance matrix estimation in econometrics, as in Den Haan and Levin (1997) and the references therein) or we want to achieve possible efficiency gains using different bandwidths for each frequency, we need a rule to choose a local bandwidth. The relative mean square error at a frequency ν

$$\text{MSE}(\nu,\ M) \stackrel{\text{def}}{=} E\left\{ \frac{\hat{f}_M(\nu) \quad f(\nu)}{f(\nu)} \right\}^2$$

is the usual criterion employed to assess non parametric spectral estimates at a single frequency.

We use the following assumptions.

ASSUMPTION 1. X_t, $t = 1, 2, \ldots$, is a Gaussian stationary time series.

ASSUMPTION 2. The spectral density $f(\lambda)$ of X_t has three uniformly bounded

derivatives in an interval around the fixed frequency λ, with $f(\lambda) > 0$ for λ in that interval and $f \in L_p[\ \pi, \pi]$ for some $p > 5/3$.

ASSUMPTION 3. K is non negative, even, bounded, zero outside $[\ \pi, \pi]$, of bounded variation and satisfies (2).

ASSUMPTION 4. K has Fourier transform $w(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} K(\lambda) \exp(i\lambda x)\, d\lambda$ satisfying

$$|w(x)| = O(|x|^{-\alpha}) \qquad \text{as } |x| \to \infty$$

for some $\alpha > 5/4$.

Assumption 1 was used also in BB, but we do not need to assume zero mean since we avoid the zero frequency periodogram ordinate in the definition of our estimates. Assumption 2 only requires smoothness properties of f around the frequency of interest, allowing for a wide class of spectral densities, including ones with zeros and poles outside a neighbourhood of ν, which are ruled out by (1). The only requirement outside this band is an integrability condition to ensure ergodicity (with respect to second moments) of the series (see Lemma 7 in the Appendix).

A compact support kernel in Assumption 3 is then the complement of Assumption 2 in order to guarantee that only information in an interval around ν is used. The rest of the conditions on K are standard, Assumption 4 being necessary to approximate $\hat{f}_M$ by a weighted autocovariance type estimate in Lemma 5. From this lemma, both estimates have the same asymptotic distribution and mean square error, so the bandwidth choice techniques for one are valid for the other. This condition is satisfied by the Bartlett Priestley and quadratic spectral kernel (with $\alpha = 2$), but not by the Daniell or uniform spectral window.

With Assumption 3, the summation in k in the definition of $\hat{f}_M$ takes values in $\{j \quad N+1, \ldots, j \quad 1\} \quad \{j\}$ due to the compact support kernel, and in $\{j \quad N+1, \ldots, j \quad 1\} \quad \{j, 2j, 2j \quad N\}$ for $\hat{f}^{\mathrm{j}}_M(\lambda_j)$.

We now present a result concerning the mean square error of the estimate $\hat{f}_M$ at Fourier frequencies λ_j, which is used to analyse a local version of the likelihood (4). We use in the proof two lemmas given in the Appendix about the discrete Fourier transform and periodogram of the observed sequence, extending and correcting some of the results of BB, assuming only local smoothness for the spectral density. We have to distinguish between estimates for Fourier frequencies λ_j close to the origin and estimates at remote frequences. Define $\|K\|_2^2 = \int K^2(x)\, dx$ and let c be a finite positive constant, not necessarily always the same.

LEMMA 1. *Under Assumptions 1, 2 and 3, if $M = cN^{1/5}$, for frequencies $\lambda_j = 2\pi j/N$ such that $|\nu \quad \lambda_j| \leqslant cm^{-1}$ for some positive sequence $m = m_N$ such that $1/m + m/M \to 0$ as $N \to \infty$, then uniformly in j, for $\nu > 0$*

$$\mathrm{MSE}(\lambda_j,\, M) = \frac{M}{N} 2\pi \|K\|_2^2 f(\lambda_j)^2 + M^{-4} \left\{ \frac{w_2}{2} f^{(2)}(\lambda_j) \right\}^2 + \mathrm{O}\left(\frac{M^2}{N^2} + N^{-1} \right),$$

where $f^{(2)}$ *is the second derivative of* f, *and for* $\nu = 0$

$$\mathrm{MSE}(\lambda_j,\, M) = \frac{M}{N} 2\pi \|K\|_2^2 f(\lambda_j)^2 \{1 + \delta_M(j)\} + M^{-4} \left\{ \frac{w_2}{2} f^{(2)}(\lambda_j) \right\}^2 + \mathrm{O}\left(\frac{M^2}{N^2} + N^{-1} \right)$$

where w_2 *is defined in (2) and*

$$\delta_M(j) = \|K\|_2^{-2} \int_{-\infty}^{\infty} K(\lambda) K(\lambda + 2M\lambda_j)\, d\lambda. \tag{5}$$

Note that all the bounds are uniform in j and that $\delta_M(j)$ measures the degree of overlapping between different kernels K at a distance $2M\lambda_j$ apart when $\lambda_j \to 0$ as $N \to \infty$, with $0 \leqslant \delta_M(j) \leqslant 1$. For $j = 0$, $\delta_M(j) = 1\ \forall M$, and for $\lambda_j > 2\pi/M$, $\delta_M(j) = 0$.

For $\nu > 0$, fixed, this is the standard result for globally smooth spectral densities (see, for example, Brillinger, 1975, Corollaries 5.6.1 and 5.6.2). However, in a degenerating band around the origin (small λ_j), the non parametric spectral estimates have variance depending on the overlapping of two kernel functions K centred at frequencies λ_j and λ_j respectively, measured by the quantity $\delta_M(j)$.

To make the bias and the variance of the same order of magnitude we would take $M = \tau N^{1/5}$, for some $0 < \tau < \infty$, and then MSE is of order $M^{-4} \sim cM/N$. From the previous lemma, the optimal constant $\tau^*(\nu)$ that minimizes the leading term of the mean square error of $\hat{f}_M(\nu)$ is

$$\tau^*(\nu) = \left\{ \frac{w_2 f^{(2)}(\nu)^2}{2\pi \|K\|_2^2 f(\nu)^2} \right\}^{1/5} \tag{6}$$

if $\nu \neq 0$ and with 4π instead of 2π for $\nu = 0$. Now it is possible to estimate the value of $\tau^*(\nu)$ using initial, pilot estimates of the spectral density and its second derivative at ν. This is the approach of several authors, including Andrews (1991), Newey and West (1994) and Bühlmann (1996), to give some recent contributions. In the following section we adopt instead an indirect approach using a cross validation argument.

3. LOCAL CROSS-VALIDATION

Consider for some positive sequence $m = m_N$ such that $1/m + m/M \to 0$ as $N \to 0$, one form of *local* integrated mean square error

$$\mathrm{IMSE}_m(\nu,\ M) \stackrel{\mathrm{def}}{=} \frac{2\pi}{N}\sum_{j\ 1}^{N-1} W_m(\lambda_j \quad \nu) E\left\{\frac{\hat{f}_M(\lambda_j) \quad f(\lambda_j)}{f(\lambda_j)}\right\}^2$$

where $W_m(\lambda) = m\sum_j W\{m(\lambda + 2\pi j)\}$ for some appropriate kernel function W satisfying Assumption 3. For the uniform kernel $W = (2\pi)^{-1} I_{[-\pi,\pi]}$ and $m = 1$, we have $\mathrm{IMSE}_m(\nu,\ M) = \mathrm{IMSE}(M)$ for all ν.

Then, from Lemma 1 and for $\nu > 0$ we can obtain under the same regularity conditions

$$\begin{aligned}\mathrm{IMSE}_m(\nu,\ M) &= \frac{M}{N} 2\pi \|K\|_2^2 + M^{-4}\left\{\frac{w_2}{2}\frac{f^{(2)}(\nu)}{f(\nu)}\right\}^2 \\ &\quad + \mathrm{O}\left\{\frac{M^2}{N^2} + \frac{1}{N} + \frac{M}{N}\left(\frac{m}{N} + \frac{1}{m}\right)\right\} \\ &= \mathrm{MSE}(\nu,\ M) + \mathrm{o}\{\mathrm{MSE}(\nu,\ M)\}\end{aligned}$$

where the errors in m arise from the continuous approximation to the sum in IMSE_m and we use that the ratio $f^{(2)}(\nu)/f(\nu)$ has bounded derivative. Therefore $\mathrm{IMSE}_m(\nu,\ M)$ approximates $\mathrm{MSE}(\nu,\ M)$ when $\nu > 0$ as $m \to \infty$ and $m/M \to 0$.

When $\nu = 0$, we can see that

$$\begin{aligned}\mathrm{IMSE}_m(0,\ M) &= \frac{2\pi}{N}\sum_{j\ 1}^{N-1} W_m(\lambda_j)\left[\frac{M}{N} 2\pi \|K\|_2^2 \{1 + \delta_M(j)\}\right] \qquad (7) \\ &\quad + M^{-4}\left\{\frac{w_2}{2}\frac{f^{(2)}(0)}{f(0)}\right\}^2 + \mathrm{O}\left\{\frac{M^2}{N^2} + N^{-1} + \frac{M}{N}\left(\frac{m}{N} + m^{-1}\right)\right\}\end{aligned}$$

Now in the summation in (7) we consider only the values of λ_j smaller than $2\pi/M$ in absolute value, because $|\delta_M(j)| = 0$ if $|\lambda_j| > 2\pi/M$ (i.e. $|j| > N/M$). Then, using that $|\delta_M(j)| \leqslant 1\ \forall j$, $\sup_{m,j}|W_m(\lambda_j)| = \mathrm{O}(m)$ and $m/M \to 0$,

$$\begin{aligned}&\frac{2\pi}{N}\sum_j W_m(\lambda_j)\left[\frac{M}{N} 2\pi \|K\|_2^2 \{1 + \delta_M(j)\}\right] \\ &= \frac{M}{N} 2\pi \|K\|_2^2 \frac{2\pi}{N}\sum_j W_m(\lambda_j) + \frac{M}{N} 2\pi \|K\|_2^2 \frac{2\pi}{N}\sum_{|j| \leqslant N/M} W_m(\lambda_j)\delta_M(j) \qquad (8) \\ &= \frac{M}{N} 2\pi \|K\|_2^2 + \mathrm{O}\left(\frac{M}{N}\frac{m}{N}\right) + \mathrm{O}\left(\frac{m}{N}\right) \\ &= \frac{1}{2}\mathrm{var}\{\hat{f}_M(0)\} + \mathrm{o}\left(\frac{M}{N}\right). \qquad (9)\end{aligned}$$

Therefore, when $\nu = 0$, $\mathrm{IMSE}_m(0, M)$ only estimates half of the asymptotic variance in $\mathrm{MSE}(0, M)$, because the kernel W_m includes estimates $\hat{f}_M(\lambda_j)$ too far away from the origin and then with much smaller variance than $\hat{f}_M(0)$. However, the second term in (8) of magnitude m/N, corresponding to the overlapping factor $\delta_M(j)$ in Lemma 1, contributes to $\mathrm{IMSE}_m(0, M)$ in finite samples and could help to approximate the true variance. A possible approach to obtaining a consistent estimate of the optimal local bandwidth which minimizes $\mathrm{MSE}(\nu, M)$, $M^* = \tau^*(\nu)N^{1/5}$, is to minimize an estimate of $\mathrm{MSE}(\nu, M)$ or of $\mathrm{IMSE}_m(\nu, M)$, which approaches the former as m increases. Some adjustments might be necessary in the case $\nu = 0$ due to the problem described in the previous paragraph. The presence of two related bandwidth parameters, m and M, seems to imply a circular argument like that present in a plug in method, where pilot estimates of the spectral density and its derivatives are used, depending on other bandwidths or parametric assumptions. To circumvent this problem we describe some procedures in the next section showing that the choice of m might not be too decisive.

The logical cross validation argument in this case would be the minimization with respect to M of the function (recalling the definition of the 'leave two out' spectral estimate in (3))

$$\mathrm{CVLL}_m(\nu, M) \stackrel{\mathrm{def}}{=} 2\pi \sum_{j=1}^{N-1} W_m(\lambda_j - \nu)\left\{\log \hat{f}^{\mathrm{j}}_M(\lambda_j) + \frac{I(\lambda_j)}{\hat{f}^{\mathrm{j}}_M(\lambda_j)}\right\}$$

which is a likelihood that tends to use only the information around ν as $m \to \infty$. Since W has compact support $[-\pi, \pi]$, about N/m frequencies around ν are used. It is likely that this procedure leads to more variability than the global procedure, since we are not using all information of the sample (see Bühlmann 1996, Section 3.1) or Brockmann *et al.* (1993) for a related problem in non parametric regression).

