

IHS Economics Series
Working Paper 62
February 1999

Optimal Bandwidth Selection in Non-Parametric Spectral Density Estimation: Review and Simulation

Ines Fortin
Christoph Kuzmics



INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna



INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

Impressum

Author(s):

Ines Fortin, Christoph Kuzmics

Title:

Optimal Bandwidth Selection in Non-Parametric Spectral Density Estimation: Review and Simulation

ISSN: Unspecified

1999 Institut für Höhere Studien - Institute for Advanced Studies (IHS)

Josefstädter Straße 39, A-1080 Wien

E-Mail: office@ihs.ac.at

Web: www.ihs.ac.at

All IHS Working Papers are available online: http://irihs.ihs.ac.at/view/ihs_series/

This paper is available for download without charge at: <http://irihs.ihs.ac.at/1139/>

Contents

1	Introduction	1
2	The Basics	2
2.1	The Spectral Density	2
2.2	The Periodogram	2
2.3	Smoothing the Periodogram – Kernel Estimates	3
3	Cross-validation Methods	9
3.1	Beltrão & Bloomfield 1987	9
3.2	Hurvich 1985	11
3.3	Hurvich & Beltrão 1990	14
4	An Iterative Procedure	18
5	A Bootstrap Approach	23
6	Simulation	25
7	Conclusion	37
	Bibliography	39

1 Introduction

The crucial step in nonparametric 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 determine this scale parameter optimally one might try to minimize some measure of the distance between the true spectral density of a process and its estimator over a the range of scale parameters. Different theoretical criteria were proposed for that, among them the mean square error (MSE) and the mean square percentage error (MSPE). The resulting optimal value for the scale parameter depends, however, on the true underlying spectral density. To overcome this problem, various approaches have been undertaken. The earlier way to go about it, is to use a likelihood motivated cross-validation criterion, which may be seen as an estimate of some distance measure (Hurvich (1985), Beltrão & Bloomfield (1987), Hurvich & Beltrão (1990)). Another line of thought has been taken up by Franke & Härdle (1992). They consider bootstrap estimates of some distance measure by resampling the residuals of a multiplicative nonparametric regression, which can be shown to be 'nearly' independent. A third variant to tackle the problem has been developed by Bühlmann (1996), who iteratively estimates the spectral density, calculating the optimal scale parameter in every step according to one of the theoretical criteria, plugging in the (step-) spectral density estimate for the true spectral density.

The paper is organized as follows. Section 2 gives basic definitions and some well-known results in spectral density estimation. Cross-validation criteria for determining the optimal window width, as developed by Hurvich (1985), Beltrão & Bloomfield (1987) and Hurvich & Beltrão (1990), are discussed in section 3. Section 4 gives an account of the iterative approach formulated by Bühlmann (1996), and section 5 deals with the bootstrap method employed by Franke & Härdle (1992). A small simulation study for a comparison of the discussed methods is presented in section 6. Section 7 concludes.

2 The Basics

2.1 The Spectral Density

Definition 2.1

Let $\{X_t, t \in \mathbb{N}\}$ be a real(-valued) strictly stationary stochastic process with autocovariance function $\gamma(\cdot)$ and zero expectation.¹ The **spectral distribution function** $F(\omega)$ of $\gamma(\cdot)$ or of the process is defined as a right-continuous, non-decreasing, bounded function on $[-\pi, \pi]$ with $F(-\pi) = 0$ satisfying

$$\gamma(k) = \int_{(-\pi, \pi)} e^{ik\nu} dF(\nu) \text{ for all } k = 0, \pm 1, \dots \quad (1)$$

Definition 2.2

The **spectral density function** $f(\omega)$ is defined by

$$F(\omega) = \int_{-\pi}^{\omega} f(\nu) d\nu, \quad \omega \in [-\pi, \pi] \quad (2)$$

and may be written as

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\omega} \gamma(k) \text{ for all } \omega \in [-\pi, \pi]. \quad (3)$$

The autocovariance function, therefore, is an inverse Fourier transform of the spectral density function, and vice versa the spectral density function is a Fourier transform of the autocovariance function. The autocovariance function, however, by nature is only defined on the set of integers.

Remark 2.1 By a Fourier transform $F(y)$ of a function $f(x)$ we understand any integral of or discrete sum over this function or its discrete values multiplied by e^{-iyx} with respect to x . We do not care about the constant in front of the integral or sum. Sometimes we call the discrete version a discrete Fourier transform if we think it necessary to stress that point. If, on the other hand, the multiplying factor in the integral or sum is e^{iyx} we call the result an inverse Fourier transform.

2.2 The Periodogram

Let x_1, \dots, x_n be the sample of a real, strictly stationary stochastic process. Let the Fourier frequencies of the sample be defined as $\omega_j = \frac{2\pi j}{n}$, where j assumes integer values such that $-\pi < \omega_j \leq \pi$. Let the sample estimate of the autocovariance function,² denoted by $\hat{\gamma}(\cdot)$, be given by

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (x_{t+k} - m)(x_t - m) \text{ for } k \geq 0 \text{ and} \quad (4)$$

$$\hat{\gamma}(k) = \hat{\gamma}(-k) \text{ for } k < 0, \text{ where} \quad (5)$$

$$m = \frac{1}{n} \sum_{t=1}^n x_t. \quad (6)$$

¹We will assume this throughout the paper.

²This estimate is biased downwards for all k , the larger $|k|$ the larger the bias.

Then this sample version of the autocovariance function yields an intuitive estimate of the spectral density by replacing $\gamma(\cdot)$ in the definition of the spectral density, as in (3), by its estimate $\hat{\gamma}(\cdot)$.

$$I(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} e^{-ik\omega} \hat{\gamma}(k) \text{ for all } \omega \in [-\pi, \pi]. \quad (7)$$

$I(\cdot)$ is called the periodogram and is very often defined on Fourier frequencies only. It is easy to see that the periodogram may also be stated in terms of the observations. (Brockwell & Davis p. 332)

$$I(\omega) = \frac{1}{n} \left| \sum_{t=1}^n x_t e^{-it\omega} \right|^2. \quad (8)$$

The periodogram $I(\cdot)$ is therefore the square of the absolute value of the discrete Fourier transform of the data x_1, \dots, x_n .

Remark 2.2 *The periodogram is apparently defined differently by different authors. Especially the constant in front of the sum differs a lot.*

2.3 Smoothing the Periodogram – Kernel Estimates

The periodogram is NOT a consistent estimator of the spectral density (Priestley p. 425) in the sense that $\text{Var}(I(\omega))$ does not converge to zero as $n \rightarrow \infty$. Also $I(\omega)$ does not converge to $f(\omega)$, the true density, in mean square. A smoothed version of the periodogram, though, may be shown, under some conditions, to be a mean square consistent estimate of the true spectral density. Estimators of the form,

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \lambda(k) \hat{\gamma}(k) e^{-ik\omega}, \quad (9)$$

(Brockwell & Davis p. 354, Priestley p. 434 6.2.54) where $\lambda(\cdot)$ is a so-called lag window, are generally called **Lag Window Estimators**. For the moment let the lag window be some general function that maps \mathbb{R} into the positive real line with some assumptions that will guarantee consistency of the above lag window estimator. Typically we would think of a lag window as an even function with one single maximum at zero and decaying smoothly and fast enough (to yield a consistent estimate) as the argument becomes greater in absolute values. As we will later consider only a special type of lag window estimators, namely scale parameter windows, we will not go into detail about the general conditions on the window to yield consistent estimates of the spectral density. Some calculation (Priestley p. 435 6.2.56) shows that, using properties of Fourier transforms, the lag window density estimate may be written as a weighted average of nearby periodogram values,

$$\hat{f}(\omega) = \int_{-\pi}^{\pi} I(\theta) W(\omega - \theta) d\theta, \quad (10)$$

where $W(\cdot)$, the **spectral window**, is the discrete Fourier transform of the corresponding lag window.

$$W(\theta) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{(n-1)} \lambda(k) e^{-ik\theta}. \quad (11)$$

The spectral density estimate at some specified frequency ω , therefore, is the weighted average of the periodogram values with the largest weight attached to ordinates in the neighborhood of $\theta = \omega$. (Priestley p. 435) For practical purposes instead of the integral in (10) we will rather use a discrete sum over all Fourier frequencies.

$$\hat{f}(\omega) \approx \frac{1}{2\pi} \sum_{j=-N}^N W(\omega - \omega_j) I(\omega_j), \quad (12)$$

where N is the largest integer less than or equal to $\frac{n-1}{2}$.

There are lots of different possible lag windows that would fulfill the conditions to obtain a consistent estimate of the spectral density (Priestley, p. 434). A rather convenient type of lag windows are the scale parameter windows (Priestley, p. 446). These involve a parameter, the scale parameter, that in some obvious way controls for the width of the window. Say, a lag window family is given as a function of a scale parameter h that controls for the width of the window and of k , $\lambda(k; h)$. Then

Definition 2.3

$\lambda(\cdot)$ is a **scale parameter window** if it can be written in the form,

$$\lambda(k; h) = \kappa(k/h), \quad (13)$$

where $\kappa(\cdot)$ is a lag window generator or lag kernel and h the scale parameter.

