Locally Adaptive Lag-Window Spectral Estimation

By

Peter Bühlmann

Technical Report No. 422
October 1994

Department of Statistics
University of California
Berkeley, California

Locally Adaptive Lag-Window Spectral Estimation

Peter Bühlmann
ETH Zürich and Univ. of Calif. Berkeley

October 1994

Abstract

We propose a procedure for the locally optimal window width in nonparametric spectral estimation, minimizing the asymptotic mean square error at a fixed frequency λ of a lag-window estimator. Our approach is based on an iterative plug-in scheme. Besides the estimation of a spectral density at a fixed frequency, e.g. at frequency $\lambda = 0$, our procedure allows to perform nonparametric spectral estimation with variable window width which adapts to the smoothness of the true underlying density.

Key words and phrases. Bandwidth, Iterative plug-in, Nonparametric spectral estimation, Strong mixing sequence, Time Series, Window width .

Short title: Adaptive Spectral Estimation

1 Introduction

The well known nonparametric estimators for the spectral density of a stationary, real-valued process all involve a smoothing parameter, the bandwidth for the smoothed periodogram or the window width for the lag-window estimators (cf. Priestley (1981)). These parameters are of crucial importance and can influence the outcome of a spectral estimate very much. We treat here the problem of local window width selection for lag-window estimators in nonparametric spectral density estimation. We introduce an iterative plug-in estimator for the optimal local window width, minimizing the mean square error of a spectral estimate at a fixed frequency λ. We derive some asymptotic results and obtain under appropriate mixing conditions on the underlying process, which implies a highly smooth behavior of the corresponding spectral density, the rate of order $n^{-1/2+\varepsilon}$ ($\varepsilon > 0$) for the relative error. Our estimation procedure is motivated by the work of Brockmann et al. (1993). They propose an iterative plug-in estimator for the optimal local bandwidth in nonparametric regression for i.i.d. observations. A different approach is given by Chiu (1990). In particular the iterative nature of their plug-in procedure is an attractive idea which we adapted for spectral estimation. A nice feature of these methods is its simple underlying idea and its fast computability.

Often the estimation at frequency $\lambda = 0$ is interesting. The asymptotic variance of the arithmetic mean of stationary observations equals the spectral density at zero of the observation process. Thus, to construct an approximate confidence interval for the expectation based on the mean we need to estimate the corresponding spectral density at zero. Brillinger (1994) proposes a wavelet estimator for reconstructing a deterministic signal from observations with additive stationary errors. For the limiting variance of this estimator one needs an estimate of the spectrum at zero of the error process. The same is true for kernel estimators in nonparametric regression with stationary errors. Another application of interest is the blockwise bootstrap in time series (cf. Künsch (1989), Politis and Romano (1992), Bühlmann (1994)). There, the optimal blocklength minimizing the mean square error of the bootstrap variance of a linear statistic depends again on a corresponding spectral density at zero.

Similarly as in Brockmann et al. (1993) we describe a modified procedure for adaptive lag-window spectral estimation over the whole interval $[0, \pi]$ with variable window width. In particular for spectra with peaks there is an advantage of using variable window widths.

In spectral estimation Beltrão and Bloomfield (1987) have given a global bandwidth choice for kernel estimators based on cross-validation, see also Beltrão and Hurvich (1990). Müller and Gasser (1986) use some autoregressive fitting (with fixed order) for the same objective and Park and Cho (1991) give some estimators of the integrated squared spectral density derivatives which could be used for a plug-in estimator for global bandwidth selection of a kernel estimator. Nothing seems to be known about optimal local bandwidth selection.

The finite sample behavior of our procedures is illustrated in a simulation study and on a published data-set.

2 Window width selection

We consider a stationary process $\{X_t\}_{t\in\mathbb{Z}}$ and for simplicity we always assume that $\mathbf{E}[X_t] = 0$. Furthermore we assume that $\{X_t\}_{t\in\mathbb{Z}}$ has a spectral density f which can be written as

$$f(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} R(k)e^{-ik\lambda}, \; -\pi \leq \lambda \leq \pi, \tag{1}$$

where $R(k) = Cov(X_0, X_k) = \mathbf{E}[X_0 X_k]$. In the following we denote by

$$\hat{R}(k) = n^{-1} \sum_{t=1}^{n-|k|} X_t X_{t+|k|} \; (|k| \leq n-1)$$

a nonparametric estimate of $R(k)$.
The representation (1) motivates the so called lag-window estimator for f:

$$\hat{f}(\lambda; b) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} w(kb)\hat{R}(k)e^{-ik\lambda}, \; -\pi \leq \lambda \leq \pi, \tag{2}$$

where the window width b is a function of n with $b = o(1)$ $(n \to \infty)$, $bn \to \infty$ $(n \to \infty)$ and w is a so called window.

Definition 1 *w is a window if $w : \mathbf{R} \to \mathbf{R}^+$ is even with $w(0) = 1$. w is a C^r-window, if the window w is r times continuously differentiable (if the support of w is finite, one-sided r-th continuous differentiability at the boundaries is sufficient) $(r \in \mathbf{N})$. A C^r-window has characteristic exponent r if $w^{(k)}(0) = 0$ $(k < r)$, $w^{(r)}(0) \neq 0$ $(r \in \mathbf{N})$ (cf. Priestley (1981)).*

The outcome of $\hat{f}(\lambda; b)$ usually depends strongly on the nuisance parameter b and therefore the problem of estimating b is of crucial importance.
Examples for windows are:

Bartlett window: $w(x) = max(0, 1 - |x|)$.

Tukey-Hanning window : $w(x) = \begin{cases} (1 + cos(\pi x))/2 & : \; |x| < 1, \\ 0 & : \; |x| \geq 1. \end{cases}$

The asymptotic bias and variance of the estimator in (2) are well known, under some regularity conditions it holds that (cf. Priestley (1981)):

Bartlett window:

$$\begin{aligned} \mathbf{E}[\hat{f}(\lambda; b)] - f(\lambda) &\sim -bf^{(1)}(\lambda), \\ Var(\hat{f}(\lambda; b)) &\sim \iota(\lambda) b^{-1} n^{-1} 2/3 (f(\lambda))^2. \end{aligned} \tag{3}$$

where $f^{(1)}(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} |k| R(k) e^{-ik\lambda}$ (generalized first derivative) and $\iota(\lambda) = 1 + 1_{[\{0,-\pi,\pi\}]}(\lambda)$.

C^2-window:

$$\mathbf{E}[\hat{f}(\lambda;b)] - f(\lambda) \sim -1/2w''(0)b^2 f^{(2)}(\lambda),$$
$$Var(\hat{f}(\lambda;b)) \sim \iota(\lambda)b^{-1}n^{-1}\int_{\infty}^{\infty} w^2(x)dx(f(\lambda))^2, \tag{4}$$

where $f^{(2)}(\lambda) = -(2\pi)^{-1}\sum_{k=-\infty}^{\infty} k^2 R(k)e^{-ik\lambda}$.

The case with the Bartlett window is interesting because this window spectral estimate is directly related to the blocklength ℓ for the blockwise bootstrap in time series. The connection is based on an equivalence of the bootstrap variance of the mean to the Bartlett-window spectral estimate at zero with window width ℓ^{-1} (see Bühlmann and Künsch (1994)).

The mean square error $MSE(\lambda;b) = \mathbf{E}[(\hat{f}(\lambda;b) - f(\lambda))^2]$ provides a local measure for the quality of the estimate at λ, the global behavior can be measured by the mean integrated square error $MISE(b) = \mathbf{E}[\int_{-\pi}^{\pi}(\hat{f}(\lambda;b) - f(\lambda))^2 d\lambda] = \int_{-\pi}^{\pi} MSE(\lambda;b)d\lambda$. We introduce now the asymptotic mean square error. Consider an estimator U_n for θ. The asymptotic bias $B_{\infty}(U_n,\theta)$ and the asymptotic variance $V_{\infty}(U_n)$ are defined as the leading terms of the bias and variance respectively such that $\mathbf{E}[U_n]-\theta \sim B_{\infty}(U_n,\theta)$ and $Var(U_n) \sim V_{\infty}(\theta)$. Then the asymptotic mean square error of U_n for θ is defined as $AMSE(U_n,\theta) = (B_{\infty}(U_n,\theta))^2 + V_{\infty}(U_n)$. By formula (3) or (4) respectively we get the asymptotic mean square error $AMSE(\lambda;b)$ and $AMISE(b)$. We define the optimal value of the local or global window width as

$$b_{opt}(\lambda) = \text{argmin}_b AMSE(\lambda;b), \quad b_{opt} = \text{argmin}_b AMISE(b).$$

Set down by Parzen (1957,1958), we obtain now by simple calculus:

for the Bartlett window:

$$b_{opt}(\lambda) = n^{-1/3}(\iota(\lambda)\frac{(f(\lambda))^2}{3(f^{(1)}(\lambda)^2)})^{1/3},$$
$$b_{opt} = n^{-1/3}(\frac{\int_{-\pi}^{\pi}(f(\lambda))^2 d\lambda}{3\int_{-\pi}^{\pi} f^{(1)}(\lambda)^2 d\lambda})^{1/3}. \tag{5}$$

for the C^2-window:

$$b_{opt}(\lambda) = n^{-1/5}(\iota(\lambda)\frac{\int_{-\infty}^{\infty} w^2(x)dx(f(\lambda))^2}{(w''(0))^2(f^{(2)}(\lambda))^2})^{1/5},$$
$$b_{opt} = n^{-1/5}(\frac{\int_{-\infty}^{\infty} w^2(x)dx \int_{-\pi}^{\pi}(f(\lambda))^2 d\lambda}{(w''(0))^2\int_{-\pi}^{\pi}(f^{(2)}(\lambda))^2 d\lambda})^{1/5}. \tag{6}$$

The formulas (5) and (6) are somewhat disappointing since the optimal window widths depend on the unknown spectral density. One way out of this impasse is to plug in some estimates for the unknown spectral density and its (generalized) derivatives. Our approach relies on an iterative plug-in method.

2.1 Iterative plug-in for local window width selection

Our procedure is in its ideas similar to the one of Brockmann et al. (1993). However, our context is completely different, the estimators we consider are in the frequency domain of stationary time series, whereas their estimates are in nonparametric regression for i.i.d. observations. We use an iterative scheme which consists of global- and local iteration-steps, i.e., we first estimate b_{opt} and then use this global estimate for a local estimate of $b_{opt}(\lambda)$ at some frequency λ.

