A note on wavelet density deconvolution for weakly dependent data

Harry van Zanten · Pawel Zareba

Abstract   In this paper we investigate the performance of a linear wavelet-type deconvolution estimator for weakly dependent data. We show that the rates of convergence which are optimal in the case of i.i.d. data are also (almost) attained for strongly mixing observations, provided the mixing coefficients decay fast enough. The results are applied to a discretely observed continuous-time stochastic volatility model.

1 Introduction

In this paper we consider the problem of nonparametric deconvolution for weakly dependent data. The general setup is that we observe random variables $Y_1, \ldots, Y_n$ which can be represented as $Y_i = X_i + \varepsilon_i$, where for every fixed i, the X_i and ε_i are independent. The hidden sequence (X_i) is assumed to be marginally strictly stationary and the aim is to estimate its unknown marginal density f. The ε_i are marginally strictly stationary as well, and their density q is known to the observer.

Density deconvolution problems of this type were initially studied for the case that (X_i) and (ε_i) are independent i.i.d. sequences. This case was considered, among others, by Stefanski and Carroll (1990) and Carroll and Hall (1998). They define a kernel-type estimator and provide optimal rates of convergence for mean-square integrated error for some special cases of the noise distribution (normal, Cauchy, gamma, etc.). For the i.i.d. case with

H. van Zanten · P. Zareba(✉)
Department of Mathematics,
Faculty of Sciences, Vrije Universiteit Amsterdam,
De Boelelaan 1081a, Amsterdam 1081 HV, The Netherlands
e-mail: pzareba@cs.vu.nl

H. van Zanten
e-mail: harry@cs.vu.nl

arbitrary noise distribution, Fan (1991) gave lower bounds for the estimation of the unknown density f at a fixed point x_0 and proved that the kernel-type estimators achieve the optimal rate. This optimal rate depends on the smoothness of the density f and on the smoothness of the error distribution. The latter is expressed in terms of the tail of the characteristic function $\tilde{q}$ of the error distribution q. Suppose that for some constants $B, \beta, \gamma \geq 0$, it holds that

$$|\tilde{q}(\omega)| \sim \frac{e^{-B|\omega|^\beta}}{(1+\omega^2)^{\gamma/2}}. \tag{1}$$

If $B = 0$, i.e. if $\tilde{q}$ has power tails, we say the noise is "ordinary smooth". The case $B > 0$ is called the "supersmooth" case. Fan (1991) proved that the optimal rate in the supersmooth case is $(\log n)^{-\alpha/\beta}$, where α is the order of smoothness of the unknown density f. In the ordinary smooth case the optimal rate is $n^{-\alpha/(1+2\alpha+2\gamma)}$. Moreover, the kernel estimator is adaptive in the supersmooth case, in the sense that it does not depend on the unknown smoothness level α of f. In the ordinary smooth case however, the optimal bandwidth does depend on α.

An alternative estimation method was proposed in the paper (Pensky and Vidakovic 1999), using wavelet methods instead of kernel estimators. First a linear wavelet estimator was constructed and it was shown that it has the same properties as the kernel estimators. The optimal rates are achieved, and the estimator is adaptive in the supersmooth case. The only difference is that global L^2-errors were considered instead of pointwise errors. The main contribution of Pensky and Vidakovic (1999) was the construction of a nonlinear wavelet estimator which was shown to be adaptive in the ordinary smooth case as well. We also refer to Pensky and Vidakovic (1999) for a discussion of the estimation in the case of a supersmooth density f. The optimality of the pointwise rates of such estimators was recently proved by Butucea (2004). It is also worth mentioning that Blanke and Pumo (2003) consider the optimal choice of the data collecting scheme depending on the features of the bivariate density $f_{(Y_i, Y_j)}$. See also the papers Fan (2002) and Walter (1999) and the references therein for related work on wavelet deconvolution.

Obviously, the i.i.d. assumptions on the data (X_i) and the errors (ε_i) is often to stringent in applications. Observations from popular non-linear GARCH-type time series models for instance, are typically of the form

$$S_i = \sigma_i Z_i,$$

where $Z_1, Z_2, \ldots$ are i.i.d., and the process (σ_i^2) is strictly stationary and predictable (see for instance Carrasco and Chen (2002) for an overview of models of this type). Squaring and taking logarithms puts the equation in the form $Y_i = X_i + \varepsilon_i$, where $Y_i = \log S_i^2$, $X_i = \log \sigma_i^2$, and $\varepsilon_i = \log Z_i^2$. Since (σ_i^2) is usually only known to be predictable and stationary, the sequence (X_i) will typically not be i.i.d., and X_i and ε_i will only be independent for each fixed i, not as sequences. Hence, the problem of nonparametric estimation of the volatility density in a GARCH-type model transforms into a deconvolution problem for dependent data, which does not fall into the classical i.i.d. framework.

