

Nonparametric methods for volatility density estimation

Bert van Es, Peter Spreij, and Harry van Zanten

¹ Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, PO Box 94248, 1090GE Amsterdam, The Netherlands, a.j.vanes@uva.nl

² Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, PO Box 94248, 1090GE Amsterdam, The Netherlands, spreij@uva.nl

³ Department of Mathematics, Eindhoven University of Technology, PO Box 513, 5600MB Eindhoven, The Netherlands, j.h.v.zanten@tue.nl

Abstract. Stochastic volatility modelling of financial processes has become increasingly popular. The proposed models usually contain a stationary volatility process. We will motivate and review several nonparametric methods for estimation of the density of the volatility process. Both models based on discretely sampled continuous time processes and discrete time models will be discussed.

The key insight for the analysis is a transformation of the volatility density estimation problem to a deconvolution model for which standard methods exist. Three types of nonparametric density estimators are reviewed: the Fourier-type deconvolution kernel density estimator, a wavelet deconvolution density estimator and a penalized projection estimator. The performance of these estimators will be compared.

Key words: stochastic volatility models, deconvolution, density estimation, kernel estimator, wavelets, minimum contrast estimation, mixing

AMS subject classification: 62G07, 62G08, 62M07, 62P20, 91G70

1.1 Introduction

We discuss a number of nonparametric methods that come into play when one wants to estimate the density of the volatility process, given observations of the price process of some asset. The models that we treat are mainly formulated in continuous time, although we pay some separate attention to discrete time models. The observations of the continuous time models will always be in discrete time however and may occur at low frequency (fixed lag between observation instants), or high frequency (vanishing time lag). In this review, for simplicity we focus on the univariate marginal distribution of the volatility process, although similar results can be obtained for multivariate marginal distributions.

Although the underlying models differ in the sense that they are formulated either in continuous or in discrete time, in all cases the observations are given by a discrete time process. Moreover, as we shall see, the observation scheme can always (approximately) be cast as of ‘signal plus noise’ type

$$Y_i = X_i + \varepsilon_i,$$

where X_i is to be interpreted as the ‘signal’. If for fixed i the random variables X_i and ε_i are independent, the distribution of the Y_i is a convolution of the distributions of X_i and ε_i . The density of the ‘signal’ X_i is the object of interest, while the density of the ‘noise’ ε_i is supposed to be known to the observer. The statistical problem is to recover the density of the signal by *deconvolution*. Classically for such models it was often also assumed that the processes (X_i) and (ε_i) are i.i.d. Under these conditions Fan [12] gave lower bounds for the estimation of the unknown density f at a fixed point x_0 and showed that kernel-type estimators achieve the optimal rate. An alternative estimation method was proposed in the paper Pensky and Vidakovic [23], using wavelet methods instead of kernel estimators and where global L^2 -errors were considered instead of pointwise errors.

However, for the stochastic volatility models that we consider, the i.i.d. assumption on the X_i is violated. Instead, the X_i may be modelled as stationary random variables, that are allowed to exhibit some form of weak dependence, controlled by appropriate mixing properties, strongly mixing or β -mixing. These mixing conditions are justified by the fact that they are satisfied for many popular GARCH-type and stochastic volatility models (see e.g. Carrasco and Chen [6]), as well as for continuous time models, where σ^2 solves a stochastic differential equation, see e.g. Genon-Catalot et al. [17]. The estimators that we discuss are based on kernel methods, wavelets and penalized contrast estimation, also referred to as penalized projection estimation. We will review the performance of these deconvolution estimators under weaker than i.i.d. assumptions and show that this essentially depends on the smoothness and mixing conditions of the underlying process and the frequency of the observations. For a survey of other nonparametric statistical problems for financial data we refer to Franke et al. [14]

The paper is organized as follows. In Section 1.2 we introduce the continuous time model. In Section 1.3 we consider a kernel type estimator of the invariant volatility density and apply it to a set of real data. Section 1.4 is devoted to a wavelet density estimator and in Section 1.5 a minimum contrast estimator is discussed. Some related results for discrete time models are reviewed in Section 1.6 and Section 1.7 contains some concluding remarks.

1.2 The continuous time model

Let S denote the log price process of some stock in a financial market. It is often assumed that S can be modelled as the solution of a stochastic differen-

tial equation or, more general, as an Itô diffusion process. So we assume that we can write

$$dS_t = b_t dt + \sigma_t dW_t, \quad S_0 = 0, \quad (1.1)$$

or, in integral form,

$$S_t = \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \quad (1.2)$$

where W is a standard Brownian motion and the processes b and σ are assumed to satisfy certain regularity conditions (see Karatzas and Shreve [22]) to have the integrals in (1.2) well-defined. In a financial context, the process σ is called the volatility process. One often takes the process σ independent of the Brownian motion W .

Adopting this common assumption throughout the paper, unless explicitly stated otherwise, we also assume that σ is a strictly stationary positive process satisfying a mixing condition, for example an ergodic diffusion on $(0, \infty)$. The standing assumption in all what follows is that the one-dimensional marginal distribution of σ admits an invariant density w.r.t. Lebesgue measure on $(0, \infty)$. This is typically the case in virtually all stochastic volatility models that are proposed in the literature, where the evolution of σ is modelled by a stochastic differential equation, mostly in terms of σ^2 , or $\log \sigma^2$ (cf. e.g. Wiggins [31], Heston [20]). Often σ_t^2 is a function of a process X_t satisfying a stochastic differential equation of the type

$$dX_t = b(X_t) dt + a(X_t) dB_t, \quad (1.3)$$

with B_t a Brownian motion. Under regularity conditions, the invariant density of X is up to a multiplicative constant equal to

$$x \mapsto \frac{1}{a^2(x)} \exp \left(2 \int_{x_0}^x \frac{b(y)}{a^2(y)} dy \right), \quad (1.4)$$

where x_0 is an arbitrary element of the state space, see e.g. Gihman and Skorohod [19] or Skorokhod [25]. From formula (1.4) one sees that the invariant distribution of the volatility process (take X for instance equal to σ^2 or $\log \sigma^2$) may take on many different forms, as is the case for the various models that have been proposed in the literature. In absence of parametric assumptions on the coefficients a and b , we will investigate nonparametric procedures to estimate the corresponding densities, even refraining from an underlying model like (1.3), partly aimed at recovering possible ‘stylized facts’ exhibited by the observations.