To justify the above ideas we have the following proposition, proved in the Appendix.

PROPOSITION 1. Under Assumptions 1, 2, 3 and 4, with W satisfying Assumption 3, $M = cN^{1/5}$ and $1/m + m/M \to 0$ as $N \to \infty$,

$$\mathrm{CVLL}_m(\nu, M) = 2\pi \sum_{j=1}^{N-1} W_m(\lambda_j - \nu)\left\{\log f(\lambda_j) + \frac{I(\lambda_j)}{f(\lambda_j)}\right\}$$

$$+ \frac{N}{2}\mathrm{IMSE}_m(\nu, M) + \mathrm{o_P}\{N\ \mathrm{IMSE}_m(M)\}.$$

where $0 < \mathrm{IMSE}_m(\nu, M)/\mathrm{IMSE}(M) < \infty$ as $N \to \infty$, and the first term on the right hand side depends only on m (but not on M).

Then, under regularity conditions, $\mathrm{CVLL}_m(\nu, M)$ is a consistent estimator of $\mathrm{IMSE}_m(\nu, M)$ up to a constant not depending on M. From there, minimization

of CVLL_m should be approximately equal to minimization of IMSE_m. Since the latter approximates $\mathrm{MSE}(\nu)$ under similar conditions on m, we can expect to obtain reasonable estimates of the local optimal M using the local cross validation criterion with $\hat{M}(\nu) = \arg\min_M \mathrm{CVLL}_m(\nu, M)$.

BB did not require to estimate explicitly IMSE or its asymptotic rate of convergence, but in our case we need to do so because we estimate a local MSE from an IMSE calculated from estimates around the frequency of interest. To this end, additional stronger conditions are required for the spectrum at that frequency, but we do not need to make global assumptions for the spectral density or for the autocovariance sequence, allowing for the presence of peaks at remote frequencies due to persistent or cyclic behaviours.

4. MONTE CARLO WORK

In this section we assess whether all the asymptotic arguments given in the previous sections are good approximations for reasonable finite sample sizes and whether the cross validation leads to sensible bandwidth estimations. We concentrated first on the special case of the estimation of the bandwidth for non parametric spectral estimates at the origin ($\nu = 0$), and then on the estimation of the spectral density for all $\lambda \in [\ \pi, \pi]$, following Bühlmann (1996, Section 3).

We simulated Gaussian sequences following five different models and three sample sizes, $N = 120$, 256 and 480. The models considered are the following AR processes:

$$X_t = \sum_{j\ 1}^{p} \alpha_j X_{t-j} + \epsilon_j \qquad \epsilon_t \sim \text{iidN}(0, 1)$$

with parameters

MODEL 1 AR(3): $\alpha_1 = 0.6$, $\alpha_2 = \ \ 0.6$, $\alpha_3 = 0.3$

MODEL 2 AR(2): $\alpha_1 = 0.6$, $\alpha_2 = \ \ 0.9$

MODEL 3 AR(1): $\alpha_1 = 0.8$

MODEL 4 AR(2): $\alpha_1 = 1.372$, $\alpha_2 = \ \ 0.677$

MODEL 5 AR(5): $\alpha_1 = 0.9$, $\alpha_2 = \ \ 0.4$, $\alpha_3 = 0.3$, $\alpha_4 = \ \ 0.5$, $\alpha_5 = 0.3$.

The last three parameter sets were used also by Bühlmann (1996). These models are convenient because of their simplicity and the different spectra they represent. From Figure 1, Model 1's spectral density exhibits a small peak at the origin and a larger one at $\lambda \approx 1.5$. Model 2 is flat at the origin but with a very sharp peak at frequency $\lambda \approx 1.3$. The AR(1) Model 3 has the typical spectral density of an AR(1) series with positive autocorrelation and a maximum at zero

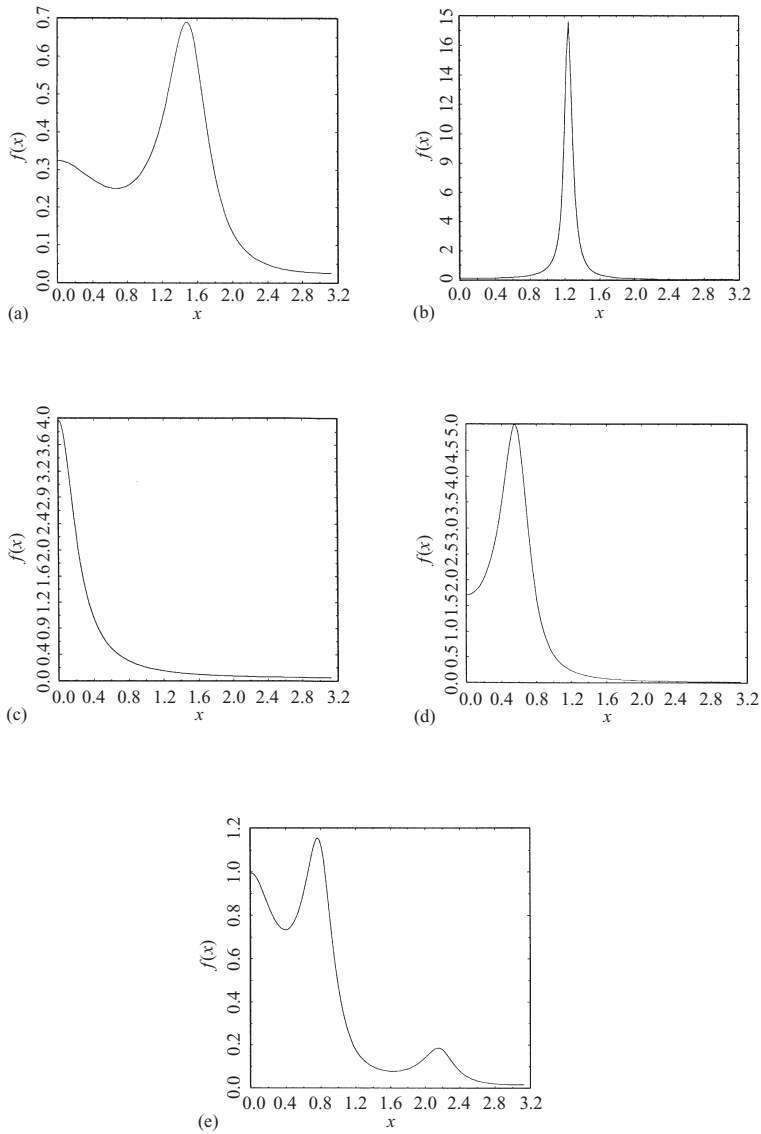

FIGURE 1. Spectral densities for Models 1 5: (a) Model 1, AR(3), $\alpha_1 = 0.6$, $\alpha_2 = 0.6$ $\alpha_3 = 0.3$; (b) Model 2, AR(2), $\alpha_1 = 0.6$, $\alpha_2 = 0.9$; (c) Model 3, AR(1), $\alpha_1 = 0\,8$; (d) Model 4, AR(2), $\alpha_1 = 1.372$, $\alpha_2 = 0.677$; (e) Model 5, AR(5), $\alpha_1 = 0.9$, $\alpha_2 = 0.4$, $\alpha_3 = 0.3$, $\alpha_4 = 0.5$, $\alpha_5 = 0.3$.

frequency. The AR(2) spectrum of Model 4 is similar to the first one but with a minimum at the origin and a closer peak, whereas Model 5's spectrum shows several peaks, including one at the origin.

With these processes we hope to gauge the performance of our approximations in situations where global bandwidths might not be very appropriate due to the presence of special features in the spectral density at the frequency of interest or at remote frequencies which could distort global procedures. We did not normalize the time series to have equal variance or the same spectral density at the origin, since this would only imply multiplying the periodogram of the observed time series by a fixed constant and does not affect any of the methods used.

For the local choice at $\lambda = 0$, we employ the Bartlett–Priestley kernel (for both K and W), with spectral window

$$K(\lambda) = \begin{cases} (3/4\pi)\{1 - (\lambda/\pi)^2\} & |\lambda| \leqslant \pi \\ 0 & |\lambda| \geqslant \pi \end{cases}$$

and lag window

$$w(x) = \frac{3}{(\pi x)^2}\left(\frac{\sin \pi x}{\pi x} - \cos \pi x\right).$$

The uniform kernel was also tried for K, with much less smooth results as a consequence of the non continuity in the boundaries of its support and a lag window with tails slowly decreasing to zero. For the choice at all frequencies $\lambda_j \in [-\pi, \pi]$ we report the results for W equal to the uniform kernel, this case not being very different from those with the Bartlett–Priestley window.

4.1. *Spectral estimation at the origin*

From Equation (9) we know that, for the frequency $\nu = 0$ in particular, $\mathrm{IMSE}_m(0, M)$ does not approach $\mathrm{MSE}(0, M)$ asymptotically due to the different variance of the spectral density estimates around the origin. Nevertheless, from Lemma 1, the transition from the variance of $\hat{f}_M(0)$ to the variance of an estimate at a frequency apart from the origin (one half of the previous one) is smooth, depending on the shape of the kernel used (see the definition of $\delta_M(j)$). Then we can expect that the approximation behaves moderately well also for this case.

We used the following equivalent version of the cross-validated log likelihood, given the periodicity and symmetry of W_m, f and I,

$$\mathrm{CVLL}_m(0, M) = 2\pi \sum_{j=-[N/2]}^{[N/2]} W_m(\lambda_j)\left\{\log \hat{f}^{\mathrm{j}}_M(\lambda_j) + \frac{I(\lambda_j)}{\hat{f}^{\mathrm{j}}_M(\lambda_j)}\right\}$$

dropping the frequency $\lambda_j = 0$ (since due to mean correction $I(0) = 0$ and $\hat{f}_M(0) = \hat{f}^0_M(0)$) and we calculated IMSE_m accordingly.

We first estimate the function $\mathrm{CVLL}_m(0, M)$ for a grid of values of m and M and then we report in Table I the bias (with respect to the minimizer $M^*(0) = \tau^*(0)N^{1/5}$ of the asymptotic MSE$(0, M)$ given by (6)), standard deviation (sd) and mean square error (mse) across replications of the $\hat{M}$ estimated by the minimization of CVLL_m (1000 replications and sample size $N = 256$). The number of different Fourier frequencies around λ_0 over which the kernel W_m averages for each value of m is 'band' $= N/(2m)$. We cover all the values of 'band' from 1 to 115 with increments of size 6, which correspond to values of m from 128 to 1.11. We also report the finite sample optimal $\hat{M}^*(0)$ which minimizes the sample MSE across replications. This value turned out to be quite close to $M^*(0)$ for Models 2 and 4, but much smaller than the asymptotic value for Models 1, 3 and 5. The results are also summarized in the bi and tri dimensional plots of Figures 2 6, where we can only give some of the CVLL_m lines because of the very different scales. In the bi dimensional graphs each line corresponds to the plot of $\mathrm{CVLL}_m(0, M)$ against M for a particular m.

The lag number M estimated by CVLL_m shows relatively stable standard deviations for all values of m between 1 and 5, though increasing slightly with m as expected. This bias displays a different behaviour for each of the models used. In the case of Model 2, higher values of $\hat{M}$ are estimated than the asymptotic optimal ($M^*(0) = 7.13$), and a similar observation holds in inverse direction for Model 3, where much smaller values than $M^*(0) = 20.34$ are found. In general, for moderate values of m only Models 4 and 5 have small bias with respect to the asymptotic $M^*(0)$, while for Models 1 and 2 there is a large positive bias and for Model 3 a negative one.

Focusing on $\hat{M}^*(0)$ instead, we can observe that, for most values of m between 5.12 and 9.85 and for all models, CVLL_m is correcting the estimations of M substantially in the direction of the finite sample minimizer of MSE$(0, M)$. These large values of m correspond to the use of local information for the estimation of $M^*(0)$, but due to the use of a small number of spectral estimates this leads to quite imprecise estimates (high standard deviations) in many cases. Nevertheles, occasionally these estimates of M based on CVLL_m with high m have smaller mse calculated with respect to $M^*(0)$, and this reduction of the mse would extend to all cases if we consider the empirical $\hat{M}^*(0)$.

