Penalized projection estimator for volatility density.

F. Comte[*], V. Genon-Catalot[†]

Abstract

We consider the problem of estimating the stationary density of the process V_t in the stochastic volatility model $dY_t = \sqrt{V_t}dW_t$ where W_t is a standard Brownian motion and V_t a Markov stationary mixing process. We propose a nonparametric adaptive strategy for which we give non asymptotic risk bounds. We discuss the resulting rate and show that it is quite good in some classical examples of volatility models.

Keywords. Adaptive estimation. Density deconvolution. Diffusion processes. Penalized projection estimator. Stochastic volatility

1 Introduction

In this paper, we consider a stochastic volatility (SV) model:

$$dY_t = \sqrt{V_t}dW_t, Y_0 = 0,$$

where W_t is a standard Brownian motion, independent of the process (V_t) assumed to be stationary and ergodic. Our aim is to estimate the stationary density of the so-called volatility process V. When Y is a log-price, this model is widely used in finance to take into account the stochastic feature of the volatility, which was modelled as a constant in the original Black and Scholes model.

Stochastic volatility models have been introduced in finance by Hull and White (1987) and popularized by Renault (1997) or Shephard (2005) and the references therein. Statistical properties of the volatility process or of the integrated volatility process have been studied by Genon-Catalot *et al.* (1999, 2000), by Bibby *et al.* (2004), by Gloter (2000a, b), by Gloter and

[*]Université Paris V, MAP5, UMR CNRS 8145.
Corresponding author: Fabienne Comte, MAP5 UMR 8145, Université René Descartes-Paris V, 45 rue des Saints Pères, 75270 PARIS cedex 06, FRANCE. email: fabienne.comte@univ-paris5.fr

[†]Université Paris V, MAP5, UMR CNRS 8145. email: Valentine.Genon-Catalot@math-info.univ-paris5.fr

Jacod (2001), in different settings: parametric rather than nonparametric, discrete observations with fixed or tending to zero time interval between two observations. In this work, we consider a discretely observed process at times $k\Delta_n$, for $k = 1, \dots, n$ where Δ_n tends to zero and $n\Delta_n$ tends to infinity when n tends to infinity and we are interested in the problem of the nonparametric estimation of the stationary density of the process (V_t).

The tools we use to solve the problem are related to the deconvolution literature. Indeed, this model can be written as if the variables with unknown density where not directly observed but observed up to an additive noise having known density. There is a wide literature on deconvolution methods, which were first based in the nonparametric setting on kernel estimators, see e.g. Carroll and Hall (1988), Stefansky and Carroll (1990), Fan (1991). These works were concerned with the estimation of densities with standard smoothness (i.e. belonging to Sobolev or Hölder spaces) and were constrained by minimax rates (Fan (1991)) which were very slow in the context of what is usually called "super-smooth" errors. Our model belongs to this class, and this explains why van Es *et al.* (2003) obtain logarithmic rates with the kernel estimator they study in the context of a SV model. But recent developments in this field show that, if the function to be estimated has the same kind of smoothness as the noise, then the rates can be significantly improved. This context has been first considered by Pensky and Vidakovic (1999) with a wavelet estimator which is adaptive in some cases. Lower bounds for the rates have been studied by Butucea (2004) and Butucea and Tsybakov (2004). The work by Comte *et al.* (2005a) proposes an adaptive projection estimator reaching automatically the optimal rates in most cases and with negligible loss in the other cases. Comte *et al.* (2005b) study intensively the very good finite sample properties of the estimator.

These improvements lead to a natural construction of a nonparametric, projection estimator in the context of the SV model seen as a convolution model. Heuristical justifications are easy to provide and the risk bounds can be much better than logarithmic, if the function to be estimated has the same smoothness as the noise. The procedure can be made adaptive by penalization of the projection contrast, leading to a data-driven selection of the projection space. The resulting estimator has very good theoretical properties in term of its mean square integrated risk.

The model and the assumptions are detailed in Section 2, the estimator is defined and heuristically discussed in Section 3. Section 4 studies the risk bound for the estimator whereas Section 5 generalizes the non asymptotic risk bound to an adaptive estimator. The rates are discussed and examples are provided in Section 6, for which we can check that the assumptions are fulfilled and compute the asymptotic rates. Most proofs are gathered in Section 7.

2 Model and assumptions

2.1 Stationarity conditions

Consider the process (Y_t) given by

$$dY_t = \sqrt{V_t}dW_t, \quad Y_0 = 0 \tag{1}$$

where:

($\mathbf{A}_0$) (W_t) is a Wiener process; (V_t) is a process with values in $(0, +\infty)$, independent of (W_t).

($\mathbf{A}_1$) (V_t) is a time-homogeneous Markov process, with continuous sample paths, strictly stationary and ergodic. Moreover its stationary (marginal) distribution admits a density $\pi(v)$ with respect to the Lebesgue measure on $(0, +\infty)$ (hereafter denoted by dv).

($\mathbf{A}_2$) Either (V_t) is β-mixing and its β-mixing coefficient β_V satisfies $\int_0^{+\infty} \beta_V(t)dt = A_2 < +\infty$; or (V_t) is ρ-mixing. In this case, we set $\int_0^{+\infty} \rho_V(t)dt = A_2(< +\infty)$ where ρ_V is the ρ-mixing coefficient.

Note that if V_t is ρ-mixing under **($\mathbf{A}_1$)**, then ρ_V tends to zero exponentially fast. Under **($\mathbf{A}_1$)**, if we denote by $P_t(v, dv')$ the transition probability of (V_t), then it is well-known that the β-mixing coefficient $\beta_V(t)$ of (V_t) has the following explicit expression that will allow to check condition **($\mathbf{A}_2$)** on our examples:

$$\beta_V(t) = \int_0^{+\infty} \pi(v)dv \|P_t(v, dv') - \pi(v')dv'\|_{TV}$$

where $\|\cdot\|_{TV}$ denotes the total variation distance between probability measures.

Below, we focus on the case where (V_t) is a diffusion process, defined by a stochastic differential equation:

$$dV_t = b(V_t)dt + a(V_t)dB_t, \quad V_0 = V, \tag{2}$$

where (W_t, B_t) is a Brownian motion of $\mathbb{R}^2$, and in that case, we can consider the following assumption:

($\mathbf{A'}_1$) (i) b and a are continuous real valued functions on $\mathbb{R}$ and C^1 functions on $(0, +\infty)$ with:

$$\exists K > 0, \qquad \forall v > 0, |b(v)| + |a(v)| \leq K(1 + v) \text{ and } \forall v > 0, a(v) > 0. \tag{3}$$

(ii) For $v_0 > 0$, let

$$s(v) = \exp\left[-2\int_{v_0}^{v} \frac{b(u)}{a^2(u)}du\right]$$

be the scale density of (2), then

$$\int_{0^+} s(v)dv = +\infty, \int^{+\infty} s(v)dv = +\infty, \int_0^{+\infty} dv/(a^2(v)s(v)) = M_0 < +\infty.$$

(iii) Let $\pi(v) = (M_0 a^2(v)s(v))^{-1}\mathbf{1}_{]0,+\infty[}(v)$, then the initial random variable V has distribution $\pi(v)dv$.

(iv) The random variable V is independent of the two-dimensional Brownian motion (W_t, B_t).

In this case, Assumption $(\mathbf{A}_1)$ holds, and (V_t) is automatically β-mixing, that is $\lim_{t\to+\infty} \beta_V(t) = 0$. The checking of the integrability condition of $\beta_V(t)$ may be obtained on specific examples under criteria detailed e.g. in Pardoux and Veretennikov (2001). Under $(\mathbf{A'}_1)$, there exists a necessary and sufficient condition (see Genon-Catalot *et al.* (2000)) to prove that the process is ρ-mixing and in that case, the mixing is exponential (i.e. $\rho_V(t) = e^{-\lambda t}, \forall t \geq 0$, with $\lambda > 0$), and is therefore such that $\int_0^{+\infty} \rho_V(t)dt < +\infty$.

2.2 Moment-type conditions

The sample path of (Y_t) is observed at regularly spaced instants $t_i = i\Delta$, $i = 1, \ldots, n$ with sampling interval Δ. From the observations $(Y_{i\Delta}, 1 \leq i \leq n)$, we want to estimate the stationary density π of the diffusion (V_t). More precisely, we shall build a nonparametric estimator of the stationary density f of

$$X_t = \ln(V_t). \tag{4}$$

This density is linked with π through the relation $f(x) = e^x \pi(e^x)$.
Let us first note that, setting, for $\Delta > 0$,

$$\bar{V}_i = \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} V_s ds, \qquad \bar{X}_i = \ln(\bar{V}_i). \tag{5}$$

we have, using $(\mathbf{A}_0)$,

$$\frac{1}{\sqrt{\Delta}}(Y_{i\Delta} - Y_{(i-1)\Delta}) = \bar{V}_i^{1/2}\varepsilon_i$$

where $(\varepsilon_i, i = 1, \ldots, n)$ are independent identically distributed random variables having distribution $\mathcal{N}(0,1)$, independent of $(V_t, t \geq 0)$. Now we set $\eta_i = \ln(\varepsilon_i^2)$ and we consider the equation

$$Z_i = \ln\left\{\left[\frac{1}{\sqrt{\Delta}}(Y_{i\Delta} - Y_{(i-1)\Delta})\right]^2\right\} = \ln(\bar{V}_i) + \eta_i = \bar{X}_i + \eta_i. \tag{6}$$

The following assumptions on f and on the moments of X_0 are required:

$(\mathbf{A}_3)$ $M_f = \sup_x f(x) < +\infty$.

($\mathbf{A}_4$) $\mathbb{E}[\ln^2(V_0)] = \mathbb{E}(X_0^2) < +\infty$ and $\mathbb{E}\{[\ln(\bar{V}_1) - \ln(V_0)]^2\} \leq c\Delta$, $\Delta \leq 1$.

($\mathbf{A}_5$) The variables $\bar{V}_i$ (and therefore $\bar{X}_i$) admit a density, we denote by g_Δ the density of $\bar{X}_i$ and we assume that g_Δ belongs to $\mathbb{L}^2(\mathbb{R})$.

Comments on the assumptions.

• Assumption **($\mathbf{A}_3$)** is not very strong and can be slightly weakened since we only need that

$$\int x^2 f^2(x)dx < C_f < \infty. \tag{7}$$

Under **($\mathbf{A}_3$)** and the first part of **($\mathbf{A}_4$)**, $C_f = M_f \mathbb{E}(X_0^2)$ suits. Moreover, this ensures that f belongs to $\mathbb{L}^2(\mathbb{R})$.

