Noname manuscript No.
(will be inserted by the editor)

Nonparametric estimation for a stochastic volatility model.

F. Comte · V. Genon-Catalot · Y. Rozenholc

Received: 12-2007. Revised version: 06-2007

Corresponding author: F. Comte, MAP5 UMR 8145,
Université Paris Descartes,
45 rue des Saints-Pères,
75270 Paris cedex 06, FRANCE.
email: fabienne.comte@univ-paris5.fr

Abstract Consider discrete time observations $(X_{\ell\delta})_{1\leq\ell\leq n+1}$ of the process X satisfying $dX_t = \sqrt{V_t}dB_t$, with V_t a one-dimensional positive diffusion process independent of the Brownian motion B. For both the drift and the diffusion coefficient of the unobserved diffusion V, we propose nonparametric least square estimators, and provide bounds for their risk. Estimators are chosen among a collection of functions belonging to a finite dimensional space whose dimension is selected by a data driven procedure. Implementation on simulated data illustrates how the method works. June 23, 2008

Keywords Diffusion coefficient · Drift · Mean square estimator · Model selection · Nonparametric estimation · Penalized contrast · Stochastic volatility

Mathematics Subject Classification (2000) 62G08 · 62M05 · 62P05

1 Introduction

In this paper, we consider the stochastic volatility model $(X_t, V_t)_{t\geq 0}$ with dynamics described by the following equations:

$$\begin{cases} dX_t = \sqrt{V_t}dB_t, \quad X_0 = 0, \\ dV_t = b(V_t)dt + \sigma(V_t)dW_t, \quad V_0 = \eta, \quad V_t > 0, \ \text{for all } t \geq 0, \end{cases} \tag{1}$$

where $(B_t, W_t)_{t\geq 0}$ is a standard bidimensional Brownian motion and η is independent of $(B_t, W_t)_{t\geq 0}$. Discrete time observations of the X process are available whereas the stochastic volatility V is unobservable. Our aim is to propose and study nonparametric

F. Comte E-mail: fabienne.comte@univ-paris5.fr · V. Genon-Catalot E-mail: genon@math-info.univ-paris5.fr · Y. Rozenholc E-mail: yves.rozenholc@math-info.univ-paris5.fr
MAP5 UMR 8145, University Paris Descartes.

estimators of the drift function $b(.)$ and the diffusion coefficient $\sigma^2(.)$ of the unobserved volatility process V.

Statistical inference for stochastic volatility models is often parametric: the functions $b(.)$ and $\sigma^2(.)$ are specified up to a few unknown parameters, see the popular examples of Heston (1993) or Cox, Ingersoll and Ross (1985). General statistical parametric approaches of the problem are studied in Genon-Catalot *et al.* (1999), Hoffmann (2002), Gloter (2007), Aït-Sahalia and Kimmel (2007). A recent proposal for nonparametric estimation of the drift and diffusion coefficients of V can be found in Renó (2006), who studies the empirical performance of a Nadaraya-Watson kernel strategy on two parametric simulated examples. Our approach is new and different. It is based on a nonparametric mean square strategy. We follow the ideas developed in Comte *et al.* (2007, 2008), where discrete observations of the process (V_t) or discrete observations of the integrated process $(\int_0^t V_s ds)$ are considered. We assume that (V_t) is stationary and we consider discrete time observations $(X_{\ell\delta})_{1\leq\ell\leq n+1}$ of the process (X_t) in the high frequency context: δ is small, n is large and $n\delta = T$, the time interval where observations are taken, is large. Such high frequency data of the process X cannot be used directly to estimate nonparametrically the drift and diffusion coefficients of the volatility process. In order to "eliminate" the effect of the Brownian motion B, one must first compute quadratic variations based on the discrete observations of X and then use these to estimate b and σ^2. This is done as follows. Given $n = kN$ observations of X with sampling interval δ, groups of k observations are used to compute quadratic variations. As it is usual, we define, for $i = 0, 1, \ldots, N-1$, the realized quadratic variation associated with $(X_{\ell\delta})_{ik+1\leq\ell<(i+1)k}$:

$$\hat{\bar{V}}_i = \frac{1}{k\delta}\sum_{j=0}^{k-1}\left(X_{(ik+j+1)\delta} - X_{(ik+j)\delta}\right)^2.$$

Setting $\varDelta = k\delta$, $\hat{\bar{V}}_i$ provides an approximation of the integrated volatility:

$$\bar{V}_i = \frac{1}{\varDelta}\int_{i\varDelta}^{(i+1)\varDelta} V_s ds, \tag{2}$$

which in turn may be, for well chosen k, δ, a satisfactory approximation of $V_{i\varDelta}$. One may think that N is a number of "days" and k is the number of observations "per day". Then the basic idea is to regress changes (square changes) of the quadratic variation of X from period $(i+1)$ to period $(i+2)$ on the quadratic variation of period $(i-1)$ to period (i), to get an estimate of the drift (diffusion) function. More precisely, we obtain regression-type equations, for $\ell = 1, 2$:

$$Y_{i+1}^{(\ell)} = f^{(\ell)}(\hat{\bar{V}}_i) + \text{noise} + \text{remainder},$$

where

$$f^{(1)} = b,\ Y_i^{(1)} = \frac{\hat{\bar{V}}_{i+1} - \hat{\bar{V}}_i}{\varDelta} \text{ and } f^{(2)} = \sigma^2,\ Y_i^{(2)} = \frac{3}{2}\frac{(\hat{\bar{V}}_{i+1} - \hat{\bar{V}}_i)^2}{\varDelta}. \tag{3}$$

Choosing a collection of finite dimensional spaces, we use the regression-type equations to construct estimators of the functions $b(.)$ and $\sigma^2(.)$ on these spaces. Then, we propose a data driven procedure to select a relevant estimation space in the collection. As it is usual with these methods, the risk of an estimator $\tilde{f}$ of $f = b$ or σ^2 is measured via $\mathbb{E}(\|f-\tilde{f}\|_N^2)$ where $\|f-\tilde{f}\|_N^2 = (1/N)\sum_{i=0}^{N-1}(f-\tilde{f})^2(\hat{\bar{V}}_i)$. We obtain risk bounds which can be interpreted as n, N tend to infinity, δ, Δ tend to 0 and $T = n\delta = N\Delta$ tends to infinity. These bounds are compared with Hoffmann's (1999) minimax rates in the case of direct observations of V. For what concerns b, our method leads to the best rate that can be expected. For what concerns σ^2, no benchmark is available in this asymptotic framework. Indeed, Gloter (2000) and Hoffmann (2002) only treat the case of observations within a fixed length time interval, in a parametric setting.

The paper is organized as follows. Section 2 describes the regression equations, the collection of estimation spaces, and the estimators, defined as minimizers of mean square contrast functions. In Section 3, the assumptions on the model are explained and the risks of the estimators are studied. Section 4 completes the procedure by the data driven selection of the estimation space. Examples of models and simulation results are presented in Section 5. Lastly, proofs are gathered in Section 6.

2 The mean square approach

2.1 The regression equations

Our estimation strategy is based on the idea that, if one observes (Y_i, X_i) with $Y_i = f(X_i) + \varepsilon_i$ where ε_i is a white noise, then nonparametric mean square contrasts lead to a good estimation of the regression function f. Let us explain how to use this idea for the case of the drift estimation.
Suppose we observe directly the $(V_{i\Delta})$, then, we can write:

$$\begin{aligned}\frac{V_{(i+1)\Delta} - V_{i\Delta}}{\Delta} &= \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} dV_s = \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} b(V_s)ds + \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta}\sigma(V_s)dW_s \\ &= b(V_{i\Delta}) + \underbrace{\frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta}\sigma(V_s)dW_s}_{\text{noise}} + \underbrace{\frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta}[b(V_s) - b(V_{i\Delta})]ds}_{\text{Residual term}}\,.\end{aligned}$$

This regression of the $(V_{(i+1)\Delta} - V_{i\Delta})/\Delta$ on the $V_{i\Delta}$ allows to estimate b (see Comte *et al.* (2007)).
Suppose we observe the $(\bar{V}_i)$, then, we can write

$$\begin{aligned}\bar{V}_i &= \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} V_s ds = \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta}\left(V_{i\Delta} + \int_{i\Delta}^{s} dV_u\right)ds \\ &= V_{i\Delta} + \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta}[(i+1)\Delta - u]dV_u.\end{aligned}$$

So we have

$$\frac{\bar{V}_{i+1} - \bar{V}_i}{\Delta} = \frac{V_{(i+1)\Delta} - V_{i\Delta}}{\Delta} + \frac{1}{\Delta^2}\left[\int_{(i+1)\Delta}^{(i+2)\Delta}((i+2)\Delta - u)dV_u + \int_{i\Delta}^{(i+1)\Delta}(u - (i+1)\Delta)dV_u\right].$$

Introducing

$$\psi_{i\Delta}(u) = (u - i\Delta)\mathbf{1}_{[i\Delta,(i+1)\Delta[}(u) + [(i+2)\Delta - u]\mathbf{1}_{[(i+1)\Delta,(i+2)\Delta[}(u) \qquad (4)$$

leads to

$$\frac{\bar{V}_{i+1}-\bar{V}_i}{\Delta} = b(V_{i\Delta}) + \underbrace{\frac{1}{\Delta^2}\int_{i\Delta}^{(i+2)\Delta} \psi_{i\Delta}(u)\sigma(V_u)dW_u}_{\text{noise}} + \underbrace{\frac{1}{\Delta^2}\int_{i\Delta}^{(i+2)\Delta} \psi_{i\Delta}(u)[b(V_u)-b(V_{i\Delta})]du}_{\text{residual}}.$$

The last step is to use the quadratic variations $(\hat{\bar{V}}_i)$ built using our effective observations (recall that we have set $k\delta = \Delta$). We write:

$$\hat{\bar{V}}_i = \bar{V}_i + u_{i,k},$$

where

$$u_{i,k} = \frac{1}{\Delta}\sum_{j=0}^{k-1}\left[\left(\int_{(ik+j)\delta}^{(ik+j+1)\delta} \sqrt{V_s}dB_s\right)^2 - \int_{(ik+j)\delta}^{(ik+j+1)\delta} V_s ds\right].$$

This yields

$$Y_i^{(1)} = \frac{\hat{\bar{V}}_{i+1}-\hat{\bar{V}}_i}{\Delta} = \frac{\bar{V}_{i+1}-\bar{V}_i}{\Delta} + \frac{u_{i+1,k}-u_{i,k}}{\Delta}.$$

Finally, we obtain the development,

$$Y_{i+1}^{(1)} = b(\hat{\bar{V}}_i) + Z_{i+1}^{(1)} + R^{(1)}(i+1), \tag{5}$$

where $Z_{i+1}^{(1)}$ is a noise term (with martingale properties):

$$Z_{i+1}^{(1)} = \frac{1}{\Delta^2}\int_{(i+1)\Delta}^{(i+3)\Delta} \psi_{(i+1)\Delta}(u)\sigma(V_u)dW_u + (u_{i+2,k}-u_{i+1,k})/\Delta,$$

and $R^{(1)}(i+1)$ is a sum of negligible residual terms given by

$$R^{(1)}(i+1) = [b(V_{(i+1)\Delta}) - b(\hat{\bar{V}}_i)] + \frac{1}{\Delta^2}\int_{(i+1)\Delta}^{(i+3)\Delta} \psi_{(i+1)\Delta}(s)(b(V_s)-b(V_{(i+1)\Delta}))ds.$$

The lag in (5) is to avoid some cumbersome correlations. On the other hand, following the same steps, we have, for $\ell = 2$ $(f^{(2)} = \sigma^2)$,

$$Y_{i+1}^{(2)} = \frac{3}{2}\frac{(\hat{\bar{V}}_{i+2}-\hat{\bar{V}}_{i+1})^2}{\Delta} = \sigma^2(\hat{\bar{V}}_i) + Z_{i+1}^{(2)} + R^{(2)}(i+1), \tag{6}$$

with $Z_{i+1}^{(2)} = Z_{i+1}^{(2,1)} + Z_{i+1}^{(2,2)} + Z_{i+1}^{(2,3)}$ and

$$Z_{i+1}^{(2,1)} = \frac{3}{2\Delta^3}\left[\left(\int_{(i+1)\Delta}^{(i+3)\Delta} \psi_{(i+1)\Delta}(s)\sigma(V_s)dW_s\right)^2 - \int_{(i+1)\Delta}^{(i+3)\Delta} \psi^2_{(i+1)\Delta}(s)\sigma^2(V_s)ds\right],$$

$$\begin{aligned} Z_{i+1}^{(2,2)} &= \frac{3}{\Delta}b(V_{(i+1)\Delta})\int_{(i+1)\Delta}^{(i+3)\Delta} \psi_{(i+1)\Delta}(s)\sigma(V_s)dW_s \\ &\quad + \frac{3}{\Delta^3}\int_{(i+1)\Delta}^{(i+3)\Delta}\left(\int_s^{(i+3)\Delta} \psi^2_{(i+1)\Delta}(u)du\right)[(\sigma^2)'\sigma](V_s)dW_s, \end{aligned}$$

where $\psi_{i\Delta}$ is given in (4), and

$$Z_{i+1}^{(2,3)} = \frac{3}{\Delta}(\bar{V}_{i+2}-\bar{V}_{i+1})(u_{i+2,k}-u_{i+1,k}).$$

The residual term $R^{(2)}(i+1)$ is detailed in Section 6.3.

2.2 Spaces of approximation

The functions b and σ^2 are estimated only on a compact subset A of the state space of (V_t). For simplicity and without loss of generality, we assume from now on that

$$A = [0,1], \text{ and we set } \quad b_A = b1_A, \quad \sigma_A = \sigma 1_A. \tag{7}$$

To estimate $f = b, \sigma^2$, we consider a family $S_m, m \in \mathcal{M}_n$ of finite dimensional subspaces of $\mathbb{L}_2([0,1])$ and compute a collection of estimators $\hat{f}_m$ where for all m, $\hat{f}_m$ belongs to S_m. Afterwards, a data driven procedure chooses among the collection of estimators the final estimator $\hat{f}_{\hat{m}}$.