For instance, one could think of *volatility clustering*. This may be cast by saying that for different time instants t_1, t_2 that are close, the corresponding values of $\sigma_{t_1}, \sigma_{t_2}$ are close again. This can partly be explained by assumed continuity of the process σ , but it might also result from specific areas around the diagonal where the multivariate density of $(\sigma_{t_1}, \sigma_{t_2})$ assumes high values if t_1 and t_2 are relatively close. It is therefore conceivable that the density of

$(\sigma_{t_1}, \sigma_{t_2})$ has high concentrations around points (ℓ, ℓ) and (h, h) , with $\ell < h$, a kind of bimodality of the joint distribution, with the interpretation that clustering occurs around a low value ℓ or around a high value h . This in turn may be reflected by bimodality of the univariate marginal distribution of σ_t .

A situation in which this naturally occurs is the following. Consider a regime switching volatility process. Assume that for $i = 0, 1$ we have two stationary processes X^i having stationary densities f^i . We assume these two processes to be independent, and also independent of a two-state stationary homogeneous Markov chain U with states 0, 1. The stationary distribution of U is given by $\pi_i := P(U_t = i)$. The process ξ is defined by

$$\xi_t = U_t X_t^1 + (1 - U_t) X_t^0.$$

Then ξ is stationary too and it has a stationary density f given by

$$f(x) = \pi_1 f^1(x) + \pi_0 f^0(x).$$

Suppose that the volatility process is defined by $\sigma_t^2 = \exp(\xi_t)$ and that the X^i are both Ornstein-Uhlenbeck processes given by

$$dX_t^i = -b_i(X_t^i - \mu_i) dt + a_i dW_t^i,$$

with W^1, W^2 independent Brownian motions, $\mu_1 \neq \mu_2$ and $b_1, b_2 > 0$. Suppose that the X^i start in their stationary $N(\mu_i, \frac{a_i^2}{2b_i})$ distributions. Then the stationary density f is a bimodal mixture of normal densities with μ_1 and μ_2 as the locations of the local maxima. Nonparametric procedures are able to detect such a property and are consequently by all means sensible tools to get some first insights into the shape of the invariant density.

A first object of study is the marginal univariate distribution of the stationary volatility process σ . We will also consider the invariant density of the integrated squared volatility process over an interval of length Δ . By stationarity of σ this is the density of $\int_0^\Delta \sigma_t^2 dt$. We will consider density estimators and assess their quality by giving results on their mean squared or integrated mean squared error. For kernel estimators, we rely on Van Es et al. [10], where this problem has been studied for the marginal *univariate* density of σ . In Van Es and Spreij [9] one can find results for multivariate density estimators. Results on wavelet estimators will be taken from Van Zanten and Zareba [32]. Penalized contrast estimators have been treated in Comte and Genon-Catalot [7].

The observations of log-asset price S process are assumed to take place at the time instants $0, \Delta, 2\Delta, \dots, n\Delta$. In case one deals with low frequency observations, Δ is fixed. For high frequency observations, the time gap satisfies $\Delta = \Delta_n \rightarrow 0$ as $n \rightarrow \infty$. To obtain consistency for the estimators that we will study in the latter case, we will make the additional assumption $n\Delta_n \rightarrow \infty$.

To explain the origin of the estimators that we consider in this paper, we often work with the simplified model, which is obtained from (1.1) by taking

$b_t = 0$. We then suppose to have discrete-time data $S_0, S_\Delta, S_{2\Delta}, \dots$ from a continuous-time stochastic volatility model of the form

$$dS_t = \sigma_t dW_t.$$

Under this additional assumption, we will see that we (approximately) deal with stationary observations Y_i that can be represented as $Y_i = X_i + \varepsilon_i$, where for each i the random variables X_i and ε_i are independent.

1.3 Kernel deconvolution

In this section we consider kernel deconvolution density estimators. We construct them, give expressions for bias and variance and give an application to real data.

1.3.1 Construction of the estimator

To motivate the construction of the estimator, we first consider (1.1) without the drift term, so we assume to have the simplified model

$$dS_t = \sigma_t dW_t, \quad S_0 = 0. \quad (1.5)$$

It is assumed that we observe the process S at the discrete time instants $0, \Delta, 2\Delta, \dots, n\Delta$, satisfying $\Delta \rightarrow 0, n\Delta \rightarrow \infty$. For $i = 1, 2, \dots$ we work, as in Genon-Catalot et al. [15, 16], with the normalized increments

$$X_i^\Delta = \frac{1}{\sqrt{\Delta}}(S_{i\Delta} - S_{(i-1)\Delta}).$$

For small Δ , we have the rough approximation

$$\begin{aligned} X_i^\Delta &= \frac{1}{\sqrt{\Delta}} \int_{(i-1)\Delta}^{i\Delta} \sigma_t dW_t \\ &\approx \sigma_{(i-1)\Delta} \frac{1}{\sqrt{\Delta}} (W_{i\Delta} - W_{(i-1)\Delta}) \\ &= \sigma_{(i-1)\Delta} Z_i^\Delta, \end{aligned} \quad (1.6)$$

where for $i = 1, 2, \dots$ we define

$$Z_i^\Delta = \frac{1}{\sqrt{\Delta}} (W_{i\Delta} - W_{(i-1)\Delta}).$$

By the independence and stationarity of Brownian increments, the sequence $Z_1^\Delta, Z_2^\Delta, \dots$ is an i.i.d. sequence of standard normal random variables. Moreover, the sequence is independent of the process σ by assumption.

Writing $Y_i = \log(X_i^\Delta)^2$, $\xi_i = \log \sigma_{(i-1)\Delta}^2$, $\varepsilon_i = \log(Z_i^\Delta)^2$ and taking the logarithm of the square of X_i^Δ we get

$$Y_i \approx \xi_i + \varepsilon_i,$$

where the terms in the sum are independent. Assuming that the approximation is sufficiently accurate we can use this approximate convolution structure to estimate the unknown density f of $\log \sigma_{i\Delta}^2$ from the transformed observed $Y_i = \log(X_i^\Delta)^2$. The characteristic functions involved are denoted by ϕ_Y , ϕ_ξ and ϕ_k , where k is the density of the ‘noise’ $\log(Z_i^\Delta)^2$. One obviously has $\phi_Y = \phi_\xi \phi_k$ and one easily sees that the density k is given by

$$k(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x} e^{-\frac{1}{2}e^x},$$

and its characteristic function by

$$\phi_k(t) = \frac{1}{\sqrt{\pi}} 2^{it} \Gamma\left(\frac{1}{2} + it\right).$$

The idea of getting a deconvolution estimator of f is simple. Using a kernel function w , a bandwidth h , and the Y_i , the density g of the Y_i is estimated by

$$g_{nh}(y) = \frac{1}{nh} \sum_j w\left(\frac{y - Y_j}{h}\right).$$

Denoting $\phi_{g,nh}$ the characteristic function of g_{nh} , one estimates ϕ_Y by $\phi_{g,nh}$ and ϕ_ξ by $\phi_{g,nh}/\phi_k$. Following a well-known approach in statistical deconvolution theory (see e.g. Section 6.2.4 of Wand and Jones [30]), Fourier inversion then yields the density estimator of f . By elementary calculations one obtains from this procedure

$$f_{nh}(x) = \frac{1}{nh} \sum_{j=1}^n v_h\left(\frac{x - \log(X_j^\Delta)^2}{h}\right), \quad (1.7)$$

where v_h is the kernel function, depending on the bandwidth h ,

$$v_h(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\phi_w(s)}{\phi_k(s/h)} e^{-isx} ds. \quad (1.8)$$

One easily verifies that the estimator f_{nh} , is real-valued.