Definition 2.4

A **lag kernel** or lag window generator is an even function $\kappa : \mathbb{R} \rightarrow \mathbb{R}^+$ with

$$\kappa(0) = 1, \quad (14)$$

$$|\kappa(x)| \leq 1, \quad \text{for all } x, \quad (15)$$

and

$$\kappa(x) = 0, \quad \text{for } x > h. \quad (16)$$

The scale parameter h is very often, a bit confusingly, called the 'bandwidth' of the lag window and describes the shape or concentration of the lag window. Confusingly, as on the one hand it in fact has got something to do with the term bandwidth defined differently by various authors (cf. Priestley, p. 520 ff.) as it also measures or controls for the width of a window, but on the other hand in general it is not the same as the bandwidth in any of the given definitions.

For spectral window density estimation the lag kernel transforms into a spectral kernel.

Definition 2.5

The corresponding **spectral kernel** or *spectral window generator* for a given lag kernel is given by its Fourier transform,

$$K(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \kappa(x)e^{-ix\omega} dx. \tag{17}$$

The spectral window may then at least approximately be written as

$$W(\omega) \approx hK(h\omega). \tag{18}$$

For the Bartlett-Priestley window this holds exactly true.

As already mentioned some windows, like the Lomnicki-Zaremba window (Priestley, p. 445), cannot be put into this scale parameter window generating framework. Examples of some lag windows that can be put into the kernel framework are the Bartlett or Triangular Window and the Bartlett-Priestley Window. Their respective lag and spectral kernels are given by,

$$\kappa(x) = \begin{cases} 1 - |x| & \text{if } |x| \leq 1, \\ 0 & \text{if } |x| > 1, \end{cases} \tag{19}$$

with corresponding spectral kernel (Fejer kernel)

$$K(\omega) = \frac{1}{2\pi} \left(\frac{\sin(\omega/2)}{\omega/2} \right)^2 \tag{20}$$

for the Bartlett window and

$$\kappa(x) = \frac{3}{\pi^2 x^2} \left\{ \frac{\sin(\pi x)}{\pi x} - \cos(\pi x) \right\} \tag{21}$$

with

$$K(\omega) = \begin{cases} \frac{3}{4\pi} \left(1 - \left(\frac{\omega}{\pi} \right)^2 \right) & \text{if } |\omega| \leq \pi \\ 0 & \text{if } |\omega| > \pi. \end{cases} \tag{22}$$

for the Bartlett-Priestley Window (see Priestley, p. 447/8).

To better distinguish the various windows with respect to some crucial aspects, we formulate the following definitions.

Definition 2.6

A lag window is a **C^r window** if its lag kernel is r times continuously differentiable in the neighborhood of zero and Lipschitz-continuous on \mathbb{R} .

Definition 2.7

A window λ has **characteristic exponent r** if its lag kernel has the properties

$$\kappa^{(s)} = 0 \quad \text{for all } s < r \tag{23}$$

$$\kappa^{(r)} \neq 0, \tag{24}$$

where

$$\kappa^{(s)} = \lim_{x \rightarrow 0} \left(\frac{1 - \kappa(x)}{|x|^s} \right) \tag{25}$$

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). The generalized derivative is NOT the same as the (ordinary) derivative of $\kappa(k)$ at zero. For even s , however, there is an obvious relationship between the two.

$$\kappa^{(s)} = -\frac{1}{s!} \left(\frac{d^s(\kappa(x))}{dx^s} \right)_{x=0}. \quad (26)$$

The Bartlett window, therefore, would be a C^0 window with characteristic exponent 1, whereas the Bartlett-Priestley window would be C^2 with characteristic exponent 2.

Given a specific window one still has to choose the scale parameter. Different scale parameters yield very 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 distance between the estimator and the true spectral density. Various different measures were suggested in the literature (Priestley, p. 510 ff.). We will only mention two.

Definition 2.8

The **mean square error** or **MSE** of a spectral density estimate \hat{f} at a fixed frequency ω is given by

$$MSE(\hat{f}, \omega) = E \left(\hat{f}(\omega) - f(\omega) \right)^2. \quad (27)$$

Definition 2.9

The **mean square percentage error** or **MSPE** of a spectral density estimate \hat{f} at a fixed frequency ω is given by

$$MSPE(\hat{f}, \omega) = E \left(\frac{\hat{f}(\omega) - f(\omega)}{f(\omega)} \right)^2. \quad (28)$$

Minimizing one of the two above criteria would produce an optimal **local** scale parameter. That is, optimally, the scale parameter and therefore the window will in general be different for different frequencies. If one would like to employ the same smoothing window for the whole spectrum, one may choose to select the scale parameter such as to minimize the integrated version of the above criteria.

Definition 2.10

The **mean integrated square error** or **MISE** of a spectral density estimate \hat{f} is given by

$$MISE(\hat{f}) = E \int_{-\pi}^{\pi} \left(\hat{f}(\omega) - f(\omega) \right)^2 d\omega. \quad (29)$$

Definition 2.11

The **mean integrated square percentage error** or **MISPE** of a spectral density estimate \hat{f} is given by

$$MISPE(\hat{f}) = E \int_{-\pi}^{\pi} \left(\frac{\hat{f}(\omega) - f(\omega)}{f(\omega)} \right)^2 d\omega. \quad (30)$$

MISE and MISPE are just two more or less arbitrary definitions of a global distance between the true spectral density function and its estimate. Other measures are possible (see e. g. Hurvich (1985), $MISE_3$). The measure to be employed should be determined by the qualities the induced measure-minimizing estimate is wanted to have. The difference between MISE and MISPE is the weight they put on different frequencies with differently high values of the true spectral density. Whereas MISE gives equal weight to all frequencies, MISPE will depend a lot on the fit at frequencies ω with low spectral density value $f(\omega)$, as $f(\omega)$ is in the denominator of MISPE. If we would like to have a good estimate of the spectral density at peaks especially, one should probably not consider MISPE as the appropriate criterion, but rather MISE or even another distance where the squared distance between true and estimated density is multiplied (instead of divided) by some positive-valued, monotone increasing function of $f(\omega)$, the value of the true spectral density. E. g.

$$MISME(\hat{f}) = E \int_{-\pi}^{\pi} (\hat{f}(\omega) - f(\omega))^2 f^a(\omega) d\omega, \quad (31)$$

for some positive a (maybe 1 or 2).

Considering local distance measures (MSE, MSPE or others) we realize that it will not matter which one of them we will choose as long as, at a fixed frequency, one measure is just a monotone transformation of the other, which is the case for MSE and MSPE. We will now only consider minimizing the MSE. By the usual variance decomposition the MSE can be written as the sum of the squared bias and the variance.

$$E (\hat{f}(\omega) - f(\omega))^2 = \left(E (\hat{f}(\omega)) - f(\omega) \right)^2 + E (\hat{f}(\omega) - E (\hat{f}(\omega)))^2 \quad (32)$$

$$MSE(\hat{f}, \omega) = \text{BIAS}^2(\hat{f}, \omega) + \text{VAR}(\hat{f}, \omega). \quad (33)$$

For various lag or spectral windows, given the true spectral density, it is possible to at least asymptotically assess bias and variance of the corresponding estimators. We will again closely follow Priestley (p. 457 ff.). The bias as well as the variance generally will depend on the spectral density and its (generalized) derivatives and the form of the lag (or spectral) window.

Definition 2.12

The s -th generalized derivative of a spectral density $f(\omega)$ is given by

$$f^{(s)}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} |k|^s \gamma(k) e^{-ik\omega}. \quad (34)$$

This generalized derivative is again NOT the same as the (ordinary) derivative of $f(\omega)$, but for s even there again is an obvious relationship between the two.

$$f^{(s)}(\omega) = (-1)^{\frac{s}{2}} \left(\frac{d}{d\omega} \right)^s f(\omega). \quad (35)$$

Priestley (p. 459, 6.2.124) derives an asymptotic formula for the bias of a kernel spectral estimate³

$$\text{ABIAS}(h, \omega) = -\frac{1}{h^r} \kappa^{(r)} f^{(r)}(\omega), \quad (36)$$

and for the variance (Priestley p. 457, 6.2.113),

$$\text{AVAR}(h, \omega) = \iota(\omega) \frac{h}{n} f^2(\omega) \int_{-\infty}^{\infty} \kappa^2(x) dx, \quad (37)$$

where r is the characteristic exponent of a lag window generated by kernel κ and where

$$\iota(\omega) = \begin{cases} 2 & \text{if } \omega \in \{-\pi, 0, \pi\} \\ 1 & \text{otherwise.} \end{cases} \quad (38)$$

For general C^2 windows with characteristic exponent 2 (I think Bühlmann implicitly assumes that or forgot about it) these are given by (see Bühlmann, p. 249, equation 4).

$$\text{ABIAS}(h, \omega) = -\frac{1}{2h^2} f^{(2)}(\omega) \kappa''(0), \quad (39)$$

$$\text{AVAR}(h, \omega) = \iota(\omega) \frac{h}{n} f^2(\omega) \int_{-\infty}^{\infty} \kappa^2(x) dx. \quad (40)$$

Minimizing the sum of the squared asymptotic bias and the asymptotic variance over h will yield the optimal asymptotic local scale parameter (see section 4 and 5). This optimal value for h does, however, depend on the true spectral density and its (generalized) derivatives. As we do not know the true spectrum we will have to adopt some kind of estimation of either the MISE directly (see section 3) or the MISE indirectly by iteratively estimating the true density and its generalized derivatives (see section 4).

³The different estimators are represented by the scale parameter. Therefore we replace \hat{f} in expressions like $\text{BIAS}(\hat{f}, \omega)$ by the scale parameter h .