• Assumption **($\mathbf{A}_4$)** can be checked by using one of the lemmas:

Lemma 2.1 *Let (X_t) be a strictly stationary and ergodic diffusion process on $\mathbb{R}$ satisfying $dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$ with $\mu, \sigma \in C^1(\mathbb{R})$, $|\mu(x)| + |\sigma(x)| \leq K(1+|x|), \forall x \in \mathbb{R}$, and $\mathbb{E}(X_0^2) < +\infty$, then* **($\mathbf{A}_4$)** *holds for $V_t = e^{X_t}$.*

Lemma 2.2 *Assume that* **($\mathbf{A'}_1$)** *holds, that $\mathbb{E}(V_0^4) + \mathbb{E}[\sup_{t\in[0,1]} 1/V_t^4] < +\infty$. Then* **($\mathbf{A}_4$)** *is fulfilled.*

In particular, we use some bounds given in Gloter (2000b) implying that $\mathbb{E}[\sup_{t\in[0,\Delta]}(1/V_t^4)] \leq c\mathbb{E}[1/V_0^4]$ in almost all the cases studied in the examples below (see Section 6).

• Let us now turn to Assumption **($\mathbf{A}_5$)**. On the one hand, the existence and the properties of such a density is known in some particular cases (for instance, for the Cox-Ingersoll-Ross process, see Cox *et al.* (1985)). On the other hand, for diffusion processes given by (2), the existence of g_Δ can be obtained as a consequence of the existence of a transition density for the two-dimensional diffusion process $(V_t, \int_0^t V_s ds)$. This in turn is obtained by checking the Hörmander condition for $(V_t, \int_0^t V_s ds)$ (see Gloter (2000a)). For instance, we may use the following proposition:

Proposition 2.1 *Assume that V_t is solution of (2) with a and b infinitely differentiable with bounded derivatives of any order $n \geq 1$. Assume moreover that for any $x \in \mathbb{R}$, $a(x) \neq 0$ or ($b(x) \neq 0$ and $\exists n \geq 1, a^{(n)}(x) \neq 0$). Then $\bar{V}_1$ admits a density.*

Concerning the square integrability condition, an available criterion is to check that

$$\sup_{u\in\mathbb{R}} |u^2 \mathbb{E}(e^{iu\bar{X}_1})| < +\infty.$$

Such a condition implies both that g_Δ exists and is bounded. This criterion is natural to get regularity conditions on densities. In our case, it can be checked by means of Malliavin calculus through an integration by part formula, as it is done e.g. in Yoshida (1997).

3 The estimation method

Equation (6) suggests to build first, by a deconvolution approach a nonparametric estimator of g_Δ. Then, we will simply consider it as an estimator of f and study its asymptotic properties as $\Delta = \Delta_n$ tends to zero.

3.1 Some notations

We denote by $u^*(x) = \int e^{itx}u(t)dt$ the Fourier transform of a square-integrable function, by $\|u\|^2 = \int |u(x)|^2 dx$ the square of the $\mathbb{L}^2(\mathbb{R})$-norm (for possibly complex valued functions) and by $\langle u, v\rangle = \int u(x)\bar{v}(x)dx$ the scalar product ($\bar{z}$ denotes the conjugate of z). We recall that for any two functions $u(x)$ and $v(x)$ in $\mathbb{L}^2(\mathbb{R})$ and complex valued, the following relation holds

$$\langle u, v\rangle = \int u(x)\bar{v}(x)dx = \frac{1}{2\pi}\langle u^*, v^*\rangle.$$

If u is real valued, $u^*(-.) = \bar{u}^*(.)$ and if u^* is in $\mathbb{L}^2(\mathbb{R})$ then $(u^*)^*(.) = 2\pi u(-.)$. We also use the following notation for the convolution: $u \star v(x) = \int u(t)v(x-t)dt$, and recall that $(u \star v)^* = u^* v^*$.

3.2 Heuristical approach

For any function t belonging to $\mathbb{L}^2(\mathbb{R})$, such that t^* is compactly supported, we can define

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n}\sum_{k=1}^{n} u_t(Z_k), \text{ where } u_t(x) = \frac{1}{2\pi}\int e^{iux}\frac{t^*(-u)}{f_\eta^*(u)}du = \frac{1}{2\pi}\left[\frac{t^*(-.)}{f_\eta^*(.)}\right]^*, \tag{8}$$

where the Z_k's are defined by (6).

Elementary computations lead to the following density for $\eta_i = \ln(\varepsilon_i^2)$ with $\varepsilon_i \sim \mathcal{N}(0,1)$:

$$f_\eta(x) = \frac{1}{\sqrt{2\pi}}e^{x/2}e^{-e^x/2}$$

and it is proved in Comte (2004) that

$$f_\eta^*(t) = \frac{1}{\sqrt{\pi}}2^{it}\Gamma(\frac{1}{2} + it), |f_\eta^*(t)| = \sqrt{2/e}\ \ e^{-\pi|t|/2}\left(1 + O(1/|t|)\right) \text{ for } |t| \to +\infty. \tag{9}$$

The following notation is useful:

$$\Phi_\eta(L) = \int_{|x|\le \pi L}\frac{dx}{|f_\eta^*(x)|^2}. \tag{10}$$

Observe that $\mathbb{E}(u_t(Z_k)) = \langle u_t, g_\Delta \star f_\eta\rangle = \frac{1}{2\pi}\langle u_t^*, g_\Delta^* f_\eta^*\rangle$. Due to the definition of u_t, $u_t^* = t^*(.)/f_\eta^*(-.)$ and therefore

$$\mathbb{E}(u_t(Z_k)) \;=\; \frac{1}{2\pi}\langle\frac{t^*(.)}{f_\eta^*(-.)}, g_\Delta^* f_\eta^*\rangle = \frac{1}{2\pi}\int \frac{t^*(u)}{f_\eta^*(-u)}g_\Delta^*(-u)f_\eta^*(-u)du = \frac{1}{2\pi}\langle t^*, g_\Delta^*\rangle = \langle t, g_\Delta\rangle.$$

Consequently $\mathbb{E}\gamma_n(t) = \|t\|^2 - 2\langle t, g_\Delta\rangle = \|t - g_\Delta\|^2 - \|g_\Delta\|^2$. This equality justifies the choice of an estimator of g_Δ obtained by minimizing the criterion $t \mapsto \gamma_n(t)$. Actually, our aim is to estimate f (and not g_Δ). Proposition 3.1 below explains why the choice of γ_n is relevant to estimate f.

3.3 The contrast property.

Let us first state a simple convergence result that explains why our function γ_n may be called a contrast function for the estimation of f.

Proposition 3.1 *Under Assumptions* $(\mathbf{A}_0)$-$(\mathbf{A}_5)$*, for any function* t *in* $\mathbb{L}^2(\mathbb{R})$ *such that* t^* *is compactly supported, then*

$$\frac{1}{n}\sum_{k=1}^{n} u_t(Z_k) \to \langle t, f\rangle \text{ in probability, if } n \to +\infty, \Delta = \Delta_n \to 0, n\Delta_n \to +\infty.$$

This implies straightforwardly if f^* is compactly supported that

$$\gamma_n(t) - \gamma_n(f) \to \|t - f\|^2 \text{ in probability, if } n \to +\infty, \Delta = \Delta_n \to 0, n\Delta_n \to +\infty.$$

The constraint of compact support for t^* or f^* is strong and due the exponential decay of f_η^* near infinity (see equation (9)). But, as we shall see below, this constraint is not useful for f. In particular $\gamma_n(f)$ is not assumed to exist in the following. However, this explains why the functions t which are considered are chosen such that t^* is compactly supported. These considerations just give a glance to the heuristic of the nonparametric contrast method.

The proof of Proposition 3.1 is obtained as a by-product of the general study of our estimator below. We extract this simple result to establish the link with parametric methods based on contrast functions (see e.g. Genon-catalot *et al.* (1999)).

3.4 The projection method

To construct our estimator of the density f, we use a projection method on an adequate sequence of subspaces of $\mathbb{L}^2(\mathbb{R})$.

Let $\varphi(x) = \sin(\pi x)/(\pi x)$ so that $\varphi^*(x) = \mathbb{I}_{[-\pi,\pi]}(x)$. For $L \in \mathbb{N}$ and $j \in \mathbb{Z}$, set $\varphi_{L,j} = \sqrt{L}\varphi(Lx - j)$. The functions $\{\varphi_{L,j}, j \in \mathbb{Z}\}$ constitute an orthonormal system in $\mathbb{L}^2(\mathbb{R})$. For $L = 2^k$, it is also known as the Shannon basis. Moreover, we choose here integer values of L but a grid with smaller step (for instance with step 1/10 or 1/4) would also be possible. Let us define

$$S_L = \mathrm{V}ect\,(\varphi_{L,j},\ \ j \in \mathbb{Z})\,, L \in \mathbb{N}.$$

The space S_L is exactly the subspace of $\mathbb{L}^2(\mathbb{R})$ of the square integrable functions t with Fourier transform compactly supported with support $[-\pi L, \pi L]$. It can be proved that if f is a square integrable function, then $f_L^* = f^* \mathbb{I}_{[-\pi L, \pi L]}$ and this yields:

$$\|f - f_L\|^2 = \frac{1}{2\pi} \int_{|x| \geq \pi L} |f^*(x)|^2 dx.$$

This evaluation of what becomes in the following a squared bias term is most useful.

Thus, for all $t \in S_L$, u_t is well defined in spite of the exponential order of $1/f_\eta^*$ (see (8) and (9)). We will define an estimator $\hat{f}_L$ as an element of S_L. To obtain a representation of $\hat{f}_L$ having a finite number of "coordinates", we must introduce

$$S_L^{(n)} = \mathrm{V}ect\left(\varphi_{L,j}, \quad |j| \leq K_n\right). \tag{11}$$

Now we may define

$$\hat{f}_L = \arg \min_{t \in S_L^{(n)}} \gamma_n(t). \tag{12}$$

These estimators admit a very simple expression:

$$\hat{f}_L = \sum_{|j| \leq K_n} \hat{a}_{L,j} \varphi_{L,j}, \text{ with } \hat{a}_{L,j} = \frac{1}{n} \sum_{k=1}^{n} u_{\varphi_{L,j}}(Z_k). \tag{13}$$

In a second step, the problem of the optimal and automatic choice of L leads to consider a penalized contrast estimator

$$\hat{L} = \arg \min_{L \in \{1, \ldots, m_n\}} \left[\gamma_n(\hat{f}_L) + \mathrm{pen}(L)\right].$$

This procedure selects a relevant estimator in the collection $(\hat{f}_L)_{L \in \{1, \ldots, m_n\}}$ provided that the penalty function pen is well chosen.