To justify the approximation in (1.6), we quantify a stochastic continuity property of σ^2 . In addition to this we make the mixing condition explicit. We impose

Condition 1.3.1 The process σ^2 satisfies the following conditions.

1. It is L^1 -Hölder continuous of order one half, $\mathbb{E}|\sigma_t^2 - \sigma_0^2| = O(t^{1/2})$ for $t \rightarrow 0$.
2. It is strongly mixing with coefficient $\alpha(t)$ satisfying, for some $0 < q < 1$,

$$\int_0^\infty \alpha(t)^q dt < \infty. \quad (1.9)$$

The kernel function w is assumed to satisfy the following conditions (an example of such a kernel is given in (1.12) below, see also Wand [29]), which includes in particular the behavior of ϕ_w at the boundary of its domain.

Condition 1.3.2 Let w be a real symmetric function with real valued symmetric characteristic function ϕ_w with support $[-1,1]$. Assume further

1. $\int_{-\infty}^{\infty} |w(u)|du < \infty$, $\int_{-\infty}^{\infty} w(u)du = 1$, $\int_{-\infty}^{\infty} u^2|w(u)|du < \infty$,
2. $\phi_w(1-t) = At^\rho + o(t^\rho)$, as $t \downarrow 0$ for some $\rho > 0$, $A \in \mathbb{R}$.

The first part of Condition 1.3.1 is motivated by the situation where $X = \sigma^2$ solves a SDE like (1.1). It is easily verified that for such processes it holds that $\mathbb{E}|\sigma_t^2 - \sigma_0^2| = O(t^{1/2})$, provided that $b \in L_1(\mu)$ and $a \in L_2(\mu)$, where μ is the invariant probability measure. Indeed we have $\mathbb{E}|\sigma_t^2 - \sigma_0^2| \leq \mathbb{E} \int_0^t |b(\sigma_s^2)| ds + (\mathbb{E} \int_0^t a^2(\sigma_s^2) ds)^{1/2} = t\|b\|_{L_1(\mu)} + \sqrt{t}\|a\|_{L_2(\mu)}$.

The main result we present for this estimator concerns its mean squared error at a fixed point x . Although the motivation of the estimator was based on the simplified model (1.5), the result below applies to the original model (1.1). For its proof and additional technical details, see Van Es et al. [10].

Theorem 1.3.3 *Assume that $\mathbb{E}b_t^2$ is bounded. Let the process σ satisfy Condition 1.3.1, and let the kernel function w satisfy Condition 1.3.2. Moreover, let the density f of $\log \sigma_t^2$ be twice continuously differentiable with a bounded second derivative. Also assume that the density of σ_t^2 is bounded in a neighbourhood of zero. Suppose that $\Delta = n^{-\delta}$ for given $0 < \delta < 1$ and choose $h = \gamma\pi/\log n$, where $\gamma > 4/\delta$. Then the bias of the estimator (1.7) satisfies*

$$\mathbb{E}f_{nh}(x) - f(x) = \frac{1}{2}h^2 f''(x) \int u^2 w(u)du + o(h^2), \tag{1.10}$$

whereas, the variance of the estimator satisfies the order bounds

$$\text{Var } f_{nh}(x) = O\left(\frac{1}{n} h^{2\rho} e^{\pi/h}\right) + O\left(\frac{1}{nh^{1+q}\Delta}\right). \tag{1.11}$$

Remark 1.3.4 The choices $\Delta = n^{-\delta}$, with $0 < \delta < 1$ and $h = \gamma\pi/\log n$, with $\gamma > 4/\delta$ render a variance that is of order $n^{-1+1/\gamma}(1/\log n)^{2\rho}$ for the first term of (1.11) and $n^{-1+\delta}(\log n)^{1+q}$ for the second term. Since by assumption $\gamma > 4/\delta$ we have $1/\gamma < \delta/4 < \delta$ so the second term dominates the first term. The order of the variance is thus $n^{-1+\delta}(\log n)^{1+q}$. Of course, the order of the bias is logarithmic, hence the bias dominates the variance and the mean squared error of $f_{nh}(x)$ is of order $(\log n)^{-4}$.

Remark 1.3.5 It can then be shown that for the characteristic function ϕ_k one has the behavior

$$|\phi_k(s)| = \sqrt{2} e^{-\frac{1}{2}\pi|s|} (1 + O(\frac{1}{|s|})), \quad |s| \rightarrow \infty.$$

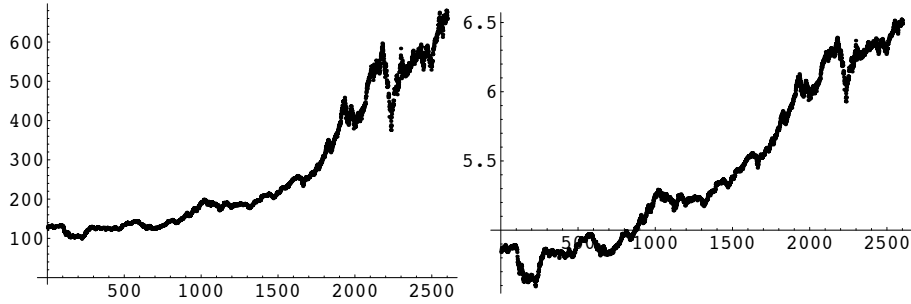


Fig. 1.1. AEX. Left: daily closing values. Right: log of the daily closing values.

This means that k is supersmooth in the terminology of Fan [12] which explains the slow logarithmic rate at which the bias vanishes. Sharper results on the variance can be obtained when σ^2 is strongly mixing, see Van Es et al. [11] for further details. The orders of the bias and of the MSE remain unchanged though.

