NONPARAMETRIC ESTIMATION FOR A DISCRETELY OBSERVED INTEGRATED DIFFUSION MODEL.

F. COMTE[1], V. GENON-CATALOT[1], AND Y. ROZENHOLC[1]

ABSTRACT. We consider here nonparametric estimation for integrated diffusion processes. Let (V_t) be a stationary and β-mixing diffusion with unknown drift and diffusion coefficient. The integrated process $X_t = \int_0^t V_s ds$ is observed at discrete times with regular sampling interval Δ. For both the drift function and the diffusion coefficient of the unobserved diffusion (V_t), we propose nonparametric estimators based on a penalized least square approach. Estimators are chosen among a collection of functions belonging to a finite dimensional space selected by an automatic data-driven method. We derive non asymptotic risk bounds for the estimators. Interpreting these bounds through the asymptotic framework of high frequency data, we show that our estimators reach the minimax optimal rates of convergence. The algorithms of estimation are implemented for several examples of diffusion models that can be exactly simulated.

October 30, 2006

Keywords. Adaptive estimation. Diffusion process. Drift. Diffusion coefficient. Mean square estimator. Model selection. Integrated process. Discrete observation.

1. INTRODUCTION

In this paper, we consider the following two-dimensional process

$$\begin{aligned} dX_t &= V_t dt \quad X_0 = 0 \\ dV_t &= b(V_t)dt + \sigma(V_t)dW_t \;\; t \geq 0, \quad V_0 = \eta \end{aligned} \tag{1}$$

where (W_t) is a standard Brownian motion and η a real random variable independent of (W_t). This model is a special case of two-dimensional diffusion process without noise in the first equation. Our aim is to estimate the unknown functions b and σ^2 when only the first component (X_t) is observed at discrete equispaced times, $k\Delta$, $k = 1, \ldots, n+2$. Our estimation procedure will be based on the following equivalent set of data

$$\frac{1}{\Delta}(X_{(k+1)\Delta} - X_{k\Delta}) = \bar{V}_k = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} V_s ds, \; k \leq n+1. \tag{2}$$

Integrated diffusion processes are of common use for modelling purposes in the field of engineering and physics. For instance, (V_t) may represent the velocity of a particle and (X_t) its coordinate (see *e.g.* Rogers and Williams (1987, 114-115)). Other concrete examples where these processes are considered can be found in Lefebvre (1997) or in Ditlevsen and Sørensen (2004). It is worth noting that the component (X_t) provides a simple model for non Markovian observations or increasing observations when V_t is positive. Now, the most popular field of applications is certainly the field of finance with the stochastic volatility

[1] Université Paris V, MAP5, UMR CNRS 8145.

models. In this context, the model of interest $(\xi_t, V_t)_{t\geq 0}$ is a bivariate diffusion process, (V_t) is nonnegative, and the dynamics is described by the following equations:

$$\begin{cases} d\xi_t = \rho(\xi_t)dt + \sqrt{V_t}dB_t, \\ dV_t = b(V_t)dt + \sigma(V_t)dW_t \quad t \geq 0, \end{cases} \tag{3}$$

where (B_t, W_t) is a standard two-dimensional Brownian motion. The first component (ξ_t) describes the logarithm of a stock or asset price. It is observed while the volatility process (V_t) is unobserved. Practitioners generally approximate the integrated volatility by quadratic variations of (ξ_t) (realized volatility). Or, they derive the integrated volatility using option prices (implied volatility) (see *e.g.* Renault and Touzi (1996), Barndorff-Nielsen and Shephard (2002), Bollerslev and Zhou (2002)).

Statistical inference for discretely observed diffusion processes has been widely investigated recently (see *e.g.* Yoshida (1992), Kessler (1997), Genon-Catalot *et al.* (1999), Elerian *et al.* (2001), Bibby *et al.* (2002), Aït-Sahalia and Mykland (2004), Aït-Sahalia (2006), Beskos *et al.* (2006b)). For what concerns integrated diffusions, parametric frameworks have been considered. Ditlevsen and Sørensen (2004) use prediction-based estimating functions (see Sørensen (2000)) and special parametric models for the underlying diffusion. For general models, parametric inference for integrated diffusion processes has been extensively addressed by Gloter (2000, 2006) and Gloter and Gobet (2005). For ergodic underlying diffusion models, in the high frequency framework, Gloter (2006) introduces a general contrast function and proves the consistency and asymptotic normality of the resulting estimators of the parameters.

To our knowledge, nonparametric inference for these models has never been studied up to now. In contrast, nonparametric estimation of b and σ^2 when discrete observations $(V_{k\Delta})_{1\leq k\leq n}$ are available has been the subject of several contributions. In particular, in Hoffmann (1999), minimax rates of convergence are exhibited (over Besov smoothness classes) and adaptive estimators based on wavelet thresholding are built. These estimators achieve optimal rates of convergence (up to a logarithmic factor) but are difficult to implement in concrete. In a previous work (Comte *et al.* (2005)), we proposed nonparametric estimators based on a penalized mean square approach. These estimators have optimality properties and can be implemented through feasible algorithms. In the present paper, we use analogous tools to build nonparametric estimators of b and σ^2 based on the observations (2). The process given by (1) is supposed to be strictly stationary and β-mixing. Relying on regression-type equations for the drift and for the diffusion coefficient, we build mean-square contrasts. These allow to construct a collection of estimators belonging to finite dimensional spaces including piecewise polynomials spaces. Model selection techniques using penalization devices enable us to exhibit a data-driven choice of the estimator among the collection. As it is usual with these methods, the risk of an estimator $\hat{f}$ of $f = b, \sigma^2$ is measured by the expectation of an empirical norm $\mathbb{E}(\|\hat{f} - f\|_n^2)$ where $\|\hat{f} - f\|_n^2 = \frac{1}{n}\sum_{k=1}^n (\hat{f}(\bar{V}_k) - f(\bar{V}_k))^2$. We obtain bounds for the risks which are non asymptotic in the sense that they are expressed as functions of n, Δ and constants. Interpreting these bounds when n tends to infinity while $\Delta = \Delta_n$ tends to 0, we prove that our estimators achieve the minimax optimal rates under some constraints on the rate of Δ_n, up to logarihtmic factors in some cases for σ^2. The optimality is evaluated in comparison with Hoffmann's results.

The paper is organized as follows. The model, assumptions and finite dimensional spaces on which estimators are built are described in Section 2. The spaces of approximation include piecewise polynomials on irregular partitions of the interval where the unknown functions are estimated. Sections 3 and 4 concern respectively the drift and the diffusion coefficients. The first step is to establish the regression-type equations which are the basement of the estimation method. Then, we present the penalized mean square contrasts allowing the automatic selection of the best adaptive estimators and state the risk bounds. For estimating the drift function, the regression-type equation has the form

$$Y_{k+1} := \frac{\bar{V}_{k+2} - \bar{V}_{k+1}}{\Delta} = b(\bar{V}_k) + \text{ noise} + \text{remainder},$$

where the lag of order 2Δ avoids cumbersome correlations due to integrated data. For estimating σ^2, the regression-type equation has the form

$$\bar{U}_{k+1} := \frac{3}{2\Delta}(\bar{V}_{k+2} - \bar{V}_{k+1})^2 = \sigma^2(\bar{V}_k) + \text{ noise} + \text{remainder}.$$

The correcting factor $3/2$ is specific to integrated observations and appears also in Gloter (2000). The study of the remainder term (see Proposition 4.4) is surprisingly difficult and induces constraints on the bases and on the sampling interval Δ which must be small enough. One assumption ([A6]) is especially discussed and illustrated in Section 5. Section 6 presents some simulation results illustrated by plots and tables. Proofs are gathered in Sections 7 for the drift, 8 for the diffusion coefficient, 9 for the results of Section 5. Lastly a technical proof is given in the appendix.

2. The assumptions

2.1. **Model assumptions.** Let $(V_t)_{t\geq 0}$ be given by (1) and assume that only integrals $(\bar{V}_k)_{1\leq k\leq n+1}$ given by (2) are observed. We want to estimate the drift function b and the square of the diffusion coefficient σ^2 when V is stationary and geometrically β-mixing. We assume that the state space of (V_t) is a known open interval (r_0, r_1) of the real line and consider the following set of assumptions.

[A1] $-\infty \leq r_0 < r_1 \leq +\infty$, $\overset{\circ}{I} = (r_0, r_1)$, b and σ belong to $C^1(\overset{\circ}{I})$, with $\sigma(v) > 0$, for all $v \in \overset{\circ}{I}$.

[A2] For all $v_0, v \in \overset{\circ}{I}$,
(i) the scale density

$$s(v) = \exp\left[-2\int_{v_0}^{v} \frac{b(u)}{\sigma^2(u)}du\right] \tag{4}$$

satisfies $\int_{r_0} s(x)dx = +\infty = \int^{r_1} s(x)dx$,
(ii) the speed density $m(v) = 1/(\sigma^2(v)s(v))$ satisfies $\int_{r_0}^{r_1} m(v)dv = M < +\infty$.

When the initial random variable satisfies $\mathbb{P}(\eta \in \overset{\circ}{I}) = 1$, Assumption [A1] implies the existence and unicity of the solution of (1) until a possible explosion time at r_0 or r_1. Assumption [A2] implies that the process never reaches r_0 nor r_1, is positive recurrent on $\overset{\circ}{I}$ and that $d\pi(v) = (m(v)/M)\mathbb{I}_{(r_0,r_1)}(v)dv$ is the unique stationary density. We assume moreover that

[A3] $\eta \sim \pi$ and $\mathbb{E}(\eta^{12}) < \infty$.

Under [A1]-[A3], (V_t) is strictly stationary, 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 *e.g.* Genon-Catalot *et al*, 2000 for a review). We need in fact a stronger mixing condition which is satisfied in most standard examples:

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

Lastly, we strengthen Assumption [A1] as follows in order to deal altogether with finite or infinite boundaries (see *e.g.* Ethier and Kurtz (1986, chap.8)):

[A5] (i) Let $I=[r_0,r_1]\cap\mathbb{R}$. Suppose $b\in C^1(I)$, b' bounded on I, $\sigma^2\in C^2(I)$, $(\sigma^2)''$bounded on I, $\sigma^2(r_i)=0\leq(-1)^ib(r_i)$ if $r_i\in\mathbb{R}$, $i=0,1$,
(ii) $\sigma^2(v)\leq\sigma_1^2$ for all v in I.

Assumption [A5](i) immediately implies that, for some positive constant K, for all v,v' in I,

$$|b(v)|\leq K(1+|v|),\ \ \sigma^2(v)\leq K(1+v^2),|b(v)-b(v')|\leq K|v-v'|. \tag{5}$$

The functions b and σ^2 are estimated only on a compact subset A of the state space $\overset{\circ}{I}$. For simplicity and without loss of generality, we assume from now on that

$$A=[0,1], \tag{6}$$

and set

$$b_A=b1_A,\quad \sigma_A=\sigma 1_A. \tag{7}$$

Under [A1]-[A4], for fixed Δ, $(\bar V_k)_{k\geq0}$ is a strictly stationary process. Since its β-mixing coefficients $\beta_{\bar V}(k)$ satisfy $\beta_{\bar V}(k)\leq\beta_V(k\Delta)$, $\bar V_k$ is geometrically β-mixing.

It follows from [A1]-[A3] that the stationary density π of (V_t) is bounded from below and above on any compact subset of $\overset{\circ}{I}$. We need the analogous property for the marginal density of the stationary process $(\bar V_k)_{k\geq0}$ and state it as an additional assumption:

[A6] The process $(\bar V_k)_{k\geq0}$ admits a stationary density $\bar\pi_\Delta$ and there exist two positive numbers $\bar\pi_0$ and $\bar\pi_1$ (independent of Δ) such that

$$0<\bar\pi_0\leq\bar\pi_\Delta(x)\leq\bar\pi_1,\ \ \forall x\in[0,1]. \tag{8}$$

The existence of a density for $\bar V_k$ is obtained under mild regularity conditions on b and σ (see *e.g.* Rogers and Williams (2000) or Comte and Genon-Catalot (2006)). In Section 5, sufficient conditions ensuring (8) are given together with some examples for which exact computations can be done. Assumption [A6] associated with [A4] is used in the proofs of Theorem 3.1 and 4.1 to obtain the risk bounds.

Below, we use the following notations:

$$\|t\|_\pi^2=\int t^2(x)\pi(x)dx=\mathbb{E}(t^2(V_0))\text{ and }\|t\|_{\bar\pi}^2=\int t^2(x)\bar\pi_\Delta(x)dx=\mathbb{E}(t^2(\bar V_0)). \tag{9}$$

2.2. **Spaces of approximation.** We aim at estimating functions b and σ^2 of Model (3) on $[0,1]$ using a data driven procedure. For that purpose, we consider families of finite dimensional linear subspaces of $\mathbb{L}^2([0,1])$ and compute for each space an associated least-squares estimator. Afterwards, an adaptive procedure chooses among the resulting collection of estimators the "best" one, in a sense that will be later specified, through a penalization device.

Let us describe now the collection of spaces that are considered below.
We start by describing the collection of dyadic regular piecewise polynomial spaces with constant degree, denoted hereafter by [DP]. We fix an integer $r \geq 0$. Let $p \geq 0$ an integer. On each subinterval $I_j = [(j-1)/2^p, j/2^p]$, $j = 1, \ldots, 2^p$, consider $r+1$ polynomials of degree $0, 1, \ldots, r$, $\varphi_{j,\ell}(x)$, $\ell = 0, 1, \ldots r$ and set $\varphi_{j,\ell}(x) = 0$ outside I_j. The space S_m, $m = (p, r)$, is defined as generated by the $D_m = 2^p(r+1)$ functions $(\varphi_{j,\ell})$. A function t in S_m may be written as

$$t(x) = \sum_{j=1}^{2^p} \sum_{\ell=0}^{r} t_{j,\ell} \varphi_{j,\ell}(x).$$

The collection [DP] is composed of the spaces $(S_m, m \in \mathcal{M}_n)$ where

$$\mathcal{M}_n = \{m = (p, r), p \in \mathbb{N}, 2^p(r+1) \leq N_n\}. \tag{10}$$

In other words, $D_m \leq N_n$ and $N_n \leq n$. We denote by $\mathcal{S}_n$ the largest space of this collection of nested spaces and set $\dim \mathcal{S}_n = N_n$. The maximal dimension N_n is subject to additional constraints given below.

To be more concrete, consider the orthogonal collection in $\mathbb{L}^2([-1,1])$ of Legendre polynomials $(Q_\ell, \ell \geq 0)$, where the degree of Q_ℓ is equal to ℓ, generating $\mathbb{L}^2([-1,1])$ (see Abramowitz and Stegun (1972), p.774). They satisfy $|Q_\ell(x)| \leq 1, \forall x \in [-1,1]$, $Q_\ell(1) = 1$ and $\int_{-1}^{1} Q_\ell^2(u)du = 2/(2\ell+1)$. Let us set $P_\ell(x) = \sqrt{2\ell+1} Q_\ell(2x-1)$ to get an orthonormal basis of $\mathbb{L}^2([0,1])$. And finally,

$$\varphi_{j,\ell}(x) = 2^{p/2} P_\ell(2^p x - j + 1) \mathbb{1}_{I_j}(x), \quad j = 1, \ldots, 2^p, \ \ell = 0, 1, \ldots, r.$$

The space S_m has dimension $D_m = 2^p(r+1)$. Its orthonormal basis described above satisfies

$$\left\| \sum_{j=1}^{2^p} \sum_{\ell=0}^{r} \varphi_{j,\ell}^2 \right\|_\infty \leq D_m(r+1). \tag{11}$$

Hence, for all $t \in S_m$, $\|t\|_\infty \leq \sqrt{r+1}\sqrt{D_m}\|t\|$, where

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

This connection property between the sup-norm and the $\mathbb{L}^2$-norm for functions in S_m is essential for the proofs. The order $\sqrt{D_m}$ is specific to the case of regular subdivisions of $[0,1]$.

A more general family can be described, the collection of general piecewise polynomials spaces denoted by [GP]. We first build the largest space $\mathcal{S}_n$ of the collection whose dimension is denoted as above by N_n ($N_n \leq n$ and is subject to other constraints appearing later on). For this, we fix an integer R_{max} and let D_{max} be an integer such that

$D_{max}(R_{max}+1) = N_n$. The space $\mathcal{S}_n$ is linearly spanned by piecewise polynomials of degree R_{max} on the regular subdivision of $[0,1]$ with step $1/D_{max}$. Any other space S_m of the collection is described by a multi-index $m = (d, j_1, \ldots, j_{d-1}, r_1, \ldots, r_d)$ where d is the number of intervals of the partition, $j_0 := 0 < j_1 < \cdots < j_{d-1} < j_d := 1$ are integers such that $j_i \in \{1, \ldots, D_{max} - 1\}$ for $i = 1, \ldots d-1$. The latter integers define the knots j_i/D_{max} of the subdivision. Lastly $r_i \leq R_{max}$ is the degree of the polynomial on the interval $[j_{i-1}/D_{max}, j_i/D_{max}[$, for $i = 1, \ldots, d$. A function t in S_m can thus be described as

$$t(x) = \sum_{i=1}^{d} P_i(x) \mathbb{I}_{[j_{i-1}/D_{max}, j_i/D_{max})}(x),$$

with P_i a polynomial of degree r_i. The dimension of S_m is still denoted by D_m and equals $\sum_{i=1}^{d}(r_i+1)$ for all the $\binom{D_{max}-1}{d-1}$ choices of the knots $(j_1, \ldots, j_{d-1})$. Note that the P_i's can still be decomposed by using the Legendre basis rescaled on the intervals $[j_{i-1}/D_{max}, j_i/D_{max})$.

It is easy to see that now, for $t \in S_m \subset \mathcal{S}_n$,

$$\|t\|_\infty \leq \sqrt{(R_{max}+1)N_n}\|t\|.$$

The collection [GP] of models $(S_m)_{m \in \mathcal{M}_n}$ is described by the set of indexes

$$\mathcal{M}_n = \{m = (d, j_1, \ldots, j_{d-1}, r_1, \ldots, r_d),\ 1 \leq d \leq D_{max}, j_i \in \{1, \ldots, D_{max}-1\}, \\ r_i \in \{0, \ldots, R_{max}\}\}.$$

Obviously, collection [GP] has higher complexity than [DP]. The complexity of a collection is usually evaluated through a set of weights (L_m) that must satisfy $\sum_{m \in \mathcal{M}_n} e^{-L_m D_m} < \infty$. For [DP], it is easy to see that $L_m = 1$ suits. For [GP], we have to look at

$$\sum_{m \in \mathcal{M}_n} e^{-L_m D_m} = \sum_{d=1}^{D_{max}} \sum_{1 \leq j_1 < \cdots < j_{d-1} < D_{max}} \sum_{0 \leq r_1, \ldots, r_d \leq R_{max}} e^{-L_m \sum_{i=1}^{d}(r_i+1)}$$

From the equality above, we deduce that the choice

$$L_m D_m = D_m + \ln \begin{pmatrix} D_{max}-1 \\ d-1 \end{pmatrix} + d \ln(R_{max}+1) \tag{12}$$

can suit. Actually, it is the term inspiring the penalty function used in the practical implementation. To see more clearly what orders of magnitude are involved, let us set $L_m = L_n$ for all $m \in \mathcal{M}_n$. Then, we have a further bound for the series:

$$\begin{aligned}
\sum_{m \in \mathcal{M}_n} e^{-L_m D_m} &\leq \sum_{d=1}^{D_{max}} \begin{pmatrix} D_{max}-1 \\ d-1 \end{pmatrix} (R_{max}+1)^d e^{-dL_n} \\
&\leq \sum_{d=0}^{D_{max}-1} \begin{pmatrix} D_{max}-1 \\ d \end{pmatrix} [(R_{max}+1)e^{-L_n}]^{d+1} \\
&\leq (R_{max}+1)\left[1 + (R_{max}+1)e^{-L_n}\right]^{D_{max}-1} \\
&\leq (R_{max}+1)\exp(D_{max}(R_{max}+1)e^{-L_n}) \leq (R_{max}+1)\exp(N_n e^{-L_n}).
\end{aligned}$$

Thus $L_m = L_n = \ln(N_n)$ ensures that the series is bounded. (For more details on these collections, see *e.g.* Comte and Rozenholc (2004) or Baraud *et al* (2001b)).

Other spaces of approximation can be considered as, for example:
[T] *Trigonometric spaces*: S_m is generated by $\{\ 1,\ \sqrt{2}\cos(2\pi jx), \sqrt{2}\sin(2\pi jx)$ for $j = 1,\dots,m\ \}$, has dimension $D_m = 2m+1$ and $m \in \mathcal{M}_n = \{1,\dots,[n/2]-1\}$ with $D_m \leq N_n$.
[W] *Dyadic wavelet generated spaces* with smoothness $r \geq 2$ and compact support, as described e.g. in Cohen *et al.* (1993), Donoho et al. (1996) or Hoffmann (1999). The spaces are also denoted by S_m, with $\dim(S_m) = D_m \leq N_n$.