Formula (5) or (6) respectively indicate, that one has to estimate different quantities depending on the unknown spectral density f. Denote by $\tilde{w}(.)$ and $\bar{w}(.)$ windows with finite support which can be different from the window $w(.)$ of the estimator defined in (2). We consider the following estimators:

$$\frac{1}{2}\int_{-\pi}^{\pi}((2\pi)^{-1}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-ik\lambda})^2 d\lambda \quad \text{for} \quad \int_{-\pi}^{\pi}(f(\lambda))^2 d\lambda$$

($(2\pi)^{-1}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-ik\lambda}$ is the periodogram);

$$\hat{f}_{\tilde{w}}(\lambda;b) = (2\pi)^{-1}\sum_{k=-\infty}^{\infty}\tilde{w}(kb)\hat{R}(k)e^{-ik\lambda} \quad \text{for} \quad f(\lambda);$$

$$\hat{f}_{\bar{w}}^{(1)}(\lambda;b) = (2\pi)^{-1}\sum_{k=-\infty}^{\infty}\bar{w}(kb)|k|\hat{R}(k)e^{-ik\lambda} \quad \text{for} \quad f^{(1)}(\lambda),$$

$$\hat{f}_{\bar{w}}^{(2)}(\lambda;b) = -(2\pi)^{-1}\sum_{k=-\infty}^{\infty}\bar{w}(kb)k^2\hat{R}(k)e^{-ik\lambda} \quad \text{for} \quad f^{(2)}(\lambda).$$

Remark 1. The estimate $((2\pi)^{-1}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-ik\lambda})^2/2$ is not consistent for $(f(\lambda))^2$, whereas the integral is $\sqrt{n}$-consistent (see Lemma 1, Section 4). For the integral of the (generalized) derivatives of the spectral density one has to use again a window having width b such that $b = o(1)$, $bn \to \infty$.

Remark 2. Because $k\hat{R}(k)$ and $k^2\hat{R}(k)$ are usually not decaying very fast as a function of the lag k we propose to take a splitted rectangular-cosine window for estimating $f^{(1)}(.)$ or $f^{(2)}(.)$ respectively, e.g.,

$$\bar{w}(x) = \begin{cases} 1 & : \quad |x| < 0.8 \\ (1+cos(5(x-0.8)\pi))/2 & : \quad 0.8 \le |x| < 1 \\ 0 & : \quad \text{otherwise} \end{cases}$$

The specific choice for the cut-off point 0.8 is based on some simulation results.

Procedure for the Bartlett window:
We denote by $\tilde{w}$ and $\bar{w}$ C^2-windows with characteristic exponent ≥ 2.

(I) Let $b_0 = n^{-1}$ be the starting window width, independent from the data.

(II) Global steps: According to (5) iterate

$$b_i = n^{-1/3}(\frac{1/2\int_{-\pi}^{\pi}((2\pi)^{-1}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-ik\lambda})^2 d\lambda}{3\int_{-\pi}^{\pi}(\hat{f}_{\tilde{w}}(\lambda;b_{i-1}n^{4/21}))^2 d\lambda})^{1/3}, \; i = 1,\ldots,4.$$

(III) Local step: According to (5),

$$\hat{b}_{opt}(\lambda) = n^{-1/3}(\iota(\lambda)\frac{(\hat{f}_{\tilde{w}}(\lambda; b_4 n^{4/21}))^2}{3(\hat{f}_{\bar{w}}^{(1)}(\lambda; b_4 n^{4/21}))^2})^{1/3}.$$

Procedure for the C^2-window:
To have a faster algorithm we start with $n^{-1/2}$; note that b_{opt} is also of a smaller order in this case. Denote by $\tilde{w}$ and $\bar{w}$ C^2-windows with characteristic exponent ≥ 2.

(I) Let $b_0 = n^{-1/2}$ be the starting window width, independent from the data.

(II) Global steps: According to (6) iterate

$$b_i = n^{-1/5}(\frac{\int_{\infty}^{\infty} w^2(x)dx 1/2 \int_{-\pi}^{\pi}((2\pi)^{-1}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-ik\lambda})^2 d\lambda}{(w''(0))^2\int_{-\pi}^{\pi}(\hat{f}_{\bar{w}}^{(2)}(\lambda; b_{i-1}n^{4/45}))^2 d\lambda})^{1/5}, \; i = 1, \ldots, 4.$$

(III) Local step: According to (6),

$$\hat{b}_{opt}(\lambda) = n^{-1/3}(\iota(\lambda)\frac{\int_{-\infty}^{\infty} w^2(x)dx(\hat{f}_{\tilde{w}}(\lambda; b_4 n^{4/45}))^2}{(w''(0))^2(\hat{f}_{\bar{w}}^{(2)}(\lambda; b_4 n^{4/45}))^2})^{1/5}.$$

Remark 3. In the global steps the deterministic terms are dominating, that is why the global steps stabilize the procedure (compare with Brockmann et al. (1993)). After 4 global steps the procedure achieves the correct asymptotic order, i.e., $b_4 = const.n^{-1/3}(1+o_P(1))$ or $b_4 = n^{-1/5}(1+o_P(1))$ respectively. Based on the results of a simulation study we propose to perform only one local step, further local iterations do not appear to improve the rates of the error terms. Moreover, they could be dominated by stochastic terms which again could cause a loss of accuracy of the algorithm.

Remark 4. The inflation factor $n^{4/21}$ or $n^{4/45}$ is motivated by asymptotics and leads to an estimate for the window width with best possible order. It can also be seen as an adjustment to the optimal order for estimating $f^{(1)}(.)$ or $f^{(2)}(.)$ respectively; note that the error terms arising from estimating the (generalized) derivatives are dominating. Smaller factors improve the rate of bias at the expense of variability (see Theorem 2).
We also tried to inflate only the window width for the estimates of the derivative $f^{(1)}(.)$ or $f^{(2)}(.)$ respectively, and to use no inflation factor for the local estimate of $f(.)$. Based on results from our simulation study it seems that in general this version of the procedure is less stable and less accurate.

Remark 5. Based on empirical evidence, we propose to take for $\tilde{w}$ the Tukey-Hanning window (see discussion following (2)) and for $\bar{w}$ the split rectangular-cosine window as given in Remark 2. A simulation study yields similar results by using the cut-off point 1.0, i.e., the rectangular window.

Remark 6. By using Parseval's identity for expressions of the form $\int_{-\pi}^{\pi}((2\pi)^{-1}\sum_{k=-c}^{c} g(k)e^{-i\lambda k})^2 d\lambda = (2\pi)^{-1}\sum_{k=-c}^{c} g^2(k)$ we can avoid numerical integration.

We are going to make the following assumption:

(B1) $\{X_t\}_{t\in\mathbb{Z}}$ is stationary with $\mathbf{E}[X_t] = 0$, $\sum_{i_1,\ldots,i_{h-1}=0}^{\infty} cum_h(X_0, X_{i_1}, \ldots, X_{i_{h-1}}) < \infty$ for $h \leq 8$ and

$$\sum_{i=0}^{\infty}(i+1)^4|R(i)| < \infty \text{ for the Bartlett-window procedure}$$
$$\sum_{i=0}^{\infty}(i+1)^6|R(i)| < \infty \text{ for the } C^2\text{-window procedure.}$$

(B2) $\bar{w}$ and $\tilde{w}$ are C^2-windows with finite support and characteristic exponent 2 (cf. Definition 1).

(B3) The frequency of interest $\lambda \in [-\pi, \pi]$ satisfies:

$$f^{(1)}(\lambda) \neq 0 \text{ for the Bartlett window procedure,}$$
$$f^{(2)}(\lambda) \neq 0 \text{ for the } C^2\text{-window procedure.}$$

Remark 7. The summability of cumulants up to order 8 holds for example for strong-mixing (cf. Rosenblatt (1985)) processes with $\sum_{i=0}^{\infty}(i+1)^6\alpha^{\delta/(14+\delta)}(i) < \infty, \mathbf{E}|X_t|^{8+\delta} < \infty$ $(\delta > 0)$.

Remark 8. In Remark 5 we propose a window $\bar{w}$ which is C^∞ with characteristic exponent ∞. This is more than we assume in assumption (B2) but does not change the asymptotic results in the following Theorems 1 or 2.

Remark 9. Very often (B3) holds at the frequencies $-\pi, 0, \pi$, where the spectral density has local extrema.

Theorem 1 *Assume that (B1)-(B3) hold.*
Then we have:

$$\hat{b}_{opt}(\lambda) = b_{opt}(\lambda)(1 + O_P(n^{-2/7})) \textit{ for the Bartlett window procedure,}$$
$$\hat{b}_{opt}(\lambda) = b_{opt}(\lambda)(1 + O_P(n^{-2/9})) \textit{ for the } C^2\textit{-window procedure.}$$

To see the effect of the characteristic exponent r of a window and of an arbitrary inflation factor n^ζ, we give a more general result. For such factors we have to iterate over m global steps, where m is the smallest integer satisfying $m \geq 2/(3\zeta)$ for the Bartlett window procedure and $m \geq 3/(10\zeta)$ for the C^2-window procedure respectively. Our assumptions are:

(B1') $\{X_t\}_{t\in\mathbb{Z}}$ is stationary with $\mathbf{E}[X_t] = 0$, $\sum_{i_1,\ldots,i_{h-1}=0}^{\infty} cum_h(X_0, X_{i_1}, \ldots, X_{i_{h-1}}) < \infty$ for $h \leq 8$ and for $r \in \mathbf{N}$

$$\sum_{i=0}^{\infty}(i+1)^{r+2}|R(i)| < \infty \text{ for the Bartlett-window procedure}$$
$$\sum_{i=0}^{\infty}(i+1)^{r+4}|R(i)| < \infty \text{ for the } C^2\text{-window procedure.}$$

(B2') $\bar{w}$ and $\tilde{w}$ are C^r-windows with finite support and characteristic exponent r.