3 Cross-validation Methods

Beltrão & Bloomfield (1987)⁴ provide the first objective criterion for the selection of the window width in the area of cross-validation methods. They argue that, by minimizing a cross-validators version of the log-likelihood function (CVLL), one will also minimize the mean square integrated error, which is what they propose as a theoretical figure of merit for a spectrum estimate. The class of potential estimates they consider is still limited, though, as it only contains non-parametric estimates. This is a restriction which is relaxed by Hurvich (1985). He extends the class of candidate estimates to include any estimate derived from the observed data. In particular, the class of estimates now includes both Yule-Walker and periodogram-based type estimates. As Hurvich still wants to use Beltrão & Bloomfield's technique for the automatic smoothness parameter selection, he defines a leave-out-one spectrum version for any candidate estimate.⁵ Apart from the mean square integrated error used by Beltrão & Bloomfield, Hurvich examines two other distance measures to assess the quality of spectrum estimates. His main contribution, though, is certainly the introduction of a method that allows for simultaneous and objective choice of both a type of estimate and the corresponding smoothness parameter.

The last part of this section deals with computational efficiency in cross-validators spectral density estimation. While Hurvich uses the generalized CVLL to simultaneously determine the window width for non-parametric density estimation and the order for parametric density estimation, Hurvich & Beltrão (1990) suggest the use of CVLL for non-parametric density estimation only and the use of the computationally more efficient Akaike information criterion (AIC) for parametric density estimation. Hurvich & Beltrão motivate this procedure by showing that CVLL can in fact be seen as a cross-validators generalization of AIC. Finally, they suggest a computationally more efficient non-cross-validators version of CVLL for non-parametric estimates.

3.1 Beltrão & Bloomfield 1987

In probability density estimation, which is a field closely related to spectral density estimation, Marron (1985) and others have already discussed the use of a cross-validated log-likelihood function to determine the window width. Beltrão & Bloomfield adopt a similar approach to Marron's and suggest to use a slightly adapted function to determine the smoothness parameter in spectral kernel density estimation. Beltrão & Bloomfield show that their cross-validated log-likelihood criterion is asymptotically equivalent to the mean integrated squared error.

⁴To avoid confusion, let us note that although the publication year of Beltrão & Bloomfield's paper is 1987, they put down their ideas on the topic in a working paper already in 1983, that is in particular before Hurvich published his 1985 paper.

⁵Thus, he obtains a new method for autoregressive order selection. Note, however, that this new method is computationally extremely inefficient as opposed to the use of Akaike's information criterion.

The class of estimates Beltrão & Bloomfield consider (non-parametric estimates) contains all estimates that are derived from the periodogram $I(\omega)$ of the observed data $\{x_t : 0 \leq t < n\}$, as defined in (8).

The estimates \hat{f} are obtained by a discrete averaging process, using kernel $K(\cdot)$ and a window width h :

$$\hat{f}(\omega_j, h) = \frac{1}{\sigma(h)} \sum_k K(h\omega_k) I(\omega_j - \omega_k), \quad (41)$$

where

$$\sigma(h) = \sum_k K(h\omega_k) \quad (42)$$

and the summations extend over all ω_k in the support of K ($h\omega_k < \pi$). To judge the quality of a spectrum estimate Beltrão & Bloomfield propose to use the mean square percentage error (MSPE), as defined by (28), which will be integrated to provide one single global measure for the spectrum estimate. For discrete samples Beltrão & Bloomfield define the mean integrated square percentage error (MISPE) by summing up the MSPE, rather than integrating it, over the Fourier frequencies, omitting the endpoints:

$$\text{MISPE} = E \frac{1}{N} \sum_{0 < \omega_j < \pi} \left(\frac{\hat{f}(\omega_j, h) - f(\omega_j)}{f(\omega_j)} \right)^2. \quad (43)$$

An approximate expression for minus twice the logarithm of the Gaussian likelihood function for spectrum f is given by⁶

$$A = \sum_{j=0}^N \log f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)}. \quad (44)$$

It can easily be seen that straightforward substitution for $f(\omega_j)$ is not useful, as function A is clearly minimized among all f by setting $f(\omega_j)$ equal to $I(\omega_j)$, which is done by choosing sufficiently small values of h in (41). So, one way to go about it is to construct a 'leave-out-one' or cross-validated form of A , where $I(\omega_j)$ is omitted in estimating $f(\omega_j)$

$$f^{-j}(\omega_j, h) = \frac{1}{\sigma_j(h)} \sum_{k \notin J(n, j)} K(h\omega_k) I(\omega_j - \omega_k), \quad (45)$$

where

$$\sigma_j(h) = \sum_{k \notin J(n, j)} K(h\omega_k) \quad (46)$$

and $J(n, j)$ is the set of indices k for which $I(\omega_j - \omega_k) = I(\omega_j)$.

⁶Note that while the likelihood function is always maximized in an estimation procedure, the approximation employed here, denoted cross-validated log-likelihood (CVLL), is being minimized.

Substitution of $\hat{f}^{-j}(\omega)$ for $f(\omega)$ in (44) and omitting some terms figuring twice leads to the cross-validated log-likelihood function.⁷

$$\text{CVLL}_{BB}(\hat{f}) = \sum_{0 < \omega_j < \pi} \log \hat{f}^{-j}(\omega_j, h) + \frac{I(\omega_j)}{\hat{f}^{-j}(\omega_j, h)}. \quad (47)$$

In the following large sample result Beltrão & Bloomfield show that choosing h to minimize CVLL is asymptotically equivalent to minimizing MISPE.

For large n

$$\frac{1}{N} \text{CVLL}_{BB}(\hat{f}) = \frac{1}{N} \sum_{0 < \omega_j < \pi} \left(\log f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)} \right) + \frac{1}{2} \text{MISPE} + o_P(\text{MISPE}). \quad (48)$$

Thus, one may approximately minimize MISPE by choosing h to minimize CVLL. Notice that, the first term on the right-hand side being constant for any h , one can also use the difference in $2N^{-1}\text{CVLL}$ for two different window widths, h_1 and h_2 , to estimate the additional MISPE that is incurred as a result of the change in the window width.

$$\frac{2}{N} (\text{CVLL}_{BB}(h_1) - \text{CVLL}_{BB}(h_2)) \approx \text{MISPE}(h_1) - \text{MISPE}(h_2). \quad (49)$$

The actual MISPE for a given h cannot be estimated, though, as we obviously do not know the true spectrum $f(\omega)$.

The proposal to determine h such as to minimize CVLL derived from (48) is investigated by a small simulation study. Beltrão & Bloomfield generate 100 samples of length 128 for an autoregressive and a moving-average model, respectively. The simulation results show that, on average, CVLL can be viewed as a good indicator of MISPE, even though variability across samples may be substantial.

3.2 Hurvich 1985

Hurvich extends the class of potential estimates used by Beltrão & Bloomfield to a bigger class of estimates by extending the applicability of existing cross-validatory techniques through the introduction of generalized leave-out-one spectrum estimates. His main purpose is to find objective smoothness parameter selection methods that allow for comparison of both autoregressive estimates and discrete periodogram average estimates.

Hurvich presents three different forms of cross-validatory methods: the cross-validated log-likelihood method of Beltrão & Bloomfield (1987), Stuetzle's smoothed estimate (SES, see Palmer (1983)) and an adaptation of the cross-validation mean squared error

⁷The cross-validated log-likelihood function is defined slightly differently by Beltrão & Bloomfield, Hurvich, and Hurvich & Beltrão, and it will be denoted CVLL_{BB} , CVLL_H and CVLL_{HB} , respectively. As the difference in definition is negligible for the maximization problem, we will still use the term CVLL whenever referring to the technique but not to one particular method.

(CVMSE) method of Wahba and Wold (1975). By introducing two generally applicable definitions of leave-out-one versions of the spectrum estimate he extends the applicability of the CVLL, SES and CVMSE techniques. Either of these definitions in conjunction with the CVLL, SES, or CVMSE method will yield an objective choice from a general class \mathcal{C} , where \mathcal{C} includes any estimate whose leave-out-one version is defined.

The distance measure, which Hurvich quite loosely denotes MISE, for the CVLL, SES, and CVMSE methods, respectively, are defined by⁸

$$\text{MISE}_1(\hat{f}) = E \frac{1}{N} \sum_{j=1}^N \left(\frac{\hat{f}(\omega_j) - f(\omega_j)}{f(\omega_j)} \right)^2, \quad (50)$$

$$\text{MISE}_2(\hat{f}) = E \frac{1}{N} \sum_{j=1}^N \left(\hat{f}(\omega_j) - f(\omega_j) \right)^2, \quad (51)$$

$$\text{MISE}_3(\hat{f}) = E \frac{1}{N} \sum_{j=1}^N \left(\log \hat{f}(\omega_j) - \log f(\omega_j) \right)^2. \quad (52)$$

The cross-validatory estimates of $\text{MISE}_i(\hat{f})$, for $i = 1, 2, 3$ are

$$\text{CVLL}_H(\hat{f}) = \frac{1}{N} \sum_{j=1}^N \log \hat{f}^{-j}(\omega_j) + \frac{I(\omega_j)}{\hat{f}^{-j}(\omega_j)}, \quad (53)$$

$$\text{SES}(\hat{f}) = \frac{1}{N} \sum_{j=1}^N (\hat{f}^{-j}(\omega_j) - I(\omega_j))^2, \quad (54)$$

$$\text{CVMSE}(\hat{f}) = \frac{1}{N} \sum_{j=1}^N \{(\log \hat{f}^{-j}(\omega_j) - (\log I(\omega_j) + C))^2 - \pi^2/6\} \quad (55)$$

where $C = 0.577216\dots$ is Euler's constant and $\hat{f}^{-j}(\omega_j)$ is a general leave-out-one (cross-validated) version of \hat{f} , such that $\hat{f}^{-j}(\omega_j)$ is approximately independent of $I(\omega_j)$ for each j . The independence is achieved by omitting $I(\omega_j)$ from the computation of $\hat{f}^{-j}(\omega_j)$.

