ELSEVIER

Statistics & Probability Letters 32 (1997) 11–24

STATISTICS & PROBABILITY LETTERS

Minimax estimation of the diffusion coefficient through irregular samplings

Marc Hoffmann

Mathématiques, URA CNRS 1321, Université Paris VII, Couloir 45-55, 2 place Jussieu, 75251 Paris Cedex 05, France

Received October 1995

Abstract

We study the problem of estimating the time dependent diffusion coefficient of a diffusion process in a nonparametric setting, when the sample path is observed at discrete times. We look at global L^p-error loss over a wide range of function spaces (namely, Besov spaces). We exhibit the minimax rate of convergence over linear estimators and provide estimators based on fast wavelets methods which are optimal. Our method takes into account functional estimation on the interval (with edges effects) and allows to consider irregular sampling schemes.

Keywords: Minimax estimation; Diffusion processes; Irregular sampling schemes; Wavelet orthonormal bases; Wavelets on the interval; Besov spaces

1. Introduction

1.1. Estimating the diffusion coefficient

The statistics of diffusion processes, when only a discrete sampling is available, has been progressively investigated. Pioneering work dealt with the drift estimation (Dorogovcev, 1976; Ibragimov and Has'minski, 1981; and many others). The parametric estimation of the diffusion coefficient was studied by Donahl (1987), Jacod (1993) and Genon-Catalot and Jacod (1993, 1994) in a general framework, including random samplings. A nonparametric approach appeared in Florens-Zmirou (1993) for the homogeneous case.

The evolution of modeling financial data (in option pricing theory or in modeling interest rates) led to the problem of estimating the time dependent diffusion coefficient. In the nonparametric case, Genon-Catalot, Laredo and Picard (1992) proposed an estimator based on a wavelets orthonormal basis and studied its asymptotic properties in L^2-error loss. Later, Soulier (1993) made a refinement of their method, extending it to L^p-error and to kernel estimators.

This paper studies the functional estimation in the minimax framework by use of wavelets of the diffusion coefficient $\sigma(t)$ in the one dimensional model

$$dX_t = b_t(X_.)dt + \sigma(t)dW_t, \quad X_0 = x, \tag{1.1}$$

0167-7152/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved
PII S0167-7152(96)00052-1

where (W_t) is a standard Brownian motion, $x \in \mathbb{R}$ is known and the two functionals $\sigma(t)$ and $b_t(x_.) = b(x_s, 0 \leqslant s \leqslant t)$ (where x is defined on the canonical space of real-valued continuous functions) are assumed to be unknown.

The sample path $(X_t, 0 \leqslant t \leqslant 1)$ is discretely observed on the time interval $[0,1]$ at $m = 2^n$ points, which are not supposed to be equispaced.

1.2. Guidelines

Our approach is twofold:

- from a theoretical viewpoint, we exhibit the asymptotic minimax rate of convergence over linear estimators for an unknown function σ lying in a Besov space. This is motivated by the fact that in the density estimation problem – which is a commonly used test for nonparametric estimation – the case of Besov spaces is optimal (in the sense that the minimax rates for Sobolev balls are obtained because of their inclusion in Besov balls, see Kerkyacharian and Picard, 1993). Another interesting point is that wavelets basis offer unconditional basis for Besov spaces which are therefore well suited for estimating-like problems.
- from a practical point of view, we shall be concerned with methods that can be easily implemented in practice. We provide estimators based on fast wavelets methods which achieve this purpose and which are optimal in the minimax sense. Moreover, we do not restrict ourselves to equispaced data. This is also of practical interest as, for instance, irregular sampling allows us to deal with missing data.

Furthermore, our procedure provides a 'true' estimator on $[0,1]$ thanks to the fast wavelet transform on the interval, as developed by Cohen et al. (1994). This enables us to make an accurate estimation near the edges of the interval, and to avoid penalizing by a test function with compact support strictly included in $[0,1]$ when looking at the integrated error of the estimator, in contrast to previous authors (Genon–Catalot et al., 1992; Soulier, 1993). This is of practical relevance too: the treatment of the edges when analysing a signal on an interval is a well known problem (in numerical analysis and image processing) and we refer to Cohen et al. (1994) for further details.

1.3. Contents

Section 2 describes the model and hypotheses. Section 3 presents the construction of our estimators and the minimax results obtained when working with different kinds of sampling schemes: we will develop, in particular, sampling procedures which achieve optimal results (the classical $m^{-sp/(1+2s)}$ rate of convergence) that we will call *regular schemes*. In short, these are the discretization procedures which are not 'too sparse'. The proofs of these results are given in Section 4.

Finally, we wish to mention that our model can be extended by a classical argument, derived from Ito's formula, to the more general equation

$$\mathrm{d}X_t = b_t(X_.)\mathrm{d}t + h(X_t)\sigma(t)\mathrm{d}W_t, \quad X_0 = x,$$

where h is a known function which is assumed to be smooth (see e.g. Genon–Catalot et al., 1992).

2. Statistical model

2.1. Basic assumptions

We consider the model derived from the stochastic differential equation (1.1) defined in Section 1. $E_{\sigma,b}$ denotes the mathematical expectation w.r.t. the law of the observations. We make the following assumptions:

A0. (1.1) admits a unique solution with almost surely continuous paths.

A1. σ is a nonvanishing positive function defined on $[0,1]$ (nondegeneracy of the model). Thus, we may either work with σ or σ^2. We will see in Section 3.2. the relevance of choosing σ^2 instead of σ.