Summarizing, we find that CVLL_m reflects the different characteristics of the spectral density for a range of values of m and can be useful for studying local properties of the spectral density in order to make a choice for M. The variability of the estimates is relatively high, as in most bandwidth choice methods (characterized by slow rates of convergence) and as in any non parametric method this variance tends to increase in general with the value of m (which is proportional to the inverse of the actual bandwidth of the kernel W_m).

From a theoretical point of view, the choice of m does not have a definitive answer, though it might not be too decisive for approximating $\hat{M}^*(\nu)$. In

TABLE I

$\hat{M}$ MINIMIZING $\mathrm{CVLL}_m(0, M)$

		Model 1 $M^*(0)$: 7.158 $\hat{M}^*(0)$: 1.829			Model 2 $M^*(0)$: 7.131 $\hat{M}^*(0)$: 6.481			Model 3 $M^*(0)$: 20.340 $\hat{M}^*(0)$: 11.130			Model 4 $M^*(0)$: 11.076 $\hat{M}^*(0)$: 8.983			Model 5 $M^*(0)$: 11.969 $\hat{M}^*(0)$: 2.943		
m	band	bias	sd	mse	bias	sd	mse	bias	sd	mse	bias	sd	mse	bias	sd	mse
1.11	115	3.42	4.20	29.35	9.54	4.54	111.52	10.98	4.08	137.08	0.49	3.72	14.10	0.13	3.70	13.74
1.17	109	3.47	4.22	29.82	9.76	4.59	116.36	10.85	4.24	135.62	0.45	3.68	13.76	0.04	3.85	14.81
1.24	103	3.49	4.19	29.74	9.96	4.57	120.13	10.74	4.32	134.08	0.28	3.88	15.11	0.01	3.94	15.54
1.32	97	3.52	4.27	30.64	10.11	4.67	124.04	10.74	4.27	133.53	0.13	4.00	15.98	0.05	3.85	14.84
1.41	91	3.68	4.38	32.69	10.54	4.85	134.69	10.55	4.37	130.44	0.03	3.99	15.94	0.06	3.89	15.15
1.51	85	3.80	4.36	33.48	11.01	5.04	146.66	10.41	4.46	128.19	0.12	4.11	16.90	0.30	4.05	16.52
1.62	79	3.83	4.36	33.69	11.55	5.52	163.94	10.34	4.27	125.14	0.32	4.12	17.11	0.18	4.15	17.22
1.75	73	3.91	4.45	35.09	11.88	5.56	172.01	10.22	4.38	126.63	0.47	4.09	16.98	0.00	4.44	19.72
1.91	67	3.92	4.63	36.79	12.33	5.67	184.10	10.05	4.54	121.58	0.74	4.45	20.39	0.01	4.55	20.71
2.10	61	3.61	5.05	38.56	12.80	6.12	201.33	9.90	4.52	118.52	0.95	4.53	21.42	0.02	4.76	22.70
2.33	55	3.08	5.62	41.13	13.08	6.55	213.86	9.73	4.62	116.11	1.19	4.87	25.15	0.11	5.10	26.06
2.61	49	2.80	6.18	46.00	10.83	5.93	152.52	9.41	4.94	112.97	1.28	4.89	25.49	0.12	5.37	28.83
2.98	43	2.85	6.13	45.74	9.95	5.31	127.17	9.17	5.08	109.82	1.52	5.09	28.21	0.42	5.68	32.46
3.46	37	2.98	6.23	47.65	6.26	4.64	60.72	8.66	5.53	105.52	1.91	5.87	38.12	0.81	6.18	38.88
4.13	31	2.92	6.27	47.85	4.32	4.48	38.69	8.58	5.57	104.62	2.12	5.68	36.79	0.27	6.60	43.68
5.12	25	3.02	6.83	55.78	2.97	5.16	35.43	7.84	6.10	98.70	1.86	6.20	41.91	1.45	7.11	52.70
6.74	19	3.58	3.44	24.66	0.87	4.12	17.70	8.26	6.05	104.87	4.50	5.25	47.76	8.14	2.90	74.61
9.85	13	2.18	3.82	19.32	0.14	4.01	16.08	6.64	7.10	94.61	2.95	6.04	45.25	7.54	3.04	66.15
18.29	7	0.38	4.65	21.76	0.06	5.51	30.40	6.85	8.29	115.77	1.34	6.21	40.33	4.58	5.31	49.24
128.0	1	9.15	10.39	191.59	4.00	9.41	104.49	3.97	9.78	111.51	5.54	10.11	132.95	4.24	10.08	119.50

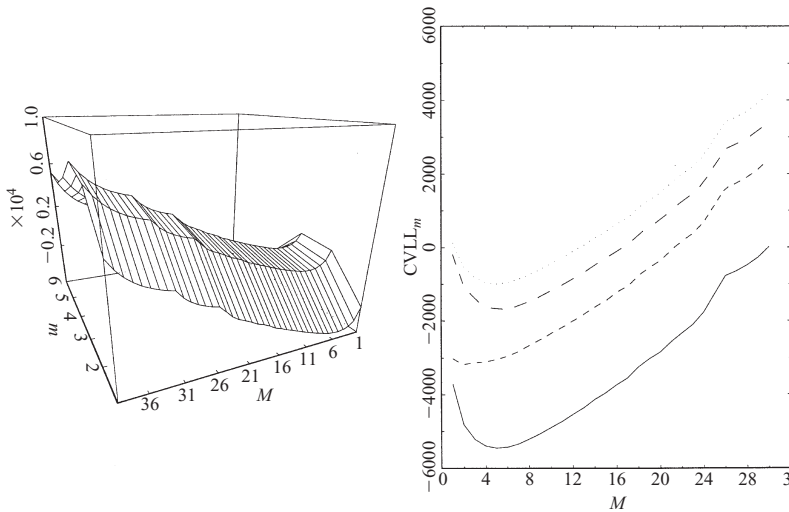

FIGURE 2. $\mathrm{CVLL}_m(M, 0)$ for Model 1. Cross-validated likelihood, AR(3), $\alpha_1 = 0.6$, $\alpha_2 = 0.6$, $\alpha_3 = 0.3$.

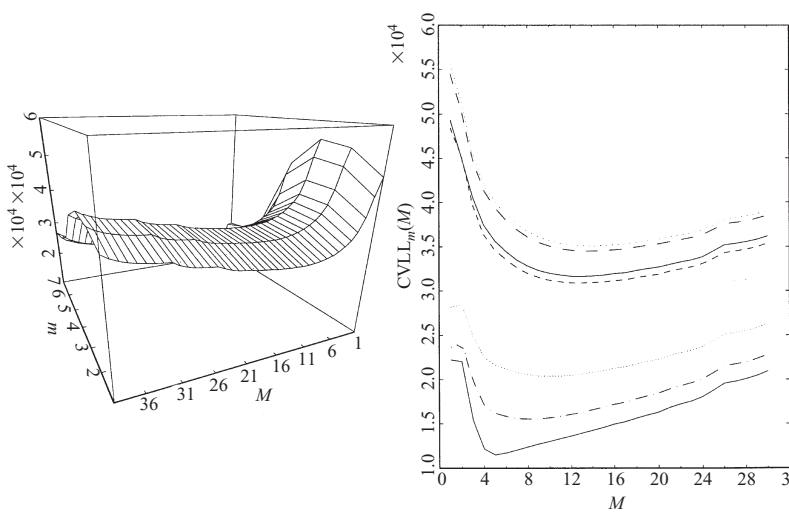

FIGURE 3. $\mathrm{CVLL}_m(M, 0)$ for Model 2. Cross-validated likelihood, AR(2), $\alpha_1 = 0.6$, $\alpha_2 = 0.9$.

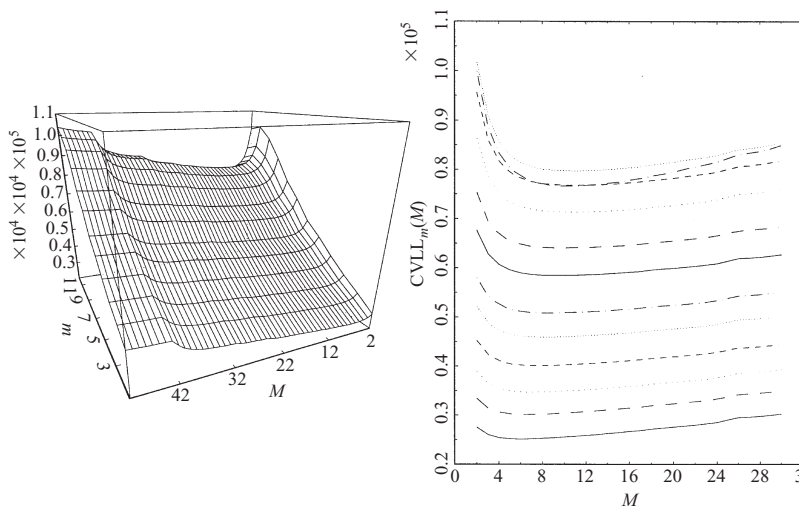

FIGURE 4. $\mathrm{CVLL}_m(M, 0)$ for Model 3. Cross-validated likelihood, AR(1), $\alpha_1 = 0.8$.

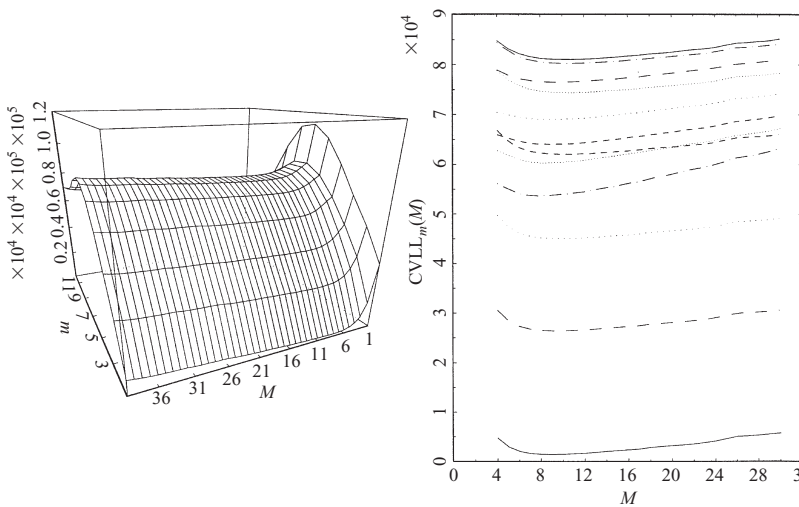

FIGURE 5. $\mathrm{CVLL}_m(M, 0)$ for Model 4. Cross-validated likelihood, AR(2), $\alpha_1 = 1.372$, $\alpha_2 = 0.677$.

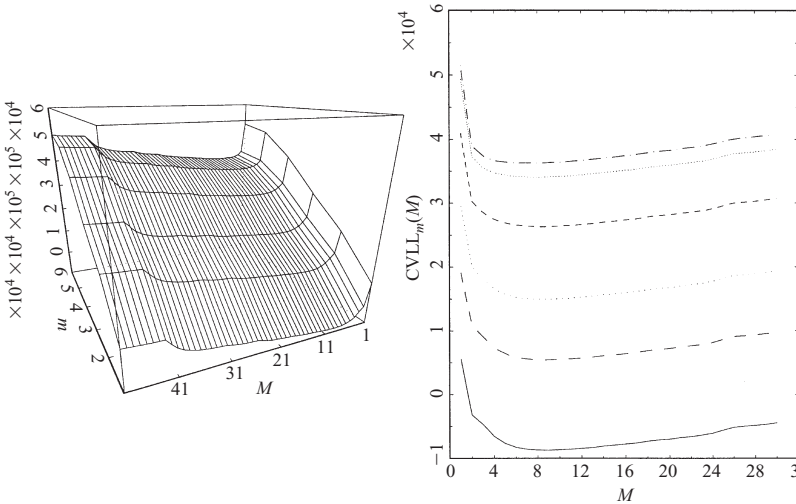

FIGURE 6. $\mathrm{CVLL}_m(M, 0)$ for Model 5. Cross-validated likelihood, AR(5), $\alpha_1 = 0.9$, $\alpha_2 = \ 0.4$, $\alpha_3 = 0.3$, $\alpha_4 = \ 0.5$, $\alpha_5 = 0.3$.

practical applications a simple possibility is a selection criterion depending only on a fixed m, possibly a function of N. For any sample size this would imply the use of about N/m Fourier frequencies in CVLL_m. In Table II VI we tried the following choices for sample sizes $N = 120$, 256, 480 and Models 1 5:

Method	GLOBAL	(a)	(b)	(c)	(d)	(e)	LOCAL
m	1	$N^{0.05}$	$N^{0.1}$	$N^{0.2}$	$N^{0.3}$	$N^{0.4}$	$N/6$

These cover all reasonable values for the three sample sizes in the light of the behaviour of CVLL_m. GLOBAL is the same as BB's global procedure, with the use of a kernel W, not necessarily the uniform one. Procedures (a) and (b) satisfy the conditions of Proposition 1 since $M = cN^{0.2}$, but options (d) and (e) do not, (c) being on the borderline. For comparison purposes the LOCAL method selects a quite local choice for all sample sizes. We employed 1000 replications.

