Journal of Multivariate Analysis 102 (2011) 683–697

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Estimation of a multivariate stochastic volatility density by kernel deconvolution

Bert Van Es, Peter Spreij *

Korteweg–de Vries Institute for Mathematics, Universiteit van Amsterdam, PO Box 94248, 1090GE Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 19 November 2009
Available online 14 December 2010

AMS subject classifications:
62G07
62M07
62P20

Keywords:
Stochastic volatility models
Multivariate density estimation
Kernel estimator
Deconvolution
Mixing

ABSTRACT

We consider a continuous time stochastic volatility model. The model contains a stationary volatility process. We aim to estimate the multivariate density of the finite-dimensional distributions of this process. We assume that we observe the process at discrete equidistant instants of time. The distance between two consecutive sampling times is assumed to tend to zero.

A multivariate Fourier-type deconvolution kernel density estimator based on the logarithm of the squared processes is proposed to estimate the multivariate volatility density. An expansion of the bias and a bound on the variance are derived.

1. Introduction

Let S denote the log price process of some stock in a financial market. It is often assumed that S can be modeled as the solution of a stochastic differential equation or, more generally, as an Itô diffusion process. So we assume that we can write

$$dS_t = b_t\, dt + \sigma_t\, dW_t, \qquad S_0 = 0, \tag{1}$$

or, in integral form,

$$S_t = \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s, \tag{2}$$

where W is a standard Brownian motion and the processes b and σ are assumed to satisfy certain regularity conditions (see [16]) to have the integrals in (2) well defined. In a financial context, the process σ is called the volatility process. In the literature, the process σ is often taken to be independent of the Brownian motion W.

In this paper we adopt this independence assumption and we furthermore assume that σ is a strictly stationary positive process satisfying a mixing condition, for example an ergodic diffusion on $(0, \infty)$. We will assume that all p-dimensional marginal distributions of σ have invariant densities with respect to the Lebesgue measure on $(0, \infty)^p$. This is typically the case in virtually all stochastic volatility models that are proposed in the literature, where the evolution of σ is modeled by a stochastic differential equation, mostly in terms of σ^2, or $\log \sigma^2$ (cf. e.g. [27,14]).

Therefore, think of X as $X = \sigma^2$ or $X = \log \sigma^2$, as a motivation for nonparametric estimation procedures, consider stochastic differential equations of the type

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

* Corresponding author.
E-mail address: spreij@uva.nl (P. Spreij).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.12.003

with B equal to Brownian motion. Focusing on the invariant univariate density of X_t, we recall that it is up to a multiplicative constant equal to

$$x \mapsto \frac{1}{a^2(x)} \exp\left(2\int_{x_0}^{x} \frac{b(y)}{a^2(y)}\,\mathrm{d}y\right), \tag{3}$$

where x_0 is an arbitrary element of the state space (l, r); see e.g. [12] or [21]. From formula (3) one sees that the invariant distribution of X may take on many different forms, as is the case for the various models that have been proposed in the literature. Refraining from parametric assumptions on the functions a and b, nonparametric statistical procedures may be used to obtain information about the shape of the (one-dimensional) invariant distribution.

A phenomenon that is often observed in practice, is *volatility clustering*. This means that for different time instants $t_1, \ldots, t_p$ that are close, the corresponding values of $\sigma_{t_1}, \ldots, \sigma_{t_p}$ 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 $(\sigma_{t_1}, \ldots, \sigma_{t_p})$ assumes high values. For instance, it is conceivable that for $p = 2$, 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 on the diagonal of the joint distribution, with the interpretation that clustering occurs around a low value ℓ or around a high value h.

Here is an example where this happens. We consider a regime switching volatility process. Assume that for $i = 0, 1$ we have two stationary processes X^i, each of them having multivariate invariant distributions having densities. Call these $f^i_{t_1,\ldots,t_p}(x_1, \ldots, x_p)$, whereas for $p = 1$ we simply write f^i. We assume these two processes to be independent, and also independent of a two-state homogeneous Markov chain U with states 0, 1. Let $Q(t)$ be the matrix of transition probabilities $q_{ij}(t) = P(X_t = i|X_0 = j)$. Let A be the matrix of transition intensities and write

$$A = \begin{pmatrix} -a_0 & a_1 \\ a_0 & -a_1 \end{pmatrix},$$

with $a_0, a_1 > 0$. Then $\dot{Q}(t) = AQ(t)$, and

$$Q(t) = \frac{1}{a_0 + a_1}\begin{pmatrix} a_1 + a_0\mathrm{e}^{-(a_0+a_1)t} & a_1 - a_1\mathrm{e}^{-(a_0+a_1)t} \\ a_0 - a_0\mathrm{e}^{-(a_0+a_1)t} & a_0 + a_1\mathrm{e}^{-(a_0+a_1)t} \end{pmatrix}.$$

The stationary distribution of U is given by $\pi_i := P(U_t = i) = \frac{a_{1-i}}{a_0+a_1}$ and we assume that U_0 has this distribution. We finally define the process ξ by

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

Then ξ is stationary too and it has a bivariate stationary distribution with a density, related by $P(\xi_s \in \mathrm{d}x, \xi_t \in \mathrm{d}y) = f_{s,t}(x, y)\,\mathrm{d}x\,\mathrm{d}y$. Elementary calculations lead to the following expression for $f_{s,t}$ for $0 < s < t$

$$f_{s,t}(x, y) = q_{11}(t-s)\pi_1 f^1_{s,t}(x, y) + q_{10}(t-s)\pi_0 f^0(x)f^1(y) + q_{01}(t-s)\pi_1 f^1(x)f^0(y) + q_{00}(t-s)\pi_0 f^0_{s,t}(x, y).$$

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

$$\mathrm{d}X_t^i = -a(X_t^i - \mu_i)\,\mathrm{d}t + b\,\mathrm{d}W_t^i,$$

with W^1, W^2 independent Brownian motions, $\mu_1 \neq \mu_2$ and $a > 0$. Suppose that the X^i start in their stationary $N(\mu_i, \frac{b^2}{2a})$ distributions. Then the center of the distribution of (X_s^i, X_t^i) is (μ_i, μ_i), whereas the center of the distribution of (X_s^0, X_t^1) is (μ_0, μ_1). Hence the density $f_{s,t}$ is a mixture of four hump shaped contours, each of them having a different center of location. If $t - s$ is small, this effectively reduces to mixture of distributions with centers (μ_1, μ_1) and (μ_2, μ_2). Similar qualitative observations can be made for models that switch between different GARCH regimes.

Nonparametric procedures are able to detect such a property of a bivariate distribution, and are consequently appropriate tools to get some partial insight into the behavior of the volatility. In the present paper we propose a nonparametric estimator for the multivariate density of the volatility process. Using ideas from deconvolution theory, we will propose a procedure for the estimation of this density at a number of fixed time instants. Related work on estimating a stationary univariate density of the volatility process has been done by Van Es et al. [22], Comte and Genon-Catalot [3], Van Zanten and Zareba [24], whereas a deconvolution approach has also been adopted to estimate a regression function for a discrete time stochastic volatility model by Franke et al. [8], Comte [1] and Comte et al. [2]. In [4] it is assumed that the volatility process solves a stochastic differential equation and nonparametric estimators for the drift and diffusion coefficients of that equation are studied. In most of these papers, one works with a simplified model with detrended log prices, which amounts to modeling the process S by Eq. (1) with a zero drift coefficient. In the present paper we adhere to this approach, which has become the tradition, as well.

The observations of the log-asset price S process are assumed to take place at the time instants $\Delta, 2\Delta, \ldots, n\Delta$, where the time gap satisfies $\Delta = \Delta_n \to 0$ and $n\Delta_n \to \infty$ as $n \to \infty$. This means that we base our estimator on the so-called *high frequency data*.

To assess the quality of our procedure, we will study how the bias and variance of the estimator behave under these assumptions. In [22] this problem has been studied for the marginal *univariate* density of σ. The multivariate study of the present paper largely builds on the approach of the cited paper, in particular we will rely not only on a number of technical results that are contained in it, but also we will borrow ideas from [23], where a multivariate problem for discrete time models has been studied. Nevertheless, we will encounter a number of technical problems that are not present in the univariate case, nor in the multivariate case for discrete time models.

The remainder of the paper is organized as follows. In Section 2, we give the heuristic arguments that motivate the definition of our estimator. In Section 3 the main results concerning the asymptotic behavior of the estimator are presented and discussed. Section 4 contains simulated examples in which the estimator is computed and whose behavior is qualitatively compared to that of another estimator. The proofs of the main theorems are given in Section 5. They are based on a number of technical lemmas, whose proofs are collected in Sections 6 and 7.

2. Construction of the estimator

As stated in the introduction, we consider (1) without the drift term, so we assume to have

$$\mathrm{d}S_t = \sigma_t \,\mathrm{d}W_t, \qquad S_0 = 0.$$

It is assumed that we observe the process S at the discrete time instants $0, \Delta, 2\Delta, \dots, n\Delta$. For $i = 1, 2, \dots$ we work, as in [9,10], with the normalized increments

$$X_i^\Delta = \frac{1}{\sqrt{\Delta}}(S_{i\Delta} - S_{(i-1)\Delta}) = \frac{1}{\sqrt{\Delta}}\int_{(i-1)\Delta}^{i\Delta} \sigma_t \,\mathrm{d}W_t. \tag{4}$$

For small Δ, we have the rough approximation (its precise meaning is not relevant at this stage, since we only develop the heuristics that eventually lead to the estimator to be presented below; but see (39) for a precise statement given the appropriate assumptions)

$$X_i^\Delta \approx \sigma_{(i-1)\Delta}\frac{1}{\sqrt{\Delta}}(W_{i\Delta} - W_{(i-1)\Delta}) = \sigma_{(i-1)\Delta} Z_i^\Delta, \tag{5}$$

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.

Let us first describe the univariate density estimator. Taking the logarithm of the square of X_i^Δ, we get from (5)

$$\log((X_i^\Delta)^2) \approx \log(\sigma^2_{(i-1)\Delta}) + \log((Z_i^\Delta)^2), \tag{6}$$

where the two terms on the right are independent. Assuming that the approximation is sufficiently accurate we can use this approximate convolution structure to estimate the unknown density f of $\log(\sigma^2_{i\Delta})$ from the observed $\log((X_i^\Delta)^2)$.

Before we can define the estimator, we need some more notation. Observe that the density of the 'noise' $\log(Z_i^\Delta)^2$, denoted by k, is given by

$$k(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{\frac{1}{2}x}\mathrm{e}^{-\frac{1}{2}\mathrm{e}^x}. \tag{7}$$

The characteristic function of the density k is denoted by ϕ_k and is given by $\phi_k(t) = \frac{1}{\sqrt{\pi}}\,2^{it}\,\Gamma(\frac{1}{2} + it)$; see [22].

