Journal of Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: www.tandfonline.com/journals/gnst20

Nonparametric volatility density estimation for discrete time models

Bert Van Es, Peter Spreij & Harry Van Zanten

To cite this article: Bert Van Es, Peter Spreij & Harry Van Zanten (2005) Nonparametric volatility density estimation for discrete time models, Journal of Nonparametric Statistics, 17:2, 237-249, DOI: 10.1080/1048525042000267752

To link to this article: https://doi.org/10.1080/1048525042000267752

Published online: 31 Jan 2007.

Nonparametric Statistics
Vol. 17, No. 2, March 2005, 237–251

Taylor & Francis
Taylor & Francis Group

Nonparametric volatility density estimation for discrete time models

BERT VAN ES†, PETER SPREIJ*‡ and HARRY VAN ZANTEN‡

†Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands
‡Division of Mathematics and Computer Science, Vrije Universiteit, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

(Received December 2002; in final form May 2004)

We consider discrete time models for asset prices with a stationary volatility process. We aim at estimating the multivariate density of this process at a set of consecutive time instants. A Fourier-type deconvolution kernel density estimator based on the logarithm of the squared process is proposed to estimate the volatility density. Expansions of the bias and bounds on the variance are derived.

Keywords: Stochastic volatility models; Density estimation; Kernel estimator; Deconvolution; Mixing

AMS Subject Classification: 62G07; 62M07; 62P20

1. Introduction

Suppose that we have price data of a certain asset in a financial market. By X, we denote the detrended or demeaned log-return process. It is commonly believed that stochastic volatility models of the form

$$X_t = \sigma_t Z_t \tag{1}$$

describe much of the observed behavior of this type of data. Here, Z is typically an i.i.d. noise sequence (often Gaussian) and at each time t the random variables σ_t and Z_t are independent. See the survey papers by Ghysels *et al.* [1] or Shephard [2]. We will assume that the process σ is strictly stationary and that the (multivariate) marginal distributions of σ have a density with respect to the Lebesgue measure on $(0, \infty)$. Our aim is to construct a nonparametric estimator for the multivariate density of $(\sigma_t, \ldots, \sigma_{t+p} - 1)$ and to study its asymptotic behavior.

Models that are used in the literature to describe the volatility display rather different invariant distributions. This observation lies at the basis of our point of view, which we pursue in this paper, that nonparametric estimation procedures are by all means sensible tools to get some insight in the behavior of the volatility. Quite often in models that are used in practice, the

*Corresponding author. Email: spreij@science.uva.nl

Nonparametric Statistics
ISSN 1048-5252 print/ISSN 1029-0311 online
http://www.tandf.co.uk/journals
DOI: 10.1080/10485250420002677520

invariant distributions of σ are unimodal. As it is known that volatility clustering is an often-occurring phenomenon, it is hard to believe that this can be explained by any of these models. Instead, one would expect in such a case, for instance, the distribution of (σ_t, σ_{t+1}) to have a density that has concentration regions around the diagonal with possibly peaks at certain clusters of low and high volatility, a phenomenon that may lead to, for instance, bimodal one-dimensional marginal distributions. Nonparametric density estimation could perhaps reveal such a shape of the invariant density of the volatility.

We will distinguish two classes of models in this paper. In both the classes, we will assume that the noise sequence is standard Gaussian and that σ is a strictly stationary, positive process satisfying a certain mixing condition. However, the way in which the bivariate process (σ, Z), in particular its dependence structure, is further modeled differs. In the first class of models that we consider, we assume that the process σ is predictable with respect to the filtration $\mathcal{F}_t$ generated by the process Z. Note that σ_t is independent of Z_t for each fixed time t. Furthermore, we 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. Financial data such as log-returns of stock prices or exchange rates are believed to share a number of stylized features, including heavy-tailedness and long-range dependence. Models of the type (1) have been proposed to capture those features. A well-known family included in the class (1) is the family of GARCH-models, introduced by Bollerslev [3]. For the GARCH$(p,\ q)$-model the sequence $\{\sigma_t\}$ in equation (1) is assumed to satisfy the equation

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{2}$$

where the α_i and β_j are non-negative constants. Under suitable assumptions (see ref. [4, 5]) GARCH processes are stationary and the statistical problem in this case would be to estimate the coefficients α_i and β_j in equation (2).

In the second class of models that we consider, we assume that the whole process σ is independent of the noise process Z. 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 [6], where σ is generated as an AR(1) process with α-stable noise ($\alpha \in (0, 1)$).

As we said earlier, we do not want to make a parametric assumption such as equation (2), but we still want to measure the volatility of the data somehow. In the present paper, we propose a nonparametric statistical procedure for this problem. Using ideas from deconvolution theory, we will propose a procedure for the estimation of the marginal density at a fixed point. To assess the quality of our procedure, we will derive expansions of the bias and bounds on the variance. This will be done separately for the two kinds of model classes outlined earlier.

2. The estimators

We first briefly review the construction of the deconvolution kernel density estimator based on i.i.d. observations. These kernel type estimators have been introduced by Carroll and Hall [7], Liu and Taylor [8], Stefanski [9], and Stefanski and Carroll [10] (see also ref. [11]). For simplicity, we treat in a standard setting the univariate case only. Let a random variable X

be equal to the sum of two independent random variables, say Y, with unknown density f, and Z, with known density k. The density g of X is the convolution of f and k and for the corresponding characteristic functions $\phi_g(t)$, $\phi_f(t)$, and $\phi_k(t)$, we have

$$\phi_g(t) = \phi_f(t)\phi_k(t). \tag{3}$$

The objective is to estimate f from i.i.d. observations of $X_1, \ldots, X_n$ having density g. In identity (3), we know $\phi_k(t)$ and we can estimate $\phi_g(t)$ by the characteristic function of a *kernel estimator* g_{nh} of g. Therefore, let

$$g_{nh}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h}w\left(\frac{x - X_j}{h}\right),$$