In both cases, the maximal dimension N_n is subject additional constraints (see below). The drawback of these spaces is their lack of flexibility. In particular, the notion of regular or irregular partitions has no sense for trigonometric bases. For what concerns wavelet bases, they are systematically built on dyadic partitions. On the other hand, the interest of these spaces is that they are generated by smooth functions contrary to piecewise polynomials. For the estimation of the diffusion coefficient, smooth bases are needed to recover the optimal nonparametric rate of convergence.

Below, we keep general notations for the spaces of approximation: an orthonormal basis of a space S_m will be denoted by $(\varphi_\lambda)_{\lambda\in\Lambda_m}$ where $|\Lambda_m| = D_m$.

3. Adaptive estimation of the drift

3.1. Estimator of the drift.

Let

$$Y_k = \frac{\bar{V}_{k+1} - \bar{V}_k}{\Delta}. \tag{13}$$

The following regression-type decomposition holds:

$$Y_{k+1} = b(\bar{V}_k) + Z_{(k+1)\Delta} + R_b((k+1)\Delta) \tag{14}$$

where $Z_{k\Delta}$ is a noise term given by

$$Z_{k\Delta} = \frac{1}{\Delta^2}\left[\int_{k\Delta}^{(k+2)\Delta} \psi_{k\Delta}(u)\sigma(V_u)dW_u\right] \tag{15}$$

with

$$\psi_{k\Delta}(u) = (u-k\Delta)\mathbb{I}_{[k\Delta,(k+1)\Delta[}(u) + [(k+2)\Delta - u]\mathbb{I}_{[(k+1)\Delta,(k+2)\Delta[}(u). \tag{16}$$

Note that, using the strict stationarity of (V_t),

$$\mathbb{E}(Z_{k\Delta}^2) = \frac{2}{3\Delta}\mathbb{E}\sigma^2(V_0). \tag{17}$$

This explains the correcting factor 3/2 appearing in (29) below. As a consequence of Proposition 3.1 below, the last term in (14) is negligible when Δ is small (see Section 7 for proofs).

Proposition 3.1. *Under Assumptions [A1]-[A2]-[A3],* $\mathbb{E}(R_b^2((k+1)\Delta) \leq c\Delta$ *and* $\mathbb{E}(R_b^4((k+1)\Delta) \leq c'\Delta^2$ *where* c *and* c' *neither depend on* k *nor on* Δ.

In light of decomposition (14), for S_m a space of the collection $\mathcal{M}_n$ and for $t \in S_m$, we consider the following regression contrast:

$$\gamma_n(t) = \frac{1}{n}\sum_{k=1}^{n}[Y_{k+1} - t(\bar{V}_k)]^2. \tag{18}$$

If we denote by

$$\mathcal{F}_t = \sigma\left(V_s, s \leq t\right),$$

it must be noticed that $Y_{k+1}, Z_{(k+1)\Delta}, R_b((k+1)\Delta)$ are $\mathcal{F}_{(k+3)\Delta}$-measurable whereas $\bar{V}_k$ is $\mathcal{F}_{(k+1)\Delta}$-measurable. This lag of order 2Δ avoids dealing with unnecessary and tedious correlations.

In a first step, the estimator belonging to S_m is defined as

$$\hat{b}_m = \arg\min_{t\in S_m} \gamma_n(t). \tag{19}$$

The second step is to ensure an automatic selection of the space S_m, which does not use any knowledge on b. This selection is standardly done by

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n}\left[\gamma_n(\hat{b}_m) + \mathrm{pen}(m)\right], \tag{20}$$

with $\mathrm{pen}(m)$ a penalty to be properly chosen. We denote by $\tilde{b} = \hat{b}_{\hat{m}}$ the resulting estimator.

Remark 3.1. *It is worth noting that in (19), $\hat{b}_m$ exists but may be non unique. Indeed minimizing γ_n over S_m often leads to an affine space of solutions. In contrast, the random $\mathbb{R}^n$-vector $(\hat{b}_m(\bar{V}_1), \ldots, \hat{b}_m(\bar{V}_n))'$ 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 $\mathbb{R}^n$, $\{(t(\bar{V}_1), \ldots, t(\bar{V}_n))', t \in S_m\}$. Then, we have $(\hat{b}_m(\bar{V}_1), \ldots, \hat{b}_m(\bar{V}_n))' = \Pi_m Y$ where $Y = (Y_2, \ldots, Y_{n+1})'$. This is the reason why we need consider a risk fitted to our problem.*

Let us define the empirical norm of a function t in some S_m by

$$\|t\|_n^2 = \frac{1}{n}\sum_{k=1}^{n} t^2(\bar{V}_k). \tag{21}$$

The risk of an estimator $\hat{b}_m$ is computed as the expectation of this empirical norm: $\mathbb{E}(\|\hat{b}_m - b\|_n^2)$.

Note that for a deterministic function $\mathbb{E}(\|t\|_n^2) = \|t\|_{\bar{\pi}}^2 = \int t^2(x)\bar{\pi}_\Delta(x)dx$ and that, under Assumption [A6], the norms $\|.\|$ and $\|.\|_{\bar{\pi}}$ are equivalent for $[0,1]$-supported functions.

3.2. **Risk of the drift estimator.** The regression contrast (18) may be written as:

$$\gamma_n(t) - \gamma_n(b) \quad = \quad \|t-b\|_n^2 - \frac{2}{n}\sum_{k=1}^{n}(Y_{k+1} - b(\bar{V}_k))(t-b)(\bar{V}_k)$$

In view of (14), let us introduce the two processes indexed by functions t:

$$\nu_n(t) = \frac{1}{n}\sum_{k=1}^{n} t(\bar{V}_k)Z_{(k+1)\Delta} \text{ and } R_n(t) = \frac{1}{n}\sum_{k=1}^{n} t(\bar{V}_k)R_b((k+1)\Delta). \tag{22}$$

Using the above notations, we obtain that

$$\gamma_n(t) - \gamma_n(b) \quad = \quad \|t-b\|_n^2 - 2\nu_n(t-b) - 2R_n(t-b).$$

Let b_m be the orthogonal projection of b_A on S_m. By (19), $\gamma_n(\hat{b}_m) \leq \gamma_n(b_m)$, hence $\gamma_n(\hat{b}_m) - \gamma_n(b) \leq \gamma_n(b_m) - \gamma_n(b)$. This implies:

$$\|\hat{b}_m - b\|_n^2 \quad \leq \quad \|b_m - b\|_n^2 + 2\nu_n(\hat{b}_m - b_m) + 2R_n(\hat{b}_m - b_m). \tag{23}$$

Since b_m and $\hat{b}_m$ are A-supported, $\|b1_{A^c}\|_n$ appears in both sides of the inequality. We can cancel it and obtain

$$\|\hat{b}_m - b_A\|_n^2 \quad \leq \quad \|b_m - b_A\|_n^2 + 2\nu_n(\hat{b}_m - b_m) + 2R_n(\hat{b}_m - b_m). \tag{24}$$

The last term, involving the residual R_b, can be controlled thanks to Proposition 3.1. And the process ν_n defined in (22) satisfies:

Proposition 3.2. *Consider S_m in collection* [DP], [T] *or* [W]. *Under Assumptions* [A1]-[A3], *for any Δ, $0 < \Delta \leq 1$,*

$$\mathbb{E}\left(\sup_{t\in S_m, \|t\|=1} \nu_n^2(t)\right) \leq c\frac{\mathbb{E}(\sigma^2(V_0))D_m}{n\Delta}.$$

For S_m in [GP], *under* [A1]-[A3]-[A5]-[A6], *for any Δ, $0 < \Delta \leq 1$,*

$$\mathbb{E}\left(\sup_{t\in S_m, \|t\|=1} \nu_n^2(t)\right) \leq c'\frac{\sigma_1^2 D_m}{n\Delta}.$$

Remark 3.2. *Propositions 3.1, 3.2 and inequality (24) are the keys to bound the risk for one estimator $\hat{b}_m$ of b belonging to a space S_m. Indeed, assume that, as n tends to infinity, $\Delta = \Delta_n$ is such that $\Delta_n \to 0$, $n\Delta_n/\ln^2(n) \to +\infty$. Under our set of assumptions, it is possible to prove that (see (7):*

$$\mathbb{E}(\|\hat{b}_m - b_A\|_n^2) \leq 7\bar{\pi}_1\|b_m - b_A\|^2 + K\frac{\mathbb{E}(\sigma^2(V_0))D_m}{n\Delta} + K'\Delta, \tag{25}$$

where K and K' are positive constants. Equation (25) holds if the maximal dimension N_n satisfies $N_n = o(n\Delta_n/\ln^2(n))$ for collections [DP] *and* [W]. *For collection* [T], *the constraint is $N_n = o(\sqrt{n\Delta_n}/\ln(n))$.*

Note that, under the standard condition $n\Delta^2 = O(1)$, the term $K'\Delta$ is negligible with respect to the previous one.

Moreover the result is easy to extend to collection [GP] provided that $\mathbb{E}(\sigma^2(V_0))$ is replaced by σ_1^2 in (25). Since Theorem 3.1 below mainly contains this result, we do not give the proof of (25).

To obtain results on the adaptive estimator, more accurate considerations on the martingale properties of ν_n must be driven. In particular, we prove the following Bernstein-type inequality:

Proposition 3.3. *Under Assumptions* [A1]-[A2]-[A3]-[A5],*for any positive numbers ϵ and v and for any function t in a space S_m, we have (see (2)-(15)-(21)-(22))*

$$\mathbb{P}\left[\nu_n(t) \geq \epsilon, \|t\|_n^2 \leq v^2\right] \leq \exp\left(-\frac{n\Delta\epsilon^2}{4\sigma_1^2 v^2}\right).$$

Proposition 3.3 enables us to obtain the adequate penalty function for (20), that leads to selecting the dimension D_m realizing the best compromise between the squared bias term $\|b_m - b_A\|^2$ and the variance term of order $D_m/(n\Delta)$ (see (25)).

Theorem 3.1. *Let $\Delta = \Delta_n$ be such that $\Delta_n \to 0$, $n\Delta_n/\ln^2(n) \to +\infty$ when $n \to +\infty$. Assume that* [A1]-[A6] *hold. Consider the nested collection of models [DP] (with $L_m = 1$) or the collection [GP] (with L_m given by (12)), both with maximal dimension satisfying $N_n = o(n\Delta/\ln^2(n))$. Then the estimator $\tilde{b} = \hat{b}_{\hat{m}}$ of b with $\hat{m}$ defined by (20) and*

$$\text{pen}(m) \geq \kappa\sigma_1^2 \frac{(1+L_m)D_m}{n\Delta}, \tag{26}$$

where κ is a universal constant, is such that

$$\mathbb{E}(\|\tilde{b} - b_A\|_n^2) \leq C \inf_{m\in\mathcal{M}_n} \big(\|b_m - b_A\|^2 + \text{pen}(m)\big) + K'\Delta + \frac{K"}{n\Delta}. \tag{27}$$

Inequality (27) holds for the basis [W], under the same assumptions, with $L_m = 1$. For [T] the additional constraint $N_n = o(\sqrt{n\Delta}/\ln(n))$ is required (with still $L_m = 1$).

Let us make some comments on Theorem 3.1. The constant κ in (26) is a numerical value that has to be calibrated by simulations (see Section 6.2). One would expect from (25) to obtain $\mathbb{E}(\sigma^2(V_0))$ instead of σ_1^2 in (26). We do not know if this is the consequence of technical problems or if this is a structural result. In practice, this term is replaced by an estimator (see Section 6.2). Inequality (27) enlights the fact that the adaptive estimator automatically realizes the bias-variance compromise in a non asymptotic way.

Let us look at rates of convergence using the asymptotic point of view. Assume that b_A belongs to a ball of some Besov space, $b_A \in \mathcal{B}_{\alpha,2,\infty}([0,1])$. Consider for instance collection [DP] with $r+1 \geq \alpha$ and weights $L_m = 1$ (see (10)) . Then $\|b_A - b_m\|^2 \leq C(\alpha, L)D_m^{-2\alpha}$, for $\|b_A\|_{\alpha,2,\infty} \leq L$ (see DeVore and Lorentz (1993) p.359 or Lemma 12 in Barron *et al.* (1999)). Therefore, if we search the dimension D_m that achieves $\inf\{D_m^{-2\alpha} + D_m/(n\Delta)\}$, we get $D_m \propto (n\Delta)^{1/(2\alpha+1)}$. Thus, we find

$$\mathbb{E}(\|\tilde{b} - b_A\|_n^2) \leq C(n\Delta)^{-2\alpha/(2\alpha+1)} + K'\Delta + \frac{K"}{n\Delta}. \tag{28}$$

The first term $(n\Delta)^{-2\alpha/(2\alpha+1)}$ is the optimal nonparametric rate proved by Hoffmann (1999) for a direct observation of V. Moreover, under the standard condition $\Delta = o(1/(n\Delta))$, the last two terms are negligible with respect to $(n\Delta)^{-2\alpha/(2\alpha+1)}$. Hence, even though V is not directly observed, the estimator $\tilde{b}$ reaches the optimal rate.

4. Adaptive estimation of the diffusion coefficient

4.1. **Estimator of the volatility.** Let us define

$$U_k = \frac{3}{2}\frac{(\bar{V}_{k+1} - \bar{V}_k)^2}{\Delta}. \tag{29}$$

The correcting factor 3/2, linked with integrated observations, is not surprising since it also appears in the parametric framework (see Gloter (2000, 2006)). Applications of Ito's

formula and Fubini's theorem yield the following regression-type decomposition:

$$\begin{aligned} U_{k+1} &= \sigma^2(V_{(k+1)\Delta}) + \breve{Z}_{(k+1)\Delta} + \breve{R}_{(k+1)\Delta} \\ &= \sigma^2(\bar{V}_k) + \breve{Z}_{(k+1)\Delta} + \breve{R}_{(k+1)\Delta} + [\sigma^2(V_{(k+1)\Delta}) - \sigma^2(\bar{V}_k)]. \end{aligned}$$

where $\breve{Z}_{k\Delta} = \breve{Z}^{(1)}_{k\Delta} + \breve{Z}^{(2)}_{k\Delta} + \breve{Z}^{(3)}_{k\Delta}$. The main component of this noise term is (see (16))

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

The two other components have negligible variance weight:

$$\begin{aligned} \breve{Z}^{(2)}_{k\Delta} &= \frac{3}{\Delta} b(V_{k\Delta}) \int_{k\Delta}^{(k+2)\Delta} \psi_{k\Delta}(s)\sigma(V_s)dW_s, \\ \breve{Z}^{(3)}_{k\Delta} &= \frac{3}{2\Delta^3} \int_{k\Delta}^{(k+2)\Delta} \left(\int_s^{(k+2)\Delta} \psi^2_{k\Delta}(u)du\right)(\sigma'\sigma^2)(V_s)dW_s. \end{aligned}$$

On the other hand, $\breve{R}_{(k+1)\Delta}$ is a residual term, as well as $\sigma^2(V_{(k+1)\Delta}) - \sigma^2(\bar{V}_k)$. The latter term raises specific problem because the rates for the estimation of σ^2 are faster than the rates for the estimation of b. Proposition 4.1 and 4.2 below rely on standard tools.

Proposition 4.1. *Under Assumptions [A1]-[A2]-[A3] and [A5],* $\mathbb{E}([\breve{Z}^{(1)}((k+1)\Delta)]^2)] \leq c_1\mathbb{E}(\sigma^4(V_0))$ *and for* $i = 2, 3$, $\mathbb{E}([\breve{Z}^{(i)}((k+1)\Delta)]^2 \leq c_i\Delta$, *where the* c_i*'s neither depend on* k *nor on* Δ.

Proposition 4.2. *Under Assumptions [A1]-[A2]-[A3]-[A5],* $\mathbb{E}(\breve{R}^2((k+1)\Delta) \leq c\Delta^2$ *and* $\mathbb{E}(\breve{R}^4((k+1)\Delta) \leq c'\Delta^4$ *where* c *and* c' *neither depend on* k *nor on* Δ.

Roughly, the last term has the following order

Proposition 4.3. *Under Assumptions [A1]-[A2]-[A3]-[A5],* $\mathbb{E}[(\sigma^2(V_{(k+1)\Delta}) - \sigma^2(\bar{V}_k))^2] \leq C\Delta$ *where* C *neither depend on* k *nor on* Δ.

The order obtained in Proposition 4.3 is worse than the one obtained in Proposition 4.2 and is not enough to reach optimal rates in the risk bounds (see Remark 4.2 below). Nevertheless, if the functions of S_m are at least twice differentiable, then we obtain a better result by using another approach.

Proposition 4.4. *Let*

$$T_n(t) = \frac{1}{n}\sum_{k=1}^{n} (\sigma^2(V_{(k+1)\Delta}) - \sigma^2(\bar{V}_k))t(\bar{V}_k). \tag{30}$$

Then, under Assumptions [A1]-[A5] and for S_m *in collection* [T], $\Delta \leq 1$, $D_m \leq N_n \leq \sqrt{n\Delta}/\ln(n)$ *or for* S_m *in collection* [W], $\Delta \leq 1$, $D_m \leq N_n \leq n\Delta/\ln^2(n)$,

$$\mathbb{E}\left(\sup_{t\in S_m, \|t\|=1} T_n^2(t)\right) \leq C\left(D_m^2\Delta^2 + \Delta^3 D_m^5\right). \tag{31}$$

If moreover $\Delta \leq n^{-2/3}$ *for* [T] *or* $\Delta \leq n^{-3/4}$ *for* [W], *then, for the maximal space* $\mathcal{S}_n$ *of the collection,*

$$\mathbb{E}\left(\sup_{t\in\mathcal{S}_n,\|t\|=1} T_n^2(t)\right) \leq \frac{c}{n}. \tag{32}$$

To estimate σ^2 on the compact set $A=[0,1]$, we define first

$$\hat{\sigma}_m^2 = \arg\min_{t\in S_m} \breve{\gamma}_n(t), \text{ with } \breve{\gamma}_n(t) = \frac{1}{n}\sum_{k=1}^n \left[U_{k+1} - t(\bar{V}_k)\right]^2. \tag{33}$$

We refer to Remark 3.1 for the existence of $\hat{\sigma}_m^2$. As previously, the second step is to ensure an automatic selection of m. For this, we define

$$\breve{m} = \arg\min_{m\in\mathcal{M}_n} \left[\breve{\gamma}_n(\hat{\sigma}_m^2) + \breve{\text{pen}}(m)\right]. \tag{34}$$

We denote by $\tilde{\sigma}^2 = \hat{\sigma}_{\breve{m}}^2$ the resulting estimator and we need to determine the adequate $\breve{\text{pen}}(m)$.

4.2. **Risk of the estimator.** Let us define

$$\breve{\nu}_n(t) = \frac{1}{n}\sum_{k=1}^n t(\bar{V}_k)\breve{Z}_{(k+1)\Delta},\ \breve{R}_n(t) = \frac{1}{n}\sum_{k=1}^n t(\bar{V}_k)\breve{R}_{(k+1)\Delta}. \tag{35}$$

As for b, we start by writing:

$$\begin{aligned}\breve{\gamma}_n(t) - \breve{\gamma}_n(\sigma^2) &= \|\sigma^2 - t\|_n^2 - \frac{2}{n}\sum_{k=1}^n (t-\sigma^2)(\bar{V}_k)[U_{k+1} - \sigma^2(\bar{V}_k)] \\ &= \|\sigma^2 - t\|_n^2 - 2\breve{\nu}_n(t-\sigma^2) - 2T_n(t-\sigma^2) - 2\breve{R}_n(t-\sigma^2).\end{aligned}$$

We denote by σ_m^2 the orthogonal projection of σ_A^2 on S_m. Writing that $\breve{\gamma}_n(\hat{\sigma}_m^2) - \breve{\gamma}_n(\sigma^2) \leq \breve{\gamma}_n(\sigma_m^2) - \breve{\gamma}_n(\sigma^2)$, we get (see (30), (35))

$$\|\hat{\sigma}_m^2 - \sigma^2\|_n^2 \leq \|\sigma_m^2 - \sigma^2\|_n^2 + 2\breve{\nu}_n(\hat{\sigma}_m^2 - \sigma_m^2) + 2T_n(\hat{\sigma}_m^2 - \sigma_m^2) + 2\breve{R}_n(\hat{\sigma}_m^2 - \sigma_m^2).$$

Cancelling $\|\sigma_{A^c}^2\|_n^2$ on both sides of the inequality, we obtain

$$\|\hat{\sigma}_m^2 - \sigma_A^2\|_n^2 \leq \|\sigma_m^2 - \sigma_A^2\|_n^2 + 2\breve{\nu}_n(\hat{\sigma}_m^2 - \sigma_m^2) + 2T_n(\hat{\sigma}_m^2 - \sigma_m^2) + 2\breve{R}_n(\hat{\sigma}_m^2 - \sigma_m^2). \tag{36}$$

The last two terms can be controlled thanks to Propositions 4.2 and 4.4. Using Proposition 4.1, we can prove the result analogous to Proposition 3.2 by a similar proof which is omitted.