Following Bühlmann (1995), for each case we report the bias, standard deviation and two relative mean squared errors for the simulated $\hat{M}$ Rmse, relative to the asymptotic $M^*(0)$, and Rmse†, relative to the empirical $\hat{M}^*(0)$,

$$\mathrm{Rmse} = \frac{E^\dagger[\{\hat{M}(0) \quad M^*(0)\}^2]}{M^*(0)^2} \qquad \mathrm{Rmse}^\dagger = \frac{E^\dagger[\{\hat{M}(0) \quad \hat{M}^*(0)\}^2]}{\hat{M}^*(0)^2}$$

where $E^\dagger$ denotes expectation with respect to the simulated samples, and similarly for $\hat{f}_{\hat{M}}(0)$ (with bias and Rmse relative to the true value $f(0)$).

TABLE II

MODEL 1: AR(3), α [0.60 0.60, 0.30]

Method	bias	sd	Rmse	Rmse†	bias	sd	Rmse	(sd)	MSE ratio	MSE† ratio
N 120		$M^*(0)$: 6.1516, $\hat{M}(0)$: 7.5					$f(0)$: 0.3248			
GLOBAL	0.3584	2.1217	0.1224	0.0975	0.0177	0.1263	0.1542	(0.0641)	1.2357	1.2438
(a)	0.1579	2.2988	0.1403	0.1191	0.0199	0.1235	0.1482	(0.0627)	1.1880	1.1958
(b)	0.2734	2.3369	0.1463	0.1438	0.0204	0.1191	0.1385	(0.0606)	1.1100	1.1173
(c)	2.8218	1.8509	0.3009	0.3701	0.0334	0.0872	0.0826	(0.0449)	0.6618	0.6661
(d)	2.7554	1.8015	0.2864	0.3571	0.0366	0.0896	0.0887	(0.0459)	0.7110	0.7156
(e)	2.8278	1.8884	0.3055	0.3734	0.0406	0.0930	0.0977	(0.0476)	0.7827	0.7878
LOCAL	2.5814	2.2423	0.3090	0.3639	0.0364	0.1028	0.1127	(0.0516)	0.0835	0.9094
N 256		$M^*(0)$: 7.1582, $\hat{M}^*(0)$: 1.8286					$f(0)$: 0.3248			
GLOBAL	1.2755	2.4679	0.1506	13.9681	0.0151	0.0938	0.0856	(0.0482)	1.3106	0.3310
(a)	1.0319	2.5011	0.1429	13.12307	0.0154	0.0933	0.0848	(0.0480)	1.2985	1.3187
(b)	0.0500	2.5712	0.1291	9.9707	0.0180	0.0877	0.0760	(0.0451)	1.1637	1.1818
(c)	3.6088	2.1191	0.3418	2.0748	0.0313	0.0604	0.0438	(0.0312)	0.6712	0.6816
(d)	3.9257	1.7523	0.3607	1.3975	0.0384	0.0598	0.0479	(0.0309)	0.7336	0.7450
(e)	4.0488	1.8927	0.3898	1.4523	0.0383	0.0625	0.0510	(0.0322)	0.7805	0.7927
LOCAL	3.7970	1.8768	0.3501	1.6301	0.0357	0.0612	0.0476	(0.0316)	0.7290	0.7403
N 480		$M^*(0)$: 8.1171, $\hat{M}^*(0)$: 1.875					$f(0)$: 0.3248			
GLOBAL	1.8742	2.6173	0.1573	20.6865	0.0101	0.0755	0.0551	(0.0390)	1.3783	1.4137
(a)	1.7283	2.7263	0.1581	20.1843	0.0102	0.0751	0.0544	(0.0388)	1.3630	1.3980
(b)	0.4309	2.7659	0.1189	14.8423	0.0126	0.0704	0.0485	(0.0364)	1.2140	1.2452
(c)	4.2859	2.3343	0.3615	2.6384	0.0263	0.0440	0.0249	(0.0228)	0.6238	0.6398
(d)	5.0850	1.5266	0.4278	1.0437	0.0349	0.0423	0.0285	(0.0219)	0.7136	0.7319
(e)	5.2591	1.4893	0.4534	0.9058	0.0333	0.0429	0.0280	(0.0222)	0.6998	0.7178
LOCAL	4.8497	1.7325	0.4025	1.4053	0.0302	0.0416	0.0251	(0.0216)	0.6282	0.6443

TABLE III

MODEL 2: AR(2), α [0.60 0.90]

Method	bias	sd	Rmse	Rmse†	bias	sd	Rmse	(sd)	MSE ratio	MSE† ratio
N 120		$M^*(0)$: 5.1285, $\hat{M}(0)$: 7.5					$f(0)$: 0.0942			
GLOBAL	4.9919	1.3243	0.7102	0.2642	0.0144	0.0580	0.4027	(0.3572)	1.4147	1.4192
(a)	5.0614	1.3056	0.7275	0.2724	0.0144	0.0582	0.4060	(0.3587)	1.4263	1.4307
(b)	5.0557	1.3236	0.7272	0.2724	0.0146	0.0582	0.4063	(0.3472)	1.4275	1.4320
(c)	4.0417	1.7333	0.5149	0.1802	0.0157	0.0563	0.3858	(0.3295)	1.3555	1.3598
(d)	2.2451	1.8955	0.2299	0.0774	0.0180	0.0534	0.3583	(0.3190)	1.2589	1.2628
(e)	0.8858	2.2370	0.1541	0.0932	0.0179	0.0517	0.3378	(0.3185)	1.1867	1.1904
LOCAL	0.1706	2.7554	0.2029	0.1606	0.0187	0.0516	0.3394	(0.0134)	1.1924	1.1961
N 256		$M^*(0)$: 7.1313, $\hat{M}^*(0)$: 6.4810					$f(0)$: 0.0942			
GLOBAL	8.4895	2.3973	1.5302	4.7419	0.0061	0.0392	0.1779	(0.2420)	1.7397	1.7562
(a)	8.5908	2.3850	1.5631	4.8261	0.0061	0.0394	0.1797	(0.2432)	1.7573	1.7739
(b)	8.3915	2.4307	1.5008	4.6654	0.0063	0.0395	0.1802	(0.2435)	1.7627	1.7795
(c)	4.9506	2.5475	0.6095	2.2657	0.0072	0.0358	0.1504	(0.2209)	1.4708	1.4848
(d)	1.3965	2.6261	0.1740	0.7736	0.0109	0.0332	0.1378	(0.2048)	1.3476	1.3604
(e)	0.0054	2.7176	0.1452	0.4762	0.0079	0.0302	0.1102	(0.1866)	1.0782	1.0884
LOCAL	0.6787	2.8666	0.1706	0.4131	0.0143	0.0341	0.1544	(0.2105)	1.5098	1.5242
N 480		$M^*(0)$: 8.0866, $\hat{M}^*(0)$: 5.0					$f(0)$: 0.0942			
GLOBAL	12.6852	3.6600	2.6656	10.4858	0.0029	0.0332	0.1249	(0.2046)	2.3300	2.3743
(a)	12.9979	3.7590	2.7996	10.9137	0.0029	0.0335	0.1276	(0.2068)	2.3806	2.4259
(b)	12.2446	3.7141	2.5037	9.9536	0.0029	0.0330	0.1241	(0.2038)	2.3140	2.3580
(c)	4.6116	2.8291	0.4476	2.6907	0.0045	0.0268	0.0830	(0.1651)	1.5481	1.5775
(d)	1.0187	2.8068	0.1363	0.9893	0.0070	0.0242	0.0718	(0.1496)	1.3400	1.3654
(e)	0.2205	3.1503	0.1525	0.7256	0.0049	0.0229	0.0618	(0.1413)	1.1529	1.1748
LOCAL	0.9393	3.3423	0.1843	0.6313	0.0128	0.0290	0.1135	(0.1791)	2.1159	2.1561

TABLE IV

MODEL 3: AR(1), α [0.8]

Method	bias	sd	Rmse	Rmse†	bias	sd	Rmse	(sd)	MSE ratio	MSE† ratio
N 120		$M^*(0)$: 17.4803, $\hat{M}(0)$: 10.0					$f(0)$: 3.9789			
GLOBAL	10.6166	2.0697	0.3829	0.1412	1.4311	1.4412	0.2609	(0.0126)	0.7774	0.7793
(a)	10.2789	2.1244	0.3606	0.1235	1.3654	1.4838	0.2568	(0.0134)	0.7651	0.7671
(b)	10.0005	2.1871	0.3430	0.1114	1.3178	1.5076	0.2533	(0.0134)	0.7545	0.7564
(c)	9.5028	2.2363	0.3119	0.0909	1.2429	1.5800	0.2553	(0.0150)	0.7605	0.7624
(d)	9.0186	2.2093	0.2822	0.0725	1.1664	1.6004	0.2477	(0.0148)	0.7380	0.7398
(e)	8.8514	2.2607	0.2731	0.0699	1.1201	1.6132	0.2436	(0.0157)	0.7258	0.7276
LOCAL	9.4723	3.1568	0.3263	0.1393	1.1735	1.7293	0.2759	(0.0173)	0.8219	0.8240
N 256		$M^*(0)$: 20.3404, $\hat{M}^*(0)$: 11.1304					$f(0)$: 3.9789			
GLOBAL	11.4808	2.4632	0.3332	0.0737	1.1175	1.1254	0.1589	(0.0053)	0.8988	0.9035
(a)	11.0161	2.6134	0.3098	0.0729	1.0485	1.1778	0.1571	(0.0058)	0.8885	0.8931
(b)	10.5917	2.7507	0.2894	0.0763	0.9975	1.1979	0.1535	(0.0059)	0.8683	0.8728
(c)	9.9272	2.9508	0.2592	0.0888	0.9171	1.2465	0.1513	(0.0068)	0.8557	0.8601
(d)	9.3162	3.0514	0.2323	0.1036	0.8459	1.2867	0.1498	(0.0075)	0.8472	0.8516
(e)	8.7241	3.3655	0.2113	0.1394	0.7494	1.3510	0.1508	(0.0083)	0.8529	0.8573
LOCAL	10.2942	4.8292	0.3125	0.2332	1.0102	1.4883	0.2044	(0.0101)	1.1562	1.1621
N 480		$M^*(0)$: 23.0653, $\hat{M}^*(0)$: 10.0					$f(0)$: 3.9789			
GLOBAL	12.7966	2.6874	0.3214	0.0729	0.9144	0.9351	0.1080	(0.0027)	0.8936	0.8998
(a)	12.3139	2.8446	0.3002	0.0866	0.8623	0.9629	0.1055	(0.0031)	0.8729	0.8789
(b)	11.8107	2.9736	0.2788	0.1042	0.8132	0.9883	0.1035	(0.0035)	0.8557	0.8617
(c)	10.8172	3.2750	0.2401	0.1578	0.7182	1.0589	0.1034	(0.0043)	0.8553	0.8612
(d)	10.0015	3.5333	0.2115	0.2187	0.6310	1.1233	0.1049	(0.0054)	0.8672	0.8733
(e)	9.3773	3.9248	0.1942	0.2901	0.5397	1.2028	0.1098	(0.0075)	0.9080	0.9143
LOCAL	11.9401	6.2338	0.3410	0.4013	0.9859	1.3978	0.1848	(0.0095)	1.5286	1.5393

TABLE V

MODEL 4: AR(2), α [1.37, 0.68]