Theorem 2 *Assume that (B1'),(B2') and (B3) hold.*
Then we have:

$$\hat{b}_{opt}(\lambda) = b_{opt}(\lambda)(1 + O(n^{-r/3+r\zeta}) + O(n^{-3\zeta}) + O_P(n^{-3\zeta/2}))(1 + o_P(max\{n^{-r/3+r\zeta}, n^{-3\zeta/2}\}))$$
$(0 < \zeta < 1/3)$ *for the Bartlett window procedure,*
$$\hat{b}_{opt}(\lambda) = b_{opt}(\lambda)(1 + O(n^{-r/5+r\zeta}) + O(n^{-\zeta}) + O_P(n^{-5\zeta/2}))(1 + o_P(max\{n^{-r/5+r\zeta}, n^{-5\zeta/2}\}))$$
$(0 < \zeta < 1/5)$ *for the C^2-window procedure.*

If the autocovariances decay exponentially, e.g. for ARMA-models, and for windows of order ∞, (B1') and (B2') hold for $r = \infty$. Then we almost achieve the parametric rate by choosing ζ arbitrarily close to 1/3 or 1/5 respectively, i.e.,

$$\hat{b}_{opt}(\lambda) = b_{opt}(\lambda)(1 + O_P(n^{-1/2+\varepsilon}))\ (\varepsilon > 0) \text{ for both window procedures.}$$

The proof of Theorem 1 is given in Section 4, the proof of Theorem 2 follows exactly the same lines.

2.2 Iterative plug-in for semi-local window width selection

Some problems remain at frequencies λ for which $f^{(1)}(\lambda) = 0$ or $f^{(2)}(\lambda) = 0$ (inflection point), respectively. In these cases the formulas (5) and (6) respectively for $b_{opt}(\lambda)$ do not hold anymore. As in Brockmann et al. (1993) we propose an alternative procedure which is a step away from the purely local algorithm above. We replace the local step (III) by

$$\hat{b}_{opt}^{semiloc.}(\lambda) = n^{-1/3}\left(\frac{\iota(\lambda)(\hat{f}_{\tilde{w}}(\lambda; b_4 n^{4/21}))^2}{3\int_{-c}^{c}(\hat{f}_{\tilde{w}}^{(1)}(\lambda; b_4 n^{4/21}))^2 d\lambda}\right)^{1/3},\ c = b_4 n^{4/21},$$

for the Bartlett window;

$$\hat{b}_{opt}^{semiloc.}(\lambda) = n^{-1/5}\left(\frac{\iota(\lambda)\int_{-\infty}^{\infty} w^2(x)dx(\hat{f}_{\tilde{w}}(\lambda; b_4 n^{4/45}))^2}{(w''(0))^2\int_{-c}^{c}(\hat{f}_{\tilde{w}}^{(2)}(\lambda; b_4 n^{4/45}))^2 d\lambda}\right)^{1/5},\ c = b_4 n^{4/45},$$

for the C^2-window.

If (B3) holds we can always use the purely local estimation and do not need the modification above. However, if we are interested to estimate the spectral density as a function of frequencies in $[0, \pi]$ with locally adaptive window widths, we suggest to use the modified version to obtain a curve without some artificial spiky peaks at frequencies, where (B3) fails to hold.

3 Simulation Study

We investigate the finite sample behavior of our procedures. The data are simulated from autoregressive processes $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$:
(M1) AR(1) model: $\phi_1 = 0.8$, ε_t i.i.d. $\mathcal{N}(0, 1)$.
(M2) AR(2) model: $\phi_1 = 1.372$, $\phi_2 = -0.677$, ε_t i.i.d. $\mathcal{N}(0, 0.4982)$.
(M3) AR(5) model: $\phi_1 = 0.9$, $\phi_2 = -0.4$, $\phi_3 = 0.3$, $\phi_4 = -0.5$, $\phi_5 = 0.3$, ε_t i.i.d. $\mathcal{N}(0, 1)$.
The spectral densities of (M1)-(M3) are given in Figure 1 and 2. The lag(1)-correlations

in (M1) and (M2) are approximately equal, whereas the lag(2)-correlations differ considerably. By choosing $\sigma^2 = 0.4982$ in (M2) we force $Var(X_t)$ in (M2) to be equal to $Var(X_t)$ in (M1). The model (M2) exhibits a 'pseudo-periodic behavior', the corresponding autocorrelation function can be described as 'damped periodic' (cf. Priestley (1981), Chapter 3.5.3); this phenomenon is not present in model (M1). The models (M1) and (M2) are considered in simulation studies in Künsch (1989) and Bühlmann (1994). We choose the sample size $n = 480$ and $n = 120$. To compare the estimation procedures at the different models we do not vary the distribution of the innovations. Our procedures are not restricted to the normal distribution. Our study is always based on 300 simulations.

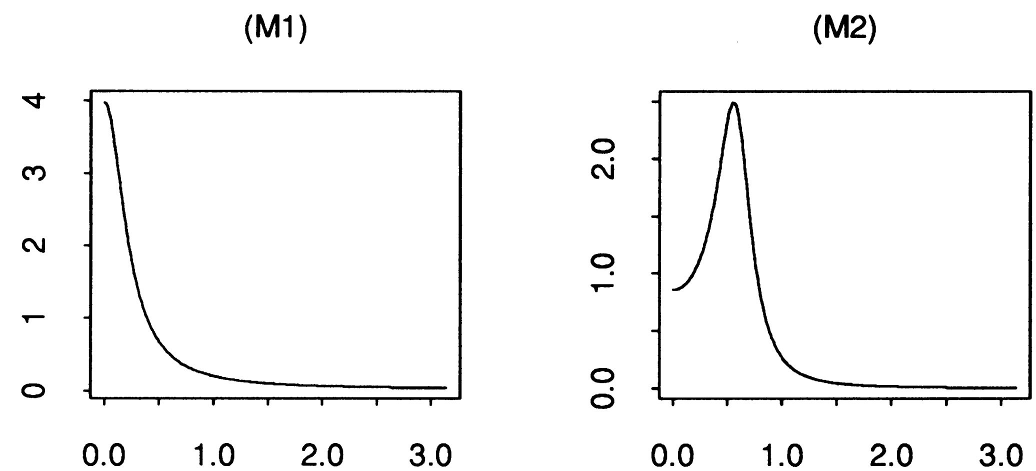

Figure 1: Spectral densities of (M1) and (M2)

3.1 Estimation at frequency zero

We consider the purely local estimates with adaptively chosen window widths for the spectrum of a stationary process at zero.

To estimate the spectrum at zero we consider $\hat{f}_w(0; b)$ as defined in (2) with
(E1) $w(x) = \max(0, 1 - |x|)$ (Bartlett window),
(E2) $w(x) = \max(0, 1 - |.|) * \max(0, 1 - |.|)(x)$ (convolution of the Bartlett window).
By using (5) and (6) we can compute the optimal window width (under the assumed model). The following table reports the performance of $\hat{f}(0; \hat{b}_{opt}(0))$ in comparison with $\hat{f}(0; b_{opt}(0))$. The estimates $\hat{b}_{opt}(0)$ are calculated according to our proposal in Remark 5, and using the inflation factors minimizing the total error as in Theorem 1. For each of the 300 simulations we compute $\hat{f}(0; \hat{b}_{opt}(0))$ with its individual window width $\hat{b}_{opt}(0)$. We denote by $RMSE$ the relative mean square error $MSE(\hat{f}(0, \hat{b}_{opt}(0)))/f(0)^2$ and by MSE-ratio the ratio $MSE(\hat{f}(0, \hat{b}_{opt}(0)))/MSE(\hat{f}(0, b_{opt}(0)))$. An estimate of the standard deviation of $RMSE$ is given in parentheses.

$n=480$	$f(0)$	$\mathbf{E}[\hat{f}(0;\hat{b}_{opt}(0))]$	$S.D.(\hat{f}(0;\hat{b}_{opt}(0)))$	$RMSE$	MSE-ratio
(M1),(E1)	3.98	2.95	0.91	0.118 (0.006)	1.09
(M1),(E2)	3.98	3.06	0.92	0.106 (0.006)	1.06
(M2),(E1)	0.85	1.00	0.22	0.098 (0.007)	1.81
(M2),(E2)	0.85	0.94	0.27	0.109 (0.008)	1.70
(M3),(E1)	0.99	0.80	0.16	0.064 (0.004)	1.33
(M3),(E2)	0.99	0.85	0.16	0.048 (0.004)	0.87

$n=120$	$f(0)$	$\mathbf{E}[\hat{f}(0;\hat{b}_{opt}(0))]$	$S.D.(\hat{f}(0;\hat{b}_{opt}(0)))$	$RMSE$	MSE-ratio
(M1),(E1)	3.98	2.00	0.91	0.299 (0.010)	1.14
(M1),(E2)	3.98	2.13	0.95	0.274 (0.010)	1.07
(M2),(E1)	0.85	1.07	0.34	0.227 (0.023)	1.27
(M2),(E2)	0.85	1.21	0.41	0.336 (0.030)	1.96
(M3),(E1)	0.99	0.77	0.23	0.102 (0.006)	0.89
(M3),(E2)	0.99	0.82	0.26	0.098 (0.007)	0.63

The method (E2) usually has smaller bias than (E1) at the expense of variability. From a mean square error point of view there is no overall optimality, (E2) is better in model (M1) and (M3). This is in accordance with the theoretical minimal relative $AMSE$. The performance measured in terms of $RMSE$ is in all cases quite satisfactory.

Our adaptive procedure works well in model (M1) considering the MSE-ratios, this might be due to the simple structure of the autocovariance function. The 'damped-periodic' autocovariance function in model (M2) might be an explanation for the bigger MSE-ratios. The large MSE-ratio in (M3),(E1) for $n = 480$ can be explained by the behavior of the spectral density around zero, in particular the fluctuation of the generalized first derivative: the optimal window width in (E1) is larger than in (E2) and hence the estimation is disturbed by other local extrema of the spectral density in a neighborhood around zero (cf. Figure 2).

The variability of $\hat{b}_{opt}(0)$ is in all the cases quite large. The same fact also occurs in the global bandwidth selection procedure of Beltrão and Bloomfield (1987). Nevertheless, our procedure works quite well for the original aim, i.e., the estimation of the spectral density at zero.

3.2 Estimation of the whole spectrum

We consider the semi-local version of our plug-in procedure for the estimation of the whole spectrum. For comparison we also studied a global plug-in procedure which is our algorithm stopped after the iteration over the global steps (II).

Figure 2 shows the estimates with semi-locally and globally adaptive chosen window widths based on a realization of model (M3) with sample size $n = 480$. The local procedure is everywhere at least a little bit better and considerably better in the peaks. This also justifies to use a local procedure for estimating the spectral density at zero, i.e. at a local extremum. The window width is adapting to the flat and rough parts of the true underlying spectral density.