where w is an integrable function with integral one, the *kernel function*, and $h > 0$ is a positive number, the *bandwidth*. The characteristic function $\phi_{g_{nh}}$ of g_{nh} is then given by

$$\phi_{g_{nh}}(t) = \phi_w(ht)\phi_{\mathrm{emp}}(t), \tag{4}$$

where $\phi_{\mathrm{emp}}(t) = 1/n\sum_{j=1}^{n}\mathrm{e}^{itX_j}$ is the *empirical characteristic function* and ϕ_w is the Fourier transform of w. From equations (3) and (4), we see that

$$\frac{\phi_w(ht)\phi_{\mathrm{emp}}(t)}{\phi_k(t)} \tag{5}$$

is an obvious candidate to estimate ϕ_f. Its inverse Fourier transform f_{nh} is then used to estimate f. Therefore, the estimator f_{nh} of f is

$$f_{nh}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-itx}\frac{\phi_w(ht)\phi_{\mathrm{emp}}(t)}{\phi_k(t)}\,\mathrm{d}t. \tag{6}$$

The inversion is allowed, if the function (5) is integrable. In general, this is not guaranteed. However, to enforce integrability, we assume that ϕ_w has a bounded support. Note that equation (6) can be rewritten as

$$f_{nh}(x) = \frac{1}{nh}\sum_{j=1}^{n}v_h\left(\frac{x - X_j}{h}\right), \tag{7}$$

where

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

It is easy to see that the function v_h, and hence the estimator $f_{nh}(x)$, is real valued.

We now turn to the model (1), therefore $X_t = \sigma_t Z_t$. If we square this equation and take logarithms, we get

$$\log X_t^2 = \log\sigma_t^2 + \log Z_t^2. \tag{9}$$

Recall that under our assumptions for each t, the random variables σ_t and Z_t are independent. The density of $\log Z_t^2$, denoted by k, is given by

$$k(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{(1/2)x}\mathrm{e}^{-(1/2)\mathrm{e}^x}. \tag{10}$$

We will use a *deconvolution kernel density estimator* to estimate the unknown density f of $\log\sigma_t^2$. An estimate of the density of σ_t^2 or σ_t can then be obtained by a simple transformation.

Computing the characteristic function ϕ_k of $\log Z^2$, we get

$$\phi_k(t) = \frac{1}{\sqrt{\pi}} 2^{it} \Gamma\left(\frac{1}{2} + it\right), \tag{11}$$

where Γ denotes the gamma function. For the model (1), this leads to the estimator

$$f_{nh}(x) = \frac{1}{nh} \sum_{j=1}^{n} v_h \left(\frac{x - \log X_j^2}{h} \right) \tag{12}$$

of the density f of $\log \sigma_t^2$, with $v_h(x)$ as in equation (8).

The expression for the estimator of the density of the p-dimensional random vector $(\log \sigma_t^2, \ldots, \log \sigma_{t-p+1}^2)$ is similar. We first introduce some auxiliary notation. Let p be fixed and write $\mathbf{x}_j$ for a vector $(x_j, \ldots, x_{j-p+1})$. We use similar boldface expressions for other (random) vectors. The kernel $\mathbf{w}$ that we 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

$$\mathbf{v}_h(\mathbf{x}) = \frac{1}{(2\pi)^p} \int_{\mathbb{R}^p} \frac{\phi_{\mathbf{w}}(\mathbf{s})}{\phi_{\mathbf{k}}(\mathbf{s}/h)} \mathrm{e}^{-i\mathbf{s}\cdot\mathbf{x}} \, \mathrm{d}\mathbf{s}, \tag{13}$$

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 that we will use is given by

$$f_{nh}(\mathbf{x}) = \frac{1}{(n-p+1)h^p} \sum_{j=p}^{n} \mathbf{v}_h \left(\frac{\mathbf{x} - \log \mathbf{X}_j^2}{h} \right), \tag{14}$$

where we use $\log \mathbf{X}_j^2$ to denote the vector $\left(\log X_j^2, \ldots, \log X_{j-p+1}^2\right)$.

3. Asymptotics

The bias of the deconvolution estimator described in section 2 will be seen to be the same as the bias of a kernel density estimator based on independent observations from f. Hence, under standard smoothness assumptions, it is of order h^2 as $h \to 0$. The variance of this type of deconvolution estimator 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. In other words, the smoother the k, the harder the estimation problem. This follows, for instance, for i.i.d. observations from results in Fan [12] and for stationary observations from the work of Masry [13–15].

The rate of decay of $|\phi_k(t)|$ for the density equation (10) is given by Lemma 5.1 in section 5, where we show that

$$|\phi_k(t)| \sim \sqrt{2}\mathrm{e}^{-(1/2)\pi|t|}, \quad \text{as } |t| \longrightarrow \infty. \tag{15}$$

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. ref. [12] for results on i.i.d. observations). Note that this rate, however slow, is faster than the one for normal deconvolution.

In the model (9), the sequence $\{\log X_t^2\}$ is not independent; therefore, results on the asymptotic behavior of the kernel estimator of section 2 are not directly applicable. In the literature, also more general deconvolution problems have been studied, where the i.i.d. assumption has been relaxed. For instance, the deconvolution model $X_j = Y_j + Z_j$, where $\{Y_j, Z_j\}$ is a stationary sequence and the sequences $\{Z_j\}$ $\{Y\}$ are independent has been treated by Masry [13–15].

Expansions for the variance of the deconvolution kernel estimator have been derived under several mixing conditions. Under the assumption that the volatility process is independent of the noise sequence, the model (9) fits into this scheme. We will obtain similar results for the estimator, when σ (as a process) is not independent of Z, but only predictable with respect to the filtration generated by Z.

Let us define the mixing conditions. For a certain process $\{X_j\}$, let $\mathcal{F}_a^b$ be the σ-algebra of events generated by the random variables X_j, $j = a, \dots, b$. Let the mixing coefficient α_k be defined by

$$\alpha_k = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |P(AB) - P(A)P(B)|. \tag{16}$$

We call a process $\{X_j\}$ *strongly mixing*, if $\alpha_k \to 0$ as $k \to \infty$.