Proposition 4.5. *For* S_m *in* [DP], [W] *or* [T], *under Assumptions* [A1]-[A3]-[A5], *for any* Δ, $0<\Delta\leq 1$,

$$\mathbb{E}\left(\sup_{t\in S_m,\|t\|=1} \breve{\nu}_n^2(t)\right) \leq c\frac{\mathbb{E}(\sigma^4(V_0))D_m}{n}.$$

Remark 4.1. *We can draw intermediate conclusions as in Remark 3.2 concerning an estimator* $\hat{\sigma}_m^2$ *with fixed* m *(see (33)). Assume* [A1]-[A6]. *Let* $\Delta = \Delta_n \to 0$, *with* $n\Delta_n/\ln^2(n) \to +\infty$ *when* $n\to+\infty$. *Let* S_m *be a space in collection* [DP] *or* [T] *with* $dim(S_m) \leq N_n$, $N_n = o(n\Delta/\ln^2(n))$ *or in collection* [T] *with* $dim(S_m) \leq N_n$, $N_n =$

$o(\sqrt{n\Delta}/\ln(n))$. *Relying on Propositions 4.1, 4.2, 4.3, 4.4 and 4.5, it is possible to derive that*

$$\mathbb{E}(\|\hat{\sigma}_m^2 - \sigma_A^2\|_n^2) \leq 7\bar{\pi}_1\|\sigma_m^2 - \sigma_A^2\|^2 + K\frac{\mathbb{E}(\sigma^4(V_0))D_m}{n} + B_n, \tag{37}$$

where $\sigma_A^2 = \sigma^2\mathbf{1}_{[0,1]}$, K *is positive constant. The remainder term* B_n *is given by* $B_n = K'\Delta$ *for collection* [DP] *and* $B_n = K'/n$ *for collection* [T] *if* $\Delta \leq n^{-2/3}$ *and for collection* [W] *if* $\Delta \leq n^{-3/4}$. *Here* K' *is a positive constant. The proof of this result is not provided, since it is mainly implied by Theorem 4.1.*

Here again, to obtain results on the adaptive estimator, some more accurate considerations on the martingale properties must be driven. In particular, we prove:

Proposition 4.6. *Under the assumptions of Theorem 4.1,*

$$\mathbb{P}\left(\sum_{k=1}^{n} t(\bar{V}_k)\breve{Z}_{(k+1)\Delta}^{(1)} \geq n\epsilon, \|t\|_n^2 \leq v^2\right) \leq \exp\left(-Cn\frac{\epsilon^2/2}{2\sigma_1^4 v^2 + \epsilon\|t\|_\infty\sigma_1^2 v}\right)$$

and

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} t(\bar{V}_k)\breve{Z}_{(k+1)\Delta}^{(1)} \geq v\sigma_1^2\sqrt{2x} + \sigma_1^2\|t\|_\infty x, \|t\|_n^2 \leq v^2\right) \leq \exp(-Cnx). \tag{38}$$

The (non trivial) link between the two inequalities is established by Birgé and Massart (1998). Using this result, we can prove the following main theorem.

Theorem 4.1. *Let* $\Delta = \Delta_n \to 0$ *and* $n\Delta/\ln^2(n) \to +\infty$ *as* $n \to +\infty$. *Assume that* [A1]-[A6] *hold. Consider the nested collection of models [DP]* ($L_m = 1$) *or the general collection [GP]* (L_m *given by (12)), both with maximal dimension* $N_n \leq n\Delta/\ln^2(n)$. *Or consider the nested collection of models [T]* ($L_m = 1$) *with maximal dimension* $N_n \leq \sqrt{n\Delta}/\ln(n)$ *and* $\Delta \leq n^{-2/3}$. *Or consider the nested collection of models [W]* ($L_m = 1$) *with maximal dimension* $N_n \leq n\Delta/\ln^2(n)$ *and* $\Delta \leq n^{-3/4}$. *Then the estimator* $\tilde{\sigma}^2 = \hat{\sigma}_{\breve{m}}^2$ *of* σ^2 *where* $\breve{m}$ *is defined by (34) with*

$$\text{pĕn}(m) \geq \tilde{\kappa}\sigma_1^4\frac{(1+L_m)D_m}{n}, \tag{39}$$

where $\tilde{\kappa}$ *is a universal constant, is such that*

$$\mathbb{E}(\|\tilde{\sigma}^2 - \sigma_A^2\|_n^2) \leq C\inf_{m\in\mathcal{M}_n}\left(\|\sigma_m^2 - \sigma_A^2\|^2 + \text{pĕn}(m)\right) + B_n', \tag{40}$$

where B_n' *is given by* $B_n' = K"\Delta$ *for collection* [DP] *or* [GP] *and* $B_n = K"/n$ *for collections* [T] *or* [W] *where* $K"$ *is a positive constant.*

Remark 4.2. *Let us discuss the rate of convergence of* $\tilde{\sigma}^2$ *in relation with Hoffmann's (1999) results. Assume that* σ_A^2 *belongs to a ball of some Besov space,* $\sigma_A^2 \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, *with* $\alpha \geq 2$.

Consider first collection [DP] with $r + 1 > \alpha$. *For* $\|\sigma_A^2\|_{\alpha,2,\infty} \leq L$, *it is known that* $\|\sigma_A^2 - \sigma_m^2\|^2 \leq C(\alpha, L)D_m^{-2\alpha}$. *The infimum in (40) is attained when* $D_m \propto n^{-1/(2\alpha+1)}$ *and this choice yields*

$$\mathbb{E}(\|\tilde{\sigma}^2 - \sigma_A^2\|_n^2) \leq Cn^{-2\alpha/(2\alpha+1)} + K"\Delta. \tag{41}$$

The first term $n^{-2\alpha/(2\alpha+1)}$ is the optimal nonparametric rate proved by Hoffmann (1999).

However, we still have to check that the optimal dimension $D_m = n^{1/(2\alpha+1)}$ can be attained, i.e. *that $n^{1/(2\alpha+1)} \leq N_n \leq n\Delta/\ln^2(n)$. This requires $\Delta \geq n^{-2\alpha/(2\alpha+1)}\ln^2(n)$. Hence, the optimal rate can at best be attained with a logarithmic loss. But we must fix Δ without knowledge of α. Since $\alpha \geq 2$, $2\alpha/(2\alpha+1) \geq 4/5$. Consequently the only admissible choice is $\Delta = n^{-4/5}\ln^2(n)$ which is consistent with the constraint $n\Delta^2 = o(1)$ found for the drift. If $\alpha = 2$, the optimal rate is attained with a logarithmic loss. Otherwise, it is not.*

Consider now collection [T]. *With $D_m \propto n^{-1/(2\alpha+1)}$, we get now*

$$\mathbb{E}(\|\tilde{\sigma}^2 - \sigma_A^2\|_n^2) \leq Cn^{-2\alpha/(2\alpha+1)} + \frac{K"}{n}. \tag{42}$$

We consider $\Delta = n^{-c}$ with $c > 2/3$ and we require $n^{1/(2\alpha+1)} \leq N_n \leq \sqrt{n\Delta}/\ln(n)$. This gives $c < (2\alpha-1)/(2\alpha+1)$. Therefore there is now a possible range of values for c: $]2/3, (2\alpha-1)/(2\alpha+1)[\neq \emptyset$ for $\alpha > 5/2$. Clearly, the collection [T] *is well fitted for estimating very smooth functions. Notice that when $\alpha \to +\infty$, the range for c tends to $]2/3, 1[$. It follows that for large values of α, the optimal nonparametric rate is reached for a wider range of values of c.*

For collection [W], *(42) still holds. An analogous discussion leads to $c \in]3/4, 2\alpha/(2\alpha+1)[$ which non empty for any $\alpha \geq 2$ and contains the interval $]3/4, 4/5[$.*

5. Discussion about Assumption [A6]

5.1. **Sufficient conditions.** Consider a diffusion model $dV_t = b(V_t)dt + \sigma(V_t)dW_t$, $V_0 = \eta$ satisfying [A1]-[A5]. We give now details on how to check Assumption [A6]. First note that the existence of the density $\bar{\pi}_\Delta$ of $\bar{V}_0$ is obtained under rather mild conditions on b and σ. For this, it is enough to check that the two-dimensional diffusion process (X_t, V_t) with $dX_t = V_t dt$ satisfies the Hörmander condition (see *e.g.* Rogers and Williams (2000), where this model is studied). Under rather strong assumptions on b and σ, the following proposition shows that [A6] holds.

Proposition 5.1. *Assume that b, σ are defined on $\mathbb{R}$ and C^1, that b, b', σ, σ' are bounded and that $\sigma(.) \geq \sigma_0 > 0$. Then, on any compact interval $K \subset \mathbb{R}$, there exist constants c, C depending only on the bounds of b and σ and their derivatives and not on Δ, such that*

$$\forall v \in K, \quad c \leq \bar{\pi}_\Delta(v) \leq C.$$

5.2. **Explicit examples.** Assumption [A6] can also be checked when explicit formulae are available. Note that, as Δ tends to 0, $\bar{V}_0$ tends to $V_0 = \eta$ almost surely, hence in distribution. Now, the characteristic functions of these random variables are often more explicit. Using the Fourier inversion formula, we can use the following standard sufficient condition.

- Let $\Phi_\Delta(s)$ and $\Phi(s)$ denote respectively the characteristic functions of $\bar{V}_0$ and $V_0 = \eta$. If $\int_{\mathbb{R}} |\Phi_\Delta(s) - \Phi(s)|ds$ tends to 0 as Δ tends to 0, then $\sup_{v\in(r_0,r_1)} |\bar{\pi}_\Delta(v) - \pi(v)|$ tends to 0.

Since the stationary density π satisfies [A6], the same will hold for $\bar{\pi}_\Delta$.

We consider two models. For Model 1, the density $\bar{\pi}_\Delta$ is explicit. For Model 2, we compute its characteristic function.

Model 1. The Ornstein-Uhlenbeck process gives evidently an explicit case. Consider $dV_t = -\theta V_t dt + cdW_t, V_0 = \eta$, with $\theta > 0$ and η centered Gaussian with variance $\rho^2 = c^2/2\theta$. Then, the solution process (V_t) is centered Gaussian, with covariance function $(s,t) \to \rho^2 \exp(-\theta|t-s|)$. The random variable $\bar{V}_0$ is centered Gaussian with variance

$$\bar{\sigma}^2_\Delta = \frac{c^2(e^{-\theta\Delta} - 1 + \theta\Delta)}{\Delta^2\theta^3} \sim \rho^2 \quad \text{as} \quad \Delta \to 0.$$

Model 2. Now, we consider the classical model used by Cox, Ingersoll and Ross (1985) to model interest rates. Let V_t be given by

$$dV_t = (-2\theta V_t + \delta c^2)dt + 2cV_t^{1/2}dW_t, V_0 = \eta. \tag{43}$$

Since this model is well known, we briefly recall some of its properties (for more details, see *e.g.* Lamberton and Lapeyre (1996) or Chaleyat-Maurel and Genon-Catalot (2006)). We assume that $\theta > 0$ and $\delta \geq 1$. When δ is integer, (V_t) is identical in law to $\sum_{i=1}^{\delta}(\xi_t^\delta)^2$ where (ξ_t^j) are i.i.d. Ornstein-Uhlenbeck processes solution of $d\xi_t^j = -\theta\xi_t^j dt + cdW_t^j$. Setting again $\rho^2 = c^2/2\theta$, the stationary distribution of (43) is the Gamma distribution $G(\delta/2, 1/2\rho^2)$. This law is exactly equal to a $\rho^2\chi^2(\delta)$. The Laplace transform is $\bar{V}_0$ is explicit and can be obtained as follows.

Proposition 5.2. *For $\lambda > 0$,*

$$\varphi_t(\lambda) = \mathbb{E}(\exp(-\lambda\int_0^t V_s ds)) = B_t(\lambda)^\delta \left(\frac{1}{1 + 2\mu_t(\lambda)\rho^2}\right)^{\delta/2}, \tag{44}$$

with

$$B_t(\lambda) = \left(1 + (\bar{c} + \theta)\frac{e^{2\bar{c}t} - 1}{2\bar{c}}\right)^{-1/2} \exp((\bar{c} + \theta)t/2), \quad \mu_t(\lambda) = \lambda\frac{e^{2\bar{c}t} - 1}{2\bar{c} + (\bar{c} + \theta)(e^{2\bar{c}t} - 1)},$$

and $\bar{c} = (\theta^2 + 2\lambda c^2)^{1/2}$.

Then we can easily deduce:

Corollary 5.1. *The characteristic function of $\bar{V}_0$ is equal to*

$$\Phi_\Delta(s) = \varphi_\Delta(-is/\Delta)$$

where $\varphi_t(\lambda)$ is given in (44).

Looking at formula (75), we see that the characteristic function of V_0 is equal to

$$\Phi(s) = (1 - 2is\rho^2)^{-(\delta/2)}.$$

This function is integrable for $\delta/2 > 1$. After some tedious computations, we can prove that

$$\sup_{\Delta \leq 1} |\Phi_\Delta(s)|$$

is also integrable for the same values of δ. So, in these cases, we get the uniform convergence of $\bar{\pi}_\Delta$ to π and [A6] holds.

5.3. **An approach well-fitted to the problem.** Actually, we only need

[A'6] $\exists \bar{\pi}_0, \bar{\pi}_1$ independent of n and Δ, such that

(i) $\forall m \in \mathcal{M}_n$, $\forall t \in S_m$,

$$\bar{\pi}_0 \|t\|^2 \leq \int t^2(x)\bar{\pi}_\Delta(x)dx \leq \bar{\pi}_1 \|t\|^2,$$

(ii) $\|b_A - b_m\|^2_{\bar{\pi}} \leq \bar{\pi}_1 \|b_A - b_m\|^2$ and $\|\sigma^2_A - \sigma^2_m\|^2_{\bar{\pi}} \leq \bar{\pi}_1 \|\sigma^2_A - \sigma^2_m\|^2$.

Obviously [A6] implies [A'6] and we can prove [A'6] (i) in our context.

Proposition 5.3. *$\forall m \in \mathcal{M}_n$, $\forall t \in S_m$ for S_m in collection* [T] *or* [W],

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

With $N_n \leq \sqrt{n\Delta}/\ln(n)$ (resp. $N_n \leq n\Delta/\ln^2(n)$), the quantity $N_n^3\Delta$ tends to zero when n tends to infinity and $\Delta = \Delta_n = o(n^{-2/3})$ (resp. $\Delta = \Delta_n = o(n^{-3/4})$). Therefore, it follows from Proposition 5.3 that, for n large enough, [A'6] (i) holds with e.g. $\bar{\pi}_1 = (3/2)\pi_1$ and $\bar{\pi}_0 = (1/2)\pi_0$ where $\pi_1 = \sup_{x\in A}\pi(x)$ and $\pi_0 = \inf_{x\in A}\pi(x)$.

6. Examples and numerical simulation results

In this section, we consider examples of diffusions and implement the estimation algorithms on simulated data.

6.1. **Examples of diffusions.** We consider the processes $V_t^{(i)}$ for $i = 1, \ldots, 7$ specified by the couples of functions $(b^{(i)}, \sigma^{(i)})$ given in Table 6.1.

To simulate sample paths of diffusions $V_t^{(1)}$ and $V_t^{(3)}$, we use the retrospective exact simulation algorithms proposed by Beskos et al. (2006a) and Beskos and Roberts (2005). Contrary to the Euler scheme, these algorithms produce exact simulation of diffusions under some assumptions on the drift and diffusion coefficient. We refer to Comte et al. (2005) for details on the way the diffusions are chosen and generated.

Then processes $V_t^{(2)}$, $V_t^{(4)}$ and $V_t^{(5)}$ are obtained as transformations of the previous ones. More precisely, $V_t^{(2)} = \sinh(V_t^{(1)}/c)$ and $V_t^{(4)} = \arg\sinh(cV_t^{(3)})$ and $V_t^{(5)} = G(V_t^{(3)})$ with $G(x) = \arg\sinh(x-5) + \arg\sinh(x+5)$. The function $G(.)$ is invertible and its inverse has the following explicit expression,

$$G^{-1}(x) = \frac{1}{\sqrt{2}\,\sinh(x)}\left[49\sinh^2(x) + 100 + \cosh(x)(\sinh^2(x) - 100)\right]^{1/2}. \tag{45}$$

The last two models are simulated by using that the exact discretization of an Ornstein-Uhlenbeck process is an autoregressive process of order one with known coefficients and noise distribution. More precisely, $V_t^{(6)} = \tanh(Y_t)$ where $dY_t = -\theta Y_t dt + c dW_t$ and

$$Y_{i\delta} = e^{-\theta\delta}Y_{(i-1)\delta} + c\left(\frac{1-e^{-2\theta\delta}}{2\theta}\right)^{1/2}\varepsilon_i,$$

$Y_0 \rightsquigarrow \mathcal{N}(0, c^2/(2\theta))$, and the ε_i's are i.i.d. $\mathcal{N}(0,1)$.

For $V_t^{(7)}$, an exact discrete path is obtained with the standard following method. If U_t is a d-dimensional Ornstein-Uhlenbeck process:

$$dU_t = -\frac{\theta}{2}U_t dt + \frac{c}{2}dW^{(d)}(t)$$

Process	Drift: $b^{(i)}(x) =$	$\sigma^{(i)}(x) =$	(θ, c)
$V_t^{(1)}$	$-(\theta/c + c/2)\tanh(cx)$	1	(4,1)
$V_t^{(2)}$	$-\theta x$	$c\sqrt{1+x^2}$	(4,1)
$V_t^{(3)}$	$-\theta \dfrac{x}{\sqrt{1+c^2x^2}}$	1	(2,2)
$V_t^{(4)}$	$-\left[\theta + \dfrac{c^2}{2\cosh(x)}\right]\dfrac{\sinh(x)}{\cosh^2(x)}$	$\dfrac{c}{\cosh(x)}$	(2,2)
$V_t^{(5)}$	$G'(G^{-1}(x))b^{(3)}(G^{-1}(x)) + \frac{1}{2}G"(G^{-1}(x))$ (*)	$G'(G^{-1}(x))$ (**)	(1,10)
$V_t^{(6)}$	$-(1-x^2)\left[c^2x + \frac{\theta}{2}\ln\left(\frac{1+x}{1-x}\right)\right]$	$c(1-x^2)$	(1, 0.75)
$V_t^{(7)}$	$\frac{dc^2}{4} - \theta x$, $d = 9$	$c\sqrt{x}$	(0.75, 1/3)

TABLE 1. List of the simulated diffusion processes.
(*) $G(x) = \arg\sinh(x-5) + \arg\sinh(x+5)$, G^{-1} is given by (45),
(**) $G'(u) = \frac{1}{(1+(u-5)^2)^{1/2}} + \frac{1}{(1+(u+5)^2)^{1/2}}$

where $W^{(d)}$ is a d-dimensional standard brownian motion, then $V_t^{(7)} = |U_t|^2 = \sum_{i=1}^d U_{i,t}^2$ where $U_{i,t}$ are the coordinates of U_t, satisfies the equation

$$dV_t^{(7)} = \left[\frac{dc^2}{4} - \theta V_t^{(7)}\right]dt + c\sqrt{V_t^{(7)}}dW^*(t)$$

where W^* is another one-dimensional Brownian motion built on the coordinates of $W^{(d)}$.

Therefore, we build $U_0 \sim (c/2\sqrt{\theta})\mathcal{N}(0, I_d)$, where I_d denote the $d \times d$ identity matrix and

$$U_{(p+1)\delta} = e^{-\theta\delta/2}U_{p\delta} + \frac{c\sqrt{1-e^{-\theta\delta}}}{2\sqrt{\theta}}\varepsilon_{p+1}$$

where the ε_i's are i.i.d. $\mathcal{N}(0, I_d)$ random vectors and take $V_{k\delta}^{(7)} = |U_{k\delta}|^2$.

It can be checked that all the above processes satisfy assumptions [A1]-[A6], with $I = \mathbb{R}$ for $V_t^{(j)}$ with $j = 1, \ldots, 5$ and $I = [-1, 1]$ for $V_t^{(6)}$, $I = (0, +\infty)$ for $V_t^{(7)}$.