A common criterion for comparison is the (one sided) mean integrated relative square error $(MIRSE)\,\mathbf{E}[\int_0^\pi((\hat{f}(\lambda)-f(\lambda))/f(\lambda))^2 d\lambda] = \int_0^\pi RMSE(\lambda)d\lambda$ (cf. Beltrão and Bloomfield (1987)). We actually computed the integrated square error as

$$(n+1)^{-1}\sum_{j=0}^{n}((\hat{f}(\lambda_j)-f(\lambda_j))/f(\lambda_j))^2,\ \lambda_j=\pi j/n.$$

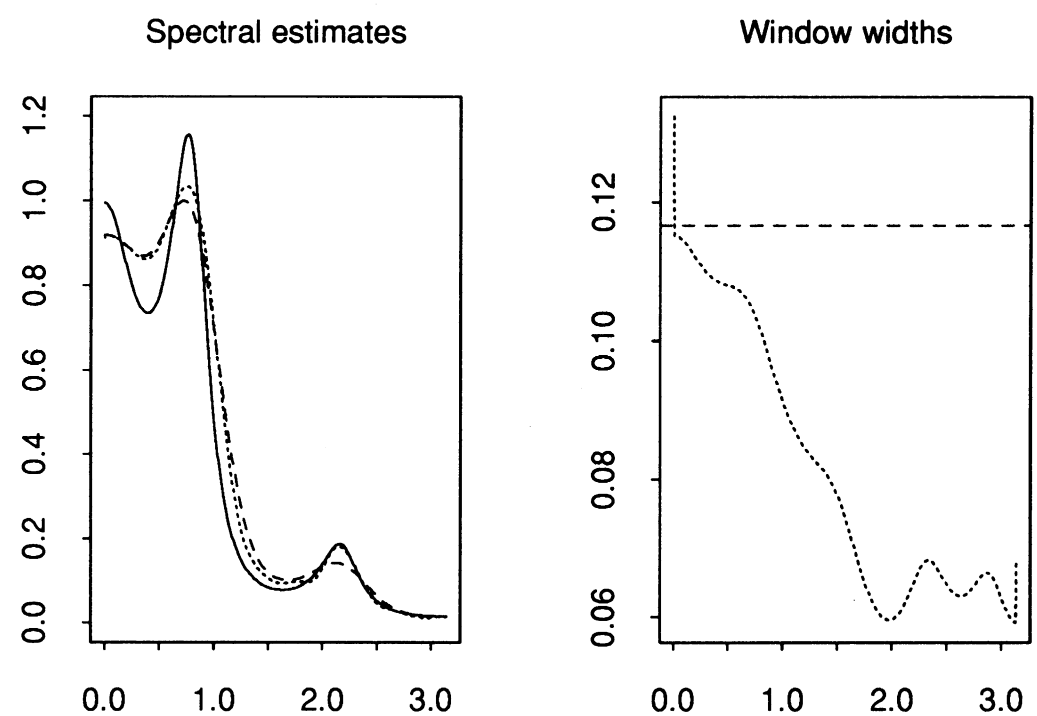

Figure 2: Estimation (E2) for (M3) with n=480; true spectral density (solid line), semi-local procedure (small dashes), global procedure (large dashes).

The following tables show the MIRSE, computed over 300 simulations. The estimates are according to Remark 5 with cut-off point 1.0 for $\bar{w}$. An estimate of the standard deviation of $MIRSE$ is given in parentheses.

$n=480$	$MIRSE$, local	$MIRSE$, global
(M1),(E1)	0.075 (0.002)	0.077 (0.002)
(M1),(E2)	0.059 (0.001)	0.040 (0.001)
(M2),(E1)	0.172 (0.003)	1.011 (0.023)
(M2),(E2)	0.082 (0.002)	0.063 (0.003)
(M3),(E1)	0.119 (0.003)	0.441 (0.010)
(M3),(E2)	0.070 (0.002)	0.132 (0.010)

$n = 120$	$MIRSE$, local	$MIRSE$, global
(M1),(E1)	0.201 (0.007)	0.246 (0.010)
(M1),(E2)	0.164 (0.005)	0.132 (0.005)
(M2),(E1)	0.677 (0.023)	4.168 (0.169)
(M2),(E2)	0.357 (0.013)	0.486 (0.031)
(M3),(E1)	0.473 (0.017)	1.580 (0.052)
(M3),(E2)	0.292 (0.012)	0.590 (0.028)

The estimation method (E2) (C^2-window) outperforms (E1). The semi-local procedure is always better for method (E1). The global estimate usually has a smaller variability (cf. Lemma 1 (iii) or Lemma 7 (ii)) than its local counterpart. However, the global procedure estimates a window width which is not optimal with respect to the $MIRSE$. The tendency of the global procedure to have a smaller variability explains the better performance of the global procedure in (M1), (E2) for both sample sizes and in (M2), (E2) for $n = 480$. But the semi-local procedure is in these cases only slightly worse, whereas the gain for more spiky densities as in (M3) is much bigger, and we conjecture that it yields better results particularly in the peaks which is also interesting from an explanatory point of view.

Finally we tested our method on the delicate AR(4) data-set from Percival and Walden (1993) (equation (46a) and figures 45, 313, 314). Tapering of this data-set is crucial (see

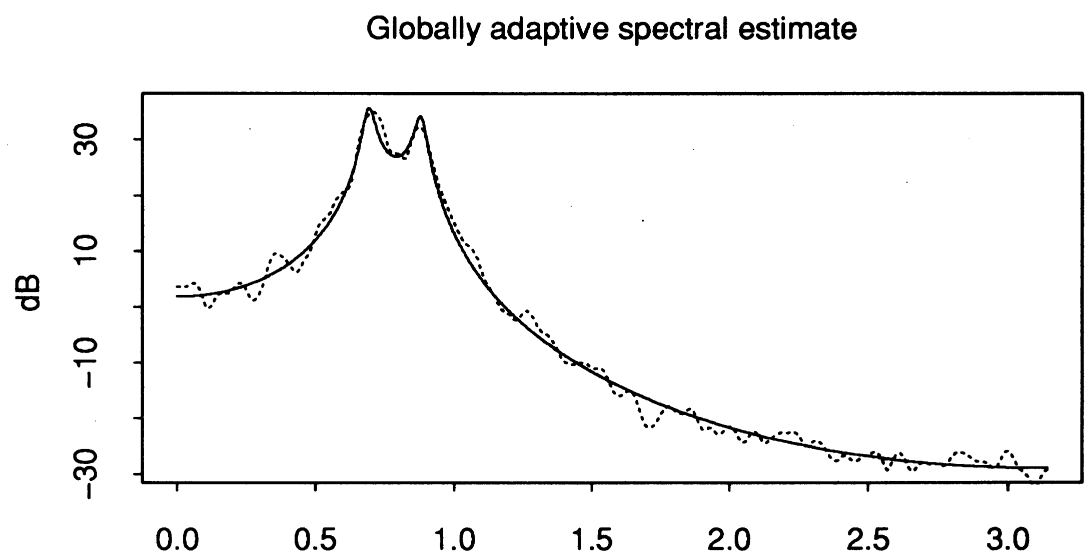

Figure 3: AR(4) series from Percival and Walden: true spectral density (solid line), global procedure (small dashes).

Percival and Walden, chapter 6.18), so we modified our procedure in the obvious way by noting that formula (6) will change under tapering. We used a Parzen lag-window estimator, as in Percival and Walden, and for ease of computability the Hanning taper which differs a little from the NW=2 dpss-taper in Percival and Walden. The global procedure does at least as good as the method in Percival and Walden, our inverse adaptive window width is $1/\hat{b}_{opt} = 170.9$, whereas they mainly consider $1/b = 128$ and $1/b = 256$ (see Figure 3). We point out that our method is not designed for estimating the spectrum on the variance-stabilizing dB-scale, however for reasons of comparison with Percival and Walden and because the true underlying spectral density is almost zero and flat at low and higher frequencies, we show our result on this scale. The almost constant zero-ness of the spectral density at low and higher frequencies makes it more difficult for the local procedure to work well. The ratio $(f/f^{(2)})^2$, which governs the optimal window width, takes values in a reasonable range, and thus compensates the flat almost zero-ness (i.e., $f \approx 0$, $f^{(2)} \approx 0$). But estimation of $f^{(2)}$ is less accurate than estimation of f and usually destroys the latter compensation effect. The locally adaptive procedure selects at these frequencies a very small window width, yielding almost the periodogram-estimates for the lag-window estimator. The global procedure overcomes this problem.

4 Proofs

We first give the proof in the case of the Bartlett window procedure. We denote by $G_B(b) = n^{-1/3}\left(\frac{\frac{1}{2}\int_{-\pi}^{\pi}((2\pi)^{-1}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-i\lambda k})^2 d\lambda}{3\int_{-\pi}^{\pi}(\hat{f}_{\tilde{w}}^{(1)}(\lambda;b))^2 d\lambda}\right)^{1/3}$ and $L_B(\lambda;b) = n^{-1/3}\left(\iota(\lambda)\frac{(\hat{f}_{\tilde{w}}(\lambda;b))^2}{3(\hat{f}_{\tilde{w}}^{(1)}(\lambda;b))^2}\right)^{1/3}$ the expressions corresponding to the global and local steps for the Bartlett window. For the iteration steps in our procedure we have to evaluate $G_B(.)$ and $L_B(\lambda;.)$ at some stochastic b. This causes some additional difficulties (see Lemma 2 and 3).

In the sequel we denote by *const.* different constants. These constants are functions of higher order cumulants of $\{X_t\}_{t\in\mathbb{Z}}$ and we implicitly assume that these constants do not vanish, which usually holds under dependence. Otherwise we would even get better rates of convergence.