To obtain expansions for the bias and variance, we also need conditions on the kernel function w such as bounded support of its characteristic function $\phi_w(t)$. Moreover, the rate of decay to zero of $\phi_w(t)$ at the boundary of its support turns up in the asymptotics. The complete list of assumptions on w that we use is as follows.

Condition W Let w be a real symmetric function satisfying

1. $\int_{-\infty}^{\infty} |w(u)| \, \mathrm{d}u < \infty$,
2. $\int_{-\infty}^{\infty} w(u) \, \mathrm{d}u = 1$,
3. $\int_{-\infty}^{\infty} u^2 w(u) \, \mathrm{d}u < \infty$,
4. $\lim_{|u| \to \infty} w(u) = 0$,
5. ϕ_w, the characteristic function of w has support $[-1, 1]$,
6. $\phi_w(1-t) = At^\rho + o(t^\rho)$, as $t \downarrow 0$ for some $\rho > 0$.

Note that by Fourier inversion these conditions imply that w is bounded and Lipschitz. More precisely, we have

$$|w(x)| \le \frac{1}{2\pi} \quad \text{and} \quad |w(x+u) - w(x)| \le \frac{1}{2\pi}|u|. \tag{17}$$

An example of such a kernel, from Wand [16], with $\rho = 3$ and $A = 8$, is

$$w(x) = \frac{48x(x^2 - 15)\cos x - 144(2x^2 - 5)\sin x}{\pi x^7}. \tag{18}$$

It has characteristic function

$$\phi_w(t) = (1 - t^2)^3, \quad |t| \le 1. \tag{19}$$

The next theorem, whose proof can be found in section 5, 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 is a Markov chain, since then convergence of α_k to zero takes place at an *exponential rate* (see Theorems 4.2 and 4.3 of Bradley, [17] for precise statements). Similar behavior occurs for ARMA processes with absolutely continuous distributions of the noise terms (ref. [17], Example 6.1).

THEOREM 3.1 *Assume that the process X 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 W and let the density f of the p-vector $(\log \sigma_1^2, \ldots, \log \sigma_p^2)$ be bounded and twice continuously differentiable with bounded second-order partial derivatives. Assume that σ is a predictable process with respect to the filtration generated by the process Z. Then, we have for the estimator of the multivariate density defined as in equation (14) and $h \to 0$

$$E\, f_{nh}(\mathbf{x}) = f(\mathbf{x}) + \frac{1}{2} h^2 \int \mathbf{u}^{\top} \nabla^2 f(\mathbf{x}) \mathbf{u}\, \mathbf{w}(\mathbf{u})\, \mathrm{d}\mathbf{u} + o(h^2) \tag{20}$$

and

$$\mathrm{Var}\, f_{nh}(\mathbf{x}) = O\left(\frac{1}{n} (h^{2\rho-\beta} \mathrm{e}^{\pi/h})^p \right). \tag{21}$$

THEOREM 3.2 *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 W and let the density f of the p-vector $(\log \sigma_1^2, \ldots, \log \sigma_p^2)$ be bounded and twice continuously differentiable with bounded second-order partial derivatives. Assume that σ and Z are independent processes. Then, the multivariate density estimator f_{nh} satisfies the same bias expansion as in Theorem 3.1. *For the variance, we have the sharper bound*

$$\mathrm{Var}\, f_{nh}(\mathbf{x}) = O\left(\frac{1}{n} (h^{2\rho} \mathrm{e}^{\pi/h})^p \right). \tag{22}$$

Remark 3.3 Because of the exponential factor in the variance bound, in order to obtain consistency, one has to take essentially $h \geq p\pi/\log n$ (see also ref. [9] for a related problem). On the other hand, we would like to minimize the bias, therefore, the choice $h = p\pi/\log n$ is optimal. Both the bias and the variance decay at a logarithmic rate for this choice of bandwidth. This seems disappointing; however, Fan [12] shows for the i.i.d. situation of Section 2 that we cannot expect anything better.

Remark 3.4 Notice that the results in Masry [14, 15] establishing strong consistency, rates of convergence and asymptotic normality are not useful here, because the condition that ϕ_k has either purely real or purely imaginary tails is not satisfied.

Remark 3.5 Note that our assumptions in Theorem 3.1 are slightly different from those of Masry [13]. One of the essential facts that is used in the proof is the mixing property of X. If σ and Z are independent processes that is implied by a similar assumption on the σ process itself as in Masry [13].

Remark 3.6 In the case where the processes σ and Z are independent, the estimators $f_{nh}(x)$ have the following property:

$$\tilde{f}_{nh}(x) := E[f_{nh}(x)|\mathcal{F}^\sigma] = \frac{1}{nh}\sum_{j=1}^{n} w\left(\frac{x - \log\sigma_j^2}{h}\right), \tag{23}$$

where $\mathcal{F}^\sigma$ denotes the σ-algebra generated by the whole process σ. Thus, the $\tilde{f}_{nh}(x)$ would be ordinary kernel density estimators, if the σ_j^2 could be observed.