We consider here simple projection spaces, namely trigonometric spaces, $S_m, m \in \mathcal{M}_n$. The space S_m is linearly spanned in $\mathbb{L}_2([0,1])$ by $\varphi_1, \ldots, \varphi_{2m+1}$ with $\varphi_1(x) = 1_{[0,1]}(x)$, $\varphi_j(x) = \sqrt{2}\cos(2\pi jx)1_{[0,1]}(x)$ for even j's and $\varphi_j(x) = \sqrt{2}\sin(2\pi jx)1_{[0,1]}(x)$ for odd j's larger than 1. We have $D_m = 2m+1 = \dim(S_m) \leq \mathcal{D}_n$ and $\mathcal{M}_n = \{1, 3, \ldots, \mathcal{D}_n\}$. The largest space in the collection has maximal dimension $\mathcal{D}_n$, which is subject to constraints appearing later. Note that, for all $x \in [0,1]$, $\sum_{j=1}^{2m+1} \varphi_j^2(x) = 2m+1 = D_m$. Thus, for any function $t \in S_m$, $\sup_{x\in[0,1]} |t(x)|^2 \leq D_m \int_0^1 t^2(x)dx$.

2.3 The collection of mean squares estimators

Equations (5)-(6) give the adequate regression equations to estimate $f^{(\ell)}$. We consider the collection of spaces (S_m) described above. For each m, and for a function $t \in S_m$, we introduce, for $\ell = 1, 2$, the following contrast:

$$\gamma_N^{(\ell)}(t) = \frac{1}{N} \sum_{i=0}^{N-1} [Y_{i+1}^{(\ell)} - t(\hat{\bar{V}}_i)]^2. \tag{8}$$

Then the mean squares estimators are defined as

$$\hat{f}_m^{(\ell)} = \arg\min_{t \in S_m} \gamma_N^{(\ell)}(t). \tag{9}$$

In a first step, we study the risk of the above estimators (with m fixed). For this, we have to consider a well-defined risk. Let us remark that the minimization of $\gamma_N^{(\ell)}$ over S_m may lead to several solutions. In contrast, the random $\mathbb{R}^N$-vector $(\hat{f}_m^{(\ell)}(\hat{\bar{V}}_0), \ldots, \hat{f}_m^{(\ell)}(\hat{\bar{V}}_{N-1}))'$ is always uniquely defined. Indeed, let us denote by Π_m the orthogonal projection (with respect to the inner product of $\mathbb{R}^N$) onto the subspace of $\mathbb{R}^N$, $\{(t(\hat{\bar{V}}_0), \ldots, t(\hat{\bar{V}}_{N-1}))', t \in S_m\}$, then $(\hat{f}_m^{(\ell)}(\hat{\bar{V}}_0), \ldots, \hat{f}_m^{(\ell)}(\hat{\bar{V}}_{N-1}))' = \Pi_m Y^{(\ell)}$ where $Y^{(\ell)} = (Y_1^{(\ell)}, \ldots, Y_N^{(\ell)})'$. This is the reason why we consider a specific risk for $\hat{f}_m^{(\ell)}$ based on the design points, i.e.

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=0}^{N-1} (\hat{f}_m^{(\ell)}(\hat{\bar{V}}_i) - f^{(\ell)}(\hat{\bar{V}}_i))^2\right].$$

Thus, the error is measured via the risk $\mathbb{E}(\|\hat{f}_m^{(\ell)} - f^{(\ell)}\|_N^2)$ where

$$\|t\|_N^2 = \frac{1}{N}\sum_{i=0}^{N-1} t^2(\hat{\bar{V}}_i). \tag{10}$$

Now, we use the contrasts (8) and the definition (9) to bound the empirical norms $\|\hat{f}_m^{(\ell)} - f^{(\ell)}\|_N^2$. The following decomposition of the contrasts holds:

$$\gamma_N^{(\ell)}(t) - \gamma_N^{(\ell)}(f^{(\ell)}) = \|t - f^{(\ell)}\|_N^2 - \frac{2}{N}\sum_{i=0}^{N-1}(Y_{i+1}^{(\ell)} - f^{(\ell)}(\hat{\bar{V}}_i))(f^{(\ell)}(\hat{\bar{V}}_i) - t(\hat{\bar{V}}_i)).$$

In view of (5)-(6), we define the centered empirical processes, for $\ell = 1, 2$:

$$\nu_N^{(\ell)}(t) = \frac{1}{N}\sum_{i=0}^{N-1} t(\hat{\bar{V}}_i)Z_{i+1}^{(\ell)},$$

and the residual processes:

$$R_N^{(\ell)}(t) = \frac{1}{N}\sum_{i=0}^{N-1} t(\hat{\bar{V}}_i)R^{(\ell)}(i+1).$$

Using these notations, we obtain that

$$\gamma_N^{(\ell)}(t) - \gamma_N^{(\ell)}(f^{(\ell)}) = \|t - f^{(\ell)}\|_N^2 - 2\nu_N^{(\ell)}(t - f^{(\ell)}) - 2R_N^{(\ell)}(t - f^{(\ell)}).$$

Let $f_m^{(\ell)}$ be the orthogonal projection of $f^{(\ell)}$ on S_m. By definition of the estimators, the following inequality holds: $\gamma_N^{(\ell)}(\hat{f}_m^{(\ell)}) \leq \gamma_N^{(\ell)}(f_m^{(\ell)})$. Therefore, $\gamma_N^{(\ell)}(\hat{f}_m^{(\ell)}) - \gamma_N^{(\ell)}(f^{(\ell)}) \leq \gamma_N^{(\ell)}(f_m^{(\ell)}) - \gamma_N^{(\ell)}(f^{(\ell)})$. This yields

$$\|\hat{f}_m^{(\ell)} - f^{(\ell)}\|_N^2 \leq \|f_m^{(\ell)} - f^{(\ell)}\|_N^2 + 2\nu_N^{(\ell)}(\hat{f}_m^{(\ell)} - f_m^{(\ell)}) + 2R_N^{(\ell)}(\hat{f}_m^{(\ell)} - f_m^{(\ell)}).$$

The functions $\hat{f}_m^{(\ell)}$ and $f_m^{(\ell)}$ being A-supported, we can cancel the terms $\|f^{(\ell)}\mathbf{1}_{A^c}\|_N^2$ that appears in both sides of the inequality. So, we get

$$\|\hat{f}_m^{(\ell)} - f_A^{(\ell)}\|_N^2 \leq \|f_m^{(\ell)} - f_A^{(\ell)}\|_N^2 + 2\nu_N^{(\ell)}(\hat{f}_m^{(\ell)} - f_m^{(\ell)}) + 2R_N^{(\ell)}(\hat{f}_m^{(\ell)} - f_m^{(\ell)}). \tag{11}$$

Let us denote by $\|t\|^2 = \int_0^1 t^2(x)dx$. To find the rate of the risks, we have to take expectations and find upper bounds for

$$\mathbb{E}(\sup_{t\in S_m, \|t\|=1}[\nu_N^{(\ell)}(t)]^2) \quad \text{and} \quad \mathbb{E}(\sup_{t\in S_m, \|t\|=1}[R_N^{(\ell)}(t)]^2).$$

3 The assumptions

3.1 Model assumptions.

Let $(X_t, V_t)_{t\geq 0}$ be given by (1) and assume that only discrete time observations of X, $(X_{\ell\delta})_{1\leq \ell\leq n+1}$ are available. We want to estimate the drift function b and the square of the diffusion coefficient σ^2. We assume that the state space of (V_t) is a known open interval (r_0, r_1) of $\mathbb{R}^+$ and consider the following set of assumptions.

[A1] $0 \leq r_0 < r_1 \leq +\infty$, $\overset{\circ}{I} = (r_0, r_1)$, with $\sigma(v) > 0$, for all $v \in \overset{\circ}{I}$. Let $I = [r_0, r_1] \cap \mathbb{R}$. The function b belongs to $C^1(I)$, b' is bounded on I, $\sigma^2 \in C^2(I)$, $(\sigma^2)'\sigma$ is Lipschitz on I, $(\sigma^2)''$ is bounded on I and $\sigma^2(v) \leq \sigma_1^2$ for all v in I.

[A2] For all $v_0, v \in \overset{\circ}{I}$, the scale density $s(v) = \exp\left[-2\int_{v_0}^{v} b(u)/\sigma^2(u)du\right]$ satisfies $\int_{r_0} s(x)dx = +\infty = \int^{r_1} s(x)dx$, and the speed density $m(v) = 1/(\sigma^2(v)s(v))$ satisfies $\int_{r_0}^{r_1} m(v)dv = M < +\infty$.

[A3] $\eta \sim \pi$ and $\forall i, \mathbb{E}(\eta^i) < \infty$, where $\pi(v)dv = (m(v)/M)\mathbf{1}_{(r_0,r_1)}(v)dv$.

Under [A1]-[A3], (V_t) is strictly stationary with marginal distribution π, ergodic and β-mixing, *i.e.* $\lim_{t\to+\infty}\beta_V(t) = 0$. Here, $\beta_V(t)$ denotes the β-mixing coefficient of (V_t) and is given by

$$\beta_V(t) = \int_{r_0}^{r_1} \pi(v)dv\|P_t(v, dv') - \pi(v')dv'\|_{TV}.$$

The norm $\|.\|_{TV}$ is the total variation norm and P_t denotes the transition probability of (V_t) (see Genon-Catalot *et al.* (2000)). To prove our main result, we need the following stronger mixing condition:

[A4] The process (V_t) is exponentially β-mixing, *i.e.*, there exist constants $K > 0, \theta > 0$, such that, for all $t \geq 0$, $\beta_V(t) \leq Ke^{-\theta t}$.

Assumption [A4] is satisfied in most standard examples. Under [A1]-[A4], for fixed Δ, $(\bar{V}_i)_{i\geq 0}$ is a strictly stationary process. And we have:

Proposition 3.1 *Under* [A1]-[A4], *for fixed k and δ, $(\hat{\bar{V}}_i)_{i\geq 0}$ is strictly stationary and $\beta_{\hat{\bar{V}}}(i) \leq c\beta_V(i\Delta)$ for all $i \geq 1$.*

In connection with the collection of spaces S_m, we need an additional assumption on the marginal density of the stationary process $(\hat{\bar{V}}_i)_{i\geq 0}$:

[A5] The process $(\hat{\bar{V}}_i)_{i\geq 0}$ admits a stationary density π^* and there exist two positive constants π_0^* and π_1^* (independent of n, δ) such that $\forall m \in \mathcal{M}_n$, $\forall t \in S_m$,

$$\pi_0^*\|t\|^2 \leq \mathbb{E}(t^2(\hat{\bar{V}}_0)) \leq \pi_1^*\|t\|^2. \tag{12}$$

The existence of the density π^* is easy to obtain. The checking of (12) is more technical. See the discussion on [A5] in Section 6.2. Below, we use the notations:

$$\|t\|_{\pi^*}^2 = \int t^2(x)\pi^*(x)dx, \quad \|t\|^2 = \int_0^1 t^2(x)dx \quad \text{and} \quad \|t\|_\infty = \sup_{x\in[0,1]}|t(x)|. \tag{13}$$

Let us mention that for a deterministic function $\mathbb{E}(\|t\|_N^2) = \|t\|_{\pi^*}^2 = \int t^2(x)\pi^*(x)dx$, where $\|.\|_N$ is defined by (10). Moreover, under Assumption [A5], the norms $\|.\|$ and $\|.\|_{\pi^*}$ are equivalent for functions in S_m (see notations (13)).

3.2 Risk for the collection of drift estimators

For the estimation of b, we obtain the following result.

Proposition 3.2 *Assume that $N\Delta \geq 1$ and $1/k \leq \Delta$. Assume that* [A1]-[A5] *hold and consider a model S_m in the collection of models with $\mathcal{D}_n \leq O(\sqrt{N\Delta}/\ln(N))$ where $\mathcal{D}_n$ is the maximal dimension (see Section 2.2). Then the estimator $\hat{f}_m^{(1)} = \hat{b}_m$ of $f^{(1)} = b$ is such that*

$$\mathbb{E}(\|\hat{b}_m - b_A\|_N^2) \leq 7\|b_m - b_A\|_{\pi^*}^2 + K\frac{\mathbb{E}(\sigma^2(V_0))D_m}{N\Delta} + K'\Delta, \tag{14}$$

where $b_A = b\mathbf{1}_{[0,1]}$ and K and K' are some positive constants.

Note that the condition on $\mathcal{D}_n$ implies that $\sqrt{N\Delta}/\ln(N)$ must be large enough.

It follows from (14) that it is natural to select the dimension D_m that leads to the best compromise between the squared bias term $\|b_m - b_A\|_{\pi^*}^2$ (which decreases when D_m increases) and the variance term of order $D_m/(N\Delta)$.