We obtain samples of direct observations of the processes $(V_{k\delta}^{(j)})_{1\le k\le N}$ for $j = 1, \ldots, 7$, from which we approximate the $(\bar{V}_k^{(j)})_{1\le k\le n}$, by taking the mean of every $p = N/n$ observations, the new step being $\Delta = p\delta$. We shall compare the estimation procedure using these $(\bar{V}_k^{(j)})$ with the one using the direct observations $V_{k\Delta}^{(j)}$. Note that the regression equations for the estimation based on the exact observations $V_{k\Delta}$ are the following:

$$\frac{1}{\Delta}(V_{(k+1)\Delta} - V_{k\Delta}) = b(V_{k\Delta}) + \text{ noise } + \text{ remainder}, \tag{46}$$

	b (integ)	b (exact)	σ^2 (integ)	σ^2 (exact)
$V1$	$1.5e-01$	$7.5e-02$	$3.4e-02$	$2.4e-02$
$V2$	$2.2e-01$	$1.6e-01$	$2.9e-02$	$2.5e-02$
$V3$	$3.5e-02$	$3.3e-02$	$1.3e-03$	$8.8e-04$
$V4$	$1.9e-01$	$1.2e-01$	$2.8e-01$	$1.8e-01$
$V5$	$1.0e-02$	$1.1e-02$	$7.9e-03$	$7.5e-03$
$V6$	$1.5e-02$	$1.4e-02$	$3.3e-03$	$1.5e-03$
$V7$	$2.6e-03$	$2.5e-03$	$5.1e-05$	$4.6e-05$

TABLE 2. Empirical risks obtained for the estimation of b and σ^2 with 100 paths of the integrated and exact discretized processes when using the trigonometric basis.

	b (integ)	b (exact)	σ^2 (integ)	σ^2 (exact)
$V1$	$1.0e-01$	$3.8e-02$	$3.5e-02$	$2.4e-02$
$V2$	$9.3e-02$	$2.5e-02$	$2.8e-02$	$2.5e-02$
$V3$	$5.7e-02$	$5.4e-02$	$4.3e-03$	$2.4e-03$
$V4$	$1.9e-01$	$1.3e-01$	$4.2e-01$	$1.9e-01$
$V5$	$9.8e-03$	$1.0e-02$	$7.0e-03$	$5.9e-03$
$V6$	$1.6e-02$	$1.3e-02$	$3.2e-03$	$1.7e-03$
$V7$	$2.9e-04$	$3.3e-04$	$1.0e-05$	$6.6e-06$

TABLE 3. Empirical risks obtained for the estimation of b and σ^2 with 100 paths of the integrated and the exact discretized processes when using the piecewise polynomial basis.

	b (integ)	b (exact)	σ^2 (integ)	σ^2 (exact)
$V1$	$1.0e-01$	$3.7e-02$	$3.6e-02$	$2.5e-02$
$V2$	$9.3e-02$	$2.5e-02$	$2.9e-02$	$2.7e-02$
$V3$	$5.1e-02$	$4.8e-02$	$4.4e-03$	$2.4e-03$
$V4$	$1.9e-01$	$1.3e-01$	$2.9e-01$	$1.8e-01$
$V5$	$1.0e-02$	$1.1e-02$	$8.1e-03$	$6.9e-03$
$V6$	$1.7e-02$	$1.6e-02$	$3.2e-03$	$1.7e-03$
$V7$	$2.9e-04$	$3.3e-04$	$1.0e-05$	$6.6e-06$

TABLE 4. Empirical risks obtained for the estimation of b and σ^2 with 100 paths of the integrated and the exact discretized processes when using a mixed trigonometric-piecewise polynomial strategy.

$$\frac{1}{\Delta}(V_{(k+1)\Delta} - V_{k\Delta})^2 = \sigma^2(V_{k\Delta}) + \text{ noise } + \text{ remainder}, \tag{47}$$

see Comte *et al.* (2005). Obviously, risks are computed using $V_{k\Delta}$ instead of $\bar{V}_k$.

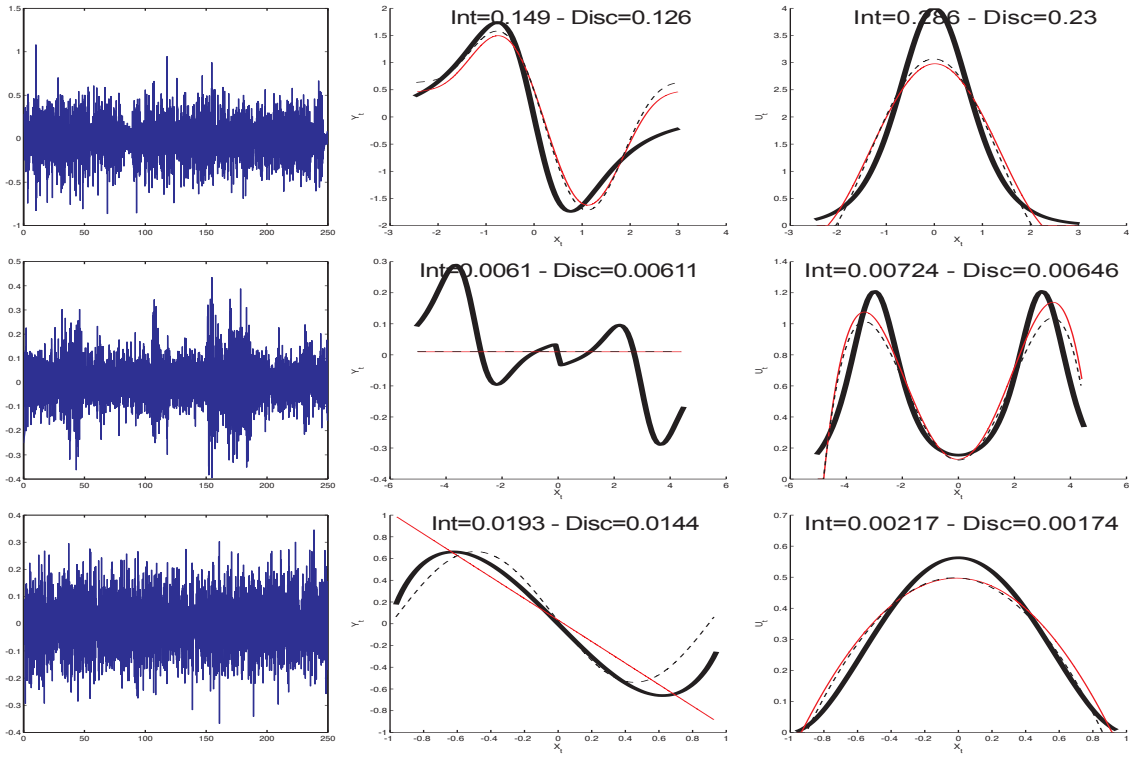

FIGURE 1. Processes $V^{(i)}$, $i = 4, 5, 6$ given in Table 6.1. First column: Difference between the integrated and discretized. True (bold), estimates using the integrated (thin grey) and the exact discretized (dotted thin) for b (second column) and σ^2 (third column). Error values: "Int" for the integrated and "Disc" for the exact discretized.

	b			σ^2		
	[T]	[GP]	[M]	[T]	[GP]	[M]
$V^{(1)}$	49	0	1	0	2	4
$V^{(2)}$	140	0	0	4	0	5
$V^{(3)}$	0	65	46	0	224	231
$V^{(4)}$	0	1	0	0	47	1
$V^{(5)}$	0	7	14	13	0	16
$V^{(6)}$	0	9	14	3	0	0
$V^{(7)}$	797	0	0	400	0	0

TABLE 5. (Risk−Best Risk)/Best Risk with Trigonometric [T], General Piecewise Polynomial [GP] or Mixed [M] bases.

6.2. **Estimation algorithms and numerical results.**

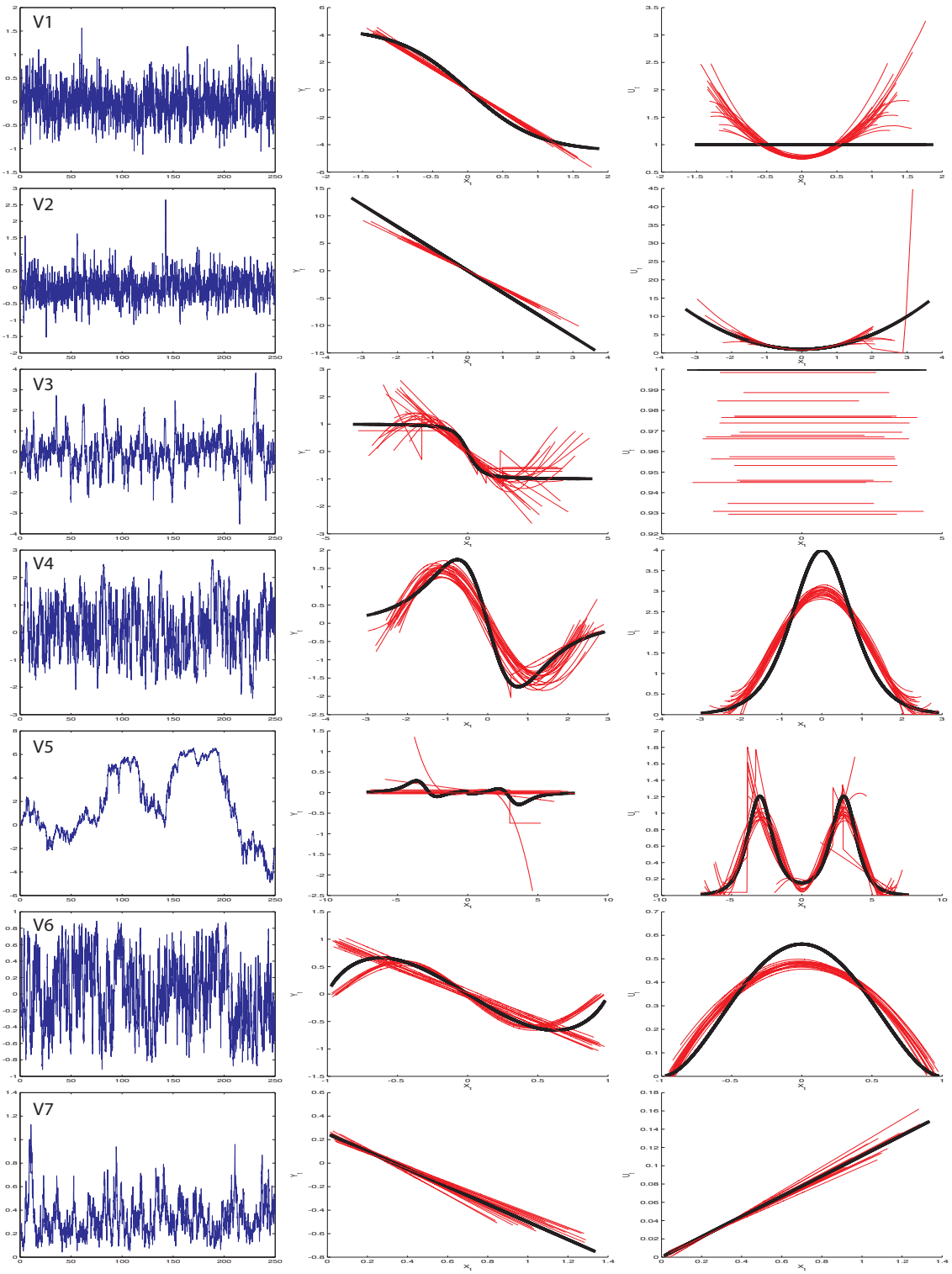

FIGURE 2. First column: one path of the processes $V^{(i)}$, $i = 1, \ldots, 7$ given in Table 6.1. Second column: true b (bold) and 20 estimations of b. Third column: true σ^2 and 20 estimations of σ^2.

We use the denoising algorithm described in full details in Comte and Rozenholc (2004). The algorithm minimizes the mean-square contrast and selects the space of approximation. There is a difficulty for precise calibration of the penalties. This is done for bases [GP] and [T] and this is the reason why our implementation focuses on those spaces.

Additive correcting terms are involved in the penalty (see Comte and Rozenholc (2004)). Such terms avoid under-penalization and are in accordance with the fact that the theorems provide lower bounds for the penalty. The correcting terms are asymptotically negligible so they do not affect the rate of convergence. Both penalties contain additional logarithmic terms which have been calibrated in other contexts by intensive simulation experiments (see Comte and Rozenholc (2002, 2004)).

More precisely, for collection [GP], the drift penalty ($i = 1$) and the diffusion penalty ($i = 2$) are given by

$$2\frac{\hat{s}_i^2}{n}\left(d - 1 + \ln\begin{pmatrix} D_{max} - 1 \\ d - 1 \end{pmatrix} + \ln^{2.5}(d) + \sum_{j=1}^{d}(r_j + \ln^{2.5}(r_j + 1))\right).$$

These penalties are valid for collection [T], with $d = D_{max} = 1$ and $r_1 = D_m$. For [GP], $D_{max} = [n\Delta/\ln^{1.5}(n)]$, $R_{max} = 5$ and for [T], r_1 is at most $[n\Delta/\ln^{1.5}(n)]$.

The constants κ and $\tilde{\kappa}$ in both drift and diffusion penalties have been set equal to 2. 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 basis [T] and with a preliminary penalty where $\hat{s}_2^2$ is taken equal to $2\max_m(\breve{\gamma}_n(\hat{\sigma}_m^2))$. This gives a preliminary estimator $\tilde{\sigma}_0^2$. Now, we take $\hat{s}_2$ equal to twice the 99.5%-quantile of $\tilde{\sigma}_0^2$. The use of the quantile is here to avoid extreme values. We get $\tilde{\sigma}^2$. We use this estimate and set $\hat{s}_1^2 = \max_{1\le k\le n}(\tilde{\sigma}^2(\bar{V}_k))/\Delta$ for the penalty of b.

In all the examples, parameters have been chosen in the admissible range of ergodicity (see Table 6.1). The sample size $n = 5000$ and the step $\Delta = 1/20$ are in accordance with the asymptotic context (great n's and small Δ's) and may be relevant for applications in finance. They are obtained with $N = 50000$ initial observations and blocks of size $p = 10$ to compute the integrated process.

First, Tables 2, 3, 4 give empirical risks estimated over 100 simulated paths. In Tables 2 and 3, we give the results of the estimation procedure when the $V_{k\Delta}$'s are observed or when only the $\bar{V}_k$'s are available, using either the trignometric basis [T] or the general piecewise polynomials basis [GP]. In addition, we also made another attempt denoted by [M] (mixed) whose results are stated in Table 4. In [M], the algorithm chooses between the basis [T] and [GP], looking at the global penalized least square criterion value. It appears that the results are slightly better with the exact observations, which was to be expected. One can notice that the risks are in most cases smaller for the estimation of σ^2 than for the estimation of b, which is in accordance with the theoretical rates.

Figure 1 shows in a few cases (for $V^{(4)}$, $V^{(5)}$ and $V^{(6)}$) the differences $\bar{V}_k - V_{k\Delta}$ (first column). Clearly, these differences look like white noises for $V^{(4)}$ and $V^{(6)}$ and this was also true for $V^{(1)}$, $V^{(2)}$, $V^{(3)}$ and $V^{(7)}$. Only $V^{(5)}$ seems to suffer from a lack of stationarity implying some picks. In any case, the approximation of $V_{k\Delta}$ by $\bar{V}_k$ does not suffer from any systematic bias. The last columns of Figure 1 plot the estimated curves obtained when using the $V_{k\Delta}$'s or the $\bar{V}_k$'s, with associated error values. The estimated curves are very close.

Table 5 compares more directly the performances of the different bases [T], [GP] and the mixed basis [M]. The table gives the relative differences 100[(risk - smallest risk)/smallest risk], which is a percentage of degradation with respect to the best score. Consequently, the best basis corresponds to a null value. The basis [GP] appears to be better than [T]: both have approximately the same number of null scores but errors with [T] may be large. The mixed strategy is slightly better, but it does not really outperforms [GP].

Lastly, in Figure 2, we have plotted the sample paths of $V^{(1)}, \ldots, V^{(7)}$, the true functions b and σ^2 (bold lines) together with 20 estimated functions based on the data points $\bar{V}_k$ using the mixed strategy [M].

7. Proofs for the estimation of the drift

We shall need all along the proofs the following results and decompositions. First

$$V_{(k+1)\Delta} = \bar{V}_k + \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} (u - k\Delta) dV_u. \tag{48}$$

Noting that $V_{(k+3)\Delta} - V_{(k+2)\Delta} = \int_{(k+2)\Delta}^{(k+3)\Delta} dV_u$, and using (48), we get

$$Y_{k+1} = \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{(k+1)\Delta}(u) dV_u \tag{49}$$

where $\psi_{k\Delta}$ is given in (16). Second

Lemma 7.1. *Under assumptions [A1]-[A3], for all s, t, $|t-s| \leq 1$, $\mathbb{E}(V_t - V_s)^{2i} \leq c|t-s|^i$ and $\mathbb{E}(V_{(k+1)\Delta} - \bar{V}_k)^{2i} \leq c\Delta^i$ for $i \leq 6$ and for any integer k.*

Proof of Lemma 7.1. From the strict stationarity, it is enough to prove that for $0 \leq t \leq 1$, $\mathbb{E}(V_t - V_0)^{2i} \leq ct^i$. This follows from [A3], (5), and standard applications of Hölder and Burkholder-Davis-Gundy inequalities.□

7.1. **Proof of Proposition 3.1.** Using (49), we can see that, in decomposition (14), the residual term can be written $R_b(k\Delta) = \sum_{i=1}^5 R^{(i)}(k\Delta)$ with

$$\begin{aligned}
R_b^{(1)}((k+1)\Delta) &= b(V_{(k+1)\Delta}) - b(\bar{V}_k), \quad R_b^{(2)}((k+1)\Delta) = \frac{1}{\Delta}\int_{(k+1)\Delta}^{(k+2)\Delta} [b(V_s) - b(V_{(k+1)\Delta})] ds \\
R_b^{(3)}((k+1)\Delta) &= \frac{1}{2}(b(V_{(k+2)\Delta}) - b(V_{(k+1)\Delta}) \\
R_b^{(4)}((k+1)\Delta) &= \frac{1}{\Delta^2}\int_{(k+2)\Delta}^{(k+3)\Delta} ((k+3)\Delta - s)(b(V_s) - b(V_{(k+2)\Delta})) ds \\
R_b^{(5)}((k+1)\Delta) &= -\frac{1}{\Delta^2}\int_{(k+2)\Delta}^{(k+1)\Delta} ((k+2)\Delta - s)(b(V_s) - b(V_{(k+1)\Delta})) ds.
\end{aligned}$$

Four terms are under study, the fifth one being the same as the fourth. For the first one, use Taylor formula, Lemma 7.1 and [A5](i) to obtain

$$\begin{aligned}
\mathbb{E}[(R_b^{(1)}((k+1)\Delta))^2] &= \mathbb{E}\left\{\left[(V_{(k+1)\Delta} - \bar{V}_k)\int_0^1 b'(\bar{V}_k + u(V_{(k+1)\Delta} - \bar{V}_k)) du\right]^2\right\} \\
&\leq K\mathbb{E}\left[(V_{(k+1)\Delta} - \bar{V}_k)^2\right] \leq K'\Delta. \qquad (50)
\end{aligned}$$

It follows from (5) and Lemma 7.1 that

$$\mathbb{E}\left(\int_{(k+1)\Delta}^{(k+2)\Delta}(b(V_s)-b(V_{(k+1)\Delta}))ds\right)^2 \le \Delta\int_{(k+1)\Delta}^{(k+2)\Delta} c'\Delta ds = c'\Delta^3.$$

Thus, $\mathbb{E}[(R_b^{(2)}((k+1)\Delta))^2] \le c'\Delta$. The third term is obvious. Lastly

$$\mathbb{E}[(R_b^{(4)}((k+1)\Delta))^2] \le \frac{1}{\Delta^4}\int_{(k+1)\Delta}^{(k+2)\Delta}\Delta[(k+2)\Delta-u]^2\mathbb{E}[(b(V_u)-b(V_{(k+1)\Delta}))^2]du$$

so that with (5) again, we obtain $\mathbb{E}[(R_b^{(4)}((k+1)\Delta))^2] \le c_4\Delta$. Analogous tools lead to $\mathbb{E}[(R_b((k+1)\Delta))^4] \le c\Delta^2$. □

7.2. **Proof of Proposition 3.2.** For S_m in [DP] or [GP], we can write

$$\begin{aligned}
&\mathbb{E}\left(\sup_{t\in S_m,\|t\|=1}[\nu_n(t)]^2\right) \le \sum_{\lambda\in\Lambda_m}\mathbb{E}[\nu_n^2(\varphi_\lambda)] = \sum_{\lambda\in\Lambda_m}\mathrm{Var}\left(\nu_n(\varphi_\lambda)\right)\\
=\ &\sum_{\lambda\in\Lambda_m}\frac{1}{n^2}\sum_{1\le k,l\le n}\mathrm{cov}(\varphi_\lambda(\bar V_k)Z_{(k+1)\Delta},\varphi_\lambda(\bar V_l)Z_{(l+1)\Delta})\\
=\ &\sum_{\lambda\in\Lambda_m}\left(\frac{1}{n}\mathrm{Var}(\varphi_\lambda(\bar V_1)Z_{2\Delta})+\frac{1}{n^2}\sum_{k=1}^{n-1}\mathrm{cov}(\varphi_\lambda(\bar V_k)Z_{(k+1)\Delta},\varphi_\lambda(\bar V_{k+1})Z_{(k+2)\Delta})\right)\\
\le\ &\sum_{\lambda\in\Lambda_m}\frac{2}{n}\mathbb{E}(\varphi_\lambda^2(\bar V_1)Z_{2\Delta}^2),
\end{aligned}$$

as the variances of the $\varphi_\lambda(\bar V_k)Z_{(k+1)\Delta}$'s do not depend on k. When S_m belongs to collection [DP], we use (11) and (17) and get

$$\mathbb{E}\left(\sup_{t\in S_m,\|t\|=1}[\nu_n(t)]^2\right) \le \frac{2(r+1)D_m}{n}\mathbb{E}(Z_{2\Delta}^2) = \frac{4(r+1)D_m}{3n\Delta}\mathbb{E}(\sigma^2(V_0)). \tag{51}$$