Equation (23) is seen to be true as follows. Write $u_j = \log X_j^2$ and use similar notation for $\zeta_j = \log Z_j^2$ and $\tau_j = \log\sigma_j^2$. Then,

$$\begin{aligned} E\left[v_h\left(\frac{x-u_j}{h}\right)|\mathcal{F}^\sigma\right] &= \frac{1}{2\pi}\int E\mathrm{e}^{is\zeta_j/h}\frac{\phi_w(s)}{\phi_k(s/h)}\mathrm{e}^{-is(x-\tau_j)/h}\,\mathrm{d}s \\ &= \frac{1}{2\pi}\int \phi_w(s)\mathrm{e}^{-is(x-\tau_j)/h}\,\mathrm{d}s \\ &= w\left(\frac{x-\tau_i}{h}\right). \end{aligned}$$

The result now follows. Of course, the analogous statement for the multivariate density estimator is equally true. One has

$$\tilde{f}_{nh}(\mathbf{x}) := E[f_{nh}(\mathbf{x})|\mathcal{F}^\sigma] = \frac{1}{nh^p}\sum_{j=p}^{n} w\left(\frac{\mathbf{x} - (\log\sigma_j^2, \ldots, \log\sigma_{j-p+1}^2)}{h}\right).$$

Remark 3.7 Better bounds on the asymptotic variance in Theorem 3.1 can be obtained under stronger mixing conditions. Consider for instance *uniform mixing*. In this case, the mixing coefficient ϕ_t is defined for $t > 0$ as

$$\phi_t = \sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_t^\infty} |P(A|B) - P(A)|. \tag{24}$$

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

$$\alpha_t \leq \frac{1}{2}\phi_t.$$

See Doukhan [18] for this inequality and many other mixing properties. If $\{\sigma_t\}$ is uniform mixing with coefficient ϕ satisfying $\sum_{j=1}^\infty \phi(j)^{1/2} < \infty$, then the variance bound equation (21) can be replaced with the sharper bound equation (22) of Theorem 3.2. The proof of the latter bound runs similarly to the strong-mixing bound as given in section 5. The essential difference is that in equation (28) we use Theorem 17.2.3 of Ibragimov and Linnik [19] with $\tau = 0$ instead of a lemma by Deo [20], as in the proof of Theorem 2 in Masry [21]. The result is that we can now bound the term M_{nh} of equation (28) at a constant times $\sum_{j=1}^{n-p+1}\varphi_j^{1/2} E\, W_0^2$. After this step, the proof is essentially unchanged. Use the estimate $E\, W_0^2 \leq Ch^p||v||_2^2$ to finish the proof.

The bound on the variance in Theorem 3.2 cannot be improved upon by strengthening the assumption that σ is strongly mixing to uniform mixing.

Remark 3.8 An example of an observed process that is strongly mixing and that belongs to the first model class is a GARCH(p, q) process. It has been shown in Carasso and Chen [22] (see also ref. [23]) that such a process is β-mixing with exponentially decaying β-mixing coefficients. Hence, this process is also α-mixing, as the β-mixing coefficient $\beta_k = E \text{ ess sup}\{|P(A|\mathcal{F}_k^{\infty}) - P(A)|: A \in \mathcal{F}_{-\infty}^{0}\}$ satisfies the inequality $2\alpha_k \leq \beta - k$ (see ref. [18]). Notice that we also have that the assumption of Theorem 3.1 on the α's is satisfied in this case.

4. Numerical examples

In this section, we present some results that give an indication of the quality of the deconvolution estimator. In a number of simulation examples, we compared the obtained estimated density with the true density, and if this density is unknown, with an estimate of the true density based on direct observations. An example using real data closes this section. The kernel used to compute the estimates is the kernel (18) with the characteristic function (19). The bandwidths are chosen by hand. The estimates have been computed by fast Fourier transforms using the Mathematica 4.2 package.

4.1 *Simulations*

We present three estimates based on simulated samples. The processes σ are equal to a GARCH(1, 1) process, a process that switches between two GARCH based regimes and an AR(1)-type process described in De Vries [6].

Example 4.1 (a GARCH(1,1) process) The values of the σ process have been generated according to equation (2), with parameter values $p = q = 1, \alpha_0 = 1, \alpha_1 = 0.7$, and $\beta_1 = 0.2$. For this process, there are no explicit expressions for the true densities of σ^2 and $\log(\sigma^2)$. The sample size n is equal to 1000. In figure 1, the dashed line represents a direct kernel estimate of the density of $\log(\sigma^2)$ based on the 1000 realizations of the $\log(\sigma^2)$ process in the left plot.

Example 4.2 (a process with two regimes) In this example, the $\log(\sigma^2)$ process switches between two GARCH(1, 1) regimes of the type (2) that have been multiplied by a factor 0.3 to obtain a smaller signal-to-noise ratio. The corresponding parameter values for the first process

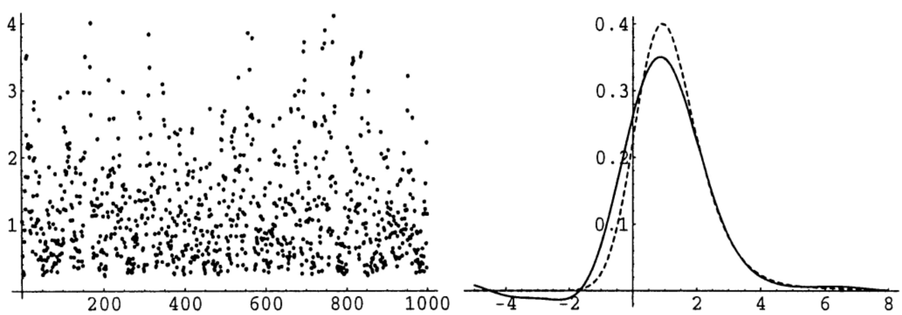

Figure 1. Left: the simulated values of $\log(\sigma^2)$ in Example 4.1. Right: estimate of the density of $\log(\sigma^2)$, $n = 1000$, $h = 0.4$. The solid line is the deconvolution estimate and the dashed line a kernel estimate based on the direct data in the left plot.