A second example arises from observing discrete-time data $S_0, S_\Delta, S_{2\Delta}, \ldots$ from a continuous-time stochastic volatility model of the form

$$dS_t = \sigma_t \, dW_t, \tag{2}$$

where W is a Brownian motion and σ^2 is a strictly stationary, predictable process which is independent of the driving noise process W. The transformed increments $\log(S_{i\Delta} - S_{(i-1)\Delta})^2$ are then distributed as $Y_i = X_i + \varepsilon_i$, where

$$X_i = \log \int_{(i-1)\Delta}^{i\Delta} \sigma_u^2 \, du, \quad \varepsilon_i = \log Z_i^2,$$

and Z_i is an i.i.d. sequence of standard normal random variables, independent of σ. Hence, the problem of estimation of the density of the aggregated volatility $\int_0^\Delta \sigma_u^2 \, du$ reduces to a deconvolution problem for dependent data as well.

Motivated by these examples we study how the performance of deconvolution estimators initially proposed for i.i.d. data changes if the observations are not i.i.d., but weakly dependent. We consider the linear wavelet-type estimators of Pensky and Vidakovic (1999) and focus on strongly mixing data, since strong mixing conditions are satisfied for many popular GARCH-type and stochastic volatility models (cf. e.g. Carrasco and Chen (2002)). Our main results roughly state that if the data are strictly stationary and strongly mixing with mixing coefficients that vanish "faster than polynomially", then the linear wavelet estimator (almost) achieves the optimal rates of convergence of Fan (1991).

Let us mention that Masry (1991) studied the properties of the deconvolution kernel estimators under various mixing assumptions. In particular, he showed that under a uniform mixing condition, certain kernel estimators still achieve the optimal i.i.d. rates of Fan (1991) in the case that f has two bounded derivatives. Masry did not study the effect of different smoothness levels of the unknown density f on the rate of convergence.

To illustrate the main results we apply them to the model (2). The results we find in this case extend and complement results from the papers Van Es et al. (2003, 2004). In the first of these papers the same model is studied, but it is assumed that the time Δ between the observations vanishes as the number of observations grows. In the present paper we keep Δ fixed (low frequency data). The main difference with Van Es et al. (2004) is that in that paper, the smoothness level of the unknown density is fixed. In the present paper we investigate the dependence of the rate of convergence on the level of smoothness.

The organization of the paper is as follows. In the next section we present the construction of the linear wavelet deconvolution estimator, following Pensky and Vidakovic (1999). In Sect. 3 we introduce the weak dependence assumptions on the observations, the smoothness assumptions on the unknown density and we present the main results. The application to the model (2) is given in Sect. 4. The proofs of the main results are collected in Sect. 5.

2 Construction of the estimator

In this section we recall the construction of the wavelet estimator proposed in Pensky and Vidakovic (1999). For the necessary background on wavelet theory, see for instance Jawerth and Sweldens (1994), Blatter (1998), and the references therein.

For the construction of deconvolution estimators we need to use band-limited wavelets. As in Pensky and Vidakovic (1999) we use a Meyer-type wavelet (see also Walter (1994), Walter and Zayed (1996)). We consider an orthogonal scaling function and wavelet φ and ψ, respectively, associated with an orthogonal multiresolution analysis of $L^2(\mathbb{R})$, and suppose that for a symmetric probability measure μ with support contained in $[-\pi/3, \pi/3]$ it holds that

$$\tilde{\varphi}(\omega) = (\mu(\omega - \pi, \omega + \pi])^{1/2}, \quad \tilde{\psi}(\omega) = e^{-i\omega/2} \left(\mu(|\omega|/2 - \pi, |\omega| - \pi]\right)^{1/2}.$$

Here, and elsewhere in the paper, we denote the Fourier transform of a function g by $\tilde{g}$, i.e.

$$\tilde{g}(\omega) = \int_{\mathbb{R}} e^{-i\omega x} g(x) dx.$$

Observe that the assumptions imply that ϕ and ψ are indeed band-limited. For the supports of their Fourier transforms we have supp $\tilde{\varphi} \subset [-4\pi/3, 4\pi/3]$ and supp $\tilde{\psi} \subset [-8\pi/3, -2\pi/3] \cup [2\pi/3, 8\pi/3]$. By choosing μ smooth enough we ensure that $\tilde{\varphi}$ and $\tilde{\psi}$ are at least twice continuously differentiable.

For any integer m, the unknown density f can now be written as

$$f(x) = \sum_{k \in \mathbb{Z}} a_{m,k} \varphi_{m,k}(x) + \sum_{k \in \mathbb{Z}} \sum_{j=m}^{\infty} b_{j,k} \psi_{j,k}(x), \tag{3}$$

where $\varphi_{m,k}(x) = 2^{m/2} \varphi(2^m x - k)$, $\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$ and the coefficients are given by

$$a_{m,k} = \int_{\mathbb{R}} \varphi_{m,k}(x) f(x) \, dx, \qquad b_{j,k} = \int_{\mathbb{R}} \psi_{j,k}(x) f(x) \, dx.$$