Now, for collection [GP], we use [A5] to write

$$\mathbb{E}(\varphi_\lambda^2(\bar V_1)Z_{2\Delta}^2) = \mathbb{E}\varphi_\lambda^2(\bar V_1)\frac{1}{\Delta^4}\int_{2\Delta}^{4\Delta}\psi_{2\Delta}^2(u)\sigma^2(V_u)du \le \frac{2\sigma_1^2}{3\Delta}\mathbb{E}(\varphi_\lambda^2(\bar V_1)).$$

By [A6], $\mathbb{E}(\varphi_\lambda^2(\bar V_1)) \le \bar\pi_1\int_A\varphi_\lambda^2(x)dx = \bar\pi_1$. Thus, for S_m in [GP],

$$\mathbb{E}\left(\sup_{t\in S_m,\|t\|=1}[\nu_n(t)]^2\right) \le \frac{4\sigma_1^2\bar\pi_1 D_m}{3n\Delta}.\square$$

7.3. **Proof of Proposition 3.3.** We use that $\sum_{k=1}^n t(\bar V_k)Z_{(k+1)\Delta}$ can be written as a stochastic integral. Consider the process $H_u^n = H_u$ defined by

$$H_u = \sum_{k=1}^n \psi_{(k+1)\Delta}(u)t(\bar V_k)\sigma(V_u)$$

with $\psi_{k\Delta}$ given by (16). Note that $0 \le \psi_{k\Delta}(u) \le 1$ for all u and k and $\|\psi_{k\Delta}\|^2 = \int\psi_{k\Delta}^2(u)du = 2\Delta/3$. Then, H_u satisfies $H_u^2 \le \sigma_1^2\|t\|_\infty^2$ for all $u \ge 0$. Then, denoting by

$M_s = \int_0^s H_u dW_u$, we get that $M_{(n+1)\Delta} = \Delta \sum_{k=1}^n t(\bar{V}_k) Z_{(k+1)\Delta}$, and $\langle M \rangle_{(n+1)\Delta}$ is less than or equal to

$$2\sum_{k=1}^n t^2(\bar{V}_k) \left[\int_{(k+1)\Delta}^{(k+2)\Delta} [1 - ((k+2) - \frac{u}{\Delta})]^2 \sigma^2(V_u) du + \int_{(k+2)\Delta}^{(k+3)\Delta} [(k+3) - \frac{u}{\Delta}]^2 \sigma^2(V_u) du \right].$$

Moreover, $\langle M \rangle_s \leq 2n\sigma_1^2 \Delta \|t\|_n^2$, $\forall s \geq 0$, so that (M_s) and $\exp(\lambda M_s - \lambda^2 \langle M \rangle_s / 2)$ are martingales with respect to the filtration $\mathcal{F}_s = \sigma(X_u, u \leq s)$. Therefore, for all $s \geq 0$, $c > 0$, $d > 0$, $\lambda > 0$,

$$\mathbb{P}(M_s \geq c, \langle M \rangle_s \leq d) \leq \mathbb{P}\left(e^{\lambda M_s - \frac{\lambda^2}{2}\langle M \rangle_s} \geq e^{\lambda c - \frac{\lambda^2}{2} d} \right) \leq e^{-(\lambda c - \frac{\lambda^2}{2} d)}.$$

Therefore, $\mathbb{P}(M_s \geq c, \langle M \rangle_s \leq d) \leq \inf_{\lambda > 0} e^{-(\lambda c - \frac{\lambda^2}{2} d)} = e^{-\frac{c^2}{2d}}$. Finally,

$$\begin{aligned} \mathbb{P}\left[\sum_{k=1}^n t(\bar{V}_k) Z_{(k+1)\Delta} \geq n\epsilon, \|t\|_n^2 \leq v^2 \right] &= \mathbb{P}(M_{(n+1)\Delta} \geq n\Delta\epsilon, \langle M \rangle_{(n+1)\Delta} \leq 2nv^2\sigma_1^2\Delta) \\ &\leq \exp\left(-\frac{(n\Delta\epsilon)^2}{4nv^2\sigma_1^2\Delta} \right) = \exp\left(-\frac{n\epsilon^2\Delta}{4v^2\sigma_1^2} \right). \quad \square \end{aligned}$$

7.4. **Proof of Theorem 3.1.** Recall that $\|t\|_{\bar{\pi}}^2 = \int t^2(x) \bar{\pi}_\Delta(x) dx$ (see (9)). We start as for getting (24). By simply writing that $\gamma_n(\hat{b}_{\hat{m}}) + \text{pen}(\hat{m}) \leq \gamma_n(b_m) + \text{pen}(m)$, for all m in $\mathcal{M}_n$, we obtain

$$\begin{aligned} \|\hat{b}_{\hat{m}} - b_A\|_n^2 &\leq \|b_m - b_A\|_n^2 + 2\|\hat{b}_{\hat{m}} - b_m\|_{\bar{\pi}} \sup_{t \in S_{\hat{m}} + S_m, \|t\|_{\bar{\pi}} = 1} \nu_n(t) \\ &\quad + 2\|\hat{b}_{\hat{m}} - b_m\|_n \sqrt{\frac{1}{n}\sum_{k=1}^n R_b^2((k+1)\Delta)} + \text{pen}(m) - \text{pen}(\hat{m}) \\ &\leq \|b_m - b_A\|_n^2 + \frac{1}{8}\|\hat{b}_{\hat{m}} - b_m\|_{\bar{\pi}}^2 + 8 \sup_{t \in S_{\hat{m}} + S_m, \|t\|_{\bar{\pi}} = 1} [\nu_n(t)]^2 \\ &\quad + \frac{1}{8}\|\hat{b}_{\hat{m}} - b_m\|_n^2 + \frac{8}{n}\sum_{k=1}^n R_b^2((k+1)\Delta) + \text{pen}(m) - \text{pen}(\hat{m}) \end{aligned}$$

Let us consider the set

$$\Omega_n = \left\{ \omega / \left| \frac{\|t\|_n^2}{\|t\|_{\bar{\pi}}^2} - 1 \right| \leq \frac{1}{2}, \ \forall t \in \cup_{m,m' \in \mathcal{M}_n} (S_m + S_{m'}) / \{0\} \right\}. \tag{52}$$

We use that, on Ω_n, $\|t\|_{\bar{\pi}} \leq \sqrt{2}\|t\|_n$, and that $\|\hat{b}_{\hat{m}} - b_m\|_n^2 \leq 2(\|\hat{b}_{\hat{m}} - b_A\|_n^2\| + \|b_A - b_m\|_n^2)$. After some elementary computations, we get

$$\begin{aligned} \frac{1}{4}\|\hat{b}_{\hat{m}} - b_A\|_n^2 \mathbb{1}_{\Omega_n} &\leq \frac{7}{4}\|b_m - b_A\|_n^2 + 8 \sup_{t \in S_{\hat{m}} + S_m, \|t\|_{\bar{\pi}} = 1} [\nu_n(t)]^2 \mathbb{1}_{\Omega_n} + \frac{8}{n}\sum_{k=1}^n R_b^2((k+1)\Delta) \\ &\quad + \text{pen}(m) - \text{pen}(\hat{m}) \end{aligned}$$

By Proposition 3.1, $\mathbb{E}R_b^2((k+1)\Delta) \leq c\Delta$. Therefore, using [A6],

$$\begin{aligned}\mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_n^2 \mathbb{I}_{\Omega_n}) &\leq 7\bar{\pi}_1\|b_m - b_A\|^2 + 32\mathbb{E}\left(\sup_{t\in S_{\hat{m}}+S_m, \|t\|_{\bar{\pi}}=1} [\nu_n(t)]^2 \mathbb{I}_{\Omega_n}\right) \\ &\quad +32c'\Delta + 4(\mathrm{pen}(m) - \mathbb{E}(\mathrm{pen}(\hat{m}))).\end{aligned}$$

The difficulty here is to control the supremum of $\nu_n(t)$ on a random ball (which depends on the random $\hat{m}$). This is done by using the martingale property of $\nu_n(t)$. Let us set

$$G_m(m') = \sup_{t\in S_m+S_{m'}, \|t\|_{\bar{\pi}}=1} \nu_n(t).$$

Introducing a function $p(m, m')$, we first write

$$\begin{aligned}G_m^2(\hat{m})\mathbb{I}_{\Omega_n} &\leq [(G_m^2(\hat{m}) - p(m,\hat{m}))\mathbb{I}_{\Omega_n}]_+ + p(m,\hat{m}) \\ &\leq \sum_{m'\in\mathcal{M}_n} [(G_m^2(m') - p(m,m'))\mathbb{I}_{\Omega_n}]_+ + p(m,\hat{m}).\end{aligned}$$

Then the penalty pen(.) is chosen such that $32p(m, m') \leq 4(\mathrm{pen}(m) + \mathrm{pen}(m'))$. More precisely, the next proposition determines the choice of $p(m, m')$ which in turn will fix the penalty.

Proposition 7.1. *Under the assumptions of Theorem 3.1, there exists a numerical constant κ_1 such that, for $p(m, m') = \kappa_1\sigma_1^2[D_m + (1 + L_{m'}D_{m'})]/(n\Delta)$, we have*

$$\mathbb{E}[(G_m^2(m') - p(m,m'))\mathbb{I}_{\Omega_n}]_+ \leq c\sigma_1^2 \frac{e^{-L_{m'}D_{m'}}}{n\Delta}.$$

Proof of Proposition 7.1. The result of Proposition 7.1 follows from the inequality of Proposition 3.3 by the $\mathbb{L}^2$-chaining technique used in Baraud et al. (2001b) (see Section 7 p.44-47, Lemma 7.1, with $s^2 = \sigma_1^2/\Delta$). □

The result of Theorem 3.1 on Ω_n follows from Proposition 7.1 with $\mathrm{pen}(m) \geq \kappa\sigma_1^2(1 + L_m)D_m/(n\Delta)$, and $\kappa = 32\kappa_1$. Indeed, this choice ensures that for all m, m' in $\mathcal{M}_n$, $32p(m, m') \leq \mathrm{pen}(m) + \mathrm{pen}(m')$. Now, the weights given in (12) ensure that $\Sigma = \sum_{m'\in\mathcal{M}_n} e^{-L_{m'}D_{m'}} < +\infty$. Thus,

$$\mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_n^2 \mathbb{I}_{\Omega_n}) \leq 7\bar{\pi}_1\|b_m - b_A\|^2 + 8\mathrm{pen}(m) + c'\sigma_1^2\frac{\Sigma}{n\Delta} + 32c'\Delta. \tag{53}$$

Now, we look at Ω_n^c. In Lemma 6.1 of Comte et al. (2005), it is proved that, under our set of assumptions, $\mathbb{P}(\Omega_n^c) \leq \tilde{c}/n^4$. The constraint on N_n (i.e. $N_n = o(n\Delta/\ln^2(n)$ for [DP], [GP] or $N_n = o(\sqrt{n\Delta}/\ln(n)$ for [T]) is imposed here. The existence of the maximal space $\mathcal{S}_n$, [A4] and [A6] are especially needed also and the constant $\tilde{c}$ depends on $\bar{\pi}_0, \bar{\pi}_1$ and the rate of mixing θ.

Lastly we need to check that $\mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_n^2 \mathbb{I}_{\Omega_n^c}) \leq c/n$. Write the regression model (14) as $Y_{k+1} = b(\bar{V}_k) + \varepsilon_{(k+1)\Delta}$ with $\varepsilon_{k\Delta} = Z_{k\Delta} + R_b(k\Delta)$. Let us recall that Π_m denotes the orthogonal projection (with respect to the inner product of $\mathbb{R}^n$) onto the subspace of $\mathbb{R}^n$, $\{(t(\bar{V}_1), \ldots, t(\bar{V}_n))', t \in S_m\}$. By definition of $\hat{b}_m$, we have $(\hat{b}_m(\bar{V}_1), \ldots, \hat{b}_m(\bar{V}_n))' = \Pi_m Y$ where $Y = (Y_2, \ldots, Y_{n+1})'$. Denoting in the same way a function t and the vector

$(t(\bar{V}_1),\dots,t(\bar{V}_n))'$, we can see that $\|b_A-\hat{b}_{\hat{m}}\|_n^2=\|b_A-\Pi_{\hat{m}}b_A\|_n^2+\|\Pi_{\hat{m}}\varepsilon\|_n^2\leq\|b_A\|_n^2+n^{-1}\sum_{k=2}^{n+1}\varepsilon_{k\Delta}^2$. Using that $\mathbb{P}(\Omega_n^c)\leq\tilde{c}/n^4$ and [A6], we have:

$$\mathbb{E}(\|b_A\|_n^2\mathbb{1}_{\Omega_n^c})\leq\mathbb{E}^{1/2}(b^4(\bar{V}_0))\mathbb{P}^{1/2}(\Omega_n^c)\leq\frac{c}{n^2}. \tag{54}$$

Next, $\mathbb{E}(\varepsilon_{k\Delta}^2\mathbb{1}_{\Omega_n^c})\leq\mathbb{E}^{1/2}(\varepsilon_{k\Delta}^4)\mathbb{P}^{1/2}(\Omega_n^c)$. From Lemma 3.1 we know that $\mathbb{E}(R_b^4(k\Delta))\leq c'\Delta^2$. Moreover $\mathbb{E}(Z_{k\Delta}^4)\leq c\mathbb{E}(\sigma^4(V_0))/\Delta^2$. Thus, $\mathbb{E}(\varepsilon_{k\Delta}^4)\leq C'/\Delta^2$. This implies, by using that $n\Delta\geq 1$, that

$$\mathbb{E}\left[\|b_A-\hat{b}_{\hat{m}}\|_n^2\mathbb{1}_{\Omega_n^c}\right]\leq\frac{c"}{n}. \tag{55}$$

Inequality (27) of Theorem 3.1 follows by gathering (53) and (55). □

8. Proofs for the estimation of the diffusion coefficient

8.1. **Proof of Proposition 4.1.** First by using Burkholder-Davis-Gundy's inequality,

$$\begin{aligned}\mathbb{E}[(\breve{Z}_{k\Delta}^{(1)})^2] &\leq \frac{c}{\Delta^6}\mathbb{E}\left[\left(\int_{k\Delta}^{(k+2)\Delta}\psi_{k\Delta}^2(u)\sigma^2(V_u)du\right)^2\right]\\ &\leq \frac{c}{\Delta^5}\int_{k\Delta}^{(k+1)\Delta}\psi_k^4(u)\mathbb{E}[\sigma^4(V_0)]du\leq\frac{2c\mathbb{E}[\sigma^4(V_0)]}{5}.\end{aligned}$$

$$\mathbb{E}[(\breve{Z}_{k\Delta}^{(2)})^2] = \frac{9}{\Delta^4}\mathbb{E}\left(\int_{k\Delta}^{(k+2)\Delta}\psi_{k\Delta}^2(u)\sigma^2(V_u)b^2(V_{k\Delta})du\right)\leq 6\Delta\mathbb{E}^{1/2}[b^4(V_0)]\mathbb{E}^{1/2}[\sigma^4(V_0)].$$

Lastly, by using that $0\leq\psi_{k\Delta}(u)\leq\Delta$,

$$\begin{aligned}\mathbb{E}[(\breve{Z}_{k\Delta}^{(3)})^2] &= \frac{9}{4\Delta^6}\mathbb{E}\left\{\left[\int_{k\Delta}^{(k+2)\Delta}\left(\int_s^{(k+2)\Delta}\psi_{k\Delta}^2(u)du\right)(\sigma'\sigma^2)(V_s)dW_s\right]^2\right\}\\ &\leq \frac{9}{4\Delta^6}\int_{k\Delta}^{(k+2)\Delta}[2\Delta^3]^2\mathbb{E}[(\sigma'\sigma^2)^2(V_0)]ds=18\Delta\mathbb{E}[(\sigma'\sigma^2)^2(V_0)].\end{aligned}$$

For the moments of order 4, they are bounded for $\breve{Z}^{(1)}$ and of order Δ^2 for $\breve{Z}^{(2)}$ and $\breve{Z}^{(3)}$. □

8.2. **Proof of Proposition 4.2.** We use again (49) to compute U_{k+1} (29) and exhibit the rmainder term $\breve{R}_{(k+1)\Delta}$. More precisely, we have: $\breve{R}_{k\Delta}=\breve{R}_{k\Delta}^{(1)}+\breve{R}_{k\Delta}^{(2)}+\breve{R}_{k\Delta}^{(3)}$ with

$$\begin{aligned}\breve{R}_{k\Delta}^{(1)} &= \frac{3}{2\Delta^3}\left(\int_{k\Delta}^{(k+2)\Delta}\psi_{k\Delta}(s)b(V_s)ds\right)^2\\ \breve{R}_{k\Delta}^{(2)} &= \frac{3}{\Delta^3}\left(\int_{k\Delta}^{(k+2)\Delta}\psi_{k\Delta}(u)(b(V_u)-b(V_{k\Delta}))du\right)\left(\int_{k\Delta}^{(k+2)\Delta}\psi_{k\Delta}(u)\sigma(V_u)dW_u\right)\\ \breve{R}_{k\Delta}^{(3)} &= \frac{3}{2\Delta^3}\int_{k\Delta}^{(k+2)\Delta}\left(\int_s^{(k+2)\Delta}\psi_{k\Delta}^2(u)du\right)\tau_{b,\sigma}(V_s)ds,\end{aligned}$$

where $\tau_{b,\sigma}=2(\sigma'\sigma b)+[(\sigma')^2+\sigma\sigma"]\sigma^2$.