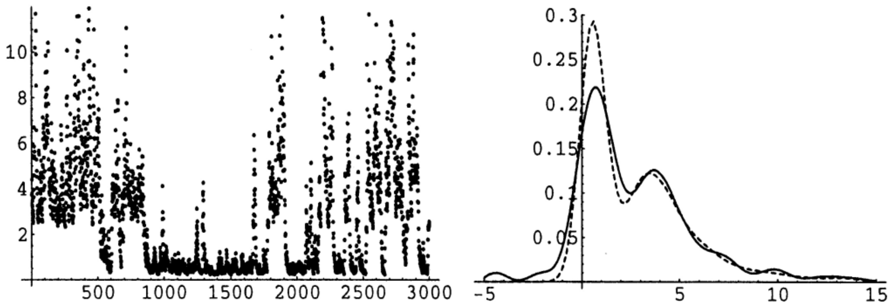

Figure 2. Left: the simulated values of $\log(\sigma^2)$ in Example 4.2. Right: estimate of the density of $\log(\sigma^2)$, $n = 3000$, $h = 0.3$. The solid line is the deconvolution estimate and the dashed line a kernel estimate based on direct data.

are $p = q = l, \alpha_0 = 1, \alpha_1 = 0.2$, and $\beta_1 = 0.7$ and for the second process $p = q = 1, \alpha_0 = 2$, $\alpha_1 = 0.2$, and $\beta_1 = 0.7$. Hence, the switch is in the α_0 parameter of the GARCH processes. At each time instant, a switch occurs with probability $p = 0.01$, independent of the running process. Hence, the time intervals between the switches have a geometric(p) distribution. Again for the resulting process, there are no explicit expressions for the true densities of σ^2 and $\log(\sigma^2)$, but they turn out to be bimodal. The sample size n is equals to 3000. In figure 2, the dashed line represents a direct kernel estimate of the density of $\log(\sigma^2)$ based on the 3000 realizations of the $\log(\sigma^2)$ process.

Example 4.3 (an AR(1) type process) Here, the σ^2 process has been generated according to a model described in De Vries [6]. In our notation, the model is as follows:

$$X_t = \sigma_t Z_t,$$
$$\sigma_t = \frac{1}{Z_{t-1}^2} + \alpha \sigma_{t-1}^2,$$

where the Zs are i.i.d. standard normal random variables and $0 \leq \alpha < 1$. We have chosen $\alpha = 0.5$. The sample size n is equal to 1000. The true density of σ^2 and $\log(\sigma^2)$ in this model are known. The graph of the density of $\log(\sigma^2)$ is represented by the dashed line in figure 3.

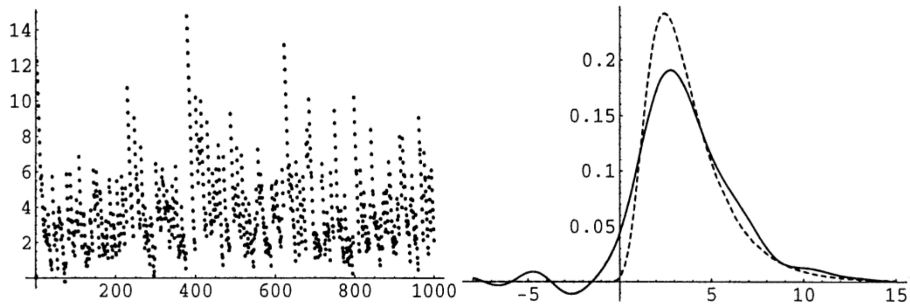

Figure 3. Left: the simulated values of $\log(\sigma^2)$ in Example 4.3. Right: The estimated density of $\log(\sigma^2)$, $n = 1000$, $h = 0.4$. The solid line is the deconvolution estimate and the dashed line the true density.

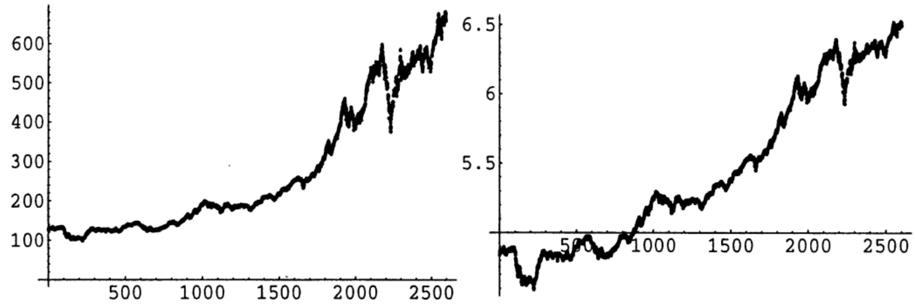

Figure 4. AEX. Left: daily closing values. Right: log of the daily closing values.

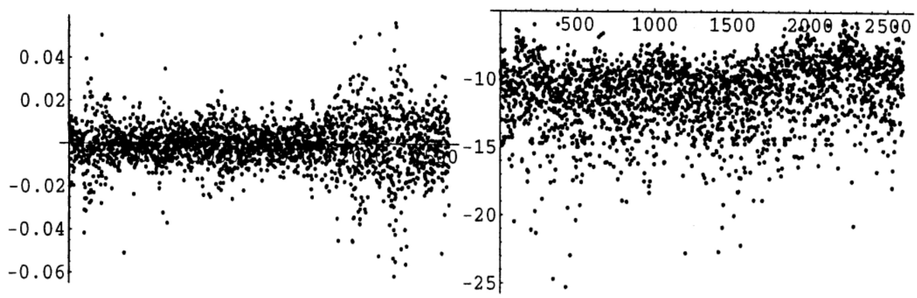

Figure 5. AEX. Left: the values of X_t, *i.e.*, the centered daily log returns. Right: $\log(X_t^2)$.

4.2 *Amsterdam AEX index*

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 4. We have centered the daily log returns, *i.e.*, we have subtracted the mean (which equaled 0.000636), see figure 5. We have computed the deconvolution estimator (with $h = 0.7$) and, based on computations of the mean and the variance of the estimate, we have also fitted a normal density by hand. The graphs of both estimators are given in figure 6 and show a remarkable resemblance.