The idea behind the linear wavelet estimator is simple. We first approximate f by the orthogonal projection given by the first term on the right-hand side of (3). For m large enough the second term will be small, and can be controlled by using the approximation properties of the specific family of wavelets that is being used. The projection of f is estimated by replacing the coefficients $a_{m,k}$ by consistent estimators and truncating the sum. Using the fact that the density p of an observation Y_i is the convolution of f and q it is easily verified that

$$a_{m,k} = \int_{\mathbb{R}} 2^{m/2} U_m(2^m x - k) p(x) \, dx = 2^{m/2} \mathbb{E} U_m(2^m Y_i - k),$$

where U_m is the function with Fourier transform

$$\tilde{U}_m(\omega) = \frac{\tilde{\varphi}(\omega)}{\tilde{q}(-2^m \omega)}. \tag{4}$$

To justify the above definition we assume throughout the paper that $|\tilde{q}(\omega)| > 0$ for almost all ω. We estimate the coefficient $a_{m,k}$ by its empirical counterpart

$$\hat{a}_{m,k,n} = \frac{1}{n} \sum_{l=1}^{n} 2^{m/2} U_m(2^m Y_l - k).$$

Under the mixing assumptions that we will impose on the sequence Y it will be strictly stationary and ergodic. Hence, by the ergodic theorem, $\hat{a}_{m,k,n}$ is a consistent estimator for $a_{m,k}$. The wavelet estimator is now defined by

$$\hat{f}_n(x) = \sum_{|k| \leq K_n} \hat{a}_{m_n,k,n} \varphi_{m_n,k}(x), \tag{5}$$

where the detail level m_n and the truncation point K_n will be chosen appropriately later.

It is easy to see that the bias of this estimator is the same as in the i.i.d. no-noise case. This is a well-known phenomenon for a kernel-type estimators.

3 Main results

3.1 Weak dependence assumptions

If (X_i) and (ε_i) are assumed to be independent i.i.d. sequences, the observations $Y_i = X_i + \varepsilon_i$ are of course i.i.d. as well. In the present paper we relax this requirement. We only assume that the sequence $Y = (Y_i)$ is strictly stationary and strongly mixing. The strong mixing coefficients α_k of Y are defined by

$$\alpha_k = \sup_{\substack{A \in \mathcal{F}^Y_{-\infty,0} \\ B \in \mathcal{F}^Y_{k,\infty}}} |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|,$$

where $\mathcal{F}^Y_{-\infty,a}$ is the σ-algebra generated by the random variables $\ldots, Y_{a-1}, Y_a$ and, similarly, $\mathcal{F}^Y_{a,\infty}$ is the σ-algebra generated by the random variables $Y_a, Y_{a+1}, \ldots$. The process Y is said to be strongly mixing if $\alpha_k \to 0$ as $k \to \infty$. In the main results of this paper we assume that mixing coefficients converge fast enough to guarantee that $\sum \alpha_k^p < \infty$ for some (or all) $p \in (0, 1)$. This holds in particular (for *all* $p \in (0, 1)$) if $\alpha_k \leq c\rho^k$ for some $c > 0$ and $\rho \in (0, 1)$. We refer to e.g. (Doukhan 1994) for the precise relations with other mixing conditions and examples. Let us just mention here that strong mixing is weaker than β-mixing and that there exist numerous results giving easily verifiable conditions under which Markov chains, GARCH-type models, stochastic volatility models and discretely observed diffusions are β-mixing with geometrically decreasing coefficients. See for instance Carrasco and Chen (2002), Genon-Catalot et al. (2002), Meyn and Tweedie (1993).

3.2 Smoothness assumptions on the unknown density

As in Pensky and Vidakovic (1999), the smoothness of the unknown marginal density f of the sequence X is controlled by assuming that it belongs to a Sobolev ball of a certain order. The Sobolev space H^α is defined for $\alpha > 0$ by

$$H^\alpha = \left\{ g : \|g\|_\alpha = \left(\int_{\mathbb{R}} |\tilde{g}(\omega)|^2 (\omega^2 + 1)^\alpha d\omega \right)^{1/2} < \infty \right\}.$$

Roughly speaking, $g \in H^\alpha$ means that the first α derivatives of g belong to $L^2(\mathbb{R})$. The Sobolev ball of radius A is defined by

$$\mathscr{S}_\alpha(A) = \left\{ g \in H^\alpha : \|g\|_\alpha \leq A \right\}.$$

For technical reasons we also have to restrict the rate of decay of the unknown density. We will always assume that f belongs to a space of the form

$$\mathscr{S}^*_\alpha(A, A') = \mathscr{S}_\alpha(A) \cap \left\{ g : \sup_x |x| g(x) \leq A' \right\}.$$

Of course, the additional requirement that $f(x)$ should decay at least like $1/x$ is not restrictive at all.