4 Quadratic risk of the estimators

This section is devoted to the study of the quadratic risk of $\hat{f}_L$, $\mathcal{R}(\hat{f}_L, f) = \mathbb{E}(\|\hat{f}_L - f\|^2)$, leading to an evaluation of the optimal L.

4.1 Decomposition of the risk

We denote in the following by f_L and by $f_L^{(n)}$ the orthogonal projections of f on S_L and on $S_L^{(n)}$ respectively. Let

$$\nu_n(t) = \frac{1}{n} \sum_{k=1}^{n} (u_t(Z_k) - \langle t, g_\Delta \rangle).$$

For t in $S_L^{(n)}$, $\gamma_n(t)$ is well defined. Therefore, for $s, t \in S_L^{(n)}$, we have the following decomposition of the contrast:

$$\gamma_n(t) - \gamma_n(s) = \|t - f\|^2 - \|s - f\|^2 - 2\nu_n(t - s) + 2\langle t - s, f - g_\Delta\rangle.$$

Then since, by (12), $\gamma_n(\hat{f}_L) \leq \gamma_n(f_L^{(n)})$, we obtain

$$\|\hat{f}_L - f\|^2 \leq \|f - f_L^{(n)}\|^2 + 2\nu_n(\hat{f}_L - f_L^{(n)}) + 2\langle \hat{f}_L - f_L^{(n)}, g_\Delta - f\rangle.$$

The first right-hand-side term is the standard deterministic projection bias term, the second one corresponds to a kind of variance term and the last one to a random bias term due to the implicit approximation of f by g_Δ.

4.2 Projection bias term

The regularity conditions for f are described as follows and the functions considered in our examples will belong to such classes.

(Reg) There exist some positive real numbers s, r, b such that the density f belongs to

$$\mathcal{S}_{s,r,b}(M) = \left\{ t \text{ density } : \int_{-\infty}^{+\infty} |t^*(x)|^2 (x^2+1)^s \exp\{2b|x|^r\} dx \leq M \right\}.$$

Note that densities satisfying **(Reg)** with $r = 0$ belong to some Sobolev class of order s, whereas densities satisfying **(Reg)** with $r > 0, b > 0$ are infinitely many times differentiable. Moreover, such densities admit analytic continuation on a finite width strip when $r = 1$ and on the whole complex plane if $r = 2$.

We know that under Assumption ($\mathbf{A}_3$) and the first part of ($\mathbf{A}_4$), f is square integrable, so

$$\|f - f_L^{(n)}\|^2 \leq \|f - f_L\|^2 + \|f_L - f_L^{(n)}\|^2. \tag{14}$$

The term $\|f - f_L\|^2$ depends on the smoothness of the function f and has the standard order for classical smoothness classes since it is given by the distance between f and the classes of entire functions having Fourier transform compactly supported on $[-\pi L, \pi L]$ (see Ibragimov and Hasminskii (1983)). More precisely, the following inequalities hold:

Proposition 4.1 *Assume that f belongs to a class $\mathcal{S}_{s,r,b}(M)$, that* ($\mathbf{A}_3$) *is fulfilled and that $\mathbb{E}[\ln^2(V_0)] = \mathbb{E}(X_0^2) < +\infty$, then*

$$\|f - f_L\|^2 \leq \frac{M}{2\pi}(L^2\pi^2 + 1)^{-s} \exp\{-2b(\pi L)^r\}, \quad \|f_L - f_L^{(n)}\| \leq \frac{(1 + C_f^{1/2})L^2}{K_n}, \tag{15}$$

where C_f is defined by (7).

In our case, the values of L which must be considered are of order at most $\ln(n)$. This implies that the term of order L^2/K_n is less than $\ln^2(n)/K_n$ and is negligible as soon as K_n is of order n. The choice of K_n is free from a theoretical point of view, but in practice, the larger K_n, the slower our estimation procedure.

4.3 Bound for the risk

The two other terms are controlled as follows:

Lemma 4.1 *Under Assumptions* $(\mathbf{A}_0)$-$(\mathbf{A}_2)$, $(\mathbf{A}_4)$ *and* $(\mathbf{A}_5)$,

$$2\mathbb{E}(\langle \hat{f}_L - f_L^{(n)}, g_\Delta - f\rangle) \leq \frac{1}{4}\mathbb{E}(\|\hat{f}_L - f\|^2) + \frac{1}{4}\|f - f_L^{(n)}\|^2 + C\Delta L^3. \tag{16}$$

Lemma 4.2 *Under Assumptions* $(\mathbf{A}_0)$-$(\mathbf{A}_2)$ *and* $(\mathbf{A}_5)$,

$$2\mathbb{E}[\nu_n(\hat{f}_L - f_L^{(n)})] \leq \frac{1}{2}\mathbb{E}\|\hat{f}_L - f\|^2 + \frac{1}{2}\|f - f_L^{(n)}\|^2 + \frac{2}{\pi n}\Phi_\eta(L) + \frac{8\pi L(1+A_2)}{n\Delta}, \tag{17}$$

where $\Phi_\eta(L)$ *is defined by (10).*

By gathering the bounds in (14), (16) and (17), we obtain the following result:

Theorem 4.1 *Under Assumptions* $(\mathbf{A}_0)$-$(\mathbf{A}_5)$,

$$\mathbb{E}(\|\hat{f}_L - f\|^2) \leq 7\|f - f_L\|^2 + \frac{8}{\pi n}\Phi_\eta(L) + \frac{CL^2}{K_n} + \frac{C'L(1+A_2)}{n\Delta} + C"\Delta L^3.$$

The first three right-hand-side terms are the usual ones in deconvolution theory, and the last ones are due to the implicit approximation of f by g_Δ.
Therefore, if f belongs to a class described by $\mathcal{S}_{s,r,b}(M)$, then the bound for $\|f - f_L\|^2$ is given by (15). On the other hand, it follows from (9) that

$$\Phi_\eta(L) = \int_{-\pi L}^{\pi L} \frac{dx}{|f_\eta^*(x)|^2} \leq Ce^{\pi^2 L}.$$

Note that we have as a consequence the consistency of our estimator:

Corollary 4.1 *Under Assumptions* $(\mathbf{A}_0)$-$(\mathbf{A}_5)$, *and if* $K_n \geq n$, *and* $L = L_n \leq \ln(n)/(2\pi^2)$ *with* $L_n \to +\infty$ *when* $n \to +\infty$, *then the quadratic risk* $\mathbb{E}(\|\hat{f}_L - f\|^2)$ *tends to zero if* $\Delta = \Delta_n$ *is such that* $\Delta_n \ln^3(n) \to 0$ *and* $n\Delta_n/\ln(n) \to +\infty$ *when* $n \to +\infty$.

Some rates can be deduced from Theorem 4.1.

4.4 Discussion about the rates

We want to illustrate here that Sobolev classes are not *large enough* for this kind of problem and more precisely in deconvolution. Indeed, with a noise of super-smooth type as we have here (i.e. with exponential rate of decay of its Fourier transform), the optimal rates for the estimation of a function belonging to a classical Sobolev class are logarithmic. An a priori selection of the space (i.e. of L) can then be done.

Now if more general classes are considered, and if the Fourier transform of the function to estimate is allowed to have *the same rate of decay as the noise*, then the rates can be improved and are much faster than logarithmic. However, these rates are very complicated to compute (see the discussion in Comte *et al.* (2005a)). Moreover the choice of the optimal L depends on the unknown function. Thus, since the regularity of the function to estimate is unknown (in particular it may or may not be of Sobolev type), an adaptive procedure is required to select L. An additional advantage of the adaptive selection of L is that it avoids the uneasy computation of the optimal value of L. Actually, the adaptive procedure automatically selects a value of L close to the optimal one. At the same time, it also makes the bias-variance trade-off. Let us show it more precisely by some examples.

If f belongs to a Sobolev space with order s ($b = r = 0$) then, by using (15) and Theorem 4.1, the orders are the following, for $K_n = n$:

$$\mathbb{E}(\|\hat{f}_L - f\|^2) \leq C_1 L^{-2s} + C_2 \frac{e^{\pi^2 L}}{n} + \frac{C_4 L^2}{n} + \frac{C_3 L}{n\Delta} + C_5 \Delta L^3.$$

In that case, choose $L = \ln(n)/(2\pi^2)$ and this yields

$$\mathbb{E}(\|\hat{f}_L - f\|^2) \leq C_1'[\ln(n)]^{-2s} + \frac{C_2'}{\sqrt{n}} + \frac{C_3' \ln(n)}{n\Delta} + C_4' \Delta \ln(n)^3.$$

If we have $\Delta = \Delta = n^{-\delta}$ for $0 < \delta < 1$,

$$\mathbb{E}(\|\hat{f}_L - f\|^2) \leq C_1'[\ln(n)]^{-2s} + \frac{C_2'}{\sqrt{n}} + \frac{C_3' \ln(n)}{n^{1-\delta}} + C_4' \frac{\ln(n)^3}{n^\delta}.$$

In that case, the discretization terms are negligible with respect to the deconvolution bias. The rate $[\ln(n)]^{-2s}$ corresponds to the optimal deconvolution rate for a density belonging to this class of regularity (see Fan (1991)). This is the rate exhibited by van Es. *et al.* (2003) with a non adaptive kernel method and $s = 2$.

Now, if f is supersmooth i.e. $b \neq 0$, $r \neq 0$, then the rate can be improved. Let us take $s = 0$ for simplicity. By using (15) and Theorem 4.1 again, the orders are now the following, for $K_n = n$:

$$\mathbb{E}(\|\hat{f}_L - f\|^2) \leq C_1 e^{-2b(\pi L)^r} + C_2 \frac{e^{\pi^2 L}}{n} + \frac{C_4 L^2}{n} + \frac{C_3 L}{n\Delta} + C_5 \Delta L^3.$$

The deconvolution rate can reach any of the rates $n^{-\alpha}$ for $0 < \alpha \leq 1$. For instance, if $r = 1$, the order of the bias is $e^{-\beta L}$. The optimal choice of L is $L = \ln(n)/(\beta + \pi^2)$ and the associated deconvolution rate is $n^{-\beta/(\beta+\pi^2)}$. This is faster than $n^{-1/2}$ if $\beta \geq \pi^2$ and slower if $\beta < \pi^2$.