Now, let us consider the classical high frequency data setting: let $\Delta = \Delta_n$, $k = k_n$ and $N = N_n$ be, in addition, such that $\Delta_n \to 0$, $N = N_n \to +\infty$, $N_n\Delta_n/\ln^2(N_n) \to +\infty$ when $n \to +\infty$ and that $1/(k_n\Delta_n) \leq 1$. Assume for instance that b_A belongs to a ball of some Besov space, $b_A \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, $\alpha \geq 1$, and $\|b_A\|_{\alpha,2,\infty} \leq L$. Assume also that $\|b_m - b_A\|_{\pi^*}^2 \leq \pi_1^*\|b_m - b_A\|^2$. Applying Lemma 12 in Barron *et al.* (1999)), we get that $\|b_A - b_m\|_{\pi^*}^2 \leq C(\alpha, L, \pi_1^*)D_m^{-2\alpha}$. Therefore, if we choose $D_m = (N_n\Delta_n)^{1/(2\alpha+1)}$, we obtain

$$\mathbb{E}(\|\hat{b}_m - b_A\|_n^2) \leq C(\alpha, L, \pi_1^*)(N_n\Delta_n)^{-2\alpha/(2\alpha+1)} + K'\Delta_n. \tag{15}$$

The first term $(N_n\Delta_n)^{-2\alpha/(2\alpha+1)} = T_n^{-2\alpha/(2\alpha+1)}$ is the optimal nonparametric rate proved by Hoffmann (1999) for direct observations of V. Note that, under the standard condition $\Delta_n = O(1/(N_n\Delta_n))$, the last term Δ_n in the risk bound is negligible with respect to $(N_n\Delta_n)^{-2\alpha/(2\alpha+1)}$.

Finally, we must look at the step δ_n. Consider the choices $k_n = 1/\Delta_n$ and $\delta_n = n^{-c}$. Let us see if there are possible choices of c for which all our constraints are fulfilled. To have $n\delta_n \to +\infty$ requires $0 < c < 1$. As $\Delta_n = k_n\delta_n = \delta_n/\Delta_n$, we have $\Delta_n = \sqrt{\delta_n} = n^{-c/2}$ and $N_n = n/k_n = n^{1-c/2}$. Thus, $\Delta_n \to 0$ and $N_n, N_n\Delta_n \to +\infty$. The last constraint to fulfill is that $N_n\Delta_n^2 = n^{1-3c/2} = O(1)$. Thus for $2/3 \leq c < 1$, the dominating term in (15) is $(N_n\Delta_n)^{-2\alpha/(2\alpha+1)}$, *i.e.* the minimax optimal rate. We have thus obtained a possible "bandwidth" of steps δ_n.

3.3 Risk for the collection of volatility estimators

For the collection of volatility estimators, we have the result

Proposition 3.3 *Assume that* [A1]-[A5] *hold and consider a model S_m in the collection of models with maximal dimension $\mathcal{D}_n \leq O(\sqrt{N\Delta}/\ln(N))$. Assume also that $1/k \leq \Delta$ and $N\Delta \geq 1$, $\Delta \leq 1$. Then the estimator $\hat{f}_m^{(2)} = \hat{\sigma}_m^2$ of $f^{(2)} = \sigma^2$ is such that*

$$\mathbb{E}(\|\hat{\sigma}_m^2 - \sigma_A^2\|_N^2) \leq 7\|\sigma_m^2 - \sigma_A^2\|_{\pi^*}^2 + K\frac{\mathbb{E}(\sigma^4(V_0))D_m}{N} + K'Res(D_m, k, \Delta), \tag{16}$$

where the residual term is given by

$$Res(D_m, k, \Delta) = D_m^2\Delta^2 + D_m^5\Delta^3 + \frac{D_m^3}{k^2} + \frac{1}{k^2\Delta^2}, \tag{17}$$

where $\sigma_A^2 = \sigma^2\mathbf{1}_{[0,1]}$, and K, K' are some positive constants.

The discussion on rates is much more tedious. Consider the asymptotic setting described for b. Assume that σ_A^2 belongs to a ball of some Besov space, $\sigma_A^2 \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, and that $\|\sigma_m^2 - \sigma_A^2\|_{\pi^*}^2 \le \pi_1^* \|\sigma_m^2 - \sigma_A^2\|^2$, then $\|\sigma_A^2 - \sigma_m^2\|_{\pi^*}^2 \le C(\alpha, L, \pi_1^*) D_m^{-2\alpha}$, for $\|\sigma_A^2\|_{\alpha,2,\infty} \le L$. Therefore, if we choose $D_m = N_n^{1/(2\alpha+1)}$, and $k_n \le 1/\Delta_n$, we obtain

$$\mathbb{E}(\|\hat{\sigma}_m^2 - \sigma_A^2\|_N^2) \le C(\alpha, L, \pi_1^*) N_n^{-2\alpha/(2\alpha+1)} + K' Res(N_n^{1/(2\alpha+1)}, k_n, \Delta_n). \tag{18}$$

The first term $N_n^{-2\alpha/(2\alpha+1)}$ is the optimal nonparametric rate proved by Hoffmann (1999) when N_n discrete time observations of V with sampling step $1/N_n$ are available (the time interval has fixed length).

For the second term, let us set $k_n = n^a$, $\Delta_n = n^{-b}$, $\delta_n = n^{-c}$, and recall that $n\delta_n = N_n \Delta_n$ and $n/N_n = k_n$, so that $N_n = n^{1-a}$ and $a + b = c$. We look for a, b such that

$$Res(N_n^{1/(2\alpha+1)}, k_n, \Delta_n) \le N_n^{-2\alpha/(2\alpha+1)}.$$

For this, we take $1/(k_n^2 \Delta_n^2) = N_n^{-2\alpha/(2\alpha+1)}$ which implies $2(a-b)/(1-a) = 2\alpha/(2\alpha+1)$. We get

$$a = \frac{(2\alpha+1)c + \alpha}{5\alpha+2}, \quad b = \frac{(3\alpha+1)c - \alpha}{5\alpha+2}.$$

Then we impose $N_n^{2/(2\alpha+1)} \Delta_n^2 \le N_n^{-2\alpha/(2\alpha+1)}$ which is equivalent to

$$2b \ge [(2\alpha+2)/(2\alpha+1)](1-a) \Rightarrow c \ge (3\alpha+2)[2(2\alpha+1)].$$

Next $N_n^{5/(2\alpha+1)} \Delta_n^3 \le N_n^{-2\alpha/(2\alpha+1)}$ leads to

$$3b \ge [(2\alpha+5)/(2\alpha+1)](1-a) \Rightarrow c \ge (7\alpha+5)/(11\alpha+8).$$

Lastly $N_n^{3/(2\alpha+1)}/k_n^2 \le N_n^{-2\alpha/(2\alpha+1)}$ holds for $-2a \le -[(3+2\alpha)/(2\alpha+1)](1-a)$, *i.e.* $c \ge 2(\alpha+3)/(6\alpha+5)$.

The optimal dimension has also to fulfill $N_n^{1/(2\alpha+1)} \le \mathcal{D}_n \le \sqrt{N_n \Delta_n}$ *i.e.* $-[(2\alpha - 1)/[2(2\alpha+1)]](1-a) \le -b/2$ which implies $c \le (5\alpha-2)/(5\alpha)$. Finally, we must have

$$c \in \left[\frac{3\alpha+2}{2(2\alpha+1)}, \frac{5\alpha-2}{5\alpha}\right] \to_{\alpha\to+\infty} \left]\frac{3}{4}, 1\right[.$$

This interval is nonempty as soon as $\alpha > 2$.

In terms of the initial number n of observations, the rate is now $(n^{1-a})^{-2\alpha/(2\alpha+1)}$ where $1-a$ is at most $1/2$. This is consistent with Gloter's (2000) result: in the parametric case, he obtains $n^{-1/2}$ instead of n^{-1} for the quadratic risk.

4 Data driven estimator of the coefficients

The second step is to ensure an automatic selection of D_m, which does not use any knowledge on $f^{(\ell)}$, and in particular which does not require to know the regularity α. This selection is standardly done by setting

$$\hat{m}^{(\ell)} = \arg\min_{m\in\mathcal{M}_n} \left[\gamma_n^{(\ell)}(\hat{f}_m^{(\ell)}) + \mathrm{pen}^{(\ell)}(m)\right], \tag{19}$$

with $\text{pen}^{(\ell)}(m)$ a penalty to be properly chosen. We denote by $\tilde{f}^{(\ell)} = \hat{f}^{(\ell)}_{\hat{m}^{(\ell)}}$ the resulting estimator and we need to determine pen such that, ideally,

$$\mathbb{E}(\|\tilde{f}^{(\ell)} - f^{(\ell)}_A\|^2_N) \leq C \inf_{m \in \mathcal{M}_n} \left(\|f^{(\ell)}_A - f^{(\ell)}_m\|^2 + \frac{\mathbb{E}(\sigma^{2\ell}(V_0))D_m}{N\Delta^{2-\ell}} \right) + \text{ negligible terms,}$$

with C a constant which should not be too large.

4.1 Result for the data driven estimator of b

We almost reach this aim for the estimation of b.

Theorem 4.1 *Assume that* [A1]-[A5] *hold,* $1/k \leq \Delta$, $\Delta \leq 1$ *and* $N\Delta \geq 1$. *Consider the collection of models with maximal dimension* $\mathcal{D}_n \leq O(\sqrt{N\Delta}/\ln(N))$. *Then the estimator* $\tilde{b} = \hat{f}^{(1)}_{\hat{m}^{(1)}}$ *of* b *where* $\hat{m}^{(1)}$ *is defined by (19) with*

$$\text{pen}^{(1)}(m) \geq \kappa\sigma_1^2 \frac{D_m}{N\Delta}, \tag{20}$$

where κ *is a universal constant, is such that*

$$\mathbb{E}(\|\tilde{b} - b_A\|^2_N) \leq C \inf_{m \in \mathcal{M}_n} \left(\|b_m - b_A\|^2_{\pi^*} + \text{pen}^{(1)}(m) \right) + K\left(\Delta + \frac{1}{N\Delta} + \frac{1}{\ln^2(N)k\Delta} \right). \tag{21}$$

For comments on the practical calibration of the penalty, see Section 5.2.

It follows from (21) that the adaptive estimator automatically realizes the bias-variance compromise, provided that the last terms can be neglected as discussed above. Here, the bandwidth for the choices of δ_n is slightly narrowed because of a stronger constraint. More precisely, we choose $1/(k_n\Delta_n) = \Delta_n$ (instead of 1 previously), that is $k_n = \Delta_n^{-2}$, so that $\Delta_n = k_n\delta_n = \Delta_n^{-2}\delta_n^{-1}$. Therefore $\Delta_n = \delta_n^{1/3}$ and if $\delta_n = n^{-c}$, then $\Delta_n = n^{-c/3}$. Also, $N_n = n/k_n = n^{1-2c/3}$, $N_n\Delta_n = n\delta_n = n^{1-c}$, $N_n\Delta_n^2 = n^{1-4c/3}$. Hence if $3/4 < c < 1$, we have altogether: N_n, $N_n\Delta_n/\ln^2(N_n)$ tend to infinity with n, Δ_n, $N_n\Delta_n^2$ tend to zero.

In that case, whenever b_A belongs to some Besov ball (see (15)), and if $\|b_m - b_A\|^2_{\pi^*} \leq \pi_1^*\|b_m - b_A\|^2$, then $\tilde{b}$ achieves the optimal corresponding nonparametric rate. Note that, in the parametric framework, Gloter (2007) obtains an efficient estimation of b in the same asymptotic context.

The above discussion is the basement of the choice of numerical values of Section 5. For $c = 0.75, n = O(10^6)$ and $\delta = n^{-c}$, we get $k = \delta^{-2/3} = O(10^3)$ and $N = 10^3$. Thinking of N as a number of days and k as a number of data per day, these values are in accordance with real financial data (e.g. if 2 data per minute are collected during 8 hours every day, this implies 960 intra day data).

4.2 Result for the data driven estimator of the volatility

We can prove the following Theorem.

Theorem 4.2 *Assume that* [A1]-[A5] *hold,* $1/k \leq \Delta$, $\Delta \leq 1$ *and* $N\Delta \geq 1$. *Consider the collection of models with maximal dimension* $\mathcal{D}_n \leq \sqrt{N\Delta}/\ln(N)$. *Then the estimator* $\tilde{\sigma}^2 = \hat{f}^{(2)}_{\hat{m}^{(2)}}$ *of* σ^2 *where* $\hat{m}^{(2)}$ *is defined by (19) with*

$$\mathrm{pen}^{(2)}(m) \geq \kappa\sigma_1^4 \frac{D_m}{N}, \tag{22}$$

where κ *is a universal constant, is such that*

$$\mathbb{E}(\|\tilde{\sigma}^2 - \sigma_A^2\|_N^2) \leq C \inf_{m\in\mathcal{M}_n} \left(\|\sigma_m^2 - \sigma_A^2\|_{\pi^*}^2 + \mathrm{pen}^{(2)}(m)\right) + C'\widetilde{Res}(N,k,\Delta), \tag{23}$$

where

$$\widetilde{Res}(N,k,\Delta) = N\Delta^3 + N^{5/2}\Delta^{11/2} + \frac{(N\Delta)^{3/2}}{k^2} + \frac{1}{k^2\Delta^2}. \tag{24}$$

Now, if σ_A^2 belongs to a ball of some Besov space, $\sigma_A^2 \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, then automatically,

$$\inf_{m\in\mathcal{M}_n} \left(\|\sigma_m^2 - \sigma_A^2\|_{\pi^*}^2 + \mathrm{pen}^{(2)}(m)\right) = O(N_n^{-2\alpha/(2\alpha+1)})$$

without requiring the knowledge of α. Therefore,

$$\mathbb{E}(\|\tilde{\sigma}^2 - \sigma_A^2\|_N^2) \leq C(\alpha, L)N_n^{-2\alpha/(2\alpha+1)} + C'\widetilde{Res}(N_n, k_n, \Delta_n).$$

It remains to study the residual term. Notice that we do not know the optimal minimax rate for estimating σ^2, under our set of assumptions on the models and on the asymptotic framework. However, Gloter (2000) and Hoffmann (2002), with observations within a fixed length time interval, obtain the parametric rate $n^{-1/2}$ (in variance). Taking this as a benchmark, we try to make the residual less than $O(n^{-1/2})$. Let us set $k_n = n^a$, $\Delta_n = n^{-b}$, hence $N_n = n/k_n = n^{1-a}$ and $N_n\Delta_n = n^{1-(a+b)}$. This yields that $1-a-3b, (5-5a-11b)/2, (3-7a-3b)/2, 2(b-a)$ must all be less than or equal to $-1/2$, in association with $a+b<1$ and $N_n^{1/(2\alpha+1)} \leq \sqrt{N_n\Delta_n}$. This set of constraint is not empty (e.g. $a = 9/16, b = 5/16$ fits).