1.3.2 An application to the Amsterdam AEX index

In this section we present an example using real data of the Amsterdam AEX stock exchange. We have estimated the volatility density from 2600 daily closing values of the Amsterdam stock exchange index AEX from 12/03/1990 until 14/03/2000. These data are represented in Figure 1.1. We have centered the daily log returns, i.e we have subtracted the mean (which equaled 0.000636), see Figure 1.2. The deconvolution estimator is given as the left hand picture in Figure 1.3. Observe that the estimator strongly indicates that the underlying density is unimodal. Based on computations of the mean and variance of the estimate, with $h = 0.7$, we have also fitted a normal density by hand and compared it to the kernel deconvolution estimator. The result is given as the right hand picture in Figure 1.3. The resemblance is remarkable.

The kernel used to compute the estimates is a kernel from Wand [29], with $\rho = 3$ and $A = 8$,

$$w(x) = \frac{48x(x^2 - 15) \cos x - 144(2x^2 - 5) \sin x}{\pi x^7}. \quad (1.12)$$

It has characteristic function

$$\phi_w(t) = (1 - t^2)^3, \quad |t| \leq 1. \quad (1.13)$$

The bandwidths are chosen by hand. The estimates have been computed by fast Fourier transforms using the Mathematica 4.2 package.

This is actually the same example as in our paper Van Es et al. [11] on volatility density estimation for discrete time models. The estimator (1.7)

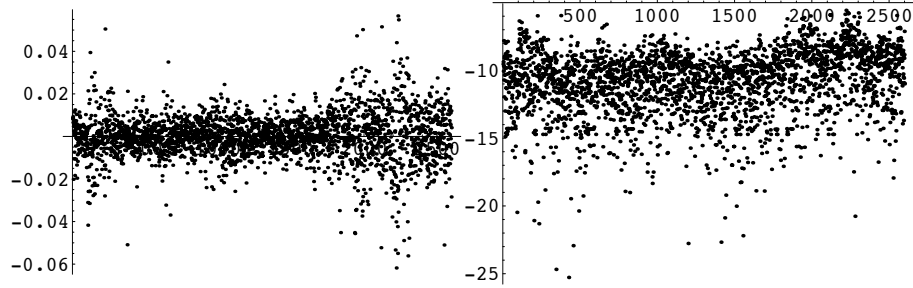


Fig. 1.2. AEX. Left: the values of X_t , i.e. the centered daily log returns. Right: $\log(X_t^2)$.

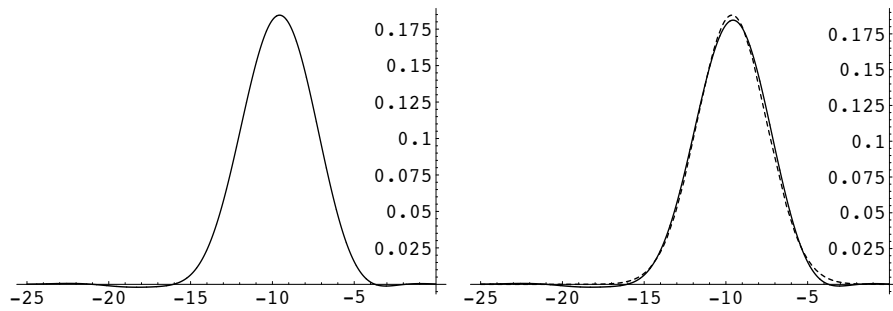


Fig. 1.3. AEX. Left: The estimate of the density of $\log(\sigma_t^2)$ with $h = 0.7$. Right: The normal fit to the $\log(\sigma_t^2)$. The dashed line is the normal density and the solid line the kernel estimate.

presented here is, as a function of the sampled data, exactly the same as the one for the discrete time models. The difference lies in the choice of underlying model. In the present paper the model is a discretely sampled continuous time process, while in Van Es et al. [11] it is a discrete time process. For the latter type of models the discretization step in the beginning of this section is not necessary since these models satisfy an exact convolution structure.

1.4 Wavelet deconvolution

As an alternative to kernel methods, in this section we consider estimators based on wavelets. Starting point is again the simplified model (1.5). Contrary to the previous section, we are now interested in estimating the accumulated squared volatility over an interval of length Δ . We assume having observations of S at times $i\Delta$ to our disposal, but now with Δ fixed (low

frequency observations). Let, as before, $X_i^\Delta = \Delta^{-1/2}(S_{i\Delta} - S_{(i-1)\Delta})$ and let $\bar{\sigma}_i^2 = \Delta^{-1} \int_{(i-1)\Delta}^{i\Delta} \sigma_t^2 dt$. Denote by \mathcal{F}_σ the σ -algebra generated by the process σ . By the assumed independence of the processes σ and W , we have for the characteristic function of X_i^Δ given \mathcal{F}_σ

$$\mathbb{E}[\exp(isX_i^\Delta)|\mathcal{F}_\sigma] = \exp(-\frac{1}{2}\bar{\sigma}_i^2 s^2).$$

Consider also the model $\tilde{X}_i^\Delta = \bar{\sigma}_i Z_i$, with $\bar{\sigma}_i$ and Z_i independent for each i and Z_i a standard Gaussian random variable. Then

$$\mathbb{E}[\exp(is\tilde{X}_i^\Delta)|\mathcal{F}_{\sigma_i}] = \exp(-\frac{1}{2}\bar{\sigma}_i^2 s^2).$$

It follows that X_i^Δ and \tilde{X}_i^Δ are identically distributed. From this observation we conclude that the transformed increments $\log(\Delta^{-1}(S_{i\Delta} - S_{(i-1)\Delta})^2)$ are then distributed as $Y_i = \xi_i + \varepsilon_i$, where

$$\xi_i = \log \bar{\sigma}_i^2, \quad \varepsilon_i = \log Z_i^2,$$

and Z_i is an i.i.d. sequence of standard Gaussian random variables, independent of σ . The sequence ξ_i is stationary and we assume that its marginal density g exists, i.e. g is the density of $\log(\Delta^{-1} \int_0^\Delta \sigma_u^2 du)$. The density of the ε_i is again denoted by k . Of course, estimating g is equivalent to estimating the density of the aggregated squared volatility $\int_0^\Delta \sigma_u^2 du$.

In the present section the main focus is on the quality of the estimator in terms of the mean integrated squared error, as opposed to establishing results for the (pointwise) mean squared error as in Section 1.3. At the end of this section we compare the results presented here to those of Section 1.3.

First we recall the construction of the wavelet estimator proposed in Pensky and Vidakovic [23]. For the necessary background on wavelet theory, see for instance Blatter [1], Jawerth and Sweldens [21], and the references therein. For the construction of deconvolution estimators we need to use band-limited wavelets. As in Pensky and Vidakovic [23] we use a Meyer-type wavelet (see also Walter [27], Walter and Zayed [28]). We consider an orthogonal scaling function and wavelet φ and ψ , respectively, associated with an orthogonal multiresolution analysis of $L^2(\mathbb{R})$. We denote in this section the Fourier transform of a function f by \tilde{f} , i.e.

