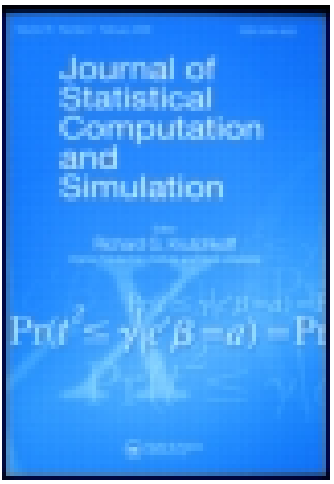

Journal of Statistical Computation and Simulation

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

Optimal window width choice in spectral density estimation

Fortin Ines [a] & Kuzmics Christoph [a]

[a] Institut für HÖhere Studien (Institute for Advanced Studies) , Stumpergasse 56, Vienna, A-1060, Austria

Published online: 20 Mar 2007.

To cite this article: Fortin Ines & Kuzmics Christoph (2000) Optimal window width choice in spectral density estimation, Journal of Statistical Computation and Simulation, 67:2, 109-131, DOI: 10.1080/00949650008812038

To link to this article: http://dx.doi.org/10.1080/00949650008812038

J. Statist. Comput. Simul., 2000, Vol. 67, pp. 109–131
Reprints available directly from the publisher
Photocopying permitted by license only

OPTIMAL WINDOW WIDTH CHOICE IN SPECTRAL DENSITY ESTIMATION

Review and Simulation

INES FORTIN and CHRISTOPH KUZMICS*

Institut für Höhere Studien (Institute for Advanced Studies), Stumpergasse 56, A-1060 Vienna, Austria

(Received 25 June 1999; In final form 9 February 2000)

This paper deals with optimal window width choice in non-parametric lag or spectral window estimation of the spectral density of a stationary zero-mean process. Several approaches are reviewed: cross-validation-based methods as described by Hurvich (1985), Beltrão and Bloomfield (1987) and Hurvich and Beltrão (1990); an iterative procedure developed by Bühlmann (1996); and a bootstrap approach followed by Franke and Härdle (1992). These methods are compared in terms of the mean square error, the mean square percentage error, and a third measure of the distance between the true spectral density and its estimate. The comparison is based on a simulation study, the simulated processes being in the class of ARMA (5,5) processes. On the basis of simulation evidence we suggest to use a slightly modified version of Bühlmann's (1996) iterative method. This paper also makes a minor correction of the bootstrap criterion by Franke and Härdle (1992).

Keywords: Bandwidth; non-parametric estimation; spectral density; simulation

1. INTRODUCTION

The crucial step in non-parametric spectral density estimation is the choice of the window width, or 'bandwidth', of some specified lag window or spectral window employed for smoothing the periodogram. To optimally determine this scale parameter one may attempt to minimize some measure of the distance between the true spectral

*Corresponding author.

density of a process and its estimator over a range of scale parameters. Various theoretical criteria have been proposed, among them the mean square error (MSE) and the mean square percentage error (MSPE). The resulting optimal value for the scale parameter, however, depends on the unknown true underlying spectral density. One way to tackle this problem is to use a likelihood motivated cross-validation criterion. This may be seen as an estimate of some distance measure (Hurvich, 1985; Beltrão and Bloomfield, 1987; Hurvich and Beltrão, 1990). Franke and Härdle (1992), on the other hand, consider bootstrap estimates of some distance measure by resampling the residuals of a multiplicative non-parametric regression, which can be shown to be 'nearly' independent. Another way to deal with the problem was developed by Bühlmann (1996). He iteratively estimates the spectral density by calculating the optimal scale parameter in every step according to one of the theoretical criteria at hand, and plugging in the (step-) spectral density estimate for the true spectral density.

This paper aims to provide a brief review of the literature mentioned and to compare the various methods on the basis of a simulation study using a set of ARMA (5,5) processes. The cross-validation-based methods compare poorly to both the iterative methods and the bootstrap method. While for the bootstrap approach the issue of finding an initial global estimate remains a problem, the iterative procedures are self-contained. A slightly modified version of Bühlmann's (1996) procedure seems to be superior to the others.

2. SPECTRAL DENSITY ESTIMATES

We consider a real(-valued) strictly stationary stochastic process $\{X_t, t \in \mathbb{N}\}$ with autocovariance function $\gamma(k) = \mathrm{Cov}(X_0, X_k)$ and zero expectation. The spectral density function of the process X_t is assumed to exist and is given by the discrete inverse Fourier transform of the autocovariance function,

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathrm{e}^{-ik\omega} \gamma(k) \quad \text{for all } \omega \in [-\pi, \pi]. \tag{1}$$

Let $x_1, \ldots, x_n$ be a sample of this process. The Fourier frequencies of the sample are defined as $\omega_j = (2\pi j/n)$, where j assumes integer values such that $-\pi < \omega_j \leq \pi$. A sample estimate of the autocovariance function may be given by $\hat{\gamma}(k) = (1/n)\sum_{t=1}^{n-|k|} x_{t+k}x_t$ for $-n < k < n$. This sample version of the autocovariance function yields the periodogram, an intuitive estimate of the spectral density, by replacing $\gamma(\cdot)$ in the definition of the spectral density, as in (1), by its estimate $\hat{\gamma}(\cdot)$,

$$I(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} e^{-ik\omega}\hat{\gamma}(k) \quad \text{for all } \omega \in [-\pi, \pi]. \tag{2}$$

The periodogram is not a consistent estimator of the spectral density (see *e.g.*, Priestley, p. 425) in the sense that $\mathrm{Var}(I(\omega))$ does not converge to zero as $n \to \infty$. A smoothed version of the periodogram, however, may be shown to be a mean square consistent estimator of the true spectral density. This generally biased estimator is given by

$$\hat{f}(\omega; h) = \frac{1}{2\pi h} \sum_{j=-N}^{N} K(h(\omega - \omega_j))I(\omega_j), \tag{3}$$

where N is the largest integer less than or equal to $(n-1)/2$ and h is the scale parameter that controls for the width of the window. This window is generated by the spectral kernel $K(\cdot)$, which satisfies (5) and conditions below.

Every spectral kernel estimate approximately equals a corresponding lag kernel estimate,

$$\hat{f}(\omega; h) \approx \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \kappa\left(\frac{k}{h}\right)\hat{\gamma}(k)e^{-ik\omega}, \tag{4}$$

(see Brockwell and Davis, p. 354, Priestley, p. 434, 6.2.54), where the spectral kernel $K(\cdot)$ and the lag kernel $\kappa(\cdot)$ relate *via* a Fourier transformation,

$$K(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \kappa(x)e^{-ix\omega}dx, \tag{5}$$

and $\kappa(\cdot)$ (see Priestley, p. 446) is an even function $\kappa : \mathbb{R} \to \mathbb{R}^+$ with $(0) = 1$, $|\kappa(x)| \leq 1$ for all x, and $\kappa(x) = 0$ for $x > h$.