In a first step, Hurvich defines the general leave-out-one spectrum estimate for any estimate that is a function of the sample autocovariances $\hat{\gamma}(k)$ as defined in (4). In particular, this class of estimates includes both all non-parametric estimates (lag window estimates and spectral window estimates) and the Yule-Walker autoregressive estimates.

Let any estimate of this class be written as $\hat{f}(\omega, \{\hat{\gamma}(k)\})$.

$$\begin{aligned} I^{-j}(\omega) &= I(\omega) && \omega \notin \{(\omega_{j-1}, \omega_{j+1}) \cup (\omega_{-j-1}, \omega_{-j+1})\} \\ &= \theta_{1,\omega} I(\omega_{j-1}) + \theta_{2,\omega} I(\omega_{j+1}) && \omega \in (\omega_{j-1}, \omega_{j+1}) \\ &= I^{-j}(-\omega) && \omega \in (\omega_{-j-1}, \omega_{-j+1}) \end{aligned} \quad (56)$$

⁸Actually, MISE_1 corresponds to what we defined as MISPE, and MISE_2 corresponds to what we defined as MISE.

for $\omega \in [-\pi, \pi]$, where

$$\theta_{1,\omega} = 1 - \frac{\omega - \omega_{j-1}}{\omega_{j+1} - \omega_{j-1}} \text{ and } \theta_{2,\omega} = \frac{\omega - \omega_{j-1}}{\omega_{j+1} - \omega_{j-1}}.$$

In general, the periodogram is only evaluated at the Fourier frequencies. If it is evaluated on a sufficiently fine grid, though, it completely determines the $\{\hat{\gamma}(k)\}$ sequence by

$$\hat{\gamma}(k) = \frac{2\pi}{n'} \sum_{k=0}^{n'-1} I(\omega'_k) e^{ir\omega'_k}, \quad (57)$$

where $n' = 2n$ and $\omega'_k = 2\pi k/n'$. Here the ω'_k are defined on a grid exactly twice as finely spaced as the Fourier frequencies. Hurvich then defines the sequence $\{\hat{\gamma}(k)^{-j}\}$ by

$$\hat{\gamma}(k)^{-j} = \frac{2\pi}{n'} \sum_{k=0}^{n'-1} I^{-j}(\omega'_k) e^{ir\omega'_k}. \quad (58)$$

Finally he defines the general leave-out-one version of the spectrum estimate $\hat{f}^{-j}(\omega_j)$ for $1 \leq j \leq N$ as follows:

$$\hat{f}^{-j}(\omega_j) = \hat{f}(\omega_j; \{\hat{\gamma}(k)^{-j}\}). \quad (59)$$

It is important to note that $\hat{f}^{-j}(\omega_j)$ and $I(\omega_j)$ will be approximately independent for each j , as the computation of $\hat{f}^{-j}(\omega_j)$ does not involve $I(\omega)$ for ω in the intervals $(\omega_{j-1}, \omega_{j+1})$ and $(\omega_{-j-1}, \omega_{-j+1})$.

Now, Hurvich defines a second general leave-out-one spectrum estimate which can be applied to any estimate whatsoever and is denoted by $\hat{f}(\omega; \{x_t\})$. First, he defines $\{J_k\}_{k=1}^n$, the Fourier transform of $\{x_t\}_{t=1}^n$, by

$$J_k = \frac{1}{n} \sum_{t=1}^n x_t e^{-i\omega_k t}. \quad (60)$$

This sequence completely determines the data sequence, through the relation

$$x_t = \sum_{k=1}^n J_k e^{i\omega_k t}. \quad (61)$$

Then, Hurvich defines the leave-out-one version of J_k , J_k^{-j} , for $1 \leq j \leq N$:

$$\begin{aligned} J_k^{-j} &= J_k && k \neq j, k \neq n-j \\ &= \frac{1}{2}(J_{k-1} + J_{k+1}) && k = j, k = n-j \end{aligned} \quad (62)$$

and the leave-out-one- ω_j version of the data sequence $\{x_t^{-j}\}_{t=0}^{n-1}$ by

$$x_t^{-j} = \sum_{k=1}^n J_k^{-j} e^{i\omega_k t}. \quad (63)$$

Finally, the general leave-out-one spectrum estimate is defined as:

$$\hat{f}^{-j}(\omega_j) = \hat{f}(\omega_j; \{x_t^{-j}\}). \quad (64)$$

Note that (59) and (64) do not coincide, even when \hat{f} can be written in terms of the $\{\hat{\gamma}(k)\}$ sequence.

In his simulation study Hurvich generates 40 samples of length 100 for a superposition of sinusoids with random phases and amplitudes, for an AR(3) process and an MA(3) process, respectively. The results suggest that if one does not have any indication as to the type of stochastic process, it seems unwise to arbitrarily choose one specific type of spectrum estimate. Hurvich argues that if the type of estimate and the corresponding smoothness parameter is determined objectively by minimizing the cross validatory log-likelihood over all candidate estimates, one will usually get much better estimates than one would get by (unluckily) making the wrong guess about the type of estimate, and almost as good ones as one would get by (luckily) making the right guess about the type of estimate. Thus, Hurvich proposes to apply a generalized type smoothness parameter selector, regardless of how well or poorly various restricted versions of the method do for particular processes.

Hurvich is already indicating that it would be nice to modify the generalized cross-validatory method to incorporate existing model order selection techniques like the AIC criterion. This is the subject of the next section.

3.3 Hurvich & Beltrão 1990

Hurvich & Beltrão (1990) motivate the use of a generalized version of the cross-validated log-likelihood criterion (CVLL) for selecting a spectrum estimate from an arbitrary class of estimates theoretically. It is shown that both CVLL and the non-cross validatory Akaike information criterion (AIC) are asymptotically equivalent to the Kullback-Leibler information. As the application of AIC is restricted to parametric estimates, CVLL can be viewed as a cross-validatory generalization of AIC.

To save computation time when the class of potential estimates includes both non-parametric and parametric (autoregressive) estimates, CVLL need only be evaluated for the non-parametric estimates, while the computationally more efficient AIC is evaluated for the parametric estimates. In this situation all criteria will be directly comparable, as CVLL and AIC both estimate the same information measure.

First, Hurvich & Beltrão show that the generalized CVLL is an estimate of the expected Kullback-Leibler information. CVLL can be written as the sum of two terms, a log estimated white noise variance and a random penalty term. In the case of parametric spectrum estimation, the first term of CVLL is identical to the first term of AIC, and the mean of the random penalty term of CVLL is asymptotically equivalent to the constant penalty term of AIC.

Then, Hurvich & Beltrão motivate the use of a computationally more efficient non cross-validatory version of CVLL, CVLL2, by showing that CVLL2 can be viewed as an approximately unbiased estimate of the Kullback-Leibler information.

A general way how to choose a specific spectrum estimate $\hat{f}(\omega) = \hat{f}(\{x_t\}_{t=1}^n, \omega)$ from a class of candidates \mathcal{C} is to construct a discrepancy function $d(f, g)$ such that $d(f, \hat{f}) \leq d(f, g)$ for all f, g , obtain an approximately unbiased estimate $\hat{d}(f, \hat{f})$ of $E\{d(f, \hat{f})\}$ which is applicable for all $\hat{f} \in \mathcal{C}$ and finally choose the estimate \hat{f} which minimizes $\hat{d}(f, \hat{f})$ over the class \mathcal{C} .

The discrepancy function Hurvich & Beltrão choose is

$$d(f, g) = 2n \log(2\pi) + \frac{n}{2\pi} \int_{-\pi}^{\pi} \left(\log g(\omega) + \frac{f(\omega)}{g(\omega)} \right) d\omega. \quad (65)$$

To understand the motivation to select d as a discrepancy function, note the following shown by Parzen (1983, p. 231)

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} E\{-2 \text{ loglikelihood } (g)\} &= 2 \log(2\pi) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\log g(\omega) + \frac{f(\omega)}{g(\omega)} \right) d\omega \\ &= \frac{1}{n} d(f, g). \end{aligned} \quad (66)$$

$d(f, g)$ thus approximates $E\{-2 \text{ loglikelihood } (g)\}$, which is by definition the Kullback-Leibler information.

The Akaike information criterion

$$\text{AIC} = -2 \text{ loglikelihood } (g_{\hat{\theta}}) + 2(m + 1) \quad (67)$$

can be seen as an estimate of $E\{d(f, g_{\hat{\theta}})\}$ where the expectation is taken with respect to the true joint distribution of the sample observations. Hurvich & Beltrão now define a slightly modified version of CVLL_{BB} , which is valid for any (parametric and non-parametric) spectrum estimate \hat{f}

$$\text{CVLL}_{HB}(\hat{f}) = 2n \log(2\pi) + \frac{n}{N} \sum_{j=1}^N \left(\log \hat{f}(\omega_j) + \frac{I(\omega_j)}{\hat{f}^{-j}(\omega_j)} \right). \quad (68)$$

The function $\hat{f}^{-j}(\omega)$ is the leave-out-one version of the spectrum estimate \hat{f} defined by Hurvich in (64). It can be shown that $\hat{f}^{-j}(\omega) \approx \hat{f}(\omega)$, but that $\hat{f}^{-j}(\omega)$ is approximately independent of $I(\omega_j)$. Note that (68) differs from the original definition by Beltrão & Bloomfield in two ways. First, Hurvich & Beltrão include a constant term in the likelihood function to make it comparable to AIC, and second, they use $\log \hat{f}(\omega_j)$ in place of $\log \hat{f}^{-j}(\omega_j)$. The reason they give for the latter change is to save computation time.