In the present paper we will estimate the density $f(\mathbf{x}) = f_{t_1,\dots,t_p}(\mathbf{x})$, with $\mathbf{x} = (x_1, \dots, x_p)$, of a vector $(\log\sigma^2_{t_1}, \dots, \log\sigma^2_{t_p})$. Here the $0 < t_1 < \cdots < t_p$ denote p pre-specified time points. For clarity of exposition we first briefly outline the construction of an estimator of the univariate ($p = 1$) density f of $\log\sigma_t^2$, which is by assumed stationarity the same for all t.

Following a well-known approach in statistical deconvolution theory, we use a *deconvolution kernel density estimator*; see e.g. Section 6.2.4 of [26]. To that end we will use a kernel function w, satisfying certain regularity conditions to be explained further down (Condition 3.3). Having the characteristic functions ϕ_k and ϕ_w at our disposal, choosing a positive *bandwidth* h, we introduce the bandwidth dependent kernel function

$$v_h(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\phi_w(s)}{\phi_k(s/h)}\,\mathrm{e}^{-\mathrm{i}sx}\,\mathrm{d}s \tag{8}$$

and the density estimator of the univariate invariant density f of $\log\sigma_t^2$ given by

$$f_{nh}(x) = \frac{1}{nh}\sum_{j=1}^{n} v_h\left(\frac{x - \log((X_j^\Delta)^2)}{h}\right). \tag{9}$$

One easily verifies that the function v_h, and therefore also the estimator f_{nh}, is real valued. In [22] bias expansion and bounds on the variance of $f_{nh}(x)$ have been obtained.

In the present paper we will extend these results to a multivariate setting, in which we will estimate the density $f(\mathbf{x}) = f_{t_1,\dots,t_p}(\mathbf{x})$. In all that follows, we will adopt the convention to use boldface expressions for (random) vectors. The expression for the estimator of this density will be seen to be analogous to the estimator in the univariate case, that has been analyzed in [22], and exhibits some similarity with the estimator of a similar multivariate density in a discrete time framework as treated in [23]. The high frequency regime of the observations ($\Delta \to 0$) introduces non-trivial complications as compared to the discrete time framework of [23], whereas the estimation of a multivariate density also substantially complicates the analysis further as compared to the univariate estimation problem of [22].

What one ideally needs to estimate $f(\mathbf{x})$ are observations of p-dimensional random vectors that all have a density equal to f. This happens under the observation scheme that we have introduced previously, if the t_k are multiples of Δ, $t_k = i_k\Delta$ say. In that case, one should use $(X^\Delta_{t_1+j\Delta}, \dots, X^\Delta_{t_p+j\Delta})$ for all the values of j that are given by the observations. The first complicating factor is however that the t_k are not given as fixed multiples of Δ. On the other hand, if this would be the case, it would lead to an uninteresting estimation problem, as $\Delta \to 0$ in our setup, resulting in a degenerate density in the limit. Note that this kind of problem is not present, when one aims at estimating a univariate marginal density of $\log\sigma_t^2$, since all $\log\sigma_t^2$, $t > 0$ have the same marginal density.

We approach the problem as follows. Write $(i_1^\Delta, \dots, i_p^\Delta)$ for the vector $([t_1/\Delta], \dots, [t_p/\Delta])$ where $[.]$ denotes the floor function. We use $\mathbf{X}_j^\Delta$ to denote the random vectors of length p

$$\mathbf{X}_j^\Delta = (X_j^\Delta, \dots, X^\Delta_{i_p^\Delta - i_1^\Delta + j}), \quad j = 1, \dots, n - i_p^\Delta + i_1^\Delta.$$

Hence its kth component is $X^\Delta_{i_k^\Delta - i_1^\Delta + j}$, $k = 1, \dots, p$. Analogously, $\log((\mathbf{X}_j^\Delta)^2)$ denotes the vector

$$\log(\mathbf{X}_j^\Delta)^2 = (\log(X_j^\Delta)^2, \dots, \log(X^\Delta_{i_p^\Delta - i_1^\Delta + j})^2).$$

Anywhere else in what follows, we adhere to a similar notation. Functions of a vector are assumed to be evaluated componentwise, yielding again a vector.

Note that $\mathbf{X}_j^\Delta$ is, by virtue of (5), approximately equal to the vector

$$\tilde{\mathbf{X}}_j := (\sigma_{(j-1)\Delta} Z_j^\Delta, \dots, \sigma_{(i_p^\Delta - i_1^\Delta + j - 1)\Delta} Z^\Delta_{i_p^\Delta - i_1^\Delta + j}) \tag{10}$$

and that $(\log\sigma^2_{(j-1)\Delta}, \dots, \log\sigma^2_{(i_p^\Delta - i_1^\Delta + j-1)\Delta})$ has density equal to $f_{i_1^\Delta\Delta,\dots,i_p^\Delta\Delta}$ for every j, because of the assumed stationarity. Since $\Delta \to 0$, one can expect that $f_{i_1^\Delta\Delta,\dots,i_p^\Delta\Delta}(\mathbf{x}) \approx f_{t_1,\dots,t_p}(\mathbf{x})$. This motivates us to use the observations $\mathbf{X}_j^\Delta$, or rather the $\log(\mathbf{X}_j^\Delta)^2$, in the construction of a kernel estimator.

The kernel $\mathbf{w}$ that we will use in the multivariate case is just a product kernel, $\mathbf{w}(\mathbf{x}) = \prod_{j=1}^p w(x_j)$. Likewise we take $\mathbf{k}(\mathbf{x}) = \prod_{j=1}^p k(x_j)$ and the Fourier transforms $\phi_{\mathbf{w}}$ and $\phi_{\mathbf{k}}$ factorize as well. Let $\mathbf{v}_h$ be defined by

$$\mathbf{v}_h(\mathbf{x}) = \frac{1}{(2\pi)^p} \int_{\mathbb{R}^p} \frac{\phi_{\mathbf{w}}(\mathbf{s})}{\phi_{\mathbf{k}}(\mathbf{s}/h)} \mathrm{e}^{-\mathrm{i}\mathbf{s}\cdot\mathbf{x}}\, \mathrm{d}\mathbf{s}, \tag{11}$$

where $\mathbf{s} \in \mathbb{R}^p$ and $\cdot$ denotes inner product. Notice that we also have the factorization $\mathbf{v}_h(\mathbf{x}) = \prod_{j=1}^p v_h(x_j)$. The multivariate density estimator $\mathbf{f}_{nh}(\mathbf{x})$ that we will use to estimate $f(\mathbf{x})$ is given by

$$\mathbf{f}_{nh}(\mathbf{x}) = \frac{1}{(n - i_p^\Delta + i_1^\Delta)h^p} \sum_{j=1}^{n - i_p^\Delta + i_1^\Delta} \mathbf{v}_h\left(\frac{\mathbf{x} - \log((\mathbf{X}_j^\Delta)^2)}{h}\right). \tag{12}$$

Note that this estimator bears some similarity to, but also differs from the corresponding one for a discrete time model in [23], where the multivariate density of $(\sigma_{t+1}, \dots, \sigma_{t+p})$ at consecutive time points is the object under study for a discrete time model. Let us explain the most important differences with [23]. In that paper one works with discrete time models, whereas the focus of the present paper is on *continuous time* models, that are discretely observed under a sampling regime in which the sampling interval tends to zero, i.e. high frequency observations. We emphasize that this situation cannot occur in a discrete time setup. The high frequency assumption introduces an additional problem in that the consecutive observations $S_{i\Delta}$ will exhibit a stronger dependence when $\Delta \to 0$. This introduces additional technical complications as compared to the discrete time setup of [23], that cannot be tackled with the techniques used in that paper. The problem of estimating a *multivariate* density is another source of technical complications as compared to univariate density estimation treated in [22]. These will become apparent from the proofs in Section 5, although this more complex situation should already have become clear from the construction of the estimator as outlined above.

As a final note in this section we briefly discuss marginalization. Under the condition that the integral of v_h is equal to one, an estimator of $f_{(t_1,\dots,t_{p-1})}(x_1, \dots, x_{p-1})$ is obtained by integrating out the variable x_p in (12). The resulting expression resembles, but is slightly different from, the analog of (12) for a $(p-1)$-dimensional density. To describe the precise result, let $\check{\mathbf{x}} = (x_1, \ldots, x_{p-1})$, $\check{\mathbf{v}}_h = \prod_{i=1}^{p-1} v_h(x_i)$, $\check{\mathbf{X}}_j^\Delta = (X_j^\Delta, \ldots, X^\Delta_{i^\Delta_{p-1} - i^\Delta_1 + j})$. Then the estimator of $f(\check{\mathbf{x}})$ obtained by integrating $\mathbf{f}_{nh}(\mathbf{x})$ over x_p becomes

$$\mathbf{f}_{nh}(\check{\mathbf{x}}) := \frac{1}{(n - i_p^\Delta + i_1^\Delta)h^p} \sum_{j=1}^{n-i_p^\Delta+i_1^\Delta} \check{\mathbf{v}}_h\left(\frac{\check{\mathbf{x}} - \log((\check{\mathbf{X}}_j^\Delta)^2)}{h}\right),$$

which indeed slightly differs from $\mathbf{f}_{n-1,h}(\check{\mathbf{x}})$. Further integration over the variables $x_2, \ldots, x_{p-1}$ then reduces this estimator to the estimator of the univariate density given by (9) upon the substitution of n by $n - i_p^\Delta + i_1^\Delta$.

Let us emphasize that v_h is in general not an L^1-function, so care has to be taken with the integration above. The integral of v_h has to be understood as an improper Riemann integral, or even as its principal value ($\lim_{N\to\infty} \int_{-N}^N v_h(x)\,\mathrm{d}x$). However, if one of these concepts of the integral is well defined, the resulting value of the integral of v_h is equal to 1.

3. Results

To derive the asymptotic behavior of the estimator, we need a mixing condition on the process σ. For the sake of clarity, we recall the basic definitions. For a certain process X let $\mathcal{F}_a^b$ be the σ-algebra of events generated by the random variables $X_t,\ a \le t \le b$. The mixing coefficient $\alpha(t)$ is defined by

$$\alpha(t) = \sup_{A\in\mathcal{F}_{-\infty}^0,\ B\in\mathcal{F}_t^\infty} |P(A\cap B) - P(A)P(B)|. \tag{13}$$

The process X is called *strongly mixing* if $\alpha(t) \to 0$ as $t \to \infty$.