A lag kernel is a C^r kernel if it is r times continuously differentiable in the neighborhood of zero and Lipschitz-continuous on $\mathbb{R}$. A lag kernel κ has characteristic exponent r if $\kappa^{(s)} = 0$ for all $s < r$ and $\kappa^{(r)} \neq 0$, where $\kappa^{(s)} = \lim_{x \to 0} ((1 - \kappa(x))/(|x|^s))$ is the generalized s-th derivative of a lag kernel $\kappa(\cdot)$ at zero (see Bühlmann, p. 249, def. 1; Priestley, p. 459, 6.2.121).

As the choice of the kernel is less important than that of the scale parameter (see Priestley, p. 449) in terms of density estimation we only consider the Bartlett–Priestley window lag kernel

$$\kappa(x) = \frac{3}{\pi^2 x^2} \left\{ \frac{\sin(\pi x)}{\pi x} - \cos(\pi x) \right\} \tag{6}$$

and spectral kernel

$$K(\omega) = \begin{cases} \frac{3}{4\pi} \left(1 - \left(\frac{\omega}{\pi}\right)^2\right) & \text{if } |\omega| \leq \pi \\ 0 & \text{if } |\omega| > \pi. \end{cases} \tag{7}$$

The Bartlett–Priestley window is C^2 with characteristic exponent 2.

Given a specific window one has to choose the scale parameter. Different scale parameters may yield utterly different estimates of the spectral density. Basically, one may get all estimates between a straight line with slope zero and the wildly fluctuating periodogram for different choices of h. Optimally, one would choose the scale parameter such as to minimize some measure of the expected distance between the estimator and the true spectral density. Various different measures were suggested in the literature (Priestley, p. 510ff). We consider distance measurers as follows:

$$\text{MSE}_a(\omega) = E\{(\hat{f}(\omega) - f(\omega))^2 (f(\omega))^a\} \tag{8}$$

as a measure of the local distance at a fixed frequency ω, where $a = 0$ yields the mean square error (MSE) and $a = -2$ the mean square percentage error (MSPE), and

$$\text{MISE}_a = E \int_{-\pi}^{\pi} \{(\hat{f}(\omega) - f(\omega))^2 (f(\omega))^a\} d\omega \tag{9}$$

as a global measure of the expected distance between the true density and its estimator.

The minimization one of the local criteria produces an optimal local choice of the scale parameter. If we look at our local distance measures we realize that it does not matter which of them we choose: At a fixed frequency, one measure is just a monotone transformation of the other, and this is true for all distance measures in the class we consider. The advantage of allowing for locally different scale parameters (window widths) is the possibility to adjust for the shape of the actual density function at different frequencies. While for flat regions of the density a high amount of smoothing may be called for, for peaky regions it seems better not to smooth too much. Not too much weight should be given to periodogram values that are far apart from the considered frequency, as their mean value would be very different from the one at the considered frequency. If one liked to employ the same smoothing window for the whole spectrum, one might decide to select the scale parameter such as to minimize the global, *i.e.*, integrated, criteria.

For the time being we will consider minimizing the MSE. By the usual variance decomposition the MSE can be written as the sum of the squared bias and the variance, $\mathrm{MSE}(\hat{f},\omega) = \mathrm{BIAS}^2(\hat{f},\omega) + \mathrm{VAR}(\hat{f},\omega)$. As different estimators of the spectral density are determined by different scale parameters only, we replace $\hat{f}$ in expressions like $\mathrm{BIAS}(\hat{f},\omega)$ by the scale parameter h. For the Bartlett–Priestley window asymptotic bias and variance may be found. These are given by (see Priestley, pp. 457–463)

$$\mathrm{ABIAS}(h,\omega) = -\frac{\pi^2}{10h^2} f''(\omega) \tag{10}$$

and

$$\mathrm{AVAR}(h,\omega) = \iota(\omega)\frac{6h}{5n} f^2(\omega), \tag{11}$$

respectively, where

$$\iota(\omega) = \begin{cases} 2 & \text{if } \omega \in \{-\pi, 0, \pi\} \\ 1 & \text{otherwise.} \end{cases} \tag{12}$$

Minimizing the sum of squared bias and the variance yields asymptotically optimal values for the local and global scale parameters,

$$h_{\text{opt}}(\omega)=n^{1/5}\left\{\frac{\pi^4}{30}\frac{\{f''(\omega)\}^2}{\iota(\omega)f^2(\omega)}\right\}^{1/5} \tag{13}$$

and

$$h_{\text{opt}}=n^{1/5}\left\{\frac{\pi^4}{30}\frac{\int_{-\pi}^{\pi}\{f''(\omega)\}^2 d\omega}{\int_{-\pi}^{\pi}f^2(\omega)d\omega}\right\}^{1/5}, \tag{14}$$

respectively.

These optimal values for h, however, depend on the true spectral density and its derivatives. As we do not know the true spectrum, we will have to adopt some kind of either direct or indirect (iterative procedure) estimation of the MISE.

3. SCALE PARAMETER SELECTION METHODS

3.1. Cross-validation Methods

Following Marron (1985), who developed a cross-validation-based selection criterion to determine the window width in non-parametric probability density estimation, Beltrão and Bloomfield (1987) provide a cross-validated log-likelihood criterion (CVLL) for determining the smoothness parameter in non-parametric spectral density estimation.

$$\text{CVLL}(\hat{f})=\sum_{0<\omega_j<\pi}\log\hat{f}^{-j}(\omega_j;h)+\frac{I(\omega_j)}{\hat{f}^{-j}(\omega_j,h)}, \tag{15}$$

where

$$\hat{f}^{-j}(\omega_j,h)=\frac{1}{\sigma_j(h)}\sum_{k\notin J(n,j)}K(h\omega_k)I(\omega_j-\omega_k), \tag{16}$$

$\sigma_j(h)=\sum_{k\notin J(n,j)}K(h\omega_k)$, and $J(n,j)$ is the set of indices k for which $I(\omega_j-\omega_k)=I(\omega_j)$.

Beltrão and Bloomfield show that by minimizing CVLL one approximately minimizes the mean integrated square percentage error MISPE (*i.e.*, MISE_{-2} as in (9)), as the difference in CVLL for two different window widths h_1 and h_2 is approximately proportional to the difference in MISPE.

$$\frac{2}{N}(\text{CVLL}(h_1)-\text{CVLL}(h_2)) \approx \text{MISPE}(h_1)-\text{MISPE}(h_2). \tag{17}$$

Hurvich (1985) presents two other cross-validation-based selection criteria: Stuetzle's smoothed estimate (SES, see Palmer, 1983), and an adaptation of the cross-validated mean square error (CVMSE) method of Wahba and Wold (1975).

$$\text{SES}(\hat{f}) = \frac{1}{N}\sum_{j=1}^{N}(\hat{f}^{-j}(\omega_j) - I(\omega_j))^2, \tag{18}$$

$$\text{CVMSE}(\hat{f}) = \frac{1}{N}\sum_{j=1}^{N}\{(\log\hat{f}^{-j}(\omega_j) - (\log I(\omega_j)+C))^2 - \pi^2/6\}. \tag{19}$$

where $C = 0.577216\ldots$ is Euler's constant. SES is motivated by MISE_0 as in (9) and CVMSE by another distance measure $E(1/N)\sum_{j=1}^{N}(\log\hat{f}(\omega_j) - \log f(\omega_j))^2$.

