Spectral Density Estimation via Nonlinear Wavelet Methods for Stationary non-Gaussian Time Series

Michael H. Neumann
Weierstrass Institute for Applied Analysis and Stochastics
Mohrenstr. 39
D-10117 Berlin, Germany

Abstract. In the present paper we consider nonlinear wavelet estimators of the spectral density f of a zero mean, not necessarily Gaussian stochastic process, which is stationary in the wide sense. It is known in the case of Gaussian regression that these estimators outperform traditional linear methods if the degree of smoothness of the regression function considerably varies over the interval of interest. Such methods are based on a nonlinear treatment of empirical coefficients that arise from an orthonormal series expansion according to a wavelet basis.
The main goal of this paper is to transfer these methods to spectral density estimation. This is done showing asymptotic normality of certain empirical coefficients based on the tapered periodogram. Using these results we can conclude the risk equivalence to the Gaussian case for monotone estimators based on such empirical coefficients. The resulting estimator of f keeps all interesting properties like high spatial adaptivity that are already known for wavelet estimators in the case of Gaussian regression.
It turns out that optimally tuned versions of this estimator attain the optimal uniform rate of convergence of their L_2-risk in a wide variety of Besov smoothness classes, including such classes, where linear estimators (Kernel, Spline) are not able to attain this rate. Some simulations indicate the usefulness of the new method in cases of high spatial inhomogeneity.

1991 Mathematics Subject Classification. Primary 62M15; Secondary 62M10, 62G07.
Key words and phrases. Spectral density estimation, wavelet estimators, nonlinear wavelet shrinkage, asymptotic normality, large deviations.
Short title. Wavelet spectral density estimation.

1. Introduction

In the present paper we consider the problem of estimating a spectral density of a real, zero mean, possibly non-Gaussian time series, which is stationary in the wide sense. An important goal in spectral analysis is the recognition of peaks of the spectral density, which are an indication for periodicities of the underlying stochastic process. Whereas spatially homogeneous smoothing methods (e.g. kernel estimators with global bandwidth, spline estimators) are an appropriate tool for estimating functions with a homogeneous degree of smoothness, one should prefer methods with a locally varying degree of smoothing in estimating objects that are quite smooth on one part of the domain but much less regular on another part.
Such a problem has been investigated by Donoho and Johnstone (1992), who derived minimax results in Besov classes in the Gaussian white noise model

$$dY_t \;=\; f(t)\, dt \;+\; \epsilon\, dW_t, \quad t \in [0,1], \tag{1.1}$$

where W_t is a Brownian motion and ϵ the noise level. Donoho and Johnstone rediscovered the phenomenon, originally detected by Nemirovskii, Polyak and Tsybakov (1985) and Nemirovskii (1985) for Sobolev smoothness classes, that linear estimators are unable to attain the optimal uniform convergence rate in balls of Besov spaces $B^m_{p,q}$ with $p < 2$. Moreover, they showed that thresholded wavelet estimators attain the minimax bound up to a small constant.
These estimators are based on an orthonormal system $\{\widetilde{\phi}_{l,k}\}_{k\in I_l^0} \cup \{\widetilde{\psi}_{j,k}\}_{j\geq l, k\in I_j}$ of basis functions, which are essentially generated by dilations and translations of two special functions ϕ and ψ. Having the orthogonal series expansion $f = \sum \alpha_k \widetilde{\phi}_{l,k} + \sum \alpha_{j,k} \widetilde{\psi}_{j,k}$ in mind, one determines first empirical versions $\widetilde{\alpha}_k$ and $\widetilde{\alpha}_{j,k}$ of the coefficients. These empirical coefficients are again normally distributed with homogeneous variances ϵ^2. Then one applies level-wise nonlinear shrinkage rules δ_j to these coefficients, which finally yields an estimator

$$\widehat{f} \;=\; \sum_k \widetilde{\alpha}_k \widetilde{\phi}_{l,k} \;+\; \sum_{j,k} \delta_j(\widetilde{\alpha}_{j,k}) \widetilde{\psi}_{j,k}$$

of the regression function f.
Extracting the essentials from various papers by Donoho, Johnstone, Kerkyacharian and Picard about estimators in model (1.1), we observe two main requirements for estimators based on any orthogonal series expansion of f to have high spatial adaptivity: a sufficiently localized basis and a sufficiently strong discrimination between noise and variance.
To elucidate the first point, consider estimators which are based on univariate functions of unbiased empirical coefficients w.r.t. any orthonormal basis. As usual in nonparametric regression, to tune such an estimator optimally, we have to look for a good compromise between variance that is roughly proportional to the number of coefficients included in the estimator, and bias caused by the exclusion of nonzero coefficients. Obviously, the more the signal is compressed, i.e. contained in a small number of coefficients, the better is the performance of an optimally tuned estimator.

In Donoho (1993) it is shown that unconditional bases in the space under consideration are optimal in this context. Whereas the sine/cosine Fourier basis is such an unconditional basis of L_2-Sobolev spaces, wavelet bases are still unconditional in L_2-Sobolev, but unlike the sine/cosine basis also unconditional in a wide variety of Besov spaces that contain functions with substantial inhomogeneous smoothness properties. Let us further explain the need of nonlinear estimation rules. If we have some function with homogeneous smoothness over the whole domain (and the L_2-norm stronger suppresses inhomogeneity than an L_p-norm with $p < 2$), then we have a certain homogeneity in the size of the wavelet coefficients at each resolution scale. Hence, we do not lose very much by choosing simply between the inclusion or exclusion of each particular level. This heuristics leads us to *linear* estimation rules. In contrast, under strong inhomogeneity, the coefficients within a fixed resolution level are of considerably different order of magnitude. Moreover, a small number of coefficients from quite high levels represent that much signal, that their empirical counterpart must be included in an efficient estimator. It is known for such cases of sparse signals, that nonlinear estimators dramatically outperform linear methods, cf. Donoho and Johnstone (1992).
We intend to transfer this locally adaptive estimation technique to spectral density estimation. First we define an appropriate wavelet basis for $\widetilde{L}_2(\Pi)$, the collection of 2π-periodic functions of $L_2(\Pi)$, $\Pi = [-\pi, \pi]$. On the basis of a, possibly tapered, periodogram we obtain empirical versions of the coefficients of f, which are then treated with the same shrinkage methods as known from Donoho and Johnstone (1992) in the case of Gaussian regression.
Our aim is to show that the resulting estimator of the spectral density behaves as good as a corresponding estimator in the Gaussian white noise model. For that we derive uniform estimates of the cumulants of the empirical wavelet coefficients, which are the basis for showing the equivalence to the Gaussian case in terms of probabilities of large deviations. These strong results allow us to conclude the risk equivalence between all monotone estimators based on the empirical coefficients and the same estimators with Gaussian random variables inserted. Therefore, thresholded wavelet estimators keep all their appealing properties known from the Gaussian case. It is shown that optimally thresholded wavelet estimators attain the minimax rate of convergence in a large scale of Besov smoothness classes. Finally, we discuss a possibility to adapt the smoothing parameters involved in the procedure.
The paper is organized as follows. In Section 2 we introduce the basic notions, as the tapered periodogram, an appropriate periodic wavelet basis and empirical versions of the wavelet coefficients. The results in the present article are mainly based on asymptotic normality of these empirical wavelet coefficients. In Section 3 we derive a quite general lemma on the cumulants of quadratic forms, which is of independent interest as well. This will be the basis for the derivation of asymptotic normality of the empirical wavelet coefficients in Section 4. In Section 5 we introduce two types of commonly used nonlinear wavelet estimators and show, using the large deviations results, the risk equivalence of these estimators to their counterpart in an idealized Gaussian model. This gives us minimax results for these estimators in a wide range of Besov classes. In particular, we describe some cases of practical interest, where our estimators outperform traditional linear methods, as e.g. kernel or spline estimators. In the next section we discuss a universal method for a data-driven choice of the

thresholds, and show its near-optimality. Section 7 contains results of a simulations study, which provides a comparison between our methods with a classical kernel estimator. Some final remarks are contained in Section 8 and the proofs are deferred to Section 9.

2. Basic tools

We assume that we observe a stretch $\{X_1, \dots, X_T\}$ of a real zero mean process $\{X_t\}_{t=1,\dots,\infty}$, which is stationary in the wide sense. We intend to estimate the spectral density

$$f(\omega) \;=\; \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} c_k \exp(i\omega k) \;=\; \frac{1}{2\pi}c_0 \;+\; \frac{1}{\pi}\sum_{k=1}^{\infty} c_k \cos(\omega k) \qquad (2.1)$$

on the interval $\Pi = [-\pi, \pi]$, where $c_k = cov(X_t, X_{t+k})$.
The basis for our estimates will be a tapered periodogram

$$I_T(\omega) \;=\; \frac{1}{2\pi H_2^{(T)}}\left|\sum_{t=1}^{T} h_t X_t e^{it\omega}\right|^2 \;=\; \frac{1}{2\pi H_2^{(T)}}\sum_{s,t=1}^{T} h_s h_t X_s X_t \cos(\omega(t-s)), \quad (2.2)$$

where $H_k^{(T)} = \sum_{t=1}^{T} h_t^k$, $h_t = h(t/T)$. (We put $I_T(\omega) \equiv 0$ if $H_2^{(T)} = 0$.)
For the data taper we assume

(A1) $\quad H \;=\; \int_0^1 h^2(x)\,dx \;>\; 0$, h is of bounded variation.

In particular, we obtain then $H_2^{(T)} \sim TH$. It is shown in Dahlhaus (1983, 1990) that the use of a smooth data taper reduces the finite sample bias of the periodogram, which pays off in an improved performance of nonparametric estimators of the spectral density. Choosing $h \equiv 1$ we obtain in (2.2) the ordinary (nontapered) periodogram I_T^0 as a special case. For the sake of greater generality we consider throughout the paper the tapered one, keeping in mind that this includes the nontapered analog as well.
It is well-known that $I_T(\omega)$ is asymptotically unbiased for $f(\omega)$ under quite general assumptions, however it is not a consistent estimator of $f(\omega)$. On the other hand, for $\omega_1 \neq \omega_2$, $I_T(\omega_1)$ and $I_T(\omega_2)$ are asymptotically uncorrelated. Therefore, there is some hope that one can obtain via smoothing estimators that are consistent under certain smoothness conditions on f.
Since the object of interest is 2π-periodic, we do not need any boundary correction of the wavelet basis. We start with two special compactly supported functions ϕ and ψ with the property that the collection

$$\left\{\tilde{\phi}_{l,k}\right\}_{k\in\mathbb{Z}} \cup \left\{\tilde{\psi}_{j,k}\right\}_{j\geq l, k\in\mathbb{Z}}$$

with $\tilde{\phi}_{l,k}(x) = 2^{l/2}\phi(2^l x - k)$ and $\tilde{\psi}_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ forms an orthonormal basis of $L_2(\mathbb{R})$. It is easy to see that

$$\{\phi_{l,k}\}_{k\in I_l} \cup \{\psi_{j,k}\}_{j\geq l, k\in I_j}$$

with $I_j = \{1, \dots, 2^j\}$,

$$\phi_{l,k}(x) \;=\; \sum_{n\in\mathbb{Z}} (2\pi)^{-1/2} \tilde{\phi}_{l,k}\left((2\pi)^{-1}x + n\right)$$

and

$$\psi_{j,k}(x) \;=\; \sum_{n\in\mathbb{Z}} (2\pi)^{-1/2} \tilde{\psi}_{j,k}\left((2\pi)^{-1}x + n\right)$$

is an orthonormal basis of $\widetilde{L}_2(\Pi)$, the L_2-space of 2π-periodic functions on Π, cf. Daubechies (1992, Chapter 9.3). (The above sums are finite, since ϕ and ψ are compactly supported.)
For $f \in \widetilde{L}_2(\Pi)$ we have the representation

$$f \;=\; \sum_{k\in I_l} \alpha_k \phi_{l,k} \;+\; \sum_{j\geq l} \sum_{k\in I_j} \alpha_{j,k} \psi_{j,k},$$

where $\alpha_k = \int f(t)\phi_{l,k}(t)\,dt$ and $\alpha_{j,k} = \int f(t)\psi_{j,k}(t)\,dt$ are the usual Fourier coefficients, which are also called wavelet coefficients.
Now we are in position to define empirical versions of these coefficients, which will be the starting point for our estimation procedures. First, we define an integral version as

$$\tilde{\alpha}_{j,k} \;=\; \int \psi_{j,k}(\omega) I_T(\omega)\,d\omega. \tag{2.3}$$