As we mentioned in the introduction, it is common practice to model the volatility process $V = \sigma^2$ as the stationary, ergodic solution of an SDE of the form

$$\mathrm{d}V_t = b(V_t)\,\mathrm{d}t + a(V_t)\,\mathrm{d}B_t.$$

It is easily verified that for such processes it holds that $\mathrm{E}\,|V_t - V_0| = O(t^{1/2})$, provided that $b \in L_1(\mu)$ and $a \in L_2(\mu)$, where μ is the invariant probability measure. Indeed we have $\mathrm{E}\,|V_t - V_0| \le \mathrm{E}\int_0^t |b(V_s)|\,\mathrm{d}s + (\mathrm{E}\int_0^t a^2(V_s)\,\mathrm{d}s)^{1/2} = t\|b\|_{L_1(\mu)} + \sqrt{t}\|a\|_{L_2(\mu)}$. In this setup, the process V is strongly mixing; see for instance Corollary 2.1 of [11]. Although we will not assume explicitly that σ^2 solves an SDE, the above observations motivate the following condition.

Condition 3.1. The process σ is stationary and is

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

$$\int_0^\infty \alpha(t)^q\,\mathrm{d}t < \infty \tag{14}$$

(iii) independent of the Brownian motion W.

Remark 3.2. Since the mixing coefficients $\alpha(t)$ are non-increasing in t, condition (14) is equivalent to the following. For all $t \in \mathbb{R}$ there exists $C(q,t)$ such that for all $\Delta > 0$

$$\sum_{k=0}^\infty \alpha(k\Delta + t)^q \le \frac{C(q,t)}{\Delta}, \tag{15}$$

where $\alpha(t)$ is set equal to 1 for $t \le 0$.

The kernel function w is subject to the following condition.

Condition 3.3. Let w be a real symmetric function with real valued symmetric Fourier transform ϕ_w ($\phi_w(u) = \int_{-\infty}^\infty \mathrm{e}^{\mathrm{i}ux} w(x)\,\mathrm{d}x$) having support $[-1,1]$. Assume further that

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

One example of such a kernel is

$$w(x) = \frac{48x(x^2-15)\cos x - 144(2x^2-5)\sin x}{\pi x^7}, \tag{16}$$

whose characteristic function is $\phi_w(t) = (1-t^2)^3$, $|t| \le 1$. For other examples of such kernels see [25].
Our main theorems are multivariate generalizations of results in [22] which describe the asymptotic behavior of the univariate density estimator.

Theorem 3.4. *Let the kernel function w satisfy Condition 3.3. Let the density $f_{t_1,\ldots,t_p}(\mathbf{x})$ of $(\log\sigma_{t_1}^2, \ldots, \log\sigma_{t_p}^2)$ be continuous, twice continuously differentiable with a bounded second derivative and Lipschitz in $t_1, \ldots, t_p$, uniformly in $\mathbf{x}$. Assume that the* *first of Condition 3.1 holds and that the invariant density of σ_t^2 is bounded in a neighborhood of zero. Suppose that $\Delta = n^{-\delta}$ for given $0 < \delta < 1$ and choose $h = \gamma\pi / \log n$, where $\gamma > 4p/\delta$. Then the bias of the estimator (9) satisfies*

$$\mathrm{E}\,\mathbf{f}_{nh}(\mathbf{x}) = f_{t_1,\ldots,t_p}(\mathbf{x}) + \frac{1}{2}h^2 \int \mathbf{u}^\top \nabla^2 f(\mathbf{x})\mathbf{u}\,\mathbf{w}(\mathbf{u})\,\mathrm{d}\mathbf{u} + o(h^2) + O(\Delta), \tag{17}$$

where $\nabla^2 f(\mathbf{x})$ denotes the Hessian of f at $\mathbf{x}$.

Theorem 3.5. *Let the kernel function w satisfy Condition 3.3. Assume that Condition 3.1 holds, that $\int |w(u)|^{2/(1-q)}\,\mathrm{d}u < \infty$, where q is as in (14), and that the invariant density of σ_t^2 is bounded in a neighborhood of zero. Suppose that $\Delta = n^{-\delta}$ for given $0 < \delta < 1$ and choose $h = \gamma\pi / \log n$, where $\gamma > 4p/\delta$. The variance of the estimator satisfies*

$$\mathrm{Var}\,\mathbf{f}_{nh}(\mathbf{x}) = O\left(\frac{1}{n}h^{2p\rho}\mathrm{e}^{p\pi/h}\right) + O\left(\frac{1}{nh^{(1+q)p}\Delta}\right). \tag{18}$$

Corollary 3.6. *Under the assumptions of Theorems 3.4 and 3.5 the bias satisfies $\gamma^2\pi^2(\log n)^{-2}(1+o(1))$ and the order of the variance is $n^{-1+\delta}(\log n)^{p(1+q)}$. Hence the mean squared error of the estimator $\mathbf{f}_{nh}(\mathbf{x})$ is of order $(\log n)^{-4}$.*

Proof. The choices $\Delta = n^{-\delta}$, with $0 < \delta < 1$ and $h = \gamma\pi / \log n$, with $\gamma > 4p/\delta$ render a variance that is of order $n^{-1+p/\gamma}(1/\log n)^{2p\rho}$ for the first term of (18) and $n^{-1+\delta}(\log n)^{p(1+q)}$ for the second term. Since by assumption $\gamma > 4p/\delta$ we have $1/\gamma < \delta/4p < \delta$ so the second term dominates the first term. The order of the variance is thus $n^{-1+\delta}(\log n)^{p(1+q)}$. 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 also logarithmic. □

The proof of the theorems are deferred to the next section. We conclude the present section by a number of comments on the result.

Remark 3.7. We observe some features that parallel some findings for the univariate case. The expectation of the deconvolution estimator is, apart from the $O(\Delta)$-term equal to the expectation of an ordinary kernel density estimator, as becomes clear from the proof of Lemma 5.1. From that proof we also see that this term is due to estimation of a multivariate density, in the univariate case, it vanishes. It is well known that the variance of kernel-type deconvolution estimators heavily depends on the rate of decay to zero of $|\phi_k(t)|$ as $|t| \to \infty$. The faster the decay the larger the asymptotic variance. This follows for instance for i.i.d. observations from results in [7] and for stationary observations from the work of [17]. The rate of decay of $|\phi_k(t)|$ for the density (7) is given by $|\phi_k(t)| = \sqrt{2}\,\mathrm{e}^{-\frac{1}{2}\pi|t|}(1 + O(\frac{1}{|t|}))$; see Lemma 5.3 in [22]. This shows that k is supersmooth; cf. [7]. By the similarity of the tail of this characteristic function to the tail of a Cauchy characteristic function we can expect the same order of the mean squared error as in Cauchy deconvolution problems, where it decreases logarithmically in n; cf. [7] for results on i.i.d. observations. Note that this rate, however slow, is faster than the one for normal deconvolution. Fan [7] also shows that we cannot expect anything better. Note also that in the first term of the order bound on the variance in (18) the decay rate of $|\phi_k|$ can be recognized. It is a consequence of the asymptotic behavior of the function $\gamma_0(h)$ of (25) as given in Lemma 6.1.

Remark 3.8. The first order bound for the variance coincides with the order bound for the variance of the multivariate density estimator in discrete time models under the assumption that the volatility process and the error process are independent; see Theorem 3.2 in [23]. The second order bound is of the same nature as in the case of estimating a univariate density in continuous time models, see Theorem 3.1 in [22], the difference being that in the multivariate case of the present paper one has $h^{p(1+q)}$ in the denominator instead of h^{1+q}.

Remark 3.9. The rate of convergence $(\log n)^{-4}$ for the mean squared error as in Corollary 3.5 has also been found for other estimators. Comte and Genon-Catalot [3] use (penalized) projection estimators for f. These estimators are obtained by computing certain projections on large, growing with n, but finite-dimensional subspaces of $L^2(\mathbb{R})$. Under similar assumptions as ours, they also find the rate of convergence $(\log n)^{-4}$. By sharpening the assumed smoothness properties of f, i.e. fast enough exponential decay of the characteristic function of f, so that f itself is a supersmooth density, they were able to obtain rates that are essentially negative powers of n.

Van Zanten and Zareba [24] consider wavelet estimators of the density of the accumulated squared volatility over intervals of length Δ with Δ fixed for the model without drift and with the same observation scheme. Under similar conditions, they found this rate for the supremum of the mean integrated squared error, the supremum taken over densities in some Sobolev ball. For densities satisfying stronger smoothness conditions, their estimators obtained better rates, albeit still negative powers of $\log n$. Both papers deal with estimating a univariate density only.

Franke et al. [8] consider a discrete time model, where the evolution of $\log\sigma_t$ is described by a nonlinear autoregression. By adopting a deconvolution approach they estimate the unknown regression function and establish tightness of the normalized estimators, where the normalization again corresponds to the rate that we found.

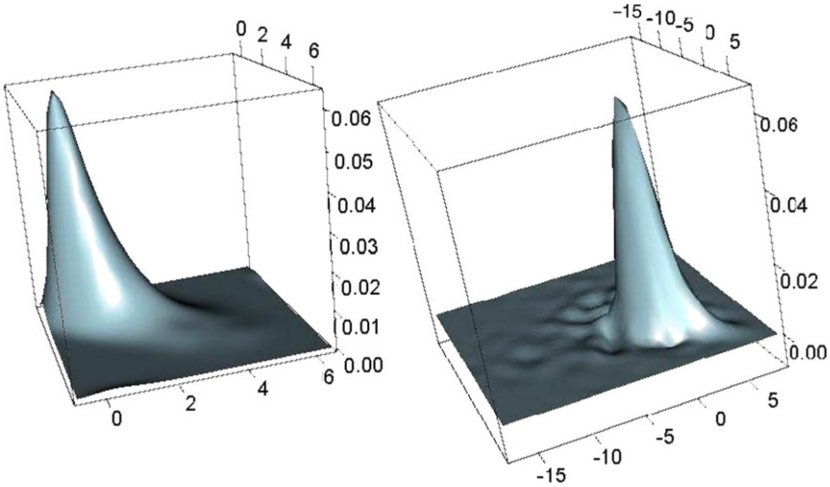

Fig. 1. Left: estimator based on direct observations, Right: deconvolution estimator.