Hurvich and Beltrão (1990) propose a computationally more efficient estimate, which they call CVLL2, but which in fact is not a cross-validation estimate:

$$\text{CVLL2} = 2n\log(2\pi) + \frac{n}{N}\sum_{\omega_j}\log\hat{f}(\omega_j) + \frac{nv}{v-2}, \tag{20}$$

where $v = 2/(\sum_{-\pi/h<\omega_j<\pi/h} K^2(h\omega_j))$. Note that computing time for calculating the non-cross validatory CVLL2 is comparably short, as it requires $O(n\log n)$ computations, whereas the cost of computing CVLL is $O(n\log n + n(N/h))$.

3.2. An Iterative Procedure

Bühlmann (1996) builds on the work by Brockmann *et al.* (1993) who employed a similar idea in the context of non-parametric regression. He estimates the optimal local and global window widths which

minimize the asymptotic mean square error and the asymptotic mean integrated square error by an iterative procedure.

As for the Bartlett–Priestley window, asymptotically optimal window widths are known and given by Eqs. (13) and (14). The optimal scale parameter, however, depends on the true spectral density and its second derivative with respect to ω. Both, of course, are unknown. Bühlmann's idea is to start with an initial estimate for the spectral density as well as its second derivative and iteratively improve these estimates in the following way. He iteratively plugs the estimates into the formula for the theoretical optimal window width, thus receiving an improved window width and uses this for the estimates of the next step.

For the various spectral density related objects Bühlmann considers the following estimators. For the density he suggests to use the estimator as given by (4) with lag kernel $\bar{\kappa}$ as defined in (23). For the integrated squared density he suggests to use the integrated squared periodogram divided by two,

$$\frac{1}{2}\int_{-\pi}^{\pi}\left\{\frac{1}{2\pi}\sum_{k=-n+1}^{n-1}\hat{\gamma}(k)\mathrm{e}^{-ik\omega}\right\}^2 d\omega. \tag{21}$$

For the second generalized derivative of the spectral density he suggests to use the estimator

$$\hat{f}^{(2)}(\omega,h)=\frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}\bar{\kappa}\left(\frac{k}{h}\right)k^2\hat{\gamma}(k)\mathrm{e}^{-ik\omega}, \tag{22}$$

where the second generalized derivative is defined as $f^{(2)}(\omega)=(1/2\pi)\sum_{k=-\infty}^{\infty}k^2\gamma(k)\mathrm{e}^{-ik\omega}$ and satisfies the identity $f^{(2)}=-f''$ and where $\bar{\kappa}$ is defined by (23).

In his Remark 2, Bühlmann proposes to use different windows for estimating the generalized derivatives of the spectral density. He argues that the terms $k^2\hat{\gamma}(k)$ usually do not decay very fast for increasing k. The lag window he proposes is a specific splitted rectangular-cosine window with lag kernel

$$\bar{\kappa}(x)=\begin{cases}1 & \text{if } |x|<0.8\\ \{1+\cos(5(x-0.8)\pi)\}/2 & \text{if } 0.8\le|x|<1\\ 0 & \text{otherwise.}\end{cases} \tag{23}$$

The iteration scheme employed by Bühlmann is the following.

ALGORITHM 3.1

1. $h_0 = n^{-1/2}$, *the initial window width*
2. $i = 0$, *counting the number of iterations*
3. $i = i + 1$
4. *Global steps*: $h_i = n^{1/5}\left\{\frac{2\{\kappa''(0)\}^2 \sum_{k=-n+1}^{n-1} \bar{\kappa}^2\left((k/h_{i-1})n^{4/45}\right)k^4\hat{\gamma}^2(k)}{\int_{-\infty}^{\infty}\kappa^2(x)dx\sum_{k=-n+1}^{n-1}\hat{\gamma}^2(k)}\right\}^{1/5}$
5. *if* $i < 4$ *goto* 3
6. *Local step*: $h_{opt}(\omega) = n^{1/5}\left\{\frac{2\{\kappa''(0)\}^2\left\{\hat{f}^{(2)}\left(\omega;(h_4/n^{4/45})\right)\right\}^2}{\int_{-\infty}^{\infty}\kappa^2(x)dx\left\{\hat{f}\left(\omega;(h_4/n^{4/45})\right)\right\}^2}\right\}^{1/5}$.

Bühlmann motivates the inflation factor $n^{4/45}$ by some asymptotics for the local step. He argues that using the same factor in the global steps too will yield a more stable procedure. Bühlmann concludes that four global iteration steps will already yield the right order and further steps will not lead to any improvement. Performing more than one local step will not improve the estimate either.

The only problem that may arise in local smoothing is at inflection points of the spectral density. At these points, where the second derivative is zero, the formulation (13) of the locally optimal scale parameter is not true. Bühlmann suggests to employ a semi-local scale parameter selection criterion. The estimate of the second derivative in the local step in Algorithm 3.1 is replaced by its integral over a small range. Semi-local step:

$$h_{\text{opt}}(\omega) = n^{1/5}\left\{\frac{2\{\kappa''(0)\}^2\int_{\omega-c}^{\omega+c}\left\{\hat{f}^{(2)}\left(\lambda;(h_4/n^{4/45})\right)\right\}^2 d\lambda}{\int_{-\infty}^{\infty}\kappa^2(x)dx\left\{\hat{f}\left(\omega;(h_4/n^{4/45})\right)\right\}^2}\right\}^{1/5},$$

where $c = n^{4/45}/h_4$.

In our simulation we determine not only the estimates of the global as well as the semi-local window widths as given by the above algorithm, but also some estimates using the same plug-in scheme, with different estimates in the spectral density and its derivatives. In particular, we try to approximate the second derivative by differences, and we use estimates with inflated as well as non-inflated window widths.

3.3. A Bootstrap Approach

Franke and Härdle (1992) adopt a bootstrap approach in order to determine the optimal scale parameter. They introduce the bootstrap in frequency domain *via* a multiplicative regression problem,

$$I(\omega_j)=f(\omega_j)\varepsilon_j. \tag{24}$$

The residuals are approximately independent and identically distributed for large n (see Priestley, Chapter 6.2). These residuals replace the true density by the kernel estimate using an 'arbitrary' initial scale parameter h_0. They constitute the sample of independent observations to be resampled,

$$\hat{\varepsilon}_j=\frac{I(\omega_j)}{\hat{f}(\omega_j;h_0)}, \quad \text{for } j=1,\ldots,N. \tag{25}$$

In fact, the residuals actually used are the rescaled ones, given by

$$\tilde{\varepsilon}_j=\frac{\hat{\varepsilon}_j}{(1/N)\sum_{j=1}^{N}\hat{\varepsilon}_j}. \tag{26}$$

The bootstrap procedure is performed as follows. A bootstrap sample $\varepsilon_1^*,\ldots,\varepsilon_N^*$ is drawn from the empirical distribution of $\tilde{\varepsilon}_1,\ldots,\tilde{\varepsilon}_N$. Alternatively, a bootstrap sample may be drawn from an exponential distribution with scale parameter 1, as this is the limiting distribution of the regression residuals. In the simulation part we will only look at resampling from the empirical distribution. Using a bandwidth g, possibly different from h_0, bootstrap periodogram values

$$I^*(\omega_j)=I^*(\omega_{-j})=I^*(-\omega_j)=\hat{f}(\omega_j,g)\varepsilon_j^* \tag{27}$$

are obtained. The corresponding bootstrap spectral estimate is then given by

$$\hat{f}^*(\omega,h,g)=\frac{2\pi h}{n}\sum_{j=-N}^{N}K(h(\omega-\omega_j))I^*(\omega_j). \tag{28}$$

Franke and Härdle minimize the mean square percentage error MSPE, *i.e.*, MSE_{-2} in Eq. (8), which is the same as minimizing the mean square error or any other monotone transformation of it.