Method	bias	sd	Rmse	Rmse†	bias	sd	Rmse	(sd)	MSE ratio	MSE† ratio
N 120		$M^*(0)$: 9.5192, $\hat{M}(0)$: 7.5					$f(0)$: 1.7109			
GLOBAL	1.4440	2.0943	0.0717	0.0835	0.3530	0.8505	0.3091	(0.0109)	1.1074	1.1111
(a)	1.1884	2.1111	0.0650	0.0909	0.3302	0.8543	0.3058	(0.0114)	1.0956	1.0992
(b)	1.0308	2.1656	0.0638	0.1000	0.3237	0.8588	0.3071	(0.0114)	1.1001	1.1038
(c)	1.0011	2.4197	0.0760	0.1218	0.3040	0.8474	0.2955	(0.0086)	1.0586	1.0621
(d)	1.6060	2.8498	0.1186	0.1471	0.3724	0.8277	0.3003	(0.0079)	1.0758	1.0794
(e)	6.4426	1.6444	0.4900	0.3991	0.8878	0.7203	0.4765	(0.0038)	1.7071	1.7128
LOCAL	6.8877	1.3974	0.5474	0.4596	0.6785	0.6920	0.3425	(0.0071)	1.2268	1.2309
N 256		$M^*(0)$: 11.0767, $\hat{M}^*(0)$: 8.9825					$f(0)$: 1.7109			
GLOBAL	0.2179	2.4914	0.0512	5.0210	0.1639	0.5955	0.1391	(0.0074)	1.0957	1.1000
(a)	0.1269	2.6163	0.0562	5.5056	0.1551	0.6038	0.1417	(0.0074)	1.1164	1.1207
(b)	0.3525	2.7296	0.0620	5.8491	0.1465	0.6048	0.1412	(0.0075)	1.1121	1.1164
(c)	0.0168	3.0867	0.0780	5.5884	0.1550	0.6107	0.1447	(0.0075)	1.1402	1.1446
(d)	6.3663	3.2311	0.4172	0.9901	0.7953	0.5603	0.3451	(0.0077)	2.7184	2.7289
(e)	8.5984	1.0056	0.6134	0.1695	0.7132	0.4724	0.2668	(0.0076)	2.1020	2.1101
LOCAL	8.4884	1.3887	0.6055	0.2282	0.5717	0.4763	0.2019	(0.0072)	1.5905	1.5966
N 480		$M^*(0)$: 12.5607, $\hat{M}^*(0)$: 3.47					$f(0)$: 1.7109			
GLOBAL	0.2129	2.7044	0.0468	7.7550	0.1249	0.4623	0.0836	(0.0073)	1.1181	1.1304
(a)	0.8005	2.9814	0.0607	8.8200	0.1144	0.4706	0.0855	(0.0074)	1.1436	1.1562
(b)	1.2263	3.0934	0.0705	9.5891	0.1048	0.4832	0.0891	(0.0074)	1.1921	1.2052
(c)	0.8642	3.3531	0.0763	9.1201	0.1149	0.4900	0.0923	(0.0075)	1.2351	1.2486
(d)	9.7083	1.4355	0.6130	0.2056	0.8746	0.3612	0.3264	(0.0064)	4.3659	4.4138
(e)	10.3231	0.6459	0.6810	0.1663	0.6596	0.3589	0.2056	(0.0064)	2.7498	2.7800
LOCAL	10.0340	1.3202	0.6519	0.2229	0.5770	0.3824	0.1747	(0.0064)	2.3362	2.3618

TABLE VI

MODEL 5: AR(5), α [0.90, 0.40, 0.30, 0.50, 0.30]

Method	bias	sd	Rmse	Rmse†	bias	sd	Rmse	(sd)	MSE ratio	MSE† ratio
N 120		$M^*(0)$: 10.2862, $\hat{M}(0)$: 7.5					$f(0)$: 0.9947			
GLOBAL	3.1082	2.4496	0.1480	0.1085	0.1328	0.3708	0.15687	(0.0157)	0.7418	0.7446
(a)	3.2427	2.3976	0.1537	0.1059	0.1335	0.3588	0.1481	(0.0157)	0.7007	0.7034
(b)	3.5374	2.3912	0.1723	0.1117	0.1463	0.3372	0.1366	(0.0154)	0.6462	0.6486
(c)	4.3022	2.2374	0.2222	0.1299	0.1631	0.3219	0.1316	(0.0142)	0.6228	0.6252
(d)	5.5814	2.1982	0.3401	0.2248	0.1255	0.3125	0.1146	(0.0134)	0.5424	0.5444
(e)	6.5136	1.9262	0.4361	0.3130	0.0979	0.3082	0.1057	(0.0132)	0.5001	0.5020
LOCAL	6.6726	2.2301	0.4678	0.3569	0.1334	0.3445	0.1379	(0.0113)	0.6528	0.6552
N 256		$M^*(0)$: 11.9693, $\hat{M}^*(0)$: 2.9425					$f(0)$: 0.9947			
GLOBAL	1.4509	2.6378	0.0633	0.1217	0.0858	0.3052	0.1015	(0.0152)	0.9496	0.9582
(a)	1.9759	2.8375	0.0834	0.1173	0.0910	0.2925	0.0948	(0.0147)	0.8865	0.8946
(b)	2.6567	3.2062	0.1210	0.1324	0.1031	0.2777	0.0887	(0.0140)	0.8295	0.8370
(c)	5.0893	2.9881	0.2431	0.1641	0.1474	0.2251	0.0732	(0.0116)	0.6843	0.6904
(d)	7.7401	2.0833	0.4485	0.3297	0.1166	0.2169	0.0613	(0.0103)	0.5731	0.5783
(e)	7.9489	2.2450	0.4762	0.3638	0.1107	0.2382	0.0697	(0.0110)	0.6522	0.6581
LOCAL	8.3001	2.4839	0.5239	0.4229	0.1616	0.2419	0.0855	(0.0110)	0.7997	0.8069
N 480		$M^*(0)$: 13.5727, $\hat{M}^*(0)$: 8.889					$f(0)$: 0.9947			
GLOBAL	1.1108	2.7918	0.0490	0.2602	0.0557	0.2492	0.0659	(0.0130)	0.9673	0.9809
(a)	1.5008	2.9584	0.0597	0.2390	0.0575	0.2481	0.0655	(0.0129)	0.9620	0.9756
(b)	2.1549	3.2538	0.0827	0.2149	0.0657	0.2418	0.0634	(0.0126)	0.9314	0.9446
(c)	6.3322	3.0869	0.2694	0.1550	0.1295	0.1715	0.0467	(0.0091)	0.6854	0.6951
(d)	9.5034	1.8473	0.5088	0.3372	0.1121	0.1473	0.0346	(0.0078)	0.5082	0.5154
(e)	9.5768	2.0590	0.5209	0.3567	0.1166	0.1626	0.0404	(0.0085)	0.5936	0.6020
LOCAL	10.0044	2.6047	0.5801	0.4441	0.1702	0.1802	0.0621	(0.0088)	0.9115	0.9244

We also give the ratio of the mse of $\hat{f}_{\hat{M}}(0)$ with respect to the mse of $\hat{f}_{M^*(0)}(0)$ calculated using the asymptotically optimal choice $M^*(0)$ (the MSE ratio) and the same ratio with respect to $\hat{f}_{\hat{M}^*(0)}(0)$, using the finite sample minimizer $\hat{M}^*(0)$ calculated across simulations (the MSE† ratio). In both cases a value less than one would indicate a better performance than that obtained using the (usually unknown) optimal values for the bandwidth. Since we are interested finally in estimating $f(0)$, our remarks concentrate more on the diagnostics for $\hat{f}_{\hat{M}}(0)$ than on those for $\hat{M}$.

The conclusions for $\hat{M}$ can be substantialy different depending on which criteria we use, Rmse or Rmse†. While for Models 2 and 5 both give similar ranking of procedures, because the asymptotic value $M^*(0)$ is usually not very far from the finite sample minimizer $\hat{M}^*(0)$, for Models 1, 3 and 4 (and the two larger sample sizes) the best results under the two criteria are obtained for values of m in very different ranges. Thus for Models 1 and 4 the largest ms are preferred by Rmse† but the smallest by Rmse, whereas the opposite holds for Model 3.

Attending to the spectral estimation, the bias and standard deviation show the customary trade off with respect to the choice of the smoothing parameter, but with reversed pattern in most cases: small values of m lead to small biases and moderate or large values of m minimize the standard deviation. Surprisingly, GLOBAL is not the more stable method and LOCAL choice leads to large biases in $\hat{f}_{\hat{M}}(0)$. Model 3 is the exception, probably because it has the most complicated spectral density at the origin but is smooth otherwise.

The smallest values of Rmse for $\hat{f}_{\hat{M}}(0)$ can be achieved or approximated closely by the best choices of m for estimating $\hat{M}^*(0)$ (focusing on Rmse†) for all models and sample sizes. Here the exception is Model 4 and $N = 256$ and 480.

For Models 1, 2 and 5, and for the three sample sizes, almost all local choices (rows (a) to LOCAL) perform better than BB's method (GLOBAL), so they appear to adapt to the local properties of the function being estimated (the exception being (a) for Model 2). The best procedures seem to be (c) and (d) for Models 1 and 2, and (d) and (e) for Model 5, though in some cases LOCAL outperforms them by a narrow margin.

On the other hand, in the case of Model 3 all choices of m but LOCAL lead to similar results for the estimation of $f(0)$, GLOBAL never being too far from the smallest Rmse and (d) giving the best performance or being very close to it. Here LOCAL seems to work substantially worse than any other choice of m. For Model 4 BB's GLOBAL method seems to be the best alternative, while (a) (c) give very similar results to it and procedures (d) and (e) break down.

For Model 2 the small choices of $m > 1$ (rows (a) and (b)) do not improve over the GLOBAL procedure, since for those ms we are still considering the peak of the spectrum in CVLL_m, while large values of m induce some improvements in the behaviour of $\hat{f}_{\hat{M}}(0)$. However, for Model 4, which has a similar spectrum but with the peak closer to the origin, the improvement is only observed for $N = 120$.

In terms of MSE ratios, the differences are not significant looking at the MSE ratio or MSE† ratio columns. For Models 1, 3 and 5 there is room to improve over (unfeasible) optimal choices of M with local cross validation: methods (c) (e) for fixing m produce substantial reductions of the MSE of $\hat{f}_{\hat{M}}$, while for Model 2 there can be some efficiency losses, especially accentuated for Model 4, if too large a value of m is selected.

We now compare with the Bühlmann (1996) results for sample sizes $N = 120$ and 480 (cf. his Table I), although he uses a different class of non parametric estimates (lag window or continuously weighted periodogram estimates) with different weight functions. For our Model 3 (Bühlmann's Model 1) and $N = 120$ all our methods work slightly better (including BB's global choice but not LOCAL), both in terms of Rmse and MSE ratio of $\hat{f}_{\hat{M}}(0)$. For $N = 480$ the Rmse results are very similar for the two procedures, but the MSE ratios are always better for the local cross validation (except with the LOCAL choice).

In the case of Model 4 and $N = 120$ and 480, the methods GLOBAL to (c) always worked better in terms of MSE ratio, but for the smaller sample size they gave larger Rmse than Bühlmann's best estimates for his Model 2. With Model 5, the cross validated methods (c) (e) always worked better than any of Bühlmann's alternatives (for his Model 3) in terms of MSE ratio and also in terms of Rmse for $N = 480$. For $N = 120$ only the choice (e) of m can approximate the Rmse results of Bühlmann's (1996) adaptive method.

In general, it seems that the asymptotic result for the optimal choice of M for a single frequency is not especially accurate for periodogram based estimates, so local cross validation can improve even with respect to the knowledge of it. Also, cross validation never behaves much worse than the iterative plug in procedure, both having a comparable performance if a sensible choice of m is made. From a practical point of view the recommended choice of m for the construction of the local cross validation criterion is (c), $m = N^{0.2}$, the same as the asymptotic optimal rate for M. This choice works uniformly well for all models and sample sizes tried, and while (d) and (e) may be preferable in some cases, they are very inefficient in other situations.