Hurvich & Beltrão also claim that the difference between the two terms is negligible and give an upper bound.⁹

CVLL_{HB} can be written as

$$\begin{aligned} \text{CVLL}_{HB}(\hat{f}) &= d(f, \hat{f}) + (d(I, f) - d(f, f)) \\ &\quad + \frac{n}{N} \sum_{\omega_j} (I(\omega_j) - f(\omega_j)) \left(\frac{1}{\hat{f}^{-j}(\omega_j)} - \frac{1}{f(\omega_j)} \right) \\ &\quad + \frac{n}{N} \sum_{\omega_j} f(\omega_j) \left(\frac{1}{\hat{f}^{-j}(\omega_j)} - \frac{1}{\hat{f}(\omega_j)} \right). \end{aligned} \quad (69)$$

The second term in (69) is a random level term which does not depend on the potential estimate. Its expectation is approximately zero. Since $E(I(\omega_j)) \approx f(\omega_j)$, and since $I(\omega_j)$ and $\hat{f}^{-j}(\omega_j)$ are approximately independent, the expectation of the third term is also approximately zero. Finally, if we assume that $E(1/\hat{f}^{-j}(\omega_j)) \approx E(1/\hat{f}(\omega_j))$, then the expectation of the fourth term is also approximately zero, and we have

$$E\{\text{CVLL}_{HB}(\hat{f})\} \approx E\{d(f, \hat{f})\}, \quad (70)$$

which confirms that CVLL_{HB} is an asymptotically unbiased estimate of the expected Kullback-Leibler information.

In the following, Hurvich & Beltrão obtain a computationally more efficient version of the CVLL by introducing another unbiased estimate of the Kullback-Leibler information which does not require computation of the leave-out-one estimate.

Since the distribution of a classical estimate $\hat{f}(\omega_j)$ is often approximated as $(f(\omega_j)/v)\chi_v^2$ with $v = 2/(\sum_{-\pi/h < \omega_j < \pi/h} K^2(h\omega_j))$, it follows that

$$\begin{aligned} E\{d(f, \hat{f})\} &\approx 2n \log(2\pi) + E \frac{n}{N} \sum \log \hat{f}(\omega_j) + \frac{n}{N} \sum_{\omega_j} E \left(\frac{f(\omega_j)}{\hat{f}(\omega_j)} \right) \\ &\approx 2n \log(2\pi) + E \frac{n}{N} \sum_{\omega_j} \log \hat{f}(\omega_j) + \frac{nv}{v-2}. \end{aligned} \quad (71)$$

Thus CVLL2, as given in the following equation, can be view as an approximately unbiased estimate of $E\{d(f, \hat{f})\}$

$$\text{CVLL2} = 2n \log(2\pi) + \frac{n}{N} \sum_{\omega_j} \log \hat{f}(\omega_j) + \frac{nv}{v-2}. \quad (72)$$

⁹This step remains somewhat unclear, as we do not see how the computation time, applying $\log \hat{f}(\omega_j)$ in place of $\log \hat{f}^{-j}(\omega_j)$, will be reduced substantially. After all, the leave-out-one estimate has to be computed anyway.

Notice that computing time for calculating the non-cross validatory CVLL2 decreases substantially, as it requires $O(n \log n)$ computations, whereas the cost of computing $CVLL_{HB}$ is $O(n \log n + n \frac{N}{h})$.

In their simulation study Hurvich & Beltrão generate 100 samples of length 100 (500) for an AR(3) process, for a MA(1) process and a superposition of a number of complex exponentials with random phases and amplitudes, respectively. Their simulation results suggest that the automatic window width selector, makes non-parametric spectrum estimation quite competitive with parametric spectrum estimation, as long as the true process is not a finite-order autoregression. Furthermore, the combined selector, i.e. the estimator that yields the minimum CVLL or AIC, respectively, is usually capable of determining the best estimate type for the data at hand.

4 An Iterative Procedure

Bühlmann (1996) estimates the optimal local and global window widths that are minimizing the asymptotic mean square error $\text{AMSE}(\omega; h)$ or the asymptotic mean integrated square error $\text{AMISE}(h)$ by an iterative procedure. The optimal local and global window widths for specific lag windows and a given true spectral density are known and depend on the unknown spectral density (see Priestley). As we are interested in estimating the spectral density, however, we should not assume the true density to be known. One way of exploiting the fact that there are known formulas for optimal window widths depending on the true density is to iteratively estimate the density and its induced optimal window widths, using these for the next density estimate and hoping that this procedure will make the density estimates converge in some sense to the true spectral density. This approach was followed by Bühlmann (1996) building on the work by Brockmann *et al.* (1993) who employed this idea in the context of nonparametric regression.

In this section we will for the first time in this paper choose a window width, for kernel smoothing the periodogram, locally, that is possibly different at different frequencies depending on the curvature of the true spectral density. 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. As for flat regions of the density a high amount of smoothing may be called for, for peaky regions it might be better not to smooth too much, i.e. not to give too much weight 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.

The spectral density estimates considered by Bühlmann are lag window estimates of the scale parameter type as defined in (9). Bühlmann considers two types of windows, a C^0 window, specifically the Bartlett or triangular window, and general C^2 windows with characteristic exponent 2.

Bühlmann finds the optimal local and global scale parameters by minimizing the asymptotic mean square error $\text{AMSE}(h, \omega)$ or the asymptotic mean integrated square error $\text{AMISE}(h)$. The AMSE is given by the sum of the squared asymptotic bias and the asymptotic variance at the frequency ω ,

$$\text{AMSE}(h, \omega) = \text{ABIAS}^2(h, \omega) + \text{AVAR}(h, \omega). \quad (73)$$

The AMISE then is just the integral of the AMSE over all frequencies. For a general C^2 window, expressions for the asymptotic bias and the asymptotic variance are given by (39) and (40) in section 2. For the particular C^0 window considered by Bühlmann, the Bartlett window, (see Bühlmann, p. 249, equation 3) these expressions are

$$\text{ABIAS}(h, \omega) = -\frac{1}{h} f^{(1)}(\omega) \quad (74)$$

and

$$\text{AVAR}(h, \omega) = \iota(\omega) \frac{2h}{3n} \{f(\omega)\}^2, \quad (75)$$

respectively, with $\iota(\omega)$ as defined in (38).

The optimal value of the local and global scale parameters for any C^2 window by simple calculus can be shown to be

$$h_{\text{opt}}(\omega) = n^{1/5} \left\{ \frac{\{\kappa''(0)\}^2 \{f^{(2)}(\omega)\}^2}{\iota(\omega) \int_{-\infty}^{\infty} \kappa^2(x) dx \{f(\omega)\}^2} \right\}^{1/5} \quad (76)$$

and

$$h_{\text{opt}} = n^{1/5} \left\{ \frac{\{\kappa''(0)\}^2 \int_{-\pi}^{\pi} \{f^{(2)}(\omega)\}^2 d\omega}{\int_{-\infty}^{\infty} \kappa^2(x) dx \int_{-\pi}^{\pi} \{f(\omega)\}^2 d\omega} \right\}^{1/5}, \quad (77)$$

respectively. For the Bartlett window we obtain

$$h_{\text{opt}}(\omega) = n^{1/3} \left\{ \frac{3\{f^{(1)}(\omega)\}^2}{\iota(\omega)\{f(\omega)\}^2} \right\}^{1/3} \quad (78)$$

and

$$h_{\text{opt}} = n^{1/3} \left\{ \frac{3 \int_{-\pi}^{\pi} \{f^{(1)}(\omega)\}^2 d\omega}{\int_{-\pi}^{\pi} \{f(\omega)\}^2 d\omega} \right\}^{1/3}, \quad (79)$$

respectively. These optimal values for the scale parameter h apparently depend on the true spectral density and its first or second generalized derivatives, as defined in equation (34), depending on whether the considered lag window is a Bartlett one or C^2 , respectively. Bühlmann uses the above results and iteratively estimates the density and its generalized derivatives to determine first the optimal global and then the optimal local scale parameter for the window used, in a few iteration steps.

For the various spectral density related objects Bühlmann considers the following estimators. For the integrated squared density¹⁰ he suggests to use the integrated squared periodogram divided by two,

$$\frac{1}{2} \int_{-\pi}^{\pi} \left\{ \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \hat{\gamma}(k) e^{-ik\omega} \right\}^2 d\omega. \quad (80)$$

For the density, $f(\omega)$, as in (9),

$$\hat{f}(\omega, h) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \tilde{\kappa} \left(\frac{k}{h} \right) \hat{\gamma}(k) e^{-ik\omega}, \quad (81)$$

and for the first and second generalized derivatives of the spectral density, $f^{(1)}(\omega)$ and $f^{(2)}(\omega)$,

$$\hat{f}^{(1)}(\omega, h) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \tilde{\kappa} \left(\frac{k}{h} \right) |k| \hat{\gamma}(k) e^{-ik\omega} \quad (82)$$

¹⁰Bühlmann mentions that the squared periodogram is not a consistent estimator of the squared density, whereas the integral thereof is consistent for the integrated squared density.

and

$$\hat{f}^{(2)}(\omega, h) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \bar{\kappa} \left(\frac{k}{h} \right) k^2 \hat{\gamma}(k) e^{-ik\omega}, \quad (83)$$

respectively.