Minimizing MSPE with respect to h should yield the optimal scale parameter. As MSPE is not known, however, we will minimize its bootstrap estimate, given by

$$\text{MSPE}^*(\omega,h)=E^*\left\{\frac{\hat{f}^*(\omega,h,g)-\hat{f}(\omega,g)}{\hat{f}(\omega,g)}\right\}^2. \tag{29}$$

In fact, there is no need to resample, as we can calculate MSPE* explicitly.

$$\begin{aligned}\hat{f}^2(\omega;g)\text{MSPE}^*(\omega;h)&=\text{MSE}^*(\omega;h)\\&=\frac{h^2\text{var}^*(\varepsilon_1^*)}{n^2}\left(K^2(0)\hat{f}^2(0;g)+\sum_{j=1}^{N}\{K(h(\omega-\omega_j))\right.\\&\qquad\left.+K(h(\omega+\omega_j))\}^2\hat{f}^2(\omega_j;g)\right)\\&\quad+\left\{\frac{h}{n}\sum_{j=-N}^{N}K(h(\omega-\omega_j))\hat{f}(\omega_j;g)-\hat{f}(\omega;g)\right\}^2. \qquad (30)\end{aligned}$$

There appears to be a small error in Franke and Härdle's formula (6). They seem to have neglected the cross terms in the variance, which are present as $I(\omega_{-j})=I(\omega_j)=\varepsilon_j^*\hat{f}(\omega_j,g)$. Franke and Härdle (p. 135) note that $\text{var}^*(\varepsilon_1^*)\to 1$ in probability. In our simulation we will use the value 1 for the bootstrap variance of each ε_j^*. The scale parameter minimizing the above estimate of the MSPE is the one regarded optimal. Franke and Härdle prove that the resulting estimate is in fact a consistent estimate of the optimal scale parameter (p. 133, Theorem 3).

4. SIMULATION

In this section we compare the procedures discussed in the previous sections by means of a simulation study. The procedures are applied to a set of AR and MA-processes, selected such as to exhibit different shapes of spectral densities (see Figs. 1–5). For each process, we simulate 300 time series of length 120 and 480. For each process and

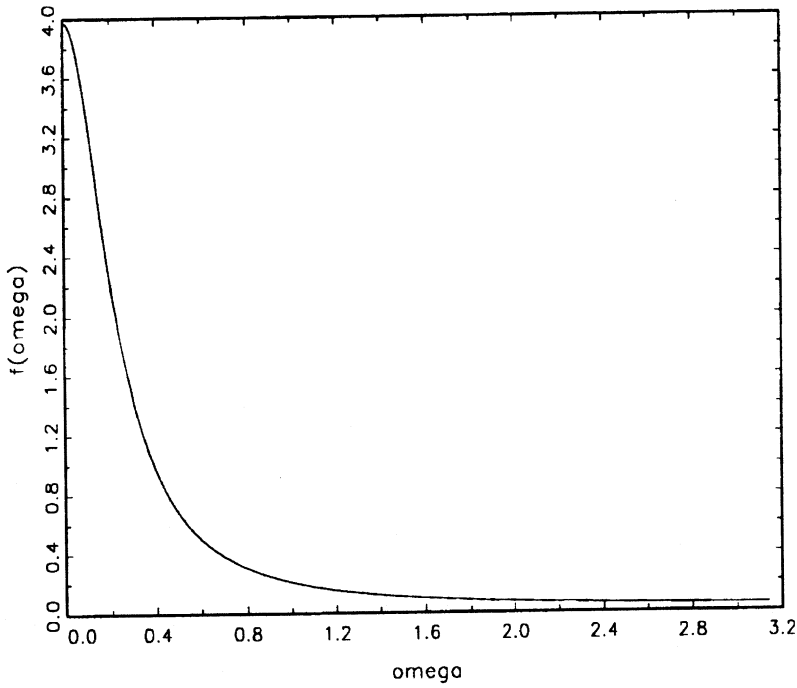

FIGURE 1 $X_t = 0.8X_{t-1} + Z_t$.

each scale parameter selection method three different distance measures are approximated: MISE_0, MISE_{-2} and MISE_2, as defined by (9). These are in fact calculated as the average over all simulations of

$$\text{ISE}_a = \frac{1}{N} \sum_{j=-N}^{N} (\hat{f}(\omega_j) - f(\omega_j))^2 (f(\omega_j))^a. \tag{31}$$

Standard normal random numbers are generated by RNDN, the normal random number generator in GAUSS. A time series of, say, length 120 is generated by setting initial values to zero, generating a sequence of 220 standard normal random numbers, recursively (if necessary) determining 220 realizations of the particular process, and finally dropping the first hundred (see appendix in Hurvich, 1985, p. 939).

In the tables which summarize the simulation results (Tabs. I–X), we abbreviate the methods employed in the following way. The

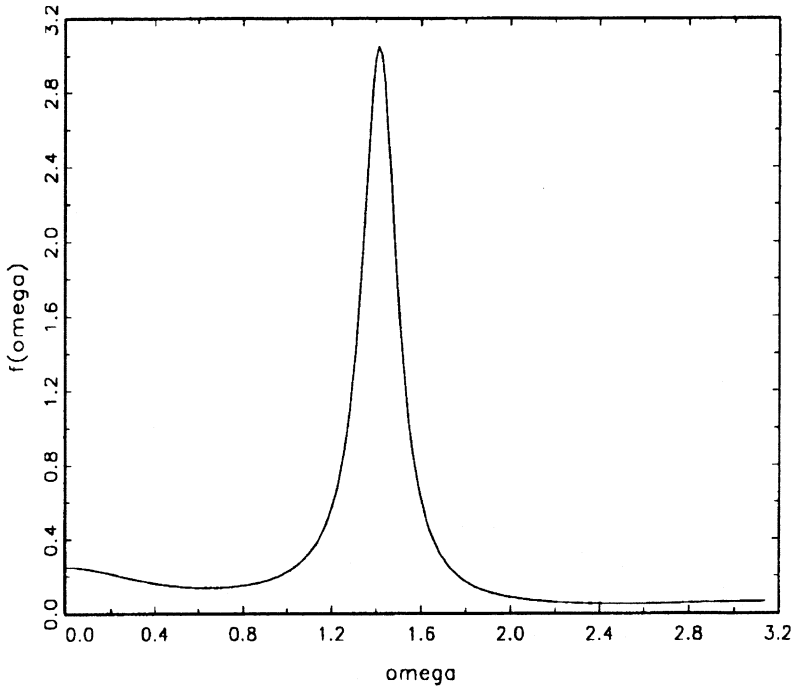

FIGURE 2 $X_t = 0.4X_{t-1} - 0.5X_{t-2} + 0.3X_{t-4} + Z_t$.