($\tilde{\alpha}_k$ is defined analogously.)
It is also possible to define discrete versions, e.g.

$$\overline{\alpha}_{j,k} \;=\; \sum_{i=1}^{T} \int_{\omega_i - 2\pi/T}^{\omega_i} \psi_{j,k}(\omega)\,d\omega\;I_T(\omega_i) \tag{2.4}$$

and

$$\overline{\overline{\alpha}}_{j,k} \;=\; \frac{2\pi}{T} \sum_{i=1}^{T} \psi_{j,k}(\omega_i)\,I_T(\omega_i), \tag{2.5}$$

where $\omega_i = 2\pi i/T - \pi$, $i = 1, \dots, T$. ($\overline{\alpha}_k$ and $\overline{\overline{\alpha}}_k$ are defined in the same way.)
One advantage of (2.4) and (2.5) in the nontapered case $h \equiv 1$ is undoubtly that, because of $\sum_{t=1}^{T} \exp\{it\omega_j\} = 0$ for $\omega_j = 2\pi j/T - \pi$, $0 < j < T$, the equation

$$I_T^0(\omega_j) \;=\; \frac{1}{2\pi T} \left| \sum_{t=1}^{T} X_t e^{it\omega_j} \right|^2 \;=\; \frac{1}{2\pi T} \left| \sum_{t=1}^{T} \left(X_t \;-\; T^{-1} \sum X_s\right) e^{it\omega_j} \right|^2, \tag{2.6}$$

holds. This means that the discrete version is robust against a deviation from our assumption $E\,X_t = 0$, since any nonzero mean of the X_t's is cancelled out. Moreover, it allows the computationally advantageous Fast Fourier Transform, see e.g. Brillinger (1975). On the other hand, we think that $\tilde{\alpha}_{j,k}$ causes perhaps less bias problems. In the following we restrict our considerations to estimators based on $\tilde{\alpha}_{j,k}$. It was shown in an earlier version of this paper, Neumann (1994), that the cumulants of $\overline{\alpha}_{j,k}$ and $\overline{\overline{\alpha}}_{j,k}$ are very close to those of $\tilde{\alpha}_{j,k}$, which implies that the corresponding estimators have the same asymptotic behavior.

3. Cumulants of the empirical wavelet coefficients

In this section we develop approximations to the mean and the variance of $\widetilde{\alpha}_{j,k}$, and we derive upper estimates for the higher order cumulants. These results are the basis for proving asymptotic normality of the empirical wavelet coefficients in the next section. First, we state a general lemma on the cumulants of quadratic forms, which is of interest in its own right. It generalizes a lemma of Rudzkis (1978), which was formulated specifically for quadratic forms that occur in periodogram-based kernel estimators of a spectral density. Our lemma does not require stationarity of the time series and it provides, moreover, a slightly improved estimate, which can turn out to be important for certain quadratic forms with matrices that do not have Toeplitz structure. Note that the rates for the decay of the cumulants could already be conjectured from the proof of Theorem 7.4.4 in Brillinger (1975, pp. 436-437). Here we derive additionally explicit *constants*, which are necessary to state asymptotic normality in a uniform manner for an increasing number of coefficients. In the sequel we impose the following assumption.

(A2) $\sup_{1\le t_1<\infty}\left\{\sum_{t_2,\dots,t_k=1}^{\infty}|cum(X_{t_1},\dots,X_{t_k})|\right\} \le C_1^k(k!)^{1+\gamma}$ for all $k=2,3,\dots,$ where $\gamma\ge 0$.

Let in the following C denote any positive constant, which may take different values at each occurence.

Remark 3.1. (i) Let $\{X_t\}$ be α-mixing with coefficients $\alpha(s)\le K\exp(-b|s|)$ and $E|X_t|^k\le C^k(k!)^{\gamma}$ for all k. It was shown in Neumann (1994) that

$$\sup_{1\le t_1<\infty}\left\{\sum_{t_2,\dots,t_k=1}^{\infty}|cum(X_{t_1},\dots,X_{t_k})|\right\} \le C^k(k!)^{3+\gamma} \quad \text{for all } k=2,3,\dots$$

(ii) Note that, besides the obvious case of the normal distribution, many of the distributions that can be found in textbooks satisfy (A2) for an appropriate choice of γ. In Johnson and Kotz (1970) we can find closed forms of higher order cumulants of the exponential, gamma, inverse Gaussian and F-distribution, which show that this condition is satisfied for $\gamma=0$. The need for a positive γ occurs in the case of heavier-tailed distributions, which could arise as the distributions of sums of weakly dependent random variables.

Lemma 3.1. *Assume* $EX_t=0$ *, (A2) and let*

$$\eta \;=\; \underline{X}_T' A \underline{X}_T,$$

where

$$\underline{X}_T \;=\; (X_1,\dots,X_T)', \quad A \;=\; ((a_{st}))_{s,t=1,\dots,T}, \quad a_{st} \;=\; a_{ts}.$$

Further, let

$$\xi \;=\; \underline{Y}_T' A \underline{Y}_T,$$

where $\underline{Y}_T\sim N(0,var(\underline{X}_T))$ *. Then*

$$cum_n(\eta) \;=\; cum_n(\xi) \;+\; R_n,$$

where

$$\text{(i)}\quad |cum_n(\xi)| \,\le\, var(\xi)2^{n-2}(n-1)!\,[\lambda_{\max}\left(A\,Cov(\underline{X}_T)\right)]^{n-2}$$

$$\text{(ii)}\quad R_n \,\le\, 2^{n-2}C^{2n}((2n)!)^{1+\gamma}\max_{s,t}\{|a_{st}|\}\,\widetilde{A}\,\|A\|_\infty^{n-2},$$

$$\widetilde{A} \,=\, \sum_s \max_t\{|a_{st}|\}, \quad \|A\|_\infty \,=\, \max_s\left\{\sum_t |a_{st}|\right\}.$$

Under the following assumptions we can derive appropriate estimates for bias and variance of the empirical coefficients:

(A3) $TV(f) \,\le\, C_2, \quad \|f\|_\infty \le C_3$,
(A4) for any $r > m$ we assume that
  (i) ϕ and ψ are in C^r,
  (ii) $\int \phi(t)\,dt \,=\, 1$,
  (iii) $\int \psi(t)t^k\,dt \,=\, 0 \quad$ for $0 \le k \le r$.

Note that the regularity r of the scaling function ϕ and the wavelet ψ has tom be chosen higher than the assumed smoothness m of the spectral density in order to make the optimal rate of convergence of the estimator possible.

Proposition 3.1. *Assume (A1) through (A4). Then*

$$\text{(i)}\quad E\widetilde{\alpha}_{j,k} \,=\, \alpha_{j,k} \,+\, O\left(2^{j/2}T^{-1}\log T\right).$$

$$\text{(ii)}\quad var(\widetilde{\alpha}_{j,k}) \,=\, 2\pi\left(H_4^{(T)}/(H_2^{(T)})^2\right)\int_\Pi \psi_{j,k}(\alpha)\,[\psi_{j,k}(\alpha)+\psi_{j,k}(-\alpha)]\,|f(\alpha)|^2\,d\alpha$$
$$+\, o(T^{-1}) \,+\, O(T^{-1}2^{-j}),$$

$$\text{(iii)}\quad |cum_n(\widetilde{\alpha}_{j,k})| \,\le\, C^n(n!)^{2+2\gamma}T^{-1}\left(2^{j/2}T^{-1}\log(T)\right)^{n-2}$$

hold uniformly over $k \in I_j$ *and* $2^j \le CT^{1-\alpha}$ *, where* $C < \infty$ *and* $\alpha > 0$ *are arbitrary but fixed constants.*

4. Asymptotic normality of the empirical wavelet coefficients

In this section we intend to state asymptotic normality of the empirical coefficients $\widetilde{\alpha}_{j,k}$. It is known (cf. Theorem 5.2.7 in Brillinger (1975)) that a value $I_T(\omega)$ of the periodogram is asymptotically chi-squared distributed. Hence, we can only expect asymptotic normality of $\widetilde{\alpha}_{j,k}$, if a certain summation effect works. On the other hand, we have for any ball $\mathcal{F}$ in a Besov space $B^m_{p,q}$ that

$$\sup_{f\in\mathcal{F}}\left\{\sum_{j>J}\sum_{k\in I_j}\alpha_{jk}^2\right\} \,=\, O\left(2^{-2J(m+1/2-1/(p\wedge 2))}\right), \tag{4.1}$$

which is $O(T^{-2m/(2m+1)})$ if $2^{-J} = O(T^{-2/3})$, cf. Theorem 8 in Donoho *et al.* (1995). Therefore, we restrict our considerations in this section to coefficients with indices (j,k) from a set

$$\mathcal{J} \,=\, \mathcal{J}(T) \,=\, \left\{(j,k)\,\Big|\, 2^j \le CT^{1-\alpha},\ k \in I_j\right\},$$

where $C < \infty$ and $0 < \alpha \le 1/3$ are arbitrary, but fixed constants. Let $\sigma_{j,k}^2$ denote the variance of $\tilde{\alpha}_{j,k}$.
By (ii) of Proposition 3.1 we obtain that

$$\sup_{j,k}\{\sigma_{j,k}\} = O(T^{-1/2}). \tag{4.2}$$

Let, for some fixed $C_0 > 0$,

$$\mathcal{J}^0 = \mathcal{J}^0(T) = \{(j,k) \in \mathcal{J} \,\big|\, \sigma_{j,k} \ge C_0 T^{-1/2}\}.$$

By (iii) of Proposition 3.1 we obtain that

$$|cum_n(\tilde{\alpha}_{j,k}/\sigma_{j,k})| \le (n!)^{2+2\gamma} C^n \left(T^{-\alpha/2}\log(T)\right)^{n-2} \tag{4.3}$$

holds uniformly in $(j,k) \in \mathcal{J}^0$. This provides by Lemma 1 in Rudzkis, Saulis and Statulevicius (1978) the following assertion.

Theorem 4.1. *Assume (A1) through (A4). Then*

$$\frac{P\left(\pm(\tilde{\alpha}_{j,k} - \alpha_{j,k})/\sigma_{j,k} \ge x\right)}{1 - \Phi(x)} \longrightarrow 1$$

holds uniformly in $(j,k) \in \mathcal{J}^0$, $-\infty < x \le \Delta_\gamma$, *where* $\Delta_\gamma = o(\Delta^{1/(3+4\gamma)})$ *and* $\Delta = T^{\alpha/2}(\log T)^{-1}$.

Note that we have those empirical coefficients excluded from the assertion of Theorem 4.1, which have a standard deviation $\sigma_{j,k}$ below the level $C_0 T^{-1/2}$. Since our cumulant estimates in (iii) of Proposition 3.1 are essentially all of the same order of magnitude, it is not possible to derive the crucial relation (4.3) for these coefficients. However, we can repair this defect. Let

$$\sigma_T = \max\{\max_{(j,k)\in\mathcal{J}}\{\sigma_{j,k}\}, C_0 T^{-1/2}\}.$$

and let $\theta_{j,k} \sim N(0, \sigma_T^2 - \sigma_{j,k}^2)$ be independent of $\tilde{\alpha}_{j,k}$. Then the new random variable $\tilde{\alpha}_{j,k} + \theta_{j,k}$ has the same mean and the same n-th order cumulants for $n \ge 3$ as $\tilde{\alpha}_{j,k}$, whereas its variance is equal to $\sigma_T^2 \asymp T^{-1}$. Therefore, we can derive in complete analogy to Theorem 4.1 the following result.

Theorem 4.2. *Assume (A1) through (A4). Then*

$$\frac{P\left(\pm((\tilde{\alpha}_{j,k} + \theta_{j,k}) - \alpha_{j,k})/\sigma_T \ge x\right)}{1 - \Phi(x)} \longrightarrow 1$$

holds uniformly in $(j,k) \in \mathcal{J}$, $-\infty < x \le \Delta_\gamma$.

On the basis of these two theorems we will conclude in the next section that the risk of monotone estimators based on $\tilde{\alpha}_{j,k}$ is asymptotically the same as in the case of Gaussian empirical coefficients. This will provide the link to existing results for wavelet estimators in Gaussian models.