Proposition 4.2 holds if $\mathbb{E}[(\breve{R}^{(i)}_{k\Delta})^2] \leq c_i\Delta^2$ for $i = 1, 2, 3$.

$$\begin{aligned}\mathbb{E}[(\breve{R}^{(1)}_{k\Delta})^2] &\leq \frac{9}{4\Delta^6}\mathbb{E}\left(\int_{k\Delta}^{(k+2)\Delta}\psi_{k\Delta}(s)b^2(V_s)ds\right)^4 \leq \frac{18}{\Delta^3}\mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta}\psi^4_{k\Delta}(s)b^4(V_s)ds\right)\\ &\leq 8\Delta^2\mathbb{E}(b^4(V_0)) \leq c\Delta^2 \text{ using [A5]}(i) \text{ and } \mathbb{E}(V_0^4) < +\infty,\end{aligned}$$

since $\int_{(k+1)\Delta}^{(k+3)\Delta}\psi^4_{k\Delta}(s)ds = 2\Delta^5/5$.

$$\mathbb{E}[(R^{(2)}_{k\Delta})^2] \leq \frac{9}{\Delta^6}\mathbb{E}\left(\int_{k\Delta}^{(k+2)\Delta}\psi_{k\Delta}(s)(b(V_s) - b(V_{k\Delta}))ds\int_{k\Delta}^{(k+2)\Delta}\psi_{k\Delta}(s)\sigma(V_s)dW_s\right)^2.$$

Both terms have already been studied and using (5), this yields, if $\mathbb{E}(V_0^8) < +\infty$,

$$\mathbb{E}[(R^{(2)}_{k\Delta})^2] \leq c'\Delta^2.$$

Lastly

$$\mathbb{E}[(R^{(3)}_{k\Delta})^2] \leq \frac{9}{2\Delta^5}\mathbb{E}\left(\int_{k\Delta}^{(k+2)\Delta}[\int_s^{(k+2)\Delta}\psi^2_{k\Delta}(u)du]^2\tau^2_{b,\sigma}(V_s)ds\right) \leq 36\mathbb{E}(\tau^2_{b,\sigma}(V_0))\Delta^2 \leq c"\Delta^2,$$

by using [A5](i) and $\mathbb{E}(V_0^{12}) < +\infty$, since $\psi_k(u)^2 \leq \Delta^2$. Therefore Lemma 4.2 is proved. □

8.3. **Proof of Proposition 4.3.** From standard results on Euler schemes, it is known that:

$$\sigma^2(V_{(k+1)\Delta}) - \sigma^2(V_{k\Delta}) = \sqrt{\Delta}(\sigma^2)'(V_{k\Delta})\sigma(V_{k\Delta})\xi_k + \tilde{R}^{(1)}_k$$

where $\xi_k = \Delta^{-1/2}(W_{(k+1)\Delta} - W_{k\Delta})$ and $\mathbb{E}[(\tilde{R}^{(1)}_k)^2] \leq c\Delta^2$. Moreover, from Gloter (2000, Proposition 2), and the Taylor formula, we easily deduce that

$$\sigma^2(\bar{V}_k) - \sigma^2(V_{k\Delta}) = \sqrt{\Delta}(\sigma^2)'(V_{k\Delta})\sigma(V_{k\Delta})\xi'_k + \tilde{R}^{(2)}_k \tag{56}$$

where $\xi'_k = \Delta^{-3/2}\int_{k\Delta}^{(k+1)\Delta}[(k+1)\Delta - s]dW_s$ and $\mathbb{E}[(\tilde{R}^{(2)}_k)^2] \leq c\Delta^2$. Therefore, the following holds:

$$\sigma^2(V_{(k+1)\Delta}) - \sigma^2(\bar{V}_k) = \sqrt{\Delta}(\sigma^2)'(V_{k\Delta})\sigma(V_{k\Delta})(\xi_k - \xi'_k) + \tilde{R}_k \tag{57}$$

with $\mathbb{E}(\tilde{R}^2_k) \leq c'\Delta^2$. Noticing that $\mathbb{E}[(\xi_k - \xi'_k)^2] = 1/3$, we get the result. □

8.4. **Proof of Proposition 4.4.** We only do the proof for [T]. Considering S_m in collection [T], we have to bound $\mathbb{E}\left(\sup_{t\in B_m(0,1)}[T_n(t)]^2\right)$ where $B_m(0,1) = \{t \in S_m, \|t\| = 1\}$. We shall use the following properties of the trigonometric basis and of collection [T] which can be checked by elementary computations.

$$\|\sum_{\lambda\in\Lambda_m}(\varphi^{(k)}_\lambda)^2\|_\infty \leq CD_m^{2k+1}, \quad \|t^{(k)}\|_\infty \leq CD_m^{k+1/2}\|t\|. \tag{58}$$

$$\|t'\| \leq CD_m\|t\|. \tag{59}$$

We use decomposition (57) to split $T_n(t)$ into

$$T_n(t) = \tilde{T}_n(t) - \mathbb{E}(\tilde{T}_n(t)) + \mathbb{E}(\tilde{T}_n(t)) + \frac{1}{n}\sum_{k=1}^n \tilde{R}_k t(\bar{V}_k)$$

with $\tilde{T}_n(t) = \frac{1}{n}\sum_{k=1}^n \sqrt{\Delta}[(\sigma^2)'\sigma](V_{k\Delta})(\xi_k - \xi'_k)t(\bar{V}_k)$. Using (58) for $k=0$, we get

$$\mathbb{E}\left[\sup_{t\in B_m(0,1)}\left(\frac{1}{n}\sum_{k=1}^n \tilde{R}_k t(\bar{V}_k)\right)^2\right] \le CD_m\left(\frac{1}{n}\sum_{k=1}^n \mathbb{E}(\tilde{R}_k^2)\right) \le C' D_m \Delta^2.$$

Then notice that

$$\mathbb{E}(\tilde{T}_n(t)) = \mathbb{E}\left(\frac{\sqrt{\Delta}}{n}\sum_{k=1}^n [(\sigma^2)'\sigma](V_{k\Delta})(\xi_k - \xi'_k)(t(\bar{V}_k) - t(V_{k\Delta}))\right).$$

Here, we have to use two derivatives of t. We use Gloter's decomposition again in order to write, as for (56), that

$$t(\bar{V}_k) - t(V_{k\Delta}) = \sqrt{\Delta} t'(V_{k\Delta})\sigma(V_{k\Delta})\xi'_k + e_k(t). \tag{60}$$

For any $t \in S_m$, (58) for $k=2$ implies $\mathbb{E}\left[\sup_{t\in B_m(0,1)} e_k^2(t)\right] \le C\Delta^2 D_m^5$. Thus, with (59), we obtain

$$\left(\sup_{t\in B_m(0,1)} \mathbb{E}[\tilde{T}_n(t)]\right)^2 \le K(D_m^2\Delta^2 + D_m^5\Delta^3). \tag{61}$$

Next we write

$$\tilde{T}_n(t) - \mathbb{E}(\tilde{T}_n(t)) = \tilde{T}_n^{(1)}(t) + \tilde{T}_n^{(2)}(t) - \mathbb{E}(\tilde{T}_n^{(2)}(t))$$

with a centered term

$$\tilde{T}_n^{(1)}(t) = \frac{1}{n}\sum_{k=1}^n \sqrt{\Delta}[(\sigma^2)'\sigma](V_{k\Delta})t(V_{k\Delta})(\xi_k - \xi'_k),$$

and the non centered term (already used above)

$$\tilde{T}_n^{(2)}(t) = \frac{1}{n}\sum_{k=1}^n \sqrt{\Delta}[(\sigma^2)'\sigma](V_{k\Delta})(\xi_k - \xi'_k)[t(\bar{V}_k) - t(V_{k\Delta})].$$

Using the Hölder inequality and the fact that $T_n^{(1)}(t)$ is a sum of uncorrelated variables, we see that

$$\begin{aligned}
\mathbb{E}\left\{\left(\sup_{t\in B_m(0,1)} \tilde{T}_n^{(1)}(t)\right)^2\right\} &\le \sum_{\lambda\in\Lambda_m} \mathbb{E}\left[\left(\tilde{T}_n^{(1)}(\varphi_\lambda)\right)^2\right] \\
&\le \frac{\Delta}{n}\sum_{\lambda\in\Lambda_m} \mathbb{E}[(\xi_1 - \xi'_1)^2]\mathbb{E}\{[(\sigma^2)'\sigma]^2(V_\Delta)\varphi_\lambda^2(V_\Delta)\} \\
&\le \frac{1}{3}\frac{CD_m\Delta}{n}\mathbb{E}\{[(\sigma^2)'\sigma]^2(V_\Delta\} := c\frac{\Delta D_m}{n}. \qquad (62)
\end{aligned}$$

Next, using (60), we introduce more terms:

$$\tilde{T}_n^{(2)}(t) = \tilde{T}_n^{(3)}(t) + \tilde{T}_n^{(4)}(t) + \tilde{T}_n^{(5)}(t)$$

with, since $\mathbb{E}[(\xi_k - \xi'_k)\xi'_k] = 1/6$,

$$\begin{aligned}
\tilde{T}_n^{(3)}(t) &= \frac{1}{n}\sum_{k=1}^{n} \Delta[(\sigma^2)'\sigma^2 t'](V_{k\Delta})[(\xi_k + \xi'_k)\xi'_k - \mathbb{E}[(\xi_k + \xi'_k)\xi'_k]] \\
\tilde{T}_n^{(4)}(t) &= \frac{\Delta}{6n}\sum_{k=1}^{n}[(\sigma^2)'\sigma^2 t'](V_{k\Delta}) \\
\tilde{T}_n^{(5)}(t) &= \frac{\sqrt{\Delta}}{n}\sum_{k=1}^{n}[(\sigma^2)'\sigma](V_{k\Delta})e_k(t).
\end{aligned}$$

The last term is bounded by

$$\mathbb{E}\left\{\left(\sup_{t\in B_m(0,1)}[\tilde{T}_n^{(5)}(t) - \mathbb{E}(\tilde{T}_n^{(5)}(t))]\right)^2\right\} \leq 4c\mathbb{E}^{1/2}\{[(\sigma^2)'\sigma]^4(V_\Delta)\}\Delta^3 D_m^5 = c'\Delta^3 D_m^5 \tag{63}$$

Moreover $\mathbb{E}(\tilde{T}_n^{(3)}(t)) = 0$ and by using (58) with $k = 1$, it follows that

$$\begin{aligned}
\mathbb{E}\left\{\left(\sup_{t\in B_m(0,1)}\tilde{T}_n^{(3)}(t)\right)^2\right\} &\leq \sum_{\lambda\in\Lambda_m}\mathbb{E}\{[\tilde{T}_n^{(3)}(\varphi_\lambda)]^2\} \\
&\leq \frac{c_\xi\Delta^2}{n}\sum_{\lambda\in\Lambda_m}\mathbb{E}\{[(\sigma^2)'\sigma^2]^2(V_\Delta)(\varphi'_\lambda)^2(V_\Delta)\} \\
&\leq \frac{CD_m^3\Delta^2}{n}\mathbb{E}\{[(\sigma^2)'\sigma^2]^2(V_\Delta)\} := c"\frac{D_m^3\Delta^2}{n}, \qquad (64)
\end{aligned}$$

where $c_\xi = \mathbb{E}[(\xi_1 - \xi'_1)^2(\xi'_1)^2]$. For the last term, we apply Viennet's mixing covariance inequality (see Theorem 2.1 p.472 and Lemma 4.2 p.481 in Viennet (1997)). There exists a function $b_\Delta^{(V)}$ such that

$$\begin{aligned}
&\mathbb{E}\left\{\left(\sup_{t\in B_m(0,1)}[\tilde{T}_n^{(4)}(t) - \mathbb{E}(\tilde{T}_n^{(4)}(t)]\right)^2\right\} \\
&\leq \Delta^2\sum_{\lambda\in\Lambda_m}\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n}[(\sigma^2)'\sigma^2\varphi'_\lambda](V_{k\Delta})\right) \\
&\leq \frac{4\Delta^2}{n}\sum_{\lambda\in\Lambda_m}\int(\varphi'_\lambda)^2(v)[(\sigma^2)'\sigma^2]^2(v)b_\Delta^{(V)}(v)d\mathbb{P}_{V_\Delta}(v) \\
&\leq \frac{4\Delta^2 CD_m^3}{n}\sqrt{\sum_k k\beta_V(k\Delta)}\mathbb{E}^{1/2}\{[(\sigma^2)'\sigma^2]^4(V_\Delta)\} := c_3\frac{\Delta D_m^3}{n}. \qquad (65)
\end{aligned}$$

It follows from (61), (62), (63), (64) and (65) that

$$\mathbb{E}\left\{\left(\sup_{t\in S_m,\|t\|=1}T_n(t)\right)^2\right\} \leq C\left(D_m^2\Delta^2 + \frac{\Delta D_m}{n} + \Delta^3 D_m^5 + \frac{D_m^3\Delta^2}{n} + \frac{\Delta D_m^3}{n}\right).$$

Since $\Delta \leq 1$, $D_m^3\Delta^2/n \leq D_m^3\Delta/n$. We have $\Delta D_m/n \leq \Delta D_m^3/n$. Using that $D_m \leq N_n \leq \sqrt{n\Delta} \leq n\Delta$, we get $\Delta D_m^3/n \leq \Delta^2 D_m^2$. This implies (31). If $\Delta \leq n^{-2/3}$, replacing D_m by N_n in the right-hand side of (31), we obtain that $N_n^2\Delta^2 + N_n^5\Delta^3 \leq c/n$ and (32) follows.

For [W], since the constraint on N_n is different ($N_n \leq n\Delta/\ln^2(n)$), we get (32) for $\Delta \leq n^{-3/4}$. □

8.5. **Proof of Proposition 4.6.** First we note that:

$$\begin{aligned}\mathbb{E}_n(u) := \mathbb{E}\left(e^{ut(\bar{V}_n)\breve{Z}^{(1)}_{(n+1)\Delta}}|\mathcal{F}_{(n+1)\Delta}\right) &= 1+\sum_{p=2}^{+\infty}\frac{u^p}{p!}\mathbb{E}\left[(t(\bar{V}_n)\breve{Z}^{(1)}_{(n+1)\Delta})^p|\mathcal{F}_{(n+1)\Delta}\right] \\ &\leq 1+\sum_{p=2}^{+\infty}\frac{u^p}{p!}|t(\bar{V}_n)|^p\mathbb{E}\left[|\breve{Z}^{(1)}_{(n+1)\Delta}|^p|\mathcal{F}_{(n+1)\Delta}\right].\end{aligned}$$

Next we use the Burkholder-Davis-Gundy inequality given in Proposition 4.2 of Barlow and Yor (1982), with optimal constant $c\sqrt{k}$: for a continuous martingale M_t, $M_0 = 0$, and $M_t^* = \sup_{s\leq t}|M_s|$, for $k \geq 2$, there exists a universal constant c such that $\|M_t^*\|_k \leq c\sqrt{k}\|\langle M\rangle_t^{1/2}\|_k$. This yields:

$$\begin{aligned}\mathbb{E}(|\breve{Z}^{(1)}_{n\Delta}|^p|\mathcal{F}_{n\Delta}) &\leq \frac{3^{p-1}}{2^p\Delta^{3p}}\left\{\mathbb{E}\left(\left|\int_{n\Delta}^{(n+2)\Delta}\psi_{n\Delta}(s)\sigma(V_s)dW_s\right|^{2p}|\mathcal{F}_{n\Delta}\right)\right. \\ &\quad \left.+\mathbb{E}\left(\left|\int_{n\Delta}^{(n+2)\Delta}\psi^2_{n\Delta}(s)\sigma^2(V_s)ds\right|^{p}|\mathcal{F}_{n\Delta}\right)\right\} \\ &\leq \frac{2^{p-1}}{\Delta^p}(c^{2p}(8p)^p\Delta^{3p}\sigma_1^{2p}+\Delta^{3p}(2\sigma_1)^{2p}) \leq (8\sigma_1 c)^{2p}p^p.\end{aligned}$$

It follows that $\mathbb{E}_n(u) \leq 1+\sum_{k=2}^{\infty}\frac{p^p}{p!}(u\sigma_1^2\tilde{c}^2)^p|t(\bar{V}_n)|^p$. Since $p^p/p! \leq e^{p-1}$, we find

$$\mathbb{E}_n(u) \leq 1+e^{-1}\sum_{k=2}^{\infty}(u\sigma_1^2\tilde{c}^2e)^p|t(\bar{V}_n)|^p \leq 1+e^{-1}\frac{(u\sigma_1^2\tilde{c}^2e)^2t^2(\bar{V}_n)}{1-(u\sigma_1^2\tilde{c}^2e\|t\|_\infty)}.$$

Let us set $a = e(\sigma_1^2\tilde{c}^2)^2$ and $b = \sigma_1^2\tilde{c}^2e\|t\|_\infty$. Since for $x \geq 0$, $1+x \leq e^x$, for $bu < 1$,

$$\mathbb{E}_n(u) \leq 1+\frac{au^2t^2(\bar{V}_n)}{1-bu} \leq \exp\left(\frac{au^2t^2(\bar{V}_n)}{1-bu}\right).$$

This can also be written:

$$\mathbb{E}\left(\exp\left(ut(\bar{V}_n)\breve{Z}^{(1)}_{(n+1)\Delta}-\frac{au^2t^2(\bar{V}_n)}{1-bu}\right)|\mathcal{F}_{(n+1)\Delta}\right) \leq 1.$$

Therefore, by iterative conditioning

$$\mathbb{E}\left\{\exp\left[\sum_{k=1}^{n}\left(ut(\bar{V}_k)\breve{Z}^{(1)}_{(k+1)\Delta}-\frac{au^2t^2(\bar{V}_k)}{1-bu}\right)\right]\right\} \leq 1.$$

Then, by using a standard method,

$$\begin{aligned}
&\mathbb{P}\left(\sum_{k=1}^{n} t(\bar{V}_k)\breve{Z}^{(1)}_{(k+1)\Delta} \geq n\epsilon, \|t\|_n^2 \leq v^2\right) \leq e^{-nu\epsilon}\mathbb{E}\left[\mathbb{1}_{\|t\|_n^2\leq v^2}\exp\left(u\sum_{k=1}^{n} t(\bar{V}_k)\breve{Z}^{(1)}_{(k+1)\Delta}\right)\right] \\
\leq\ & e^{-nu\epsilon}\mathbb{E}\left[\mathbb{1}_{\|t\|_n^2\leq v^2}\exp\left(\sum_{k=1}^{n}(ut(\bar{V}_k)\breve{Z}^{(1)}_{(k+1)\Delta} - \frac{au^2t^2(\bar{V}_k)}{1-bu})\right)e^{(au^2)/(1-bu)\sum_{k=1}^{n}t^2(\bar{V}_k)}\right] \\
\leq\ & e^{-nu\epsilon}e^{(nau^2v^2)/(1-bu)}\mathbb{E}\left[\exp\left(\sum_{k=1}^{n}(ut(\bar{V}_k)\breve{Z}^{(1)}_{(k+1)\Delta} - \frac{au^2t^2(\bar{V}_k)}{1-bu})\right)\right] \\
\leq\ & e^{-nu\epsilon}e^{(nau^2v^2)/(1-bu)}
\end{aligned}$$

The inequality holds for any u such that $bu < 1$. In particular, $u = \epsilon/(2av^2 + \epsilon b)$ gives $-u\epsilon + av^2u^2/(1-bu) = -(1/2)(\epsilon^2/(2av^2+\epsilon b)$ and therefore

$$\mathbb{P}\left(\sum_{k=1}^{n} t(\bar{V}_k)\breve{Z}^{(1)}_{(k+1)\Delta} \geq n\epsilon, \|t\|_n^2 \leq v^2\right) \leq \exp\left(-n\frac{\epsilon^2/2}{2av^2+\epsilon b}\right). \ \square$$

8.6. **Proof of Theorem 4.1.** We proceed as in Theorem 3.1. We first give the proof for collection [T]. We start from (36), with here $\gamma_n(\hat{\sigma}^2_{\breve{m}}) + \mathrm{pen}(\breve{m}) \leq \gamma_n(\sigma^2_m) + \mathrm{pen}(m)$, for all m in $\mathcal{M}_n$. We recall that Ω_n is defined by (52). Then, using that on Ω_n, $\|\hat{\sigma}^2_{\breve{m}} - \sigma^2_m\|^2_{\bar{\pi}} \leq 2\|\hat{\sigma}^2_{\breve{m}} - \sigma^2_m\|^2_n$, we find

$$\begin{aligned}
\|\hat{\sigma}^2_{\breve{m}} - \sigma^2_A\|^2_n \leq\ & \|\sigma^2_m - \sigma^2_A\|^2_n + \frac{1}{8}\|\hat{\sigma}^2_{\breve{m}} - \sigma^2_m\|^2_{\bar{\pi}} + 16\sup_{t\in S_{\breve{m}}+S_m, \|t\|_{\bar{\pi}}=1}\breve{\nu}^2_n(t) \\
&+16\sup_{t\in S_m+S_{\breve{m}}, \|t\|_{\bar{\pi}}=1}[T_n(t)]^2 \\
&+\frac{1}{8}\|\hat{\sigma}^2_{\breve{m}} - \hat{\sigma}^2_m\|^2_n + \frac{8}{n}\sum_{k=1}^{n}\breve{R}^2_{(k+1)\Delta} + \mathrm{p\breve{e}n}(m) - \mathrm{p\breve{e}n}(\breve{m}) \\
\leq\ & \|\sigma^2_m - \sigma^2_A\|^2_n + \frac{3}{8}\|\hat{\sigma}^2_{\breve{m}} - \sigma^2_m\|^2_n + 16\sup_{t\in S_{\breve{m}}+S_m, \|t\|_{\bar{\pi}}=1}\breve{\nu}^2_n(t) \\
&+16\sup_{t\in S_{\breve{m}}+S_m, \|t\|_{\bar{\pi}}=1}[T_n(t)]^2 + \frac{8}{n}\sum_{k=1}^{n}\breve{R}^2_{(k+1)\Delta} + \mathrm{p\breve{e}n}(m) - \mathrm{p\breve{e}n}(\breve{m}),
\end{aligned}$$

where $\breve{\nu}_n(t)$ is defined by (35) and $T_n(t)$ by (30). This yields on Ω_n and denoting by $B^{\bar{\pi}}_{m,m'}(0,1) = \{t \in S_m + S_{m'}, \|t\|_{\bar{\pi}} = 1\}$,$B_{m,m'}(0,1) = \{t \in S_m + S_{m'}, \|t\| = 1\}$,

$$\begin{aligned}
\frac{1}{4}\|\hat{\sigma}^2_{\breve{m}} - \sigma^2_A\|^2_n \leq\ & \frac{7}{4}\|\sigma^2_m - \sigma^2_A\|^2_n + 16\sup_{t\in B^{\bar{\pi}}_{m,\breve{m}}(0,1)}\breve{\nu}^2_n(t) + \frac{16}{\bar{\pi}_0}\sup_{t\in B_{m,\breve{m}}(0,1)}T^2_n(t) \\
&+\frac{8}{n}\sum_{k=1}^{n}\breve{R}^2_{(k+1)\Delta} + \mathrm{p\breve{e}n}(m) - \mathrm{p\breve{e}n}(\breve{m}),
\end{aligned}$$

by using (8). First Lemma 4.2 implies that $\mathbb{E}(n^{-1}\sum_{k=1}^{n}\breve{R}^2_{(k+1)\Delta}) \leq c\Delta^2$.