Then, in this case, to obtain the global rate, the discretization terms must be taken into account. For these terms, the better reachable rate is $\ln^2(n)/\sqrt{n}$ and is obtained for $\Delta_n = 1/[\ln(n)\sqrt{n}]$. The global rate is the worse of this rate and the deconvolution rate. In the previous situation of a bias of order $e^{-\beta L}$, the rate is the optimal deconvolution rate if $\beta < \pi^2$ and is $\ln^2(n)/\sqrt{n}$ if $\beta > \pi^2$.

In any case, the rates in this class of functions are much better than logarithmic and can reach the upper bound $\ln^2(n)/\sqrt{n}$. This is illustrated in the examples below that describe some standard volatility models.

4.5 Some closely related noises.

Note that the noise η has the law of $\ln(\mathcal{N}(0,1)^2) = \ln(\chi^2(1)) = \ln(G(1/2,1/2))$ where G denotes the Gamma distribution. Our result may be extended to the case where $\eta \sim \ln(G(a,\lambda))$. In that case,

$$f_\eta(x) = e^{-\lambda e^x} e^{ax} \frac{\lambda^a}{\Gamma(a)}.$$

Then it is also easy to see that $f_\eta^*(x) = \lambda^{ix}\Gamma(a+ix)/\Gamma(a)$. By using the Stirling formula, we find $|f_\eta^*(x)| \sim_{+\infty} \sqrt{2\pi}e^{-a}|x|^{a-1/2}e^{-\pi|x|/2}/\Gamma(a)$, which yields

$$\Phi_\eta(L) = \int_{-\pi L}^{\pi L} dx/|f_\eta^*(x)|^2 = O((\pi L)^{1-2a}e^{\pi^2 L}).$$

In other words, results as Theorem 4.1 are still valid for that type of extended noise, and the order of the corresponding term is slightly improved when $a > 1/2$. Nevertheless, the improvement in the rate is of a logarithmic factor.

5 Adaptive procedure: data driven selection of L

5.1 Main result

The general formula for the choice of L as a function of s, b and r is not easy to describe. Some examples are given in Comte *et al.* (2005a) that illustrate that the optimal choice of L is different when $r \leq 1/2$ or $1/2 < r \leq 2/3$, and more generally depends on k such that $k/(k+1) < r \leq (k+1)/(k+2)$. It follows that the selection of L needs to be automatic. The adaptive procedure is based on the following data-driven choice:

$$\hat{L} = \underset{L\in\{1,\ldots,\ln(n)\}}{\arg\min} \left\{\gamma_n(\hat{f}_L) + \mathrm{pen}(L)\right\}, \quad \mathrm{pen}(L) = \kappa\frac{(1+L)\Phi_\eta(L)}{n}, \tag{18}$$

where pen(L) stands for a penalty function and κ is a numerical constant. This criterion is deduced from the results of Comte *et al.* (2005a) in a pure deconvolution framework.
Theorem 5.1 shows that the optimization of the deconvolution terms (first two terms in the upper bound of Theorem 4.1) can be done automatically, and in a nonasymptotic way. This is useful because, even if in the examples of Section 6 we compute asymptotic rates, the selected dimensions are of logarithmic order and in practice often very small. For illustrations of this fact, we refer to Comte *et al.* (2005b).

Theorem 5.1 *Assume that Assumptions* $(\mathbf{A}_0)$-$(\mathbf{A}_5)$ *are fulfilled, then the estimator* $\hat{f}_{\hat{L}}$ *defined by (12) and (18) for some universal constant* κ *with a collection* $L \le \ln(n)$*, satisfies*

$$\begin{aligned}\mathbb{E}(\|\hat{f}_{\hat{L}} - f\|^2) &\le K \inf_{L\in\{1,\ldots,\ln(n)\}} \left(\|f - f_L\|^2 + \frac{(1+L)\Phi_\eta(L)}{n}\right) \\ &\quad + \frac{C'\ln(n)^2}{K_n} + \frac{4\pi\ln(n)(1+A_2)}{n\Delta_n} + C\Delta_n\ln(n)^3. \end{aligned} \tag{19}$$

Note that a (standard) loss of order L occurs in the penalty and consequently in the variance term of (19), compared to the variance term of Theorem 4.1. This implies a small loss in the rate, which can be proved to be always negligible with respect to the order of the rate. The asymptotic order for great L of the penalty is therefore $Le^{\pi^2 L}/n$ and the rate depends on the squared bias-variance compromise. Some examples follow in Section 6.

5.2 Estimation of the integrated volatility

A process of interest is often the integrated volatility itself, rather than the volatility. It is interesting from this point of view to mention that our method may be directly considered as a method of estimation of g_Δ itself. A consequence of the previous bounds is that, under Assumptions $(\mathbf{A}_0)$-$(\mathbf{A}_2)$, if g_Δ exists, is bounded, and if $\mathbb{E}(\ln^2(\bar{V}_1)) < +\infty$, then for $K_n \ge n\ln(n)$, $L \le \ln(n)$ and for any fixed Δ,

$$\begin{aligned}\mathbb{E}(\|\hat{f}_{\hat{L}} - g_\Delta\|^2) &\le K \inf_{L\in\{1,\ldots,\ln(n)\}} \left(\|g_\Delta - g_{\Delta,L}\|^2 + \frac{(1+L)\Phi_\eta(L)}{n}\right) \\ &\quad + \frac{C'\ln(n)}{n} + \frac{4\pi\ln(n)(1+A_2)}{n\Delta}. \end{aligned} \tag{20}$$

Here $g_{\Delta,L}$ denotes the orthogonal projection of g_Δ on S_L.

It follows that $\hat{f}_{\hat{L}}$ is also an adaptive estimator of g_Δ, and in this context, the rate is better when Δ is chosen as a fixed positive number, rather than as a function of n. The last two terms of inequality (20) are then negligible, and the rate is the standard deconvolution rate.
To obtain information on the regularity properties of g_Δ, we may check that $\sup_{u\in\mathbb{R}} |u^k \mathbb{E}(e^{iu\bar{X}_1})|$ is finite. See the comments on Assumption $(\mathbf{A}_5)$ in Section 2.2.

6 Examples

In the following, we consider model (1) under $(\mathbf{A}_0)$ and we propose different models for the volatility V_t. In each case, we check Assumptions $(\mathbf{A}_1)$-$(\mathbf{A}_4)$ and compute the rate of the adaptive estimator of f. For Assumption $(\mathbf{A}_5)$, we refer to the general comments and to Proposition 2.1 in Section 2.2. The checking of $(\mathbf{A}_2)$ requires a special comment: the β-mixing and the ρ-mixing coefficients are not comparable. Depending on the model, it may be easier to check $(\mathbf{A}_2)$ via the β- or the ρ-mixing coefficient. For the β-mixing coefficient we rely on a useful proposition given in Pardoux and Veretennikov (2001). For the ρ-mixing coefficient, we rely on a proposition given in Genon-Catalot *et al.* (2000). Note that $(\mathbf{A}_2)$ may hold for both the β and the ρ-mixing coefficient, or only for the β-mixing coefficient and not for the ρ-mixing.

6.1 Exponential of diffusion processes on the real line

6.1.1 Exponential of the Ornstein-Uhlenbeck process

Assume that $V_t = \exp(X_t)$ where $dX_t = -\alpha X_t dt + c dB_t$, B_t is a standard Brownian motion, X_0 is independent of (B_t, W_t). We consider $\alpha > 0$ and $X_0 \sim \mathcal{N}(0, \rho^2)$ with $\rho^2 = c^2/(2\alpha)$, which is the stationary distribution of (X_t).

Checking the Assumptions $(\mathbf{A}_1)$-$(\mathbf{A}_4)$. The process (X_t) satisfies $(\mathbf{A'}_1)$ so that (V_t) satisfies $(\mathbf{A}_1)$. For $(\mathbf{A}_2)$, the integrability property of the β-mixing coefficient can be deduced from Proposition 1 p.1063 in Pardoux and Veretennikov (2001). Consider $dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$ and the following conditions:

$(\mathbf{A'}_2)$ μ is a locally bounded borel function, $\exists M_0 \geq 0$ and $d \geq 0$ such that $\mathrm{sgn}(x)\mu(x) \leq -r|x|^d$ for $|x| \geq M_0$; σ is uniformly continuous and there exists $\sigma_0, \sigma_1 > 0$ such that $0 < \sigma_0 \leq \sigma(x) \leq \sigma_1 < +\infty$, for all $x \in \mathbb{R}$.

Then the stationary solution of the equation is β-mixing with β-mixing coefficient $\beta_X(t)$ such that $\beta_X(t) \leq ce^{-\lambda t}$. Note that we also have $\rho_X(t) \leq e^{-\lambda t}$. (Actually in this model, $\lambda = \alpha$.) Since $(\mathbf{A'}_2)$ is fulfilled by (X_t), $(\mathbf{A}_2)$ holds for (X_t) and (V_t).
Assumption $(\mathbf{A}_3)$ and the first part of $(\mathbf{A}_4)$ are fulfilled. For the second part of $(\mathbf{A}_4)$, Lemma 2.1 gives the result.

Rate of the adaptive deconvolution estimator of f. We have $f^*(x) = e^{-\rho^2 x^2/2}$, and

$$\|f - f_L\|^2 = \frac{1}{2\pi} \int_{|x| \geq \pi L} e^{-\rho^2 x^2} dx \leq \frac{1}{\pi} \int_{x \geq \pi L} e^{-\rho^2 \pi L x} dx = \frac{2}{\pi^2 L \rho^2} e^{-\pi^2 L^2 \rho^2}.$$

Then if we choose $\pi L = \sqrt{\ln(n)}/\rho$, the bias is of order $(n\sqrt{\ln(n)})^{-1}$ and the variance is of order $Le^{\pi^2 L}/n$ that is of order $\sqrt{\ln(n)}\exp(\pi^2\ln^{1/2}(n)/\rho - \ln(n))$. This term gives the rate in the infimum term of the bound (19). Note that, for any δ, $0<\delta<1$, this term is $o(n^{-\delta})$.