3.3 Main results

The main results of the paper are upper bounds for the mean integrated square error of the wavelet estimator $\hat{f}_n$, which is defined as usual by

$$\text{MISE}\left(\hat{f}_n\right) = \mathbb{E} \int_{\mathbb{R}} \left(\hat{f}_n(x) - f(x)\right)^2 dx.$$

We will specify how to choose the detail level m_n and the truncation point K_n in (5) optimally in different cases (depending on the smoothness of f and q) and give the resulting bounds for

$$\sup_{f \in \mathscr{S}_\alpha^*(A,A')} \text{MISE}\left(\hat{f}_n\right).$$

To quantify the smoothness of the noise distribution we assume that there exist nonnegative constants a, β, γ, B such that for $|\omega|$ large enough,

$$|\tilde{q}(\omega)| \geq a \left(\omega^2 + 1\right)^{-\gamma/2} e^{-B|\omega|^\beta}. \tag{6}$$

Recall that if $B > 0$ we say we are in the supersmooth case, and the case $B = 0$ is called the ordinary smooth case.

In the ordinary smooth case the main result is the following.

Theorem 3.1 *Suppose (Y_i) is strictly stationary and strongly mixing with mixing coefficients (α_k) satisfying $\sum_{k \geq 0} \alpha_k^p < \infty$ for some $p \in [0, 1)$, and that (6) holds with $B = 0$. Then with the choices*

$$2^{m_n} = n^{\frac{1}{1+2\alpha+2\gamma+p(1+2\alpha)}}, \quad K_n = n^{\frac{1+2\alpha}{1+2\alpha+2\gamma+p(1+2\alpha)}}$$

it holds that

$$\sup_{f \in \mathscr{S}_\alpha^*(A,A')} \text{MISE}\left(\hat{f}_n\right) = O\left(n^{-\frac{2\alpha}{1+2\alpha+2\gamma+p(1+2\alpha)}}\right).$$

for all α, A, $A' > 0$.

Note that the mixing condition of the theorem is satisfied for $p = 0$ if and only if (Y_i) is m-dependent for some m. In that case the estimator attains the optimal rate of Fan (1991). In particular, we recover the fact that the wavelet estimator is optimal if (Y_i) is i.i.d., cf. Pensky and Vidakovic (1999).

The optimal choices of m_n and K_n in the preceding theorem depend on p. It is desirable to know if these numbers can also be chosen independently of p in a sensible way. The following result says that if $\sum_{k \geq 0} \alpha_k^p < \infty$ for all $p \in (0, 1)$, the optimal choices for m_n and K_n for the case $p = 0$ provide an estimator which "almost" attains the optimal rate.

Theorem 3.2 *Suppose (Y_i) is strictly stationary and strongly mixing with mixing coefficients (α_k) satisfying $\sum_{k \geq 0} \alpha_k^p < \infty$ for all $p \in (0, 1)$, and that (6) holds with $B = 0$. Then with the choices*

$$2^{m_n} = n^{\frac{1}{1+2\alpha+2\gamma}}, \quad K_n = n^{\frac{1+2\alpha}{1+2\alpha+2\gamma}}$$

it holds that

$$\sup_{f \in \mathscr{S}_\alpha^*(A, A')} \text{MISE}\left(\hat{f}_n\right) = O\left(n^{\frac{-2\alpha+\varepsilon}{1+2\alpha+2\gamma}}\right).$$

for all $\varepsilon, \alpha, A, A' > 0$.

Let us remark that in Pensky and Vidakovic (1999) it is proposed to take the truncation point K_n such that $n/K_n \to 0$ as $n \to \infty$. The preceding results show that this is not necessary for obtaining the (almost) optimal rate of convergence.

We now give the main results for the supersmooth case.

Theorem 3.3 *Suppose (Y_i) is strictly stationary and strongly mixing with mixing coefficients (α_k) satisfying $\sum_{k \geq 0} \alpha_k^p < \infty$ for some $p \in [0, 1)$, and that (6) holds with $B > 0$. Then with the choices*

$$2^{m_n} = \left(\frac{\log n}{1 + 2B(4\pi/3)^\beta}\right)^{1/\beta}, \quad K_n = (\log n)^r, \quad r \geq (1 + 2\alpha)/\beta$$

it holds that

$$\sup_{f \in \mathscr{S}_\alpha^*(A, A')} \text{MISE}\left(\hat{f}_n\right) = O\left((\log n)^{-\frac{2\alpha}{\beta}}\right).$$

for all $\alpha, A, A' > 0$.

So in the super smooth case the wavelet estimator attains the optimal i.i.d. rate (cf. (Fan 1991)) if the strong mixing condition is satisfied for some $p > 0$. Moreover, the optimal choices of m_n and K_n do not depend on the parameter p. Unfortunately, the estimator loses the complete adaptivity to the smoothness level α which it has in the i.i.d. case. We can now only get adaptivity if we are given an upper bound $\overline{\alpha}$ for α a priori, by taking $r = (1 + 2\overline{\alpha})/2$.

Complete adaptivity with respect to α can be regained if we assume more regarding the decay of mixing coefficients. The following theorem covers many cases of practical interest.