Remark 3.10. Better bounds on the asymptotic variance can be obtained under stronger mixing conditions. Consider for instance *uniform mixing*. In this case the mixing coefficient $\varphi(t)$ is defined for $t > 0$ as

$$\varphi(t) = \sup_{A\in\mathcal{F}_{-\infty}^{0},\, B\in\mathcal{F}_t^{\infty}} |P(A|B) - P(A)| \tag{19}$$

and a process is called *uniform mixing* if $\varphi(t) \to 0$ for $t \to \infty$. Obviously, uniform mixing implies strong mixing. As a matter of fact, one has the relation

$$\alpha(t) \le \frac{1}{2}\varphi(t).$$

See [6] for this inequality and many other mixing properties. If σ is uniform mixing with coefficient φ satisfying $\int_0^\infty \varphi(t)^{1/2}\,\mathrm{d}t < \infty$, then the variance bound is given by

$$\operatorname{Var} f_{nh}(x) = O\left(\frac{1}{n}\, h^{2p\rho} \mathrm{e}^{p\pi/h}\right) + O\left(\frac{1}{nh^p\Delta}\right). \tag{20}$$

The proof of this bound runs similarly to the strong mixing bound. The essential difference is that in Eq. (58) one can use Theorem 17.2.3 of [15] with $\tau = 0$ instead of Deo's [5] lemma, as in the proof of Theorem 2 in [18].

4. Numerical examples

In the present section we evaluate the quality of the deconvolution estimator (12) at hand of two simulated examples. In the first example, the underlying stochastic volatility model is of continuous GARCH type, much in the spirit of [19]. For the simulations we used a discretized model (based on Euler discretization, with a discretization step normalized to 1) that amounts to an ordinary GARCH(1, 1) model. Specifically, the discretized model is given by

$$X_t = \sigma_t Z_t \tag{21}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{22}$$

where X_t stands for detrended log price, the Z_t are i.i.d. standard normal random variables, and the parameters $\alpha_0, \alpha_1, \beta$ are positive. Note that no closed form expression is available for the stationary bivariate densities of σ^2, not even for the univariate densities. We have simulated (using R) this model with the parameter values $\alpha_0 = 1.0$, $\alpha_1 = 0.7$, $\beta = 0.2$. Since the simulations also yield the values of the volatility process, we have been able to estimate the bivariate stationary density of $(\log\sigma_t^2, \log\sigma_{t-1}^2)$ *directly* by ordinary kernel estimation methods, using the kernel given in (16) and a bandwidth of $h = 0.4$. Alternatively, mimicking a realistic practical situation, we used the deconvolution estimator to estimate this bivariate density, based solely on the observations X_t. Using the same kernel and bandwidth, we have computed the estimated bivariate density estimator, using some routines based on Fast Fourier Transforms. The results for both methods are shown in Fig. 1. We conclude that the deconvolution estimators show a behavior that agrees well with that of the ordinary density estimator.

In a second example, we used a regime switching GARCH(1, 1) process. That is, we keep on having Eq. (21), but for σ_t^2 we take a process that switches according to a two-state Markov chain with all transition probabilities equal to 0.5, between two processes of type (22). The first one is 0.1 times the process given by (22) with the same parameter values, whereas for the second one we used (22) with the values $\alpha_0 = 2.0$, $\alpha_1 = 0.7$, $\beta = 0.2$. As for the previous, we compared the bivariate

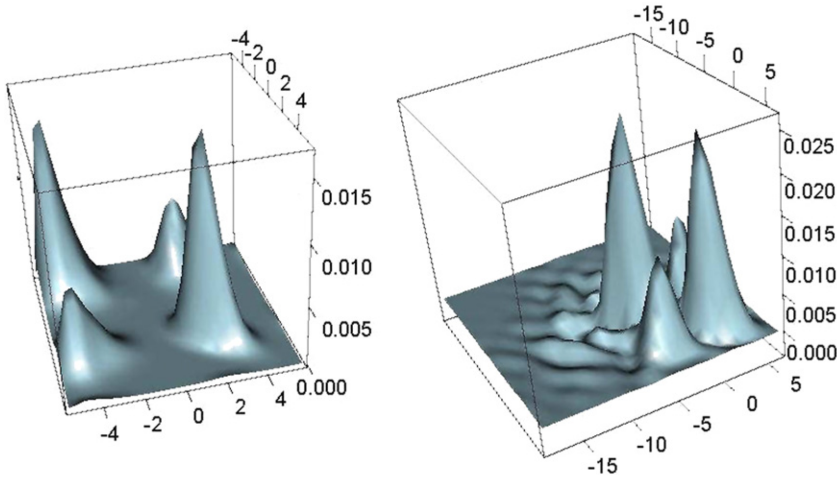

Fig. 2. Left: estimator based on direct observations, Right: deconvolution estimator.

density estimators, one directly based on the values of σ_t^2 and the second one on the convolution density estimator. From the graphs depicted in Fig. 2 we conclude that also in this case the two estimators agree rather well. Note the four peaks of the bivariate density estimators.

5. Proof of the theorems

Let $\mathcal{F}_\sigma$ denote the sigma field generated by the process σ. For $j = 1, \ldots, n - i_p^\Delta + i_1^\Delta$ we introduce, along with the $\tilde{\mathbf{X}}_j$ of (10), the following vector notation

$$\boldsymbol{\sigma}_j = (\sigma_{(j-1)\Delta}, \ldots, \sigma_{(i_p^\Delta - i_1^\Delta + j - 1)\Delta})$$
$$\mathbf{Z}_j^\Delta = (Z_j^\Delta, \ldots, Z^\Delta_{i_p^\Delta - i_1^\Delta + j}),$$

so that $\tilde{\mathbf{X}}_j$ equals the Hadamard product $\boldsymbol{\sigma}_j \circ \mathbf{Z}_j^\Delta$. Note that since the σ process is defined on the whole real line the $\boldsymbol{\sigma}$ vectors are actually well defined for all j.

Let $\tilde{\mathbf{f}}_{nh}$ denote the 'estimator' based on the approximating (but unobserved) random vectors $\tilde{\mathbf{X}}_j$, i.e.

$$\tilde{\mathbf{f}}_{nh}(\mathbf{x}) = \frac{1}{(n - i_p^\Delta + i_1^\Delta)h^p} \sum_{j=1}^{n - i_p^\Delta + i_1^\Delta} \mathbf{v}_h \left(\frac{\mathbf{x} - \log((\tilde{\mathbf{X}}_j^\Delta)^2)}{h} \right). \tag{23}$$

The proof of (17) is partly based on the following two lemmas, whose proofs are given in the next section. The first one deals with the expectation of $\tilde{\mathbf{f}}_{nh}$.

Lemma 5.1. *Let the density $f_{t_1,\ldots,t_p}(\mathbf{x})$ of $(\log \sigma_{t_1}^2, \ldots, \log \sigma_{t_p}^2)$ be Lipschitz in $t_1, \ldots, t_p$, uniformly in $\mathbf{x}$. Then*

$$\mathrm{E}\,\tilde{\mathbf{f}}_{nh}(\mathbf{x}) = \frac{1}{h^p} \int \cdots \int \mathbf{w} \left(\frac{\mathbf{x} - \mathbf{u}}{h} \right) f_{t_1,\ldots,t_p}(\mathbf{u})\, \mathrm{d}\mathbf{u} + O(\Delta). \tag{24}$$

Notice that, apart from the $O(\Delta)$ term, the equality (24) is the same as that for ordinary multivariate kernel estimators; see for instance [13,20].

The second lemma estimates the expected difference between f_{nh} and $\tilde{f}_{nh}$. The bound is in terms of the functions

$$\gamma_0(h) = \frac{1}{2\pi} \int_{-1}^{1} \left| \frac{\phi_w(s)}{\phi_k(s/h)} \right| \mathrm{d}s \tag{25}$$

and

$$\gamma_1(h, x) = \mathrm{e}^{\frac{1}{2}\pi/h} + \frac{1}{h} \exp\left(\frac{\pi}{2} \frac{1 + \pi/|x|}{h} \right) \log \frac{1 + \pi/|x|}{h}. \tag{26}$$

Lemma 5.2. *Assume Condition 3.3 and that the first of Condition 3.1 holds and that the invariant density of σ_t^2 is bounded in a neighborhood of zero. For $h \to 0$ and ε small enough we have*

$$|\mathrm{E}\,\mathbf{f}_{nh}(\mathbf{x}) - \mathrm{E}\,\tilde{\mathbf{f}}_{nh}(\mathbf{x})| = O\left(\frac{1}{h^{p+1}} \gamma_0(h)^p \frac{\Delta^{1/4}}{\varepsilon} + \frac{1}{h^p} \gamma_0(h)^p \frac{\Delta^{1/2}}{\varepsilon^2} + \frac{1}{h^{p-1}} \gamma_0(h)^{p-1} \gamma_1(h, |\log 2\varepsilon|/h) \frac{\varepsilon}{|\log 2\varepsilon|} \right).$$

Proof of Theorem 3.4. Statement (17) follows by combining standard arguments of kernel density estimation applied to expression (24) in Lemma 5.1 with Lemma 5.2. We will now show that the bound in Lemma 5.2 is essentially a negative power of n, whereas h^2 is of logarithmic order. Recall that we have assumed that $\delta > 4p/\gamma$. It follows that $p/2\gamma < \delta/4 - p/2\gamma$, so we can pick a $\beta \in (p/2\gamma, \delta/4 - p/2\gamma)$ and take $\varepsilon = n^{-\beta}$. By Lemmas 6.1 and 6.3, up to factors that are logarithmic in n, the order of $|\mathrm{E}\,\mathbf{f}_{nh}(\mathbf{x}) - \mathrm{E}\,\tilde{\mathbf{f}}_{nh}(\mathbf{x})|$ is then

$$n^{\frac{p}{2\gamma}-\frac{1}{4}\delta+\beta} + n^{\frac{p}{2\gamma}+2\beta-\frac{\delta}{2}} + n^{\frac{p}{2\gamma}-\beta}, \tag{27}$$

which is negligible to $h^2 = \gamma^2\pi^2/(\log n)^2$ for the chosen values of the parameters. □

To prove the bound (18) we use the two lemmas below, which are proved in the next section. First consider the variance of $\tilde{\mathbf{f}}_{nh}(\mathbf{x})$.

Lemma 5.3. *Assume Condition 3.3 and assume the second of Condition 3.1(ii). Assume also $\int |w(u)|^{2/(1-q)}\mathrm{d}u < \infty$ for the same q and $n\Delta \to \infty$. We have, for $h \to 0$,*

$$\mathrm{Var}\,\tilde{\mathbf{f}}_{nh}(\mathbf{x}) = O\left(\frac{1}{n}h^{2p\rho}\mathrm{e}^{p\pi/h}\right) + O\left(\frac{1}{nh^{(1+q)p}\Delta}\right). \tag{28}$$

The next lemma estimates $\mathrm{Var}\,(f_{nh}(x) - \tilde{f}_{nh}(x))$.