5. Wavelet estimators of the spectral density

5.1. Derivation of thresholding schemes. First, we recall some known facts about wavelet estimators in the white noise model. Assume observations from model (1.1). If we consider minimax estimation problems for $f \in \mathcal{F}$, it is known (for details see Donoho and Johnstone (1992)) that linear estimators are unable to attain the optimal convergence rate if $\mathcal{F}$ is some ball in a Besov space $B^m_{p,q}$ with $p < 2$. Moreover, certain nonlinear wavelet estimators attain this minimax rate. In other words, this result indicates that wavelet estimators can behave much better than traditional linear estimators whenever a certain amount of inhomogeneity in the smoothness of f is present. Since spectral densities have certain peaks in the case of the presence of some near-periodicities, the application of these methods seems to be an interesting attempt to estimate them simultaneously well in certain smooth regions as well as near peaks.
Donoho and Johnstone used in a series of papers the following nonlinear methods in the framework of model (1.1):

$$\delta^{(h)}(\tilde{\alpha}_{j,k}, \lambda) = \tilde{\alpha}_{j,k} I(|\tilde{\alpha}_{j,k}| \geq \lambda), \tag{5.1}$$

$$\delta^{(s)}(\tilde{\alpha}_{j,k}, \lambda) = (|\tilde{\alpha}_{j,k}| - \lambda)_+ \, sgn(\tilde{\alpha}_{j,k}), \tag{5.2}$$

where $\tilde{\alpha}_{j,k} = \int \tilde{\psi}_{j,k}(t)\, dY_t$ were appropriate estimates of $\alpha_{j,k}$. These two nonlinear procedures on the empirical coefficients are usually called hard and soft thresholding, respectively.
There exist two main results concerning the choice of these thresholds.

A) If the thresholds $\lambda = \lambda_j(\epsilon, \mathcal{F})$ are chosen optimally, then

$$\hat{f} = \sum_k \tilde{\alpha}_{lk}\phi_{lk} + \sum_{j \geq l} \sum_k \delta^{(.)}(\tilde{\alpha}_{j,k}, \lambda_j(\epsilon, \mathcal{F}))\psi_{jk}$$

is rate optimal in a ball $\mathcal{F} = B^m_{p,q}(C)$ of a certain Besov space.

B) If $\lambda = \epsilon\sqrt{2 \log \#\text{coefficients}}$, i.e. λ is neither depending on j nor $\mathcal{F}$, then the corresponding estimator is optimal in a wide range of smoothness classes within a factor of $\log(1/\epsilon)$.

We turn to the problem of spectral density estimation. We consider as approximating models for our empirical wavelet coefficients

$$\xi_{j,k} = \alpha_{j,k} + \sigma_{j,k}\varepsilon_{j,k} \tag{5.3}$$

and

$$\overline{\xi}_{j,k} = \alpha_{j,k} + (\sigma_{j,k} \vee \sigma_T)\varepsilon_{j,k} \tag{5.4}$$

where $\varepsilon_{j,k} \sim N(0,1)$. Then we have the following basic result for monotone coordinatewise estimators.

Theorem 5.1. *Assume (A1) through (A4). Let $\delta_{j,k} = \delta_{j,k,T}$ be monotone nondecreasing functions with*

$$|\delta_{j,k}(y)| \leq |y|. \tag{5.5}$$

Then, for $0 < p' < \infty$,

(i) $$\sum_{(j,k)\in\mathcal{J}^0} E|\delta_{j,k}(\widetilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} = (1+o(1)) \sum_{(j,k)\in\mathcal{J}^0} E|\delta_{j,k}(\xi_{j,k}) - \alpha_{j,k}|^{p'} \;+\; O(T^{-p'/2})$$

(ii) $$\sum_{(j,k)\in\mathcal{J}} E|\delta_{j,k}(\widetilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \le (2+o(1)) \sum_{(j,k)\in\mathcal{J}} E|\delta_{j,k}(\bar{\xi}_{j,k}) - \alpha_{j,k}|^{p'} \;+\; O(T^{-p'/2}).$$

Proceeding from this theorem we can obtain risk properties of thresholded wavelet estimators. Since the estimators (5.1) and (5.2) obey the assumption (5.5), we can immediately derive due to Theorem 5.1 the risk equivalence of our spectral density estimators to analogous estimators in the much simpler models (5.3) and (5.4).
Let $\delta^{(\cdot)}$ denote either the hard-threshold rule $\delta^{(h)}$ defined by (5.1) or the soft-threshold rule $\delta^{(s)}$ given by (5.2). Then we can state the following assertion.

Corollary 5.1. *Assume (A1) through (A4). Then, for nonrandom thresholds* $\lambda_{j,k}$,

(i) $$\sum_{(j,k)\in\mathcal{J}^0} E\left(\delta^{(\cdot)}(\widetilde{\alpha}_{j,k},\lambda_{j,k}) - \alpha_{j,k}\right)^2 = (1+o(1)) \sum_{(j,k)\in\mathcal{J}^0} E\left(\delta^{(\cdot)}(\xi_{j,k},\lambda_{j,k}) - \alpha_{j,k}\right)^2 \;+\; O(T^{-1}),$$

(ii) $$\sum_{(j,k)\in\mathcal{J}} E\left(\delta^{(\cdot)}(\widetilde{\alpha}_{j,k},\lambda_{j,k}) - \alpha_{j,k}\right)^2 \le (2+o(1)) \sum_{(j,k)\in\mathcal{J}} E\left(\delta^{(\cdot)}(\bar{\xi}_{j,k},\lambda_{j,k}) - \alpha_{j,k}\right)^2 \;+\; O(T^{-1}).$$

Let us now assume that the spectral density f lies in a set of the following type:

$$\mathcal{F} \;=\; \mathcal{F}(C_4) \;=\; \left\{ f = \sum_k \alpha_k \phi_{l,k} + \sum_{j,k} \alpha_{j,k}\psi_{j,k} \;\middle|\; \|\boldsymbol{\alpha}\|_{m,p,q} \le C_4 \right\}, \qquad (5.6)$$

where

$$\|\boldsymbol{\alpha}\|_{m,p,q} \;=\; \left(\sum_{j\ge l} \left(2^{js} \left(\sum_k |\alpha_{j,k}|^p \right)^{1/p} \right)^q \right)^{1/q} \qquad (5.7)$$

with $s = m + 1/2 - 1/p$. It is known that this norm is essentially equivalent to the norm in the Besov space $B^m_{p,q}$, if the basis functions ϕ and ψ satisfy (A4), cf. Donoho and Johnstone (1992). Moreover, we see by the relation $B^m_{p,1} \subseteq W^m_p \subseteq B^m_{p,\infty}$, that smoothness classes from the scale of Sobolev spaces W^m_p are also covered by our results.
Let

$$\widehat{f}^0 \;=\; \sum_{k\in I_l} \widetilde{\alpha}_k \phi_{l,k} \;+\; \sum_{(j,k)\in\mathcal{J}} \delta^{(\cdot)}(\widetilde{\alpha}_{j,k},\lambda_j^0)\psi_{j,k} \qquad (5.8)$$

be the estimator with optimal (nonrandom) thresholds $\lambda_j^0 = \lambda_j^0(T,\mathcal{F})$ and let

$$\widehat{f} \;=\; \sum_{k\in I_l} \widetilde{\alpha}_k \phi_{l,k} \;+\; \sum_{(j,k)\in\mathcal{J}} \delta^{(\cdot)}(\widetilde{\alpha}_{j,k},\lambda_{j,k})\psi_{j,k}. \qquad (5.9)$$

be an estimator with individual thresholds, which satisfy the following minimal conditions.

(A5) (i) $\sum_{(j,k)\in\mathcal{J}} \left(\frac{\lambda_{j,k}}{\sigma_{j,k}\vee\sigma_T} + 1\right) \varphi\left(\frac{\lambda_{j,k}}{\sigma_{j,k}\vee\sigma_T}\right) = O(T^{1/(2m+1)})$,

(ii) $\max_{(j,k)\in\mathcal{J}} \{\lambda_{j,k}\} = O\left(T^{-1/2}\sqrt{\log T}\right)$,

where φ denotes the standard normal density.

(A5) is motivated by the fact that one can show by Lemma 1 of Donoho and Johnstone (1994b) that for hard or soft thresholded estimators based on observations according to (5.4) the following relation holds:

$$E\left(\delta^{(\cdot)}(\overline{\xi}_{j,k},\lambda) - \alpha_{j,k}\right)^2 \le C\left((\sigma_{j,k}^2 \vee \sigma_T^2)\left(\frac{\lambda}{\sigma_{j,k}\vee\sigma_T} + 1\right)\varphi\left(\frac{\lambda}{\sigma_{j,k}\vee\sigma_T}\right) + \min\{\lambda^2, \alpha_{j,k}^2\}\right), \tag{5.10}$$

uniformly in $\lambda \ge 0$ and $\alpha_{j,k} \in \mathbb{R}$.
Two particular thresholding schemes, which obey (A5) are:

$$\lambda_{j,k} = (\sigma_{j,k}\vee\sigma_T)\sqrt{2\log(\#\mathcal{J})} \tag{5.11}$$

and

$$\lambda_{j,k} = (\sigma_{j,k}\vee\sigma_T)\sqrt{2\log((\#\mathcal{J})/2^l)}. \tag{5.12}$$

Now it can be easily shown that both of the proposed thresholding schemes lead to a rate $(\log T/T)^{2m/(2m+1)}$ for the risk of the estimator $\widehat{f}$.
The following theorem shows that we can attain the same rates in our spectral density framework as known from the case of Gaussian regression.

Theorem 5.2. *Assume (A1) through (A4). Then*

(i) $$\sup_{f\in\mathcal{F}} \left\{E\|\widehat{f}^0 - f\|_{L_2(\Pi)}^2\right\} = O\left(T^{-2m/(2m+1)}\right),$$

(ii) if additionally (A5) is satisfied, then
$$\sup_{f\in\mathcal{F}} \left\{E\|\widehat{f} - f\|_{L_2(\Pi)}^2\right\} = O\left((\log T/T)^{2m/(2m+1)}\right).$$

Remark 5.1. (i) It is known from Bentkus (1986) that $T^{-2m/(2m+1)}$ is just the optimal rate of convergence in Hölder smoothness classes with degree of smoothness m, if $\{X_t\}$ is a stationary Gaussian process. Therefore, $\widehat{f}^0$ attains the minimax rate of convergence for each smoothness class $\mathcal{F}$, which contains some ball from the corresponding Hölder space. Moreover, we see that $\widehat{f}$ is nearly minimax up to a factor of order $\log T$ in such smoothness classes.

(ii) It can easily be seen that even the simple linear (in the periodogram values) estimator

$$\widetilde{f} = \sum_k \widetilde{\alpha}_k \phi_{l,k} + \sum_{j\le j^0}\sum_k \widetilde{\alpha}_{j,k}\psi_{j,k}$$

with $j^0 = j^0(T)$ such that $2^{j^0} \asymp T^{1/(2m+1)}$ is rate optimal, i.e. it holds

$$\sup_{f\in\mathcal{F}} \left\{ E\|\widetilde{f} - f\|^2_{L_2(\Pi)} \right\} = O\left(T^{-2m/(2m+1)}\right),$$

if $\mathcal{F}$ is defined according to (5.6) and (5.7) with $p \geq 2$, i.e. $\mathcal{F}$ is some bounded set in a Besov space $B^m_{p,q}$ with $p \geq 2$.
However, it is shown in Donoho and Johnstone (1992) in the Gaussian white noise model that the optimal rate for linear estimators deteriorates to $O(T^{-2\widetilde{m}/(2\widetilde{m}+1)})$, $\widetilde{m} = m+1/2-1/p$, if $1 \leq p < 2$. Since the L_2-risk of linear estimators depends only on the first two moments of the error distributions, we can immediately conclude that linear spectral density estimators are restricted to the above-mentioned suboptimal rate for the risk.
Of course, one can make some objections to the minimax argument employed in favor of our nonlinear estimators. For example, the minimax view seems to be a bit pessimistic, because it focuses only on the worst case rather than on certain intermediate cases. However, we underline that balls in Besov spaces $B^m_{p,q}$ with small values of p serve as a caricature of functions with spatially inhomogeneous smoothness properties, which are of particular interest in spectral analysis.