5 Examples and numerical simulation results

In this section, we consider examples of diffusions and implement the estimation algorithm on simulated data for the stochastic volatility model X given by (1).

5.1 Simulated paths

We consider the processes $V_t^{(i)}$ for $i = 1, \ldots, 4$ specified by the couples of functions b_i, σ_i^2, $i = 1, \ldots, 4$:

1. $b_1(x) = x\left(-\theta\ln(x) + \frac{1}{2}c^2\right), \sigma_1^2(x) = c^2x^2$ which corresponds to $V_t^{(1)} = \exp(U_t)$ for U_t an Ornstein-Uhlenbeck process, $dU_t = -\theta U_t dt + c dW_t$. Whatever the chosen step, U_t is exactly simulated as an autoregressive process of order 1. We took $\theta = 1$ and $c = 0.75$.

2. $b_2(x) = b_0(x-2)$, $\sigma_2^2(x) = \sigma_0^2(x-2)$, where $b_0(x) = -(1-x^2)\left[c^2 x + \frac{\theta}{2}\ln\left(\frac{1+x}{1-x}\right)\right]$ and $\sigma_0^2(x) = c(1-x^2)$ are the drift and diffusion coefficients of the process $\mathrm{th}(U_t)$ ($\mathrm{th}(x) = (e^x - e^{-x})/(e^x + e^{-x})$), with the same parameters as for case 1). The process $V_t^{(2)}$ corresponds to $\mathrm{th}(U_t) + 2$ which is a positive bounded process.
3. $b_3(x) = x(b_0(\ln(x)) + \frac{1}{2}\sigma_0^2(\ln(x)))$ and $\sigma_3^2(x) = x^2\sigma_0^2(\ln(x))$ which corresponds to the process $V_t^{(3)} = \exp(\mathrm{th}(U_t))$.
4. $b_4(x) = dc^2/4 - \theta x, \sigma_4^2(x) = c^2 x$ which corresponds to the Cox-Ingersoll-Ross process, $V_t^{(4)}$. An exact simulated path is obtained by taking the Euclidean norm of a d-dimensional Ornstein-Uhlenbeck process with parameters $-\theta/2$ and $c/2$. We took $d = 9$, $\theta = 0.75$ and $c = 1/3$.

We simulate discrete data $(V_{\ell\delta'}^{(j)})_{1\leq \ell \leq n'}$ for $j = 1, \ldots, 4$ with $\delta' = \delta/6$, $n'\delta' = T$, from which we generate $(X_{\ell\delta}^{(j)})_{1\leq \ell\leq n}$, by using that

$$X_{\ell\delta} - X_{(\ell-1)\delta} = \sqrt{\int_{(\ell-1)\delta}^{\ell\delta} V_s ds}\ \varepsilon_\ell,$$

with (ε_ℓ) i.i.d. $\mathcal{N}(0,1)$ independent of $(V_s, s \geq 0)$. Approximations of the integrated processes are computed by discrete integration (with a trapeze method).

The generated $V_{j\delta'}^{(i)}$, $i = 1, \ldots, 4$ samples have length $N' = 7.2\ 10^6$, for a step $\delta' = 100/N'$, and the integrated process is computed using 6 data, therefore, we obtain $n = 1.2\ 10^6$, for $T = n\delta = 100$. Different values of k are used, but the best value, $k = 1000$, corresponds to $\Delta = k\delta = 0.083$ and $N = 1200$ realized quadratic variation.

5.2 Estimation algorithms and numerical results

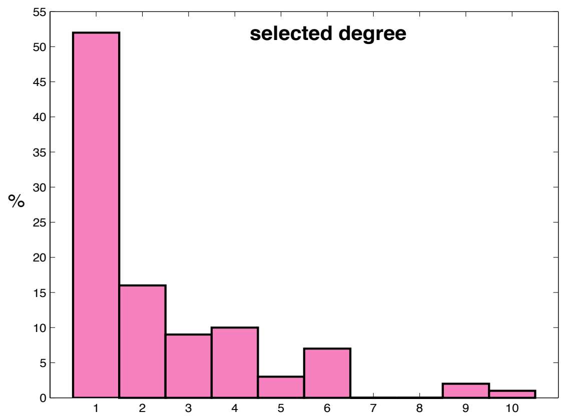

Fig. 1 Histogram of selected values $\hat{m}$ with basis T for 100 paths of $V_t^{(3)}$.

$k=$		250	500	750	1000	1250	1500	1750	2000
$V_t^{(1)}$ mean	T	0.803	**0.187**	0.196	0.253	0.271	0.291	0.302	0.341
(std)		*(0.368)*	*(0.131)*	*(0.176)*	*(0.209)*	*(0.215)*	*(0.214)*	*(0.206)*	*(0.221)*
$V_t^{(1)}$ mean	P	0.731	0.108	**0.100**	0.134	0.164	0.202	0.223	0.250
(std)		*(0.493)*	*(0.146)*	*(0.154)*	*(0.168)*	*(0.214)*	*(0.225)*	*(0.221)*	*(0.240)*
$V_t^{(2)}$ mean	T	11.15	0.784	0.181	0.076	0.049	**0.042**	0.044	0.047
(std)		*(2.086)*	*(0.190)*	*(0.063)*	*(0.043)*	*(0.026)*	*(0.020)*	*(0.017)*	*(0.017)*
$V_t^{(2)}$ mean	P	12.28	0.908	0.199	0.074	0.036	**0.032**	0.034	0.037
(std)		*(2.128)*	*(0.219)*	*(0.093)*	*(0.062)*	*(0.029)*	*(0.018)*	*(0.017)*	*(0.017)*
$V_t^{(3)}$ mean	T	0.960	0.114	**0.068**	0.070	0.074	0.076	0.081	0.084
(std)		*(0.353)*	*(0.063)*	*(0.034)*	*(0.029)*	*(0.029)*	*(0.027)*	*(0.030)*	*(0.030)*
$V_t^{(3)}$ mean	P	1.274	0.133	0.064	**0.051**	0.058	0.062	0.064	0.067
(std)		*(0.422)*	*(0.105)*	*(0.060)*	*(0.028)*	*(0.025)*	*(0.022)*	*(0.020)*	*(0.022)*
$V_t^{(4)}$ mean	T	0.090	9.7 e-3	5.9 e-3	**5.5 e-3**	6.4 e-3	6.2 e-3	7.3 e-3	7.7 e-3
(std)		*(0.036)*	*(0.005)*	*(0.003)*	*(0.003)*	*(0.004)*	*(0.003)*	*(0.005)*	*(0.005)*
$V_t^{(4)}$ mean	P	0.099	7.3 e-3	2.5 e-3	**2.2 e-3**	2.4 e-3	2.6 e-3	3.2 e-3	3.6 e-3
(std)		*(0.045)*	*(0.007)*	*(0.003)*	*(0.002)*	*(0.003)*	*(0.002)*	*(0.004*	*(0.004)*

Table 1 Mean squared errors (with standard deviations in parenthesis) for the estimation of b; 100 paths of the four examples, different values of k for the quadratic variation, when using the trigonometric basis (T) or the polynomial basis (P). In bold, the risk value corresponding to the best k.

We use the algorithm of Comte and Rozenholc (2004). In Comte *et al.* (2007), the precise calibration of penalties, which is quite difficult, is done in detail for the trigonometric basis (denoted hereafter by T) and also for another one, the piecewise polynomial basis (described in this paper, and denoted below by P; see also Comte *et al.* (2008)). The drift penalty ($\ell = 1$) and the diffusion penalty ($\ell = 2$) are given by

$$\kappa_\ell \frac{\hat{s}_\ell^2}{n} \left(D_m + \text{ additive correcting terms} \right), \text{ with } D_m \text{ at most } [N\Delta / \ln^{1.5}(N)].$$

The additive correcting terms involved in the penalty avoid under-penalization and are in accordance with the fact that the theorems provide lower bounds for the penalty. These correcting terms are asymptotically negligible and do not affect the rate of convergence (for details see Comte *et al.* (2007, 2008)).

The constants κ_1 and κ_2 in the drift and diffusion penalties have both been set equal to 6. The term $\hat{s}_1^2$ replaces σ_1^2/Δ for the estimation of b and $\hat{s}_2^2$ replaces σ_1^4 for the estimation of σ^2. Let us first explain how $\hat{s}_2^2$ is obtained. We run once the estimation algorithm of σ^2 with the trigonometric basis and with a preliminary penalty where $\hat{s}_2^2$ is taken equal to $2\max_m(\gamma_n^{(2)}(\hat{\sigma}_m^2))$. This gives a preliminary estimator $\tilde{\sigma}_0^2$. Afterwards, we take $\hat{s}_2$ equal to twice the 99.5%-quantile of $\tilde{\sigma}_0^2$. We get $\tilde{\sigma}^2$. We use this estimate and set $\hat{s}_1^2 = \max_{0\le k\le N-1}(\tilde{\sigma}^2(\hat{\bar{V}}_k))/\Delta$ for the penalty of b. In Figure 1, we give a histogram of the selected values $\hat{m}$ with the trigonometric basis T for 100 paths of $V_t^{(3)}$. This shows that the algorithm selects $\hat{m} = 1$ for 50 paths over 100, and essentially $\hat{m} \in \{1, \dots, 6\}$.

$k =$			250	500	750	1000	1250	1500	1750	2000
$V_t^{(1)}$	mean	T	7.153	0.594	0.284	**0.276**	0.311	0.364	0.399	0.397
	(std)		*(3.055)*	*(0.437)*	*(0.200)*	*(0.241)*	*(0.289)*	*(0.327)*	*(0.367)*	*(0.241)*
$V_t^{(1)}$	mean	P	7.375	0.576	0.216	**0.177**	0.216	0.222	0.304	0.331
	(std)		*(3.171)*	*(0.468)*	*(0.257)*	*(0.166)*	*(0.270)*	*(0.252)*	*(0.385)*	*(0.358)*
$V_t^{(2)}$	mean	T	24.36	1.492	0.299	0.097	0.044	0.030	0.025	**0.024**
	(std)		*(2.97)*	*(0.208)*	*(0.047)*	*(0.023)*	*(0.010)*	*(0.008)*	*(0.006)*	*(0.004)*
$V_t^{(2)}$	mean	P	24.47	1.543	0.3143	0.102	0.046	0.030	**0.024**	**0.024**
	(std)		*(3.068)*	*(0.212)*	*(0.051)*	*(0.025)*	*(0.012)*	*(0.008)*	*(0.007)*	*(0.004)*
$V_t^{(3)}$	mean	T	3.819	0.231	0.043	0.017	**0.013**	0.014	0.016	0.021
	(std)		*(0.353)*	*(0.063)*	*(0.034)*	*(0.029)*	*(0.029)*	*(0.027)*	*(0.030)*	*(0.030)*
$V_t^{(3)}$	mean	P	3.851	0.243	0.049	0.022	**0.015**	0.016	0.020	0.025
	(std)		*(0.898)*	*(0.077)*	*(0.021)*	*(0.012)*	*(0.009)*	*(0.011)*	*(0.012)*	*(0.012)*
$V_t^{(4)}$	mean	T	0.034	2.6e-3	5.63e-4	2.22e-4	1.40e-4	1.09e-4	**1.07e-4**	1.08e-4
	(std)		*(0.014)*	*(9.44 e-4)*	*(2.30 e-4)*	*(1.23 e-4)*	*(5.95 e-5)*	*(8.91 e-5)*	*(5.11 e-5)*	*(6.46 e-5)*
$V_t^{(4)}$	mean	P	0.040	2.56e-3	5.35e-4	1.87e-4	**9.09e-5**	**5.37e-5**	5.81e-5	6.82e-5
	(std)		*(0.014)*	*(0.001)*	*(0.0002)*	*(0.0002)*	*(0.0001)*	*(0.0001)*	*(0.0001)*	*(0.0001)*

Table 2 Mean squared errors (with standard deviations in parenthesis) for the estimation of σ^2; 100 paths of the four examples, different values of k for the quadratic variation, when using the trigonometric basis (T) or the polynomial basis (P). In bold, the risk value corresponding to the best k.

We give in Tables 1-2 results of Monte-Carlo type experiments. Estimated risks are computed as the mean over 100 simulated paths of the empirical norms (e.g. $(1/N\sum_{i=0}^{N-1}[b(\hat{\bar{V}}_i)-\tilde{b}(\hat{\bar{V}}_i)]^2$ for b). Results for the estimator of b are given in Table 1 and for σ^2 in Table 2. We use both bases T (trigonometric) and P (piecewise polynomials) for the four processes choosing different values of k for building the quadratic variations. Clearly, there is an optimal value. If k is too large, there are not enough observations left for the estimation algorithm. If k is too small, bias phenomena appear, related to the violation of the theoretical assumptions (mainly $1/k \le \Delta$). In general, for this sample size, the choice $k = 1000$ seems to be relevant for the basis P. In view of these tables, estimated risks are slightly better for basis P. This is why our graphs (Figures 2-5) are done for basis P.

The results given by our algorithm are also described in two types of figures. First, Figures 2-3 show the functions (true (dotted) and estimated (thick)) and the data points with coordinates $(Y_{i+1}^{(\ell)}, \hat{\bar{V}}_i), i = 1, \dots, N$, for $\ell = 1$ (left), $\ell = 2$ (right). The figures show that the points are scattered over a large area, and that, with the scale they impose, the true function is well estimated. Second, we plot in Figures 4-5 the true function (thick curve) and 10 estimated functions (thin curves) for b (left) and σ^2 (right). In the simulation, we generate a large sample of the integrated process. We compare the estimated functions using data of the integrated process (top curves) with the estimations using the realized quadratic variations (bottom curves). One must keep in mind when looking at the figures that there are 1000 times more data for estimation in the former case than in the latter. Therefore, these curves illustrate the good performance of our approach.