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

and suppose that for a symmetric probability measure μ with support contained in $[-\pi/3, \pi/3]$ it holds that

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

Observe that the assumptions imply that φ and ψ are indeed band-limited. For the supports of their Fourier transforms we have $\text{supp } \tilde{\varphi} \subset [-4\pi/3, 4\pi/3]$ and $\text{supp } \tilde{\psi} \subset [-8\pi/3, -2\pi/3] \cup [2\pi/3, 8\pi/3]$. 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 g can now be written as

$$g(x) = \sum_{l \in \mathbb{Z}} a_{m,l} \varphi_{m,l}(x) + \sum_{l \in \mathbb{Z}} \sum_{j=m}^{\infty} b_{j,l} \psi_{j,l}(x), \quad (1.14)$$

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

$$a_{m,l} = \int_{\mathbb{R}} \varphi_{m,l}(x) g(x) dx, \quad b_{j,l} = \int_{\mathbb{R}} \psi_{j,l}(x) g(x) dx.$$

The idea behind the linear wavelet estimator is simple. We first approximate g by the orthogonal projection given by the first term on the right-hand side of (1.14). 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 g is estimated by replacing the coefficients $a_{m,l}$ by consistent estimators and truncating the sum. Using the fact that the density p of an observation Y_i is the convolution of g and k it is easily verified that

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

where U_m is the function with Fourier transform

$$\tilde{U}_m(\omega) = \frac{\tilde{\varphi}(\omega)}{\tilde{k}(-2^m \omega)}. \quad (1.15)$$

We estimate the coefficient $a_{m,l}$ by its empirical counterpart

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

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

$$\hat{g}_n(x) = \sum_{|l| \leq L_n} \hat{a}_{m_n,l,n} \varphi_{m_n,l}(x), \quad (1.16)$$

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

The main results in the present section are upper bounds for the mean integrated square error of the wavelet estimator \hat{g}_n , which is defined as usual by

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

We will specify how to choose the detail level m_n and the truncation point L_n in (1.16) optimally in different cases, depending on the smoothness of g and k . The smoothness properties of g are described in terms of g belonging to certain Sobolev balls and by imposing a weak condition on its decay rate. The Sobolev space H^α is defined for $\alpha > 0$ by

$$H^\alpha = \left\{ g : \|g\|_\alpha = \left(\int_{\mathbb{R}} |\tilde{g}(\omega)|^2 (\omega^2 + 1)^\alpha d\omega \right)^{1/2} < \infty \right\}. \quad (1.17)$$

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

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

The additional assumption on the decay rate is reflected by g belonging to

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

We now have the following result, see Van Zanten and Zareba [32], for the wavelet density estimator \hat{g}_n of g defined by (1.16).

Theorem 1.4.1 *Suppose that the volatility process σ^2 is strongly mixing with mixing coefficients satisfying*

$$\sum_{k \geq 0} \alpha_{k\Delta}^p < \infty \quad (1.18)$$

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

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

the mean square error of the wavelet estimator satisfies

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

for $\alpha, A, A' > 0$. If (1.18) is satisfied for all $p \in (0, 1)$, the same bound is true if the choice for L_n is replaced by $L_n = n$.

Let us point out the relation with the results of Section 1.3 and with those in Van Es et al. [11], see also Section 1.6.1. In that paper kernel-type deconvolution estimators for discrete time stochastic volatility models were considered. When applied to the present model the results say that under the same mixing condition and assuming that g has two bounded and continuous derivatives, the (pointwise) mean squared error of the kernel estimator

is of order $(\log n)^{-4}$. The analogue of g having two bounded derivatives in our setting is that $g \in \mathcal{S}_2^*(A, A')$ for some $A, A' > 0$. Indeed, the theorem yields the same bound $(\log n)^{-4}$ for the MISE in this case. The same bound is valid for the MSE when estimating the marginal density for continuous time models, see Theorem 1.3.3 and its consequences in Remark 1.3.4. Theorem 1.4.1 is more general, because the smoothness level is not fixed at $\alpha = 2$, but allows for different smoothness levels of order $\alpha \neq 2$ as well. Moreover, the wavelet estimator is adaptive in the sense that it does not depend on the unknown smoothness level, if the condition on the mixing coefficients holds for all $p \in (0, 1)$.

1.5 Penalized projection estimators

The results of the preceding sections assume that the true (integrated) volatility density has a finite degree of regularity, either in Hölder or in Sobolev sense. Under this assumption the nonparametric estimators have logarithmic convergence rates, cf. Remark 1.3.4 and Theorem 1.4.1. Although admittedly slow, the minimax results of Fan [12] show that these rates are in fact optimal in this setting. In the paper Pensky and Vidakovic [23] it was shown however that if in a deconvolution setting the density of the unobserved variables has the same degree of smoothness as the noise density, the rates can be significantly improved, cf. also the lower bounds obtained in Butucea [4] and Butucea and Tsybakov [5]. This observation forms the starting point of the paper Comte and Genon-Catalot [7], in which a nonparametric volatility density estimator is developed that achieves better rates than logarithmic if the true density is super smooth. In the latter paper it is assumed that there are observations $S_\Delta, S_{2\Delta}, \dots, S_{n\Delta}$ of a process S satisfying the simple equation (1.5), with $V = \sigma^2$ a strictly positive process, independent of the Brownian motion W . It is assumed that we deal with high frequency observations, $\Delta \rightarrow 0$ and $n\Delta \rightarrow \infty$. We impose the following condition on V .

Condition 1.5.1 The process V is a time-homogenous, continuous Markov process, strictly stationary and ergodic. It is either β -mixing with coefficient $\beta(t)$ satisfying

$$\int_0^\infty \beta(t) dt < \infty,$$

or is ρ -mixing. Moreover, it satisfies the Lipschitz condition

$$\mathbb{E} \left(\log \left(\frac{1}{\Delta} \int_0^\Delta V_t dt \right) - \log V_0 \right)^2 \leq C\Delta,$$

for some $C > 0$.

In addition to this a technical assumption is necessary on the density f of $\log V_0$ we are interested in and on the density g_Δ of $\log \left(\frac{1}{\Delta} \int_0^\Delta V_t dt \right)$, which

is assumed to exist. Contrary to the notation of the previous section, we write g_Δ instead of g , since now Δ is not fixed.

Condition 1.5.2 The invariant density f is bounded and has a second moment and $g_\Delta \in L^2(\mathbb{R})$.