cross-validation methods, CVLL, CVLL2, SES and CVMSE are as defined in Eqs. (15), (20), (18) and (19), respectively. ITB is the global method suggested by Bühlmann (1996) and is given by the first part of Algorithm 3.1. It uses the splitted rectangular-cosine lag window estimate of the second generalized derivative and the inflation factor $n^{4/45}$, which is approximately 1.53 if n is 120 and 1.73 if n is 480. ITC refers to a global method according to the same algorithm, but without using the inflation factor. ITA is another global method following the said algorithm, which does not use the inflation factor and approximates the second generalized derivative by finite differences. The suffices 1 and 2 in *e.g.*, ITA1 and ITA2 refer to the semi-local estimation method as in the second part of Algorithm 3.1, where 1 indicates that the inflation factor was used in the semi-local step and 2 that it was not used. The global window width used in the semi-local step is always given by the corresponding global method,

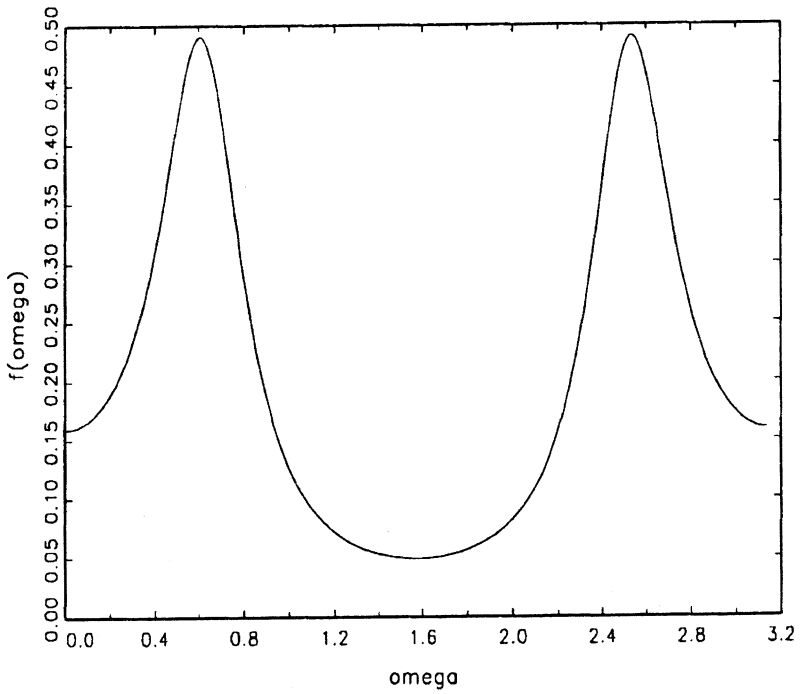

FIGURE 3 $X_t = 0.4X_{t-2} - 0.4X_{t-4} + Z_t$.

e.g., ITA for ITA1. BOOT denotes the bootstrap criterion (30), where the reference bandwidth g is determined by ITC.

The following paragraphs examine how the several methods compare. Results for the small sample ($n = 120$) and the larger sample ($n = 480$) case are discussed separately. For $n = 480$ only the iterative procedures, which seem to outperform the cross-validation-based ones, are compared.

4.1. Small Sample Results ($n = 120$)

The most adequate criterion among the cross-validation-based ones seems to be CVLL. It performs better than SES, CVMSE and CVLL2 in 4 out of 5 cases according to all three distance measures. Only for the AR(4) process with one sharp peak (see Fig. 2) the dominating cross-validation-based method is CVLL2 (see Tab. III), with SES,

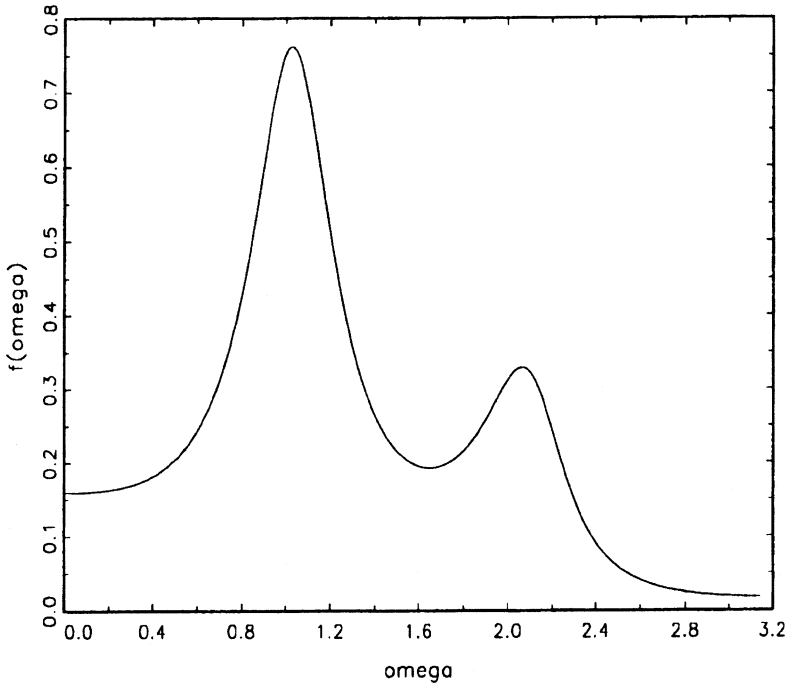

FIGURE 4 $X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + Z_t$.

CVMSE and CVLL scoring similarly compared to each other. This is due to the fact that CVLL2 in general produces rather high scale parameters h, yielding a low amount of smoothing. For a density with sharp peaks a low amount of smoothing, at least in the region of the peak, is in fact appropriate, as periodogram values that are further away from the considered frequency should not be given too much weight in the smoothing process.

Among the iterative procedures Bühlmann's (1996) original, ITB, performs well only in 2 out of 5 cases (see Tabs. I and IX). It is the worst criterion in another 2 out of 5 cases (see Tabs. III and V) and does quite badly in one more case (see Tab. VII). It seems that ITB performs well for densities with only broad bumps and no sharp peaks as is the case for the AR(1) process as in Figure 1 and the MA(5) process as in Figure 5. This may be due to the fact that the inflation factor tends to favor low scale parameters h, *i.e.*, a high amount of smoothing. For densities with broad bumps and no sharp