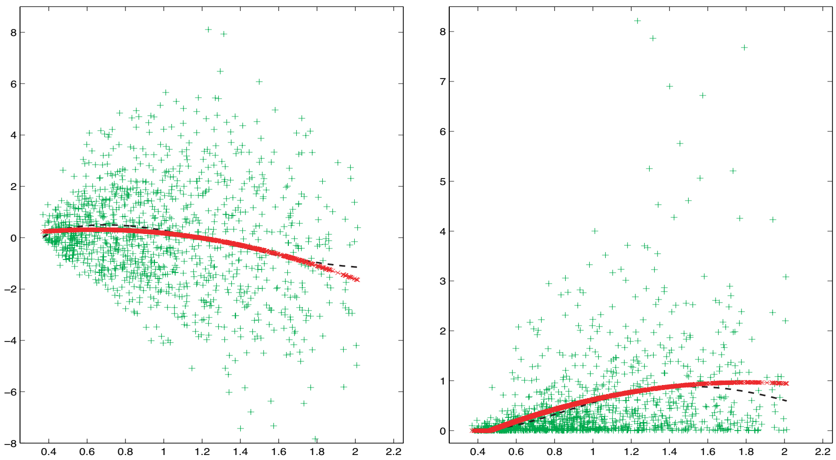

Fig. 2 Estimation of b (left) and σ^2 (right) for one path of the exponential OU process $(V_t^{(1)})$ with $N = 1200$ realized volatilities in the SV model $(k = 1000, T = 100)$. True function (dotted), estimated (thick). The scatter plot gives the data points used for the regressions.

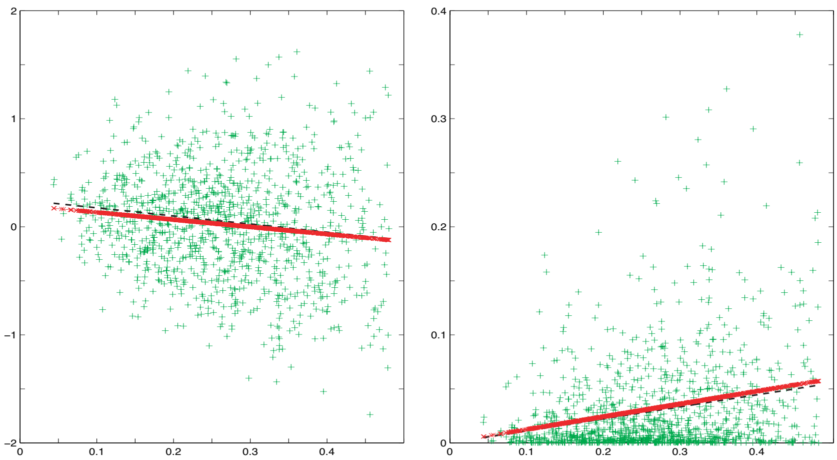

Fig. 3 Estimation of b (left) and σ^2 (right) for one path of the CIR process $(V_t^{(4)})$ with $N = 1200$ realized volatilities in the SV model (bottom) $(k = 1000, T = 100)$. True function (dotted), estimated (thick). The scatter plot gives the data points used for the regressions.

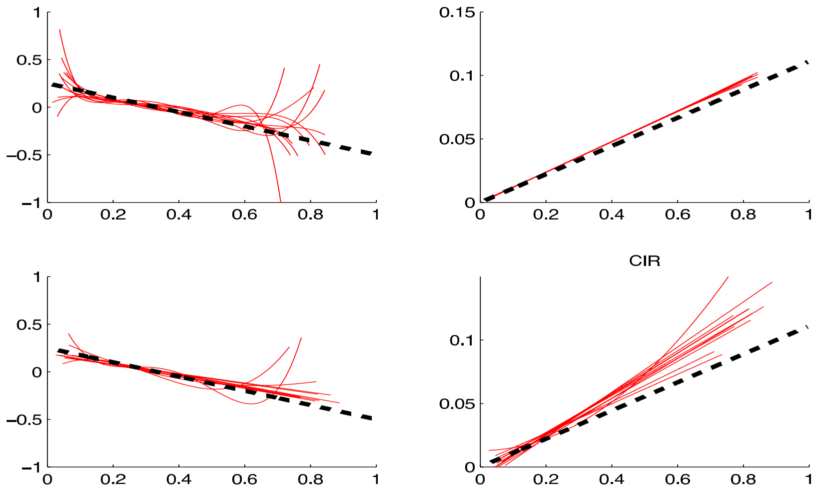

Fig. 4 Estimation of b (left) and σ^2 (right) for 10 paths of the CIR process $(V_t^{(4)})$ with $1.2\ 10^6$ observations of the integrated process (top) and 1200 realized volatilities in the SV model (bottom) $(k = 1000, T = 100)$. True: bold dashed, estimated: full grey (red) lines.

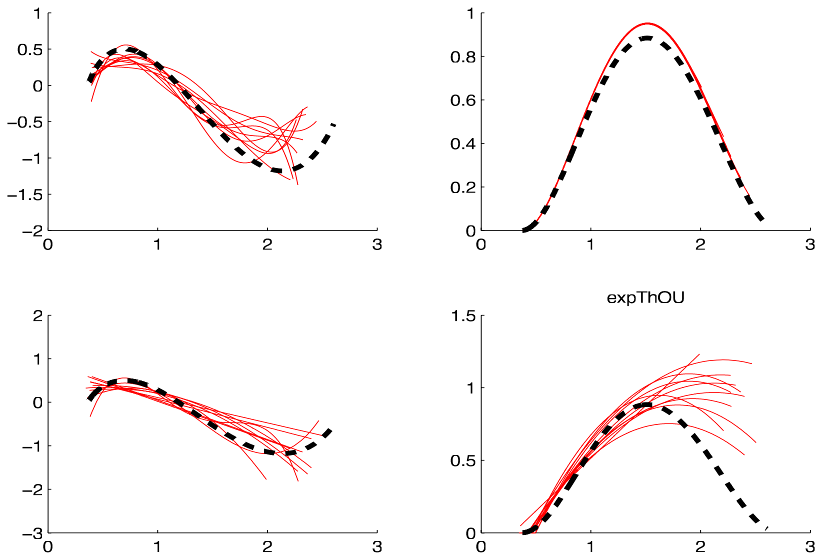

Fig. 5 Estimation of b (left) and σ^2 (right) for 10 paths of $(V_t^{(3)})$ with $1.2\ 10^6$ observations of the integrated process (top) and 1200 realized volatilities in the SV model (bottom) $(k = 1000, T = 100)$. True: bold dashed, estimated: full grey (red) lines.

5.3 Concluding remarks

In this paper, we have described a new nonparametric method for estimating both the drift and the diffusion coefficient of the volatility in a stochastic volatility model. Many papers have shown that this is a difficult task that requires first to transform discrete data by computing realized quadratic variations. A large number of observations are thus needed for this purpose. Our theoretical method provides adaptive estimators that can be implemented by feasible algorithms. Numerical results based on simulated data demonstrate that the method performs well for sample sizes corresponding to real data context ($k = 1000, N = 1200$).

Let us mention that the simulation method may generate some micro-structure type noise which we do not take into account. Indeed, we generate exact samples of the discretized processes ($V_{j\delta'}$), but we approximate the integrals of V. Corrections inspired by Zhang *et al.* (2005) may thus be experimented.

6 Discussion on the assumptions and proofs

6.1 Proof of Proposition 3.1

We start with some preliminaries. Let $I_t = \int_0^t V_s ds$. The joint process $(V_t, I_t)_{t\geq 0}$ is a two dimensional diffusion satisfying:

$$\begin{cases} dV_t = b(V_t)dt + \sigma(V_t)dW_t, \ V_0 = \eta, \\ dI_t = V_t dt, \ \ I_0 = 0. \end{cases}$$

Under regularity assumptions on b and σ, this process admits a transition density, say $q_t(v_0, i_0; v, i)$ for the conditional density of (V_t, I_t) given $V_0 = v_0$, $I_0 = i_0$. This density is w.r.t. the Lebesgue measure on $(0, +\infty)^2$ (see Rogers and Williams (2000)). We assume that these assumptions hold.

Now, let us set

$$J_{\ell\delta} = \int_{(\ell-1)\delta}^{\ell\delta} V_s ds, \ \ell \geq 1. \tag{25}$$

The discrete time process $(V_{\ell\delta}, J_{\ell\delta})_{\ell\geq 1}$ is strictly stationary and Markov. Its one step transition operator is given by the density:

$$(v, j) \to q_\delta(v_0, 0; v, j) := q_\delta(v_0; v, j).$$

Its stationary density is given by $\int \pi(v_0)dv_0 q_\delta(v_0; v, j) := \pi_\delta(v, j)$.

Let us set, for $\ell \geq 1$,

$$Z_\ell = X_{\ell\delta} - X_{(\ell-1)\delta} \tag{26}$$

and define ε_ℓ by the relation: $Z_\ell = J_{\ell\delta}^{1/2}\varepsilon_\ell$. Conditionally on $(V_t)_{t\geq 0}$, the random variables (r.v.) $Z_\ell, \ell \geq 1$ are independent and Z_ℓ has distribution $\mathcal{N}(0, J_{\ell\delta})$. Consequently, the r.v. $(\varepsilon_\ell, \ell \geq 1)$ are i.i.d. with distribution $\mathcal{N}(0, 1)$ and the sequence $(\varepsilon_\ell, \ell \geq 1)$ is independent of $(V_t)_{t\geq 0}$. Hence $(Z_\ell)_{\ell\geq 1}$ and $(\hat{\bar{V}}_i)_{i\geq 0}$ are strictly stationary processes. From the preliminaries and the above remarks, we deduce that the process $(V_{\ell\delta}, J_{\ell\delta}, \varepsilon_\ell)_{\ell\geq 1}$ is stationary Markov. Its ℓ-step transition operator is given by:

$$Q_\ell^\delta(v_0; dv, dj, du) = q_\delta^{(\ell)}(v_0; v, j)n(u)dvdjdu$$

where $q_\delta^{(\ell)}(v_0; v, j)$ is the ℓ-step transition density of $(V_{\ell\delta}, J_{\ell\delta})$ and $n(u)$ is the standard gaussian density. The stationary density of $(V_{\ell\delta}, J_{\ell\delta}, \varepsilon_\ell)_{\ell\geq 1}$ is $\pi_\delta(v,j)n(u)$. Hence

$$\begin{aligned}\|Q_\delta^{(\ell)}(v_0; dv, dj, du) - \pi_\delta(v,j)n(u)dvdjdu\|_{TV} &= \int |q_\delta^{(\ell)}(v_0, v_j) - \pi_\delta(v,j))|n(u)dvdjdu \\ &= \int |q_\delta^{(\ell)}(v_0; v, j) - \pi_\delta(v,j))|dvdj.\end{aligned}$$

We may now use the representation of the β-mixing coefficient of strictly stationary Markov processes (see e.g. Genon-Catalot *et al.* (2000)) to compute

$$\begin{aligned}\beta_{V_{.\delta},J_{.\delta},\varepsilon}(\ell) &= \int \pi_\delta(v_0, j_0)n(u_0)du_0dv_0dj_0\|Q_\delta^{(\ell)}(v_0; dv, dj, du) - \pi_\delta(v,j)n(u)dvdjdu\|_{TV} \\ &= \beta_{V_{.\delta},J_{.\delta}}(\ell).\end{aligned}$$

Now, we have $\beta_Z(\ell) \leq \beta_{V_{.\delta},J_{.\delta},\varepsilon}(\ell) = \beta_{V_{.\delta},J_{.\delta}}(\ell) \leq \beta_V((\ell-1)\delta)$. Finally,

$$\beta_{\hat{\bar{V}}}(i) \leq \beta_Z(ik) \leq \beta_V((ik-1)\delta) \leq c\beta_V(i\Delta). \quad \square$$

6.2 Discussion on the assumptions

Actually, Assumption [A3] is too strong. We only need the existence of moments up to a certain order. Let us now discuss [A5]. Using the representation

$$\hat{\bar{V}}_0 = \frac{1}{k\delta}\sum_{\ell=1}^{k} J_{\ell\delta}\ \varepsilon_\ell^2,$$

we see that $\hat{\bar{V}}_0$ has a conditional density given $(V_t, t \geq 0)$. Integrating this density w.r.t. the distribution of $(J_{\ell\delta}, \ell = 1, \ldots, k)$, we get that $\hat{\bar{V}}_0$ has a density π^*. However the formula for π^* is untractable.