Then we have

$$\mathbb{E}(\|\hat{\sigma}^2_{\check{m}} - \sigma_A^2\|_n^2 \mathbb{I}_{\Omega_n}) \leq 7\bar{\pi}_1 \|\sigma^2_m - \sigma_A^2\|^2 + 4\text{pĕn}(m) + 64\mathbb{E}\left(\sup_{t\in B^{\bar{\pi}}_{m,\check{m}}(0,1)} [\check{\nu}_n(t)]^2 \mathbb{I}_{\Omega_n}\right)$$

$$-4\mathbb{E}[\text{pĕn}(\check{m})] + \frac{64}{\bar{\pi}_0}\mathbb{E}\left(\sup_{t\in B_{m,\check{m}}(0,1)} [T_n(t)]^2 \mathbb{I}_{\Omega_n}\right) + K'\Delta^2. \tag{66}$$

Then we can use Proposition 4.4 to obtain

$$\mathbb{E}\left\{\left(\sup_{t\in B_{m,\check{m}}(0,1)} T_n(t)\right)^2\right\} \leq \mathbb{E}\left\{\left(\sup_{t\in \mathcal{S}_n, \|t\|=1} T_n(t)\right)^2\right\} \leq \frac{c}{n}.$$

Let us set

$$\check{\nu}_n^{(1)}(t) = \frac{1}{n}\sum_{k=1}^n t(\bar{V}_k)\check{Z}^{(1)}_{(k+1)\Delta}, \quad \check{\nu}_n^{(2)}(t) = \frac{1}{n}\sum_{k=1}^n t(\bar{V}_k)(\check{Z}^{(2)}_{(k+1)\Delta} + \check{Z}^{(3)}_{(k+1)\Delta}), \tag{67}$$

and

$$\check{G}_m(m') = \sup_{t\in S_m + S_{m'}, \|t\|_{\bar{\pi}}=1} \check{\nu}_n^{(1)}(t). \tag{68}$$

If we write, as for the drift

$$\begin{aligned}\mathbb{E}(\check{G}^2_m(\check{m})) &\leq \mathbb{E}[(\check{G}^2_m(\check{m}) - \check{p}(m,\check{m}))\mathbb{I}_{\Omega_n}]_+ + \mathbb{E}(\check{p}(m,\check{m})) \\ &\leq \sum_{m'\in\mathcal{M}_n} \mathbb{E}[(\check{G}^2_m(m') - \check{p}(m,m'))\mathbb{I}_{\Omega_n}]_+ + \mathbb{E}(\check{p}(m,\check{m})),\end{aligned}$$

then pĕn is chosen such that $64\check{p}(m,m') \leq 4[\text{pĕn}(m) + \text{pĕn}(m')]$. More precisely, we can prove

Proposition 8.1. *Under the assumptions of Theorem 4.1, for $\check{p}(m,m') = \kappa\sigma_1^4[D_m + (1 + L_{m'}D_{m'})]/n + K\Delta^2$, where κ is a numerical constant and K is a constant depending on the collection of models and on $\bar{\pi}_0$, we have*

$$\mathbb{E}[(\check{G}^2_m(m') - \check{p}(m,m'))\mathbb{I}_{\Omega_n}]_+ \leq c\sigma_1^4 \frac{e^{-L_{m'}D_{m'}}}{n},$$

where $\check{G}_m(m')$ is defined by (68).

The proof of the result is given in appendix.
Then, we use (66) and Proposition 8.1, choose $\text{pĕn}(m) \geq \tilde{\kappa}\sigma_1^4 \frac{(1+L_m)D_m}{n}$ and recall that $\Sigma = \sum_{m\in\mathcal{M}_n} e^{-L_m D_m}$. This yields

$$\begin{aligned}\mathbb{E}(\|\hat{\sigma}^2_{\check{m}} - \sigma_A^2\|_n^2) &\leq 7\bar{\pi}_1\|\sigma^2_m - \sigma_A^2\|^2 + 8\text{pĕn}(m) + c\sigma_1^4\frac{\Sigma}{n} + K'\Delta^2 \\ &\quad + 64\mathbb{E}\left(\sup_{t\in B^{\bar{\pi}}_{m,\check{m}}(0,1)} [\check{\nu}_n^{(2)}(t)]^2\right) + \frac{K"}{n} + \mathbb{E}(\|\hat{\sigma}^2_{\check{m}} - \sigma_A^2\|_n^2 \mathbb{I}_{\Omega_n^c}).\end{aligned} \tag{69}$$

For the last term above, the bound:

$$\mathbb{E}(\|\hat{\sigma}^2_{\check{m}} - \sigma_A^2\|_n^2 \mathbb{I}_{\Omega_n^c}) \leq \frac{c}{n} \tag{70}$$

is obtained in the same way as in the end of the proof of Theorem 3.1, by introducing the regression model $U_{k+1} = \sigma^2(\bar{V}_k) + \eta_{(k+1)\Delta}$, where, now, $\eta_{(k+1)\Delta} = \sigma^2(V_{(k+1)\Delta}) - \sigma^2(\bar{V}_k) + \breve{Z}_{(k+1)\Delta} + \breve{R}_{(k+1)\Delta}$. We can bound $\mathbb{E}(\eta_{k\Delta}^4)$ with a bound independent of k and we know that $\mathbb{P}(\Omega_n^c) \leq c/n^2$.
Moreover, let $\tilde{Z}_{k\Delta} = \breve{Z}_{k\Delta}^{(2)} + \breve{Z}_{k\Delta}^{(3)}$. To study the term involving $\breve{\nu}_n^{(2)}(t)$, we proceed as in Proposition 3.2. For all $m \in \mathcal{M}_n$,

$$\mathbb{E}\left(\sup_{t\in S_m, \|t\|_{\bar{\pi}}=1} [\breve{\nu}_n^{(2)}(t)]^2\right) \leq \frac{1}{\bar{\pi}_0}\mathbb{E}\left(\sup_{t\in B_m(0,1)} [\breve{\nu}_n^{(2)}(t)]^2\right) \leq \frac{1}{\bar{\pi}_0}\sum_{\lambda\in\Lambda_m} \mathbb{E}\left([\breve{\nu}_n^{(2)}(\varphi_\lambda)]^2\right)$$

Then, as in Proposition 3.2 and using Proposition 4.1, we obtain

$$\sum_{\lambda\in\Lambda_m} \mathbb{E}\left([\breve{\nu}_n^{(2)}(\varphi_\lambda)]^2\right) \leq \frac{2}{n}\sum_{\lambda\in\Lambda_m} \mathbb{E}(\varphi_\lambda^2(\bar{V}_1)\tilde{Z}_{2\Delta}^2) \leq c\frac{D_m}{n}\Delta$$

Therefore, since the spaces are all contained in the maximal space $\mathcal{S}_n$ which has dimension $N_n \leq n\Delta/\ln^2(n)$, we have

$$\mathbb{E}\left(\sup_{t\in B_{m,\breve{m}}^{\bar{\pi}}(0,1)} [\breve{\nu}_n^{(2)}(t)]^2\right) \leq \frac{1}{\bar{\pi}_0}\mathbb{E}\left(\sup_{t\in\mathcal{S}_n, \|t\|=1} [\breve{\nu}_n^{(2)}(t)]^2\right) \leq C\frac{\Delta N_n}{n} \leq K\Delta^2 \leq \frac{K'}{n}, \tag{71}$$

since $\Delta \leq n^{-2/3}$. The result of Theorem 4.1 follows by gathering (69), (70) and (71).

For collections [DP], [GP], the proof is analogous except that there is no $T_n(t)$. Instead, $\breve{R}((k+1)\Delta)$ is changed into $\breve{R}_{\sigma^2}((k+1)\Delta) = \breve{R}_{(k+1)\Delta} + [\sigma^2(V_{(k+1)\Delta}) - \sigma^2(\bar{V}_k)]$. Then it follows from Propositions 4.2 and 4.3 that $\mathbb{E}\left[(1/n)\sum_{k=1}^n \breve{R}_{\sigma^2}^2((k+1)\Delta)\right] \leq c\Delta$. □

9. Proofs of the Propositions of Section 5

Proof of Proposition 5.1. Consider the diffusion process $(V_t^{v_0})$ given by $dV_t = b(V_t)dt + \sigma(V_t)dW_t, V_0 = v_0$ and set $x_u = \Delta^{-(1/2)}(V_{u\Delta}^{v_0} - v_0), u \in [0,1]$. Then, (x_u) is solution of

$$dx_u = \tilde{b}(x_u)du + \tilde{\sigma}(x_u)d\tilde{W}_u, \quad x_0 = 0,$$

with $\tilde{b}(x) = \Delta^{1/2}b(x\Delta^{1/2} + v_0)$, $\tilde{\sigma}(x) = \sigma(x\Delta^{1/2} + v_0)$ and $(\tilde{W}_u)$ is a standard Brownian motion. Then, setting

$$U = \int_0^1 x_u du, \quad V = x_1,$$

we have

$$(1/\Delta)\int_0^\Delta V_s^{v_0} ds = v_0 + \Delta^{1/2}U, \quad V_\Delta^{v_0} = v_0 + \Delta^{1/2}V.$$

Now, the following result is proved in Gloter and Gobet (2005, Theorem 3). The random couple (U, V) has a joint density $p_\Delta^{v_0}(u,v)$ such that

$$c_1^{-1}\exp\left(-c_1(u^2+v^2)\right) \leq p_\Delta^{v_0}(u,v) \leq c_2^{-1}\exp\left(-c_2(u^2+v^2)\right)$$

where the constants c_1, c_2 only depend on the bounds of b, σ and their derivatives. Consequently, the marginal density of U, say $p^{v_0}_{U,\Delta}(u)$ satisfies

$$c'_1 \exp(-c_1 u^2) \leq p^{v_0}_{U,\Delta}(u) \leq c'_2 \exp(-c_2 u^2), \tag{72}$$

with $c'_i = c_i^{-1}(\pi/c_i)^{1/2}, i = 1, 2$. After an elementary change of variable, we get that the conditional density of $\bar{V}_0$ given $V_0 = v_0$, which is exactly the density of $v_0 + \Delta^{1/2}U$, is equal to

$$\bar{v} \to \frac{1}{\Delta^{1/2}} p^{v_0}_{U,\Delta}((\bar{v} - v_0)/\Delta^{1/2}).$$

The density $\bar{\pi}_\Delta$ is obtained by integrating the above density with respect to $\pi(v_0)dv_0$. Using the bounds (72), we obtain

$$c'_1 \int_{\mathbb{R}} \exp(-c_1 t^2)\, \pi(\bar{v} + t\Delta^{1/2})dt \leq \bar{\pi}_\Delta(\bar{v}) \leq c'_2 \int_{\mathbb{R}} \exp(-c_2 t^2)\, \pi(\bar{v} + t\Delta^{1/2})dt \tag{73}$$

The stationary density $\pi(.)$ is bounded and this gives an upper bound for $\bar{\pi}_\Delta$. Using (73), we have, for all $t_0 > 0$,

$$\bar{\pi}_\Delta(\bar{v}) \geq c'_1 \int_0^{t_0} \exp(-c_1 t^2)\, \pi(\bar{v} + t\Delta^{1/2})dt.$$

Hence, for all $\bar{v} \in [a, b]$, $\bar{\pi}_\Delta(\bar{v}) \geq C' \inf_{u\in[a,b+t_0]} \pi(u)$ for some constant C'. This gives the result. □

Proof of Proposition 5.2. Let Q^{δ,θ,c^2}_v be the distribution on $C(\mathbb{R}^+, \mathbb{R})$ of (43) starting with $V_0 = v$. Then,

$$Q^{\delta,\theta,c^2}_v * Q^{\delta',\theta,c^2}_{v'} = Q^{\delta+\delta',\theta,c^2}_{v+v'}.$$

This property is obtained exactly as the analogous proof for the square of a δ-dimensional Bessel process which corresponds to $\theta = 0$ (see *e.g.* Revuz and Yor, 2005, p. 440). From this property, it follows analogously that, for $\lambda > 0$,

$$\mathbb{E}(\exp(-\lambda \int_0^t V_s ds))|V_0 = v) = B_t(\lambda)^\delta A_t(\lambda)^v, \tag{74}$$

where $0 < B_t(\lambda), A_t(\lambda) < 1$ have to be computed. Now, we set $A_t(\lambda) = \exp(-\mu_t(\lambda))$ with $\mu_t(\lambda > 0$. Since $V_0 = \eta$ has the stationary distribution $G(\delta/2, 1/2\rho^2)$, we have, for all $\mu > 0$,

$$\mathbb{E}(\exp(-\mu V_0)) = (1 + 2\mu\rho^2)^{-(\delta/2)}. \tag{75}$$

Hence, integrating (74) with respect to the distribution of V_0, we get (44).

This preliminaries show that it is enough make computations for $\delta = 1$, *i.e.* for $V_t = \xi_t^2$ where (ξ_t) is an Ornstein-Uhlenbeck process. Denote by P^θ_x the distribution on $C(\mathbb{R}^+, \mathbb{R})$ of (ξ_t) given by

$$d\xi_t = -\theta\xi_t dt + c dW_t, \xi_0 = x.$$

And denote by (X_t) the canonical coordinate process of $C(\mathbb{R}^+, \mathbb{R})$. For any real number $\bar{c}$, we have, by the Girsanov formula,

$$\mathbb{E}((\exp(-\lambda \int_0^t \xi_s^2 ds)) = E_{P^\theta_x}((\exp(-\lambda \int_0^t X_s^2 ds)) = E_{P^{\bar{c}}_x}((\exp(-\lambda \int_0^t X_s^2 ds))L_t)$$

where

$$L_t = \exp \frac{-\theta - \bar{c}}{c^2} \int_0^t X_s dX_s - \frac{\theta^2 - \bar{c}^2}{2c^2} \int_0^t X_s^2 ds.$$

We have $\int_0^t X_s dX_s = (1/2)(X_t^2 - x^2 - c^2 t)$ and we choose $\bar{c}$ such that $\lambda + \frac{\theta^2 - \bar{c}^2}{2c^2} = 0$. This yields

$$E_{P_x^\theta}((\exp(-\lambda \int_0^t X_s^2 ds) = \exp \frac{\bar{c} + \theta}{2c^2}(x^2 + c^2 t) E_{P_x^{\bar{c}}}(\exp \frac{-\theta - \bar{c}}{2c^2} X_t^2). \tag{76}$$

The choice $\bar{c} = (\theta^2 + 2\lambda c^2)^{1/2}$ implies that $\bar{c} + \theta > 0$. Therefore, we easily compute the above expectation since, under $P_x^{\bar{c}}$, X_t is Gaussian with mean $x \exp(\bar{c}t)$ and variance $c^2(\exp(2\bar{c}t) - 1)/2\bar{c}$. Indeed, for X a Gaussian variable with law $\mathcal{N}(m, \beta^2)$ and $\mu > 0$,

$$\mathbb{E}(\exp -\mu X^2) = (1 + 2\mu\beta^2)^{-(1/2)} \exp(-\frac{\mu m^2}{1 + 2\mu\beta^2}).$$

From this, we deduce $B_t(\lambda)$ and $\mu_t(\lambda)$.□

Proof of Corollary 5.1. The Laplace transform is well defined for all λ such that $\theta^2 + 2\lambda c^2 \geq 0$. So it is well defined on an open interval containing 0. By properties of complex functions, this is enough to prove that we obtain the characteristic function of $\int_0^t V_s ds$ by setting $\lambda = -is$ with $s \in \mathbb{R}$ in (44). The corollary follows. □

Proof of Proposition 5.3. First write that

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

where the integrals and derivatives must be understood piecewisely. Now we use that for any $t \in S_m$, there exists some constant C such that

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

Moreover,

$$\mathbb{E}[(\bar{V}_0 - V_0)(t^2)'(V_0)] = \mathbb{E}\left[\mathbb{E}\left(\bar{V}_0 - V_0 | \mathcal{F}_0\right)(t^2)'(V_0)\right]$$

and

$$\begin{aligned} \mathbb{E}\left(\bar{V}_0 - V_0 | \mathcal{F}_0\right) &= \mathbb{E}\left(\frac{1}{\Delta}\int_0^\Delta (\int_0^s (b(V_u)du + \sigma(V_u)dW_u))ds | \mathcal{F}_0\right) \\ &= \mathbb{E}\left(\frac{1}{\Delta}\int_0^\Delta (\int_0^s b(V_u)du)ds | \mathcal{F}_0\right). \end{aligned}$$

It follows that, using (5), $|\mathbb{E}\left(\bar{V}_0 - V_0 | \mathcal{F}_0\right)| = O(\Delta)$. Thus, $|\mathbb{E}[(\bar{V}_0 - V_0)(t^2)'(V_0)]| \leq C N_n^2 \Delta \|t\|^2 = O(N_n^2 \Delta)$. On the other hand,

$$\begin{aligned} \left|\mathbb{E}\left[(\bar{V}_0 - V_0)^2 \int_0^1 (t^2)"(V_0 + u(\bar{V}_0 - V_0))du\right]\right| &\leq \|(t^2)"\|_\infty \mathbb{E}[(\bar{V}_0 - V_0)^2] \\ &\leq C N_n^3 \Delta \|t\|^2 = O(N_n^3 \Delta). \end{aligned}$$

This implies a global order $N_n^3 \Delta$. □

10. Appendix. Proof of Proposition 8.1

The result of Proposition 8.1 is obtained from inequality (38) of Proposition 4.6 by a $L^2(\bar{\pi}_\Delta) - L^\infty$ chaining technique. The method is analogous to the one given in Proposition 2-4 pp.282-287 in Comte (2001), in Theorem 5 in Birgé and Massart (1998) and in Proposition 7, Theorem 8 and Theorem 9 in Barron et al. (1999). Since the context is slightly different, for the sake of completeness, we give the detail of the proof. It relies on the following Lemma (Lemma 9 in Barron *et al.* (1999)):

Lemma 10.1. *Let μ be a positive measure on $[0,1]$. Let $(\psi_\lambda)_{\lambda\in\Lambda}$ be a finite orthonormal system in $\mathbb{L}_2 \cap \mathbb{L}_\infty(\mu)$ with $|\Lambda| = D$ and $\bar{S}$ be the linear span of $\{\psi_\lambda\}$. Let*

$$\bar{r} = \frac{1}{\sqrt{D}} \sup_{\beta\neq 0} \frac{\| \sum_{\lambda\in\Lambda} \beta_\lambda \psi_\lambda \|_\infty}{|\beta|_\infty}. \tag{77}$$

For any positive δ, one can find a countable set $T \subset \bar{S}$ and a mapping p from $\bar{S}$ to T with the following properties:

- *for any ball $\mathcal{B}$ with radius $\sigma \geq 5\delta$,*

$$|T \cap \mathcal{B}| \leq (B'\sigma/\delta)^D \quad \text{with} \quad B' < 5.$$

- *$\|u - p(u)\|_\mu \leq \delta$ for all u in $\bar{S}$, and*

$$\sup_{u\in p^{-1}(t)} \|u - t\|_\infty \leq \bar{r}\delta, \quad \text{for all } t \text{ in } T.$$

To use this Lemma, the main difficulty is often to evaluate $\bar{r}$ in the different contexts. In our problem, the measure μ is $\bar{\pi}_\Delta$. We consider a collection of models $(S_m)_{m\in\mathcal{M}_n}$ which can be [DP], [GP] or [T]. Recall that $B^{\bar{\pi}}_{m,m'}(0,1) = \{t \in S_m + S_{m'}, \|t\|_{\bar{\pi}} = 1\}$. We have to compute $\bar{r} = \bar{r}_{m,m'}$ corresponding to $\bar{S} = S_m + S_{m'}$. We denote by $D(m,m') = \dim(S_m + S_{m'})$.
Collection [DP]– $S_m + S_{m'} = S_{\max(m,m')}$, $D(m,m') = \max(D_m, D_{m'})$, an orthonormal $\mathbb{L}_2(\bar{\pi}_\Delta)$-basis $(\psi_\lambda)_{\lambda\in\Lambda(m,m')}$ can be built by orthonormalisation on each sub-interval of $(\varphi_\lambda)_{\lambda\in\Lambda(m,m')}$. Then

$$\begin{aligned}
\sup_{\beta\neq 0} \frac{\| \sum_{\lambda\in\Lambda(m,m')} \beta_\lambda \psi_\lambda \|_\infty}{|\beta|_\infty} &\leq \| \sum_{\lambda\in\Lambda(m,m')} |\psi_\lambda| \|_\infty \leq (r+1) \sup_{\lambda\in\Lambda(m,m')} \|\psi_\lambda\|_\infty \\
&\leq (r+1)^{3/2} \sqrt{D(m,m')} \sup_{\lambda\in\Lambda(m,m')} \|\psi_\lambda\| \\
&\leq (r+1)^{3/2} \sqrt{D(m,m')} \sup_{\lambda\in\Lambda(m,m')} \|\psi_\lambda\|_{\bar{\pi}} / \sqrt{\bar{\pi}_0} \\
&\leq (r+1)^{3/2} \sqrt{D(m,m')/\bar{\pi}_0}.
\end{aligned}$$

Thus here $\bar{r}_{m,m'} \leq (r+1)^{3/2}/\sqrt{\bar{\pi}_0}$.
Collection [GP]– Here we have $\bar{r}_{m,m'} \leq [(R_{max}+1)\sqrt{N_n}]/\sqrt{D(m,m')\bar{\pi}_0}$.