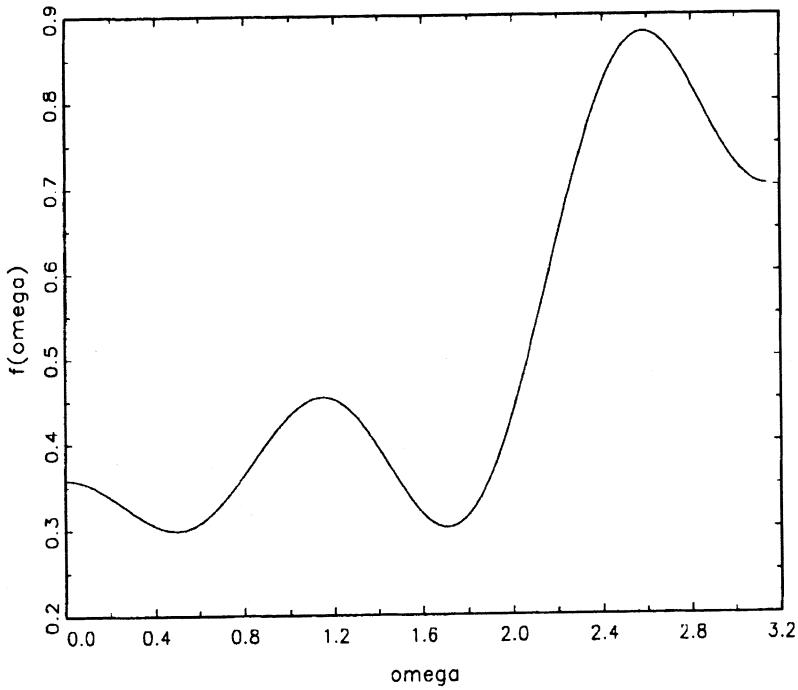

FIGURE 5 $X_t = Z_t + 0.9Z_{t-1} - 0.8Z_{t-2} + 0.6Z_{t-3} - 0.5Z_{t-4} + 0.3Z_{t-5}$.

TABLE I AR(1): $X_t = 0.8X_{t-1} + Z_t$, $Z_t \sim N(0,1)$, $(n = 120)$

Method	$MISE_0$	$MISE_{-2}$	$MISE_2$
CVLL	0.394 (0.158)	0.328 (0.048)	3.912 (21.258)
CVLL2	0.423 (0.282)	0.393 (0.074)	4.445 (39.435)
SES	0.511 (0.417)	0.424 (0.121)	5.349 (65.688)
CVMSE	0.413 (0.144)	0.363 (0.061)	4.089 (18.544)
ITA	0.384* (0.244)	0.308 (0.042)	4.108 (37.085)
ITA1	0.367 (0.081)	0.310 (0.035)	3.677 (9.597)
ITA2	0.351 (0.111)	0.295 (0.033)	3.643** (13.722)
ITB	0.385 (0.091)	0.306 (0.048)	3.715* (10.732)
ITB1	0.453 (0.061)	0.473 (0.090)	4.246 (8.626)
ITB2	0.373 (0.082)	0.325 (0.037)	3.696 (9.678)
ITC	0.391 (0.269)	0.303* (0.040)	4.255 (41.921)
ITC1	0.373 (0.085)	0.321 (0.034)	3.755 (10.327)
ITC2	0.350 (0.115)	0.279 (0.027)	3.651 (14.668)
BOOT	0.344** (0.099)	0.264** (0.023)	3.741 (14.442)

Variances are in brackets.
* Indicates the best global estimate.
** Points out the best semi-local estimate.

TABLE II AR(1): $X_t = 0.8X_{t-1} + Z_t$, $Z_t \sim N(0,1)$, $(n = 480)$

Method	$MISE_0$	$MISE_{-2}$	$MISE_2$
ITA	0.2265 (0.0744)	0.1804 (0.0052)	2.5255 (14.5014)
ITA1	0.1511 (0.0184)	0.0898 (0.0017)	1.6911 (3.0110)
ITA2	0.1605 (0.0278)	0.1165 (0.0021)	1.7677 (4.9347)
ITB	0.1434* (0.0163)	0.0782* (0.0016)	1.5675* (2.4446)
ITB1	0.1789 (0.0122)	0.1124 (0.0024)	1.8399 (2.0646)
ITB2	0.1397** (0.0154)	0.0867** (0.0015)	1.5373** (2.3410)
ITC	0.1680 (0.0472)	0.1071 (0.0028)	1.9056 (9.0932)
ITC1	0.1505 (0.0168)	0.0901 (0.0017)	1.6400 (2.7457)
ITC2	0.1462 (0.0237)	0.0890 (0.0017)	1.6223 (4.1144)

Variances are in brackets.
* Indicates the best global estimate.
** Points out the best semi-local estimate.

TABLE III AR(4): $X_t = 0.4X_{t-1} - 0.5X_{t-2} + 0.3X_{t-4} + Z_t$, $Z_t \sim N(0,1)$, $(n = 120)$

Method	$MISE_0$	$MISE_{-2}$	$MISE_2$
CVLL	0,2519 (0,0201)	0,6106 (0,1556)	1,352 (0,8559)
CVLL2	0,2125* (0,031)	0,5591 (0,1017)	1,0652* (0,8543)
SES	0,249 (0,035)	0,6662 (0,3081)	1,3258 (1,3153)
CVMSE	0,2593 (0,0239)	0,6736 (0,213)	1,389 (0,919)
ITA	0,2142 (0,0218)	0,5043 (0,0924)	1,121 (0,7175)
ITA1	0,2816 (0,0093)	0,688 (0,2006)	1,6 (0,7039)
ITA2	0,2323 (0,0118)	0,5053 (0,0841)	1,3137 (0,6714)
ITB	0,3523 (0,008)	1,5519 (2,4518)	1,9544 (0,6406)
ITB1	0,4103 (0,0052)	2,6002 (4,8588)	2,2803 (0,5484)
ITB2	0,3293 (0,0073)	1,0527 (0,7048)	1,8474 (0,6429)
ITC	0,2129 (0,0223)	0,5007* (0,0843)	1,1144 (0,7339)
ITC1	0,2722 (0,0103)	0,6229 (0,147)	1,5583 (0,7461)
ITC2	0,2282 (0,0122)	0,4886 (0,0784)	1,2943 (0,6929)
BOOT	0,2067** (0,015)	0,4287** (0,0642)	1,2209** (0,7362)

Variances are in brackets.
* Indicates the best global estimate.
** Points out the best semi-local estimate.

peaks this seems to be appropriate. ITA and ITC perform similarly well, possibly with a slight advantage for ITC.

The local iterative procedures compare to each other in a similar fashion as do their global counterparts. Bühlmann's original, ITB1, and also ITB2, are in general worse than the other procedures. Our results suggest that using no inflation factor is superior to using one, except for the MA(5) process as in Figure 5 (see Tab. IX). Again, ITA2 and ITC2 perform similarly well, with a slight advantage for ITC2.

TABLE IV AR(4): $X_t = 0.4X_{t-1} - 0.5X_{t-2} + 0.3X_{t-4} + Z_t$, $Z_t \sim N(0,1)$, $(n = 480)$

Method	$MISE_0$	$MISE_{-2}$	$MISE_2$
ITA	0.0947 (0.0062)	0.2082 (0.0047)	0.5187 (0.2716)
ITA1	0.1052 (0.0036)	0.1427** (0.0029)	0.6495 (0.2198)
ITA2	0.0909** (0.0037)	0.1542 (0.0028)	0.5360* (0.1936)
ITB	0.1314 (0.0031)	0.1796 (0.0055)	0.7819 (0.2278)
ITB1	0.2014 (0.0026)	0.3887 (0.0211)	1.1644 (0.2408)
ITB2	0.1246 (0.0031)	0.1672 (0.0036)	0.7626 (0.2187)
ITC	0.0874* (0.0054)	0.1723* (0.0040)	0.4919* (0.2394)
ITC1	0.1184 (0.0039)	0.1575 (0.0037)	0.7257 (0.2430)
ITC2	0.0917 (0.0037)	0.1455 (0.0029)	0.5458 (0.1995)

Variances are in brackets.
* Indicates the best global estimate.
** Points out the best semi-local estimate.