On the other hand, we can obtain (12) by another approach. We have

$$t^2(\bar{V}_0) = t^2(V_0) + (\bar{V}_0 - V_0)(t^2)'(V_0) + \frac{1}{2}(\bar{V}_0 - V_0)^2\int_0^1 (t^2)''(V_0 + u(\bar{V}_0 - V_0))du.$$

Now we use that, for any $t \in S_m$, there exists some constant C such that

$$\|(t^2)'\|_\infty \leq CD_m^2\|t\|^2 \text{ and } \|(t^2)''\|_\infty \leq CD_m^3\|t\|^2.$$

Noting that $|\mathbb{E}\left(\bar{V}_0 - V_0|\mathcal{F}_0\right)| = O(\Delta)$, we get $|\mathbb{E}[(\bar{V}_0 - V_0)(t^2)'(V_0)]| \leq CD_m^2\Delta\|t\|^2 = O(D_m^2\Delta)$. On the other hand,

$$\begin{aligned}\left|\mathbb{E}\left[(\bar{V}_0 - V_0)^2\int_0^1 (t^2)"(V_0 + u(\bar{V}_0 - V_0))du\right]\right| &\leq \|(t^2)''\|_\infty\mathbb{E}[(\bar{V}_0 - V_0)^2] \\ &\leq CD_m^3\Delta\|t\|^2.\end{aligned}$$

It follows that $|\mathbb{E}(t^2(\bar{V}_0) - t^2(V_0))| \leq C\Delta D_m^3\|t\|^2$. Next,

$$\begin{aligned}t^2(\hat{\bar{V}}_0) = t^2(\bar{V}_0) &+ (\hat{\bar{V}}_0 - \bar{V}_0)(t^2)'(V_0) + (\hat{\bar{V}}_0 - \bar{V}_0)[(t^2)'(\bar{V}_0) - (t^2)'(V_0)] \\ &+ \frac{1}{2}(\hat{\bar{V}}_0 - \bar{V}_0)^2\int_0^1 (t^2)''(\bar{V}_0 + u(\hat{\bar{V}}_0 - \bar{V}_0))du.\end{aligned}$$

By Gloter's (2007) Proposition 3.1, we have $|\mathbb{E}[(\hat{\bar{V}}_0 - \bar{V}_0)|V_0]| \leq c\delta(1+V_0)^c$ and $\mathbb{E}[|\hat{\bar{V}}_0 - \bar{V}_0|^2] \leq c/k$. Hence

$$|\mathbb{E}(t^2(\hat{\bar{V}}_0) - t^2(\bar{V}_0))| \leq C\|t\|^2(\Delta D_m^2 + \frac{\sqrt{\Delta}D_m^3}{\sqrt{k}} + \frac{D_m^3}{k}).$$

Since $1/k \leq \Delta$

$$|\mathbb{E}(t^2(\hat{\bar{V}}_0) - t^2(V_0))| \leq C\|t\|^2\Delta D_m^3.$$

As there exist two positive constants π_0, π_1 such that $\forall v \in A$, $\pi_0 \leq \pi(v) \leq \pi_1$, we obtain

$$(\pi_0 - C\Delta\mathcal{D}_n^3)\|t\|^2 \leq \|t\|_{\pi^*}^2 \leq (\pi_1 + C\Delta\mathcal{D}_n^3)\|t\|^2.$$

Under the constraint that $\Delta\mathcal{D}_n^3 = o(1)$, we get (12) for n large enough. This constraint is compatible with the other ones, see the discussion after Theorem 4.1.

6.3 Definition of the residuals and their properties

We have

$$R^{(1)}(i+1) = b(\bar{V}_i) - b(\hat{\bar{V}}_i) + R_*^{(1)}((i+1)),$$

where $R_*^{(1)}$ is the residual term for b studied in Comte *et al.* (2008, Proposition 3.1) and defined by

$$R_*^{(1)}(i+1) = b(V_{(i+1)\Delta}) - b(\bar{V}_i) + \frac{1}{\Delta^2}\int_{(i+1)\Delta}^{(i+3)\Delta} \psi_{(i+1)\Delta}(s)(b(V_s) - b(V_{(i+1)\Delta}))ds.$$

On the other hand,

$$R^{(2)}(i+1) = \frac{3}{2}\frac{(u_{i+1,k} - u_{i,k})^2}{\Delta} + [\sigma^2(V_{(i+1)\Delta} - \sigma^2(\hat{\bar{V}}_i)] + R_*^{(2)}(i+1),$$

where $R_*^{(2)}$ is the residual term for σ^2 studied in Comte *et al.* (2008, Propositions 4.1, 4.2 and 4.3) defined by $R_*^{(2)} = \sum_{m=1}^3 R_*^{(2,m)}$ with

$$R_*^{(2,1)}(i) = \frac{3}{2\Delta^3}\left(\int_{i\Delta}^{(i+2)\Delta} \psi_{i\Delta}(s)b(V_s)ds\right)^2,$$

$$R_*^{(2,2)}(i) = \frac{3}{\Delta^3}\left(\int_{i\Delta}^{(i+2)\Delta} \psi_{i\Delta}(u)(b(V_u) - b(V_{i\Delta}))du\right)\left(\int_{i\Delta}^{(i+2)\Delta} \psi_{i\Delta}(u)\sigma(V_u)dW_u\right),$$

$$R_*^{(2,3)}(i) = \frac{3}{2\Delta^3}\int_{i\Delta}^{(i+2)\Delta}\left(\int_s^{(i+2)\Delta} \psi_{i\Delta}^2(u)du\right)\tau_{b,\sigma}(V_s)ds,$$

where $\tau_{b,\sigma} = (\sigma^2/2)(\sigma^2)" + b(\sigma^2)'$. This decomposition is obtained by applying Ito's formula and Fubini's theorem.
We may now summarize the following useful results, proved in Comte *et al.* (2008, Propositions 3.1, 4.1, 4.2 and 4.3):

Lemma 6.1 *Under Assumptions [A1]-[A2]-[A3],*

1. *For* $\ell = 1, 2$, *for* $m = 1, 2$, *for all* i, $\mathbb{E}\{[R_*^{(\ell)}(i)]^{2m}\} \leq c\Delta^{2m\ell}$ *where* c *is a constant.*

2. Let $Z_*^{(1)}(i) = (1/\Delta^2)\int_{i\Delta}^{(i+2)\Delta} \psi_{i\Delta}(s)\sigma(V_s)dW_s$. For all i, $\mathbb{E}([Z_*^{(1)}(i)]^2) \le (2/3\Delta)\mathbb{E}(\sigma^2(V_0))$.

3. For all i, $\mathbb{E}([Z_i^{(2,1)}]^2) \le c_1\mathbb{E}(\sigma^4(V_0))$ and $\mathbb{E}([Z_i^{(2,2)}]^2) \le c_2\sigma_1^2\Delta$.

We also need the following result:

Lemma 6.2 *Under assumptions [A1]-[A3], for any integer i,* $\mathbb{E}[(\bar{V}_i-\hat{\bar{V}}_i)^2] = \mathbb{E}(u_{i,k}^2) \le 2\mathbb{E}(V_0^2)/k$ *and* $\mathbb{E}[(\bar{V}_i - \hat{\bar{V}}_i)^4] = \mathbb{E}(u_{i,k}^4) \le 56\mathbb{E}(V_0^4)/k^2$.

Proof of Lemma 6.2. This follows from Proposition 3.1 p.504 in Gloter (2007).□

6.4 Proof of Propositions 3.2 and 3.3

For sake of brevity, we give both proofs at the same time. The main difference lies in the orders of the expectations and in the appearance of a specific term in the study of the estimator of σ^2. Let us thus define $R_{**}^{(\ell)}$ for $\ell = 1,2$ as $R_{**}^{(1)} = R^{(1)}$ and

$$R_{**}^{(2)}(i+1) = R^{(2)}(i+1) - [\sigma^2(V_{(i+1)\Delta}) - \sigma^2(\hat{\bar{V}}_i)].$$

Moreover let $T_N^{(1)}(t) = 0$ and

$$T_N^{(2)}(t) = \frac{1}{N}\sum_{i=0}^{N-1}(\sigma^2(V_{(i+1)\Delta}) - \sigma^2(\hat{\bar{V}}_i))t(\hat{\bar{V}}_i).$$

Let us consider the set

$$\Omega_N = \left\{\omega/\left|\frac{\|t\|_N^2}{\|t\|_{\pi^*}^2} - 1\right| \le \frac{1}{2},\ \ \forall t \in \cup_{m,m'\in\mathcal{M}_n}(S_m + S_{m'})/\{0\}\right\}. \tag{27}$$

On Ω_N, $\|t\|_{\pi^*} \le \sqrt{2}\|t\|_N$. From (11), we deduce

$$\begin{aligned}\|\hat{f}_m^{(\ell)} - f_A^{(\ell)}\|_N^2 \le \|f_m^{(\ell)} - f_A^{(\ell)}\|_N^2 + \frac{1}{8}\|\hat{f}_m^{(\ell)} - f_m^{(\ell)}\|_{\pi^*}^2 + 16\sup_{t\in S_m,\|t\|_{\pi^*}=1}[\nu_N^{(\ell)}]^2(t)\\ +16\sup_{t\in S_m,\|t\|_{\pi^*}=1}[T_N^{(\ell)}(t)]^2\\ +\frac{1}{8}\|\hat{f}_m^{(\ell)} - f_m^{(\ell)}\|_N^2 + \frac{8}{N}\sum_{i=0}^{N-1}[R_{**}^{(\ell)}(i+1)]^2\\ \le \|f_m^{(\ell)} - f_A^{(\ell)}\|_N^2 + \frac{3}{8}\|\hat{f}_m^{(\ell)} - f_m^{(\ell)}\|_N^2 + 16\sup_{t\in S_m,\|t\|_{\pi^*}=1}[\nu_N^{(\ell)}]^2(t)\\ +\frac{16}{\pi_0^*}\sup_{t\in S_m,\|t\|=1}[T_N^{(\ell)}(t)]^2 + \frac{8}{N}\sum_{i=0}^{N-1}[R_{**}^{(\ell)}(i+1)]^2.\end{aligned}$$

In the last line above, we use the lower bound π_0^* introduced in [A5].
Setting $B_m(0,1) = \{t \in S_m, \|t\| = 1\}$ and $B_m^{\pi^*}(0,1) = \{t \in S_m, \|t\|_{\pi^*} = 1\}$, the following holds on the set Ω_N:

$$\frac{1}{4}\|\hat{f}_m^{(\ell)}-f_A^{(\ell)}\|_N^2 \le \frac{7}{4}\|f_m^{(\ell)}-f_A^{(\ell)}\|_N^2+16\sup_{t\in B_m^{\pi^*}(0,1)}[\nu_N^{(\ell)}]^2(t)+\frac{16}{\pi_0^*}\sup_{t\in B_m(0,1)}[T_N^{(\ell)}(t)]^2+\frac{8}{N}\sum_{i=0}^{N-1}[R_{**}^{(\ell)}(i+1)]^2.$$

We have the following result:

Lemma 6.3 *Under assumptions [A1]-[A3] and [A5], if $1/k \leq \Delta$, we have, for $\ell = 1, 2$*

$$\mathbb{E}\left(\sup_{t\in B_m^{\pi^*}(0,1)} [\nu_N^{(\ell)}]^2(t)\right) \leq K\frac{C_\ell D_m}{N\Delta^{2-\ell}},$$

with $C_\ell = \mathbb{E}(\sigma^{2\ell}(V_0))$.

The Lipschitz condition on b and Lemma 6.2 imply that

$$\mathbb{E}[(b(\bar{V}_i) - b(\hat{\bar{V}}_i))^2] \leq c_b\mathbb{E}[(\bar{V}_i - \hat{\bar{V}}_i)^2] \leq 2c_b\mathbb{E}(V_0^2)/k.$$

Consequently, there exists a constant c such that

$$\mathbb{E}\left(\frac{8}{N}\sum_{i=0}^{N-1}[R_{**}^{(1)}(i+1)]^2\right) \leq c(\Delta + k^{-1}).$$

Thus

$$\mathbb{E}(\|\hat{b}_m - b_A\|_N^2 \mathbf{1}_{\Omega_N}) \leq 7\|b_m - b\|_{\pi^*}^2 + \frac{32}{\pi_0^*}\mathbb{E}\left(\sup_{t\in S_m, \|t\|=1}[\nu_N^{(1)}(t)]^2\right) + c"(\Delta + k^{-1}).$$

By gathering all bounds, we find

$$\mathbb{E}(\|\hat{b}_m - b\|_N^2 \mathbf{1}_{\Omega_N}) \leq 7\|b_m - b\|_{\pi^*}^2 + K\frac{\mathbb{E}(\sigma^2(V_0))D_m}{N\Delta}(1 + \frac{1}{k\Delta}) + K'(\Delta + k^{-1}).$$

On the other hand, Lemma 6.1 and Lemma 6.2 imply that

$$\begin{aligned}\mathbb{E}(\frac{1}{N}\sum_{i=0}^{N-1}[R_{**}^{(2)}(i+1)]^2 &\leq 2\mathbb{E}\left[\frac{1}{N}\sum_{i=0}^{N-1}\left([R_*^{(2)}(i+1)]^2 + \frac{9}{4}\frac{(u_{i+1,k} - u_{i,k})^4}{\Delta^2}\right)\right] \\ &\leq 2c\Delta^2 + \frac{36}{\Delta^2}\mathbb{E}(u_{1,k}^4) \leq C(\Delta^2 + \frac{1}{k^2\Delta^2}).\end{aligned}$$

Next we need to bound $\mathbb{E}\left(\sup_{t\in S_m, \|t\|=1}[T_N^{(2)}(t)]^2\right)$. This is obtained in the following Lemma:

Lemma 6.4 *Under the Assumptions of Proposition 3.3 and if $1/k \leq \Delta$, there exists a constant C such that*

$$\mathbb{E}\left(\sup_{t\in S_m, \|t\|=1}[T_N^{(2)}(t)]^2\right) \leq C(D_m^2\Delta^2 + D_m^5\Delta^3 + D_m^3/k^2 + D_m/(Nk)).$$

We can use Lemma 6.1 in Comte *et al.* (2005) to obtain that, if $\mathcal{D}_n \leq C\sqrt{N\Delta}/\ln(N)$, then

$$\mathbb{P}(\Omega_N^c) \leq \frac{c}{N^4}.$$

This enables to check that $\mathbb{E}(\|\hat{f}_m^{(\ell)} - f^{(\ell)}\|_N^2 \mathbf{1}_{\Omega_n^c}) \leq c/N$ using the same lines as the analogous proof given p.532 in Comte *et al.* (2007). For this reason, details are omitted. □