Lemma 5.4. *Assume that Conditions 3.1 and 3.3 hold and let σ_t^2 have a bounded density in a neighborhood of zero. We have, for $h \to 0$ and $\varepsilon > 0$ small enough,*

$$\mathrm{Var}\,(\mathbf{f}_{nh}(\mathbf{x}) - \tilde{\mathbf{f}}_{nh}(\mathbf{x})) = O\left(\frac{1}{nh^{2p+2}}\gamma_0(h)^{2p}\frac{\Delta^{1/2}}{\varepsilon^2} + \frac{1}{nh^{2p-2}}\gamma_0(h)^{2p-2}\,\gamma_1(h, |\log 2\varepsilon|/h)^2\frac{\varepsilon}{|\log 2\varepsilon|^2}\right) \tag{29}$$

$$+\frac{1}{nh^{2p}\Delta}O\left(\frac{\Delta^{(1-q)/2}}{h^2\varepsilon^2} + \varepsilon^{1-q}\right). \tag{30}$$

Remark 5.5. For $p = 1$, the order bounds of Lemma 5.4 reduce to those of Lemma 4.3 in [22].

Proof of Theorem 3.5. The bound of (18) follows as soon as we show that the estimate in Lemma 5.4 is of lower order than the one in Lemma 5.3. Up to terms that are logarithmic in n, the bound in Lemma 5.3 is of order $n^{\delta-1}$. Choosing again $\varepsilon = n^{-\beta}$, by Lemmas 6.1 and 6.3, one finds that, up to logarithmic factors, the order of $\mathrm{Var}\,(\mathbf{f}_{nh}(\mathbf{x}) - \tilde{\mathbf{f}}_{nh}(\mathbf{x}))$ is

$$n^{-1+\frac{p}{\gamma}-\frac{\delta}{2}+2\beta} + n^{-1+\frac{p}{\gamma}-\beta} + n^{-1+2\beta+\frac{1+q}{2}\delta} + n^{-1+\delta-\beta(1-q)}. \tag{31}$$

Recall our assumption $\delta\gamma > 4p$. If we pick β less than $\frac{1}{4}\delta(1-q)$, then all these terms are indeed of lower order than $n^{\delta-1}$. □

6. Some technical results

We need expansions and order estimates for the functions ϕ_k, the kernel v_h as defined in (8), γ_0 as defined in (25) and the function γ_1 as defined in (26). These are collected in the next technical lemmas, that are partially taken from [22,23].

Lemma 6.1. *Assume Condition 3.3. For $h \to 0$ we have*

$$\gamma_0(h) = O(h^{1+\rho}\mathrm{e}^{\frac{1}{2}\pi/h}). \tag{32}$$

Proof. See the proof of Lemma 5.3 in [22]. □

Lemma 6.2. *Assume Condition 3.3. The functions v_h and $\mathbf{v}_h$ are bounded and Lipschitz. More precisely, for all x we have $|v_h(x)| \le \gamma_0(h)$ and for all x and u $|v_h(x+u) - v_h(x)| \le \gamma_0(h)\,|u|$. For all p-vectors $\mathbf{x}$ we have*

$$|\mathbf{v}_h(\mathbf{x})| \le \gamma_0(h)^p \tag{33}$$

and for all p-vectors $\mathbf{x}$ and $\mathbf{u}$

$$|\mathbf{v}_h(\mathbf{x}+\mathbf{u}) - \mathbf{v}_h(\mathbf{x})| \le \gamma_0(h)^p\sum_{j=1}^p |u_j| \tag{34}$$

and for some $C > 0$,

$$|\mathbf{w}(\mathbf{x}+\mathbf{u}) - \mathbf{w}(\mathbf{x})| \le C\sum_{j=1}^p |u_j|. \tag{35}$$

Proof. The results for $|v_h(\cdot)|$ are known from Lemma 5.4 in [22]. The bound (33) follows by the product structure of $\mathbf{v}_h$. Inequality (34) follows by induction and the same technique can be used to prove inequality (35). □

Lemma 6.3. *Assume Condition 3.3. For $x \to \infty$ we have the following estimate on the behavior of v_h. For some positive constant D it holds that*

$$|v_h(x)| \le D\frac{\gamma_1(h,x)}{|x|} \quad \text{as } |x| \to \infty, \tag{36}$$

and

$$\gamma_1(h,x) = O\left(\frac{|\log h|}{h}\,\mathrm{e}^{\frac{1}{2}\pi(1+\pi/|x|)/h}\right) \quad \text{as } h \to 0. \tag{37}$$

Let $x^* = \max\{|x_1|,\ldots,|x_p|\}$. We have the following estimate on the behavior of $\mathbf{v}_h$. For some positive constant D it holds that, if x^* tends to infinity,

$$|\mathbf{v}_h(\mathbf{x})| \le D\gamma_0(h)^{p-1}\frac{\gamma_1(h,x^*)}{x^*}. \tag{38}$$

Proof. The estimates of (36) and (37) are taken from Lemma 5.5 of [22]. To show (38), we argue as follows. Without loss of generality we may assume that $x^* = x_p > 0$. Use the bound on γ_0 of Lemma 6.2 and the bound in (36) to get $|\mathbf{v}_h(\mathbf{x})| = \prod_{i=1}^{p-1} v_h(x_i)v_h(x_p) \le D\gamma_0(h)^{p-1}\gamma_1(h,x_p)/x_p = D\gamma_0(h)^{p-1}\gamma_1(h,x^*)/x^*$. □

7. Proof of Lemmas 5.1–5.4

Recall that $\mathcal{F}^\sigma$ is the σ-algebra generated by the process σ.

Proof of Lemma 5.1. By Condition 3.1(iii),

$$\begin{aligned}
\mathrm{E}\,(\tilde{\mathbf{f}}_{nh}(x)|\mathcal{F}_\sigma) &= \frac{1}{(n-i_p^\Delta+i_1^\Delta)h^p}\sum_{j=1}^{n-i_p^\Delta+i_1^\Delta}\mathrm{E}\left(\mathbf{v}_h\left(\frac{\mathbf{x}-\log\boldsymbol{\sigma}_j^2-\log(\mathbf{Z}_j^\Delta)^2}{h}\right)\middle|\,\mathcal{F}_\sigma\right)\\
&= \frac{1}{(n-i_p^\Delta+i_1^\Delta)h^p}\sum_{j=1}^{n-i_p^\Delta+i_1^\Delta}\frac{1}{(2\pi)^p}\int\cdots\int\frac{\phi_{\mathbf{w}}(\mathbf{s})}{\phi_{\mathbf{k}}(\mathbf{s}/h)}\,\mathrm{E}\,(\mathrm{e}^{-\mathrm{i}\mathbf{s}\cdot(\mathbf{x}-\log\boldsymbol{\sigma}_j^2-\log(\mathbf{Z}_j^\Delta)^2)/h}|\mathcal{F}_\sigma)\,\mathrm{d}\mathbf{s}\\
&= \frac{1}{(n-i_p^\Delta+i_1^\Delta)h^p}\sum_{j=1}^{n-i_p^\Delta+i_1^\Delta}\frac{1}{(2\pi)^p}\int\cdots\int\frac{\phi_{\mathbf{w}}(\mathbf{s})}{\phi_{\mathbf{k}}(\mathbf{s}/h)}\,\mathrm{e}^{-\mathrm{i}\mathbf{s}\cdot\mathbf{x}/h}\mathrm{e}^{\mathrm{i}\mathbf{s}\cdot\log\boldsymbol{\sigma}_j^2/h}\,\phi_{\mathbf{k}}(\mathbf{s}/h)\,\mathrm{d}\mathbf{s}\\
&= \frac{1}{(n-i_p^\Delta+i_1^\Delta)h^p}\sum_{j=1}^{n-i_p^\Delta+i_1^\Delta}\frac{1}{(2\pi)^p}\int\cdots\int\phi_{\mathbf{w}}(\mathbf{s})\,\mathrm{e}^{-\mathrm{i}\mathbf{s}\cdot(\mathbf{x}-\log\boldsymbol{\sigma}_j^2)/h}\,\mathrm{d}\mathbf{s}\\
&= \frac{1}{(n-i_p^\Delta+i_1^\Delta)h^p}\sum_{j=1}^{n-i_p^\Delta+i_1^\Delta}\mathbf{w}\left(\frac{\mathbf{x}-\log\boldsymbol{\sigma}_j^2}{h}\right).
\end{aligned}$$

By taking the expectation and exploiting the stationarity of σ we get, also using $|(i_j^\Delta-1)\Delta-t_j| \le 2\Delta$, for $j=1,\ldots,p$, and the uniform Lipschitz continuity of f

$$\begin{aligned}
\mathrm{E}\,\tilde{\mathbf{f}}_{nh}(\mathbf{x}) &= \mathrm{E}\,\tilde{\mathbf{f}}^\sigma_{nh}(x) = \mathrm{E}\,\frac{1}{h^p}\mathrm{E}\,\mathbf{w}\left(\frac{\mathbf{x}-\log\boldsymbol{\sigma}_0^2}{h}\right)\frac{1}{h^p}\int\cdots\int\mathbf{w}\left(\frac{\mathbf{x}-\mathbf{u}}{h}\right)f_{(i_1^\Delta-1)\Delta,\ldots,(i_p^\Delta-1)\Delta}(\mathbf{u})\,\mathrm{d}\mathbf{u}\\
&= \frac{1}{h^p}\int\cdots\int\mathbf{w}\left(\frac{\mathbf{x}-\mathbf{u}}{h}\right)f_{t_1,\ldots,t_p}(\mathbf{u})\,\mathrm{d}\mathbf{u} + \frac{1}{h^p}\int\cdots\int\mathbf{w}\left(\frac{\mathbf{x}-\mathbf{u}}{h}\right)(f_{(i_1^\Delta-1)\Delta,\ldots,(i_p^\Delta-1)\Delta}(\mathbf{u})-f_{t_1,\ldots,t_p}(\mathbf{u}))\,\mathrm{d}\mathbf{u}\\
&= \frac{1}{h^p}\int\cdots\int\mathbf{w}\left(\frac{\mathbf{x}-\mathbf{u}}{h}\right)f_{t_1,\ldots,t_p}(\mathbf{u})\,\mathrm{d}\mathbf{u} + O(\Delta). \quad \square
\end{aligned}$$