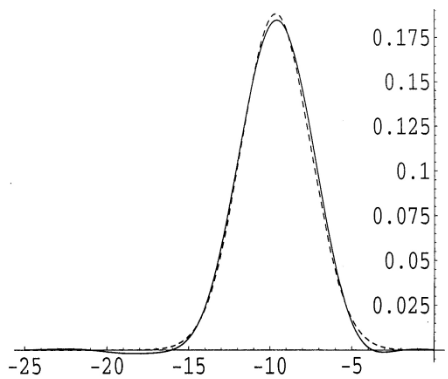

Figure 6. AEX. The deconvolution estimate of the density of $\log(\sigma_t^2)$ with $h = 0.7$ (the solid line) and its normal fit (the dashed line).

5. Proofs

All the estimators that we proposed involve the functions ϕ_k and ϕ_w. For these functions and related ones, we need expansions and order estimates. These will be given first.

LEMMA 5.1 *For $|t| \to \infty$, we have*

$$|\phi_k(t)| = \sqrt{2}\mathrm{e}^{-1/2\pi|t|}\left(1 + O\left(\frac{1}{|t|}\right)\right).$$

Proof By the Stirling formula for the complex gamma function (cf. [24] Chap. 6), we have

$$\log\Gamma(z) = \left(z - \frac{1}{2}\right)\log z - z + \frac{1}{2}\log 2\pi + O\left(\frac{1}{|z|}\right), \tag{25}$$

as $|z| \to \infty$ and $|\mathrm{Arg}\, z| < \pi$ for some $\delta > 0$. Therefore, for $z = 1/2 + it$ and $|t| \to \infty$, we get

$$\begin{aligned}
\log\Gamma\left(\frac{1}{2} + it\right) &= it\log\left(\frac{1}{2} + it\right) - \left(\frac{1}{2} + it\right) + \frac{1}{2}\log 2\pi + O\left(\frac{1}{|t|}\right)\\
&= it\left(\log\left|\frac{1}{2} + it\right| + i\mathrm{Arg}\left(\frac{1}{2} + it\right)\right) - \left(\frac{1}{2} + it\right) + \frac{1}{2}\log 2\pi + O\left(\frac{1}{|t|}\right)\\
&= -t\mathrm{Arg}\left(\frac{1}{2} + it\right) - \frac{1}{2} + \frac{1}{2}\log 2\pi + i\left(t\log\left|\frac{1}{2} + it\right| - t\right) + O\left(\frac{1}{|t|}\right).
\end{aligned}$$

Taking the modulus of the exponent the imaginary part vanishes and we get

$$\begin{aligned}
\left|\Gamma\left(\frac{1}{2} + it\right)\right| &= \exp\left(-t\ \mathrm{Arg}\left(\frac{1}{2} + it\right)\right) - \frac{1}{2} + \frac{1}{2}\log 2\pi + O\left(\frac{1}{|t|}\right)\\
&= \sqrt{2\pi}\exp\left(-t\ \arctan 2t - \frac{1}{2} + O\left(\frac{1}{|t|}\right)\right)\\
&= \sqrt{2\pi}\exp\left(-\frac{1}{2}\pi|t| + O\left(\frac{1}{|t|}\right)\right)\\
&= \sqrt{2\pi}\exp\left(-\frac{1}{2}\pi|t|\right)\left(1 + O\left(\frac{1}{|t|}\right)\right).
\end{aligned}$$

In the third equality, we have used the expansion $t\arctan t = t(\pi/2 - \arctan(1/t)) = \pi t/2 - 1 + O(1/t)$, as t tends to infinity. For negative t, a similar expansion holds. As $2^{it} = \exp(it\log 2)$ has modulus 1, substituting this expansion in equation (11) now proves the lemma. ■

LEMMA 5.2 *We have the following order estimate for the L^2 norms of the functions v_h of equation* (8) *and $\mathbf{v}_h$ of equation* (13). *For $h \to 0$*

$$||v_h||_2 = O(h^{1/2+\rho}\mathrm{e}^{\pi/2h})$$
$$||\mathbf{v}||_2 = O(h^{p(1/2+\rho)}\mathrm{e}^{p\pi/2h}).$$

Proof By Parseval's identity

$$2\pi||v_h||^2 = \int_{-1}^{1} \left| \frac{\phi_w(s)}{\phi_k(s/h)} \right|^2 \mathrm{d}s \le \frac{1}{2} \int_{-1}^{1} |\phi_w(s)|^2 w \mathrm{e}^{\pi|s/h|} \mathrm{d}s \tag{26}$$

$$+ \frac{1}{2} \int_{-1}^{1} |\phi_w(s)|^2 \left| \frac{2}{|\phi_k(s/h)|^2} - \mathrm{e}^{\pi|s/h|} \right| \mathrm{d}s \tag{27}$$

The integral in equation (26) can be rewritten as

$$\begin{aligned}\int_{-1}^{1} |\phi_w(s)|^2 \mathrm{e}^{\pi|s/h|} \mathrm{d}s &= 2\mathrm{e}^{\pi/h} \int_0^1 |\phi_w(s)|^2 \mathrm{e}^{\pi(s-1)/h} \mathrm{d}s \\ &= 2\mathrm{e}^{\pi/h} h \int_0^{1/h} |\phi_w(1-hv)|^2 \mathrm{e}^{-\pi v} \mathrm{d}v \\ &\sim 2\mathrm{e}^{\pi/h} h^{1+2\rho} A^2 \int_0^{\infty} v^{2\rho} \mathrm{e}^{-\pi v} \mathrm{d}v \\ &= 2\mathrm{e}^{\pi/h} h^{1+2\rho} \pi^{-1-2\rho} A^2 \Gamma(2\rho+1),\end{aligned}$$