Theorem 3.4 *Suppose (Y_i) is strictly stationary and strongly mixing with mixing coefficients (α_k) satisfying $\sum_{k \geq 0} \alpha_k^p < \infty$ for all $p \in (0, 1)$, and that (6) holds with $B > 0$. Then with the choices*

$$2^{m_n} = \left(\frac{\log n}{1 + 2B(4\pi/3)^\beta}\right)^{1/\beta}, \quad K_n = n$$

it holds that

$$\sup_{f \in \mathscr{S}_\alpha^*(A, A')} \text{MISE}\left(\hat{f}_n\right) = O\left((\log n)^{-\frac{2\alpha}{\beta}}\right).$$

for all $\alpha, A, A' > 0$.

Note that this theorem includes the i.i.d. result of Pensky and Vidakovic (1999), but as before, we see that K_n can be chosen smaller than proposed in the latter paper.

The following tables summarize the most useful results, which are valid under the assumption that the observation process (Y_i) is strongly mixing with mixing coefficients satisfying

$\sum \alpha_k^p < \infty$ for all $p \in (0, 1)$. In particular, the results apply if the coefficients decrease geometrically.

Supersmooth case:

choice of m_n	choice of K_n	MISE
$2^{m_n} = \left(\frac{\log n}{1+2B(4\pi/3)^\beta} \right)^{1/\beta}$	$K_n = n$	$O\left((\log n)^{-2\alpha/\beta} \right)$

Ordinary smooth case:

choice of m_n	choice of K_n	MISE
$2^{m_n} = n^{\frac{1}{1+2\alpha+2\gamma}}$	$K_n = n^{\frac{1+2\alpha}{1+2\alpha+2\gamma}}$	$O\left(n^{\frac{-2\alpha+\varepsilon}{1+2\alpha+2\gamma}} \right)$ for all $\varepsilon > 0$

So in the super smooth case, the wavelet deconvolution estimator attains the optimal i.i.d. rate (cf. (Fan 1991; Pensky and Vidakovic 1999)) as soon as the data are strongly mixing with mixing coefficients that vanish "faster than polynomially". Moreover, it has the same adaptivity property as in the i.i.d. case: the optimal m_n and K_n do not depend on the unknown smoothness level α. Note also that we take $K_n = n$, whereas Pensky and Vidakovic (1999) propose to take K_n such that $n/K_n \to 0$. In practice, taking a smaller K_n reduces the computational effort.

In the ordinary smooth case we "almost" attain the optimal rate under the same mixing condition. The linear wavelet estimator is not adaptive with respect to α in this case. Adaptivity can perhaps be achieved by using a non-linear wavelet estimator instead. This approach works in the i.i.d. case, but the proof of this fact uses Talagrand's inequality (see (Talagrand 1994)). It is not immediately clear how to extend this to the weakly dependent case.

4 Example: discretely observed stochastic volatility model

Suppose we have discrete-time data $S_0, S_\Delta, S_{2\Delta}, \ldots$ from a continuous-time stochastic volatility model of the form

$$dS_t = \sigma_t \, dW_t,$$

where W is a Brownian motion and σ^2 is a strictly stationary, predictable process which is independent of the driving noise process W. The time $\Delta > 0$ between the observations is fixed. As already mentioned in the introduction, the transformed increments $\log(S_{i\Delta} - S_{(i-1)\Delta})^2$ are then distributed as $Y_i = X_i + \varepsilon_i$, where

$$X_i = \log \int_{(i-1)\Delta}^{i\Delta} \sigma_u^2 \, du, \quad \varepsilon_i = \log Z_i^2,$$

and Z_i is an i.i.d. sequence of standard Gaussian random variables, independent of σ. The sequence X_i is strictly stationary and we assume that its marginal density f exists, i.e. f is

the density of

$$\log \int_0^\Delta \sigma_u^2 \, du.$$

Of course, estimating f is equivalent with estimating the density of the aggregated volatility $\int_0^\Delta \sigma_u^2 \, du$.

We have the following result for the wavelet density estimator $\hat{f}_n$ of f defined by (5).

Theorem 4.1 *Suppose that the volatility process σ^2 is strongly mixing with mixing coefficients satisfying*

$$\sum_{k \geq 0} \alpha_{k\Delta}^p < \infty \tag{7}$$

for some $p \in (0, 1)$. Then with the choices

$$2^{m_n} = \frac{\log n}{1 + (4\pi^2/3)}, \quad K_n = (\log n)^r, \quad r \geq 1 + 2\alpha$$

the mean square error of the wavelet estimator satisfies

$$\sup_{f \in \mathscr{S}_\alpha^*(A, A')} \mathrm{MISE}\left(\hat{f}_n\right) = O\left((\log n)^{-2\alpha}\right)$$

for $\alpha, A, A' > 0$. If (7) is satisfied for all $p \in (0, 1)$, the same bound is true if the choice for K_n is replaced by $K_n = n$.