Lemma 1 *Assume that (B1)-(B3) hold. Let b be deterministic with $b = o(1)$, $bn \to \infty$ $(n \to \infty)$.*
Then:

(i)

$$\mathbf{E}[1/2\int_{-\pi}^{\pi}((2\pi)^{-1}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-ik\lambda})^2 d\lambda] - \int_{-\pi}^{\pi}(f(\lambda)^2 d\lambda \sim const.n^{-1},$$

$$Var(1/2\int_{-\pi}^{\pi}((2\pi)^{-1}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-ik\lambda})^2 d\lambda) \sim const.n^{-1}.$$

(ii)

$$\mathbf{E}[\int_{-\pi}^{\pi}(\hat{f}_{\tilde{w}}^{(1)}(\lambda;b))^2 d\lambda] - \int_{-\pi}^{\pi}(f^{(1)}(\lambda)^2 d\lambda \sim const.b^2 + const.b^{-3}n^{-1},$$

$$Var(\int_{-\pi}^{\pi}(\hat{f}_{\tilde{w}}^{(1)}(\lambda;b))^2 d\lambda) \sim const.b^{-5}n^{-2} + const.n^{-1},$$

(iii)

$$G_B(b) = \begin{cases} b_{opt}(1 + O(\max\{b^2, b^{-3}n^{-1}\}) + O_P(\max\{b^{-5/2}n^{-1}, n^{-1/2}\})) & , \quad b^{-3}n^{-1} = O(1), \\ const.b(1 + o_P(1)) & , \quad b^{-3}n^{-1} \to \infty. \end{cases}$$

Proof:
(i) By Parseval's identity we have:

$$\int_{-\pi}^{\pi} ((2\pi)^{-1} \sum_{k=-n+1}^{n-1} \hat{R}(k) e^{-ik\lambda})^2 d\lambda = (2\pi)^{-1} \sum_{k=-n+1}^{n-1} \hat{R}^2(k).$$

Assertion (i) follows now by Lemma 11 (i) and (ii).
(ii) As above we get:

$$\int_{-\pi}^{\pi} (\hat{f}_{\bar{w}}^{(1)}(\lambda; b))^2 d\lambda = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \bar{w}^2(kb) k^2 \hat{R}^2(k).$$

Assertion (ii) follows now from Lemma 11 (iii) and (iv).
(iii) The assertion follows by a Taylor expansion and (i) and (ii). □

Lemma 1 tells us only about the behavior for some deterministic window width b. The hope is that for a stochastic $\tilde{b} = b(1 + o_P(1))$ (b deterministic) the iteration-function $G_B(\tilde{b})$ behaves asymptotically the same. In the next Lemma we give a uniform result in an arbitrary small compact set, i.e., we show stochastic equicontinuity for $G_B(.)$ which allows us to describe the behavior of the iteration-function for some stochastic $\tilde{b}$.

Lemma 2 *Assume that (B1)-(B3) hold. Let b_0 be deterministic with $b_0 = o(1)$, $b_0 n \to \infty$ $(n \to \infty)$. Denote by $U_{b_0}(\delta) = [b_0(1-\delta), b_0(1+\delta)]$ $(\delta > 0)$. Let $a(n, b_0)$ be a normalizing constant such that $a(n, b_0) n^{1/3} G_B(b_0) = O_P(1)$, $(a(n, b_0) n^{1/3} G_B(b_0))^{-1} = O_P(1)$.*
Then: $\forall \varepsilon > 0\ \forall \eta > 0\ \exists \delta\ (0 < \delta < 1)$ and $\exists n_0 \in \mathbf{N}$ such that

$$\mathbf{P}[\sup_{b \in U_{b_0}(\delta)} a(n, b_0) n^{1/3} |G_B(b) - G_B(b_0)| > \varepsilon] < \eta \quad \forall n \geq n_0 \ (b \text{ deterministic}).$$

Proof: The proof is given in Section 4.1. □

Lemma 3 *Assume that (B1)-(B3) hold. Let $\tilde{b} = b(1 + o_P(1))$, where b is deterministic with $b = o(1)$, $bn \to \infty$ $(n \to \infty)$.*
Then: $G_B(\tilde{b}) = G_B(b)(1 + o_P(1))$.

Proof: We have:

$$\begin{aligned} & a(n,b)|G_B(\tilde{b}) - G_B(b)| \\ \leq & \sup_{\tilde{b} \in U_b(\delta)} a(n,b)|G_B(\tilde{b}) - G_B(b)| 1_{[\tilde{b} \in U_b(\delta)]} + a(n,b)(|G_B(\tilde{b})| + |G_B(b)|) 1_{[\tilde{b} \notin U_b(\delta)]} \\ =: & F_1(n) + F_2(n), \end{aligned} \tag{7}$$

$a(n, b)$ as in Lemma 2.
Because $|\tilde{b} - b| = o_P(b)$ it holds that for every $\delta > 0$: $n^{1/3} F_2(n) = o_P(1)$. Hence by Lemma 2 and (7) we obtain: $\forall \varepsilon > 0\ \forall \eta > 0\ \exists n_0 \in \mathbf{N}$ such that

$$\mathbf{P}[a(n,b) n^{1/3} |G_B(\tilde{b}) - G_B(b)| > \varepsilon] < \eta,$$

i.e., $G_B(\tilde{b}) - G_B(b) = o_P(n^{-1/3}(a(n,b))^{-1})$ and therefore:
$G_B(\tilde{b}) = G_B(b) + o_P(n^{-1/3}(a(n,b))^{-1}) = G_B(b)(1 + o_P(1))$. □

By Lemma 1 (iii) and Lemma 3 we are able to control the global steps (II) of our procedure. We denote by c_i some constants in $\mathbf{R}$.

$$b_1 = c_1 n^{-17/21}(1 + o_P(1)) \text{ (by Lemma 1 (iii))};$$
$$b_k = G_B(c_{k-1} n^{-1+k4/21}(1 + o_P(1)) = c_k n^{-1+k4/21}(1 + o_P(1)), \; k = 2, 3;$$
$$b_4 = G_B(c_3 n^{-5/21})(1 + o_P(1)) = b_{opt}(1 + o_P(1)).$$

Now, let us in turn consider the local step (III). As for the global steps we analyze first the behavior for some deterministic window width b and show then that the effect of a stochastic width $\tilde{b}$ with the same order is negligible.

Lemma 4 *Assume that (B1)-(B3) hold. Let b be deterministic with $b = o(1)$, $bn \to \infty$ $(n \to \infty)$.*
Then:

(i)

$$\mathbf{E}[(\hat{f}_{\tilde{w}}(\lambda; b))^2] - (f(\lambda))^2 \sim const.b^2 + const.b^{-1}n^{-1},$$
$$Var((\hat{f}_{\tilde{w}}(\lambda; b))^2) \sim const.b^{-1}n^{-1}.$$

(ii)

$$\mathbf{E}[(\hat{f}^{(1)}_{\tilde{w}}(\lambda; b))^2] - (f^{(1)}(\lambda))^2 \sim const.b^2 + const.b^{-3}n^{-1},$$
$$Var((\hat{f}^{(1)}_{\tilde{w}}(\lambda; b))^2) \sim const.b^{-3}n^{-1}.$$

(iii)

$$L_B(\lambda; b) = b_{opt}(\lambda)(1 + O(b^2) + O_P(b^{-3/2}n^{-1/2})).$$

Proof:
(i) It is well known that

$$\mathbf{E}[\hat{f}_{\tilde{w}}(\lambda; b)] - f(\lambda) \sim const.b^2,$$
$$Var(\hat{f}_{\tilde{w}}(\lambda; b)) \sim const.b^{-1}n^{-1},$$

(see also (3)). Hence the assertion for the expectation follows. For the variance we write:

$$Var((\hat{f}_{\tilde{w}}(\lambda; b))^2) =$$
$$(\frac{1}{2\pi})^2 \sum_{k_1,\ldots,k_4=-\infty}^{\infty} \tilde{w}(k_1 b) \ldots \tilde{w}(k_4 b) e^{-i\lambda(k_1+\ldots+k_4)} Cov(\hat{R}(k_1)\hat{R}(k_2), \hat{R}(k_3)\hat{R}(k_4)).$$

The assertion follows now by Lemma 11 (v) (see Section 4.1).
(ii) By (8) and (9) (see Section 4.1)) we get:

$$\mathbf{E}[\hat{f}^{(1)}_{\tilde{w}}(\lambda; b)] - f^{(1)}(\lambda) \sim const.b^2,$$
$$Var(\hat{f}^{(1)}_{\tilde{w}}(\lambda; b)) \sim const.b^{-3}n^{-1}.$$

Hence the assertion for the expectation follows. Similarly as for $Var(\hat{f}_{\tilde{w}}(\lambda; b))$ the result for the variance follows by Lemma 11 (v).
(iii) follows directly from (i), (ii) and a Taylor expansion. □

Very similarly to the global steps we give the results for the local step which involves a stochastic window width.

Lemma 5 *Assume that (B1)-(B3) hold. Denote by $U_{b_0}(\delta) = [b_0(1-\delta), b_0(1+\delta)]$ $(\delta > 0)$, where $b_0 \sim const.n^{-1/3+\zeta}$ $(0 < \zeta < 1/3)$ is deterministic. Let $d_n = min\{b_0^{-2}, b_0^{3/2} n^{1/2}\}$.
Then: $\forall \varepsilon > 0$ $\forall \eta > 0$ $\exists \delta$ $(0 < \delta < 1)$ and $\exists n_0 \in \mathbf{N}$ such that*

$$\mathbf{P}[\sup_{b \in U_{b_0}(\delta)} n^{1/3} d_n |L_B(b) - L_B(b_0)| > \varepsilon] < \eta \ \forall n \geq n_0 \ (b \ deterministic).$$

Proof: The proof is given in Section 4.1. □

Lemma 6 *Assume that (B1)-(B3) hold. Let $\tilde{b} = b(1+o_P(1))$, where $b \sim const.n^{-1/3+\zeta}$ $(0 < \zeta < 1/3)$ is deterministic.
Then: $L_B(\lambda; \tilde{b}) = L_B(\lambda; b)(1 + o_P(max\{n^{-2/3+2\zeta}, n^{-3\zeta/2}\}))$.*

Proof: As in the proof of Lemma 3 we can show by using Lemma 4 and 5 that $L_B(\lambda; \tilde{b}) - L_B(\lambda; b) = o_P(n^{-1/3} d_n^{-1})$, with $d_n = min\{b_0^{-2}, b_0^{3/2} n^{1/2}\}$. This completes the proof. □

By Lemma 4 and 6 we obtain for the local step (III):

$$\begin{aligned}\hat{b}_{opt}(\lambda) &= L_B(\lambda; b_4 n^{4/21}) = L_B(\lambda; const.n^{-1/3+4/21})(1 + o_P(n^{-2/7})) \\ &= b_{opt}(\lambda)(1 + O_P(n^{-2/7})).\end{aligned}$$

Note that the inflation factor $n^{4/21}$ is of optimal order leading to the error term $O_P(n^{-2/7})$. Lemma 4 (i) and (ii) tell that the O-terms in Lemma 4 (iii) cannot be improved (unless one of the constants vanishes). This explains why further iterations do not improve the asymptotic rate in Theorem 1.