We take $K_n \geq n\ln^2(n)$. In that case, asymptotically for any δ, $0<\delta<1$,

$$\mathbb{E}(\|\hat{f}_{\hat{L}} - f\|^2) \leq o(n^{-\delta}) + \frac{4\pi\ln(n)(1+A_2)}{n\Delta_n} + C\Delta_n\ln(n)^3.$$

The rate is therefore determined by the choice of $\Delta = \Delta_n = 1/(\sqrt{n}\ln(n))$ and is asymptotically of order (much better than logarithmic):

$$\mathbb{E}(\|\hat{f}_{\hat{L}} - f\|^2) = O(\ln(n)^2/\sqrt{n}).$$

6.1.2 Exponential of an hyperbolic diffusion.

We consider here $V_t = \exp(X_t)$ with $dX_t = -\alpha X_t dt + c\sqrt{1+X_t^2}dB_t$.

Checking the Assumptions $(\mathbf{A}_1)$-$(\mathbf{A}_4)$. The stationary solution of this equation, X_t, satisfies $(\mathbf{A'}_1)$ for $\alpha + c^2/2 > 0$ so that V_t satisfies $(\mathbf{A}_1)$. Here the stationary density of X is given by $f(x) = A(1+x^2)^{-1-\alpha/c^2}$ for a constant A of normalization. We have $(\mathbf{A}_3)$ and the first part of $(\mathbf{A}_4)$ for $\alpha > c^2/2$. And we have $(\mathbf{A}_4)$ by Lemma 2.1. For the ρ-mixing condition, we use Proposition 2.8 of Genon-Catalot *et al.* (2000), namely for $dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$, it is required that

(i) $\lim_{x\to\pm\infty}\sigma(x)m(x) = 0$, for $m(x) = \exp(2\int^x \mu(u)/\sigma^2(u)du)/\sigma^2(x)$,

(ii) Let $\gamma(x) = \mu'(x) - 2\mu(x)/\sigma(x)$. The limits $\lim_{x\to\pm\infty} 1/\gamma(x)$ exist and are finite.

Here $m(x) = (1+x^2)^{-(1+\alpha/c^2)}$, and $\sigma(x)m(x) = (1+x^2)^{-(1/2+\alpha/c^2)}$ tend to zero when x tends to $+\infty$ and $-\infty$, as soon as $\alpha + c^2/2 > 0$, which holds for $\alpha > 0$. Moreover $\gamma(x) = (c^2+4\alpha)x/(c\sqrt{1+x^2})$ so that $\lim_{x\to-\infty} 1/\gamma(x) = -1/[c(c^2+4\alpha)]$ and $\lim_{x\to+\infty} 1/\gamma(x) = 1/[c(c^2+4\alpha)]$. It follows that (i) and (ii) are fulfilled, so that the process is ρ-mixing and therefore exponentially ρ-mixing. We can conclude that $(\mathbf{A}_2)$ is satisfied.

Rate of the adaptive deconvolution estimator of f. We consider only the cases where $\alpha/c^2 = m \in \mathbb{N}$ with $m \geq 1$. Let us denote by f_m the stationary density associated with the value $\alpha/c^2 = m$. For $m=0$, $f = f_0$ corresponds to the Cauchy density, with Fourier Transform $f_0^*(x) = e^{-|x|}$ and in the general case f_m^* satisfies $(f_m^*)" = f_m^* - f_{m-1}^*$ so that is can be proved (by recursion) that $f_m^*(t) = P_m(|t|)e^{-|t|}$ where P_m is a polynomial of degree m. The optimal asymptotic rate is obtained by the minimization of $L^{-2m}e^{-2\pi L} + Le^{\pi^2 L}/n$. Therefore choosing $\pi L = \ln(n)/(2+\pi) - (2m+1)/(2+\pi)\ln(\ln(n))$ leads to the rate $n^{-2/(2+\pi)}\ln(n)^{-[2/(2+\pi)](\pi m-1)}$ for the deconvolution terms.

Here, the discretization terms can be made negligible for instance when $\Delta = \Delta_n = 1/(\sqrt{n}\ln(n))$. Therefore the rate is entirely determined by the deconvolution rate and is of order

$$\mathbb{E}(\|\hat{f}_{\hat{L}} - f_m\|^2) = O\left(n^{-\frac{2}{2+\pi}} \ln(n)^{-\frac{2(\pi m-1)}{2+\pi}}\right). \tag{21}$$

6.1.3 A non ρ-mixing model

We consider here $V_t = \exp(X_t)$ with $X_t = -\alpha X_t/(1+X_t^2)dt + cdB_t$.

Checking the Assumptions $(\mathbf{A}_1)$-$(\mathbf{A}_4)$. The stationary solution of this equation, X_t, satisfies $(\mathbf{A'}_1)$ for $2\alpha/c^2 > 1$ so that V_t satisfies $(\mathbf{A}_1)$. The stationary density of X is given by $f(x) = A(1+x^2)^{-\alpha/c^2}$ for a constant A of normalization. We have $(\mathbf{A}_3)$ and the first part of $(\mathbf{A}_4)$ for $2\alpha/c^2 > 3$. And we have $(\mathbf{A}_4)$ by Lemma 2.1. Here, the process is not ρ-mixing $(\rho_X(t) \equiv 1)$, but $b(x) = -\alpha x/(1+x^2)$, $\sigma(x) = c$ satisfy $(\mathbf{A'}_2)$ so that the process is geometrically β-mixing.

Rate of the adaptive deconvolution estimator of f. As previously, we consider only the cases where $\alpha/c^2 = m+1 \in \mathbb{N}$ with $m \geq 1$ and $f_m(x) = A(1+x^2)^{-(m+1)}$, for which we know that the rate is given by (21).

6.2 The Cox-Ingersoll-Ross process

Another standard modelization of the volatility is the square-root process, often called the Cox-Ingersoll-Ross (CIR) process. Then, V_t is solution of the equation $dV_t = \alpha(\beta - V_t)dt + c\sqrt{V_t}dB_t$. This equation has been widely studied, in particular in Cox *et al.* (1985). The stationary density of V_t is $\pi(x) = [(\alpha/c^2)^{2\alpha\beta/c^2}/\Gamma(2\alpha\beta/c^2)]x^{2\alpha\beta/c^2-1}e^{-\alpha/c^2}\mathbb{I}_{x\geq 0}$, which is a Gamma density. It follows that

$$f(x) = e^x\pi(e^x) = [(\alpha/c^2)^{2\alpha\beta/c^2}/\Gamma(2\alpha\beta/c^2)]e^{2\alpha\beta/c^2 x}e^{-\alpha e^x/c^2}.$$

Checking the Assumptions $(\mathbf{A}_1)$-$(\mathbf{A}_4)$. Condition $(\mathbf{A'}_1)$, and thus Condition $(\mathbf{A}_1)$, is satisfied if $\alpha > 0$ and $2\alpha\beta/c^2 \geq 1$. It follows as a consequence of Proposition 2.8 in Genon-Catalot *et al.* (2000), see the examples p.1064, that the process is ρ-mixing and therefore exponentially ρ-mixing under the same condition $\alpha > 0$ and $2\alpha\beta/c^2 \geq 1$. If this holds, $(\mathbf{A}_2)$ is satisfied.
Moreover f is bounded so that $(\mathbf{A}_3)$ is fulfilled. For the first part of $(\mathbf{A}_4)$, it is clear that $X_0 = \ln^2(V_0)$ admits moments of any order. For the second part of $(\mathbf{A}_4)$, we use Lemma 2.2, where the only condition to check is $\mathbb{E}[\sup_{t\in[0,1]} 1/V_t^4] < +\infty$. It follows from Gloter (2000b)

that if $2\alpha\beta/c^2 > 1$, then,

$$\forall k \in [0, 2\alpha\beta/c^2 - 1[, \quad \mathbb{E}\left[\sup_{s\in[0,1]} V_s^{-k}\right] \le c\,\mathbb{E}\left(V_0^{-k}\right).$$

Therefore, for $k = 4$, this requires that $4 < 2\alpha\beta/c^2 - 1$ and $\mathbb{E}(V_0^{-4}) < +\infty$. The first condition gives $2\alpha\beta/c^2 > 5$ and the second $2\alpha\beta/c^2 > 4$. Therefore, $(\mathbf{A}_4)$ is satisfied if $2\alpha\beta/c^2 > 5$.

Rate of the adaptive deconvolution estimator of f. Let us consider then the CIR model with $\alpha > 0$, $2\alpha\beta/c^2 > 5$. Then $f^*(x) = \left(c^2/\alpha\right)^{ix}\Gamma(2\alpha\beta/c^2 + ix)/\Gamma(2\alpha\beta/c^2)$. By using the Stirling formula ($\Gamma(z) \sim \sqrt{2\pi}z^{z-1/2}e^{-z}$ for $|z| \to +\infty$), we find that

$$2\int_{x\ge\pi L} |f^*(x)|^2 dx \sim_{L\to+\infty} \frac{2\pi e^{-4\alpha\beta/c^2}}{\Gamma^2(2\alpha\beta/c^2)}(\pi L)^{2(2\alpha\beta/c^2-1/2)}e^{-\pi^2 L}.$$

Then choose $\pi^2 L = \ln(n)/2(2\alpha\beta/c^2 - 1)\ln\ln(n)$ to obtain for the infimum term the order $[\ln(n)]^{2\alpha\beta/c^2}/\sqrt{n}$. The same choice as previously for $\Delta = \Delta_n = 1/(\sqrt{n}\ln(n))$ gives for the squared risk the global rate (which is also better than logarithmic):

$$\mathbb{E}(\|\hat{f}_{\hat{L}} - f\|^2) = O([\ln(n)]^{2\alpha\beta/c^2}/\sqrt{n}).$$

6.3 The bilinear process

We consider here the case where V_t is the solution of $dV_t = \alpha(\beta - V_t)dt + cV_t dB_t$. The stationary density of V_t is given by $\pi(x) = Cx^{-(2\alpha)/c^2-2}\exp\left(-\alpha\beta/c^2 x\right)$, an inverse Gamma density.