5.2. Further improvement of the estimators. The estimator $\widehat{f}$ attains already the desired near-optimal rate $(\log T/T)^{2m/(2m+1)}$, but there are two obvious possibilities to improve it further.
First, in contrast to usual kernel estimators of f, the wavelet estimators are not translation-invariant. If we shift the periodogram by a certain amount s on the real line, apply nonlinear thresholding, and shift the estimate back by s, this new estimator $\widehat{f}^{(s)}$ will differ from the unshifted variant $\widehat{f}$ in most cases. The only shift lengths, which do not alter the estimator $\widehat{f}$, are multiples of the shift length of the wavelet basis at the coarsest scale, i.e. $2\pi/2^l$. On the other hand, there is no reason to assume that any of the possible shifts is always superior to other ones. To weaken the effect of not being translation-invariant we apply the well-known idea of stationary wavelet transforms (see, e.g., Coifman and Donoho (1995), Nason and Silverman (1995)) and define, with shifts $s_i = i2\pi/(2^l I)$, $i = 0, \ldots, I-1$, the new estimator

$$\widehat{f}^*(\omega) = I^{-1} \sum_{i=0}^{I-1} \widehat{f}^{(s_i)}(\omega). \tag{5.13}$$

Then we obtain by Jensen's inequality that

$$\|\widehat{f}^* - f\|^2_{L_2(\Pi)} \leq I^{-1} \sum_{i=0}^{I-1} \|\widehat{f}^{(s_i)} - f\|^2_{L_2(\Pi)}, \tag{5.14}$$

where strict inequality holds if $\widehat{f}^{(s_i)} \not\equiv \widehat{f}^{(s_j)}$ for any (i, j). (5.14) means in particular that $\widehat{f}^*$ also satisfies (ii) of Theorem 5.2. Moreover, in view of the possible strict inequality in (5.14), we hope to get a significant improvement for finite sample sizes.
Second, note that the spectral density f is necessarily symmetric w.r.t. 0, whereas $\widehat{f}^*$ is asymmetric, if compactly supported wavelets different from the Haar wavelet are

used. Further, $\widehat{f}^*$ as well as $\widehat{f}^{*,-}(\omega) = \widehat{f}^*(-\omega)$ have the same L_2-loss. Hence, we have again by Jensen's inequality that the new estimator $\widehat{f}^{**}(\omega) = (\widehat{f}^*(\omega) + \widehat{f}^{*,-}(\omega))/2$ satisfies

$$\|\widehat{f}^{**} - f\|^2_{L_2(\Pi)} \leq \left(\|\widehat{f}^* - f\|^2_{L_2(\Pi)} + \|\widehat{f}^{*,-} - f\|^2_{L_2(\Pi)}\right)/2 = \|\widehat{f}^* - f\|^2_{L_2(\Pi)}, \tag{5.15}$$

where strict inequality holds if $\widehat{f}^* \not\equiv \widehat{f}^{*,-}$.
We tested these two postprocessing methods in our simulation study described in Section 7, and noticed a significant improvement over the raw estimator $\widehat{f}$.

6. Adaptive threshold choice

Although the results of Theorem 5.2 are certainly of some theoretical interest, in particular in conjunction with Remark 5.1, they are not helpful for practical applications. The optimal as well as the log-thresholds depend on a priori assumptions on the set $\mathcal{F}$, or on f itself via the variances of the empirical wavelet coefficients, respectively.
To make the method applicable, it is necessary to find some completely data-driven rule for the thresholds, which works well over an as wide as possible range of smoothness classes.
There already exist many proposals for the tuning of linear estimators of a spectral density. Wahba and Wold (1975) proposed some cross-validation criterion to choose the degree of smoothing of a periodic spline estimator automatically. By showing that the expectation of the cross-validation function is asymptotically close to the mean squared error they gave some indication that this criterion yields an asymptotically optimal smoothing parameter. Wahba (1980) proposed for spline estimators of the log periodogram to adapt the smoothing parameter by minimizing an asymptotically unbiased risk estimate. Hurvich (1985) considered in the framework of quite general classes of estimators some refinement of the usual leave-one-out technique used in cross-validation methods. For kernel estimators, Beltrão and Bloomfield (1987) proposed some cross-validated likelihood approach to determine the bandwidth, and gave also some indications that this method can work. However, since it is based on Whittle's approximation of the likelihood, its applicability seems to be restricted to Gaussian time series only. Franke and Härdle (1992) proposed to use a bootstrap estimate of the mean square percentage error and showed that the bootstrap distribution approaches the distribution of the estimation errors in probability. This gives certainly some hope that the risk of the corresponding estimator is asymptotically close to the risk of the estimator with optimal bandwidth. However, the construction of the bootstrap estimate requires, besides the bandwidth of interest, two additional bandwidths that must tend to zero with some suboptimal rates. It is an open problem to find a reasonable data-driven rule to determine these bandwidths, which would then provide a completely defined rule.
The development of a completely data-driven, asymptotically optimal rule for the smoothing parameters seems to be more involved in our case. The classical leave-one-out cross-validation is known to work in the case of independent observations in regression. In the case of a Gaussian time series it is known that different periodogram

values are almost independent, and therefore it is not surprising that cross-validation works for linear estimators; see the literature cited above.
However, the case under consideration here shows some essential differences, which make life more difficult. One of our main intentions is to allow also non-Gaussian time series, which makes a rigorous proof of independence properties between different periodogram values more involved. Further, an essential feature of our estimators is their nonlinearity. Hence, we would presumably need different techniques for getting an efficient risk estimate. Finally, our wavelet estimators are *multiscale* estimators, which include scales with quite a large number of small coefficients. Usually, in such cases of sparse signals the relative performance of level-wise risk estimates, compared with the minimal risks at these levels themselves, deteriorates, as claimed in Donoho and Johnstone (1993) for Stein's Unbiased Risk Estimate or in Neumann and Spokoiny (1995) for leave-one-out cross-validation. We guess that the same effect would annoy us in the spectral density estimation context.
To end up with a practicable rule, we propose a much simpler, admittedly slightly suboptimal, approach. Since we exploit asymptotic normality of the empirical wavelet coefficients, we need only appropriate preliminary estimates for their variances in order to tune the wavelet estimator. Assume that any consistent estimator $\widetilde{f}$ of f is available. To be more specific, we will not need certain rates of convergence for $\widetilde{f}$; rather some kind of uniform convergence, with sufficiently high probability, will be enough. According to (ii) of Proposition 3.1 we can then use

$$\widehat{\sigma}_{j,k}^2 \;=\; 2\pi \left(H_4^{(T)}/(H_2^{(T)})^2\right) \int_\Pi \psi_{j,k}(\alpha)\,[\psi_{j,k}(\alpha)+\psi_{j,k}(-\alpha)]\,|\widetilde{f}(\alpha)|^2\,d\alpha \qquad (6.1)$$

as an estimate of $var(\widetilde{\alpha}_{j,k})$ for all levels $j=j(T)$ with $2^j \gg 1$.
In analogy to (5.11) and (5.12) we obtain adaptive thresholds as

$$\widehat{\lambda}_{j,k} \;=\; \widehat{\sigma}_{j,k}\sqrt{2\log(\#\mathcal{J})} \qquad (6.2)$$

and

$$\widehat{\lambda}_{j,k} \;=\; \widehat{\sigma}_{j,k}\sqrt{2\log((\#\mathcal{J})/2^l)}. \qquad (6.3)$$

A sufficient condition for random thresholds, which ensures the desired rate for the estimator, is the following one.

(A6) (i) $\sum_{(j,k)\in\mathcal{J}} P\left(\widehat{\lambda}_{j,k} \;<\; \alpha_T(\sigma_{j,k}\vee\sigma_T)\sqrt{2\log(C(\#\mathcal{J}))}\right) \;=\; O(T^\nu)$
for any $C>0$, $\nu<1/(2m+1)$ and $\alpha_T\to 1$,
(ii) $\sum_{(j,k)\in\mathcal{J}} P\left(\widehat{\lambda}_{j,k} \;>\; \widetilde{C}T^{-1/2}\sqrt{\log T}\right) \;=\; O(T^{-2m/(2m+1)})$
for any $\widetilde{C}<\infty$.

The discrepancy between the rates in (i) and (ii) reflects the fact that a too small thresholds is less dramatic than a too large one; see the proof of Theorem 6.1 for details. The above bounds for the probabilities of the appearance of "forbidden thresholds" are actually easily fulfilled under weak conditions. The following lemma shows two particular cases, where we can easily satisfy (A6) plugging in suitable kernel estimators into (6.1).

Lemma 6.1. *Let* $\widetilde{f}(\omega) = \int h_T^{-1} k((\omega - \lambda)/h_T) I_T(\lambda)\, d\lambda$ *be a kernel estimator with nonrandom bandwidth* h_T *based on the tapered periodogram* I_T. *Assume (A1) through (A4) and let* $f(\omega) \geq C > 0$ *for all* $\omega \in \Pi$. *Then*

(i) *if* $m > 1/p$ *and* $h_T \to 0$, $h_T^{-1} = O(T^{1-\delta})$ *for any* $\delta > 0$, *then*

$$P\left(\alpha_T f(\omega) \leq \widetilde{f}(\omega) \leq \widetilde{C} \quad \text{for all } \omega \in \Pi\right) = 1 - O(T^{-2})$$

holds uniformly in $\mathcal{F}$ *for some* $\alpha_T \to 1$ *and* $\widetilde{C} < \infty$,

(ii) *if* $h_T = O(T^{\nu-1})$ *and* $h_T^{-1} = O(T^{1-\delta})$ *for any* $\nu < 1/(2m+1)$ *and* $\delta > 0$, *then*

(a) $\sum_{(j,k)\in\mathcal{J}} P\left(\widetilde{f}(\omega) < \alpha_T f(\omega) \quad \text{for any } \omega \in supp(\psi_{j,k})\right) = O(T^{\nu})$,

(b) $P\left(\widetilde{f}(\omega) > \widetilde{C} \quad \text{for any } \omega \in \Pi\right) = O(T^{-2})$

holds uniformly in $\mathcal{F}$ *for some* $\alpha_T \to 1$ *and* $\widetilde{C} < \infty$,

Part (i) of this lemma shows that it is actually not difficult to find an appropriate kernel estimator if $m > 1/p$: Since f is then continuous, the conditions $h_T \to 0$ and $h_T/T \to \infty$ ensure pointwise consistency of $\widetilde{f}$. Moreover, $h_T^{-1} = O(T^{1-\delta})$ ensures a sufficiently fast decay of probabilities of large deviations, which leads to the tight bound for the probability of the event $\{\widetilde{f}(\omega) \notin [\alpha_T f(\omega), \widetilde{C}]$ for any $\omega\}$.
The case of a spectral density which is only of bounded variation is slightly more involved: Since f could be discontinuous, there is no hope to obtain a uniformly consistent estimator. However, the number of jumps of f whose height exceeds $c_T > 0$ is of order $O(c_T^{-1})$. Since each $\psi_{j,k}$ overlaps only with a finite number of $\psi_{j,k'}$'s from the same scale j, we are able to prove (ii)a, if h_T decays sufficiently fast. (ii)b follows again from the boundedness of f.
Now it can be easily seen that, in the cases considered in Lemma 6.1, the thresholds according to (6.1) satisfy (A6). The performance of the resulting estimator

$$\widehat{\widehat{f}} = \sum_{k\in I_l} \widetilde{\alpha}_k \phi_{l,k} + \sum_{(j,k)\in\mathcal{J}} \delta^{(\cdot)}(\widetilde{\alpha}_{j,k}, \widehat{\lambda}_{j,k}) \psi_{j,k}.$$

is described in the following theorem.

Theorem 6.1. *Assume (A1) through (A4) and (A6). Then*

$$\sup_{f\in\mathcal{F}} \left\{ E\|\widehat{\widehat{f}} - f\|^2 \right\} = O((\log T/T)^{2m/(2m+1)}).$$

For simplicity of its proof, we formulated this theorem for the unmodified estimator $\widehat{\widehat{f}}$. In practice one could also use a shifted and symmetrized version of it as described in the previous section.
Perhaps one could feel a little bit disappointed in view of the slight suboptimality of the estimator $\widehat{\widehat{f}}$, which misses the minimax rate by a logarithmic factor. We suspect however, that it is quite difficult to find a fully data-driven estimator, which attains the optimal rate simultaneously in different smoothness classes. Here we have a unique estimator, which behaves reasonably well over a wide range of smoothness

classes, whenever an initial estimator $\tilde{f}$, which is a bit better than only consistent, is available.