For the proof of Lemma 5.2 we need two properties of the process σ, valid under Condition 3.1(i). There exists a constant $C > 0$ such that

$$\mathrm{E}\,(X_1^\Delta - \sigma_0 Z_1^\Delta)^2 \le C\Delta^{1/2} \quad \text{for } \Delta \to 0, \tag{39}$$

and

$$\mathrm{E}\left|\frac{1}{\Delta}\int_0^\Delta \sigma_t^2\,\mathrm{d}t - \sigma_0^2\right| \le C\Delta^{1/2} \quad \text{for } \Delta \to 0. \tag{40}$$

We only show Eq. (39), because the proof of (40) does not involve additional arguments. We will use the basic inequality $(x-y)^2 \le x^2 - y^2$, valid for $x \ge y \ge 0$. Since $\sqrt{\Delta}(X_1^\Delta - \sigma_0 Z_1^\Delta) = \int_0^\Delta (\sigma_t - \sigma_0)\,\mathrm{d}W_t$, we have

$$
\begin{aligned}
\Delta \mathrm{E}\,(X_1^\Delta - \sigma_0 Z_1^\Delta)^2 &= \mathrm{E}\int_0^\Delta (\sigma_t - \sigma_0)^2\,\mathrm{d}t \\
&\le \mathrm{E}\int_0^\Delta |\sigma_t^2 - \sigma_0^2|\,\mathrm{d}t \\
&= O\left(\int_0^\Delta t^{1/2}\,\mathrm{d}t\right) \\
&= O(\Delta^{3/2}).
\end{aligned}
$$

Proof of Lemma 5.2. We follow the lines of thought as in the proof of Lemma 4.2 of [22], now applied in a multivariate setting. Let $\|\cdot\|$ denote the Euclidean norm. Writing

$$
\mathbf{W}_j = \mathbf{v}_h\left(\frac{\mathbf{x} - \log((\mathbf{X}_j^\Delta)^2)}{h}\right) - \mathbf{v}_h\left(\frac{\mathbf{x} - \log(\tilde{\mathbf{X}}_j^2)}{h}\right), \tag{41}
$$

so that $\mathbf{f}_{nh}(\mathbf{x}) - \tilde{\mathbf{f}}_{nh}(\mathbf{x}) = \frac{1}{(n-i_p^\Delta+i_1^\Delta)h^p}\sum_{j=1}^{n-i_p^\Delta+i_1^\Delta}\mathbf{W}_j$, and defining the event A as the event where all components of $|\mathbf{X}_1^\Delta|$ and $|\tilde{\mathbf{X}}_1|$ are larger than or equal to ε, we have

$$
\begin{aligned}
|\mathrm{E}\,\mathbf{f}_{nh}(\mathbf{x}) - \mathrm{E}\,\tilde{\mathbf{f}}_{nh}(\mathbf{x})| &\le \frac{1}{h^p}\,\mathrm{E}\,|\mathbf{W}_1| && (42)\\
&= \frac{1}{h^p}\,\mathrm{E}\,|\mathbf{W}_1|I_A && (43)\\
&\quad + \frac{1}{h^p}\,\mathrm{E}\,|\mathbf{W}_1|I_{A^c}I_{[\|\mathbf{X}_1^\Delta - \tilde{\mathbf{X}}_1\| \ge \varepsilon]} && (44)\\
&\quad + \frac{1}{h^p}\,\mathrm{E}\,|\mathbf{W}_1|I_{A^c}I_{[\|\mathbf{X}_1^\Delta - \tilde{\mathbf{X}}_1\| < \varepsilon]}. && (45)
\end{aligned}
$$

Recall that $|\log x - \log y| \le |x-y|/\varepsilon$ for $x, y \ge \varepsilon$. By Lemma 6.2, the bound (39) and stationarity, the term (43) can be bounded by

$$
\frac{2}{h^{p+1}}\,\gamma_0(h)^p\sum_{j=1}^p \mathrm{E}\,|\log(|X_{i_j^\Delta}^\Delta|) - \log(|\tilde{X}_{i_j^\Delta}|)|I_A \le \frac{2p}{h^{p+1}}\,\frac{1}{\varepsilon}\gamma_0(h)^p\mathrm{E}\,|X_1^\Delta - \tilde{X}_1| \le \frac{2p}{h^{p+1}}\,\gamma_0(h)^p\sqrt{C}\,\frac{\Delta^{1/4}}{\varepsilon}.
$$

This gives the first term in the order bound of Lemma 5.2.

The boundedness of the function $\mathbf{v}_h$ as stated in Lemma 6.2 yields $|\mathbf{w}_1| \le 2\gamma_0(h)^p$. Using also Chebychev's inequality and (39), we bound the term (44) by

$$
\begin{aligned}
\frac{2}{h^p}\,\gamma_0(h)^p P(\|\mathbf{X}_1^\Delta - \tilde{\mathbf{X}}_1\| \ge \varepsilon) &\le \frac{2}{h^p}\,\gamma_0(h)^p pP\left(|X_1^\Delta - \tilde{X}_1| \ge \frac{\varepsilon}{\sqrt{p}}\right) \\
&\le \frac{2p^2}{h^p}\,\gamma_0(h)^p C\,\frac{\Delta^{1/2}}{\varepsilon^2},
\end{aligned}
$$

which gives the second term in order bound of Lemma 5.2.

Consider the two arguments of the $\mathbf{v}_h$ functions in $\mathbf{W}_1$. Since at least one of them (and then the same for both arguments) is in absolute value eventually larger than $|\log 2\varepsilon|/h$, by Lemma 6.3 the term (45) can be bounded by

$$
2D\frac{1}{h^p}\gamma_0(h)^{p-1}\,\gamma_1(h, |\log 2\varepsilon|/h)\frac{1}{(|\log 2\varepsilon|/h)}\,pP(|\tilde{X}_1| \le 2\varepsilon) \le C_2\frac{1}{h^{p-1}}\gamma_0(h)^{p-1}\,\gamma_1(h, |\log 2\varepsilon|/h)\frac{\varepsilon}{|\log 2\varepsilon|},
$$

for some constant C_2, where we used in the last inequality the fact that the density of $\tilde{X}_1$ is bounded. This follows from the assumption that σ_0^2 has a bounded density in a neighborhood of zero, as can easily be verified. □

Proof of Lemma 5.3. Consider the decomposition

$$
\mathrm{Var}\,(\tilde{\mathbf{f}}_{nh}(\mathbf{x})) = \mathrm{Var}\,(\mathrm{E}\,(\tilde{\mathbf{f}}_{nh}(\mathbf{x})|\mathcal{F}_\sigma)) + \mathrm{E}\,(\mathrm{Var}\,(\tilde{\mathbf{f}}_{nh}(\mathbf{x})|\mathcal{F}_\sigma)). \tag{46}
$$

By the proof of Lemma 5.1 the conditional expectation $\mathrm{E}\,(\tilde{\mathbf{f}}_{nh}(\mathbf{x})|\mathcal{F}_\sigma)$ is equal to a multivariate kernel estimator of the density of $\log \boldsymbol{\sigma}_1^2$. Adapting the proof of Theorem 3 of [18] to the multivariate situation, we can bound its variance by

$$\frac{20(1+o(1))}{nh^{(1+q)p}\Delta} f(x)^{1-q} \left(\int_{-\infty}^{\infty} |w(u)|^{2/(1-q)}\, \mathrm{d}u \right)^{1-q} \int_0^\infty \alpha(\tau)^q \, \mathrm{d}\tau,$$

which is of the order $O(1/(nh^{(1+q)p}\Delta))$. This gives the second order bound in (28).

We turn to the expectation of the conditional variance. Using Lemma 6.2, we can bound the 'diagonal terms' of the conditional variance in (46) by

$$\frac{1}{(n-i_p^\Delta+i_1^\Delta)h^{2p}} \mathrm{E}\left(\mathbf{v}_h \left(\frac{\mathbf{x} - \log \tilde{\mathbf{X}}_1^2}{h} \right) \right)^2 = O\left(\frac{1}{nh^{2p}} \gamma_0(h)^{2p} \right),$$

where we also used that $i_p^\Delta/n \to 0$.

Next we consider the 'cross terms' of the conditional variance. Since nonzero covariance can only occur if the vectors $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$ have common elements, we investigate a 'worst case'. For fixed i, there are at most $p-1$ among the $\mathbf{x}_j$ that have elements in common with $\mathbf{x}_i$, which yields

$$\begin{aligned}
&\frac{1}{(n-i_p^\Delta+i_1^\Delta)^2h^{2p}} \sum_{i\neq j} \mathrm{E}\,\mathrm{Cov}\left(\mathbf{v}_h\left(\frac{\mathbf{x}-\log\tilde{\mathbf{X}}_i^2}{h}\right), \mathbf{v}_h\left(\frac{\mathbf{x}-\log\tilde{\mathbf{X}}_j^2}{h}\right) \middle| \mathcal{F}_\sigma \right) \\
&= \frac{2}{(n-i_p^\Delta+i_1^\Delta)^2h^{2p}} \sum_{i=1}^{n-i_p^\Delta+i_1^\Delta} \sum_{j=i+1}^{i+i_p^\Delta-i_1^\Delta} \mathrm{E}\,\mathrm{Cov}\left(\mathbf{v}_h\left(\frac{\mathbf{x}-\log\tilde{\mathbf{X}}_i^2}{h}\right), \mathbf{v}_h\left(\frac{\mathbf{x}-\log\tilde{\mathbf{X}}_j^2}{h}\right) \middle| \mathcal{F}_\sigma \right) \\
&\leq \frac{2(p-1)}{(n-i_p^\Delta+i_1^\Delta)h^{2p}} \gamma_0(h)^{2p} = O\left(\frac{1}{nh^{2p}}\gamma_0(h)^{2p} \right),
\end{aligned}$$

where in the last inequality we used that the expectation of the conditional covariance is bounded in absolute value by $\mathrm{E}\,(\mathbf{v}_h(\frac{\mathbf{x}-\log\tilde{\mathbf{X}}_1^2}{h}))^2$, due to stationarity. The first order bound in (28) follows by an application of Lemma 6.1. □

Proof of Lemma 5.4. We will use arguments similar to those in the proof of Lemma 5.2. With $\mathbf{W}_j$ as in (41) we have, using the ordinary variance decomposition and stationarity of the $\mathbf{W}_j$,

$$\mathrm{Var}\,(\mathbf{f}_{nh}(\mathbf{x}) - \tilde{\mathbf{f}}_{nh}(\mathbf{x})) = \frac{1}{(n-i_p^\Delta+i_1^\Delta)h^{2p}} \mathrm{Var}\,\mathbf{W}_1 + \frac{1}{(n-i_p^\Delta+i_1^\Delta)^2h^{2p}} \sum_{i\neq j} \mathrm{Cov}\,(\mathbf{W}_i, \mathbf{W}_j). \tag{47}$$