Checking the Assumptions $(\mathbf{A}_1)$-$(\mathbf{A}_4)$. It is easy to see that Assumption $(\mathbf{A'}_1)$, and therefore $(\mathbf{A}_1)$, is satisfied if $\alpha\beta > 0$ and $\alpha + c^2/2 > 0$. For $(\mathbf{A}_2)$, we use the results in Genon-Catalot *et al.* (2000, p.1064), where it is proved that this process is ρ-mixing under the same condition ($\alpha\beta > 0$ and $\alpha + c^2/2 > 0$). Since $f(x) = C\exp\left(-(2\alpha/c^2 + 1)x - (\alpha\beta/c^2)e^{-x}\right)$, it is clear that f is bounded and thus $(\mathbf{A}_3)$ is satisfied.
The first part of $(\mathbf{A}_4)$ is satisfied since clearly $\int x^2 f(x)dx < +\infty$. For the second part of $(\mathbf{A}_4)$, we use Lemma 2.2. From the formula for π, is easy to see that $\mathbb{E}(V_0^4) < +\infty$ if $2\alpha/c^2 > 3$. Moreover, it follows from Gloter (2000b) that for $\beta > 0, \alpha > 0$, then

$$\forall k \ge 0, \mathbb{E}\left[\sup_{s\in[0,1]} V_s^{-k}\right] \le c\left[1 + \mathbb{E}\left(V_0^{-k}\right)\right]$$

and it is easy to see that $\mathbb{E}(V_0^{-k}) < +\infty$ for all k. Therefore, $(\mathbf{A}_4)$ is fulfilled as soon as $2\alpha/c^2 > 3$.

Rate of the adaptive deconvolution estimator of f. We consider now the bilinear process with $\alpha > 0$, $\beta > 0$ and $2\alpha/c^2 > 3$. We can compute $f^*(t) = C'\Gamma(2\alpha/c^2+1-it)$. We find that $|f^*(t)| \sim_{t\to+\infty} C"t^{2\alpha/c^2+1/2}e^{-\pi t/2}$ which ensures that $\|f - f_L\|^2$ is of order $e^{-\pi^2 L}L^{4\alpha/c^2+1}$. Choosing $L = \ln(n)/2 + (2\alpha/c^2)\ln(\ln(n))$ gives, for $\Delta = \Delta_n = 1/(\sqrt{n}\ln(n))$, a rate of order:

$$\mathbb{E}(\|\hat{f}_{\hat{L}} - f\|^2) = O(\ln(n)^{\frac{2\alpha}{c^2}+1}/\sqrt{n}).$$

7 Appendix: Proofs

Proof of Lemma 2.1. We have $\ln(\bar{V}_1)-\ln(V_0) = \ln\left(\Delta^{-1}\int_0^\Delta e^{X_s-X_0}ds\right) \le \sup_{s\in[0,\Delta]}(X_s-X_0)$ and $\ln(V_0)-\ln(\bar{V}_1) = -\ln\left(\Delta^{-1}\int_0^\Delta e^{X_s-X_0}\right)ds \le -\inf_{s\in[0,\Delta]}(X_s-X_0)$. Therefore, $\mathbb{E}(\ln\bar{V}_1-\ln V_0)^2 \le \mathbb{E}\left[\sup_{s\in[0,\Delta]}(X_s-X_0)^2\right]$. Applying the Burkholder-Davis-Gundy inequality, we get:

$$\mathbb{E}\left[\sup_{s\in[0,\Delta]}(X_s - X_0)^2\right] \le c\left\{\mathbb{E}\left(\int_0^\Delta \sigma^2(X_u)du\right) + \mathbb{E}\left[\left(\int_0^\Delta \mu(X_u)du\right)^2\right]\right\}.$$

Using the linear growth condition on μ and σ, we obtain $\mathbb{E}\left[\sup_{s\in[0,\Delta]}(X_s - X_0)^2\right] \le c[1 + \mathbb{E}(X_0^2)]\Delta$. □

Proof of Lemma 2.2. Our assumptions imply $\mathbb{E}[V_0^4 + V_0^{-4}] < +\infty$. Noting that $|\ln(V_0)| \le V_0 + V_0^{-1}$, we deduce that $\mathbb{E}[\ln^2(V_0)] < +\infty$. Now, let $\Delta \le 1$. By the Taylor formula,

$$\ln(\bar{V}_1) - \ln(V_0) = (\bar{V}_1 - V_0)\int_0^1 \frac{du}{V_0 + u(\bar{V}_1 - V_0)}.$$

Since, for $u \in [0,1]$, $V_0 + u(\bar{V}_1 - V_0) = V_0(1-u) + u\bar{V}_1 \ge \inf_{s\in[0,\Delta]} V_s$. We get:

$$\mathbb{E}\left[(\ln\bar{V}_1 - \ln V_0)^2\right] \le \left[\mathbb{E}(\bar{V}_1 - V_0)^4\mathbb{E}\left(\sup_{s\in[0,1]}\frac{1}{V_s^4}\right)\right]^{1/2}. \quad (22)$$

Using the Hölder inequality, we obtain the bound

$$\mathbb{E}(\bar{V}_1 - V_0)^4 \le \frac{1}{\Delta}\int_0^\Delta \mathbb{E}(V_s - V_0)^4 ds. \quad (23)$$

Now, we note that: $(V_s - V_0)^4 \le 4\left[\left(\int_0^s b(V_u)du\right)^4 + \left(\int_0^s a(V_u)dW_u\right)^4\right]$. For the ordinary integral term, we use the Hölder inequality and **(A'$_1$)** to obtain: for all $s \le \Delta$,

$$\mathbb{E}\left(\int_0^s b(V_u)du\right)^4 \le C\Delta^3\mathbb{E}\left(\int_0^s (1+V_u^4)du\right) \le C\Delta^4(1+\mathbb{E}(V_0^4)).$$

For the stochastic integral, we use the Burkholder-Davis-Gundy inequality and **(A'$_1$)** to obtain:

$$\mathbb{E}\left(\int_0^s a(V_u)dW_u\right)^4 \leq C\mathbb{E}\left(\int_0^s a^2(V_u)du\right)^2 \leq C'\Delta^2(1+\mathbb{E}V_0^4).$$

Finally, for $s \leq \Delta$, $\mathbb{E}(V_s - V_0)^4 \leq C"\Delta^2$. Joining this, (22) and (23), we get $\mathbb{E}(\ln \bar{V}_1 - \ln V_0)^2 \leq C\Delta$. □

Proof of Proposition 4.1. The bound for $\|f - f_L\|^2$ is straightforward from the definition of the class $\mathcal{S}_{s,r,b}(M)$ and the inequality $\|f - f_L\|^2 \leq \frac{1}{2\pi}\int_{|x|\geq \pi L} |f^*(x)|^2 dx$. The bound for $\|f_L - f_L^{(n)}\|^2$ can be found in Comte et al. (2005a) and is recalled here for the sake of completeness. First note that $\|f_L - f_L^{(n)}\|^2 = \sum_{|j|\geq K_n} a_{L,j}^2 \leq (\sup_j j a_{L,j})^2 \sum_{|j|\geq K_n} j^{-2}$. Now we write that

$$\begin{aligned} j a_{L,j} &= j\sqrt{L}\int \varphi(Lx - j) f(x)dx \\ &\leq L^{3/2}\int |x||\varphi(Lx-j)|f(x)dx + \sqrt{L}\int |Lx - j||\varphi(Lx-j)|f(x)dx \\ &\leq L^{3/2}\left(\int |\varphi(Lx-j)|^2 dx\right)^{1/2}\left(\int x^2 f^2(x)dx\right)^{1/2} + \sqrt{L}\sup_x |x\varphi(x)|. \end{aligned}$$

This implies finally that $j a_{L,j} \leq L(C_f)^{1/2} + \sqrt{L}$, and (15) follows. □

Proof of Lemma 4.1.

$$\begin{aligned} 2\mathbb{E}(\langle \hat{f}_L - f_L^{(n)}, g_\Delta - f\rangle) &\leq 2\mathbb{E}\|\hat{f}_L - f_L^{(n)}\| \sup_{t\in S_L^{(n)}, \|t\|=1} |\langle t, g_\Delta - f\rangle| \\ &\leq \frac{1}{8}\mathbb{E}(\|\hat{f}_L - f_L^{(n)}\|^2) + 8\sup_{t\in S_L^{(n)}, \|t\|=1} \langle t, g_\Delta - f\rangle^2 \\ &\leq \frac{1}{4}\mathbb{E}(\|\hat{f}_L - f\|^2) + \frac{1}{4}\|f - f_L^{(n)}\|^2 + 8\sum_{|j|\leq K_n}\langle \varphi_{L,j}, g_\Delta - f\rangle^2. \end{aligned}$$

Note that $\varphi_{L,j}^*(x) = e^{ixj/L}\varphi^*(x/L)/\sqrt{L}$. Then

$$\sum_{|j|\leq K_n}\langle \varphi_{L,j}, g_\Delta - f\rangle^2 \leq \frac{1}{(2\pi)^2}\sum_{j\in\mathbb{Z}}\langle \varphi_{L,j}^*, g_\Delta^* - f^*\rangle^2$$

and since $\langle \varphi_{L,j}^*, g_\Delta^* - f^*\rangle = \int e^{-ixj}\varphi^*(x)(g_\Delta^* - f^*)(Lx)\sqrt{L}dx$, Parseval's formula yields

$$\sum_{|j|\leq K_n}\langle \varphi_{L,j}, g_\Delta - f\rangle^2 \leq \frac{L}{2\pi}\int_{-\pi}^{\pi}|g_\Delta^* - f^*|^2(Lx)dx = \frac{1}{2\pi}\int_{-\pi L}^{\pi L}|g_\Delta^* - f^*|^2(x)dx.$$

Since $|g_\Delta^* - f^*(x)| = |\mathbb{E}\left(e^{ix\ln(\bar{V}_1)} - e^{ix\ln(V_0)}\right)|$ and $|e^{iux} - e^{ivx}| \leq |x(u-v)|$,

$$\sum_{|j|\leq K_n}\langle \varphi_{L,j}, g_\Delta - f\rangle^2 \leq \mathbb{E}\left|\ln\left(\bar{V}_1/V_0\right)\right|^2\frac{\pi^2 L^3}{3}.$$

Then under Assumption $(\mathbf{A}_4)$ if follows that $\sum_{|j|\le K_n}\langle \varphi_{L,j}, g_\Delta - f\rangle^2 \le c\Delta\pi^2 L^3/3$. This yields to (16). $\square$

Proof of Lemma 4.2. We first give here a lemma which is used in the sequel.