4.2. *Estimation of the whole spectrum*

Finally we tried local cross validation for estimation of the optimal bandwidth at all Fourier frequencies λ_j, $j = 0, \ldots, N \quad 1$, for the same models and sample sizes as before. Here the computation costs are much greater, so we only implemented 200, 100 and 50 simulations for sample sizes 120, 256 and 480, respectively. Given the conclusions of the previous section, we tried the GLOBAL procedure of BB ($m = 1$) and the three initial choices of $m = 2$, 3, 4, which adapt to the roughness of f at each point (and correspond approximately to options (b), (c) and (d) for $N = 256$). We give in Table VII the sample mean of the relative IMSE estimated with the simulations

TABLE VII

$\hat{f}(\lambda)$, $\lambda \in [\ -\pi, \pi]$

		$N = 120$		$N = 256$		$N = 480$	
Model	m	IMSE	sd	IMSE	sd	IMSE	sd
1	1	0.13646	(0.0983)	0.06545	(0.0322)	0.03805	(0.0153)
	2	0.15758	(0.1187)	0.07140	(0.0412)	0.03659	(0.0157)
	3	0.13433	(0.0971)	0.06390	(0.0324)	0.03528	(0.0148)
	4	0.13528	(0.0938)	0.06436	(0.0313)	0.03503	(0.0154)
2	1	0.74030	(0.7346)	0.25241	(0.1734)	0.11225	(0.0446)
	2	0.60013	(0.5815)	0.23447	(0.1472)	0.11188	(0.0415)
	3	0.66356	(0.6146)	0.23208	(0.1565)	0.09898	(0.0394)
	4	0.66240	(0.6101)	0.22940	(0.1591)	0.09612	(0.0382)
3	1	0.20289	(0.1684)	0.09384	(0.0846)	0.04196	(0.0224)
	2	0.14133	(0.0935)	0.05981	(0.0330)	0.03116	(0.0140)
	3	0.12857	(0.0831)	0.05855	(0.0320)	0.03082	(0.0140)
	4	0.13006	(0.0840)	0.06068	(0.0352)	0.03085	(0.0141)
4	1	0.47824	(0.7407)	0.14747	(0.2198)	0.06533	(0.0534)
	2	0.31678	(0.3897)	0.13110	(0.0990)	0.06334	(0.0489)
	3	0.29050	(0.3906)	0.12113	(0.0919)	0.05858	(0.0462)
	4	0.28964	(0.3977)	0.14747	(0.2198)	0.05835	(0.0472)
5	1	0.13666	(0.0951)	0.09814	(0.0857)	0.04842	(0.0187)
	2	0.15576	(0.1115)	0.10457	(0.0517)	0.05267	(0.0219)
	3	0.13401	(0.0933)	0.09938	(0.0494)	0.05180	(0.0194)
	4	0.13444	(0.0917)	0.09809	(0.0480)	0.04991	(0.0198)

$$\frac{1}{N}\sum_{j\ 0}^{N-1}\left\{\frac{\hat{f}_{\hat{M}}(\lambda_j)\quad f(\lambda_j)}{f(\lambda_j)}\right\}^2$$

and its sample standard deviation.

Almost uniformly the local cross validation procedures beat the global one ($m = 1$), in some situations by a wide margin; in the worst cases (Models 1 and 5) they perform roughly in the same way. The improvement with repsect to the global choice is generally greater the smaller the sample size and, against intuition, in many cases the more local choices also lead to less variable procedures. There are no significative dissimilarities for the three different values of $m > 1$, but $m = 3$ and $m = 4$ seem to do slightly better, in agreement with the estimation at a single point.

Comparing with Bühlmann's Table II, for $N = 120$ and 480 all the local cross validation IMSEs (and in many cases also BB's global choice) are equal to or lower than that of the best plug in alternative, though they have apparently greater variability, at least in our simulations.

5. FINAL REMARKS

In this paper we justify a local bandwidth choice procedure for non parametric spectral estimation and show its performance in finite sample sizes. We assumed Guassianity throughout, but this does not seem essential, except perhaps in the proof for the supremum of the periodogram in Lemma 4. We conjecture that this condition can be avoided using Robinson's (1991) techniques and assuming summability conditions on higher order cumulants as in Brillinger (1975), except for the second order ones (autocovariances), imposing here only local conditions on the (second order) spectral density.

A multivariate version of the method will be very useful in practical work, but if we want to stress the specific characteristics of each univariate time series it could be better to apply the method to each of them separately or to a fixed linear combination of the series, as in Newey and West (1994).

Further investigation seems necessary on automatic selection of m and on the design of (possibly iterative) algorithms that, linking m and M, reduce the variability inherent to bandwidth choice procedures. Then additional finite sample evidence should be investigated for other models and distributions.

APPENDIX

PROOF OF LEMMA 1. An equivalent lemma is evidently valid for more general choices of M, but we are especially interested in this particular case. We can take an $\epsilon > 0$ as small as we want, in such a way that in the interval $I_\nu \quad [\nu \quad \epsilon, \nu + \epsilon]$ the conditions of Assumption 2 are satisfied. Then for m big enough we have that $|\nu \quad \lambda_j| \leqslant cm^{-1}$ implies $\lambda_j \in I_\nu$. Therefore when $\nu > 0$ we have that, for N big enough, $0 < \lambda_j \sim \nu$, so $(\lambda_j)^{-1} \quad \mathrm{O}(1)$, where $a \sim b$ means $a/b \to 1$ as $N \to \infty$. We study first the bias and the variance.

Bias. Similarly to Theorem 5.6.1 of Brillinger (1975, p. 147) and using now Lemma 2 with $\alpha \quad 1$, we get

$$E\{\hat{f}_M(\lambda_j)\} \quad \int_{-\pi}^{\pi} K(\lambda) f\left(\lambda_j \quad \frac{\beta}{M}\right) d\beta + \mathrm{O}\left(\frac{M}{N}\right)$$

$$f(\lambda_j) + \frac{w_2}{2} f^{(2)}(\lambda_j) M^{-2} + \mathrm{O}\left(\frac{M}{N} + M^{-3}\right).$$

The bounded variation condition on K and the differentiability of f are used to approximate the discrete average of K and f by an integral with error $\mathrm{O}(M/N)$, since by Assumption 2 and for M big enough we are only averaging inside I_ν, thanks to the compact support of K.

Variance. First, it is more convenient to write the spectral estimate using only N frequencies in this way:

$$\hat{f}_M(\lambda_j) \quad \frac{\sigma_M^{-1}}{M} \sum_{k \quad 1}^{N-1} K_M(\lambda_k \quad \lambda_j) I(\lambda_k)$$

where $K_M(\lambda) \quad \sum_j MK(M[\lambda + 2\pi j])$ is the periodic extension of $MK(M\lambda)$. Then we have

$$\operatorname{var}\{\hat{f}_M(\lambda_j)\} \quad \frac{\sigma_M^{-2}}{M^2} \sum_k K_M(\lambda_k \quad \lambda_j)^2 \operatorname{var}\{I(\lambda_k)\} \tag{10}$$

$$+ \frac{\sigma_M^{-2}}{M^2} \sum_k \sum_{i/k} K_M(\lambda_k \quad \lambda_j) K_M(\lambda_i \quad \lambda_j) \operatorname{cov}\{I(\lambda_k),\, I(\lambda_i)\}. \tag{11}$$

Then, from Lemma 3 we get $\operatorname{var}\{I(\lambda_k)\} \quad f(\lambda_k)^2 + \mathrm{O}(N^{-1} \log N)$ and, for $k \;/\; i$,

$$\operatorname{cov}\{I(\lambda_k),\, I(\lambda_i)\} \quad \begin{cases} f(\lambda_k)^2 + \mathrm{O}(N^{-1} \log N) & \text{if } k \quad N \quad i \\ \mathrm{O}(N^{-1} \log N) & \text{otherwise.} \end{cases}$$

Also we have that $\sigma_M \quad N/(2\pi M) + \mathrm{O}(1)$. Then (10) is

$$\frac{(2\pi)^2}{N^2} \sum_k K_M(\lambda_k \quad \lambda_j)^2 f(\lambda_k)^2 + \mathrm{O}\left(\frac{M}{N^2} \log N\right)$$

$$\frac{2\pi M}{N} \int_{-\pi}^{\pi} MK(M\lambda)^2 f(\lambda_j + \lambda)^2 \, d\lambda + \mathrm{O}\left(\frac{M^2}{N^2}\right)$$

$$\frac{2\pi M}{N} f(\lambda_j)^2 \int_{-\pi}^{\pi} K(\lambda)^2 \, d\lambda + \mathrm{O}\left\{\frac{M}{N}\left(\frac{M}{N} + M^{-2}\right)\right\}.$$

In (11) we only have to consider the situation where $k \quad N \quad i$, since for the other frequencies we have a bound of $\mathrm{O}(N^{-1} \log N)$ for the covariance from Lemma 3. Then, if $\nu \quad 0$ and $\lambda_j \quad 0$, (11) is similar to (10). In general, if $\nu \quad 0$ and $|\lambda_j| \leqslant 2\pi/M$ then the two kernels in (11) overlap in some interval for all M. Taking into account only the frequencies $i \quad N \quad k$, for which the leading term of the covariance is also $f(\lambda_k)$, we have that (11) is equal to, using the periodicity of K_M,

$$\frac{\sigma_M^{-2}}{M^2} \sum_k K_M(\lambda_k \quad \lambda_j) K_M(\lambda_k + \lambda_j) \{f(\lambda_k)^2 + \mathrm{O}(N^{-1} \log N)\}$$

$$\frac{2\pi}{N} \int_{-\pi}^{\pi} K_M(\lambda) K_M(\lambda + 2\lambda_j) f(\lambda \quad \lambda_j)^2 \, d\lambda + \mathrm{O}\left\{\left(\frac{M}{N}\right)^2 + \left(\frac{M}{N^2}\right) \log N\right\}$$

$$\delta_M(j) f(\lambda_j)^2 \frac{2\pi M}{N} \int_{-\pi}^{\pi} K(\lambda)^2 \, d\lambda + \mathrm{O}\left\{MN^{-1}\left(\frac{M}{N} + M^{-2}\right)\right\} \tag{12}$$

for $0 < \delta_M(j) \leqslant 1$. If $|\lambda_j| > 2\pi/M$ then the two kernels in (12) do not overlap at all and the covariance terms do not contribute to the leading term in the variance of $\hat{f}_M$, and the lemma follows. ∎

LEMMA 2. *Under Assumption 1, if f satisfies a uniform Lipschitz condition of order* $0 < \alpha \leqslant 1$ *in an interval around a fixed frequency ν, then for Fourier frequencies such that* $\sup_{\lambda l} |\nu \quad \lambda_l| \leqslant cm^{-1}$, $l \in \{j, k\}$, *for some positive sequence m such that* $1/m + m/n \to 0$, *uniformly in j and k* ($j, k \;/\; 0$),

$$E\{d_x(\lambda_j)\overline{d_x(\lambda_k)}\} \quad \delta_{jk} 2\pi N f(\lambda_j) \quad \mathrm{O}(N^{1-\alpha} \log N)$$

where $d_x(\lambda_j)$ *is the discrete Fourier transform of the series* X_t,

$$d_x(\lambda_j) \quad \sum_{t \quad 1}^{N} X_t \exp(\quad i\lambda_j t).$$

PROOF. This lemma is a restatement of, for example, the Lemma on p. 835 of Hannan and Nicholls (1977), assuming only local conditions on f. As in the proof of Lemma 1 we can fix one $\epsilon > 0$ such that, if $I_\nu \quad [\nu \quad \epsilon, \nu + \epsilon]$, $\lambda_j, \lambda_k \in I_\nu$ for N big enough. Defining the Dirichlet kernel $\varphi_N(\lambda)$

$$\varphi_N(\lambda) \quad \sum_{j \quad 1}^{N} \exp(i\lambda j)$$

we have that for $j \ / \ k$, mod(N),

$$\int_{-\pi}^{\pi} \varphi_N(\lambda_j \quad \lambda)\varphi_N(\lambda \quad \lambda_k)\, d\lambda \quad 0.$$

Then, if $j \ / \ k$, mod(N),

$$E\{d_x(\lambda_j)\overline{d_x(\lambda_k)}\} \quad \int_{-\pi}^{\pi} \varphi_N(\lambda_j \quad \lambda)\varphi_N(\lambda \quad \lambda_k)\{f(\lambda) \quad f(\lambda_j)\}\, d\lambda. \tag{13}$$