7. Simulations

In this section we present some results of a simulation study, which support the claims made in the theoretical part of this paper.
We used as a convenient programming environment the *XploRe* system, which has been developed by W. Härdle and coworkers, and runs on personal computers. A description of this is contained in Härdle *et al.* (1995).
We simulated a time series, which was a superposition of an ARMA(2,2)-process and a Gaussian white noise:

$$X_t \;=\; Y_t \;+\; c_0 Z_t, \tag{7.1}$$

where

$$Y_t \;+\; a_1 Y_{t-1} \;+\; a_2 Y_{t-2} \;=\; b_0 \varepsilon_t \;+\; b_1 \varepsilon_{t-1} \;+\; b_2 \varepsilon_{t-2},$$

and $\{\varepsilon_t\}$, $\{Z_t\}$ are independent Gaussian white noise processes with variance 1. The constants above were chosen as $a_1 = 0.2$, $a_2 = 0.9$, $b_0 = 1$, $b_1 = 0$, $b_2 = 1$ and $c_0 = 0.5$. The spectral density of $\{X_t\}$ is shown in Figure 1. It has two moderately sharp peaks and is smooth on the rest of the domain.

[Please insert Figure 1 about here]

We generated 100 samples of size $T = 1024$ according to (7.1) using the Gaussian pseudo-random number generator from *XploRe*. To reduce the influence of the initial conditions and to diminish the dependence between the samples we discarded the first 100 output values at each simulation loop.
Starting from the periodogram with a cosine taper, i.e. $h(t) = 1 - \cos(2\pi t)$, we considered the kernel estimator $\tilde{f}$ with quartic kernel $k(s) = 0.9375(1 - s^2)_+^2$ and that global bandwidth which minimizes the L_2-risk of the estimator. This estimator was compared with two versions of the proposed nonlinear wavelet estimators. We use the Symmlet 10 basis, i.e. the least asymmetric, compactly supported wavelets which are described on page 198 of Daubechies (1992). Note that knowledge of the explicit forms of the scaling function ϕ and the wavelet ψ is not necessary to calculate the wavelet estimator. Instead of this one usually uses discrete filters to calculate the empirical coefficients and after modifying them to build the estimator; see e.g. Daubechies (1992).
In the notation of this paper, we chose $l = 3$ and let the coefficients assigned to the father wavelets unthresholded. Hard thresholding was performed for the levels $j = 3, 4, 5$, and the empirical coefficients from the higher resolution scales $j > 5$ were zeroed. We investigated two possibilities for the thresholds, $\widehat{\lambda}_{j,k}^{(1)} = \widehat{\sigma}_{j,k}\sqrt{2\log \#\mathcal{J}} = \widehat{\sigma}_{j,k}\sqrt{2\log(2^5 - 2^3)}$ as well as $\widehat{\lambda}_{j,k}^{(2)} = \widehat{\sigma}_{j,k}\sqrt{2\log(\#\mathcal{J}/2^3)} = \widehat{\sigma}_{j,k}\sqrt{2\log 3}$. The estimators $\widehat{\sigma}_{j,k}$ were obtained plugging in the kernel estimator $\tilde{f}$ into the asymptotic formula given in (ii) of Proposition 3.1. We investigated these two versions of the wavelet estimator, as well as the corresponding improved (i.e. symmetrized and shifted) versions of them with 32 shifts.

In Figures 2a and b we show one realization of the kernel estimator with optimally chosen global bandwidth (which was $h = 0.16$) and the best of our wavelet estimators (i.e. the improved version with $\widehat{\lambda}_{j,k}^{(2)}$), both computed for the same set of observations. The dotted line shows the true spectral density. These pictures show the expected effect: the global kernel estimator cannot simultaneously do well at the peaks and on the smooth parts. The wavelet estimator provides some improvement in this direction: it better captures the peaks and looks slightly better on the smooth part, too.

[Please insert Figures 2a and b about here]

We also estimated the L_2-risk of these estimators on the basis of 100 simulation runs. They are contained in Table 1, together with the radii of approximate 95%-confidence intervals of the form $1.96\widehat{v}/10$, where $\widehat{v}^2$ is the estimated variance of the corresponding risk estimate.

estimator	L_2-risk	$\pm$95%-confidence bands
kernel	0.0419	0.0032
w(1,u)	0.0657	0.0042
w(2,u)	0.0536	0.0053
w(1,s)	0.0436	0.0038
w(2,s)	0.0398	0.0043

Here $w(i,u)$ and $w(i,s)$ refer to the unshifted or shifted versions of the wavelet method with thresholding scheme $\widehat{\lambda}_{j,k}^{(i)}$, respectively.

Even though the improvement of the kernel estimator is not significant, the results indicate that wavelet estimators can compete with kernel estimators also for realistic sample sizes.

Let us finally note that these simulations were a little bit unfair to the wavelet methods, since their competitor uses the L_2-optimal bandwidth, which is usually unknown. We expect at least a small deterioration of the performance of the kernel method, when the bandwidth has to be determined from the data alone.

8. Concluding remarks

1) There exist several possibilities to weaken the cumulant assumption (A3). One could obtain analogous results as in Theorem 5.1 under one of the following modifications of (A3).
   - (A3') The constant C_2 in (A3) can be allowed to increase with T with a sufficiently slow rate.
   - (A3") There exist random variables $\widetilde{X}_{t_i}$ such that
     - (i) $P(X_{t_i} \neq \widetilde{X}_{t_i}) \leq CT^{-\lambda}$ for λ large enough
     - (ii) $\sup_{1 \leq t_1 < \infty} \left\{ \sum_{t_2,\dots,t_k=1}^{\infty} \left| cum(\widetilde{X}_{t_1}, \dots, \widetilde{X}_{t_k}) \right| \right\} \leq C^k(T)(k!)^{1+\gamma}$ for all $k = 2,3,\dots,$ where $C(T) = O(T^{\delta})$ for some sufficiently small δ
     - (iii) $E|X_t|^M \leq C(M)$ for some finite, but sufficiently large M.
2) We remark that Gao (1993) obtained related results for the case of Gaussian time series. He found estimates for the L_2-risk of thresholded wavelet estimators not via an asymptotic normal approximation, but he regarded the empiri-

cal coefficients as quadratic forms of independent Gaussian random variables and used estimates for the tail probabilities of them. This leads to thresholds of order $T^{-1/2}\log T$, which cause a loss in efficiency of order $(\log T)^2$ instead of $\log T$ as in our case. For Gaussian time series, this approach allows the inclusion of all resolution levels up to a scale j^* with $2^{j^*} \asymp T$.
We also want to mention the paper by Moulin (1994) on wavelet smoothing of the log-periodogram. Starting from Wahba's approximation for the periodogram he obtained appropriate thresholds via a saddlepoint approximation of the distribution of the empirical wavelet coefficients. However, these results have been stated only for Gaussian time series, an assumption which we intend to avoid here.

3) As may be seen in Theorem 5.1, the risk equivalence to the Gaussian case can also be shown for other loss functions as the quadratic one. If we are especially interested in recognizing peaks of the true spectral density, we could take this special goal into account by an appropriate choice of the loss function. The L_p-loss for large values of p is more sensitive to relatively large estimation errors at certain small intervals than the L_2-loss, and therefore more sensitive to a misjudgement of a peak. As shown in Donoho *et al.* (1995), we can use for this purpose the same estimator as in the case considered here.
4) The use of any orthogonal series expansion in spectral density estimation suffers from the (at least aesthetical) drawback of producing estimators that may attain negative values as well. To repair this defect, one could use the corrected estimator $\widehat{\widehat{f}}(\omega) = \max\{\widehat{f}(\omega), 0\}$.
5) Explicit asymptotic expressions for integrated squared bias and integrated variance are usually available for kernel estimators. Although bias and variance alone cannot be considered as measures for the performance of an estimator, the relation between them provides some insight into the nature of the smoothing method.
Due to the application of the slightly suboptimal "denoising"-thresholds at all resolution scales, our estimator is oversmoothed. It can be seen that the integrated squared bias term is of order $O((\log T/T)^{2m/(2m+1)})$, whereas the integrated variance is of order $o(T^{-2m/(2m+1)})$.
In principle it seems to be possible to derive explicit asymptotic formulae for integrated squared bias and integrated variance. After deriving exact formulae in the Gaussian white noise model along the lines of the derivation of exact risk formulae in Donoho and Johnstone (1994a), one can employ again the asymptotic normality of the empirical coefficients to get such asymptotic approximations. However, one would not very much learn from these quite involved formulae, and therefore we do not make the attempt to derive them.

9. Proofs

Proof of Lemma 3.1. This proof is similar to that of Lemma 2 in [29]. Because of the more general assumptions on the matrix A and on the cumulants we have to modify it slightly. By splitting the estimate in a part, which occurs also with Gaussian processes, and a contribution by the non-Gaussianity we get an improved result.
The cumulants of η can be estimated with the aid of Leonov-Shiryaev's formula,

see [22]. For that we have to write η as a polynomial in certain random variables, where each of them enters at most with the power one into each of the terms of this polynomial. Therefore, we define $X_{T+t} = X_t$ for $t = 1, \ldots, T$, such that η can be written in the form

$$\eta = \sum_{\vartheta \in \Theta} a(\vartheta) X_1^{\vartheta_1} \cdots X_{2T}^{\vartheta_{2T}}, \tag{9.1}$$

where $\vartheta = (\vartheta_1, \ldots, \vartheta_{2T})$, $\Theta = \{\vartheta \,|\vartheta_t \in \{0,1\}, \sum \vartheta_t = 2\}$, and $a(\vartheta) = a_{st}$ if $\vartheta_s = \vartheta_{T+t} = 1$.
Using Leonov-Shiryaev's formula we obtain

$$cum_n(\eta) = \sum_{\vartheta^{(1)},\ldots,\vartheta^{(n)} \in \Theta} a(\vartheta^{(1)}) \cdots a(\vartheta^{(n)}) \sum_{D(\vartheta) = D_1 + \ldots + D_q}^{*} c(t_{D_1}) \cdots c(t_{D_q}), \tag{9.2}$$

where the sum $\sum^*$ is taken over all indecomposable partitions of $D(\vartheta) = \left\{(i,j) \,\middle|\, \vartheta_j^{(i)} = 1, \;\; i \in \{1, \ldots, n\}, \;\; j \in \{1, \ldots, 2T\}\right\}$. Here, $c(t_{D_p})$ is defined as $cum\left(X_{t_{m_{p,1}}}, \ldots, X_{t_{m_{p,r_p}}}\right)$, where $D_p = \left\{(m_{p,1}, t_{m_{p,1}}), \ldots, (m_{p,r_p}, t_{m_{p,r_p}})\right\}$.
For our further calculations it is not necessary to work with the general, but quite involved formula (9.2); it tells us only, which products of cumulants must be included in the calculation of $cum_n(\eta)$. We rewrite (9.2) as

$$cum_n(\eta) = \sum_{t_1,\ldots,t_{2n}=1}^{T} a_{t_1,t_2} \cdots a_{t_{2n-1},t_{2n}} \sum_{D = D_1 + \ldots + D_q}^{**} c(t_{D_1}) \cdots c(t_{D_q}), \tag{9.3}$$

where $D = \{1, \ldots, 2n\}$ and $\sum^{**}$ is the analog of $\sum^*$ from (9.2). Analogously to (9.2) we define $c(t_{D_p}) = c(t_{m_{p,1}}, \ldots, t_{m_{p,r_p}}) = cum\left(X_{t_{m_{p,1}}}, \ldots, X_{t_{m_{p,r_p}}}\right)$ if $D_p = \left\{m_{p,1}, \ldots, m_{p,r_p}\right\}$.
Since $E\,X_t = 0$ for all t, we can restrict our considerations to indecomposable decompositions with $r_p \geq 2$. We divide the right-hand side of (9.3) into two parts, R_1 and R_2, the first one containing all indecomposable partitions $D = D_1 + \ldots + D_n$, which consist of exactly n sets, and the second one containing the remaining partitions $D = D_1 + \ldots + D_q$ for any $q < n$.
Because of $r_p \geq 2$, R_1 contains all pure products of second order cumulants and is just the n-th cumulant of ξ. We obtain from (3.7) in [3] that

$$|R_1| \leq var(\xi) 2^{n-2} (n-1)! \left[\lambda_{\max}\left(A\, cov(\underline{X}_T)\right)\right]^{n-2}. \tag{9.4}$$