6.5 Proof of Lemma 6.3.

Case $\ell = 1$. Next, let us define $\mathcal{F}_t = \sigma((W_s, B_s), 0 \le s \le t, \eta)$. We can use martingale properties to see that, $\forall t \in S_m$,

$$\mathbb{E}(t(\hat{\bar{V}}_i)Z_{i+1}^{(1)}) = \mathbb{E}(\mathbb{E}(t(\hat{\bar{V}}_i)Z_{i+1}^{(1)}|\mathcal{F}_{(i+1)\Delta})) = \mathbb{E}(t(\hat{\bar{V}}_i)\mathbb{E}(Z_{i+1}^{(1)}|\mathcal{F}_{(i+1)\Delta})) = 0$$

because the last conditional expectation is zero. Moreover, the same tool shows that the covariance term $\mathbb{E}(t(\hat{\bar{V}}_i)t(\hat{\bar{V}}_\ell)Z_{i+1}^{(1)}Z_{\ell+1}^{(1)})$ for $\ell \ge i+2$ is also null by inserting a conditional expectation given $\mathcal{F}_{(\ell+1)\Delta}$. Consequently, it is now easy to see that

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in S_m, \|t\|=1}[\nu_N^{(1)}(t)]^2\right) &\le \sum_{j=1}^{D_m}\mathbb{E}[\nu_N^2(\varphi_j)] \le \sum_{j=1}^{D_m}\mathrm{Var}\left[\frac{1}{N}\sum_{i=0}^{N-1}\varphi_j(\hat{\bar{V}}_i)Z_{i+1}^{(1)}\right] \\
&\le \frac{2}{N}\sum_{j=1}^{D_m}\mathrm{Var}\left(\varphi_j(\hat{\bar{V}}_1)Z_2^{(1)}\right) \\
&\le \frac{2}{N}\sum_{j=1}^{D_m}\mathbb{E}(\varphi_j^2(\hat{\bar{V}}_1)Z_2^{(1)})^2) \le \frac{2D_m\mathbb{E}[(Z_2^{(1)})^2]}{N}.
\end{aligned}$$

Now, Lemma 6.2 implies that $\mathbb{E}[(u_{i+2,k} - u_{i+1,k})^2/\Delta^2 = \mathbb{E}[(u_{i+2,k}^2 + u_{i+1,k}^2)/\Delta^2 \le c/(k\Delta^2)$. Then, applying also Lemma 6.1 (ii), it follows that, with

$$\mathbb{E}\left(\sup_{t\in S_m, \|t\|=1}[\nu_N^{(1)}(t)]^2\right) \le K\frac{D_m}{N\Delta}\left(1+\frac{1}{k\Delta}\right).$$

Case $\ell = 2$. Next, for the martingale terms, we write

$$\begin{aligned}
\mathbb{E}(\sup_{t\in B_m^{\pi^*}(0,1)}[\nu_N^{(2)}(t)]^2) &\le \frac{1}{\pi_0^*}\mathbb{E}(\sup_{t\in B_m(0,1)}[\nu_N^{(2)}(t)]^2) \le \frac{1}{\pi_0^*}\sum_{j=1}^{D_m}\mathbb{E}([\nu_n^{(2)}(\varphi_j)]^2) \\
&= \frac{1}{\pi_0^*}\sum_{j=1}^{D_m}\mathbb{E}\left(\frac{1}{N}\sum_{i=0}^{N-1}\varphi_j(\hat{\bar{V}}_i)Z_{i+1}^{(2)}\right)^2 \\
&\le \frac{2}{\pi_0^*}\sum_{j=1}^{D_m}\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=0}^{N-1}\varphi_j(\hat{\bar{V}}_i)(Z_{i+1}^{(2,1)} + Z_{i+1}^{(2,2)})\right)^2\right. \\
&\qquad\left. + \left(\frac{9}{N\Delta}\sum_{i=0}^{N-1}\varphi_j(\hat{\bar{V}}_i)(\bar{V}_{i+2} - \bar{V}_i)(u_{i+2,k} - u_{i+1,k})\right)^2\right]
\end{aligned}$$

Both terms are bounded separately. For the first one, we use that, for $r = 1, 2$

$$\mathrm{cov}(\varphi_j(\hat{\bar{V}}_i)Z_{i+1}^{(2,r)}, \varphi_j(\hat{\bar{V}}_\ell)Z_{\ell+1}^{(2,r)}) = 0$$

if $\ell \geq i+2$, by inserting a conditional expectation with respect to $\mathcal{F}_{(\ell+1)\Delta}$. Now, for $r=1,2$,

$$\sum_{j=1}^{D_m}\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=0}^{N-1}\varphi_j(\hat{\bar{V}}_i)Z_{i+1}^{(2,r)}\right)^2\right] \leq \frac{1}{N^2}\sum_{j=1}^{D_m}\mathbb{E}\left(\sum_{0\leq i,\ell\leq N-1}\varphi_j(\hat{\bar{V}}_i)Z_{i+1}^{(2,r)}\varphi_j(\hat{\bar{V}}_\ell)Z_{\ell+1}^{(2,r)}\right)$$

$$= \frac{1}{N^2}\sum_{j=1} D_m\mathbb{E}\left\{\sum_{i=0}^{N-1}\left[\varphi_j^2(\hat{\bar{V}}_i)[Z_{i+1}^{(2,r)}]^2 + \varphi_j(\hat{\bar{V}}_i)Z_{i+1}^{(2,r)}\varphi_j(\hat{\bar{V}}_{i+1})Z_{i+2}^{(2,r)}\right]\right\}$$

$$\leq \frac{2}{N}\|\sum_{j=1}^{D_m}\varphi_j^2\|_\infty\mathbb{E}[(Z_2^{(2,r)})^2] \leq 2\frac{D_m}{N}[\tilde{c}_1\mathbb{E}(\sigma^4(V_0)) + \tilde{c}_2\Delta]$$

by using Lemma 6.1.

For the second part, let us define the filtration generated by B and the whole path of V, i.e.

$$\mathcal{G}_t^V = \sigma(V_s, s\in\mathbb{R}^+, B_s, s\leq t) = \sigma(W_s, s\in\mathbb{R}^+, B_s, s\leq t, \eta).$$

Now we observe that

$$\begin{aligned}\mathbb{E}(t(\hat{\bar{V}}_i)(\bar{V}_{i+2}-\bar{V}_{i+1})u_{i+1,k}) &= \mathbb{E}\left[\mathbb{E}(t(\hat{\bar{V}}_i)(\bar{V}_{i+2}-\bar{V}_{i+1})u_{i+1,k})|\mathcal{G}_{(i+1)\Delta}^V)\right]\\ &= \mathbb{E}\left[t(\hat{\bar{V}}_i)(\bar{V}_{i+2}-\bar{V}_{i+1})\mathbb{E}(u_{i+1,k})|\mathcal{G}_{(i+1)\Delta}^V)\right]\\ &= 0\end{aligned}$$

as $\mathbb{E}(u_{i+1,k})|\mathcal{G}_{(i+1)\Delta}^V) = 0$. Moreover for any $\ell > i$,

$$\mathbb{E}(t(\hat{\bar{V}}_i)(\bar{V}_{i+2}-\bar{V}_{i+1})u_{i+1,k}t(\hat{\bar{V}}_\ell)(\bar{V}_{\ell+2}-\bar{V}_{\ell+1})u_{\ell+1,k})) = 0$$

by inserting a conditional expectation with respect to $\mathcal{G}_{(\ell+1)\Delta}^V$. The last remark is that one can easily see that

$$\mathbb{E}[(\bar{V}_{i+1}-\bar{V}_i)^4] \leq \frac{1}{\Delta^4}\mathbb{E}\left[\left(\int_{(i+1)\Delta}^{(i+2)\Delta}(V_s - V_{s-\Delta})ds\right)^4\right] \leq C\Delta^2.$$

Now we have

$$\begin{aligned}\sum_{j=1}^{D_m}\mathbb{E}\left(\frac{1}{N\Delta}\sum_{i=0}^{N-1}\varphi_j(\hat{\bar{V}}_i)(\bar{V}_{i+2}-\bar{V}_i)u_{i+1,k}\right)^2 &= \frac{1}{N^2\Delta^2}\sum_{j=1}^{D_m}\sum_{i=0}^{N-1}\mathbb{E}\left(\varphi_j^2(\hat{\bar{V}}_i)(\bar{V}_{i+2}-\bar{V}_i)^2u_{i+1,k}^2\right)\\ &\leq \frac{D_m}{N\Delta^2}\mathbb{E}^{1/2}[(\bar{V}_2-\bar{V}_1)^4]\mathbb{E}^{1/2}[u_{2,k}^4]\\ &\leq C\frac{D_m}{N}\frac{1}{k\Delta}.\end{aligned}$$

The second part of this term can be treated in the same way, and it follows that if $1/k \leq \Delta$, then this term is less than $C'D_m/N$. □

6.6 Proof of Lemma 6.4.

Let us recall that we know from Comte *et al.* (2008) that

$$T_N^*(t) = \frac{1}{N}\sum_{i=0}^{N-1}(\sigma^2(V_{(i+1)\Delta} - \sigma^2(\bar{V}_i))t(\bar{V}_i)$$

is such that

$$\mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^*(t)]^2) \le C(D_m^2\Delta^2 + D_m^5\Delta^3).$$

Here, we write that $T_N^{(2)}(t) = T_N^{(2,1)}(t) + T_N^{(2,2)}(t) + T_N^{(2,3)}(t) + T_N^*(t)$ with

$$T_N^{(2,1)}(t) = \frac{1}{N}\sum_{i=0}^{N-1}[t(\hat{\bar{V}}_i)-t(\bar{V}_i)][\sigma^2(\hat{\bar{V}}_i)-\sigma^2(\bar{V}_i)],\ \ T_N^{(2,2)}(t) = \frac{1}{N}\sum_{i=0}^{N-1}t(\bar{V}_i)[\sigma^2(\hat{\bar{V}}_i)-\sigma^2(\bar{V}_i)],$$

$$T_N^{(2,3)}(t) = \frac{1}{N}\sum_{i=0}^{N-1}[t(\hat{\bar{V}}_i) - t(\bar{V}_i)][\sigma^2(\bar{V}_i) - \sigma^2(V_{(i+1)\Delta})].$$

We shall use the following decompositions obtained by the Taylor formula:

$$\sigma^2(\hat{\bar{V}}_i) - \sigma^2(\bar{V}_i) = (\hat{\bar{V}}_i - \bar{V}_i)(\sigma^2)'(\bar{V}_i) + R_i, t(\hat{\bar{V}}_i) - t(\bar{V}_i) = (\hat{\bar{V}}_i - \bar{V}_i)t'(\bar{V}_i) + S_i(t)$$

with $\mathbb{E}(R_i^2) \le C/k^2$ and $\mathbb{E}(R_i^4) \le C/k^4$ if (σ^2)" is bounded, and $\mathbb{E}\left(\sup_{t\in B_m(0,1)} S_i(t)^2\right) \le CD_m^5/k^2$, $\mathbb{E}^{1/2}\left(\sup_{t\in B_m(0,1)} S_i(t)^4\right) \le CD_m^5/k^2$ because $\|t"\|_\infty^2 \le CD_m^5\|t\|^2$. Now, the three terms can be studied as follows. First

$$\begin{aligned} T_N^{(2,1)}(t) &= \frac{1}{N}\sum_{i=0}^{N-1}(\hat{\bar{V}}_i - \bar{V}_i)^2(t')(\bar{V}_i)(\sigma^2)'(\bar{V}_i) + \frac{1}{N}\sum_{i=0}^{N-1}(\hat{\bar{V}}_i - \bar{V}_i)t'(\bar{V}_i)R_i \\ &\quad + \frac{1}{N}\sum_{i=0}^{N-1}(\hat{\bar{V}}_i - \bar{V}_i)(\sigma^2)'(\bar{V}_i)S_i(t) + \frac{1}{N}\sum_{i=0}^{N-1}R_iS_i(t) \\ &:= T_N^{(2,1,1)}(t) + T_N^{(2,1,2)}(t) + T_N^{(2,1,3)}(t) + T_N^{(2,1,4)}(t), \end{aligned}$$

and we bound each term successively. Clearly by Schwarz inequality applied to each term, we find,

$$\mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^{(2,1,1)}(t)]^2) \le C\mathbb{E}^{1/2}(\bar{V}_1^4)\frac{D_m^3}{k^2}$$

using that $\|t'\|_\infty^2 \le CD_m^3\|t\|^2$,

$$\mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^{(2,1,2)}(t)]^2) \le C\frac{D_m^3}{k^3},\quad \mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^{(2,1,3)}(t)]^2) \le C\mathbb{E}^{1/2}(\bar{V}_1^4)\frac{D_m^5}{k^3},$$

and

$$\mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^{(2,1,4)}(t)]^2) \le C\frac{D_m^5}{k^4}.$$

Therefore, if $1/k \le \Delta$, $\mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^{(2,1)}(t)]^2) \le C(D_m^3/k^2 + D_m^5/k^3)$.

Next, we write that

$$T_N^{(2,2)}(t) = \frac{1}{N}\sum_{i=0}^{N-1} t(\bar{V}_i)(\sigma^2)'(\bar{V}_i)(\hat{\bar{V}}_i) - \bar{V}_i) + \frac{1}{N}\sum_{i=0}^{N-1} t(\bar{V}_i)R_i$$
$$= T_N^{(2,2,1)}(t) + T_N^{(2,2,2)}(t).$$

We obtain easily that

$$\mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^{(2,2,2)}(t)]^2) \le \mathbb{E}(\sup_{t\in B_m(0,1)}\|t\|_\infty^2 \frac{1}{N}\sum_{i=1}^{N} R_i^2) \le cD_m\mathbb{E}(R_1^2) \le CD_m/k^2,$$

a term which is negligible with respect to the previous ones.