by the dominated convergence theorem. We rewrite the integral (27) as

$$\begin{aligned}&\int_{-1}^{1} |\phi_w(s)|^2 \mathrm{e}^{\pi|s/h|} \left| \frac{2\mathrm{e}^{-\pi|s/h|}}{|\phi_k(s/h)|^2} - 1 \right| \mathrm{d}s \\ &= 2\mathrm{e}^{\pi/h} \int_0^1 |\phi_w(s)|^2 \left| \frac{2\mathrm{e}^{-\pi s/h}}{|\phi_k(s/h)|^2} - 1 \right| \mathrm{e}^{\pi(s-1)/h} \mathrm{d}s \\ &= 2h^{1+2\rho} \mathrm{e}^{\pi/h} \int_0^{1/h} \left| \frac{|\phi_w(1-hv)|}{(hv)^{\rho}} \right|^2 \left| \frac{2\mathrm{e}^{-\pi(1/h-v)}}{|\phi_k(1/h-v)|^2} - 1 \right| v^{2\rho} \mathrm{e}^{-\pi v} \mathrm{d}v \\ &= 2h^{1+2\rho} \mathrm{e}^{\pi/h} o(1),\end{aligned}$$

by the dominated convergence theorem. We have used the fact that both the functions $\phi_w(1-u)/u^{\rho}$, and $|(2\exp(-\pi u)/|\phi_k(u)|^2) - 1|$ are bounded and that the second function is of order $O(1/u)$ as u tends to infinity (see Lemma 5.1). This shows that the term (27) is negligible with respect to (26). The order estimate of $||\mathbf{v}||_2$ follows by the product form of $\mathbf{v}$. ■

Proof of Theorem 3.1 The expansion (20) follows from Theorem 1 in Masry [13]. To prove the variance bound (21), we argue as in the proof of Theorem 2 in the same paper. First, we give a bound on the variance in terms of the L_2-norm of the function $\mathbf{v}_h$, and then we exploit the asymptotic expansion of the characteristic function ϕ_k as given in Lemma 5.1 to get a sharper bound on the L_2-norm of $\mathbf{v}_h$ than Masry in his Proposition 3 by taking the behavior of ϕ_w at the boundary of its support into account. Some details follow.

Arguing as in Masry [13], we can show that

$$\mathrm{Var}\, f_{nh}(\mathbf{x}) = O\left(\frac{||\mathbf{v}_h||_2^2}{nh^p} + M_{nh} \right),$$

with (up to a multiplicative constant)

$$M_{nh} = \frac{1}{nh^{2p}} \sum_{j=p}^{n} \mathrm{Cov}(W_j, W_0), \tag{28}$$

where $W_j = \mathbf{v}_h((x - \log \mathbf{X}_j)/h)$.

Applying a lemma by Deo [17], we can bound for a strong-mixing process X with mixing coefficients α_j, the term M_{nh} by a constant (not depending on n and h) times

$$\frac{1}{nh^{2p}}\sum_{j=p}^{n}\alpha_{j-p+1}^{\beta}\{E\,|W_j|^{2/(1-\beta)}E\,|W_0|^{2/(1-\beta)}\}^{(1-\beta)/2},$$

which, by stationarity, becomes

$$\frac{1}{nh^{2p}}\sum_{j=p}^{n}\alpha_{j-p+1}^{\beta}(E\,|W_0|^{2/(1-\beta)})^{1-\beta}.$$

Now, observe that, by boundedness of the density of $\log X_j^2$, the term $E\,|W_0|^{2/(1-\beta)}$ can be bounded by a constant times $h^p||\mathbf{v}_h||_{2/(1-\beta)}^{2/(1-\beta)}$ and that we can therefore write

$$\operatorname{Var} f_{nh}(\mathbf{x}) = O\left(\frac{||\mathbf{v}_h||_2^2}{nh^p} + \sum_{j=p}^{n}\alpha_j^{\beta}\frac{||\mathbf{v}_h||_{2/(1-\beta)}^2}{nh^{p(1+\beta)}}\right).$$

The proof will be finished by application of Lemma 5.2, which gives the L_2^2-norm of $\mathbf{v}_h$, and an estimate of the $L^{2/(1-\beta)}$-norm of $\mathbf{v}_h$. For the latter one, we have the inequalities $||\mathbf{v}_h||_{2/(1-\beta)} \leq ||\mathbf{v}_h||_\infty^{\beta}||\mathbf{v}_h||_2^{1-\beta}$ and $||\mathbf{v}_h||_\infty \leq C||\mathbf{v}_h||_2$, for some constant C by the fact that ϕ_w has compact support. As a result, we get $||\mathbf{v}_h||_2/ \leq C||\mathbf{v}_h||_2$ and that M_{nh} is less than a constant times $||\mathbf{v}_h||_2/nh^{p(1+\beta)}$. The bound on $\operatorname{Var} f_{nh}(\mathbf{x})$ of Theorem 3.1 now follows. ∎

Proof of Theorem 3.2 Let $\mathcal{F}^\sigma$ be the σ-algebra generated by the process σ. We use the decomposition

$$\operatorname{Var} f_{nh}(\mathbf{x}) = E\operatorname{Var}(f_{nh}(\mathbf{x})|\mathcal{F}^\sigma) + \operatorname{Var}\tilde{f}_{nh}(\mathbf{x}), \tag{29}$$

with $\tilde{f}_{nh}(\mathbf{x})$ as in Remark 3.6. Now, we consider the first term in equation (29). Let $\mathbf{z}_j = (\log Z_j^2,\ldots,\log Z_{j-p+1})$ and $\mathbf{q}_j = (\log\sigma_j^2,\ldots,\log\sigma_{j-p+1}^2)$. As the Z_i are independent given $\mathcal{F}^\sigma$, we can bound the conditional variance by

$$\frac{1}{n^2h^{2p}}\sum_{j=p}^{n}E\left[\left(\mathbf{v}_h\left(\frac{\mathbf{x}-\mathbf{q}_j-\mathbf{z}_j}{h}\right)\right)^2|\mathcal{F}^\sigma\right]$$

which is by conditional independence and stationarity equal to

$$\frac{1}{nh^{2p}}\int\left(\mathbf{v}_h\left(\frac{\mathbf{x}-\mathbf{q}_0-\mathbf{z}}{h}\right)\right)^2k(\mathbf{z})\,\mathrm{d}\mathbf{z} \leq \frac{C}{nh^p}||\mathbf{v}_h||_2^2,$$

with C the maximum of $\mathbf{k}$, the density of $\mathbf{z}_0$. Therefore, the first term in equation (29) is of order $||\mathbf{v}_h||_2^2/nh^p$, Therefore $O(h^{p(1+2\rho)}\mathrm{e}^{p\pi/h}/nh^p)$.