Now we divide the range of integration in (13) into the following intervals. First,

$$\left|\int_{\lambda_j - N^{\ -1}}^{\lambda_j + N^{\ -1}} \varphi_N(\lambda; \quad \lambda)\varphi_N(\lambda \quad \lambda_k)\{f(\lambda) \quad f(\lambda_j)\}\, d\lambda\right| \leqslant cN\int_{\lambda_j - N^{\ -1}}^{\lambda_j + N^{\ -1}} |\lambda \quad \lambda_j|^{\alpha - 1}\, d\lambda$$

$$\leqslant cN^{1-\alpha}$$

using $\sup_{\lambda \in I_\nu} |f(\lambda) \quad f(\lambda_j)| \leqslant c|\lambda \quad \lambda_j|^\alpha$ in the interval considered, and

$$|\varphi_N(\lambda)| \leqslant \min(2|\lambda|^{-1}, N).$$

Next,

$$\left|\int_{\lambda_k - N^{\ -1}}^{\lambda_k + N^{\ -1}} \varphi_N(\lambda_j \quad \lambda)\varphi_N(\lambda \quad \lambda_k)\{f(\lambda) \quad f(\lambda_j)\}\, d\lambda\right|$$

$$\leqslant cN^{-1} \sup_{|\lambda - \lambda_k| \leqslant N^{\ -1}} |\varphi(\lambda \quad \lambda_k)| \sup_{|\lambda - \lambda_k| \leqslant N^{\ -1}} |\lambda \quad \lambda_j|^{\alpha - 1} \leqslant cN^{1-\alpha}$$

since the range of integration was of order N^{-1}. Define the set $I_\nu(k, j)$ as the interval I_ν except for previous two neighbourhoods of radius N^{-1} around λ_k and λ_j. Then

$$\left|\int_{I_\nu(k,j)} \varphi_N(\lambda_j \quad \lambda)\varphi_N(\lambda \quad \lambda_k)\{f(\lambda) \quad f(\lambda_j)\}\, d\lambda\right|$$

$$\leqslant c \sup_{I_\nu(k,j)} |\lambda \quad \lambda_j|^{\alpha - 1} \int_{-\pi}^{\pi} |\varphi_N(\lambda \quad \lambda_k)|\, d\lambda$$

$$\leqslant cN^{1-\alpha} \log N$$

using $\int_{-\pi}^{\pi} |\varphi_N(\lambda)|\, d\lambda \leqslant c \log N$. Finally in the complementary set of I_ν,

$$\left|\int_{I_v^c} \varphi_N(\lambda_j \quad \lambda)\varphi_N(\lambda \quad \lambda_k)\{f(\lambda) \quad f(\lambda_j)\}\, d\lambda\right|$$

$$\leqslant c \sup_{I_v^c} |\varphi_N(\lambda_j \quad \lambda)\varphi_N(\lambda \quad \lambda_k)| \left\{ f(\lambda_j) + \int_{-\pi}^{\pi} f(\lambda)\, d\lambda \right\} \leqslant c$$

and the lemma follows in the case $j \neq k$ because any of the bounds depends on j or k. If $j \quad k$ then we can use the same methods as before together with

$$\int_{-\pi}^{\pi} |\varphi_N(\lambda)|^2\, d\lambda \quad 2\pi N$$

to get the desired result ∎

LEMMA 3. *Under Assumption 1, if f satisfies a uniform Lipschitz condition of order* $0 < \alpha \leqslant 1$ *in an interval around a fixed frequency* ν *and if* $\sup \lambda_{j_r} | \nu \quad \lambda_{j_r}| \leqslant cm^{-1}$, $r \quad 1, \ldots, q$, *for some positive sequence* m *such that* $1/m + m/N \to 0$, *then, uniformly in* $j_r \neq 0$, *with* $j_r \neq j_{r'}$, $r \neq r'$,

$$E\left\{ \prod_{r \ 1}^{q} I(\lambda_{j_r})^{p_r} \right\} \quad \prod_{r \ 1}^{q} p_r! f(\lambda_{j_r})^{p_r} + \mathrm{O}(N^{-\alpha} \log N) \tag{14}$$

and

$$E\left[\prod_{r \ 1}^{q} \left\{ \frac{I(\lambda_{j_r}) \quad f(\lambda_{j_r})}{f(\lambda_{j_r})} \right\} \right] \quad \mathrm{O}(N^{-\alpha} \log N). \tag{15}$$

PROOF. The proof is immediate in the light of the propositon on p. 31 of BB and our Lemma 2, as by the Gaussianity of X_t only cumulants of order 2 of the discrete Fourier transform of X_t have to be considered. Here the bound in (15) is only $\mathrm{O}(N^{-\alpha} \log N)$ and not this bound to the power of q as in BB. The problem with their proof is the following. At the beginning of their p. 33, for $k \in \nu_2$ in their notation, $\mathrm{cum}\{d_x(\lambda_{k1}), d_x(\lambda_{k2})\} N^{-1} \quad \mathrm{O}(1)$ at most, because we can have $\lambda_{k1} \quad \lambda_{k2}$ for all elements in one of the possible partitions. Then, the second bound in the third full paragraph formula of the same page is only O(1) and the first one is $\mathrm{O}(N^{-\alpha} \log N)$ (actually $\mathrm{O}(N^{-1})$ under their conditions), since we have $\#\nu_1 \leqslant 1$. ∎

We now give some lemmas needed for the proof of Proposition 1.

LEMMA 4. *Under Assumption 1, if f satisfies a uniform Lipschitz condition of order* $0 < \alpha \leqslant 1$ *in an interval around a fixed frequency* ν, $I_\nu \quad [\nu \quad \epsilon, \nu + \epsilon]$ *for some* $\epsilon > 0$, *then for frequencies* $\lambda_j \quad 2\pi j/N$, $j \neq 0$, *such that* $\sup_{\lambda_j} |\nu \quad \lambda_j| \leqslant cm^{-1}$, *for some positive sequence* m *such that* $1/m + m/N \to 0$, *uniformly in* $j \neq 0$,

$$\overline{\lim_{N \to \infty}} \sup_{\lambda_j} I(\lambda_j) \leqslant 2 \log N \sup_{\lambda \in I_\nu} f(\lambda)$$

with probability 1.

PROOF. We can proceed as in the proof of Theorems 4.5.1 and 5.3.2 of Brillinger (1975), taking the mean of X_t as zero, since we do not include the zero frequency. In our case, since X_t is a Gaussian series and $j \neq 0$, all the cumulants of order bigger than two are zero. From Lemma 2 we can obtain, uniformly in j, for m big enough,

$$\sigma_N \equiv \mathrm{var}\{\mathrm{Re}\, d_x(\lambda_j)\} \quad \frac{N}{2} 2\pi f(\lambda_j) + \mathrm{O}(N^{1-\alpha} \log N).$$

Then, for $\lambda_j \in I_v$ and any θ and one ϵ as small as we want, from Gaussianity, as $N \to \infty$,

$$E[\exp\{\theta\, \mathrm{Re}\, d_x(\lambda_j)\}] \leqslant \exp\{\theta^2 2\pi N f(\lambda_j)(1+\epsilon)/4\}.$$

Next,

$$\begin{aligned} E \exp\left\{\theta \sup_{\lambda_j} |\mathrm{Re}\, d_x(\lambda_j)|\right\} &\leqslant \sum_{\lambda_j \in I_v} E \exp\{\theta |\mathrm{Re}\, d_x(\lambda_j)|\} \\ &\leqslant \sum_{\lambda_j \in I_v} \exp\left\{\theta^2 2\pi N f(\lambda_j) \frac{1+\epsilon}{4}\right\} \\ &\leqslant 2 \exp\left\{\log N + \theta^2 2\pi N \sup_{\lambda \in I_v} f(\lambda) \frac{1+\epsilon}{4}\right\}. \end{aligned}$$

Now define, for $\delta > 0$,

$$a^2 \quad 2\pi(1+\epsilon)(2+\delta) N \log N \sup_{\lambda \in I_v} f(\lambda).$$

Then

$$\mathrm{Prob}\left\{\sup_{\lambda_j} |\mathrm{Re}\, d_x(\lambda_j)| \geqslant a\right\} \leqslant \exp(\quad \theta a)\, 2 \exp\left\{\log N + \theta^2 2\pi N \sup_{\lambda \in I_v} f(\lambda) \frac{1+\epsilon}{2}\right\}.$$

Taking

$$\theta \quad a\left\{2\pi N(1+\epsilon) \sup_{\lambda \in I_v} f(\lambda)\right\}^{-1}$$

this is less than or equal to

$$2 \exp\left[\quad a^2 \left\{2\pi N \sup_{\lambda \in I_v} f(\lambda)(1+\epsilon)\right\}^{-1}\right] \exp(\log N) \leqslant c N^{-1-\delta}.$$

Using this last line and the Borel Cantelli lemma, as ϵ and δ were arbitrary, we obtain that

$$\overline{\lim_{N \to \infty}} \sup_{\lambda_j} |\mathrm{Re}\, d_x(\lambda_j)| / (2\pi N \log N)^{1/2} \leqslant \left\{\sup_{\lambda \in I_v} f(\lambda_j)\right\}^{1/2}$$

with probability 1. A similar result is possible for the imaginary part of d_x and then the lemma follows from

$$|d_x(\lambda_j)| \leqslant |\mathrm{Re}\, d_x(\lambda_j)| + |\mathrm{Im}\, d_x(\lambda_j)|$$

and $I(\lambda_j) \quad (2\pi N)^{-1} |d_x(\lambda_j)|^2$. ∎

LEMMA 5. *Under Assumptions 1, 2, 3 and 4, for frequencies* $\lambda_j \quad 2\pi j/N$ *such that* $\sup_{\lambda_j} |\nu \quad \lambda_j| \leqslant c m^{-1}$, *for some positive sequence* m *such that* $1/m + m/N \to 0$, *uniformly in* j,

$$\sup_{\lambda_j} \left| \frac{\hat{f}_M(\lambda_j) \quad f(\lambda_j)}{f(\lambda_j)} \right| \quad \mathrm{O_P}(N^{-1} M^2 + N^{(1-p)/2p} M + N^{-1} \log N + M^{-1}) \quad \mathrm{o_P}(1).$$

PROOF. Define the weighted autocovariance spectral estimate corresponding to the continuous average in $\hat{f}_M$, when the mean of X_t is known (and assumed to be zero without loss of generality), as

$$\hat{f}^{\mathrm{C}}_M(\lambda_j) \quad \int_{-\pi}^{\pi} K_M(\lambda) I(\lambda_j + \lambda)\, d\lambda \quad \frac{1}{2\pi} \sum_{r\ 1-N}^{N-1} w\left(\frac{r}{M}\right) \hat{\gamma}(r) \cos r\lambda_j$$

where $K_M(\cdot) \quad MK(M\cdot)$ periodically extended and

$$\hat{\gamma}(k) \quad N^{-1} \sum_{1 \leq t, t+k \leq N} X_t X_{t+k}.$$

This estimate is unfeasible if the mean of the series is unknown, but we only need its definition for the proofs. Now we have, proceeding as in the proof of Theorem 2.1 of Robinson (1991),

$$\sup_{\lambda_j} |\hat{f}_M(\lambda_j) \quad f(\lambda_j)| \leq \sup_{\lambda_j} |\hat{f}_M(\lambda_j) \quad \hat{f}^{\mathrm{C}}_M(\lambda_j)| \tag{16}$$

$$+ \sup_{\lambda_j} |\hat{f}^{\mathrm{C}}_M(\lambda_j) \quad E\{\hat{f}^{\mathrm{C}}_M(\lambda_j)\}| \tag{17}$$

$$+ \sup_{\lambda_j} |E\{\hat{f}^{\mathrm{C}}_M(\lambda_j)\} \quad f(\lambda_j)|. \tag{18}$$

Now (16) is less than or equal to (see Robinson, 1991, p. 1353)

$$(2\pi)^{-1} \sum_{1-N}^{N-1} \left| w\left(\frac{r}{M}\right) \right| |\hat{\gamma}(N \quad r)| \quad \mathrm{O_P}\left\{ N^{-1} \sum_{1-N}^{N-1} \left| w\left(\frac{r}{M}\right) \right| |r| \right\}$$

$$\mathrm{O_P}\left(N^{-1} M^{\alpha} \sum_{1}^{N} |r|^{1-\alpha} + N^{-1} M^2 \log N \right)$$

$$\mathrm{O_P}(N^{1-\alpha} M^{\alpha} + N^{-1} M^2 \log N) \quad \mathrm{o_P}(1)$$

using Assumption 4 ($\alpha > 5/4$) and the fact that $\hat{\gamma}(N \quad r)$ is a sum of r terms whose mean exists and is uniformly bounded. Next, (17) is not bigger than

$$(2\pi)^{-1} \sum_{1-N}^{N-1} \left| w\left(\frac{r}{M}\right) \right| |\hat{\gamma}(r) \quad E\{\hat{\gamma}(r)\}| \quad \mathrm{O_P}\left\{ \sum_{1-N}^{N-1} \left| w\left(\frac{r}{M}\right) \right| N^{(1-p)/2p} \right\}$$

$$\mathrm{O_P}(N^{(1-p)/(2p)} M) \quad \mathrm{o_P}(1)$$

because of Assumptions 2 and 4, and Lemma 7 below. Finally (18) is bounded by

$$\sup_{\lambda_j} \left| \int_{-\pi}^{\pi} K_M(\lambda_j \quad \omega) [E\{I(\omega) \quad f(\omega)\}]\, d\omega \right| \tag{19}$$

$$+ \sup_{\lambda_j} \left| \int_{\mathbb{R}} K(\omega) \{ f(\lambda_j \quad \omega/M) \quad f(\omega) \}\, d\omega \right|. \tag{20}$$