Then $(\hat{\bar{V}}_i - \bar{V}_i)\psi(\bar{V}_i)$ is a martingale increment with respect to the filtration $(\mathcal{G}_t^V)$, for any measurable function ψ. In particular,

$$\mathbb{E}[(\hat{\bar{V}}_i - \bar{V}_i)\psi(\bar{V}_i)] = \mathbb{E}[\mathbb{E}[(\hat{\bar{V}}_i - \bar{V}_i)\psi(\bar{V}_i)|\mathcal{G}_{i\Delta}^V]]$$
$$= \mathbb{E}[\psi(\bar{V}_i)\mathbb{E}[(\hat{\bar{V}}_i - \bar{V}_i)|\mathcal{G}_{i\Delta}^V]] = 0$$

since $\mathbb{E}(\hat{\bar{V}}_i|\mathcal{G}_{i\Delta}^V) = \bar{V}_i$. In the same way, for $i < \ell$,

$$\mathbb{E}\left((\hat{\bar{V}}_i - \bar{V}_i)\psi(\bar{V}_i)(\hat{\bar{V}}_\ell - \bar{V}_\ell)\psi(\bar{V}_\ell)\right) = 0$$

by inserting a conditional expectation with respect to $\mathcal{G}_{\ell\Delta}^V$. Therefore

$$\mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^{(2,2,1)}(t)]^2) \le \sum_{j=1}^{D_m}\mathbb{E}\left(\frac{1}{N}\sum_{i=0}^{N-1}\varphi_j(\bar{V}_i)(\sigma^2)'(\bar{V}_i)(\hat{\bar{V}}_i - \bar{V}_i)\right)^2$$
$$= \sum_{j=1}^{D_m}\frac{1}{N}\mathbb{E}\left(\varphi_j(\bar{V}_1)(\sigma^2)'(\bar{V}_1)(\hat{\bar{V}}_1 - \bar{V}_1\right)^2$$
$$\le \frac{1}{N}\mathbb{E}\left((\sum_{j=1}^{D_m}\varphi_j^2(\bar{V}_1))[(\sigma^2)'(\bar{V}_1)]^2(\hat{\bar{V}}_1 - \bar{V}_1)^2\right)$$
$$\le \frac{D_m}{N}\mathbb{E}^{1/2}[(\sigma^2)'(\bar{V}_1)^4]\mathbb{E}^{1/2}[u_{1,k}^4] \le C\mathbb{E}^{1/2}(\bar{V}_1^4)\frac{D_m}{Nk}.$$

For the last term, we write $T_N^{(2,3)}(t) = T_N^{(2,3,1)}(t) + T_N^{(2,3,2)}(t)$ where

$$T_N^{(2,3,1)}(t) = (1/N)\sum_{i=0}^{N-1}(\hat{\bar{V}}_i - \bar{V}_i)t'(\bar{V}_i)(\sigma^2(\bar{V}_i) - \sigma^2(V_{(i+1)\Delta})),$$

$$T_N^{(2,3,2)}(t) = (1/N)\sum_{i=0}^{N-1} S_i(t)(\sigma^2(\bar{V}_i) - \sigma^2(V_{(i+1)\Delta})).$$

Moreover, we know from Comte *et al.* (2008) that $\mathbb{E}[(\sigma^2(\bar V_i)-\sigma^2(V_{(i+1)\Delta}))^2] \le \mathbb{E}^{1/2}[(\sigma^2(\bar V_i)-\sigma^2(V_{(i+1)\Delta}))^4] \le C\Delta$. Now, for $T_N^{(2,3,1)}(t)$, we proceed as for $T_N^{(2,2,1)}(t)$ since both have the same martingale property w.r.t. $\mathcal{G}_s^V$. We get

$$\begin{aligned}
\mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^{(2,3,1)}(t)]^2) &\le \sum_{j=1}^{D_m}\mathbb{E}\left(\frac{1}{N}\sum_{i=0}^{N-1}\varphi_j'(\bar V_i)(\hat{\bar V}_i-\bar V_i)(\sigma^2(\bar V_i)-\sigma^2(V_{(i+1)\Delta}))\right)^2\\
&\le \frac{1}{N}\sum_{j=1}^{D_m}\mathbb{E}\left((\varphi_j')^2(\bar V_1)(\hat{\bar V}_1-\bar V_1)^2(\sigma^2(\bar V_1)-\sigma^2(V_{2\Delta}))^2\right)\\
&\le \frac{CD_m^3}{N}\mathbb{E}^{1/2}(u_{1,k}^4)\mathbb{E}^{1/2}[(\sigma^2(\bar V_1)-\sigma^2(V_{2\Delta}))^4]\\
&\le C\frac{D_m^3\Delta}{Nk}
\end{aligned}$$

as $\sum_j(\varphi_j')^2(x)\le CD_m^3$. Using $D_m^2\le N\Delta$ and $1/k\le\Delta$ implies $\mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^{(2,3,1)}(t)]^2)\le CD_m\Delta^3$. On the other hand, $\mathbb{E}(\sup_{t\in B_m(0,1)}[T_N^{(2,3,2)}(t)]^2)\le CD_m^5\Delta/k^2\le CD_m^5\Delta^3$, as $1/k\le\Delta$.

By gathering and comparing all terms and assuming that $1/k\le\Delta$, we obtain the bound given in Lemma 6.4.□

6.7 Proof of Theorem 4.1

The proof of this theorem relies on the following Bernstein-type Inequality:

Lemma 6.5 *Under the assumptions of Theorem 4.1, for any positive numbers ϵ and v, we have*

$$\mathbb{P}\left[\sum_{i=0}^{N-1}t(\hat{\bar V}_i)Z_{(i+1)\Delta}^{(1)}\ge N\epsilon,\|t\|_N^2\le v^2\right]\le\exp\left(-\frac{N\Delta\epsilon^2}{2\sigma_1^2v^2}\right).$$

Proof of Lemma 6.5: Noting that W is a Brownian motion with respect to the augmented filtration $\mathcal{F}_s=\sigma((B_u,W_u),u\le s,\eta)$, the proof is obtained as the analogous proof in Comte *et al.* (2007), Lemma 2 p.533. □

Now we turn to the proof of Theorem 4.1.
As in the proof of Proposition 3.2, we have to split $\|\tilde b-b_A\|_N^2=\|\tilde b-b_A\|_N^2\mathbf{1}_{\Omega_N}+\|\tilde b-b_A\|_N^2\mathbf{1}_{\Omega_N^c}$. For the study on Ω_N^c, the end of the proof of Proposition 3.2 can be used.

Now, we focus on what happens on Ω_N. For simplicity, we set $\hat m^{(1)}=\hat m$. From the definition of $\tilde b$, we have, $\forall m\in\mathcal{M}_n$, $\gamma_N(\hat b_{\hat m})+\text{pen}(\hat m)\le\gamma_N(b_m)+\text{pen}(m)$. We proceed as in the proof of Proposition 3.2 with some additional penalty terms and obtain

$$\begin{aligned}
\mathbb{E}(\|\hat b_{\hat m}-b_A\|_N^2\mathbf{1}_{\Omega_N}) \le 7\|b_m-b_A\|_{\pi^*}^2+\text{pen}(m)+32\mathbb{E}\left(\sup_{t\in S_m+S_{\hat m},\|t\|_{\pi^*}=1}[\nu_N^{(1)}(t)]^2\mathbf{1}_{\Omega_N}\right)\\
-\mathbb{E}(\text{pen}(\hat m))+32c'\Delta.
\end{aligned}$$

The difficulty here is to control the supremum of $\nu_N^{(1)}(t)$ on a random ball (which depends on the random $\hat{m}$). This is done by setting $\nu_N^{(1)} = \nu_N^{(1,1)} + \nu_N^{(1,2)}$, with

$$\nu_N^{(1,1)}(t) = \frac{1}{N}\sum_{i=0}^{N-1} Z_{(i+1)\Delta}^{(1)} t(\hat{\bar{V}}_i), \;\; \nu_N^{(1,2)}(t) = \frac{1}{N}\sum_{i=0}^{N-1} t(\hat{\bar{V}}_i)\left(\frac{u_{i+2,k}-u_{i+1,k}}{\Delta}\right).$$

We use the martingale property of $\nu_N^{(1,1)}(t)$ and a rough bound for $\nu_N^{(1,2)}(t)$ as follows.

For $\nu_N^{(1,2)}$, we simply write, as previously

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in S_m+S_{\hat{m}}, \|t\|_{\pi^*}=1} [\nu_n^{(1,2)}(t)]^2\right) &\le \frac{1}{\pi_0^*}\mathbb{E}\left(\sup_{t\in \mathcal{S}_n, \|t\|=1} [\nu_n^{(1,2)}(t)]^2\right)\\
&\le \frac{1}{\pi_0^*}\sum_{j=1}^{\mathcal{D}_n}\mathbb{E}[(\nu_N^{(2)}(\varphi_j))^2]\\
&\le \frac{4\mathcal{D}_n}{\pi_0^* N}\mathbb{E}[(u_{1,k}/\Delta)^2] \le \frac{4\mathcal{D}_n}{\pi_0^* N k_n \Delta^2} \le \frac{4}{\pi_0^*}\frac{1}{k_n\Delta}.
\end{aligned}$$

For $\nu_N^{(1,1)}$, let us denote by

$$G_m(m') = \sup_{t\in S_m+S_{m'}, \|t\|_{\pi^*}=1} \nu_N^{(1,1)}(t)$$

the quantity to be studied. Introducing a function $p(m,m')$, we first write

$$\begin{aligned}
G_m^2(\hat{m})\mathbf{1}_{\Omega_N} &\le [(G_m^2(\hat{m}) - p(m,\hat{m}))\mathbf{1}_{\Omega_N}]_+ + p(m,\hat{m})\\
&\le \sum_{m'\in\mathcal{M}_n} [(G_m^2(m') - p(m,m'))\mathbf{1}_{\Omega_N}]_+ + p(m,\hat{m}).
\end{aligned}$$

Then pen is chosen such that $32p(m,m') \le \text{pen}(m)+\text{pen}(m')$. More precisely, the next Proposition determines the choice of $p(m,m')$ which in turn will fix the penalty.

Proposition 6.1 *Under the assumptions of Theorem 4.1, there exists a numerical constant* κ_1 *such that, for* $p(m,m') = \kappa_1\sigma_1^2(D_m + D_{m'})/(n\Delta)$, *we have*

$$\mathbb{E}[(G_m^2(m') - p(m,m'))\mathbf{1}_{\Omega_N}]_+ \le c\sigma_1^2\frac{e^{-D_{m'}}}{N\Delta}.$$

Proof of Proposition 6.1. The result of Proposition 6.1 follows from the inequality of Lemma 6.5 by the $\mathbb{L}^2$-chaining technique used in Baraud et al. (2001b) (see Section 7 p.44-47, Lemma 7.1, with $s^2 = \sigma_1^2/\Delta$). □

It is easy to see that the result of Theorem 4.1 follows from Proposition 6.1 with $\text{pen}(m) = \kappa\sigma_1^2 D_m/(N\Delta)$. □

6.8 Proof of Theorem 4.2

The lines of the proof are the same as the ones of Theorem 4.1. Moreover, they follow closely the analogous proof of Theorem 2 p.524 in Comte *et al.* (2007), see also Comte *et al.* (2008). Therefore, we omit it.

References

1. Aït-Sahalia, Y., Kimmel, R.: Maximum likelihood estimation of stochastic volatility models. J. Finance Econ. **83**, 413-452 (2007).
2. Andersen, T., Lund, J.: Estimating continuous-time stochastic volatility models of the short-term interest rate. J. Econom. **77**, 343-377 (1997).
3. Barron, A.R., Birgé, L., Massart, P.: Risk bounds for model selection via penalization. Probab. Theory Related Fields **113**, 301–413 (1999).
4. Comte, F., Genon-Catalot, V., Rozenholc, Y.: Penalized nonparametric mean square estimation of the coefficients of diffusion processes. Bernoulli **13**, 514-543 (2007).
5. Comte, F., Genon-Catalot, V., Rozenholc, Y.: Nonparametric adaptive estimation for integrated diffusions. To appear in Stochastic Process. Appl. (2008).
6. Comte, F., Rozenholc, Y.: A new algorithm for fixed design regression and denoising. Ann. Inst. Statist. Math. **56**, 449-473 (2004).
7. Cox, J.C., Ingersoll, J.E., Jr., Ross, S.A.: A theory of the term structure of interest rates. Econometrica **53**, 385-407 (1985).
8. Genon-Catalot, V., Jeantheau, T., Larédo, C.: Parameter estimation for discretely observed stochastic volatility models. Bernoulli **5**, 855-872 (1999).
9. Genon-Catalot, V., Jeantheau, T., Larédo, C.: Stochastic volatility models as hidden Markov models and statistical applications. Bernoulli, **6**, 1051-1079 (2000).
10. Gloter, A.: Estimation of the volatility diffusion coefficient for a stochastic volatility model. (French) C. R. Acad. Sci. Paris, Sr. I Math. **330**, 3, 243-248 (2000).
11. Gloter, A.: Efficient estimation of drift parameters in stochastic volatility models. Finance Stoch., **11**, 495-519 (2007).
12. Heston, S.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financial Stud. **6**, 327-343 (1993).
13. Hoffmann, M.: Adaptive estimation in diffusion processes. Stochastic Process. Appl. **79**, 135-163 (1999).
14. Hoffmann, M.: Rate of convergence for parametric estimation in a stochastic volatility model. Stochastic Process. Appl. **97**, 147-170 (2002).
15. Renò, R.: Nonparametric estimation of stochastic volatility models. Economic Letters **90**, 390-395 (2006).
16. Rogers, L.C.G., Williams, D.: Diffusions, Markov processes, and martingales. Vol. 2, Reprint of the second (1994) edition, Cambridge Univ. Press, Cambridge (2000).
17. Zhang, L., Mykland, P.A., Aït-Sahalia, Y.: A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. of the American Statist. Assoc. **100**, 1394-1411 (2005).