The second term of (29) is treated next. We have with $U_j = \mathbf{w}((\mathbf{x}-\mathbf{q}_j)/h)$

$$\operatorname{Var}\tilde{f}_{nh}(\mathbf{x}) = \frac{1}{n^2h^{2p}}\sum_j\operatorname{Var}U_j + \frac{2}{n^2h^{2p}}\sum_{i<j}\operatorname{Cov}(U_i,U_j).$$

The first term reduces by stationarity to $1/(nh^{2p})\operatorname{Var}U_1$ which can be bounded by a constant times $\|\mathbf{w}\|_2^2/nh^p$, as $(\log\sigma_1^2,\ldots,\log\sigma_p^2)$ has by assumption a bounded density. For the second term we proceed as in the proof of Theorem 3.1. Using stationarity, we write it as

$$\frac{2}{n^2h^{2p}}\sum_{k=1}^{n}(n-k)\mathrm{Cov}(U_k, U_0).$$

We split the summation into two parts. In the first part, we consider

$$\sum_{k=1}^{p-1}(n-k)\mathrm{Cov}(U_k, U_0),$$

whose absolute value can be bounded in view of the Cauchy–Schwarz inequality and stationarity by $(p-1)n\,E\,U_0^2$, which is bounded by $(p-1)nh^p||\mathbf{w}||_2^2$.

The absolute value of the second part

$$\sum_{k=p}^{n}(n-k)\mathrm{Cov}(U_k, U_0)$$

can be bounded by invoking once more Deo's result by

$$n\sum_{k=p}^{n}\alpha_{k-p+1}^{\beta}(E\,|U_0|^{2/(1-\beta)})^{1-\beta},$$

which is less than

$$nh^{p(1-\beta)}||w||_{2/(1-\beta)}^2\sum_{k}\alpha_{k-p+1}^{\beta}.$$

Hence, we have that Var $\tilde{f}_{nh}(\mathbf{x})$ is of order $1/nh^{p(1+\beta)}$.

Combining the obtained order estimates for the two terms of equation (29) and using the L^2-norm of the function $\mathbf{v}_h$ gives the desired result. ■

References

[1] Ghysels, E., Harvey, A. and Renault, E., 1996, Stochastic volatility. In: G.S. Maddala, and C.R. Rao (Eds) *Handbook of Statistics*, Statistical Methods in Finance, Vol. 14 (Amsterdam: North Holland), pp. 119–191.

[2] Shephard, N., 1996, Statistical aspects of ARCH and stochastic volatility. In: D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (Eds) *Time Series Models in Econometrics, Finance and Other Fields* (London: Chapman & Hall), pp. 1–67.

[3] Bollerslev, T., 1986, Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **31**, 307–321.

[4] Bougerol, P. and Picard, N., 1992a, Strict stationarity of generalized autoregressive processes. *Annals of Probability*, **20**(4), 1714–1730.

[5] Bougerol, P. and Picard, N., 1992b, Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, **52**(1–2), 115–127.

[6] De Vries, C.G., 1991, On the relation between GARCH and stable processes. *Journal of Econometrics*, **48**, 313–324.

[7] Carroll, R. and Hall, P., 1988, Optimal rates of convergence for deconvoluting a density. *Journal of the American Statistical Association*, **83**, 1184–1186.

[8] Liu, M.C. and Taylor, R.L., 1989, A consistent nonparametric density estimator for the deconvolution problem *Canadian Jounral of Statistics*, **17**, 427–438.

[9] Stefanski, L.A., 1990, Rates of convergence of some estimators in a class of deconvolution problems. *Statistical and Probability Letters*, **9**, 229–235.

[10] Stefanski, L. and Carroll, R., 1990, Deconvoluting kernel density estimators. *Statistics*, **21**, 169–184.

[11] Wand, M.P. and Jones, M.C., 1995, *Kernel Smoothing* (London: Chapman and Hall).

[12] Fan, J., 1991, On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*, **19**, 1257–1272.

[13] Masry, E., 1991, Multivariate probability density deconvolution for stationary stochastic processes. *IEEE Transactions on Information Theory*, **37**, 1105–1115.

[14] Masry, E., 1993a, Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *Journal of Multivariate Analysis*, **44**, 47–68.

[15] Masry, E., 1993b, Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Process and their Applications*, **475**, 53–74.

[16] Wand, M.P., 1998, Finite sample performance of deconvolving kernel density estimators. *Statistical and Probability Letters*, **37**, 131–139.

[17] Bradley, R.C., 1986, Basic properties of strong mixing conditions. In: E. Eberlein and M.S. Taqqu (Eds) *Dependence in Probability and Statistics* (Boston: Birkhaüser).

[18] Doukhan, P., 1994, *Mixing, Properties and Examples* (New York: Springer-Verlag).

[19] Ibragimov, I.A. and Linnik, Yu.V., 1971, *Independent and Stationary Sequences of Random Variables* (Groningen: Wolters-Noordhoff).

[20] Deo, C.M., 1973, A note on empirical processes for strong mixing processes. *Annals of Probability*, **1**, 870–875.

[21] Masry, E., 1983, Probability density estimation from sampled data. *IEEE Transactions Information Theory*, **29**, 696–709.

[22] Carasso, M. and Chen, X., 2002, Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, **18**, 17–39.

[23] Boussama, F., 1998, Ergodicité, mélange et estimation dans les modèles GARCH. PhD thesis, Université Paris 7.

[24] Abramowitz, M. and Stegun, I., 1964, *Handbook of Mathematical Functions*, 9th edn (New York: Dover).