Before we give the proof of this theorem, let us point out the relation with the results of Van Es et al. (2004). In the latter paper kernel-type deconvolution estimators for stochastic volatility models were considered. When applied to the present model the results say that under the same mixing condition and assuming that f has two bounded and continuous derivatives, the (pointwise) mean square error of the kernel estimator is of order $(\log n)^{-4}$. The analogue of f having two bounded derivatives in our setting is that $f \in \mathscr{S}_2^*(A, A')$ for some $A, A' > 0$. Indeed, the theorem yields the same bound $(\log n)^{-4}$ for the MISE in this case. Theorem 4.1 is more general because it allows for different smoothness levels as well. In particular, the density f is allowed to be only smooth of some order $\alpha < 2$. Moreover, the estimator is adaptive if the condition on the mixing coefficients holds for all $p \in (0, 1)$.

Proof In the present setting the characteristic function $\tilde{q}$ of ε_i satisfies

$$|\tilde{q}(\omega)| \sim \sqrt{2} e^{-\frac{1}{2}\pi|\omega|}$$

for $|\omega| \to \infty$, see Lemma 5.1 of Van Es et al. (2004). It follows that (6) is satisfied with $\gamma = 0$, $\beta = 1$ and $B = \pi/2$ (supersmooth case).

From the definition of the mixing coefficients it is easily seen that the process X is strongly mixing, and we have the relation $\alpha_k^X \leq \alpha_{(k-1)\Delta}^{\sigma^2}$ between the mixing coefficients of the processes X and σ^2. Since σ and Z are independent by assumption the processes X and ε are independent as well, and hence we have $\alpha_k^Y \leq \alpha_k^X$ (this is seen by conditioning on ε). It follows that

$$\sum_k (\alpha_k^Y)^p \leq \sum_k \left(\alpha_{(k-1)\Delta}^{\sigma^2}\right)^p.$$

The result now follows from Theorems 3.3 and 3.4.   □

5 Proofs for Section 3

The smoothness of the error distribution enters the basic bound for the MISE via the quantity

$$\Delta(m) = \int_{\mathbb{R}} \frac{|\tilde{\varphi}(\omega)|^2}{|\tilde{q}(2^m\omega)|^2}\, d\omega.$$

The following lemma provides the basic upper bound for the MISE. The symbol $\lesssim$ between two expressions means that the left-hand side is less than or equal to the right-hand side, up to a multiplicative constant independent of n.

Lemma 5.1 *Suppose (Y_i) is strictly stationary and strongly mixing with mixing coefficients (α_k). Then for $p \in [0, 1)$ and $\alpha, A, A' > 0$ it holds that*

$$\sup_{f \in \mathscr{S}_\alpha^*(A, A')} \mathrm{MISE}\left(\hat{f}_n\right) \lesssim \frac{2^{m_n}}{n}\Delta(m_n)\left(1 + K_n^p \sum_{k=1}^{n} \alpha_k^p\right) + \frac{2^{m_n}}{K_n} + \frac{1}{4^{m_n\alpha}}.$$

Proof To simplify the notation somewhat we suppress the dependence of m_n and K_n on n in the notation.

By construction of the estimator we have

$$f(x) - \hat{f}_n(x) = \sum_{|k|\leq K}(a_{m,k} - \hat{a}_{m,k,n})\varphi_{m,k}(x) + \sum_{|k|>K} a_{m,k}\varphi_{m,k}(x)$$
$$+ \sum_{k\in\mathbb{Z}}\sum_{j=m}^{\infty} b_{j,k}\psi_{j,k}(x).$$

Hence, since we use orthogonal wavelets and $\hat{a}_{m,k,n}$ is an unbiased estimator for $a_{m,k}$,

$$\mathrm{MISE}(\hat{f}_n) = \sum_{|k|\leq K}\mathbb{V}\mathrm{ar}\,\hat{a}_{m,k,n} + \sum_{|k|>K}|a_{m,k}|^2 + \sum_{k\in\mathbb{Z}}\sum_{j=m}^{\infty}|b_{j,k}|^2$$
$$=: I + II + III.$$

Estimate for I:

We put $h_{k,m}(x) = U_m(2^m x - k)$ and write

$$\sum_{|k|\leq K}\mathbb{V}\mathrm{ar}\,\hat{a}_{m,k,n} = \sum_{|k|\leq K}\mathbb{V}\mathrm{ar}\left(\frac{1}{n}\sum_{l=1}^{n}2^{m/2}h_{k,m}(Y_l)\right)$$
$$= \frac{2^m}{n}\sum_{|k|\leq K}\mathbb{V}\mathrm{ar}\,h_{k,m}(Y_0) + \frac{2^{m+1}}{n^2}\sum_{|k|\leq K}\sum_{\substack{i,j=1 \\ i<j}}^{n}\mathbb{C}\mathrm{ov}\,(h_{k,m}(Y_i), h_{k,m}(Y_j))$$
$$=: I_1 + I_2.$$

Arguing as in the proof of Lemma 3 of Pensky and Vidakovic (1999) it is easily shown that

$$\sup_{x\in\mathbb{R}}\left(\sum_{k\in\mathbb{Z}}|h_{k,m}(x)|^2\right) \lesssim \Delta(m).$$