TABLE V AR(4): $X_t = 0.4X_{t-2} - 0.4X_{t-4} + Z_t$, $Z_t \sim N(0,1)$, $(n = 120)$

Method	$MISE_0$	$MISE_{-2}$	$MISE_2$
CVLL	0.0175 (0.00010)	0.398 (0.079)	0.00227 (0.000002)
CVLL2	0.0226 (0.00025)	0.442 (0.080)	0.00292 (0.000005)
SES	0.0209 (0.00021)	0.447 (0.097)	0.00284 (0.000006)
CVMSE	0.0179 (0.00009)	0.409 (0.073)	0.00229 (0.000002)
ITA	0.0169 (0.00012)	0.347 (0.052)	0.00215 (0.000002)
ITA1	0.0166 (0.00005)	0.384 (0.051)	0.00222 (0.000002)
ITA2	0.0163 (0.00006)	0.350 (0.039)	0.00213** (0.000001)
ITB	0.0231 (0.00006)	0.774 (0.259)	0.00302 (0.000002)
ITB1	0.0270 (0.00005)	0.837 (0.393)	0.00372 (0.000002)
ITB2	0.0239 (0.00006)	0.584 (0.122)	0.00335 (0.000002)
ITC	0.0166* (0.00010)	0.345* (0.052)	0.00213* (0.000002)
ITC1	0.0180 (0.00005)	0.415 (0.054)	0.00245 (0.000002)
ITC2	0.0162 (0.00006)	0.352 (0.041)	0.00213** (0.000001)
BOOT	0.0161** (0.00006)	0.329** (0.0359)	0.00227 (0.000002)

Variances are in brackets.
* Indicates the best global estimate.
** Points out the best semi-local estimate.

In general, local procedures seem to be an improvement on the global ones with the exception of the AR(4) process with a single sharp peak (see Fig. 2 and Tab. III). This somewhat comes as a surprise, because one would intuitively expect that it is favorable to employ a considerably lower degree of smoothing in the region of the peak than in the flat regions. Apparently, this is not optimal, though.

We note that in all the cases both ITC2 and the bootstrap criterion score at least as well as CVLL and most of the times even better, according to all three distance measures. For the AR(4) process with

TABLE VI AR(4): $X_t = 0.4X_{t-2} - 0.4X_{t-4} + Z_t$, $Z_t \sim N(0,1)$, $(n = 480)$

Method	$MISE_0$	$MISE_{-2}$	$MISE_2$
ITA	0.0112 (2.50E−05)	0.1885 (0.0042)	0.00160 (8.00E−07)
ITA1	0.0065 (7.00E−06)	0.1148 (0.0022)	0.00094 (3.00E−07)
ITA2	0.0071 (1.00E−05)	0.1251 (0.0022)	0.00098 (3.00E−07)
ITB	0.0066* (9.00E−06)	0.1156 (0.0038)	0.00099 (3.00E−07)
ITB1	0.0100 (1.70E−05)	0.1796 (0.0065)	0.00149 (6.00E−07)
ITB2	0.0063 (8.00E−06)	0.1053** (0.0021)	0.00094 (3.00E−07)
ITC	0.0068 (1.60E−05)	0.1145* (0.0028)	0.00097* (5.00E−07)
ITC1	0.0067 (8.00E−06)	0.1165 (0.0025)	0.00099 (3.00E−07)
ITC2	0.0062** (9.00E−06)	0.1088 (0.0021)	0.00089** (3.00E−07)

Variances are in brackets.
* Indicates the best global estimate.
** Points out the best semi-local estimate.

TABLE VII AR(5): $X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + Z_t$, $Z_t \sim N(0,1)$, $(n = 120)$

Method	$MISE_0$	$MISE_{-2}$	$MISE_2$
CVLL	0.0262 (0.0004)	0.4560 (0.1205)	0.0077 (0.00006)
CVLL2	0.0351 (0.0010)	0.4743 (0.0876)	0.0100 (0.00014)
SES	0.0323 (0.0006)	0.7449 (1.2001)	0.0100 (0.00010)
CVMSE	0.0281 (0.0009)	0.4754 (0.1405)	0.0083 (0.00011)
ITA	0.0253* (0.0005)	0.3698* (0.0617)	0.0072* (0.00006)
ITA1	0.0240 (0.0002)	0.4205 (0.0818)	0.0071 (0.00003)
ITA2	0.0238** (0.0002)	0.3763 (0.0551)	0.0069** (0.00003)
ITB	0.0318 (0.0001)	1.5179 (2.2244)	0.0095 (0.00003)
ITB1	0.0366 (0.0001)	0.9600 (0.8660)	0.0114 (0.00003)
ITB2	0.0300 (0.0001)	0.6815 (0.3390)	0.0090 (0.00003)
ITC	0.0260 (0.0005)	0.4015 (0.0906)	0.0076 (0.00008)
ITC1	0.0258 (0.0002)	0.4890 (0.1407)	0.0078 (0.00003)
ITC2	0.0238** (0.0003)	0.3843 (0.0596)	0.0070 (0.00004)
BOOT	0.0243 (0.0003)	0.3705** (0.0544)	0.0075 (0.00004)

Variances are in brackets.
* Indicates the best global estimate.
** Points out the best semi-local estimate.

a single sharp peak, however, (Fig. 2) CVLL2 performs at least equally well as both ITC2 and the bootstrap method.

Considering the fact that the bootstrap criterion is far more costly to evaluate than the iterative procedures, we suggest, on ground of our simulation study, to employ ITC2 for finding the optimal window width in non-parametric density estimation. ITC2 is an iterative method according to Bühlmann's (1996) scheme, but without using inflation factors, neither in the global steps nor in the semi-local step.

TABLE VIII AR(5): $X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + Z_t$, $Z_t \sim N(0, 1)$, $(n = 480)$

Method	$MISE_0$	$MISE_{-2}$	$MISE_2$
ITA	0.0157 (6.20E−05)	0.1843 (0.0042)	0.0045 (1.00E−05)
ITA1	0.0094 (2.20E−05)	0.1145 (0.0041)	0.0029 (5.00E−06)
ITA2	0.0104 (3.00E−05)	0.1240 (0.0020)	0.0030 (5.00E−06)
ITB	0.0125 (3.70E−05)	0.1931 (0.0169)	0.0042 (8.00E−06)
ITB1	0.0180 (4.00E−05)	0.2679 (0.0160)	0.0058 (1.10E−05)
ITB2	0.0112 (3.90E−05)	0.1280 (0.0043)	0.0037 (8.00E−06)
ITC	0.0098* (3.50E−05)	0.1133* (0.0025)	0.0029* (6.00E−06)
ITC1	0.0098 (2.30E−05)	0.1222 (0.0034)	0.0032 (5.00E−06)
ITC2	0.0091** (2.20E−05)	0.1090** (0.0020)	0.0028** (5.00E−06)

Variances are in brackets.
* Indicates the best global estimate.
** Points out the best semi-local estimate.