The proof of Theorem 1 in the case of the C^2-window is very similar. Denote by $G_C(b) = n^{-1/5}(\frac{\int_{-\infty}^{\infty} w^2(x)dx 1/2 \int_{-\pi}^{\pi}((2\pi)^{-1} \sum_{k=-n+1}^{n-1} \hat{R}(k) e^{-i\lambda k})^2}{(w''(0))^2 \int_{-\pi}^{\pi} (\hat{f}_{\tilde{w}}^{(2)}(\lambda; b))^2 d\lambda})^{1/5}$ and
$L_C(\lambda; b) = n^{-1/5}(\frac{\int_{-\infty}^{\infty} w^2(x)dx (\hat{f}_{\tilde{w}}(\lambda; b))^2}{(w''(0))^2 (\hat{f}_{\tilde{w}}^{(2)}(\lambda; b))^2})^{1/5}$ the expressions corresponding to the C^2-window.
Instead of Lemma 1 (ii) and (iii) we have:

Lemma 7 *Assume that (B1)-(B3) hold. Let b be deterministic with $b = o(1)$, $bn \to \infty$ $(n \to \infty)$.
Then:*

(i)

$$\mathbf{E}[\int_{-\pi}^{\pi} (\hat{f}_{\tilde{w}}^{(2)}(\lambda; b))^2 d\lambda] - \int_{-\pi}^{\pi} (f^{(2)}(\lambda)^2 d\lambda \sim const.b^2 + const.b^{-5} n^{-1},$$

$$Var(\int_{-\pi}^{\pi} (\hat{f}_{\tilde{w}}^{(2)}(\lambda; b))^2 d\lambda) \sim const.b^{-9} n^{-2} + const.n^{-1},$$

(ii)

$$G_C(b) = \begin{cases} b_{opt}(1+O(\max\{b^2, b^{-5}n^{-1}\}) + O_P(\max\{b^{-9/2}n^{-1}, n^{-1/2}\})) & , \quad b^{-5}n^{-1} = O(1), \\ const.b(1+o_P(1)) & , \quad b^{-5}n^{-1} \to \infty. \end{cases}$$

Proof:
(i) By Parseval's identity we have:

$$\int_{-\pi}^{\pi} (f^{(2)}(\lambda))^2 d\lambda = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} k^4 R^2(k),$$
$$\int_{-\pi}^{\pi} (\hat{f}_{\bar{w}}^{(2)}(\lambda; b))^2 d\lambda = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \bar{w}^2(kb) k^4 \hat{R}^2(k).$$

Hence (i) follows by Lemma 11 (iii) and (iv).
Assertion (ii) follows by (i), Lemma 1 (i) and a Taylor expansion. □

Lemma 8 *Assume that (B1)-(B3) hold. Let $\tilde{b} = b(1+o_P(1))$, where b is deterministic with $b = o(1)$, $bn \to \infty$ $(n \to \infty)$.*
Then: $G_C(\tilde{b}) = G_C(b)(1+o_P(1))$.

Proof: The Lemma can be shown in the same way as Lemma 3. We prove stochastic equicontinuity (see Lemma 2) with the same arguments. Note that we have to replace some constants, e.g. $n^{-1/3}$ by $n^{-1/5}$, b^{-3} by b^{-5} etc... □

By Lemma 7 (ii) and Lemma 8 we analyze the global steps (II) of our procedure. We denote by c_i some constants in $\mathbf{R}$.

$$\begin{aligned} b_1 &= c_1 n^{-37/90}(1+o_P(1)) \text{ (by Lemma 7 (ii))}; \\ b_k &= G_C(c_{k-1}n^{-1/2+k4/45}(1+o_P(1)) = c_k n^{-1/2+k4/45}(1+o_P(1)), \ k = 2,3; \\ b_4 &= G_B(c_3 n^{-13/90}(1+o_P(1)) = b_{opt}(1+o_P(1)). \end{aligned}$$

For the local steps we need the following two Lemmas.

Lemma 9 *Assume that (B1)-(B3) hold. Let $b \sim const.n^{-1/5+\zeta}$ $(0 < \zeta < 1/5)$ be deterministic.*
Then:

(i)

$$\mathbf{E}[(\hat{f}_{\bar{w}}^{(2)}(\lambda; b))^2] - (f^{(2)}(\lambda))^2 \sim const.b^2 + const.b^{-5}n^{-1},$$
$$Var((\hat{f}_{\bar{w}}^{(2)}(\lambda; b))^2) \sim const.b^{-5}n^{-1}.$$

(ii)

$$L_C(\lambda; b) = b_{opt}(\lambda)(1+O(b^2) + O_P(b^{-5/2}n^{-1/2}).$$

Proof: For assertion (i), the expectation part follows from (8) and (9), the variance part follows by Lemma 11 (v). Assertion (ii) follows by (i), Lemma 4 (i) and a Taylor expansion. □

Lemma 10 *Assume that (B1)-(B3) hold. Let* $\tilde{b} = b(1 + o_P(1))$, *where* $b \sim const.n^{-1/5+\zeta}$ $(0 < \zeta < 1/5)$ *is deterministic.*
Then: $L_C(\lambda; \tilde{b}) = L_C(\lambda; b)(1 + o_P(\max\{n^{-2/5+2\zeta}, n^{-5\zeta/2}\})$.

Proof: The Lemma can be shown in the same way as Lemma 6. We prove stochastic equicontinuity (see Lemma 5) with the same arguments. Details are left to the reader. □

By Lemma 9 and 10 we complete the proof of Theorem 1 for the C^2-window, the inflation factor $n^{4/45}$ leads to the error term $O_P(n^{-2/9})$. Lemma 4 (i) and Lemma 9 (i) tell that the O-terms in Lemma 9 (ii) cannot be improved (unless one of the constants vanishes), therefore further local iterations do not improve the rate for the C^2-window in Theorem 1.

4.1 Additional proofs

We first give some results for the estimated autocovariances.

Lemma 11 *Assume that (B1)-(B3) hold, if necessary a distinction between the Bartlett- and the* C^2*-window procedure will be made with the index* $r = 1$ *and* $r = 2$, *respectively. Let* b *be deterministic with* $b = o(1)$, $bn \to \infty$ $(n \to \infty)$.
Then:

(i)

$$\frac{1}{2}\mathbf{E}[\sum_{k=-n+1}^{n-1} \hat{R}^2(k)] - \sum_{k=-\infty}^{\infty} R^2(k) \sim const.n^{-1}.$$

(ii)

$$\sum_{k_1,k_2=-n+1}^{n-1} Cov(\hat{R}^2(k_1), \hat{R}^2(k_2)) \sim const.n^{-1}.$$

(iii)

$$\mathbf{E}[\sum_{k=-\infty}^{\infty} \tilde{w}^2(kb)k^{2r}\hat{R}^2(k)] - \sum_{k=-\infty}^{\infty} k^{2r}R^2(k) \sim const.b^2 + const.b^{-2r-1}n^{-1}, r = 1, 2$$

(iv)

$$\sum_{k_1,k_2=-\infty}^{\infty} \tilde{w}^2(k_1 b)\tilde{w}^2(k_2 b)(k_1)^{2r}(k_2)^{2r} Cov(\hat{R}^2(k_1), \hat{R}^2(k_2))$$
$$\sim const.b^{-4r-1}n^{-2} + const.n^{-1},\ r = 1, 2.$$

(v) If $b^{-3}n^{-1} = o(1)$,

$$\sum_{k_1,\ldots,k_4=-\infty}^{\infty} \tilde{w}(k_1 b)\ldots\tilde{w}(k_4 b)|k_1|^r \ldots |k_4|^r e^{-i\lambda(k_1+\ldots+k_4)} Cov(\hat{R}(k_1)\hat{R}(k_2), \hat{R}(k_3)\hat{R}(k_4))$$
$$\sim const.b^{-2r-1}n^{-1},\ r = 0, 1, 2.$$

Proof: (i) Observe that

$$\mathbf{E}[\hat{R}(k)] = R(k)(1-|k|/n), \tag{8}$$

$$Cov(\hat{R}(k_1),\hat{R}(k_2)) = n^{-2}\sum_{t_1=1}^{n-|k_1|}\sum_{t_2=1}^{n-|k_2|}\{R(t_1-t_2)R(t_1-t_2+k_1-k_2)$$

$$+\quad R(t_1-t_2-k_2)R(t_1-t_2+k_1)+cum_4(X_{t_1},X_{t_1+|k_1|},X_{t_2},X_{t_2+|k_2|})\}. \tag{9}$$

By (8) and assumption (B1): $\sum_{k=-n+1}^{n-1}\mathbf{E}^2[\hat{R}(k)]-\sum_{k=-\infty}^{\infty}R^2(k)\sim const.n^{-1}$.
By (9) and again (B1) we have:

$$\sum_{k=-n+1}^{n-1} Var(\hat{R}(k)) \sim n^{-2}\sum_{k=-n+1}^{n-1}\sum_{t_1,t_2=1}^{n-|k|}R^2(t_1-t_2)$$

$$\sim\quad n^{-2}\sum_{k=n+1}^{n-1}\sum_{u=-n+|k|+1}^{n-|k|-1}R^2(u)(n-|k|-|u|) = n^{-2}\sum_{u=n+1}^{n-1}R^2(u)\sum_{k=-n+|u|+1}^{n-|u|-1}(n-|k|-|u|)$$

$$=\quad n^{-2}\sum_{u=-n+1}^{n-1}R^2(u)(n-|u|)^2\sim\sum_{u=-\infty}^{\infty}R^2(u).$$