Now we estimate $|R_2|$. Let $D(\vartheta) = D_1 + \ldots + D_q$ be any indecomposable partition, which occurs in the sum $\sum^*$. Since it is indecomposable, there exist, possibly after an appropriate renumeration of the sets D_i, numbers $p_2, \ldots, p_q \in \{1, \ldots, n\}$ and associated indices t_{p_i}, $\widetilde{t}_{p_i}$ such that

$$(p_i, t_{p_i}) \in D_i \quad \text{and} \quad (p_i, \widetilde{t}_{p_i}) \in \bigcup_{j=1}^{i-1} D_j \quad \text{for all } i = 2, \ldots, q. \tag{9.5}$$

It is easy to see that the partition

$$\widetilde{D}(\vartheta) = \widetilde{D}_1 + \ldots + \widetilde{D}_q$$

of the reduced set $\widetilde{D}(\vartheta) = \{(p_i, t_{p_i}), (p_i, \widetilde{t}_{p_i}) \mid i = 2, \dots, q\}$, where $\widetilde{D}_p = D_p \cap \widetilde{D}(\vartheta)$, is again irreducible.
Since $q < n$, there is at least one pair, say $(p_1, t_{p_1}), (p_1, \widetilde{t}_{p_1})$ left. Let w.l.o.g. $(p_1, t_{p_1}) \in D_1$. (If $(p_1, t_{p_1}) \in D_i$, $i > 1$, then we can find a rearrangement of the sets $D_1, \dots, D_q$, where D_i stands at the first place and where (9.5) can be established with the same set of pairs from $\widetilde{D}(\vartheta)$.)
Now we consider that partition $D = D_1 + \dots + D_q$ from the sum $\sum^{**}$, which corresponds to $D(\vartheta) = D_1 + \dots + D_q$ from above. According to (9.5) there exist pairs $(s_p, \widehat{s}_p)$ with $\{s_p, \widehat{s}_p\} = \{2i-1, 2i\}$ for some $i \in \{1, \dots, n\}$ with

$$s_1 \in D_1 \quad \text{and} \quad s_p \in D_p, \quad \widehat{s}_p \in \bigcup_{j=1}^{p-1} D_j \quad \text{for} \quad p = 2, \dots, q. \tag{9.6}$$

Since the sets D_p are not ordered, we set w.l.o.g. $m_{1,1} = s_1$, $m_{1,r_1} = \widehat{s}_2$, and $m_{p,r_p} = s_p$, $p = 2, \dots, q$.
Now we have, using the symmetry of A,

$$\begin{aligned}
&\left| \sum_{t_1, \dots, t_{2n}=1}^{T} a_{t_1,t_2} \cdots a_{t_{2n-1},t_{2n}} c(t_{D_1}) \cdots c(t_{D_q}) \right| \\
&\leq \max_{s,t} \left\{ |a_{s,t}|^{n-q} \right\} * \qquad\qquad (9.7) \\
&\quad * \sum_{t_{s_1}=1}^{T} \max_u \{|a_{t_{s_1},u}|\} \sum_{t_{m_{1,2}}, \dots, t_{m_{1,r_1-1}}, t_{\widehat{s}_2}=1}^{T} \left| c\left(t_{m_{s_1}}, t_{m_{1,2}}, \dots, t_{m_{1,r_1-1}}, t_{\widehat{s}_2}\right) \right| * \\
&\quad * \sum_{t_{s_2}=1}^{T} |a_{t_{s_2}, t_{\widehat{s}_2}}| \sum_{t_{m_{2,1}}, \dots, t_{m_{2,r_2-1}}=1}^{T} \left| c\left(t_{m_{2,1}}, \dots, t_{m_{2,r_2-1}}, t_{s_2}\right) \right| * \\
&\quad \vdots \\
&\quad * \sum_{t_{s_q}=1}^{T} |a_{t_{s_q}, t_{\widehat{s}_q}}| \sum_{t_{m_{q,1}}, \dots, t_{m_{q,r_q-1}}=1}^{T} \left| c\left(t_{m_{q,1}}, \dots, t_{m_{q,r_q-1}}, t_{s_q}\right) \right|.
\end{aligned}$$

Now we stepwise introduce upper estimates for the sums on the right-hand side of (9.7). The last sum is, uniformly in $\widehat{s}_q$, less or equal than $\|A\|_\infty C^{r_q} (r_q!)^{1+\gamma}$. Then we estimate the last but one sum by $\|A\|_\infty C^{r_{q-1}} (r_{q-1}!)^{1+\gamma}$, and so on. Finally we obtain that

$$\begin{aligned}
&\left| \sum_{t_1, \dots, t_{2n}=1}^{T} a_{t_1,t_2} \cdots a_{t_{2n-1},t_{2n}} c(t_{D_1}) \cdots c(t_{D_q}) \right| \\
&\leq \max_{s,t} \left\{ |a_{s,t}|^{n-q} \right\} \ \widetilde{A} \ \|A\|_\infty^{q-1} C^{2n} \left(r_1! \cdots r_q!\right)^{1+\gamma} \qquad\qquad (9.8) \\
&\leq \max_{s,t} \{|a_{s,t}|\} \ \widetilde{A} \ \|A\|_\infty^{n-2} C^{2n} \left(r_1! \cdots r_q!\right)^{1+\gamma}.
\end{aligned}$$

From the proof of Lemma 2 in [29] we get

$$\sum_{D=D_1+\cdots+D_q}^{**} (r_1!\cdots r_q!)^{1+\gamma} \le ((2n)!)^{\gamma} \sum^{**} r_1!\cdots r_q! \le 2^{2n-2}\,((2n)!)^{1+\gamma}, \quad (9.9)$$

which yields

$$|R_2| \le 2^{n-2}C^{2n}((2n)!)^{1+\gamma} \max_{s,t}\{|a_{st}|\}\ \tilde{A}\ \|A\|_\infty^{n-2}. \quad (9.10)$$

□

Proof of Proposition 3.1. *(i)*, *(ii)* The proofs of (i) and (ii) are quite similar to calculations in [6] and are therfore omitted. The interested reader can find them in an earlier version of this paper, [27].
(iii) We write $\tilde{\alpha}_{j,k}$ as $\underline{X}'A\underline{X}$ with $A = ((a_{st}))$, where

$$a_{st} = \frac{h_s h_t}{2\pi H_2^{(T)}} \int \psi_{j,k}(\omega)\cos(\omega(s-t))\,d\omega.$$

Since $TV(\psi_{j,k}) = O(2^{j/2})$, we get by integration by parts that

$$a_{st} = \begin{cases} O\left(2^{j/2}T^{-1}|s-t|^{-1}\right), & \text{if } s \ne t \\ 0, & \text{if } s = t \end{cases},$$

which implies

$$\|A\|_\infty = O\left(2^{j/2}T^{-1}\log(T)\right).$$

Further, we have $a_{st} = O(2^{-j/2}T^{-1})$, which yields

$$\tilde{A} = O\left(2^{-j/2}\right).$$

Therefore, we obtain by Lemma 3.1 that

$$\begin{aligned} |cum_n(\tilde{\alpha}_{j,k})| &= O\left(n!C^nT^{-1}(2^{j/2}T^{-1}\log(T))^{n-2}\right) \\ &\quad + O\left((n!)^{2+2\gamma}C^n 2^{-j}T^{-1}(2^{j/2}T^{-1}\log(T))^{n-2}\right), \end{aligned}$$

which proves the assertion. □

Proof of Theorem 4.1. Using (iii) of Proposition 3.1 as well as the assertion of Lemma 3.1 we obtain due to Lemma 1 in [30] that

$$\frac{P\left(\pm(\tilde{\alpha}_{j,k} - E\tilde{\alpha}_{j,k})/\sigma_{j,k} \ge x\right)}{1-\Phi(x)} = e^{L(x)}\left(1 + g(x)\frac{x+1}{\Delta^{1/(3+4\gamma)}}\right) \quad (9.11)$$

for $0 \le x \le \Delta^{1/(3+4\gamma)}$, where Δ is an appropriate constant of order $T^{1/2}2^{-j/2}(\log T)^{-1}$. Here, g is some bounded function, and by (3) and (4) in [30] we infer that $|L(x)| \le Cx^3/\Delta$. Therefore, the right-hand side of (9.11) converges uniformly to 1, if $0 \le x = o(\Delta^{1/(3+4\gamma)})$ is satisfied. (For $x < 0$ this convergence is an immediate consequence.) Since $b = (\alpha_{j,k} - E\tilde{\alpha}_{j,k})/\sigma_{j,k} = O(\Delta^{-1})$, we conclude that

$$\frac{P\left(\pm(\tilde{\alpha}_{j,k} - \alpha_{j,k})/\sigma_{j,k} \ge x\right)}{1-\Phi(x)} = (1 + o(1))\frac{1-\Phi(x+b)}{1-\Phi(x)} \quad (9.12)$$

holds uniformly over $-\infty \le x \le \Delta_\gamma$. Now it remains to estimate

$$\eta = \left| \frac{1-\Phi(x+b)}{1-\Phi(x)} - 1 \right| = \frac{|\Phi(x+b)-\Phi(x)|}{1-\Phi(x)}.$$

Let w.l.o.g. $b \ge 0$. (Otherwise we would estimate $\left|\frac{1-\Phi(x)}{1-\Phi(x+b)} - 1\right|$ in the same manner as η below.)
Fix any $c > 1$. Since $b = O(T^{-\alpha/2}\log(T))$, it is clear that η tends to zero if $x \le c$. For $c < x \le \Delta_\gamma$ we obtain by a formula for Mill's ratio (see [21], vol. 2, p. 278) that

$$\eta \le \frac{b\varphi(x)}{1-\Phi(x)} \le \frac{bx}{1-1/x^2} \longrightarrow 0 \quad \text{as} \quad T \to \infty,$$

which completes the proof. $\square$

Proof of Theorem 5.1.
(i) Since $\delta_{j,k}$ is monotone, there exists a constant $\gamma_{j,k}$ such that

$$\begin{array}{lclcl} \delta_{j,k}(y) & \ge & \alpha_{j,k}, & \text{if} & y > \gamma_{j,k} \\ \delta_{j,k}(y) & \le & \alpha_{j,k}, & \text{if} & y < \gamma_{j,k}. \end{array}$$

(W.l.o.g. we assume $\delta_{j,k}(\gamma_{j,k}) \ge \alpha_{j,k}$.)
Now we split up

$$\begin{aligned} & E\,|\delta_{j,k}(\tilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \\ &= E\,I\,(\gamma_{j,k} \le \tilde{\alpha}_{j,k} < \alpha_{j,k} + \sigma_{j,k}\Delta_\gamma)\,|\delta_{j,k}(\tilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \\ &\quad + E\,I\,(\alpha_{j,k} - \sigma_{j,k}\Delta_\gamma < \tilde{\alpha}_{j,k} < \gamma_{j,k})\,|\delta_{j,k}(\tilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \qquad (9.13) \\ &\quad + E\,I\,(|\tilde{\alpha}_{j,k} - \alpha_{j,k}| \ge \sigma_{j,k}\Delta_\gamma)\,|\delta_{j,k}(\tilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \\ &= R_1 + R_2 + R_3. \end{aligned}$$

According to the assertion of Theorem 4.1 there exist $C_T^{(l)}$, $C_T^{(u)}$, both tending to 1 as $T \to \infty$, such that

$$C_T^{(l)}(1-\Phi(x)) \le P\,(\pm(\tilde{\alpha}_{j,k} - \alpha_{j,k})/\sigma_{j,k} \ge x) \le C_T^{(u)}(1-\Phi(x)) \quad \forall x \le \Delta_\gamma. \qquad (9.14)$$