It follows that with p the marginal density of the observations Y_i, we have

$$\sum_{k \in \mathbb{Z}} \mathbb{V}\text{ar}\, h_{k,m}(Y_0) \le \sum_{k \in \mathbb{Z}} \mathbb{E}\,|h_{k,m}(Y_0)|^2 = \int_{\mathbb{R}} \sum_{k \in \mathbb{Z}} |h_{k,m}(x)|^2 \, p(x)\, dx$$

$$\le \sup_{x \in \mathbb{R}} \left(\sum_{k \in \mathbb{Z}} |h_{k,m}(x)|^2 \right) \lesssim \Delta(m). \tag{8}$$

Hence, $I_1 \lesssim \Delta(m)2^m/n$. Next, note that by stationarity,

$$\left| \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathbb{C}\text{ov}(h_{k,m}(Y_i), h_{k,m}(Y_j)) \right| = \left| \sum_{l=1}^{n} (n-l)\mathbb{C}\text{ov}(h_{k,m}(Y_0), h_{k,m}(Y_l)) \right|$$

$$\le n \sum_{l=1}^{n} |\mathbb{C}\text{ov}(h_{k,m}(Y_0), h_{k,m}(Y_l))|.$$

By a well-known covariance inequality for strongly mixing processes (cf. Deo (1973)) it holds that

$$|\mathbb{C}\text{ov}(h_{k,m}(Y_0), h_{k,m}(Y_l))| \le 8\alpha_l^p \left(\mathbb{E}\,|h_{k,m}(Y_0)|^{2/(1-p)} \right)^{1-p}$$

$$\le 8\alpha_l^p \|U_m\|_\infty^{2p} \left(\mathbb{E}\,|h_{k,m}(Y_0)|^2 \right)^{1-p},$$

so that

$$I_2 \le \frac{2^{m+4}}{n} \|U_m\|_\infty^{2p} \left(\sum_{l=1}^{n} \alpha_l^p \right) \sum_{|k| \le K} \left(\mathbb{E}\,|h_{k,m}(Y_0)|^2 \right)^{1-p}.$$

By Hölder's inequality and the bound in (8) we have

$$\sum_{|k| \le K} \left(\mathbb{E}\,|h_{k,m}(Y_0)|^2 \right)^{1-p} \le (2K+1)^p \left(\sum_{|k| \le K} \mathbb{E}\,|h_{k,m}(Y_0)|^2 \right)^{1-p}$$

$$\lesssim K^p (\Delta(m))^{1-p},$$

and hence

$$I_2 \lesssim \frac{2^m}{n} K^p (\Delta(m))^{1-p} \|U_m\|_\infty^{2p} \sum_{l=1}^{n} \alpha_l^p.$$

Using Fourier inversion and Hölder's inequality it is easy to see that we have $\|U_m\|_\infty^2 \lesssim \Delta(m)$. Thus, for $I = I_1 + I_2$ we obtain the bound

$$I \lesssim \frac{2^m}{n} \Delta(m) \left(1 + K^p \sum_{l=1}^{n} \alpha_l^p \right).$$

Estimate for II:

Under the conditions that we have imposed it holds that

$$II \lesssim \frac{2^m}{K}$$

for all $f \in \mathcal{S}_\alpha^*(A, A')$. See the proof of Theorem 3 of Pensky and Vidakovic (1999).

Estimate for III:

By Lemma 2 of Pensky and Vidakovic (1999) it holds that

$$III \leq \frac{2}{\pi} \|\tilde{\psi}\|_\infty^2 \left(\frac{2\pi}{3}\right)^{-2\alpha} A^2 2^{-2m\alpha}$$

for all $f \in \mathcal{S}_\alpha(A)$.   □

Since we want the MISE to vanish as $n \to \infty$, we will choose m_n and K_n in such a way that they tend to infinity as $n \to \infty$. In that case the bound simplifies to

$$\sup_{f \in \mathscr{S}_\alpha^*(A,A')} \mathrm{MISE}\left(\hat{f}_n\right) \lesssim \frac{2^{m_n}}{n} \Delta(m_n) K_n^p \sum_{k=0}^n \alpha_k^p + \frac{2^{m_n}}{K_n} + \frac{1}{4^{m_n\alpha}}.$$

In all main theorems we assume that $\sum_k \alpha_k^p < \infty$ for (some or all) $p \in [0, 1)$. In that case the bound simplifies further to

$$\sup_{f \in \mathscr{S}_\alpha^*(A,A')} \mathrm{MISE}\left(\hat{f}_n\right) \lesssim \frac{2^{m_n}}{n} \Delta(m_n) K_n^p + \frac{2^{m_n}}{K_n} + \frac{1}{4^{m_n\alpha}}.$$

The optimal choice of m_n and K_n depends of course on the asymptotic behaviour of $\Delta(m_n)$. It is easily seen that under the assumption (6) on the tails of $\tilde{q}$, we have the bound

$$\Delta(m) \lesssim 4^{\gamma m} \exp(2B(4\pi/3)^\beta 2^{m\beta})$$

(see also Pensky and Vidakovic (1999)).