In his remark 2 Bühlmann proposes to use different windows for estimating the generalized derivatives of the spectral density, $f^{(s)}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} |k|^s \gamma(k) e^{-ik\omega}$, $s = 1, 2$. He argues that the reason for that is the fact that the terms $|k|\hat{\gamma}(k)$ and $k^2\hat{\gamma}(k)$ usually do not decay very fast. The lag window he proposes is a specific splitted rectangular-cosine window with lag kernel

$$\kappa(x) = \begin{cases} 1 & \text{if } |x| < 0.8 \\ \{1 + \cos(5(x - 0.8)\pi)\}/2 & \text{if } 0.8 \leq |x| < 1 \\ 0 & \text{otherwise.} \end{cases} \quad (84)$$

This kernel has infinite characteristic exponent. If we would use it for estimating the spectral density also, formulations of bias and variance would be different and the above theoretically optimal window width not valid.

The fact that one has to use different windows for estimates of the density and its derivatives might seem a bit complicated and might make Bühlmann's iterative approach somewhat inelegant. One may ask whether to exploit the relationship between the generalized and the ordinary derivatives as stated in equation (35), at least for the second derivatives and therefore for the case of a C^2 window, will not be more appropriate. Approximating the second derivative by some finite difference scheme does in general not produce a very good estimate, however. In fact, this estimate might not be very smooth in contrast to the true second derivative. This depends, however, on the kernel used for the spectral density estimate. If one would use a kernel, similar to the splitted rectangular-cosine window, that has characteristic exponent of two, the numerical approximation would probably be just as good as the lag window estimate. For the Bartlett-Priestley window the numerical approximation does not seem to be a very good one as we see in our simulation study.

As the type of the window is not that much of importance as compared to the choice of the scale parameter (see Priestley p. 449), in our simulation study we only look at one specific C^2 window, the Bartlett-Priestley window as defined by (21) and (22). For this specific window bias and variance are given by (Priestley p. 463)

$$\text{ABIAS}(h, \omega) = -\frac{\pi^2}{10h^2} f''(\omega) \quad (85)$$

and

$$\text{AVAR}(h, \omega) = \iota(\omega) \frac{6h}{5n} f^2(\omega), \quad (86)$$

respectively. The optimal values for the local and global scale parameters are

$$h_{\text{opt}}(\omega) = n^{1/5} \left\{ \frac{\pi^4 \{f''(\omega)\}^2}{30 \iota(\omega) f^2(\omega)} \right\}^{1/5} \quad (87)$$

and

$$h_{\text{opt}} = n^{1/5} \left\{ \frac{\pi^4 \int_{-\pi}^{\pi} \{f''(\omega)\}^2 d\omega}{30 \int_{-\pi}^{\pi} f^2(\omega) d\omega} \right\}^{1/5}, \quad (88)$$

respectively.

The iteration scheme employed by Bühlmann is the following.

Algorithm 4.1

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: $h_i = n^{1/5} \left\{ \frac{2\{\kappa''(0)\}^2 \sum_{k=-n+1}^{n-1} \bar{\kappa}^2 \left(\frac{k}{h_{i-1}} n^{4/45} \right) k^4 \hat{\gamma}^2(k)}{\int_{-\infty}^{\infty} \kappa^2(x) dx \sum_{k=-n+1}^{n-1} \hat{\gamma}^2(k)} \right\}^{1/5}$

5. if $i < 4$ goto 3

6. Local step: $h_{\text{opt}}(\omega) = n^{1/5} \left\{ \frac{2\{\kappa''(0)\}^2 \left\{ \hat{f}^{(2)} \left(\omega; \frac{h_4}{n^{4/45}} \right) \right\}^2}{\int_{-\infty}^{\infty} \kappa^2(x) dx \left\{ \hat{f} \left(\omega; \frac{h_4}{n^{4/45}} \right) \right\}^2} \right\}^{1/5}$.

Bühlmann motivates the inflation factor $n^{4/45}$ by some asymptotics for the local step and argues that using the same factor in the global steps as well will yield a more stable procedure. This argument is based on some simulation Bühlmann mentions.

Bühlmann argues that four global iteration steps will already yield the right order and further steps will not give any improvement. Also performing more than one local step will not improve the estimate.

The only problem that might arise in local smoothing is at inflection points of the spectral density. At these points, where the second derivative is zero, the above formulation (88) of the 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 4.1 is replaced by its integral over a small range.

Algorithm 4.2

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: $h_i = n^{1/5} \left\{ \frac{2\{\kappa''(0)\}^2 \sum_{k=-n+1}^{n-1} \bar{\kappa}^2 \left(\frac{k}{h_{i-1}} n^{4/45} \right) k^4 \hat{\gamma}^2(k)}{\int_{-\infty}^{\infty} \kappa^2(x) dx \sum_{k=-n+1}^{n-1} \hat{\gamma}^2(k)} \right\}^{1/5}$

5. if $i < 4$ goto 3

$$6. \text{ Semi-local step: } h_{opt}(\omega) = n^{1/5} \left\{ \frac{2\{\kappa''(0)\}^2 \int_{\omega-c}^{\omega+c} \left\{ \hat{f}^{(2)}\left(\lambda; \frac{h_A}{n^{4/45}}\right) \right\}^2 d\lambda}{\int_{-\infty}^{\infty} \kappa^2(x) dx \left\{ \hat{f}\left(\omega; \frac{h_A}{n^{4/45}}\right) \right\}^2} \right\}^{1/5},$$

where $c = \frac{n^{4/45}}{b_4}$. 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, but with different estimates for the various objects. In particular we try to approximate the second derivative by differences and we use estimates with inflated as well as not-inflated window widths.

5 A Bootstrap Approach

Franke & Härdle (1992) adopt a bootstrap approach in order to determine the optimal scale parameter. Given the periodogram of a strictly stationary real-valued time series X_1, \dots, X_n , denoted by $I(\omega)$, Franke & Härdle consider kernel spectral density estimates of the form (compare (12))

$$\hat{f}(\omega, h) = \frac{2\pi h}{n} \sum_{j=-N}^N K(h(\omega - \omega_j)) I(\omega_j), \quad (89)$$

where K is a spectral window generating kernel as defined by (17) and N denotes the largest integer less than or equal to $n/2$. In contrast to Bühlmann (1996), Franke & Härdle determine locally optimal scale parameters according to the minimal mean square percentage error (MSPE) as defined by (28). They introduce the bootstrap in frequency domain via a multiplicative regression problem,

$$I(\omega_j) = f(\omega_j)\epsilon_j. \quad (90)$$

The residuals are approximately independent and identically distributed for large n (see Priestley Chpt. 6.2). It is these residuals, replacing the true density by the kernel estimate using an 'arbitrary' initial bandwidth, that constitute the sample of independent observations to be resampled.

$$\hat{\epsilon}_j = \frac{I(\omega_j)}{\hat{f}(\omega_j; h_0)}. \quad (91)$$

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

$$\tilde{\epsilon}_j = \frac{\hat{\epsilon}_j}{\hat{\epsilon}_\cdot}, \quad (92)$$

where

$$\hat{\epsilon}_\cdot = \frac{1}{N} \sum_{j=1}^N \hat{\epsilon}_j. \quad (93)$$

The bootstrap procedure is performed as follows. A bootstrap sample, $\epsilon_1^*, \dots, \epsilon_N^*$ from the empirical distribution of $\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_N$ is drawn. Using a bandwidth g , possibly different from h_0 , bootstrap periodogram values are obtained, which are

$$I^*(\omega_j) = \hat{f}(\omega_j, g)\epsilon_j^*. \quad (94)$$

The corresponding bootstrap spectral estimate is then given by

$$\hat{f}^*(\omega, h, g) = \frac{2\pi h}{n} \sum_{j=-N}^N K(h(\omega - \omega_j)) I^*(\omega_j). \quad (95)$$

Alluding to the fact that the rescaled residuals asymptotically follow an exponential distribution with scale parameter 1, there is an obvious second (parametric) way of performing the bootstrap. One may draw samples of the size N from an exponential

distribution with scale parameter 1, denoting them χ_1, \dots, χ_n , obtain the bootstrap periodogram values as

$$I^+(\omega_j) = \hat{f}(\omega_j; g)\chi_j, \quad (96)$$

and determine another bootstrap spectral estimate

$$\hat{f}^+(\omega, h, g) = \frac{2\pi h}{n} \sum_{j=-N}^N K(h(\omega - \omega_j)) I^+(\omega_j). \quad (97)$$

As Franke & Härdle (Theorem 1) point out, the bootstrap principle holds for both cases under some convenient assumptions.

We will now use the above bootstrap resampling schemes in order to determine the optimal scale parameter h . In contrast to the cross-validation methods used by e.g. Beltrão & Bloomfield (1987), we will not minimize the average mean square percentage error, but the local mean square percentage error for each Fourier frequency as defined by (28). The average MSPE would just be given by

$$\text{AMSPE}(h) = \frac{1}{N} \sum_{j=1}^N \text{MSPE}(\omega_j, h) \quad (98)$$

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

$$\text{MSPE}^*(\omega, h) = E^* \left\{ \frac{\hat{f}^*(\omega, h, g) - \hat{f}(\omega, g)}{\hat{f}(\omega, g)} \right\}^2. \quad (99)$$

In fact, there is no need to resample, as we may calculate MSPE^* explicitly.¹¹

$$\begin{aligned} & \hat{f}^2(\omega; g) \text{MSPE}^*(\omega; h) = \\ & \frac{h^2 \text{var}^*(\epsilon_1^*)}{n^2} \left(K^2(0) \hat{f}^2(0; g) + \sum_{j=1}^N \{K(h(\omega - \omega_j)) + K(h(\omega + \omega_j))\}^2 \hat{f}^2(\omega_j; g) \right) 100 \\ & + \left\{ \frac{h}{n} \sum_{j=-N}^N K(h(\omega - \omega_j)) \hat{f}(\omega_j; g) - \hat{f}(\omega; g) \right\}^2. \end{aligned}$$

Franke & Härdle note (p. 135) that $\text{var}^*(\epsilon_1^*) \rightarrow 1$ in probability. The scale parameter minimizing the above estimate of the MSPE is the one regarded optimal. Franke & Härdle are able to prove that the resulting estimate is in fact a consistent estimate of the optimal scale parameter (p. 133, Theorem 3).