Since $|\delta_{j,k}(y) - \alpha_{j,k}|$ is monotone nondecreasing for $y \ge \gamma_{j,k}$, we obtain by integration by parts that

$$\begin{aligned} R_1 &= -\int \left[I\,(\gamma_{j,k} \le x < \alpha_{j,k} + \sigma_{j,k}\Delta_\gamma)\,|\delta_{j,k}(x) - \alpha_{j,k}|^{p'} \right] dP\,(\tilde{\alpha}_{j,k} \ge x) \\ &= \int P\,(\tilde{\alpha}_{j,k} \ge x)\; d\left[I\,(\gamma_{j,k} \le x < \alpha_{j,k} + \sigma_{j,k}\Delta_\gamma)\,|\delta_{j,k}(x) - \alpha_{j,k}|^{p'} \right] \\ &\quad + P\,(\tilde{\alpha}_{j,k} \ge \gamma_{j,k})\,|\delta_{j,k}(\gamma_{j,k}) - \alpha_{j,k}|^{p'} \\ &\le C_T^{(u)} \left\{ \int P\,(\xi_{j,k} \ge x)\; d\left[I\,(\gamma_{j,k} \le x < \alpha_{j,k} + \sigma_{j,k}\Delta_\gamma)\,|\delta_{j,k}(x) - \alpha_{j,k}|^{p'} \right] \right. \\ &\qquad\qquad \left. + P\,(\xi_{j,k} \ge \gamma_{j,k})\,|\delta_{j,k}(\gamma_{j,k}) - \alpha_{j,k}|^{p'} \right\} \\ &= C_T^{(u)} E\,I\,(\gamma_{j,k} \le \xi_{j,k} < \alpha_{j,k} + \sigma_{j,k}\Delta_\gamma)\,|\delta_{j,k}(\xi_{j,k}) - \alpha_{j,k}|^{p'} \qquad (9.15) \end{aligned}$$

holds uniformly in $(j,k) \in \mathcal{J}^0$. Analogously we get

$$R_2 \,\leq\, C_T^{(u)} E\, I\,(\alpha_{j,k} - \sigma_{j,k}\Delta_\gamma < \xi_{j,k} < \gamma_{j,k})\, |\delta_{j,k}(\xi_{j,k}) - \alpha_{j,k}|^{p'} . \qquad (9.16)$$

Using once more (ii) of Proposition 3.1 we obtain, for arbitrary even n, that

$$E\,(\widetilde{\alpha}_{j,k} \;-\; \alpha_{j,k})^n \;=\; O\left(\sum_{r=1}^{n} \prod_{i_1,\dots,i_r : i_1+\cdots+i_r=n, i_j\geq 1} |\, cum_{i_j}(\widetilde{\alpha}_{j,k})| \right) \;=\; O(T^{-n/2}). \qquad (9.17)$$

Further we have, with $\Delta_\gamma = T^\nu$ for some $\nu > 0$, that

$$P\,(|\widetilde{\alpha}_{j,k} - \alpha_{j,k}| \geq \sigma_{j,k}\Delta_\gamma) \;\leq\; C\,(1 - \Phi(\Delta_\gamma)) \;=\; O\left(T^{-\mu}\right)$$

for arbitrary $\mu < \infty$, which implies by $|\delta_{j,k}(\widetilde{\alpha}_{j,k}) - \alpha_{j,k}| \leq |\widetilde{\alpha}_{j,k} - \alpha_{j,k}| + 2|\alpha_{j,k}|$ and the Cauchy-Schwarz inequality that

$$R_3 \;\leq\; \sqrt{P(|\widetilde{\alpha}_{j,k} - \alpha_{j,k}| \geq \sigma_{j,k}\Delta_\gamma)}\sqrt{E|\delta_{j,k}(\widetilde{\alpha}_{j,k}) - \alpha_{j,k}|^{2p'}} \;=\; O\left(T^{-p'/2-1}\right). \qquad (9.18)$$

By (9.13) and (9.15) through (9.18) we conclude that

$$E\,|\delta_{j,k}(\widetilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \;\leq\; (C_{T1} \vee C_{T2}) E\,|\delta_{j,k}(\xi_{j,k}) - \alpha_{j,k}|^{p'} \;+\; O\left(T^{-p'/2-1}\right).$$

A lower bound can be proved analogously.
(ii) Let $\theta_{j,k} \sim N(0, \sigma_T^2 - \sigma_{j,k}^2)$ be independent of $\widetilde{\alpha}_{j,k}$. Then

$$\begin{aligned}
E|\delta_{j,k}(\widetilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \;&=\; E\,I(\widetilde{\alpha}_{j,k} \geq \gamma_{j,k})|\delta_{j,k}(\widetilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \;+\; E\,I(\widetilde{\alpha}_{j,k} < \gamma_{j,k})|\delta_{j,k}(\widetilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \\
&=\; 2E\,I(\widetilde{\alpha}_{j,k} \geq \gamma_{j,k},\, \theta_{j,k} \geq 0)|\delta_{j,k}(\widetilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \;+\; 2E\,I(\widetilde{\alpha}_{j,k} < \gamma_{j,k},\, \theta_{j,k} \leq 0)|\delta_{j,k}(\widetilde{\alpha}_{j,k}) - \alpha_{j,k}|^{p'} \\
&\leq\; 2E\,|\delta_{j,k}(\widetilde{\alpha}_{j,k} + \theta_{j,k}) - \alpha_{j,k}|^{p'},
\end{aligned}$$

which yields (ii) due to Theorem 4.2. $\square$

Proof of Theorem 5.2. In view of Corollary 5.1, it remains to show that

$$\sum_{(j,k)\in\mathcal{J}} E\left(\delta^{(\cdot)}(\overline{\xi}_{j,k}, \lambda_{j,k}) \;-\; \alpha_{j,k}\right)^2 \;+\; \sum_{(j,k)\notin\mathcal{J}} \alpha_{j,k}^2, \qquad (9.19)$$

where $\overline{\xi}_{j,k}$ is given by (5.4), decays with the claimed rates.
By (4.1) we obtain for the second term of (9.19) the estimate

$$\sup_{f\in\mathcal{F}} \left\{ \sum_{(j,k)\in\mathcal{J}} \alpha_{j,k}^2 \right\} \;=\; O\left(T^{2(\alpha-1)(m+1/2-1/(p\wedge 2))}\right) \;=\; O\left(T^{-2m/(2m+1)}\right). \quad (9.20)$$

Further, we obtain by (5.10) the following estimate for the first term of (9.19):

$$\sum_{(j,k)\in\mathcal{J}} E\left(\delta^{(\cdot)}(\overline{\xi}_{j,k},\lambda_{j,k}) \;-\; \alpha_{j,k}\right)^2$$
$$\le\; C \sum_{(j,k)\in\mathcal{J}} \left((\sigma_{j,k}\vee\sigma_T)^2\left(\frac{\lambda_{j,k}}{\sigma_{j,k}\vee\sigma_T}\;+\;1\right)\varphi\left(\frac{\lambda_{j,k}}{\sigma_{j,k}\vee\sigma_T}\right)\;+\;\min\left\{\alpha_{j,k}^2,\lambda_{j,k}^2\right\}\right). \tag{9.21}$$

(i) Proof of (i)
Choose an integer j_0 such that $2^{j_0} \asymp T^{1/(2m+1)}$, and further let

$$\lambda_{j,k} \;=\; \begin{cases} 0, & \text{if } j\le j_0 \\ (\sigma_{j,k}\vee\sigma_T)\sqrt{K(j-j_0)}, & \text{if } j>j_0 \end{cases}$$

for any fixed $K>\log 4$.
Then

$$\sum_{(j,k)\in\mathcal{J}} \left((\sigma_{j,k}\vee\sigma_T)^2\left(\frac{\lambda_{j,k}}{\sigma_{j,k}\vee\sigma_T}\;+\;1\right)\varphi\left(\frac{\lambda_{j,k}}{\sigma_{j,k}\vee\sigma_T}\right)\right)$$
$$=\; O(2^{j_0}T^{-1}) \;+\; \sum_{j>j_0} O\left(T^{-1}2^j\sqrt{K(j-j_0)}\exp(-K(j-j_0)/2)\right)$$
$$=\; O(T^{-2m/(2m+1)}) \;+\; O\left(T^{-2m/(2m+1)}\sum_{j>j_0}\sqrt{K(j-j_0)}2^{(j-j_0)(1-K/\log 4)}\right)$$
$$=\; O(T^{-2m/(2m+1)}). \tag{9.22}$$

With $\widetilde{p}=\min\{p,2\}$ we obtain by Jensen's inequality

$$\left(2^{-j}\sum_k|\alpha_{j,k}|^{\widetilde{p}}\right)^{1/\widetilde{p}} \;\le\; \left(2^{-j}\sum_k|\alpha_{j,k}|^p\right)^{1/p} \;=\; O\left(2^{-j(m+1/2)}\right),$$

which implies that

$$\sum_k \min\{\alpha_{j,k}^2,\lambda_{j,k}^2\}$$
$$\le\; \lambda_{j,k}^{2-\widetilde{p}}\sum_k|\alpha_{j,k}|^{\widetilde{p}}$$
$$=\; O\left(T^{-(2-\widetilde{p})/2}(j-j_0)^{(2-\widetilde{p})/2}\;\; 2^{-j(m+1/2-1/\widetilde{p})\widetilde{p}}\right)$$
$$=\; O\left(T^{-2m/(2m+1)}(j-j_0)^{(2-\widetilde{p})/2}\;\; 2^{-(j-j_0)(m+1/2-1/\widetilde{p})\widetilde{p}}\right)$$

hold uniformly in $f\in\mathcal{F}$, $j>j_0$. Hence,

$$\sup_{f\in\mathcal{F}}\left\{\sum_{j>j_0}\sum_k\min\{\alpha_{j,k}^2,\lambda_{j,k}^2\}\right\} \;=\; O(T^{-2m/(2m+1)}), \tag{9.23}$$

which yields, in conjunction with (9.20) through (9.22), (i).

(ii) Proof of (ii)
We have, according to (A5)(i), that

$$\sum_{(j,k)\in\mathcal{J}} \left((\sigma_{j,k} \vee \sigma_T)^2 \left(\frac{\lambda_{j,k}}{\sigma_{j,k} \vee \sigma_T} + 1 \right) \varphi \left(\frac{\lambda_{j,k}}{\sigma_{j,k} \vee \sigma_T} \right) \right) = O(T^{-2m/(2m+1)}). \tag{9.24}$$

Let j_* be such that $2^{j_*} = (T/\log T)^{1/(2m+1)}$. Then we obtain, analogously to the above calculations, that

$$\sum_{j\le j_*} \sum_k \min\{\alpha_{j,k}^2, \lambda_{j,k}^2\} = O\left(2^{j_*} T^{-1} \log T\right) = O\left((\log T/T)^{2m/(2m+1)}\right) \tag{9.25}$$

and

$$\begin{aligned} &\sum_{j>j_*} \sum_k \min\{\alpha_{j,k}^2, \lambda_{j,k}^2\} \\ &= O\left((\log T/T)^{(2-\widetilde{p})/2} \sum_{j>j_*} \sum_k |\alpha_{j,k}|^{\widetilde{p}} \right) \\ &= O\left((\log T/T)^{(2-\widetilde{p})/2} \; 2^{-j_*(m+1/2-1/\widetilde{p})\widetilde{p}} \right) \\ &= O\left((\log T/T)^{(2-\widetilde{p})/2} \; 2^{j_*} (\log T/T)^{\widetilde{p}/2} \right) = O\left((\log T/T)^{2m/(2m+1)} \right), \end{aligned} \tag{9.26}$$

which prove, in conjunction with (9.20), (9.21) and (9.24), (ii). □

Proof of Lemma 6.1. We decompose the difference between $\widetilde{f}$ and f into a stochastic component, bias of the periodogram and bias due to smoothing, i.e.

$$\begin{aligned} \widetilde{f}(\omega) - f(\omega) &= \widetilde{f}(\omega) - E\widetilde{f}(\omega) \\ &\quad + \int h_T^{-1} k((\omega-\lambda)/h_T)(EI_T(\lambda) - f(\lambda))\, d\lambda \\ &\quad + \int h_T^{-1} k((\omega-\lambda)/h_T)(f(\lambda)) - f(\omega))\, d\lambda \\ &= S_1(\omega) + S_2(\omega) + S_3(\omega). \end{aligned}$$

(A) By $h_T^{-1} = O(T^{1-\delta})$ we obtain by standard calculations that $var(\widetilde{f}(\omega)) = O((Th_T)^{-1}) = O(T^{-\delta})$. Moreover, since $\widetilde{f}(\omega)$ is a weighted integral of the periodogram, we can derive, in complete analogy to (ii) and (iii) of Proposition 3.1, appropriate estimates for its cumulants. They can be used to show that there exists a sequence $c_T \to 0$ such that

$$P\left(|\widetilde{f}(\omega) - E\widetilde{f}(\omega)| > c_T \right) = O(T^{-\lambda})$$

holds for arbitrary *fixed* ω and arbitrary $\lambda < \infty$. Approximating $\widetilde{f}$ on increasingly (in T) fine grids we can further prove that

$$P\left(|\widetilde{f}(\omega) - E\widetilde{f}(\omega)| > c_T \quad \text{for all } \omega \in \Pi \right) = O(T^{-2}). \tag{9.27}$$

(B) We can show, in the same way as (i) of Proposition 3.1, that

$$S_2(\omega) \;=\; O\left(h_T^{-1} T^{-1} \log T\right) \;=\; O(T^{-\delta} \log T) \tag{9.28}$$

holds uniformly in $\omega \in \Pi$, if f has bounded total variation.