As a first step in the construction of the final estimator a preliminary estimator \hat{f}_L is constructed for $L \in \mathbb{N}$ fixed. Note that Condition 1.5.2 implies that $f \in L^2(\mathbb{R})$, hence we can consider its orthogonal projection f_L on the subspace S_L of $L^2(\mathbb{R})$, defined as the space of functions whose Fourier transform is supported on the compact interval $[-\pi L, \pi L]$. An orthonormal basis for the latter space is formed by the Shannon basis functions $\psi_{L,j}(x) = \sqrt{L}\psi(Lx-j)$, $j \in \mathbb{Z}$, with $\psi(x) = \sin(\pi x)/(\pi x)$ the sinc kernel. For integers $K_n \rightarrow \infty$ to be specified below, the space S_L is approximated by the finite-dimensional spaces $S_L^n = \text{span}\{\psi_{L,j} : |j| \leq K_n\}$. The function f_L is estimated by $\hat{f}_L = \text{argmin}_{h \in S_L^n} \gamma_n(h)$, where the contrast function γ_n is defined for $h \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ by

$$\gamma_n(h) = \|h\|_2^2 - \frac{2}{n} \sum_{i=1}^n u_h(\log(X_i^\Delta)^2), \quad u_h(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixs} \frac{\tilde{h}(-s)}{\phi_k(s)} ds.$$

Here, as before, ϕ_k is the characteristic function of $\log \varepsilon^2$, with ε standard normal and \tilde{h} is the Fourier transform of h . It is easily seen that

$$\hat{f}_L = \sum_{|j| \leq K_n} \hat{a}_{L,j} \psi_{L,j}, \quad \hat{a}_{L,j} = \frac{1}{n} \sum_{i=1}^n u_{\psi_{L,j}}(\log(X_i^\Delta)^2).$$

Straightforward computations show that, with $\langle \cdot, \cdot \rangle$ the $L^2(\mathbb{R})$ inner product, $\mathbb{E}u_h(\log(X_i^\Delta)^2) = \langle h, g_\Delta \rangle$, and hence $\mathbb{E}\gamma_n(h) = \|h - g_\Delta\|_2^2 - \|g_\Delta\|_2^2$. So in fact, \hat{f}_L is an estimator of the element of S_L^n which is closest to g_Δ . Since S_L^n approximates S_L for large n and g_Δ is close to f for small Δ , the latter element should be close to f_L .

Under Conditions 1.5.1 and 1.5.2, a bound for the mean integrated square error, or quadratic risk $\text{MISE}(\hat{f}_L) = \mathbb{E}\|\hat{f}_L - f\|_2^2$ can be derived, depending on the approximation error $\|f - f_L\|_2$, the bandwidth L and the truncation point K_n , see Comte and Genon-Catalot [7], Theorem 1. The result implies that if f belongs to the Sobolev space H^α as defined in (1.17), then the choices $K_n = n$ and $L = L_n \sim \log n$ yield a MISE of the order $(\log n)^{-2\alpha}$, provided that $\Delta = \Delta_n = n^{-\delta}$ for some $\delta \in (0, 1)$. Not surprisingly, this is completely analogous to the result obtained in Theorem 1.4.1 for the wavelet-based estimator in the fixed Δ setting. In particular the procedure is adaptive, in that the estimator does not depend on the unknown regularity parameter α .

To obtain faster than logarithmic rates and adaptation in the case that f is supersmooth, a data-driven choice of the bandwidth L is proposed. Define

$$\hat{L} = \operatorname{argmin}_{L \in \{1, \dots, \log n\}} \left(\gamma_n(\hat{f}_L) + \operatorname{pen}_n(L) \right),$$

where the penalty term is given by

$$\operatorname{pen}_n(L) = \kappa \frac{(1+L)\Phi_k(L)}{n},$$

for $\kappa > 0$ a calibration constant and

$$\Phi_k(L) = \int_{-\pi L}^{\pi L} \frac{1}{|\phi_k(s)|^2} ds.$$

For the quadratic risk of the estimator $\hat{f}_{\hat{L}}$, the following result holds (Comte and Genon-Catalot [7]).

Theorem 1.5.3 *Under Conditions 1.5.2 and 1.5.1 we have*

$$\begin{aligned} \operatorname{MISE}(\hat{f}_{\hat{L}}) &\leq C_1 \inf_{L \in \{1, \dots, \log n\}} \left(\|f - f_L\|_2^2 + \frac{(1+L)\Phi_k(L)}{n} \right) \\ &\quad + C_2 \frac{\log^2 n}{K_n} + C_3 \frac{\log n}{n\Delta} + C_4 \Delta \log^3 n, \end{aligned}$$

for constants $C_1, C_2, C_3, C_4 > 0$.

It can be seen that this bound is worse than the corresponding bound for the estimator \hat{f}_L by a factor of the order L . This is at worst a logarithmic factor which, as usual in this kind of setting, has to be paid for achieving adaptation. The examples in Section 6 of Comte and Genon-Catalot [7] show that indeed, the estimator $\hat{f}_{\hat{L}}$ can achieve algebraic convergence rates in case the true density f is supersmooth.

1.6 Estimation for discrete time models

Although the main focus of the present paper is on estimation procedures for continuous time models, in the present section we also highlight some analogous results for discrete time models. These deal with both density and regression function estimation.

1.6.1 Discrete time models

The discrete time analogue of (1.5) is

$$X_t = \sigma_t Z_t, \quad t = 1, 2, \dots \quad (1.19)$$

Here we denote by X the detrended or demeaned log-return process. Stochastic volatility models are often described in this form. The sequence Z is typically an i.i.d. noise (e.g. Gaussian) and at each time t the random variables

σ_t and Z_t are independent. See the survey papers by Ghysels et al. [18] or Shephard [24]. Also in this section we assume that the process σ is strictly stationary and that the marginal distribution of σ has a density with respect to the Lebesgue measure on $(0, \infty)$. We present some results for a nonparametric estimator of the density of $\log \sigma_t^2$, as well as results for a nonparametric estimator of a nonlinear regression function, in case σ^2 is given by a nonlinear autoregression. The standing assumption in all what follows is that for each t the random variables σ_t and Z_t are independent, the noise sequence is standard Gaussian and σ is a strictly stationary, positive process satisfying a certain mixing condition.

In principle one can distinguish two classes of models. The way in which the bivariate process (σ, Z) , in particular its dependence structure, is further modelled offers different possibilities. In the first class of models one assumes that the process σ is predictable with respect to the filtration \mathcal{F}_t generated by the process Z , and obtains that σ_t is independent of Z_t for each fixed time t . We furthermore have that (assuming that the unconditional variances are finite) σ_t^2 is equal to the conditional variance of X_t given \mathcal{F}_{t-1} . This class of models has become quite popular in the econometrics literature. It is well known that this class also contains the (parametric) family of GARCH-models, introduced by Bollerslev [2].