Lemma 7.1 *The following inequalities hold.*
In the β-mixing case, there exists a function b_Δ such that

$$\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n h(\bar X_k)\right) \le \frac{C}{n}\int h^2(v)(1+b_\Delta(v))g_\Delta(v)dv,$$

and such that $\mathbb{E}(b_\Delta(\bar X_1)) = \int b_\Delta(v)g_\Delta(v)dv \le (1+\int_0^{+\infty}\beta_V(u)du)/\Delta$.
In the ρ-mixing case,

$$\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n h(\bar X_k)\right) \le \frac{C}{n}\mathbb{E}(h^2(\bar X_1))\frac{1+\int_0^{+\infty}\rho_V(u)du}{\Delta}.$$

Note that if h is bounded, both bounds becomes $(C/n)\|h\|_\infty^2(1+A_2)/\Delta$, where $A_2 = \int_0^{+\infty}\beta_V(u)du$ or $A_2 = \int_0^{+\infty}\rho_V(u)du$.

Proof of Lemma 7.1. In all cases, we write

$$\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n h(\bar X_k)\right) = \frac{\mathrm{Var}(h(\bar X_1))}{n} + \frac{2}{n^2}\sum_{k=1}^n (n-k)\mathrm{cov}(h(\bar X_1), h(\bar X_{k+1})).$$

The ρ-mixing inequality follows then from

$$\mathrm{cov}(h(\bar X_1), h(\bar X_{k+1})) \le \rho_V((k-1)\Delta)\mathbb{E}(h^2(\bar X_1)), \tag{24}$$

which implies that

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n h(\bar X_k)\right) &\le \frac{\mathbb{E}(h^2(\bar X_1))}{n} + \frac{2\mathbb{E}(h^2(\bar X_1))}{n}\sum_{k=1}^n \rho_V((k-1)\Delta)\\ &\le C\mathbb{E}(h^2(\bar X_1))\frac{1+\int_0^{+\infty}\rho_V(u)du}{n\Delta}.\end{aligned}$$

The first inequality is a straightforward adaptation of Viennet's (1997) inequality when using that the process $(\bar X_k)$ is stationary with density g_Δ and β-mixing coefficients bounded by $\beta_V((k-1)\Delta)$. This yields

$$\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n h(\bar X_k)\right) \le \frac{\mathbb{E}(h^2(\bar X_1))}{n} + \frac{4}{n}\int h^2(u)b_\Delta(u)g_\Delta(u)du,$$

with $\int b_\Delta(u)g_\Delta(u)du \le \sum_{k=0}^n \beta_{k\Delta} \le C(1+\int \beta_s ds)/\Delta$. The β-mixing inequality follows. $\square$

Now, the term of interest can be written

$$\begin{aligned} 2\nu_n(\hat f_L - f_L^{(n)}) &\le 2\|\hat f_L - f_L^{(n)}\| \sup_{t\in S_L^{(n)}, \|t\|=1} |\nu_n(t)| \le \frac{1}{4}\|\hat f_L - f_L^{(n)}\|^2 + 4 \sup_{t\in S_L^{(n)}, \|t\|=1} \nu_n^2(t) \\ &\le \frac{1}{2}\|\hat f_L - f\|^2 + \frac{1}{2}\|f - f_L^{(n)}\|^2 + 4\sum_{j\in\mathbb{Z}} \nu_n^2(\varphi_{L,j}). \end{aligned} \tag{25}$$

Then write that

$$\sum_{j\in\mathbb{Z}} \nu_n^2(\varphi_{L,j}) = \sum_{j\in\mathbb{Z}} \frac{1}{4\pi^2}\left(\int \frac{\varphi_{L,j}^*(x)}{f_\eta^*(x)}\Psi(x)dx\right)^2, \quad \text{where} \quad \Psi(x) = \frac{1}{n}\sum_{k=1}^n \left(e^{ixZ_k} - \mathbb{E}(e^{ixZ_k})\right).$$

As previously, Parseval's formula yields

$$\sum_{j\in\mathbb{Z}} \nu_n^2(\varphi_{L,j}) = \frac{1}{2\pi}\int_{-\pi L}^{\pi L} \frac{|\Psi(x)|^2}{|f_\eta^*(x)|^2}dx. \tag{26}$$

It remains to bound $\mathbb{E}(|\Psi(x)|^2)$ which is a variance.

$$\mathbb{E}(|\Psi(x)|^2) = \frac{1}{n^2}\left(\sum_{k=1}^n \mathrm{Var}(e^{ixZ_k}) + 2\sum_{1\le k<l\le n} \mathrm{cov}(e^{ixZ_k}, e^{ixZ_l})\right).$$

Then $\mathrm{Var}(e^{ixZ_k}) = 1 - |\mathbb{E}(e^{ixZ_k})|^2 = 1 - |f_\eta^*(x)g_\Delta^*(x)|^2 \le 1$ and for $k<l$,

$$\mathrm{cov}(e^{ixZ_k}, e^{ixZ_l}) = \mathbb{E}(e^{ix(Z_k-Z_l)}) - |f_\eta^*(x)g_\Delta^*(x)|^2 = \mathbb{E}(\mathbb{E}(e^{ix(Z_k-Z_l)}|\mathcal{G})) - |f_\eta^*(x)g_\Delta^*(x)|^2,$$

where $\mathcal{G} = \sigma(V_t, t\ge 0)$. Then, conditionally on $\mathcal{G}$, Z_k and Z_l are independent for $k\ne l$ and

$$\mathbb{E}(\mathbb{E}(e^{ix(Z_k-Z_l)}|\mathcal{G})) = \mathbb{E}(\mathbb{E}(e^{ixZ_k}|\sigma)\mathbb{E}(e^{-ixZ_l}|\mathcal{G})) = |f_\eta^*(x)|^2\mathbb{E}(e^{ix\ln(\bar V_k)}e^{-ix\ln(\bar V_l)})$$

It follows that

$$\mathbb{E}(|\Psi(x)|^2) \le \frac{1}{n} + \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n e^{ix\ln(\bar V_k)}\right)|f_\eta^*(x)|^2. \tag{27}$$

Then, since $h(x) = e^{iux}$ is bounded, Lemma 7.1 implies, in both mixing contexts,

$$\mathbb{E}\left(\sum_{j\in\mathbb{Z}} \nu_n^2(\varphi_{L,j})\right) \le \frac{\Phi_\eta(L)}{2\pi n} + C\frac{(1+A_2)L}{n\Delta}. \quad \square$$

Proof of Theorem 5.1. The proof of the result follows the same line as previously. By using that for any L, $\gamma_n(\hat f_{\hat L}) + \mathrm{pen}(\hat L) \le \gamma_n(f_L^{(n)}) + \mathrm{pen}(L)$ we have:

$$\begin{aligned} \|\hat f_{\hat L} - f\|^2 &\le \|f - f_L^{(n)}\|^2 + 2\nu_n(\hat f_{\hat L} - f_L^{(n)}) + 2\langle \hat f_{\hat L} - f_L^{(n)}, g_\Delta - f\rangle + \mathrm{pen}(L) - \mathrm{pen}(\hat L) \\ &\le \frac{7}{4}\|f - f_L^{(n)}\|^2 + \frac{3}{4}\|\hat f_{\hat L} - f\|^2 + \mathrm{pen}(L) + 8\sup_{t\in S_{\ln(n)}, \|t\|=1} [\nu_n^{(1)}(t)]^2 \\ &\quad + 8\sup_{t\in S_{L\vee\hat L}, \|t\|=1} [\nu_n^{(2)}(t)]^2 - \mathrm{pen}(\hat L) + 8\sup_{t\in S_{\hat L\vee L}^{(n)}, \|t\|=1} \langle t, g_\Delta - f\rangle^2, \end{aligned}$$

where

$$\nu_n^{(1)}(t) = \frac{1}{2\pi n}\sum_{k=1}^{n}\left(\int e^{ix\bar{X}_k}t^*(-x)dx - \langle g_\Delta^*, t^*\rangle\right), \;\; \nu_n^{(2)}(t) = \frac{1}{2\pi n}\sum_{k=1}^{n}\left(u_t(Z_k) - \int e^{ix\bar{X}_k}t^*(-x)dx\right).$$

It follows straightforwardly from the study of the approximation bias term that

$$8 \sup_{t\in S^{(n)}_{\hat{L}\vee L}, \|t\|=1} \langle t, g_\Delta - f\rangle^2 \le C\Delta(L\vee\hat{L})^3 \le C\Delta\ln^3(n).$$

Moreover, by using Lemma 7.1 in the β-mixing context, we find

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in S_{\ln(n)}, \|t\|=1}[\nu_n^{(1)}(t)]^2\right) &\le \sum_{|j|\le K_n}\mathbb{E}[\nu_n^{(1)}(\varphi_{\ln(n),j})]^2 = \sum_{|j|\le K_n}\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n}\varphi_{\ln(n),j}(\bar{X}_k)\right)\\ &\le \sum_{|j|\le K_n}\frac{C}{n}\int|\varphi_{\ln(n),j}(x)|^2(1+b_\Delta(x))g_\Delta(x)dx\end{aligned}$$

and since $\|\sum_{j\in\mathbb{Z}}|\varphi_{L,j}|^2\|_\infty \le L$, we obtain

$$\mathbb{E}\left(\sup_{t\in S_{\ln(n)}, \|t\|=1}[\nu_n^{(1)}(t)]^2\right) \le \frac{C\ln(n)(1+A_2)}{n\Delta}.$$

In the ρ-mixing case, we have, still by using (24) and $\|\sum_{j\in\mathbb{Z}}|\varphi_{L,j}|^2(.)\|_\infty \le L$ and Lemma 7.1:

$$\sum_{|j|\le K_n}\mathrm{Var}\left[\frac{1}{n}\sum_{k=1}^{n}\varphi_{\ln(n),j}(\bar{X}_k)\right] \le \frac{C\ln(n)(1+A_2)}{n}.$$

We introduce here the centering quantity $p(L, L')$ which will induce the choice of the penalty:

$$\begin{aligned}\mathbb{E}(\|\hat{f}_{\hat{L}} - f\|^2) &\le 7\|f - f_L^{(n)}\|^2 + 4\mathrm{pen}(L) + C_1\Delta\ln^3(n) + \frac{C_2\ln(n)(1+A_2)}{n\Delta}\\ &\quad + \mathbb{E}\left[16\left(\sup_{t\in S_{L\vee\hat{L}}, \|t\|=1}[\nu_n^{(2)}(t)]^2 - p(L,\hat{L})\right)_+ + 16p(L,\hat{L}) - \mathrm{pen}(\hat{L})\right]\end{aligned}$$

Therefore, we fix $\mathrm{pen}(L)$ by assuming that, for all L, L', $16p(L, L') \le \mathrm{pen}(L) + \mathrm{pen}(L')$. It follows:

$$\begin{aligned}\mathbb{E}(\|\hat{f}_{\hat{L}} - f\|^2) &\le 7\|f - f_L^{(n)}\|^2 + 5\mathrm{pen}(L) + C_1\Delta\ln^3(n) + \frac{C_2\ln(n)(1+A_2)}{n\Delta}\\ &\quad + \mathbb{E}\left[16\sum_{L'=1}^{\ln(n)}(\sup_{t\in S_{L\vee L'}, \|t\|=1}[\nu_n^{(2)}(t)]^2 - p(L,L'))_+\right].\end{aligned}$$

To bound the last term, we apply the following inequality, to the variables Z_k conditionally on $\mathcal{G}$, which gives independent but non identically distributed random variables.