Let us first derive a bound on $\mathrm{Var}\,\mathbf{W}_1$. As in the proof of Lemma 5.2 we use A, the event where all components of $|\mathbf{X}_1^\Delta|$ and $|\tilde{\mathbf{X}}_1|$ are larger than or equal to ϵ. We have $\mathrm{Var}\,\mathbf{W}_1 \leq \mathrm{E}\,\mathbf{W}_1^2$, which can be split up as the sum of three terms

$$\mathrm{E}\,\mathbf{W}_1^2 = \mathrm{E}\,\mathbf{W}_1^2 I_A \tag{48}$$

$$+\,\mathrm{E}\,\mathbf{W}_1^2 I_{A^c} I_{[\|\mathbf{X}_1^\Delta - \tilde{\mathbf{X}}_1\|\geq\varepsilon]} \tag{49}$$

$$+\,\mathrm{E}\,\mathbf{W}_1^2 I_{A^c} I_{[\|\mathbf{X}_1^\Delta - \tilde{\mathbf{X}}_1\|<\varepsilon]}. \tag{50}$$

By stationarity, the Lipschitz property of $\mathbf{v}_h$ in Lemma 6.2 and (39), the term (48) can be bounded by

$$\begin{aligned}
\frac{4}{h^2}\gamma_0(h)^{2p}\mathrm{E}\left(\sum_{j=1}^p \left| \log|X_{i_j^\Delta}| - \log|\tilde{X}_{i_j^\Delta}| \right| \right)^2 I_A &\leq \frac{4p}{h^2}\gamma_0(h)^{2p}\mathrm{E}\sum_{j=1}^p (\log|X_{i_j^\Delta}| - \log|\tilde{X}_{i_j^\Delta}|)^2 I_A \\
&\leq \frac{4p^2}{h^2}\frac{1}{\varepsilon^2}\gamma_0(h)^{2p}\mathrm{E}\,(|X_1^\Delta| - |\tilde{X}_1|)^2 \\
&\leq \frac{4p^2}{h^2}\frac{1}{\varepsilon^2}\gamma_0(h)^{2p}\mathrm{E}\,(X_1^\Delta - \tilde{X}_1)^2 \leq \frac{4p^2}{h^2}\gamma_0(h)^{2p} C\frac{\Delta^{1/2}}{\varepsilon^2}.
\end{aligned} \tag{51}$$

We turn to the term (49). By the bound on $\mathbf{v}_h$ of Lemma 6.2 and by (39) again, it can be bounded by

$$\begin{aligned}
4\gamma_0(h)^{2p} P(\|\mathbf{X}_1^\Delta - \tilde{\mathbf{X}}_1\| \geq \varepsilon) &\leq 4\gamma_0(h)^{2p} pP\left(|X_1^\Delta - \tilde{X}_1| \geq \frac{\varepsilon}{\sqrt{p}} \right) \\
&\leq 4p^2\gamma_0(h)^{2p} C\,\frac{\Delta^{1/2}}{\varepsilon^2}.
\end{aligned}$$

Due to the absence of a factor h^2 in the denominator, this bound is of smaller order than the one for (48) and will therefore be neglected.

Next we consider (50). Recall from the proof of Lemma 5.2 that $P(|\tilde{X}_1| \le 2\varepsilon) = O(\varepsilon)$. Since at least one (the same) coordinate of the absolute value of both arguments of $\mathbf{v}_h$ is eventually larger than $|\log 2\varepsilon|/h$, by Lemma 6.3 the term (50) can be bounded by

$$4D^2 \; \gamma_0(h)^{2p-2}\gamma_1(h, |\log 2\varepsilon|/h)^2 \tfrac{1}{(|\log 2\varepsilon|^2/h^2)} pP(|\tilde{X}_1| \le 2\varepsilon) \le C_2 h^2 \gamma_0(h)^{2p-2}\gamma_1(h, |\log 2\varepsilon|/h)^2 \tfrac{\varepsilon}{|\log 2\varepsilon|^2}, \tag{52}$$

for some constant C_2.

Wrapping up the order bounds (51) and (52) for $\mathrm{E}\,\mathbf{W}_1^2$, we get

$$\mathrm{E}\,\mathbf{W}_1^2 = O\left(\frac{1}{h^2}\gamma_0(h)^{2p}\frac{\Delta^{1/2}}{\varepsilon^2} + h^2\gamma_0(h)^{2p-2}\,\gamma_1(h, |\log 2\varepsilon|/h)^2\frac{\varepsilon}{|\log 2\varepsilon|^2}\right), \tag{53}$$

which, substituted in (47), gives the order bounds of (29).

We now consider the covariance terms in (47), that will be seen to have the order bounds of (30). We have the decomposition

$$\mathrm{Cov}\,(\mathbf{W}_i, \mathbf{W}_j) = \mathrm{E}\,\mathrm{Cov}\,(\mathbf{W}_i, \mathbf{W}_j|\mathcal{F}_\sigma) + \mathrm{Cov}\,(\mathrm{E}\,(\mathbf{W}_i|\mathcal{F}_\sigma), \mathrm{E}\,(\mathbf{W}_j|\mathcal{F}_\sigma)). \tag{54}$$

The last term in (47) then becomes

$$\frac{2}{(n - i_p^\Delta + i_1^\Delta)^2 h^{2p}} \sum_{i=1}^{n-i_p^\Delta+i_1^\Delta} \sum_{j\neq i}^{i+i_p^\Delta-i_1^\Delta} \mathrm{E}\,\mathrm{Cov}\,(\mathbf{W}_i, \mathbf{W}_j|\mathcal{F}_\sigma) \tag{55}$$

$$+ \frac{1}{(n - i_p^\Delta + i_1^\Delta)^2 h^{2p}} \sum_{i\neq j} \mathrm{Cov}\,(\mathrm{E}\,(\mathbf{W}_i|\mathcal{F}_\sigma), \mathrm{E}\,(\mathbf{W}_j|\mathcal{F}_\sigma)). \tag{56}$$

In a first step we consider the expectation of the conditional covariances in (55). Arguing as in the proof of Lemma 5.3, we can bound it by

$$\frac{(p-1)}{(n - i_p^\Delta + i_1^\Delta)h^{2p}}\mathrm{Var}\,\mathbf{W}_1,$$

which is $p - 1$ times the first term on the right hand side of Eq. (47). Hence its contribution can be absorbed in the already obtained bounds of (29).

Next we concentrate on the sum of covariances in (56). Define

$$\bar{\sigma}_i^2 = \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} \sigma_t^2\,\mathrm{d}t \tag{57}$$

and the vector $\bar{\boldsymbol{\sigma}}_j^2$ by $\bar{\boldsymbol{\sigma}}_j^2 = (\bar{\sigma}_j^2, \dots, \bar{\sigma}^2_{i_p^\Delta - i_1^\Delta + j})$. Note that given $\mathcal{F}_\sigma$, $\mathbf{X}_i^\Delta$ is a multivariate normal vector with independent components with variances equal to the components of $\bar{\boldsymbol{\sigma}}_i^2$ and that $\tilde{\mathbf{X}}_i$ is a multivariate normal vector with independent components with variances equal to the components of $\boldsymbol{\sigma}_i^2$. As in the proof of Lemma 5.1 it follows that

$$\mathrm{E}\,(\mathbf{W}_i|\mathcal{F}_\sigma) = \mathbf{w}\left(\frac{\mathbf{x} - \log\bar{\boldsymbol{\sigma}}_i^2}{h}\right) - \mathbf{w}\left(\frac{\mathbf{x} - \log\boldsymbol{\sigma}_i^2}{h}\right).$$

We follow the line of arguments in the proof of Theorem 3 in [18]. The stationarity of $\mathbf{W}_j$ implies that the conditional expectations $\tilde{\mathbf{W}}_j := \mathrm{E}\,(\mathbf{W}_j|\mathcal{F}_\sigma)$ are also stationary. Hence we have

$$\sum_{i\neq j}\mathrm{Cov}\,(\tilde{\mathbf{W}}_i, \tilde{\mathbf{W}}_j) = 2\sum_{k=1}^{n-1}(n-k)\mathrm{Cov}\,(\tilde{\mathbf{W}}_0, \tilde{\mathbf{W}}_k).$$

Now note that the process $\tilde{\mathbf{W}}_j$ is strongly mixing with a mixing coefficient $\tilde{\alpha}(k) \le \alpha((k-2)\Delta + t_1 - t_p)$, $k = 1, 2, \dots$ if $k\Delta > t_p - t_1 + 2\Delta$ and $\tilde{\alpha}(k) = 1$ otherwise. By a lemma of [5] for strongly mixing processes it follows that for all $\tau > 0$

$$|\mathrm{Cov}\,(\tilde{\mathbf{W}}_0, \tilde{\mathbf{W}}_k)| \le 10\alpha((k-2)\Delta + t_1 - t_p)^{\tau/(2+\tau)}(\mathrm{E}\,|\tilde{\mathbf{W}}_1|^{2+\tau})^{2/(2+\tau)}. \tag{58}$$

By the equivalent condition (15) on the mixing coefficients $\alpha(t)$ (applied with $\tau = 2q/(1-q)$, a choice for τ that we will make later on as well), we get for (56)

$$\left|\frac{1}{(n-i_p^\Delta+i_1^\Delta)^2h^{2p}}\sum_{i\neq j}\operatorname{Cov}(\tilde{\mathbf{W}}_i,\tilde{\mathbf{W}}_j)\right| \le \frac{10}{(n-i_p^\Delta+i_1^\Delta)h^{2p}}(\mathrm{E}\,|\tilde{\mathbf{W}}_1|^{2+\tau})^{2/(2+\tau)}$$
$$\times\sum_{k=1}^{n-i_p^\Delta}\left(1-\frac{k}{n-i_p^\Delta+i_1^\Delta}\right)\alpha(k\Delta+t_1-t_p)^{\tau/(2+\tau)}$$
$$\le \frac{10}{(n-i_p^\Delta+i_1^\Delta)h^{2p}}\frac{C\left(\frac{\tau}{2+\tau},t_1-t_p-2\Delta\right)}{\Delta}(\mathrm{E}\,|\tilde{\mathbf{W}}_1|^{2+\tau})^{2/(2+\tau)}.$$