In the second class of models one assumes that the whole process σ is independent of the noise process Z , and one commonly refers to the resulting model as a stochastic volatility model. In this case, the natural underlying filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t \geq 0}$ is generated by the two processes Z and σ in the following way. For each t the σ -algebra \mathcal{F}_t is generated by Z_s , $s \leq t$ and σ_s , $s \leq t + 1$. This choice of the filtration enforces σ to be predictable. As in the first model the process X becomes a martingale difference sequence and we have again (assuming that the unconditional variances are finite) that σ_t^2 is the conditional variance of X_t given \mathcal{F}_{t-1} . An example of such a model is given in De Vries [26], where σ is generated as an AR(1) process with α -stable noise ($\alpha \in (0, 1)$).

As in the previous sections we refrain from parametric modelling and review some completely nonparametric approaches. We will mainly focus on results for the second class, as it is the discrete time analogue of the stochastic volatility models of the previous sections. At the heart of all what follows is again the convolution structure that is obtained from (1.19) by squaring and taking logarithms,

$$\log X_t^2 = \log \sigma_t^2 + \log Z_t^2.$$

1.6.2 Density estimation

The main result of this section gives a bias expansion and a variance bound of a kernel density type estimator of the density f of $\log \sigma_t^2$. The estimator is, analogously to (1.7),

$$f_{nh}(x) = \frac{1}{nh} \sum_{j=1}^n v_h \left(\frac{x - \log(X_j)^2}{h} \right), \quad (1.20)$$

where v_h is the kernel function of (1.8).

The next theorem is derived from Van Es et al. [11], where a multivariate density estimator is considered. It establishes the expansion of the bias and an order bound on the variance of our estimator under a strong mixing condition. Under broad conditions this mixing condition is satisfied if the process σ Markov, since then convergence of the mixing coefficients to zero takes place at an *exponential rate*, see Theorems 4.2 and Theorem 4.3 of Bradley [3] for precise statements. Similar behaviour occurs for ARMA processes with absolutely continuous distributions of the noise terms (Bradley [3], Example 6.1).

Theorem 1.6.1 *Assume that the process σ is strongly mixing with coefficient α_k satisfying*

$$\sum_{j=1}^{\infty} \alpha_j^\beta < \infty,$$

for some $\beta \in (0, 1)$. Let the kernel function w satisfy Condition 1.3.2 and let the density f of $\log \sigma_t^2$ be bounded and twice continuously differentiable with bounded second order partial derivatives. Assume furthermore that σ and Z are independent processes. Then we have for the estimator of f defined as in (1.20) and $h \rightarrow 0$

$$\mathbb{E}f_{nh}(x) = f(x) + \frac{1}{2}h^2 f''(x) \int u^2 w(u) du + o(h^2) \quad (1.21)$$

and

$$\text{Var} f_{nh}(x) = O\left(\frac{1}{n} h^{2\rho} e^{\pi/h}\right). \quad (1.22)$$

Remark 1.6.2 Comparing the above results to the ones in Theorem 1.3.3, we observe that in the continuous time case, the variance has an additional $O\left(\frac{1}{nh^{1+q}\Delta}\right)$ term.

1.6.3 Regression function estimation

In this section we assume the basic model (1.19), but in addition we assume that the process σ satisfies a nonlinear autoregression and we consider non-parametric estimation of the regression function as proposed in Franke et al. [13]. In that paper a discrete time model was proposed as a discretization of the continuous time model given by (1.1). In fact, Franke et al. include a mean parameter μ , but since they assume it to be known, without loss of generality we can still assume (1.19). Assume that the volatility process is strictly positive and consider $\log \sigma_t^2$. It is assumed that its evolution is governed by

$$\log \sigma_{t+1}^2 = m(\log \sigma_t^2) + \eta_t, \quad (1.23)$$

where the η_t are i.i.d. Gaussian random variables with zero mean. The regression function m is assumed to satisfy the stability condition

$$\limsup_{|x| \rightarrow \infty} \left| \frac{m(x)}{x} \right| < 1. \quad (1.24)$$

Under this condition the process σ is exponentially ergodic and strongly mixing, see Doukhan [8] and these properties carry over to the process X as well. Moreover, the process $\log \sigma_t^2$ admits an invariant density f .

Denoting $Y_t = \log X_t^2$, we have

$$Y_t = \log \sigma_t^2 + \log Z_t^2.$$

It is common to assume that the processes Z and η are independent, the second class of models described in Section 1.6.1, but dependence between η_t and Z_t for fixed t can be allowed for (first model class) without changing in what follows, see Franke et al. [13].

The purpose of the present section is to estimate the function m in (1.23). To that end we use the estimator f_{nh} as defined in (1.20). Since this estimator resembles an ordinary kernel density estimator, the important difference being that the kernel function v_h now depends on the bandwidth h , the idea is to mimic the classical Nadaraya-Watson regression estimator similarly, in order to obtain an estimator of $m(x)$. Doing so, one obtains the estimator

$$m_{nh}(x) = \frac{\frac{1}{nh} \sum_{j=1}^n v_h\left(\frac{x-Y_j}{h}\right) Y_{j+1}}{f_{nh}(x)}. \quad (1.25)$$

It follows that

$$m_{nh}(x) - m(x) = \frac{p_{nh}(x)}{f_{nh}(x)},$$

where

$$p_{nh}(x) = \frac{1}{nh} \sum_{j=1}^n v_h\left(\frac{x-Y_j}{h}\right) (Y_{j+1} - m(x)).$$

In Franke et al. [13] bias expansions for $p_{nh}(x)$ and f_{nh} are given that fully correspond to those in Theorem 1.6.1. They are again of order h^2 , under similar assumptions. It is also shown that the variances of p_{nh} and f_{nh} tend to zero. The main result concerning the asymptotic behavior then follows from combining the asymptotics for p_{nh} and f_{nh} .

Theorem 1.6.3 *Assume that m satisfies the stability condition (1.24), that m and f are twice differentiable and the first of Condition 1.3.2 on the kernel w . The estimator $m_{nh}(x)$ satisfies $(\log n)^2(m_{nh}(x) - m(x)) = O_p(1)$ if $h = \gamma/\log n$ with $\gamma > \pi$.*

Following the proofs in Franke et al. [13], one can conclude that e.g. the variance of p_{nh} is of order $O(\frac{\exp(\pi/h)}{nh^2})$, which tends to zero for $h = \gamma/\log n$, with $\gamma > \pi$. For the variance of f_{nh} a similar bound holds. Comparing these order bounds to the ones in Theorem 1.6.1, we see that the latter ones are sharper. This is partly due to the fact that Franke et al. [13], don't impose conditions on the boundary behavior of the function ϕ_w (the second of Condition 1.3.2), whereas their other assumptions are the same as in Theorem 1.6.1.