(C) (i) If $f \in B^m_{p,q}(C)$ with $m > 1/p$, then f is Hölder continuous with some positive exponent (cf. [32]). Hence, it is obvious that

$$\sup_{\omega \in \Pi} \{|S_3(\omega)|\} \;=\; o(1), \tag{9.29}$$

which completes the proof of (i).

(ii) If f has only bounded total variation, then

$$\sum_k \sup_{supp(\psi_{j,k})} \{|S_3(\omega)|\} \;\leq\; \sum_k TV(f)|_{supp(\psi_{j,k})+supp(k(./h_T))} \;=\; O\left((h_T 2^j \;+\; 1) \quad TV(f)|_{[-\pi,\pi]}\right).$$

Hence, we have, for each scale j, that

$$\#\left\{k \;\middle|\; \sup_{supp(\psi_{j,k})} \{|S_3(\omega)|\} > c_T\right\} \;=\; O\left((h_T 2^j + 1)/c_T\right),$$

which implies, for appropriate c_T, that

$$\#\left\{(j,k) \in \mathcal{J} \;\middle|\; \sup_{supp(\psi_{j,k})} \{|S_3(\omega)|\} > c_T\right\} \;=\; O\left(T^{\nu}\right), \tag{9.30}$$

since $h_T = O(T^{\nu-1})$. (ii) follows now by (9.27) and (9.28).

□

Proof of Theorem 6.1. Using the monotonicity of $\delta^{(\cdot)}(\tilde{\alpha}_{j,k},.)$ in the second argument we get, with $\lambda^{(l)}_{j,k} = \alpha_T(\sigma_{j,k} \vee \sigma_T)\sqrt{2\log(C(\#\mathcal{J}))}$ and $\lambda^{(u)}_{j,k} = \tilde{C} T^{-1/2}\sqrt{\log T}$, that

$$\left(\delta^{(\cdot)}(\tilde{\alpha}_{j,k}, \widehat{\lambda}_{j,k}) \;-\; \alpha_{j,k}\right)^2$$

$$\leq \left\{ \begin{array}{ll} (\tilde{\alpha}_{j,k} - \alpha_{j,k})^2 \;+\; (\delta^{(\cdot)}(\tilde{\alpha}_{j,k}, \lambda^{(l)}_{j,k}) \;-\; \alpha_{j,k})^2, & \text{if } \widehat{\lambda}_{j,k} < \lambda^{(l)}_{j,k} \\ (\delta^{(\cdot)}(\tilde{\alpha}_{j,k}, \lambda^{(l)}_{j,k}) \;-\; \alpha_{j,k})^2 \;+\; (\delta^{(\cdot)}(\tilde{\alpha}_{j,k}, \lambda^{(u)}_{j,k}) \;-\; \alpha_{j,k})^2, & \text{if } \lambda^{(l)}_{j,k} \leq \widehat{\lambda}_{j,k} \leq \lambda^{(u)}_{j,k} \\ (\delta^{(\cdot)}(\tilde{\alpha}_{j,k}, \lambda^{(u)}_{j,k}) \;-\; \alpha_{j,k})^2 \;+\; (\alpha_{j,k})^2, & \text{if } \widehat{\lambda}_{j,k} > \lambda^{(u)}_{j,k} \end{array} \right.$$

which implies the decomposition

$$\begin{aligned} E\|\widehat{\widehat{f}} - f\|^2 \;\le\; & \sum_k E(\tilde{\alpha}_k - \alpha_k)^2 \\ & + \sum_{(j,k)\in\mathcal{J}} E(\delta^{(\cdot)}(\tilde{\alpha}_{j,k}, \lambda_{j,k}^{(l)}) - \alpha_{j,k})^2 \\ & + \sum_{(j,k)\in\mathcal{J}} E(\delta^{(\cdot)}(\tilde{\alpha}_{j,k}, \lambda_{j,k}^{(u)}) - \alpha_{j,k})^2, \\ & + \sum_{(j,k)\in\mathcal{J}} EI(\widehat{\lambda}_{j,k} < \lambda_{j,k}^{(l)})(\tilde{\alpha}_{j,k} - \alpha_{j,k})^2 \\ & + \sum_{(j,k)\in\mathcal{J}} P(\widehat{\lambda}_{j,k} > \lambda_{j,k}^{(u)}) \;\alpha_{j,k}^2 \\ & + \sum_{(j,k)\notin\mathcal{J}} \alpha_{j,k}^2 \\ = \;& T_1 + \ldots + T_6. \end{aligned} \tag{9.31}$$

Since both thresholding schemes, $\{\lambda_{j,k}^{(l)}\}$ and $\{\lambda_{j,k}^{(u)}\}$, satisfy (A5), we obtain by (ii) of Theorem 5.2 that

$$T_1 + T_2 + T_3 + T_6 = O((\log T/T)^{2m/(2m+1)}). \tag{9.32}$$

Using Hölder's inequality we obtain by (9.17) that

$$\begin{aligned} T_4 \;\le\; & \sum_{(j,k)\in\mathcal{J}} \left(P(\widehat{\lambda}_{j,k} < \lambda_{j,k}^{(l)})\right)^{1-\delta} \left(E|\tilde{\alpha}_{j,k} - \alpha_{j,k}|^{2/\delta}\right)^{\delta} \\ = \;& O(T^{-1}) \sum_{(j,k)\in\mathcal{J}} \left(P(\widehat{\lambda}_{j,k} < \lambda_{j,k}^{(l)})\right)^{1-\delta} \\ = \;& O(T^{-1}) \left(\sum_{(j,k)\in\mathcal{J}} P(\widehat{\lambda}_{j,k} < \lambda_{j,k}^{(l)})\right)^{1-\delta} (\#\mathcal{J})^{\delta} \\ = \;& O\left(T^{-1} \;T^{\nu(1-\delta)} \;T^{\delta}\right) = O(T^{-2m/(2m+1)}) \end{aligned} \tag{9.33}$$

holds, if $0 < \delta \le (1/(2m+1) - \nu)/(1-\nu)$. Finally, we obviously have that

$$T_5 = O(T^{-2m/(2m+1)}), \tag{9.34}$$

which completes the proof. □

Acknowledgment. Part of the work on this paper was carried out while the author was visiting the Centre for Mathematics and its Applications of the Australian National University.
I thank Professor R. Dahlhaus and R. von Sachs for helpful discussions.
Further, I am indebted to an anonymous referee, whose numerous constructive remarks led to a considerably improved presentation of the results.

REFERENCES

1. Beltrão, K. I. and Bloomfield, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *J. Time Ser. Anal.* **8**, 21–38.
2. Bentkus, R. (1986). Rate of uniform convergence of statistical estimators of spectral density in spaces of differentiable functions. *Lithuanian Math. J* **26**, 209–219.
3. Bentkus, R. and Rudzkis, R. (1976). Large deviations for the spectral estimate of a stationary Gaussian sequence. *Lithuanian Math. J.* **16**, 519–529.
4. Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston: New York.
5. Coifman, R. R. and Donoho, D. L. (1995) Translation-invariant de-noising. in *Wavelets and Statistics*, Anestis Antoniadis ed., Lectures Notes in Statistics, Springer, to appear.
6. Dahlhaus, R. (1983). Spectral analysis with tapered data. *J. Time Ser. Anal.* **4**, 163–175.
7. Dahlhaus, R. (1990). Nonparametric high resolution spectral estimation. *Probab. Th. Rel. Fields* **85**, 147–180.
8. Daubechies, I. (1992). *Ten Lectures on Wavelets.* Philadelphia: Society for Industrial and Applied Mathematics.
9. Davydov, Y. A. (1970). The invariance principle for stationary processes. *Theory Probab. Appl.* **14**, 487–498.
10. Donoho, D. L. (1993). Unconditional bases are optimal bases for data compression and statistical estimation. *Appl. Comp. Harmonic Anal.* **1**, 100–115.
11. Donoho, D. L. and Johnstone, I. M. (1992). Minimax estimation via wavelet shrinkage. Technical Report No. 402, Department of Statistics, Stanford University.
12. Donoho, D. L. and Johnstone, I. M. (1993). Adapting to unknown smoothness via wavelet shrinkage. Technical Report, Department of Statistics, Stanford University.
13. Donoho, D. L. and Johnstone, I. M. (1994a). Minimax risk over l_p-balls for l_q-loss. *Probab. Th. Rel. Fields* **99**, 277–303.
14. Donoho, D. L. and Johnstone, I. M. (1994b). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.
15. Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion) *J. R. Statist. Soc., Ser. B* **57**, 301–369.
16. Franke, J. and Härdle, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist.* **20**, 121–145.
17. Gao, H.-Ye (1993). Wavelet estimation of spectral densities in time series analysis. Ph.D. dissertation. University of California, Berkeley.
18. Gasser, T. and Müller, H. G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation. Lecture Notes in Math.* **757**, 23–68.
19. Härdle, W., Klinke, S. and Turlach, B. A. (1995). *XploRe: an Interactive Statistical Computing Environment.* Springer, New York.
20. Hurvich, C. M. (1985). Data-driven choice of a spectrum estimate: Extending the applicability of cross-validation methods. *J. Amer. Statist. Assoc.* **80**, 933–940.
21. Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics. Continuous Univariate Distributions – 1, 2.* Wiley, New York.
22. Leonov, V. P. and Shiryaev, A. N. (1959). On a method of calculation of semi-invariants. *Theory Probab. Appl.* **4**, 319–329.
23. Moulin, P. (1994). Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. on Signal Processing*, **42**, 11, 3126–3136.
24. Nason, G. and Silverman, B. W. (1995) The stationary wavelet transform and some statistical applications. in *Wavelets and Statistics*, Anestis Antoniadis ed., Lectures Notes in Statistics, Springer, to appear.
25. Nemirovskii, A. S. (1985). Nonparametric estimation of smooth regression functions. *Izv. Akad.*

Nauk SSR Techn. Kibernet. **3**, 50–60 (in Russian).

26. Nemirovskii, A. S., Polyak, B. T. and Tsybakov, A. B. (1985). Rate of convergence of nonparametric estimates of maximum-likelihood type. *Problems of Information Transmission* **21**, 258–272.
27. Neumann, M. H. (1994). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series, Statistics Research Report No. SRR 028-94, CMA, Australian National University, Canberra.
28. Neumann, M. H. and Spokoiny, V. G. (1995). On the efficiency of wavelet estimators under arbitrary error distributions. *Math. Methods of Statist.* **4**, 137–166.
29. Rudzkis, R. (1978). Large deviations for estimates of spectrum of stationary series. *Lithuanian Math. J.* **18**, 214-226.
30. Rudzkis, R., Saulis, L. and Statulevicius, V. (1978). A general lemma on probabilities of large deviations. *Lithuanian Math. J.* **18**, 226–238.
31. Statulevicius, V. and Jakimavicius, D. (1988). Estimates of semiinvariants and centered moments of stochastic processes with mixing I. *Lithuanian Math. J.* **28**, 67–80.
32. Triebel, H. (1983). *Theory of Function Spaces.* Birkhäuser, Basel.
33. Wahba, G. (1980). Automatic smoothing of the log periodogram. *J. Amer. Statist. Assoc.* **75**, 122–132.
34. Wahba, G. and Wold, S. (1975). Periodic splines for spectral density estimation: The use of cross-validation for determining the degree of smoothing. *Comm. Statist.* 4, (2), 125–144.