¹¹There seems to be an error in Franke & Härdle's formula (6). They seem to have forgotten about the cross terms in the variance, which are present as $I(\omega_{-j}) = I(\omega_j) = \epsilon_j^* \hat{f}(\omega_j, g)$.

6 Simulation

In this section we compare several procedures, discussed in the previous sections, for obtaining optimal scale parameters by means of a small simulation study. These procedures are applied to a set of AR and MA processes, selected such as to exhibit different shapes of spectral densities. For each process, we simulate 300 time series of length 120 and 480. For each process and each scale parameter selection method three different distance measures are approximated, MISE, MISPE and MISME, as defined by (29), (30) and (31) with $a = 2$. These are in fact calculated as the average over all simulations of

$$\text{IS(P,M)E} = \frac{1}{N} \sum_{j=-N}^N \left(\hat{f}(\omega_j) - f(\omega_j) \right)^2 (f(\omega_j))^a, \quad (101)$$

where $a = 0$ yields the ISE, $a = -2$ ISPE and $a = 2$ ISME.

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 'observations' of the particular process, and dropping the first hundred at the end (see appendix in Hurvich (1985), p. 939).

In the tables that summarize the simulation results, we abbreviate the methods in the following way. The cross-validation methods, CVLL, CVLL2, SES and CVMSE are as defined in previous sections. There are several iterative methods that we discuss, however. ITB is the global method suggested by Bühlmann (1996) and is given by the first part of algorithm (4.1), that is using the splitted rectangular-cosine lag window estimate of the second generalized derivative and using 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 this inflation factor, while ITA is yet another global method, following the said algorithm, that also does not use this inflation factor and furthermore 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 (4.2), where 1 indicates that for this semi-local step the inflation factor was used and 2 that it was not used. The global window width used in the one further semi-local step are always given by the corresponding global method, e.g. ITA for ITA1. BOOT denotes the bootstrap criterion (100), where the reference bandwidth g is determined by ITC.

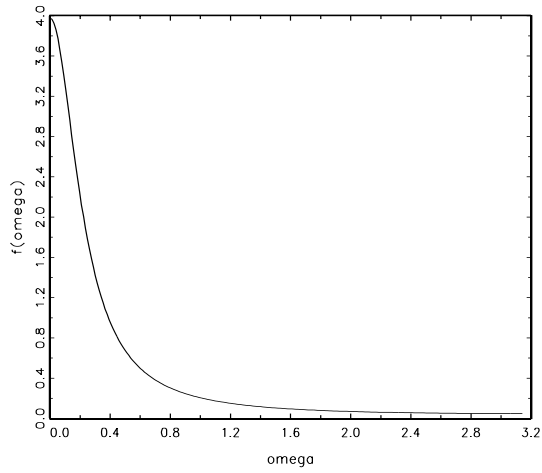


Fig.1: $X_t = 0.8X_{t-1} + Z_t$

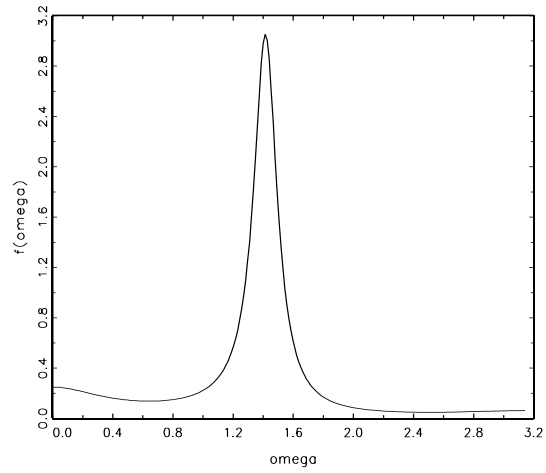


Fig.2: $X_t = 0.4X_{t-1} - 0.5X_{t-2} + 0.3X_{t-4} + Z_t$

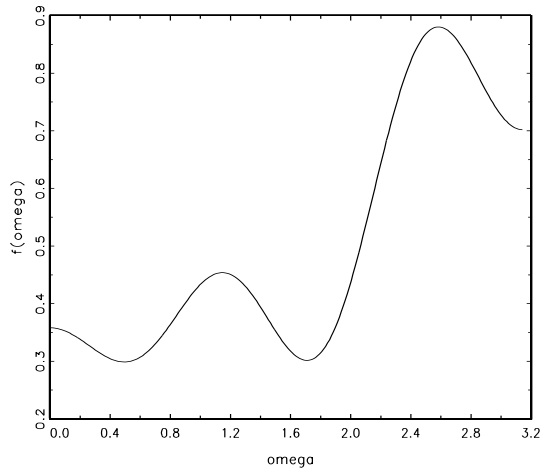


Fig.3:

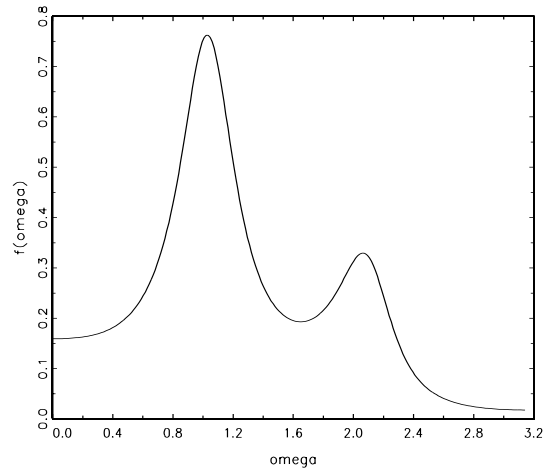


Fig.4:

$X_t = Z_t + 0.9Z_{t-1} - 0.8Z_{t-2} + 0.6Z_{t-3} - 0.5Z_{t-4} + 0.3Z_{t-5}$ $X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + Z_t$

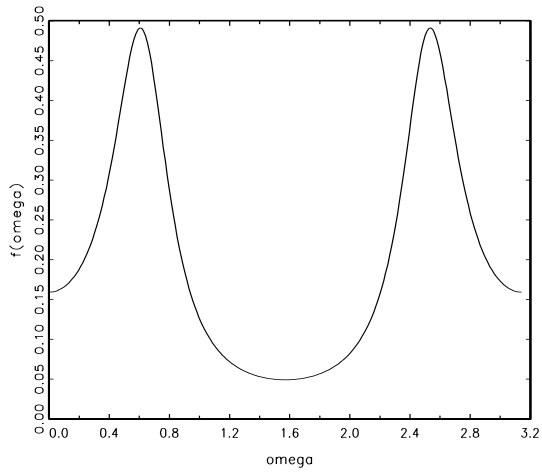


Fig.5: $X_t = 0.4X_{t-2} - 0.4X_{t-4} + Z_t$

Table 1:

Method	MISE	MISPE	MISME
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)

AR(1): $X_t = 0.8X_{t-1} + Z_t$, $Z_t \sim N(0, 1)$, ($n=120$). Variances are in brackets. A * indicates the best global estimate, whereas a ** points out the best semi-local estimate.

For the AR(1)-process in table 6, whose spectral density has one peak at frequency zero, the best method among the cross-validation ones is CVLL, which is in fact only slightly worse than all three global iterative procedures. The three iterative procedures perform similarly, the best in this case may be Bühlmann's original if we consider variances as well. Semi-local procedures seem to improve the estimate, and this is similar for all three methods, where using no inflation factor (i.e. suffix 2) seems to be superior to using it. The bootstrap criterion, however, seems to be even slightly better than ITC2.

Table 2:

Method	MISE	MISPE	MISME
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)

AR(1): $X_t = 0.8X_{t-1} + Z_t$, $Z_t \sim N(0, 1)$, ($n=480$). Variances are in brackets. A * indicates the best global estimate, whereas a ** points out the best semi-local estimate.

As we may see in table 6 Bühlmann's original, ITB, outperforms the other global methods for this AR(1)-process and $n = 480$. Semi-local window-width choice, ITB2, might be of a slight advantage. ITC2 does not perform much worse than ITB and ITB2.