Assertion (i) follows now by using the identity $\mathbf{E}[\hat{R}^2(k)] = Var(\hat{R}(k))+\mathbf{E}^2[\hat{R}(k)]$.
(ii) By using well known results about cumulants of a two way array of random variables (cf. Rosenblatt (1985), Theorem 2, Chapter II):

$$\begin{aligned}
&Cov(\hat{R}^2(k_1),\hat{R}^2(k_2)) = \\
&4\mathbf{E}[\hat{R}(k_1)]\mathbf{E}[\hat{R}(k_2)]Cov(\hat{R}(k_1),\hat{R}(k_2)) + 2Cov^2(\hat{R}(k_1),\hat{R}(k_2)) + \\
&2\mathbf{E}[\hat{R}(k_1)]cum_3(\hat{R}(k_1),\hat{R}(k_2),\hat{R}(k_2)) + 2\mathbf{E}[\hat{R}(k_2)]cum_3(\hat{R}(k_1),\hat{R}(k_1),\hat{R}(k_2)) + \\
&cum_4(\hat{R}(k_1),\hat{R}(k_1),\hat{R}(k_2),\hat{R}(k_2)).
\end{aligned} \tag{10}$$

To analyze the third- and fourth-order cumulants we use again the combinatorical results about cumulants (cf. Rosenblatt (1985), Theorem 2, Chapter II):

$$\begin{aligned}
&cum_3(\hat{R}(k_1),\hat{R}(k_2),\hat{R}(k_3)) = \\
&n^{-3}\sum_{t_1=1}^{n-|k_1|}\sum_{t_2=1}^{n-|k_2|}\sum_{t_3=1}^{n-|k_3|}\sum_{\nu}cum(X_{i_j};i_j\in\nu_1)\dots cum(X_{i_j};i_j\in\nu_p), \\
&cum_4(\hat{R}(k_1),\hat{R}(k_2),\hat{R}(k_3),\hat{R}(k_4)) = \\
&n^{-4}\sum_{t_1=1}^{n-|k_1|}\sum_{t_2=1}^{n-|k_2|}\sum_{t_3=1}^{n-|k_3|}\sum_{t_4=1}^{n-|k_4|}\sum_{\nu}cum(X_{i_j};i_j\in\nu_1)\dots cum(X_{i_j};i_j\in\nu_p),
\end{aligned} \tag{11}$$

where the summation $\sum_\nu$ is over all indecomposable partitions $\nu_1\cup\dots\cup\nu_p$ of the corresponding two-way table, i.e.,

$$\begin{array}{ll} X_{t_1} & X_{t_1+|k_1|} \\ X_{t_2} & X_{t_2+|k_2|} \\ X_{t_3} & X_{t_3+|k_3|} \end{array} \qquad\qquad \begin{array}{ll} X_{t_1} & X_{t_1+|k_1|} \\ X_{t_2} & X_{t_2+|k_2|} \\ X_{t_3} & X_{t_3+|k_3|} \\ X_{t_4} & X_{t_4+|k_4|}. \end{array}$$

By (8) - (11) and assumption (B1) assertion (ii) follows.
(iii) By a Taylor expansion we obtain:

$$\begin{aligned}\bar{w}(kb) &= \bar{w}(0) + 1/2\bar{w}''(\xi)(kb)^2 \\ &= 1 + 1/2\bar{w}''(\xi)(kb)^2,\ 0 < \xi < kb.\end{aligned}$$

Now assertion (iii) follows by (8), (9) and (B1).
(iv) By (8)-(11) and (B1) we get by a straightforward calculation:

$$\begin{aligned}\text{left hand side of (iv)} &\sim (const.n^{-1} + const.n^{-2}\sum_{k=-\infty}^{\infty} \bar{w}^4(kb)k^{4r})(1+o(1)) \\ &\sim const.b^{-4r-1}n^{-2} + const.n^{-1}\end{aligned}$$

(the $const.n^{-1}$-term stems from the first term on the right-hand side of (10), the next term from the second term on the right-hand side of (10), the third and fourth term on the right-hand side of (10) and the fourth-order cumulant term in (10) are of negligible order).
(v) More generally than (10) we have:

$$\begin{aligned}&Cov(\hat{R}(k_1)\hat{R}(k_2), \hat{R}(k_3)\hat{R}(k_4)) = \\ &\sum_{\mathcal{P}_1}\mathbf{E}[\hat{R}(k_{i_1})]\mathbf{E}[\hat{R}(k_{i_2})]Cov(\hat{R}(k_{i_3}), \hat{R}(k_{i_4})) + \\ &Cov(\hat{R}(k_1), \hat{R}(k_3))Cov(\hat{R}(k_2), \hat{R}(k_4)) + Cov(\hat{R}(k_1), \hat{R}(k_4))Cov(\hat{R}(k_2), \hat{R}(k_3)) + \\ &\sum_{\mathcal{P}_2}\mathbf{E}[\hat{R}(k_{i_1})]cum_3(\hat{R}(k_{i_2}), \hat{R}(k_{i_3}), \hat{R}(k_{i_4})) + cum_4(\hat{R}(k_1), \hat{R}(k_2), \hat{R}(k_3), \hat{R}(k_4)),\end{aligned}$$

where $\mathcal{P}_1 = \{$indecomposable partitions $i_1 \cup i_2 \cup \{i_3, i_4\}$ of $\begin{matrix}1 & 2\\ 3 & 4\end{matrix}$ $\}$,

$\mathcal{P}_2 = \{$indecomposable partitions $i_1 \cup \{i_2, i_3, i_4\}$ of $\begin{matrix}1 & 2\\ 3 & 4\end{matrix}$ $\}$.
By a straightforward calculation using (8), (9), (11) and (B1) we get:

$$\begin{aligned}\text{left hand side of (v)} &\sim \{4(f^{(r)}(\lambda))^2\iota(\lambda)n^{-1}b^{-2r}\sum_{k=-\infty}^{\infty}\tilde{w}^2(kb)|kb|^{2r}(f(\lambda))^2 \\ &+ 2(\iota(\lambda)n^{-1}b^{-2r}\sum_{k=-\infty}^{\infty}\tilde{w}^2(kb)|kb|^{2r}(f(\lambda))^2)^2\}(1+o(1)) \\ &\sim const.b^{-2r-1}n^{-1}\end{aligned}$$

(the first term stems from the terms in $\sum_{\mathcal{P}_1}$, the second from the product covariance terms, the terms in $\sum_{\mathcal{P}_2}$ and the fourth-order cumulant term are of negligible order; the assumption $b^{-3}n^{-1} = o(1)$ allows us to use some rough bounds for the negligible terms, we do not strive for maximal generality and use this additional assumption which actually yields no restriction for our procedures). □

Proof of Lemma 2:
Denote by $D(b) = \int_{-\pi}^{\pi}(\hat{f}_{\bar{w}}^{(1)}(\lambda; b))^2 d\lambda,\ D^{\dagger}(b) = c(n, b_0)D(b)$ with $c(n, b_0) = a(n, b_0)^{-1/3}$.

Then $c(n,b_0)$ is a normalizing constant for $D(b)$ ($b \sim const.b_0$), i.e., $D^\dagger(b) = O_P(1)$ and $(D^\dagger(b))^{-1} = O_P(1)$ for $b \sim const.b_0$. We will show stochastic equicontinuity of $D^\dagger(b)$, i.e., $\forall \varepsilon > 0\ \forall \eta > 0\ \exists \delta\ (0 < \delta < 1)$ and $\exists n_0 \in \mathbf{N}$ such that

$$\mathbf{P}[\sup_{b \in U_{b_0}(\delta)} |D^\dagger(b) - D^\dagger(b_0)| > \varepsilon] < \eta. \tag{12}$$

To prove (12) we replace the supremum by a discrete maximum plus an error term, the idea of this approach is well known (cf. Billingsley (1968), Theorem 12.2 and Theorem 22.1). The discrete neighborhoods of b_0 are defined by $V_{upp} = \{b_0(1+\delta i/K_n; i = 0, \ldots K_n\}$, $V_{low} = \{b_0(1 - \delta i/K_n); i = 0, \ldots, K_n\}$l, where $K_n = [b_0^{-4}c(n,b_0)] + 1$ and $[x]$ denotes the greatest integer $\leq x$. Then:

$$\sup_{b \in U_{b_0}(\delta)} |D^\dagger(b) - D^\dagger(b_0)| \leq \max_{b \in V_{low} \cup V_{upp}} |D^\dagger(b) - D^\dagger(b_0)| + \sup_{b \in U_{b_0}(\delta)} |D^\dagger(b) - D^\dagger(z_b)|, \tag{13}$$

where $z_b \in V_{upp} \cup V_{low}$ is the closest point of $b \in U_{b_0}(\delta)$.
Let us now analyze the difference $D^\dagger(b_1) - D^\dagger(b_2)$ for $b_1 \sim const.b_0$, $b_2 \sim const.b_0$. We have:

$$D^\dagger(b_1) - D^\dagger(b_2) = c(n,b_0) \sum_{k=-\infty}^{\infty} (\bar{w}^2(kb_1) - \bar{w}^2(kb_2))k^2\hat{R}^2(k), \tag{14}$$

where by the Lipschitz continuity of the window $|\bar{w}^2(kb_1) - \bar{w}^2(kb_2)| \leq const.b_0^{-1}|b_1 - b_2|$. From (14) we obtain by the Cauchy-Schwarz inequality the estimate:

$$|D^\dagger(b) - D^\dagger(z_b)| \leq const.c(n,b_0)b_0^{-4}|b - z_b|\hat{R}^2(0).$$

Since $\sup_{b \in U_{b_0}(\delta)} |b - z_b| \leq \delta/K_n \leq b_0^4(c(n,b_0))^{-1}\delta$ we arrive at:
$\forall \varepsilon > 0\ \forall \eta > 0\ \exists \delta\ (0 < \delta < 1)$ and $\exists n_0 \in \mathbf{N}$ such that

$$\mathbf{P}[\sup_{b \in U_{b_0}(\delta)} |D^\dagger(b) - D^\dagger(z_b)| > \varepsilon] < \eta. \tag{15}$$

On the other hand:

$$\max_{b \in V_{upp} \cup V_{low}} |D^\dagger(b) - D^\dagger(b_0)| \leq \max_{b \in V_{upp}} |D^\dagger(b) - D^\dagger(b_0)| + \max_{b \in V_{low}} |D^\dagger(b) - D^\dagger(b_0)|. \tag{16}$$

From (14), the Lipschitz continuity of $\bar{w}$, (8), (9) and (B1) we obtain:

$$|\mathbf{E}[D^\dagger(b_1) - D^\dagger(b_2)]| \leq const.b_0^{-1}|b_1 - b_2|\ (b_1 \sim const.b_0,\ b_2 \sim const.b_0). \tag{17}$$