1.7 Concluding remarks

In recent years, many different parametric stochastic volatility models have been proposed in the literature. To investigate which of these models are best supported by observed asset price data, nonparametric methods can be useful. In this paper we reviewed a number of such methods that have recently been proposed. The overview shows that ideas from deconvolution theory can be instrumental in dealing with this statistical problem and that both for high and for low frequency data, methods are now available for nonparametric estimation of the (integrated) volatility density at optimal convergence rates.

On a critical note, the methods available so far all assume that the volatility process is independent of the Brownian motion driving the asset price dynamics. This is a limitation, since in several interesting models non-zero correlations are assumed between the Brownian motions driving the volatility dynamics and the asset price dynamics.

References

1. C. Blatter, *Wavelets, a primer*, (A. K. Peters Ltd, Natick, MA, 1998).
2. T. Bollerslev, Generalized autoregressive conditional heteroscedasticity, *J. Econometrics* **31**, 307–321 (1986).
3. R.C. Bradley, Basic properties of strong mixing conditions, in: *Dependence in Probability and Statistics*, ed. by E. Eberlein and M.S. Taqqu Eds (Birkhäuser Boston, Inc., Boston, MA, 1986).
4. C. Butucea, Deconvolution of supersmooth densities with smooth noise, *Can. J. Statist.* **32**, 181–192 (2004).
5. C. Butucea, and A.B. Tsybakov, Sharp optimality in density deconvolution with dominating bias. I, *Theory Probab. Appl.* **52**(1), 24–39 (2008).
6. M. Carrasco, and X. Chen, Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory* **18**, 17–39 (2002).
7. F. Comte, and V. Genon-Catalot, Penalized projection estimator for volatility density, *Scand. J. Statist.* **33**(4), 875–893 (2006).
8. P. Doukhan, *Mixing: Properties and Examples*, Lecture Notes in Statistics, 85 (Springer-Verlag, New York, 1994).
9. A.J. van Es and P.J.C. Spreij, Multivariate Nonparametric Volatility Density Estimation, available at *Mathematics ArXiv*, <http://arxiv.org/abs/0910.4337> (2009).

10. A.J. van Es, P.J.C. Spreij and J.H. van Zanten, Nonparametric volatility density estimation, *Bernoulli* **9**(3), 451–465 (2003).
11. A.J. van Es, P.J.C. Spreij and J.H. van Zanten, Nonparametric volatility density estimation for discrete time models, *J. Nonparametr. Stat.* **17**, 237–251 (2005).
12. J. Fan, On the optimal rates of convergence of nonparametric deconvolution problems, *Ann. Statist.* **19**, 1257–1272 (1991).
13. J. Franke, W. Härdle and J.P. Kreiss, Nonparametric estimation in a stochastic volatility model, in: *Recent Advances and Trends in Nonparametric Statistics*, ed. by M.G. Akritas and D.N. Politis (Elsevier, Amsterdam, 2003).
14. J. Franke, J.P. Kreiss and E. Mammen E, Nonparametric modelling in financial time series, in: *Handbook of Financial Time Series*, ed. by T. Mikosch, J.P. Kreiss, R.A. Davis and T.G. Andersen Eds (Springer, New York, 2009).
15. V. Genon-Catalot, T. Jeantheau, and C. Larédo, Limit theorems for discretely observed stochastic volatility models, *Bernoulli* **4**, 283–303 (1998).
16. V. Genon-Catalot, T. Jeantheau, and C. Larédo, Parameter estimation for discretely observed stochastic volatility models, *Bernoulli* **5**, 855–872 (1999).
17. V. Genon-Catalot, T. Jeantheau, and C. Larédo, Stochastic volatility models as hidden Markov models and statistical applications, *Bernoulli* **6**(6),1051–1079 (2000).
18. E. Ghysels, A. Harvey, and E. Renault, Stochastic Volatility, in: *Statistical Methods in Finance*, ed. by G.S. Maddala, and C.R. Rao, Handbook of Statistics, 14 (North-Holland Publishing Co., Amsterdam, 1996), 119–191.
19. I.I. Gihman, and A.V. Skorohod, *Stochastic Differential Equations* (Springer-Verlag, New York-Heidelberg, 1972).
20. S.L. Heston, A closed-form solution for options with stochastic volatility with applications to Bond and Currency options, *The Review of Financial Studies* **6**(2), 327–343 (1993).
21. B. Jawerth and W. Sweldens, An overview of wavelet based multiresolution analyses, *SIAM Rev.* **36**, 377–412 (1994).
22. I. Karatzas, and S.E. Shreve, *Brownian Motion and Stochastic Calculus* (Springer Verlag, New York, 1991).
23. M. Pensky, and B. Vidakovic, Adaptive wavelet estimator for nonparametric density deconvolution, *Ann. Statist.* **27**, 2033–2053 (1999).
24. N. Shephard, Statistical aspects of ARCH and stochastic volatility, in: *Time Series Models in Econometrics, Finance and Other Fields*, ed. by D.R. Cox, D.V Hinkley and O.E. Barndorff-Nielsen (Chapman & Hall, London, 1996), 1–67.
25. A.V. Skorokhod, *Asymptotic Methods in the Theory of Stochastic Differential Equations*, Translations of Mathematical Monographs 78, (American Mathematical Society, 1989).
26. C.G. de Vries, On the relation between GARCH and stable processes, *J. Econometrics* **48**, 313–324 (1991).
27. G.G. Walter, *Wavelets and Other Orthogonal Systems with Applications* (CRC Press, Boca Raton, FL, 1994).
28. G.G. Walter and A.I. Zayed, Characterization of analytic functions in terms of their wavelet coefficients, *Complex Variables* **29**, 265–276 (1996).
29. M.P. Wand, Finite sample performance of deconvolving kernel density estimators, *Statist. Probab. Lett.* **37**, 131–139 (1998).
30. M.P. Wand, and M.C. Jones, *Kernel Smoothing* (Chapman and Hall, London, 1995)

31. J.B. Wiggins, Option valuation under stochastic volatility, *Journal of Financial Economics* **19**, 351–372 (1987).
32. J.H. van Zanten, and P. Zareba, A note on wavelet density deconvolution for weakly dependent data, *Stat. Inference Stoch. Process.* **11**, 207–219 (2008).