Collection [T]– For trigonometric polynomials, we write

$$\begin{aligned}\sup_{\beta\neq 0}\frac{\|\sum_{\lambda\in\Lambda(m,m')}\beta_\lambda\psi_\lambda\|_\infty}{|\beta|_\infty} &\leq \frac{C\sqrt{D(m,m')}\|\sum_\lambda\beta_\lambda\psi_\lambda\|}{|\beta|_\infty}\leq\frac{C\sqrt{D(m,m')}\|\sum_\lambda\beta_\lambda\psi_\lambda\|_{\bar\pi}}{\sqrt{\bar\pi_0}|\beta|_\infty}\\ &\leq \frac{C\sqrt{D(m,m')}\sqrt{\sum_\lambda\beta_\lambda^2}}{\sqrt{\bar\pi_0}|\beta|_\infty}\leq\frac{CD(m,m')}{\sqrt{\bar\pi_0}}.\end{aligned}$$

Therefore, $\bar r_{m,m'}\leq C\sqrt{D(m,m')/\bar\pi_0}$.

We may now prove Proposition 8.1. We apply Lemma 10.1 to the linear space $S_m+S_{m'}$ of dimension $D(m,m')$ and norm connection measured by $\bar r_{m,m'}$ bounded above. We consider δ_k-nets, $T_k=T_{\delta_k}\cap B^{\bar\pi}_{m,m'}(0,1)$, with $\delta_k=\delta_0 2^{-k}$ with $\delta_0\leq 1/5$, to be chosen later and we set $H_k=\ln(|T_k|)\leq D(m,m')\ln(5/\delta_k)=D(m,m')[k\ln(2)+\ln(5/\delta_0)]$. Given some point $u\in B^{\bar\pi}_{m,m'}(0,1)$, we can find a sequence $\{u_k\}_{k\geq 0}$ with $u_k\in T_k$ such that $\|u-u_k\|^2_{\bar\pi}\leq\delta_k^2$ and $\|u-u_k\|_\infty\leq\bar r_{m,m'}\delta_k$. Thus we have the following decomposition that holds for any $u\in B^{\bar\pi}_{m,m'}(0,1)$,

$$u=u_0+\sum_{k=1}^{\infty}(u_k-u_{k-1}).$$

Clearly $\|u_0\|_{\bar\pi}\leq 1$, $\|u_0\|_\infty\leq\bar r_{(m,m')}$ and for all $k\geq 1$, $\|u_k-u_{k-1}\|^2_{\bar\pi}\leq 2(\delta_k^2+\delta_{k-1}^2)=5\delta_{k-1}^2/2$ and $\|u_k-u_{k-1}\|_\infty\leq 3\bar r_{(m,m')}\delta_{k-1}/2$. In the sequel we denote by $\mathbb{P}_n(.)$ the measure $\mathbb{P}(.\cap\Omega_n)$, see (52), (actually only the inequality $\|t\|_n^2\leq\frac{3}{2}\|t\|^2_{\bar\pi}$ holding for any $t\in S_m+S_{m'}$ is required).
Let $(\eta_k)_{k\geq 0}$ be a sequence of positive numbers that will be chosen later on and η such that $\eta_0+\sum_{k\geq 1}\eta_k\leq\eta$. Recall that $\breve\nu_n^{(1)}$ is defined by (67). We have

$$\begin{aligned}&\mathbb{P}_n\left[\sup_{u\in B^{\bar\pi}_{m,m'}(0,1)}\breve\nu_n^{(1)}(u)>\eta\right]\\ =\;&\mathbb{P}_n\left[\exists(u_k)_{k\in\mathbb{N}}\in\prod_{k\in\mathbb{N}}T_k\;/\;\breve\nu_n^{(1)}(u_0)+\sum_{k=1}^{+\infty}\breve\nu_n^{(1)}(u_k-u_{k-1})>\eta_0+\sum_{k\geq 1}\eta_k\right]\\ \leq\;&\mathbb{P}_1+\mathbb{P}_2\end{aligned}$$

where

$$\mathbb{P}_1=\sum_{u_0\in T_0}\mathbb{P}_n(\breve\nu_n^{(1)}(u_0)>\eta_0),\quad\mathbb{P}_2=\sum_{k=1}^{\infty}\sum_{\substack{u_{k-1}\in T_{k-1}\\ u_k\in T_k}}\mathbb{P}_n(\breve\nu_n^{(1)}(u_k-u_{k-1})>\eta_k).$$

Then using Inequality (38), we straightforwardly infer that $\mathbb{P}_1\leq\exp(H_0-Cnx_0)$ and $\mathbb{P}_2\leq\sum_{k\geq 1}\exp(H_{k-1}+H_k-Cnx_k)$ if we choose

$$\begin{cases}\eta_0=\sigma_1^2(\sqrt{3x_0}+\bar r_{(m,m')}x_0)\\ \eta_k=(\sigma_1^2/\sqrt{2})\delta_{k-1}(\sqrt{15x_k}+3\bar r_{(m,m')}x_k).\end{cases}$$

Fix $\tau > 0$ and choose x_0 such that

$$Cnx_0 = H_0 + L_{m'} D_{m'} + \tau$$

and for $k \geq 1$, x_k such that

$$Cnx_k = H_{k-1} + H_k + kD_{m'} + L_{m'} D_{m'} + \tau.$$

If $D_{m'} \geq 1$, we infer that

$$\mathbb{P}_n \left(\sup_{t \in B^{\bar{\pi}}_{m,m'}(0,1)} \breve{\nu}_n^{(1)}(t) > \eta_0 + \sum_{k \geq 1} \eta_k \right) \leq e^{-L_{m'} D_{m'} - \tau} \left(1 + \sum_{k=1}^{\infty} e^{-kD_{m'}} \right) \leq 1.6 e^{-L_{m'} D_{m'} - \tau}.$$

Now, it remains to compute $\sum_{k \geq 0} \eta_k$. We note that $\sum_{k=0}^{\infty} \delta_k = \sum_{k=0}^{\infty} k\delta_k = 2\delta_0$. This implies

$$\begin{aligned}
& x_0 + \sum_{k=1}^{\infty} \delta_{k-1} x_k \\
\leq & \left[\ln(5/\delta_0) + \delta_0 \sum_{k=1}^{\infty} 2^{-(k-1)} [(2k-1)\ln(2) + 2\ln(5/\delta_0) + k] \right] \frac{D(m,m')}{nC} \\
& + \left(1 + \delta_0 \sum_{k \geq 1} 2^{-(k-1)} \right) \frac{L_{m'} D_{m'}}{nC} + \left(1 + \delta_0 \sum_{k \geq 1} 2^{-(k-1)} \right) \frac{\tau}{nC} \\
\leq & \; a(\delta_0) \frac{D(m,m')}{n} + \frac{1 + 2\delta_0}{C} \frac{L_{m'} D_{m'}}{n} + \frac{1 + 2\delta_0}{C} \frac{\tau}{n}, \qquad (78)
\end{aligned}$$

where $Ca(\delta_0) = \ln(5/\delta_0) + \delta_0(4\ln(5/\delta_0) + 6\ln(2) + 4)$. This leads to

$$\begin{aligned}
\left(\sum_{k=0}^{\infty} \eta_k \right)^2 & \leq \frac{\sigma_1^4}{2} \left[\sqrt{2} \left(\sqrt{3x_0} + + \bar{r}_{m,m'} x_0 \right) + \sum_{k=1}^{\infty} \delta_{k-1} \left(\sqrt{15x_k} + 3\bar{r}_{m,m'} x_k \right) \right]^2 \\
& \leq \frac{\sigma_1^4}{2} \left[\left(\sqrt{6x_0} + \sum_{k=1}^{\infty} \delta_{k-1} \sqrt{15x_k} \right) + \bar{r}_{m,m'} \left(\sqrt{2} x_0 + 3 \sum_{k=1}^{\infty} \delta_{k-1} x_k \right) \right]^2 \\
& \leq 15\sigma_1^4 \left[\left(\sqrt{x_0} + \sum_{k=1}^{\infty} \delta_{k-1} x_k \right)^2 + \bar{r}_{m,m'}^2 \left(x_0 + \sum_{k=0}^{\infty} \delta_{k-1} x_k \right)^2 \right] \\
& \leq 15\sigma_1^4 \left[\left(1 + \sum_{k=1}^{\infty} \delta_{k-1} \right) \left(x_0 + \sum_{k=1}^{\infty} \delta_{k-1} x_k \right) + \bar{r}_{m,m'}^2 \left(x_0 + \sum_{k=1}^{\infty} \delta_{k-1} x_k \right)^2 \right].
\end{aligned}$$

Now, fix $\delta_0 \leq 1/5$ (say, $\delta_0 = 1/10$) and use the bound (78). The bound for $(\sum_{k=0}^{+\infty} \eta_k)^2$ is less than a quantity proportional to

$$\sigma_1^4 \left[\frac{D(m,m')}{n} + \frac{L_{m'} D_{m'}}{n} + \bar{r}_{m,m'}^2 \left(\frac{D(m,m')}{n} + \frac{L_{m'} D_{m'}}{n} \right)^2 + \frac{\tau}{n} + \bar{r}_{m,m'}^2 \frac{\tau^2}{n^2} \right].$$

Now in the case of collection [DP], we have $L_m = 1$, $\bar{r}_{m,m'}$ is bounded uniformly with respect to m and m' and $(D(m,m')/n)^2 \leq (N_n/n)^2 \leq \Delta^2 / \ln^4(n)$ with $N_n \leq n\Delta / \ln^2(n)$.

Thus the bound for $(\sum \eta_k)^2$ reduces to

$$C'\sigma_1^4\left[\frac{D(m,m')}{n}+(1+r)^3\Delta^2/\bar{\pi}_0+\frac{\tau}{n}+\bar{r}^2_{m,m'}\frac{\tau^2}{n^2}\right].$$

Next, for collection [GP], we use the $L_m \le c\ln(n)$, $\bar{r}^2_{m,m'} \le (R_{max}+1)^3 N_n/(\bar{\pi}_0 D(m,m'))$ and $N_n \le n\Delta/\ln^2(n)$ to obtain the bound

$$\begin{aligned}\bar{r}^2_{m,m'}\left(\frac{D(m,m')}{n}+\frac{L_{m'}D_{m'}}{n}\right)^2 &\le (R_{max}+1)^3\frac{N_n}{\bar{\pi}_0 D(m,m')}\frac{D(m,m')^2}{n^2}(1+\ln(n))^2\\ &\le (R_{max}+1)^3\frac{N_n D(m,m')}{\bar{\pi}_0 n^2}(1+\ln(n))^2\\ &\le (R_{max}+1)^3\frac{N_n^2}{n^2}(1+\ln(n))^2 \le 2(R_{max}+1)^3\Delta^2/\bar{\pi}_0.\end{aligned}$$

Thus, the bound for $(\sum \eta_k)^2$ is proportional to

$$\sigma_1^4\left[\frac{D(m,m')}{n}+\frac{L_{m'}D_{m'}}{n}+2(R_{max}+1)^3\Delta^2/\bar{\pi}_0+\frac{\tau}{n}+\bar{r}^2_{m,m'}\frac{\tau^2}{n^2}\right].$$

This term defines $\breve{p}(m,m')$ as given in Proposition 4.6.
The last case corresponds to collection [T]. Here $L_m = 1$, $\bar{r}_{m,m'} \le C\sqrt{D(m,m')}$ and $N_n \le \sqrt{n\Delta}/\ln(n)$. We get

$$\begin{aligned}\bar{r}^2_{m,m'}\left(\frac{D(m,m')}{n}+\frac{L_{m'}D_{m'}}{n}\right)^2 &\le C"\frac{D(m,m')^3}{n^2} \le C"\frac{N_n^3}{n^2} \le C"\frac{\Delta^{3/2}}{\sqrt{n}}\\ &\le C"\Delta^2,\end{aligned}$$

since $1/\sqrt{n} \le \sqrt{\Delta}$. Thus, the bound for $(\sum \eta_k)^2$ is proportional to

$$\sigma_1^4\left[\frac{D(m,m')}{n}+\frac{L_{m'}D_{m'}}{n}+C"\Delta^2+\frac{\tau}{n}+\bar{r}^2_{m,m'}\frac{\tau^2}{n^2}\right].$$

This term defines $\breve{p}(m,m')$ as given in Proposition 4.6.

We obtain, for $K=(R_{max}+1)^3/\bar{\pi}_0$ or $K=C"$,

$$\begin{aligned}&\mathbb{P}_n\left[\sup_{u\in B^{\bar{\pi}}_{m,m'}(0,1)}[\breve{\nu}_n^{(1)}(u)]^2 > \kappa\sigma_1^4\left(\frac{D_m+D_{m'}(1+L_{m'})}{n}+K\Delta^2+2(\frac{\tau}{n}\vee 2\bar{r}^2_{m,m'}\frac{\tau^2}{n^2})\right)\right]\\ &\quad\le \mathbb{P}_n\left[\sup_{u\in B^{\bar{\pi}}_{m,m'}(0,1)}[\breve{\nu}_n^{(1)}(u)]^2 > \eta^2\right] \le 2\mathbb{P}_n\left[\sup_{u\in B^{\bar{\pi}}_{m,m'}(0,1)}\breve{\nu}_n^{(1)}(u) > \eta\right] \le 3.2e^{-L_{m'}D_{m'}-\tau}\end{aligned}$$

so that, reminding that $\breve{G}_m(m')$ is defined by (68),

$$\begin{aligned}
&\mathbb{E}\left[\left(\breve{G}_m^2(m') - \kappa\sigma_1^4\frac{D_m + D_{m'}(1+L_{m'})}{n} + K\Delta^2\right)_+ \mathbb{I}_{\Omega_n}\right] \\
\leq\ & \int_0^\infty \mathbb{P}_n\left(\breve{G}_m^2(m') > \kappa\sigma_1^4\frac{D_m + D_{m'}(1+L_{m'})}{n} + K\Delta^2 + \tau\right)d\tau \\
\leq\ & e^{-L_{m'}D_{m'}}\left(\int_{2\kappa\sigma_1^4/\bar{r}^2_{(m,m')}}^\infty e^{-n\tau/(2\kappa\sigma_1^4)}d\tau + \int_0^{2\kappa\sigma_1^4/\bar{r}^2_{m,m'}} e^{-n\sqrt{\tau}/(2\sqrt{\kappa}\bar{r}_{m,m'}\sigma_1^2)}d\tau\right) \\
\leq\ & e^{-L_{m'}D_{m'}}\frac{2\kappa\sigma_1^4}{n}\left(\int_0^\infty e^{-v}dv + \frac{2\bar{r}^2_{m,m'}}{n}\int_0^\infty e^{-\sqrt{v}}dv\right) \\
\leq\ & e^{-L_{m'}D_{m'}}\frac{2\kappa\sigma_1^4}{n}(1 + \frac{4\bar{r}^2_{m,m'}}{n}) \leq \kappa' e^{-L_{m'}D_{m'}}\frac{\sigma_1^4}{n} \text{ which ends the proof.}\square
\end{aligned}$$

References

[1] Abramowitz, M. and Stegun, I.A. (1972). Handbook of mathematical functions with formulas, graphs, and mathematical tables. Edited by Milton Abramowitz and Irene A. Stegun. John Wiley and Sons, Inc., New York.

[2] Aït-Sahalia, Y. (2006) Estimating Continuous-Time Models Using Discretely Sampled Data, Econometric Society 2005 World Congress Invited Lecture, forthcoming in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress,* edited by Richard Blundell, Persson Torsten and Whitney K. Newey, Econometric Society Monographs, Cambridge University Press.

[3] Aït-Sahalia, Y. and Mykland, P. A. (2004) Estimators of diffusions with randomly spaced discrete observations: a general theory. *Ann. Statist.* **32**, 2186-2222.

[4] Baraud, Y., Comte, F. and Viennet, G. (2001a). Adaptive estimation in an autoregression and a geometrical β-mixing regression framework. *Ann. Statist.* **39**, 839-875.

[5] Baraud, Y., Comte, F. and Viennet, G. (2001b). Model selection for (auto)-regression with dependent data. *ESAIM Probab. Statist.* **5**, 33-49.

[6] Barlow, M.T. and Yor, M. (1982). Semi-martingale inequalities via the Garsia-Rodemich-Rumsey Lemma and applications to local times. *Journal of Functional Analysis* **49**, 198-229.

[7] Barndorff-Nielsen, O.-E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J.R. Stat. Soc. Ser. B. Stat. Methodol.*, **64**, 253-280.

[8] Barron, A.R., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113**, 301–413.

[9] Beskos, A., Papaspiliopoulos, O. and Roberts, G.O. (2006a) Retrospective exact simulation of diffusion sample paths with applications. *To appear in Bernoulli.*

[10] Beskos, A., Papaspiliopoulos, O., Roberts, G.O. and Fearnhead P. (2006b) Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68**, 333-382.

[11] Beskos, A. and Roberts, G.O. (2005). Exact simulation of diffusions. *Ann. Appl. Probab.* **15**, 2422-2444.

[12] Bibby, B.M., Jacobsen, M. and Sørensen, M. (2002). Estimating functions for discretely sampled diffusion-type models. In *Handbook of Financial Econometrics.* Amsterdam: North-Holland.

[13] Birgé, L. and Massart, P. (1998) Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329–375.

[14] Bollerslev, T. and Zhou, H. (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *J. of Econom.* **109**, 33-65.

[15] Chaleyat-Maurel, M. and Genon-Catalot, V. (2006) Computable infinite dimensional filters with applications to discretized diffusion processes. *Stochastic Process. Appl.* **116**, 1447-1467.

[16] Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1**, 54-81.
[17] Comte, F. (2001) Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Bernoulli* **7**, 267-298.
[18] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2005). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. Preprint MAP5 2005-21, to appear in *Bernoulli*.
[19] Comte, F. and Genon-Catalot (2006). Penalized projection estimator for volatility density. *Preprint MAP5* 2005-9,to appear in *Scand. J. Statist.* (online early issue).
[20] Comte, F. and Rozenholc, Y. (2002). Adaptive estimation of mean and volatility functions in (auto-)regressive models. *Stochastic Process. Appl.* **97**, 111-145.
[21] Comte, F. and Rozenholc, Y. (2004) A new algorithm for fixed design regression and denoising. *Ann. Inst. Statist. Math.* **56** 449-473.
[22] 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.
[23] DeVore, R. A. and Lorentz, G.G. (1993) *Constructive approximation.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 303. Springer-Verlag, Berlin.
[24] Ditlevsen, S. and Sørensen, M. (2004). Inference for observations of integrated diffusion processes. *Scand. J. Statist.* **31**, 417-429.
[25] Donoho, D.L. Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996) Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508-539.
[26] Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69**, 959-993.
[27] Ethier, S. N. and Kurtz, T. G. (1986) *Markov processes. Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
[28] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, **6**, 1051-1079.
[29] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **5**, 885-872.
[30] Gloter, A. (2000). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.* **4**, 205-227.
[31] Gloter, A. (2006). Parameter estimation for a discretely observed integrated process. *Scand. J. Statist.* **33**, 83-104.
[32] Gloter, A. and Gobet, E. (2005). LAMN property for integrated diffusions. *Preprint CMAP* R.I. 570, http://www.cmap.polytechnique.fr/preprint/repository/570.pdf.
[33] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79**, 135-163.
[34] Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211-229.
[35] Lamberton, D. and Lapeyre, B. (1996) *Introduction to stochastic calculus applied to finance*, Translated from the 1991 French original by Nicolas Rabeau and Francois Mantion, Chapman & Hall, London.
[36] Lefebvre, M. (1997) On the inverse of the first hitting time problem for bidimensional processes. *J. Appl. Probab.* **34**, 610-622.
[37] Renault, E. and Touzi, N (1996). Option Hedging and Implicit Volatilities. *Mathematical Finance* **6**, 279-302.
[38] D. Revuz, D. and Yor, M. (2005) *Continuous martingales and Brownian motion*, Corrected Third edition (1999), Springer, Berlin.
[39] Rogers, L.C.G. and Williams, D. (2000) *Diffusions, Markov processes, and martingales. Vol. 2*, Reprint of the second (1994) edition, Cambridge Univ. Press, Cambridge.
[40] Sørensen, M. (2000). Prediction-based estimating functions. *Econometrics Journal* **3**, 123-147.
[41] Viennet, G. (1997) Inequalities for absolutely regular sequences: application to density estimation. *Prob. Th. Rel. Fields* **107**, 467-492.
[42] Yoshida, N. (1992) Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.* **41**, 220–242.