From (14), the Lipschitz continuity of $\bar{w}$, (8)-(11) and (B1) we obtain (compare with the proof of Lemma 11 (iv)):

$$Var(D^\dagger(b_1) - D^\dagger(b_2)) \leq const.b_0^{-2}|b_1 - b_2|^2\ (b_1 \sim const.b_0,\ b_2 \sim const.b_0). \tag{18}$$

Let us consider $\max_{b \in V_{upp}} |D^\dagger(b) - D^\dagger(b_0)|$. Denote by $v_i = b_0(1 + \delta i/K_n)\ i = 0, \ldots, K_n$ and $S_j = \sum_{r=1}^{j}(D^\dagger(v_r) - D^\dagger(v_{r-1})) = D^\dagger(v_j) - D^\dagger(b_0)$. Then for $j > i$: $S_j - S_i =$

$D^\dagger(v_j) - D^\dagger(v_i)$, $v_j - v_i = b_0\delta(j-i)/K_n$. By (17) and (18) we get for $j > i$:
$\forall\varepsilon > 0\ \forall\eta > 0\ \exists\delta\ (0 < \delta < 1)$ and $\exists n_0 \in \mathbf{N}$ such that

$$\mathbf{P}[|S_j - S_i| > \varepsilon] \le \varepsilon^{-2}C\delta^2|j-i|^2/K_n^2 = (\sum_{i<r\le j} u_r)^2,\ u_r = \varepsilon^{-1}\sqrt{C}\delta/K_n\ (C \text{ a constant}).$$

Hence by Theorem 12.2 from Billingsley (1968): $\forall\varepsilon > 0\ \forall\eta > 0\ \exists\delta\ (0 < \delta < 1)$ and $\exists n_0 \in \mathbf{N}$ such that

$$\mathbf{P}[\max_{b\in V_{upp}} |D^\dagger(b) - D^\dagger(b_0)| > \varepsilon] = \mathbf{P}[\max_{0<j\le K_n} |S_j| > \varepsilon] \le \varepsilon^{-2}C\delta^2 < \eta, \tag{19}$$

by choosing δ sufficiently small. Analogously for $\mathbf{P}[\max_{b\in V_{upp}} |D^\dagger(b) - D^\dagger(b_0)| > \varepsilon]$.
By (13), (15) and (19) we have shown (12).

By using a Taylor expansion we have:

$$a(n,b_0)(G_B(b) - G_B(b_0)) =$$
$$n^{-1/3}[1/6\int_{-\pi}^{\pi}((2\pi)^{-1}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-i\lambda k})^2 d\lambda]^{1/3}(-1/3)[c(n,b_0)I(n)]^{-4/3}(D^\dagger(b) - D^\dagger(b_0)),$$

where $c(n,b_0)I(n) = D^\dagger(b_0) + \tau(D^\dagger(b) - D^\dagger(b_0))$ $(0 < \tau < 1)$. By (12) we have: $c(n,b_0)I(n) = D^\dagger(b_0) + o_P(1)$ uniformly for $b \in U_{b_0}(\delta)$, $D^\dagger(b) - D^\dagger(b_0) = o_P(1)$ uniformly for $b \in U_{b_0}(\delta)$. We complete the proof by Lemma 1 (i). □

Proof of Lemma 5:
The proof follows closely the lines of the proof of Lemma 2. Denote by $Q(\lambda;b) = (\hat{f}_{\tilde{w}}(\lambda;b))^2$, $S(\lambda;b) = (\hat{f}_{\bar{w}}^{(1)}(\lambda;b))^2$. By Lemma 4 (i), (ii) and since $b_0 \sim const.n^{-1/3+\zeta}$, $(0 < \zeta < 1/3)$ we don't need any normalizing constants for $Q(\lambda;b)$ and $S(\lambda;b)$. We show: $\forall\ \varepsilon > 0\ \forall\eta > 0\ \exists\ \delta\ (0 < \delta < 1)$ and $\exists n_0 \in \mathbf{N}$ such that

$$\mathbf{P}[\sup_{b\in U_{b_0}(\delta)} d_n|Q(\lambda;b) - Q(\lambda;b_0)| > \varepsilon] < \eta, \tag{20}$$

$$\mathbf{P}[\sup_{b\in U_{b_0}(\delta)} d_n|S(\lambda;b) - S(\lambda;b_0)| > \varepsilon] < \eta. \tag{21}$$

To show (20) and (21) we proceed in the same way as in the proof of Lemma 2. Instead of (14) we have:

$$\begin{aligned} &Q(\lambda;b_1) - Q(\lambda;b_2) \\ &= \sum_{k_1,k_2=-\infty}^{\infty}(\tilde{w}(k_1b_1)\tilde{w}(k_2b_1) - \tilde{w}(k_1b_2)\tilde{w}(k_2b_2))e^{-i\lambda(k_1+k_2)}\hat{R}(k_1)\hat{R}(k_2), \qquad (22)\\ &S(\lambda;b_1) - S(\lambda;b_2) \\ &= \sum_{k_1,k_2=-\infty}^{\infty}(\bar{w}(k_1b_1)\bar{w}(k_2b_1) - \bar{w}(k_1b_2)\bar{w}(k_2b_2))|k_1||k_2|e^{-i\lambda(k_1+k_2)}\hat{R}(k_1)\hat{R}(k_2). \qquad (23)\end{aligned}$$

We write

$$\begin{aligned} &\tilde{w}(k_1b_1)\tilde{w}(k_2b_1) - \tilde{w}(k_1b_2)\tilde{w}(k_2b_2) \\ &= (\tilde{w}(k_1b_1) - \tilde{w}(k_1b_2))\tilde{w}(k_2b_1) + (\tilde{w}(k_2b_1) - \tilde{w}(k_2b_2))\tilde{w}(k_1b_2) \\ &\text{(the same holds for } \bar{w}). \qquad (24)\end{aligned}$$

By the Lipschitz continuity of $\bar{w}$ and $\tilde{w}$ this leads to

$$d_n|Q(\lambda;b) - Q(\lambda;z_b)| \leq const.d_n b_0^{-3}|b - z_b|\hat{R}^2(0),$$
$$d_n|S(\lambda;b) - S(\lambda;z_b)| \leq const.d_n b_0^{-5}|b - z_b|\hat{R}^2(0).$$

This suggests to take $K_n^Q = [d_n b_0^{-3}]+1$, $K_n^S = [d_n b_0^{-5}]+1$ for the neighborhoods V_{low}^Q, V_{upp}^Q, V_{low}^S, V_{upp}^S (the capital letters Q, S indicate the correspondence). For the analogue of (17) we use (22)-(24) and $d_n \leq const.min\{b_0^{-2}, b_0^3 n^1\}$. Then we get in a straightforward way by using (8), (9), (B1) and the Lipschitz continuity of $\bar{w}$ and $\tilde{w}$:

$$|\mathbb{E}[d_n(Q(\lambda;b_1) - Q(\lambda;b_2))]| \leq const.b_0^{-1}|b_1 - b_2|,$$
$$|\mathbb{E}[d_n(S(\lambda;b_1) - S(\lambda;b_2))]| \leq const.b_0^{-1}|b_1 - b_2|.$$

Instead of (18) we get from (22)-(24) and analogously to the proof of Lemma 11 (v):

$$Var(d_n(Q(\lambda;b_1) - Q(\lambda;b_2))) \leq const.b_0^{-2}|b_1 - b_2|^2,$$
$$Var(d_n(S(\lambda;b_1) - S(\lambda;b_2))) \leq const.b_0^{-2}|b_1 - b_2|^2,$$

(these bounds are sufficient).
Then (20) and (21) follow by the same arguments as in the proof of Lemma 2. Finally we complete the proof by a Taylor expansion (compare with the proof of Lemma 2). □

Acknowledgments. I thank H. Künsch for useful suggestions and clarifying remarks. I appreciate also interesting discussions with Th. Gasser. Further thanks are going to D. Brillinger for some stimulating comments.

References

[1] Billingsley,P. (1968). Convergence of Probability Measures. Wiley, New York.

[2] Beltrão,K.I. and Bloomfield,P. (1987). Determining the bandwidth of a kernel spectrum estimate. J. Time Ser. Anal. **8** 21-38.

[3] Beltrão,K.I. and Hurvich,C.M (1990). Cross-validatory choice of a spectrum estimate and its connection with AIC. J. Time Ser. Anal. **11** 121-137.

[4] Brillinger,D.R. (1994). Some asymptotics of wavelet fits in the stationary error case. Tech. Rep. 415, Statistics Dept., Univ. of Calif. Berkeley.

[5] Brockmann,M., Gasser,Th. and Herrmann,E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. J. Amer. Statist. Assoc. **88** 1302-1309.

[6] Bühlmann,P. (1994). Blockwise bootstrapped empirical process for stationary sequences. To appear in Ann. Statist. **22**

[7] Bühlmann,P. and Künsch,H.R. (1994). Block length selection in the bootstrap for time series. Res.Report 72, Seminar für Statistik, ETH Zürich.

[8] Chiu, S.T. (1990). Why bandwidth selectors tend to choose smaller bandwidths, and a remedy. Biometrika **77** 222-226.

[9] Künsch,H.R. (1989). The jackknife and the bootstrap for general stationary observations. Ann.Statist. **17** 1217-1241.

[10] Müller,H.G. and Gasser,Th. (1986). Non parametric estimation of spectral densities and their derivatives: integrated mean square error, choice of window width and applications. Cahiers du C.E.R.O. **28** 163-173.

[11] Park,B.U. and Cho,S. (1991). Estimation of integrated squared spectral density derivatives. Statist. Probab. Lett. **12** 65-72.

[12] Parzen,E. (1957). On consistent estimates of the spectrum of a stationary time series. Ann.Math.Statist. **28** 329-348.

[13] Parzen,E. (1958). On asymptotically efficient consistent estimates of the spectral density function of a stationary time series. J.Roy.Statist.Soc., B. **20** 303-322.

[14] Percival,D.B. and Walden,A.T. (1993). Spectral Analysis for Physical Applications. University Press, Cambridge.

[15] Politis,D.N. and Romano,J.P. (1992). A general resampling scheme for triangular arrays of α-mixing random variables. Ann.Statist. **20** 1985-2007.

[16] Priestley,M.B. (1981). Spectral Analysis and Time Series **1**. Academic, New York.

[17] Rosenblatt,M. (1985). Stationary Sequences and Random Fields. Birkhäuser, Boston.

Department of Statistics
Evans Hall
University of California
Berkeley, CA 94720
USA