Table 3:

Method	MISE	MISPE	MISME
CVLL	0.02619 (0.00041)	0.456 (0.120)	0.00770 (0.00006)
CVLL2	0.03509 (0.00100)	0.474 (0.088)	0.00996 (0.00014)
SES	0.03233 (0.00063)	0.745 (1.200)	0.00997 (0.00010)
CVMSE	0.02806 (0.00090)	0.475 (0.141)	0.00830 (0.00011)
ITA	0.02531* (0.00052)	0.370* (0.062)	0.00716* (0.00006)
ITA1	0.02395 (0.00016)	0.420 (0.082)	0.00712 (0.00003)
ITA2	0.02377** (0.00024)	0.376 (0.055)	0.00689** (0.00003)
ITB	0.03179 (0.00014)	1.518 (2.224)	0.00947 (0.00003)
ITB1	0.03658 (0.00010)	0.960 (0.866)	0.01137 (0.00003)
ITB2	0.03001 (0.00014)	0.681 (0.339)	0.00902 (0.00003)
ITC	0.02604 (0.00054)	0.402 (0.091)	0.00764 (0.00008)
ITC1	0.02576 (0.00019)	0.489 (0.141)	0.00776 (0.00003)
ITC2	0.02380 (0.00025)	0.384 (0.060)	0.00697 (0.00004)
BOOT	0.02427 (0.00027)	0.370** (0.054)	0.00752 (0.00004)

AR(4): $X_t = 0.4X_{t-1} - 0.5X_{t-2} + 0.3X_{t-4} + Z_t$, $Z_t \sim N(0, 1)$, ($n=120$). Variances are in brackets. A * indicates the best global estimate, whereas a ** points out the best semi-local estimate.

The spectral density of this particular AR(4)-process as in table 6 contains one sharp peak. This peak calls for a rather low amount of smoothing (high h) at least in the region around the peak. CVLL performs best among the cross-validation methods. The global estimates induced by ITC and ITA are slightly better. Contradicting our belief semi-local or local (BOOT) window width choice does yield only slightly better estimates than global procedures. This is somewhat surprising, because in the region of this one single sharp peak it would be of advantage to use much higher scale parameters than in the flat region further apart from the peak. Bühlmann's original methods, ITB and ITB1, do not perform very well for this process. ITA2, ITC2 and BOOT perform very similarly.

Table 4:

Method	MISE	MISPE	MISME
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)

AR(4): $X_t = 0.4X_{t-1} - 0.5X_{t-2} + 0.3X_{t-4} + Z_t$, $Z_t \sim N(0, 1)$, ($n=480$). Variances are in brackets. A * indicates the best global estimate, whereas a ** points out the best semi-local estimate.

Bühlmann's original can not compete with the other two methods for the process in table 6. ITC seems to yield the best global estimate. One can hardly tell whether semi-local procedures yield better estimates than ITC.

Table 5:

Method	MISE	MISPE	MISME
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)

AR(5): $X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + Z_t$, $Z_t \sim N(0, 1)$, ($n=120$). Variances are in brackets. A * indicates the best global estimate, whereas a ** points out the best semi-local estimate.

The best cross-validation method is clearly CVLL for the AR(5)-process in table 6, with spectral density that exhibits two rather broad and rather low peaks (bumps), and is not much worse than the two iterative procedures ITA and ITC. ITB as well as the other cross-validatory methods are not competitive for this process. ITA and ITC perform rather similarly with a slight advantage for ITA. BOOT performs similar to ITC2. The best estimate seems to be determined by ITA2, which does not much better than ITC2, though.

Table 6:

Method	MISE	MISPE	MISME
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)

AR(5): $X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + Z_t$, $Z_t \sim N(0, 1)$, ($n=480$). Variances are in brackets. A * indicates the best global estimate, whereas a ** points out the best semi-local estimate.

ITC and ITC2 are clearly dominating for the process from as in table 6. Only ITA1 performs equally well. There is not much difference between ITC and ITC2, i.e. between the global and the semi-local procedure.

Table 7:

Method	MISE	MISPE	MISME
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)

MA(5): $X_t = Z_t + 0.9Z_{t-1} - 0.8Z_{t-2} + 0.6Z_{t-3} - 0.5Z_{t-4} + 0.3Z_{t-5}$, $Z_t \sim N(0, 1)$, ($n=120$). Variances are in brackets. A * indicates the best global estimate, whereas a ** points out the best semi-local estimate.

The spectral density of the MA(5)-process in table 6 is rather flat with three differently sized bumps. CVLL performs best among the cross-validation methods, but is worse than Bühlmann's original global estimate, ITB, worse than ITC2 and worse than all semi-local estimates that use the inflation factor. These perform in fact similarly for all three iterative procedures. BOOT performs similar to ITC2. The best estimate is determined by ITC1.

Table 8:

Method	MISE	MISPE	MISME
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)

MA(5): $X_t = Z_t + 0.9Z_{t-1} - 0.8Z_{t-2} + 0.6Z_{t-3} - 0.5Z_{t-4} + 0.3Z_{t-5}$, $Z_t \sim N(0, 1)$, ($n=480$). Variances are in brackets. A * indicates the best global estimate, whereas a ** points out the best semi-local estimate.

The best global method for the process of table 6 seems to be Bühlmann's original, ITB. Dominating, however, is ITC1.

Table 9:

Method	MISE	MISPE	MISME
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)

AR(4): $X_t = 0.4X_{t-2} - 0.4X_{t-4} + Z_t$, $Z_t \sim N(0, 1)$, ($n=120$). Variances are in brackets. A * indicates the best global estimate, whereas a ** points out the best semi-local estimate.

For the AR(4)-process in table 6, with a spectral density that exhibits two equal-sized peaks that are quite distant from each other, the best cross-validation method is again CVLL. ITB is not competitive, whereas ITA, ITA2, ITC and ITC2 seem to perform quite similarly and better than CVLL. There seems to be a slight improvement if one considers semi-local estimates (without inflation factor). The best estimate, however, seems to be obtained by BOOT.

Table 10:

Method	MISE	MISPE	MISME
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)

AR(4): $X_t = 0.4X_{t-2} - 0.4X_{t-4} + Z_t$, $Z_t \sim N(0, 1)$, ($n=480$). Variances are in brackets. A * indicates the best global estimate, whereas a ** points out the best local estimate.

For the process of table 6 all methods perform rather similarly, with the exception of ITA and ITB1. The best procedure might be ITC2.

7 Conclusion

In this paper we reviewed and compared several 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 & Bloomfield (1987) and Hurvich & Beltrão (1990), iterative estimates following Bühlmann (1996) and a bootstrap estimate following Franke & 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 looked at, for reasons of speed and because we do not expect the cross-validation based procedures to suddenly perform better than the iterative ones for larger sample sizes.

It seems that best among the cross-validation methods in general is CVLL, no matter which of the three distance measures we use.

Concerning the iterative methods it depends a bit on the criterion, as well as on the shape of the spectral density function, and on the sample size, which of the three procedures is better adapted for determining the optimal window width. In general, however, Bühlmann's original method is worse than the other iterative methods and even worse than CVLL, unless the spectral density is rather flat. This seems to stem from the fact that this method tends to yield rather low scale parameters induced by the inflation factor. For larger sample sizes, Bühlmann's original seems to be more competitive, but still worse than the other two methods, in general. The other two iterative procedures seem in general to perform better or nearly as good as the best cross-validatory one. Semi-local procedures seem to improve the estimate, unless the spectral density exhibits sharp peaks. The best among them is probably ITC2, which is using the non-inflation factor global estimate for determining the non-inflation factor semi-local estimate.

The bootstrap procedure performs similarly to ITC2. We think that this is additional evidence that the inflation factor should not be used in the iterative methods.

Generally, it seems most appropriate to use method ITC for the global estimates and ITC2 for the local one, that is using Bühlmann's iterative scheme as in algorithm (4.1), but without the inflation factor. ITC2, very often, seems to perform best in terms of at least one of the three criteria, and is never much worse than other methods. It is furthermore a very fast method compared to the bootstrap procedure and the cross-validatory ones, as it only needs five iterations in order to determine the optimal scale parameter as compared to many more steps necessary when searching for the e.g. CVLL minimizing scale parameter.

Our insight, however, depends on the particular set of simulations we chose, where we tried to pick processes that are rather different from each other in terms of the shape of their spectral density. To be able to draw valid conclusions one should probably

perform a much more rigorous simulation study. For example one might do a total of about one hundred simulations for different processes in the ARMA(5,5) class, where in a first step the one hundred times ten parameters are chosen randomly according to a uniform distribution taking into consideration some stationarity conditions. With these one hundred simulation results one might in fact do some statistical inference in order to find the best general method for this class of processes, provided there is one. One may argue, however, that, in practice, we are not confronted with one in the uniformly distributed mass of processes, but rather with very specific types only. This belief might make our small simulation approach more reasonable.

References

- [1] Beltrão, K. I. & Bloomfield, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *J. Time Ser. An.*, **8**, 21-38.
- [2] Brockmann, M., Gasser, T. & Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *J. Am. Statist. Assoc.*, **88**, 1302-9.
- [3] Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed., Springer, New York.
- [4] Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *J. Time Ser. An.*, **17**, 3, 247-70.
- [5] Franke, J. & Härdle, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist.*, **20**, 1, 121-45.
- [6] 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.
- [7] Hurvich, C. M. & Beltrão, K. I. (1990). Cross-validatory choice of a spectrum estimate and its connection with AIC. *J. Time Ser. An.*, **11**, 121-37.
- [8] Marron, J. S. (1985). An asymptotically efficient solution to the bandwidth problem of kernel density estimation. *Ann. Statist.*, **13**, 1011-23.
- [9] Palmer, F. (1983). Adaptive spectral smoothing, technical report ONR 30, Massachusetts Institute of Technology, Statistics Center.
- [10] Parzen, E. (1983). Autoregressive spectral estimation, in Handbook of Statistics, Vol. 3 (eds. D. R. Brillinger and P. R. Krishnaiah), New York: Elsevier, 221-47.
- [11] Priestley, M. B. (1981). *Spectral Analysis and Time Series 1*. Academic, New York.
- [12] Wahba, G. & 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.