TABLE IX MA(5): $X_t = Z_t + 0.9Z_{t-1} - 0.8Z_{t-2} + 0.6Z_{t-3} - 0.5Z_{t-4} + 0.3Z_{t-5}$, $Z_t \sim N(0, 1)$, $(n = 120)$

Method	$MISE_0$	$MISE_{-2}$	$MISE_2$
CVLL	0.0577 (0.0022)	0.2204 (0.0290)	0.0262 (0.0006)
CVLL2	0.0936 (0.0055)	0.3208 (0.0535)	0.0454 (0.0017)
SES	0.0691 (0.0053)	0.2387 (0.0418)	0.0334 (0.0018)
CVMSE	0.0625 (0.0025)	0.2372 (0.0299)	0.0288 (0.0008)
ITA	0.0818 (0.0044)	0.2748 (0.0383)	0.0383 (0.0012)
ITA1	0.0460 (0.0011)	0.1708** (0.0136)	0.0207 (0.0003)
ITA2	0.0604 (0.0020)	0.2118 (0.0187)	0.0277 (0.0006)
ITB	0.0478* (0.0007)	0.2051* (0.0168)	0.0213* (0.0003)
ITB1	0.0489 (0.0007)	0.2175 (0.0166)	0.0213 (0.0002)
ITB2	0.0478 (0.0008)	0.2000 (0.0155)	0.0213 (0.0003)
ITC	0.0640 (0.0028)	0.2190 (0.0259)	0.0301 (0.0008)
ITC1	0.0452** (0.0009)	0.1721 (0.0122)	0.0204** (0.0003)
ITC2	0.0529 (0.0015)	0.1897 (0.0162)	0.0242 (0.0004)
BOOT	0.0518 (0.0012)	0.1884 (0.0127)	0.0241 (0.0004)

Variances are in brackets.
* Indicates the best global estimate.
** Points out the best semi-local estimate.

4.2. Large Sample Results ($n = 480$)

ITB, ITB1 and ITB2 perform poorly in 2 out of 5 cases (see Tabs. IV and VIII). ITA is worse than ITC in all 5 cases and worse than ITB most of the times. ITB and ITB2 perform quite well in some cases, especially for the AR(1) process, where ITB2 seems to be the most adequate method. ITC2 is the most adequate procedure in 2 out of 5

TABLE X MA(5): $X_t = Z_t + 0.9Z_{t-1} - 0.8Z_{t-2} + 0.6Z_{t-3} - 0.5Z_{t-4} + 0.3Z_{t-5}$, $Z_t \sim N(0,1)$, $(n = 480)$

Method	$MISE_0$	$MISE_{-2}$	$MISE_2$
ITA	0.05529 (0.00058)	0.18844 (0.00450)	0.02654 (0.00020)
ITA1	0.02028 (0.00014)	0.07211 (0.00108)	0.00940 (0.00005)
ITA2	0.03221 (0.00024)	0.11190 (0.00193)	0.01520 (0.00008)
ITB	0.02111* (0.00007)	0.09361 (0.00163)	0.00914* (0.00003)
ITB1	0.02537 (0.00007)	0.11948 (0.00226)	0.01072 (0.00003)
ITB2	0.01981 (0.00006)	0.08491 (0.00163)	0.00866 (0.00002)
ITC	0.02702 (0.00030)	0.09132* (0.00231)	0.01299 (0.00010)
ITC1	0.01527** (0.00009)	0.05777** (0.00088)	0.00693** (0.00003)
ITC2	0.02080 (0.00017)	0.07198 (0.00127)	0.00982 (0.00006)

Variances are in brackets.
* Indicates the best global estimate.
** Points out the best semi-local estimate.

cases (see Tabs. VIII and VI) and not much worse than other methods in the other cases, except for the MA(5) process (see Tab. X), where ITC1 is clearly dominating.

5. CONCLUSION

In this paper we reviewed and compared various methods for determining optimal scale parameters for non-parametric lag or spectral window estimation of a spectral density of a stationary zero-mean process. These are cross-validation-based estimates following Hurvich (1985); Beltrão and Bloomfield (1987) and Hurvich and Beltrão (1990); iterative estimates following Bühlmann (1996); and a bootstrap estimate following Franke and Härdle (1992). The means of comparison was a simulation study performed for selected ARMA(5,5) processes with simulation size 300 and time series length 120 and 480. In the case of $n = 480$, only iterative methods were examined, for reasons of their comparable computational efficiency, and because there is no reason to assume that cross-validation-based procedures suddenly perform better than the iterative ones for larger sample sizes.

All three distance measures provide similar rankings of the optimal bandwidth selection criteria. It seems that best among the cross-validation methods in general is CVLL, possibly except for sharp peak

density processes. More adequate is the global iterative procedure ITC, using Bühlmann's (1996) scheme, disregarding the inflation factor. Local selection criteria, in general, seem to be an improvement on global criteria. Among the discussed methods, ITC2, a local selection criterion using Bühlmann's (1996) scheme, without inflation factors, and the bootstrap method, originated by Franke and Härdle (1992) and slightly corrected in this paper, seem to be the most adequate criteria for determining the optimal window width in non-parametric spectral density estimation.

In general terms, method ITC2 seems to be the most appropriate to use. First, it performs best in choosing the optimal window width in most of the cases and only slightly worse in the rest. Second, it is a computationally efficient method compared with the bootstrap procedure and the cross-validatory methods.

Acknowledgements

We are most grateful to Alexander Samarov for insightful discussions and suggestions that led to improvements of this work. We also thank Jürgen Franke, Peter Bühlmann, Robert Kunst, Manfred Deistler, and participants at the econometric research seminar, Institut für Höhere Studien, Vienna, for helpful comments.

References

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

Brockmann, M., Gasser, T. and Herrmann, E. (1993) Locally adaptive bandwidth choice for kernel regression estimators. *J. Am. Statist. Assoc.*, **88**, 1302–9.

Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*, 2nd edn., Springer, New York.

Bühlmann, P. (1996) Locally adaptive lag-window spectral estimation. *J. Time Ser. An.*, **17**(3), 247–70.

Franke, J. and Härdle, W. (1992) On bootstrapping kernel spectral estimates. *Ann. Statist.*, **20**(1), 121–45.

Hurvich, C. M. (1985) Data-driven choice of a spectrum estimate: Extending the applicability of cross-validation methods. *J. Am. Statist. Assoc.*, **80**, 933–40.

Hurvich, C. M. and Beltrão, K. I. (1990) Cross-validatory choice of a spectrum estimate and its connection with AIC. *J. Time Ser. An.*, **11**, 121–37.

Marron, J. S. (1985) An asymptotically efficient solution to the bandwidth problem of kernel density estimation. *Ann. Statist.*, **13**, 1011–23.

Palmer, F. (1983) Adaptive spectral smoothing. *Technical report ONR 30*, Massachusetts Institute of Technology, Statistics Center.

Parzen, E. (1983) Autoregressive spectral estimation. In: *Handbook of Statistics* (Eds.) Brillinger, D. R. and Krishnaiah, P. R., New York: Elsevier, **3**, 221–47.
Priestley, M. B. (1981) *Spectral Analysis and Time Series 1*, Academic, New York.
Wahba, G. and Wold, S. (1975) Periodic splines for spectral density estimation: the use of cross validation for determining the degree of smoothing. *Communications in Statistics*, **4**, 125–141.