Lemma 7.2 *Let $U_1, \dots, U_n$ be independent random variables and $\nu_n(r) = (1/n)\sum_{i=1}^n [r(U_i) - \mathbb{E}(r(U_i))]$ for r belonging to a countable class $\mathcal{R}$ of uniformly bounded measurable functions. Then for $\epsilon > 0$*

$$\mathbb{E}\left[\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 - 2(1+2\epsilon)H^2\right]_+ \leq \frac{6}{K_1}\left(\frac{v}{n} e^{-K_1\epsilon\frac{nH^2}{v}} + \frac{8M_1^2}{K_1 n^2 C^2(\epsilon)} e^{-\frac{K_1 C(\epsilon)\sqrt{\epsilon}}{\sqrt{2}}\frac{nH}{M_1}}\right), \tag{28}$$

with $C(\epsilon) = \sqrt{1+\epsilon} - 1$, K_1 is a universal constant, and where

$$\sup_{r\in\mathcal{R}} \|r\|_\infty \leq M_1, \quad \mathbb{E}\left(\sup_{r\in\mathcal{R}} |\nu_n(r)|\right) \leq H, \quad \sup_{r\in\mathcal{R}} \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(r(U_i)) \leq v.$$

The inequality (28) is a straightforward consequence of Talagrand's (1996) inequality given in Ledoux (1996) (or Birgé and Massart (1997)), with f replaced by $r = f - \mathbb{E}f(X_1)$ and M_1 by $2M_1$, and by taking $\eta = (\sqrt{1+\epsilon} - 1) \wedge 1 = C(\epsilon) \leq 1$. Moreover, standard density arguments allow to apply it to the unit ball of a finite dimensional linear space. Therefore, denoting by $\mathbb{E}_V$ the conditional expectation given $\mathcal{G}$, and by $B_{L,L'}(0,1) = \{t \in S_{L\vee L'},, \|t\| = 1\}$, we have

$$\mathbb{E}_V\left[\sup_{t\in B_{L,L'}(0,1)} |\nu_n^{(2)}(t)|^2 - 2(1+2\epsilon_1)\mathbb{H}^2\right]_+ \leq \kappa_1\left(\frac{v}{n}e^{-K_1\epsilon_1\frac{n\mathbb{H}^2}{v}} + \frac{M_1^2}{n^2}e^{-K_2\sqrt{\epsilon_1}C(\epsilon_1)\frac{n\mathbb{H}}{M_1}}\right), \tag{29}$$

where $K_2 = K_1/\sqrt{2}$ and $\mathbb{H}$, v and M_1 are defined by $\mathbb{E}_V(\sup_{t\in B_{L,L'}(0,1)} |\nu_n^{(2)}(t)|^2) \leq \mathbb{H}^2$,

$$\sup_{t\in B_{L,L'}(0,1)} \mathrm{Var}_V(u_t(Z_1)) \leq v, \text{ and } \sup_{t\in B_{L,L'}(0,1)} \|u_t(Z_1)\|_\infty \leq M_1.$$

Let $L^* = L \vee L'$.

$$\begin{aligned}
&\mathbb{E}_V\left(\sup_{t\in B_{L,L'}(0,1)} |\nu_n^{(2)}(t)|^2\right) \leq \sum_{|j|\leq K_n} \mathrm{Var}_V\left(\frac{1}{n}\sum_{k=1}^n u_{\varphi_{L^*,j}}(Z_k)\right) \\
\leq\ & \frac{1}{(2\pi n)^2}\sum_{k=1}^n\sum_{j\in\mathbb{Z}} \mathbb{E}_V\left(\int e^{ixZ_k}\frac{\varphi_{L^*,j}(-x)}{f_\eta^*(-x)}dx\right)^2 \leq \frac{1}{2\pi n^2}\sum_{k=1}^n \Phi_\eta(L^*) = \frac{\Phi_\eta(L^*)}{2\pi n} := \mathbb{H}^2
\end{aligned}$$

$$\begin{aligned}
\sup_{t\in B_{L,L'}(0,1)} \sup_x |u_t(x)| &\leq \frac{1}{2\pi}\sup_{t\in B_{L,L'}(0,1)} \int \left|\frac{t^*(-x)}{f_\eta^*(x)}\right| dx \leq \frac{1}{2\pi}\sup_{t\in B_{L,L'}(0,1)} \|t^*\|\sqrt{\Phi_\eta(L^*)} \\
&\leq \sqrt{\frac{1}{2\pi}\Phi_\eta(L^*)} := M_1
\end{aligned}$$

Note that $M_1 = \sqrt{n\mathbb{H}^2}$. Lastly, for v, the crude bound $v = n\mathbb{H}^2$ holds straightforwardly. It follows that,

$$\mathbb{E}_V\left[\sup_{t\in B_{L,L'}(0,1)} |\nu_n^{(2)}(t)|^2 - 2(1+2\xi)\mathbb{H}^2\right]_+ \leq \kappa_1\left(\frac{\Phi(L^*)}{n}e^{-K_1\xi} + \frac{\Phi(L^*)}{n^2}e^{-K_2\sqrt{\xi}C(\xi)\sqrt{n}}\right),$$

Then by choosing $\xi = 2\pi L^*/K_1$ and by considering a collection such that $\Phi(L)/n$ is bounded (say by K), we find

$$\mathbb{E}_V \left[\sup_{t \in B_{L,L'}(0,1)} |\nu_n^{(2)}(t)|^2 - 2(1 + 4\pi L^*/K_1) \frac{\Phi(L^*)}{n} \right]_+ \leq \kappa_1 \left(\frac{e^{-\pi L^*}}{n} + \frac{1}{n} e^{-K_2 \sqrt{n}} \right),$$

and

$$\sum_{L'=1}^{\ln(n)} \mathbb{E}_V \left[\sup_{t \in B_{L,L'}(0,1)} |\nu_n^{(2)}(t)|^2 - 2(1 + 4\pi L^*/K_1) \frac{\Phi(L^*)}{n} \right]_+ \leq \frac{C_3}{n}$$

since $\sum_{L'} e^{-\pi L^*} = \sum_{L' \leq L} e^{-\pi L} + \sum_{L' > L} \frac{e^{-\pi L'}}{n} \leq L e^{-\pi L} + 1/(e-1) \leq K$. Since the bound does not depend on V, we can take the expectation with respect to V without changing it. By gathering all terms, we obtain the bound given in the Theorem. □

References

Bibby, M.B., Jacobsen, M. and Sørensen, M. (2004). Estimating Functions for Discretely Sampled Diffusion-Type Models. Chapter of the *Handbook of financial econometrics*, Ait-Sahalia and Hansen eds. http://home.uchicago.edu/ lhansen/handbook.htm

Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation, in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics* (D. Pollard, E. Torgersen and G. Yang, eds), 55-87, Springer-Verlag, New-York.

Butucea, C. (2004) Deconvolution of supersmooth densities with smooth noise. *Canadian J. of Statistics* **32**, 181-192.

Butucea, C. and Tsybakov, A.B. (2004). Fast asymptotics in density deconvolution. Working paper. Preprint LPMA-898, http://www.proba.jussieu.fr/mathdoc/preprints/index.html#2004.

Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184-1186.

Comte, F. (2004). Kernel Deconvolution of Stochastic Volatility Models. *J. Time Ser. Anal.* **25**, 563-582.

Comte, F., Rozenholc, Y. and Taupin, M.-L.(2005a). Penalized contrast estimator for density deconvolution with mixing variables. http://www.math-info.univ-paris5.fr/ comte/publi.html.

Comte, F., Rozenholc, Y. and Taupin, M.-L. (2005b) Finite sample penalization in density deconvolution. Preprint.

Cox, J. C., Ingersoll, J. E., Jr. and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385-407.

van Es, B., Spreij, P. and van Zanten, H. (2003). Nonparametric volatility density estimation.

Bernoulli **9**, 451-465.
Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.* **19**, 1257-1272.
Genon-Catalot, V., Jeantheau, T. and Laredo, C. (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **5**, 855-872.
Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, **4**, 1051-1079.
Gloter, A. (2000a). Estimation des paramètres d'une diffusion cachée. PhD Thesis of the University of Marne-La-Vallée.
Gloter, A. (2000b). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.* **4**, 205-227.
Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 3, 281-300.
Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Statist.* **5**, 225-242.
Ibragimov, I. A. and Hasminskii, R. Z. (1983). Estimation of distribution density. *J. Soviet. Math.* **21** 40-57.
Ledoux, M. (1996). On Talagrand's deviation inequalities for product measure. *ESAIM Probab. Statist.*, **1**, 63-87.
Pardoux, E. and Veretennikov, A. Yu. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29**, 3, 1061-1085.
Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27**, 6, 2033-2053.
Renault, E.(1997). Econometric models of option pricing errors in advances econometrics. Seventh World Congress, Vol.3. Eds. D.M. Kreps and K.F. Wallis), Cambridge University Press.
Shephard, N. (2005). *Stochastic volatility selected readings.* Series: Advanced Texts in Econometrics. Neil Shephard ed. Oxford University Press.
Stefansky, L. and Carroll, R.J. (1990). Deconvolution kernel density estimators. *Statistics* **21**, 169-184.
Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126**, 505-563.
Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. *Prob. Th. Rel. Fields* **107**, 467-492.
Yoshida, N. (1997). Malliavin calculus and asymptotic expansions for martingales. *Prob. Th. Rel. Fields* **109**, 301-342.