Denote by $\Phi_N(\lambda) \quad (2\pi N)^{-1} |\varphi_N(\lambda)|^2$. Similarly to Lemma 2, we have that in (19) ω lies in the interior of I_v as $M \to \infty$ due to the compact support of K, and for fixed $\delta > 0$ small enough

$$\sup_{\omega\in I_v}|E\{I(\omega)\} \quad f(\omega)| \leqslant \sup_{\omega\in I_v}\left|\int_{-\pi}^{\pi}\Phi_N(\alpha \quad \omega)\{f(\alpha) \quad f(\omega)\}\,d\alpha\right|$$

$$\leqslant \sup_{\omega\in I_v}|f'(\omega)|\int_{|\omega-\alpha|\leqslant\delta}|\Phi_N(\alpha \quad \omega)\|\alpha \quad \omega|\,d\alpha$$

$$+ \sup_{\omega\in I_v}\int_{|\omega-\alpha|>\delta}|\Phi_N(\alpha \quad \omega)|\{f(\alpha)+f(\omega)\}\,d\alpha$$

$$\mathrm{O}(N^{-1}\log N)+\mathrm{O}(N^{-1})$$

$$\mathrm{O}(N^{-1}\log N)$$

uniformly in $\omega \in I_v$, so (19) is $\mathrm{O}(N^{-1}\log N)$ since $\int|K_M(\alpha)|\,d\alpha<\infty$. Next, as $M\to\infty$, (20) is bounded by (denoting by λ^* a value between λ_j and $\lambda_j \quad \omega/M$),

$$\sup_{\lambda_j}\int K(\omega)\{f(\lambda_j \quad \omega/M) \quad f(\lambda_j)\}\,d\omega \leqslant \sup_{\lambda_j}\int|K(\omega)\|f'(\lambda^*)|\left|\frac{\omega}{M}\right|d\omega \quad \mathrm{O}(M^{-1}),$$

using the compact support of K and that f' is bounded in I_v. ∎

LEMMA 6. *Under the assumptions of Lemma 5, uniformly in j,*

$$\sup_{\lambda_j}\left|\frac{\hat{f}^{\mathrm{j}}_M(\lambda_j) \quad f(\lambda_j)}{f(\lambda_j)}\right| \quad \mathrm{O_P}(N^{-1}M^2+N^{(1-p)/(2p)}M+N^{-1}\log N+M^{-1}).$$

PROOF. The proof is exactly the same as that of Lemma 4 of BB, using now our Lemma 5. ∎

LEMMA 7. *Under Assumptions 1 and 2, uniformly in r, for* $p>1$,

$$\mathrm{var}\{\hat{\gamma}(r)\} \quad \mathrm{O}(N^{(1-p)/p})$$

where $\hat{\gamma}(r)$ *is the (biased) estimate of the lag-r autocovariance* $\gamma(r)$ *when the expectation of* X_t *is known:*

$$\hat{\gamma}(r) \quad \frac{1}{N}\sum_{1\leqslant t,t+r\leqslant N}\{X_t \quad E(X_1)\}\{X_{t+r} \quad E(X_1)\}.$$

PROOF. From, for example, Anderson (1971, p. 452), denoting as before the Fejér kernel by Φ_N,

$$N\,\mathrm{var}\{\hat{\gamma}(r)\} \quad \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\Phi_N(\alpha \quad \omega)[1+\exp\{ \quad i(\alpha+\omega)r\}]f(\alpha)f(\omega)\,d\alpha\,d\omega$$

and now the result follows on applying the Hölder inequality twice, with $|\Phi_N(\omega)| \quad \mathrm{O}(N)$ uniformly in ω, $|1+\exp\{ \quad i\alpha+\omega)r\}|\leqslant 2$ uniformly in r and $\int_{-\pi}^{\pi}f^p<\infty$ by Assumption 2. ∎

PROOF OF PROPOSITION 1. From the proof of Theorem 3.1 in BB the proposition will follow, using their definitions, if we show

$$N^{-1}T_i \quad \mathrm{o_P}(\mathrm{IMSE}_m) \qquad i \quad 1,\,2$$

$$N^{-1}T_3 \quad \mathrm{IMSE}_m+\mathrm{o_P}(\mathrm{IMSE}_m).$$

First we have, denoting now σ_j $\sigma_{j,M}$, from the last steps in the proofs of BB,

$$E(T_1) \quad 2\pi \sum_{j\ 1}^{N} W_m(\lambda_j \quad \nu)\sigma_j^{-1} \sum_k{}' K(M\lambda_k)\mathrm{O}(N^{-1}) \quad \mathrm{O}(1)$$

and, denoting as $\mathrm{IMSE}'_m(\nu, M)$ the IMSE_m calculated from the modified spectral estimate (3),

$$\begin{aligned} E(T_1^2) \quad & N\,\mathrm{IMSE}'_m + 2\pi \sum_j W_m(\lambda_j \quad \nu)^2\sigma_j^{-2} \sum_k{}' \sum_n{}' K(M\lambda_k)K(M\lambda_n)\mathrm{O}(N^{-1}) \\ & + 2\pi \sum_j \sum_{i/j} W_m(\lambda_j \quad \nu)W_m(\lambda_i \quad \nu)\sigma_j^{-1}\sigma_i^{-1} \sum_k{}' \sum_n{}' K(M\lambda_k)K(M\lambda_n)\mathrm{O}(N^{-2}) \\ & + 2\pi \sum_j W_m(\lambda_j \quad \nu)^2\sigma_j^{-2} \sum_k{}' \sum_n{}' K(M\lambda_k)^2 \\ & N\,\mathrm{IMSE}'_m + \mathrm{O}(m) + \mathrm{O}(m) + \mathrm{O}(m\,N\,\mathrm{IMSE}_m) \\ & \mathrm{O}(m\,N\,\mathrm{IMSE}_m) \end{aligned}$$

since $\sup_{\lambda,m}|W_m(\mathrm{l})| \quad \mathrm{O}(m)$. Then, using $\mathrm{IMSE}_m \quad \mathrm{O}(M/N)$ we can obtain

$$T_1 \quad \mathrm{O_P}\{\mathrm{IMSE}_m(m/M)^{1/2}\} \quad \mathrm{o_P}(\mathrm{IMSE}_m)$$

because $m/M \to 0$. Now, in a similar fashion,

$$E(T_2) \quad 2\pi \sum_{j\ 1}^{N} W_m(\lambda_j \quad \nu)\sigma_j^{-2} \sum_k{}' \sum_n{}' K(M\lambda_k)K(M\lambda_n)\mathrm{O}(N^{-1}) \quad \mathrm{O}(1)$$

and, as before,

$$E(T_2) \quad \mathrm{O}(m\,N\,\mathrm{IMSE}_m^2).$$

(Note that in BB's expression they have N^{-1} instead of N in the corresponding formula, although in the statements in the main part of their paper they give the right bounds.) Then $N^{-1}T_2 \quad \mathrm{O_P}\{\mathrm{IMSE}_m(m/N)^{1/2}\} \quad \mathrm{o_P}(\mathrm{IMSE}_m)$. Next,

$$\begin{aligned} E(T_3) \quad & N\,\mathrm{IMSE}_m + \mathrm{O}\{(N/M)^{-1}N\,\mathrm{IMSE}_m\} + \mathrm{O}\{(N/M)^{-1}\} + \mathrm{O}\{(N/M)^{-1}M\} \\ & N\,\mathrm{IMSE}_m + \mathrm{O}\{(N/M)^{-1}N\,\mathrm{IMSE}_m\} \end{aligned}$$

and reasoning in the same way as before

$$\mathrm{var}(T_3) \quad \mathrm{O}(m\,N\,\mathrm{IMSE}_m).$$

Then $N^{-1}T_3 \quad \mathrm{IMSE}_m + \mathrm{O_P}\{\mathrm{IMSE}_m(m/M)^{1/2}\} \quad \mathrm{IMSE}_m + \mathrm{o_P}(\mathrm{IMSE}_m)$. The proof for the remainder term in BB's expression (3.2) continues the same here, using now our Lemmas 2, 4 and 5 instead of their references, since the bound for the third term in the expansion still holds for the modified (local) cross validation. ■

ACKNOWLEDGEMENTS

This paper is a revised version of the Chapter 4 of the author's University of London Ph.D. Thesis. I am grateful to Professor P. M. Robinson for his guidance

and advice. I also thank the Associate Editor and a referee for helpful comments. The research was supported by ESRC Grant R000235892 and Spanish Dirección General de Enseñanza Superior Ref. PB95-0292.

REFERENCES

ANDERSON, T. W. (1971) *The Statistical Analysis of Time Series.* New York: Wiley.
ANDREWS, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 60, 953–66.
BELTRAO, K. I. and BLOOMFIELD, P. (1987) Determining the bandwidth of a kernel spectrum estimate. *J. Time Ser. Anal.* 8, 21–38.
BRILLINGER, D. R. (1975) *Time Series: Data Analysis and Theory.* New York: Holt, Rinehart and Winston.
BROCKMANN, M., GASSER, T. and HERRMANN, E. (1993) Locally adaptive bandwidth choice for kernel regression estimators. *J. Am. Stat. Assoc.* 88, 1302–9.
BÜHLMANN, P. (1996) Locally adaptive lag-window spectral estimation. *J. Time Ser. Anal.* 17, 247–70.
DEN HAAN, W. J. and LEVIN, A. (1997) A practitioners guide to robust covariance matrix estimation. In *Handbook of Statistics*, Vol. 15: *Robust Inference* (eds G. S. Maddala and C. R. Rao). Amsterdam: North-Holland, pp. 291–341.
EFROMOVICH, S. (1998) Data-driven efficient estimation of the spectral density. *J. Am. Stat. Assoc.* 92, 762–69.
FAN, J. and GIJBELS, I. (1995) Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. R. Stat. Soc. Ser. B.* 57, 371–394.
—— and KREUTZBERGER, E. (1998) Automatic local smothing for spectral density estimation. *Scand. J. Stat.* 25, 359–69.
HANNAN, E. J. and NICHOLLS, D. F. (1977) The estimation of the prediction error variance. *J. Am. Stat. Assoc.* 72, 834–40.
HERRMANN, E. (1997) Local bandwidth choice in kernel regression estimation. *J. Comput. Graph. Stat.* 6, 35–54.
HURVICH, C. M. (1985) Data driven choice of a spectrum estimate: extending the applicability of cross-validation methods. *J. Am. Stat. Assoc.* 80, 933–40.
—— and BELTRAO, K. I. (1994) Automatic semiparametric estimation of the memory parameter of a long-memory time series. *J. Time Ser. Anal.* 15, 285–302.
KOOPERBERG, C. STONE, C. J. and TRUONG, Y. K. (1995) Rate of convergence for logspline spectral density estimation. *J. Time Ser. Anal.* 16, 389–401.
LEPSKII, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997) Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Stat.* 25, 929–47.
NEWEY, W. K. and WEST, K. D. (1994) Automatic lag selection in covariance matrix estimation. *Rev. Econ. Stud.* 61, 631–53.
PAWITAN, Y. and O'SULLIVAN, F. (1994) Nonparametric spectral density estimation using penalized Whittle likelihood. *J. Am. Stat. Assoc.* 89, 600–10.
ROBINSON, P. M. (1991) Automatic frequency domain inference on semiparametric and non-parametric models. *Econometrica* 59, 1329–63.
WAHBA, G. (1980) Automatic smoothing of the log periodogram. *J. Am. Stat. Assoc.* 75, 122–32.