Next we derive a bound on $\mathrm{E}\,|\tilde{\mathbf{W}}_1|^{2+\tau}$. Fix $\kappa\in(0,1]$ and define the event B as the event where all components of $\bar{\boldsymbol{\sigma}}_1^2$ and $\boldsymbol{\sigma}_1^2$ are larger than or equal to ϵ. We have

$$\mathrm{E}\,|\tilde{\mathbf{W}}_1|^{2+\tau} = \mathrm{E}\left|\mathbf{w}\left(\frac{\mathbf{x}-\log(\bar{\boldsymbol{\sigma}}_1^2)}{h}\right)-\mathbf{w}\left(\frac{\mathbf{x}-\log(\boldsymbol{\sigma}_1^2)}{h}\right)\right|^{2+\tau}I_B \tag{59}$$

$$+\mathrm{E}\left|\mathbf{w}\left(\frac{\mathbf{x}-\log(\bar{\boldsymbol{\sigma}}_1^2)}{h}\right)-\mathbf{w}\left(\frac{\mathbf{x}-\log(\boldsymbol{\sigma}_1^2)}{h}\right)\right|^{2+\tau}I_{B^c}I_{[\|\bar{\boldsymbol{\sigma}}_1^{2\kappa}-\boldsymbol{\sigma}_1^{2\kappa}\|\ge\varepsilon]} \tag{60}$$

$$+\mathrm{E}\left|\mathbf{w}\left(\frac{\mathbf{x}-\log(\bar{\boldsymbol{\sigma}}_1^2)}{h}\right)-\mathbf{w}\left(\frac{\mathbf{x}-\log(\boldsymbol{\sigma}_1^2)}{h}\right)\right|^{2+\tau}I_{B^c}I_{[\|\bar{\boldsymbol{\sigma}}_1^{2\kappa}-\boldsymbol{\sigma}_1^{2\kappa}\|<\varepsilon]}. \tag{61}$$

By Lemma 6.2 the term (59) can be bounded by a constant times

$$\frac{1}{h^{2+\tau}}\mathrm{E}\left(\sum_{j=1}^p|\log(\bar\sigma^2_{i_j^\Delta})-\log(\sigma^2_{(i_j^\Delta-1)\Delta})|\right)^{2+\tau}I_B \le \frac{p^{1+\tau}}{h^{2+\tau}}\mathrm{E}\sum_{j=1}^p|\log(\bar\sigma^2_{i_j^\Delta})-\log(\sigma^2_{(i_j^\Delta-1)\Delta})|^{2+\tau}I_B$$
$$\le \frac{p^{2+\tau}}{h^{2+\tau}}\mathrm{E}\,|\log(\bar\sigma^2_{i_1^\Delta})-\log(\sigma^2_{(i_j^\Delta-1)\Delta})|^{2+\tau}I_B$$
$$\le \frac{p^{2+\tau}}{(\kappa\epsilon h)^{2+\tau}}\mathrm{E}\,|\bar\sigma_1^{2\kappa}-\sigma_0^{2\kappa}|^{2+\tau}. \tag{62}$$

The term (60) can be bounded by

$$pP\left(|\bar\sigma_1^{2\kappa}-\sigma_0^{2\kappa}|\ge\frac{\varepsilon}{\sqrt{p}}\right)\le\frac{p^{2+\tau/2}}{\varepsilon^{2+\tau}}\mathrm{E}\,|\bar\sigma_1^{2\kappa}-\sigma_0^{2\kappa}|^{2+\tau}.$$

Since this is for $h\to0$ of smaller order than (62), it will be neglected in what follows.

Finally we analyze the term (61). On the complement of B there is at least one component of either $\bar{\boldsymbol{\sigma}}_1^2$ or $\boldsymbol{\sigma}_0^2$ that is smaller than or equal to ϵ. Together with $\|\bar{\boldsymbol{\sigma}}_1^{2\kappa}-\boldsymbol{\sigma}_1^{2\kappa}\|<\varepsilon$ this implies that there is at least one pair of corresponding components of the vectors that are both smaller than $\varepsilon(1+\varepsilon^{1-\kappa})^{1/\kappa}$. Using the stationarity, we bound the term (61) by

$$pP(\bar\sigma_1^2\le\varepsilon(1+\varepsilon^{1-\kappa})^{1/\kappa}\quad\text{and}\quad\sigma_0^2\le\varepsilon(1+\varepsilon^{1-\kappa})^{1/\kappa}),$$

which is bounded by

$$pP(\sigma_0^2\le2\varepsilon)=O(\varepsilon), \tag{63}$$

since σ_0^2 was assumed to have a bounded density in a neighborhood of zero. Combining (62) and (63) with $\tau=2q/(1-q)$ and $\kappa=\frac{1}{2+\tau}=\frac{1-q}{2}$, we have with an application of the basic inequality $|u^\kappa-v^\kappa|\le|u-v|^\kappa$ for $u,v\ge0$ and $\kappa\in(0,1]$ in the second equality below and (40) in the fourth equality for the term (56)

$$\left|\frac{1}{(n-i_p^\Delta+i_1^\Delta)^2h^{2p}}\sum_{i\neq j}\operatorname{Cov}(\tilde{\mathbf{W}}_i,\tilde{\mathbf{W}}_j)\right| = \frac{1}{(n-i_p^\Delta+i_1^\Delta)h^{2p}\Delta}O\left(\frac{1}{h^{2+\tau}}\frac{1}{\varepsilon^{2+\tau}}\mathrm{E}\,|\bar\sigma_1^{2\kappa}-\sigma_0^{2\kappa}|^{2+\tau}+\varepsilon\right)^{2/(2+\tau)}$$
$$=\frac{1}{(n-i_p^\Delta+i_1^\Delta)h^{2p}\Delta}O\left(\frac{1}{h^{2+\tau}}\frac{1}{\varepsilon^{2+\tau}}\mathrm{E}\,|\bar\sigma_1^2-\sigma_0^2|^{\kappa(2+\tau)}+\varepsilon\right)^{2/(2+\tau)}$$
$$=\frac{1}{(n-i_p^\Delta+i_1^\Delta)h^{2p}\Delta}O\left(\frac{(\mathrm{E}\,|\bar\sigma_1^2-\sigma_0^2|)^{2/(2+\tau)}}{h^2\varepsilon^2}+\varepsilon^{2/(2+\tau)}\right),$$

$$= \frac{1}{(n - i_p^{\Delta} + i_1^{\Delta})h^{2p}\Delta} O\left(\frac{\Delta^{1/(2+\tau)}}{h^2\varepsilon^2} + \varepsilon^{2/(2+\tau)}\right)$$
$$= \frac{1}{(n - i_p^{\Delta} + i_1^{\Delta})h^{2p}\Delta} O\left(\frac{\Delta^{(1-q)/2}}{h^2\varepsilon^2} + \varepsilon^{1-q}\right).$$

Hence the last term in (47) now gives the third order bound (30). □

Acknowledgment

We thank Galyna Ignatenko for providing us with the illustrations.

References

[1] F. Comte, Kernel deconvolution of stochastic volatility models, J. Time Ser. Anal. 25 (2004) 563–582.
[2] F. Comte, J. Dedecker, M.L. Taupin, Adaptive density estimation for general ARCH models, Econometric Theory 24 (2008) 1628–1662.
[3] F. Comte, V. Genon-Catalot, Penalized projection estimator for volatility density, Scand. J. Statist. 33 (4) (2006) 875–893.
[4] F. Comte, V. Genon-Catalot, Y. Rozenholc, Nonparametric estimation for a stochastic volatility model, Finance Stoch. 14 (1) (2010) 49–80.
[5] C.M. Deo, A note on empirical processes for strong mixing processes, Ann. Probab. 1 (1973) 870–875.
[6] P. Doukhan, Mixing, Properties and Examples, Springer-Verlag, 1994.
[7] J. Fan, On the optimal rates of convergence for nonparametric deconvolution problems, Ann. Statist. 19 (1991) 1257–1272.
[8] J. Franke, W. Härdle, J.P. Kreiss, Nonparametric estimation in a stochastic volatility model, in: M.G. Akritas, D.N. Politis (Eds.), Recent Advances and Trends in Nonparametric Statistics, Elsevier, 2003.
[9] V. Genon-Catalot, T. Jeantheau, C. Larédo, Limit theorems for discretely observed stochastic volatility models, Bernoulli 4 (1998) 283–303.
[10] V. Genon-Catalot, T. Jeantheau, C. Larédo, Parameter estimation for discretely observed stochastic volatility models, Bernoulli 5 (1999) 855–872.
[11] V. Genon-Catalot, T. Jeantheau, C. Larédo, Stochastic volatility models as hidden Markov models and statistical applications, Bernoulli 6 (2000) 1051–1079.
[12] I.I. Gihman, A.V. Skorohod, Stochastic Differential Equations, Springer, 1972.
[13] W. Härdle, Smoothing Techniques, Springer Verlag, New York, 1990.
[14] S.L. Heston, A closed-form solution for options with stochastic volatility with applications to Bond and Currency options, Rev. Financ. Stud. 6 (2) (1993) 327–343.
[15] I.A. Ibragimov, Yu.V. Linnik, Independent and Stationary Sequences of Random Variables, Wolters-Noordhoff, 1971.
[16] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, Springer Verlag, New York, 1991.
[17] E. Masry, Strong consistency and rates for deconvolution of multivariate densities of stationary processes, Stochastic Process. Appl. 475 (1993) 53–74.
[18] E. Masry, Probability density estimation from sampled data, IEEE Trans. Inform. Theory 29 (1983) 696–709.
[19] D.B. Nelson, ARCH models as diffusion approximations, J. Econometrics 45 (1990) 7–38.
[20] D.W. Scott, Multivariate Density Estimation. Theory, Practice, and Visualization, Wiley, New York, 1992.
[21] A.V. Skorokhod, Asymptotic Methods in the Theory of Stochastic Differential Equations, AMS, 1989.
[22] B. Van Es, P. Spreij, H. Van Zanten, Nonparametric volatility density estimation, Bernoulli 9 (2003) 451–465.
[23] B. Van Es, P. Spreij, H. Van Zanten, Nonparametric volatility density estimation for discrete time models, J. Nonparametr. Stat. 17 (2005) 237–251.
[24] H. Van Zanten, P. Zareba, A note on wavelet density deconvolution for weakly dependent data, Stat. Inference Stoch. Process. 11 (2008) 207–219.
[25] M.P. Wand, Finite sample performance of deconvolving kernel density estimators, Statist. Probab. Lett. 37 (1998) 131–139.
[26] M.P. Wand, M.C. Jones, Kernel Smoothing, Chapman and Hall, London, 1995.
[27] J.B. Wiggins, Option valuation under stochastic volatility, J. Financ. Econ. 19 (1987) 351–372.