We can now prove the theorems in Sect. 3.

Proof of Theorems 3.1 and 3.2 In this case $B = 0$, so $\Delta(m) \lesssim 4^{\gamma m}$. For the MISE we obtain the bound

$$\sup_{f \in \mathscr{S}_\alpha^*(A,A')} \mathrm{MISE}\left(\hat{f}_n\right) \lesssim \frac{L_n^{1+2\gamma}}{n} K_n^p + \frac{L_n}{K_n} + \frac{1}{L_n^{2\alpha}}, \quad L_n = 2^{m_n}.$$

The last two terms are balanced in we take $K_n = L_n^{1+2\alpha}$. To balance the first term and the last two we then have to choose L_n such that $L_n^{1+2\gamma+p+2\alpha p}/n = 1/L_n^{2\alpha}$, i.e.

$$L_n = n^{\frac{1}{1+2\alpha+2\gamma+p(1+2\alpha)}}.$$

This leads to the bound

$$n^{\frac{-2\alpha}{1+2\alpha+2\gamma+p(1+2\alpha)}}$$

for the MISE, proving Theorem 3.1.

The proof of Theorem 3.2 is a straightforward adaptation of the preceding one.   □

Proof of Theorems 3.3 and 3.4 For the proof of Theorem 3.3 note that in this case we have $B > 0$, which implies that $\Delta(m) \lesssim \exp(2B(4\pi/3)^\beta 2^{m\beta})$, whence

$$\sup_{f \in \mathscr{S}_\alpha^*(A,A')} \mathrm{MISE}\left(\hat{f}_n\right) \lesssim \frac{K_n^p}{n} e^{2B(4\pi/3)^\beta L_n^\beta} + \frac{L_n}{K_n} + \frac{1}{L_n^{2\alpha}}, \quad L_n = 2^{m_n}.$$

Again, the last two terms are balanced if we take $K_n = L_n^{1+2\alpha}$. With $L_n^\beta = \log n/(1 + 2B(4\pi/3)^\beta)$ the first term on the right-hand side is then bounded by a negative power of n

and $L_n^{-2\alpha} \lesssim (\log n)^{-2\alpha/\beta}$. The same rate is obtained if we let K_n tend to infinity faster than $L_n^{1+2\alpha}$, but still logarithmically. This proves Theorem 3.3. Theorem 3.4 is proved in the same manner. □

References

Blanke D, Pumo B (2003) Optimal sampling for density estimation in continuous time. J Time Ser Anal 24(1): 1–23

Blatter C (1998) Wavelets, a primer. A K Peters Ltd

Butucea C (2004) Deconvolution of supersmooth densities with smooth noise. Canad J Stat 32(2): 181–192

Carrasco M, Chen X (2002) Mixing and moment properties of various GARCH and stochastic volatility models. Economet Theory 18: 17–39

Carroll RJ, Hall P (1988) Optimal rates of convergence for deconvolving a density. J Amer Stat Assoc 83(404): 1184–1186

Deo CM (1973) A note on empirical processes of strong-mixing sequences. Ann Probability 1(5): 870–875

Doukhan P (1994) Mixing: properties and examples. Springer

Fan J (1991) On the optimal rates of convergence of nonparametric deconvolution problems. Ann Stat 19: 1257–1272

Fan J, Koo J (2002) Wavelet deconvolution. IEEE Trans Inform Theory 48(3): 734–747

Genon-Catalot V, Jeantheau T, Larédo C (2000) Stochastic volatility models as hidden Markov models and statistical applications. Bernoulli 6(6): 1051–1079

Jawerth B, Sweldens W (1994) An overview of wavelet based multiresolution analyses. SIAM Rev 36: 377–412

Masry E (1991) Multivariate probability density deconvolution for stationary random processes. IEEE Inf Th 37: 1105–1115

Meyn SP, Tweedie RL (1993) Markov Chains and Stochastic Stability. Springer

Pensky M, Vidakovic B (1999) Adaptive wavelet estimator for nonparametric density deconvolution. Ann Stat 27: 2033–2053

Stefanski L, Carroll RJ (1990) Deconvoluting kernel density estimators. Statistics 21(2): 169–184

Van Es AJ, Spreij PJC, Van Zanten JH (2003) Nonparametric volatility density estimation. Bernoulli 9(3): 451–465

Van Es AJ, Spreij PJC, Van Zanten JH (2004) Nonparametric volatility density estimation for discrete time models. J. Nonparam. Stat (To appear)

Walter GG (1994) Wavelets and other orthogonal systems with applications. CRC Press

Walter GG (1999) Density estimation in the presence of noise. Stat. Probab Lett 41: 237–246

Walter GG, Zayed AI (1996) Characterization of analytic functions in terms of their wavelet coefficients. Complex Variables 29: 265–276

Talagrand M (1994) Sharper bounds for empirical processes. Ann Probab 22: 28–76