σ^2 belongs to the Besov ball V of radius $M > 0$, included in the Besov space $B^s_{p,\lambda}([0,1])$, with $1 \leqslant p < +\infty, s > 1, 1 \leqslant \lambda \leqslant +\infty$. We assume s is large enough so that σ^2 has its first derivative bounded.

A2. The drift b is a nonanticipative fonctional and the pair (σ^2, b) belongs to the class $\mathscr{H}$ of functionals for which there exists a constant $v > 0$ such that

$$E_{\sigma,b}\left(\exp\left(v\int_0^1 \frac{b_s^2(X.)}{\sigma^2(s)}\mathrm{d}s\right)\right) \leqslant K < +\infty,$$

for a constant K which does not depend on b or σ. Note that this can be obtained by classical assumptions such as σ bounded away from 0 and b satisfying a uniform linear growth condition.

2.2. Sampling schemes

We do not restrict ourselves to equispaced data. We will call *sampling scheme* a family of increasing points $(t_{i,n}, 1 \leqslant i \leqslant 2^n, n \geqslant 0)$ of the interval $[0,1]$, which represent the observation times of the process X. We will simply write t_i instead of $t_{i,n}$ when no confusion is possible.

- We shall always assume that for fixed n, the $t_{i,n}$'s are all distinct.
- A special case of interest to us will be the so-called *regular sampling schemes* for which we will be able to construct optimal estimators. A sampling will be *regular* if any of the three following conditions is fulfilled:

Regular case B1.

$$\forall p > 1: \quad \sum_i (\Delta t_i)^p = \mathrm{O}(2^{-n(p-1)}),$$

where $\Delta t_i = t_i - t_{i-1}$ denotes the length between two observation times.

Regular case B2. For some $q \in]1, +\infty[$,

$$\left(\sum_i |t_i - i2^{-n}|^q\right)^{1/q} = \mathrm{O}(w_{n,q}),$$

where $w_{n,q} = 2^{ns(1+2s)(2/q-1)}$. In particular, a sampling such that $\sum_i |t_i - i2^{-n}|^2$ is bounded is B2-*regular*.

Regular case B.3. There exists a sub-sampling $(t_i,\ i \in S_n)$, $S_n \subset \{1, \ldots, 2^n\}$ such that:

$$\forall i \in S_n : |t_{\varsigma(i)} - i2^{-n}| \leqslant h_n,$$

where ς is a reordering of S_n and $h_n = 2^{-ns/(1+2s)}$. In addition, we will ask that $\#S_n^c = \mathrm{O}(h_n^{-1})$.

B1 measures the gaps between the observation times. It is less restrictive than the usual conditions like $\sup_{t_i} \Delta t_i = \mathrm{O}(2^{-n})$ and is of course true for equispaced data.

B2 measures the l^q distance between the sampling and the uniform sampling. We will hereafter see that if the $(t_{i,n},\ i \leqslant 2^n)$ are 'close' enough to the $(i2^{-n},\ i \leqslant 2^n)$ a standard procedure of estimation is still optimal.

B3 is a refinement of the approach described in B2. The idea is still to compare the $t_{i,n}$'s to the uniform sampling. One defines a threshold h_n and omits some 'irrelevant' $t_{i,n}$'s in such a way that the admissible $t_{i,n}$'s are close to some corresponding $i2^{-n}$'s, within h_n. The resulting sub-sampling scheme is then closer – in the l^q sense – to the uniform scheme than the previous sampling, up to the omitted points. We will construct an estimating procedure based on such samplings which is optimal, provided h_n and $\#S_n$ are chosen properly.

Now, consider a B1-*regular* sampling scheme and let $\mu_n = 2^{-n}\sum_i \delta_{t_{i,n}}$ be the associated empirical sampling measure. One can look at the weak convergence of μ_n to a measure μ on $[0,1]$. For instance, the equispaced sampling scheme ($t_{i,n} = i2^{-n},\ 1\leqslant i\leqslant 2^n$) converges weakly to the Lebesgue measure on $[0,1]$. Conversely, given a measure μ with no atom, one can construct a sampling scheme converging to μ in an obvious way: let F_μ be the repartition function of μ. Then, setting $t_{i,n} = F_\mu^{-1}(i2^{-n})$ (where F^{-1} denotes as usual the right continuous inverse of F) defines such a procedure. If we focus on such sampling schemes, they will be B1-*regular* as soon as μ has some smoothness. More precisely, if we assume that μ is absolutely continuous, with density f w.r.t. the Lebesgue measure, B1 will hold if f is continuous and nonvanishing (we can even weaken this to the (Riemann) integrability of f^{-p}).

Finally, let us just mention a standard procedure to obtain a sub-sampling as described for B3-*regular schemes*:

$$\text{set } t_1^* = \inf\{t_i\colon |t_i - 2^{-n}|\leqslant h_n\}$$

and for $i\geqslant 2\colon t_i^* = \inf\{t_j \neq t_k^*,\ k\leqslant i-1\colon |t_j - i2^{-n}|\leqslant h_n\}$, with the usual notation $\inf\emptyset = \infty$.

3. Main results

3.1. Lower bounds

Proposition 1. *Let $\mathscr{F}$ denote the set of all estimators. Set $m = 2^n$ and*

$$R_m(\hat\sigma^2, V) = \sup_{(\sigma^2,b)\in\mathscr{H}} E_{\sigma,b}\int_{[0,1]} |\hat\sigma^2 - \sigma^2|^p \mathrm{d}x.$$

Let $t_{i,n}$ be an arbitrary sampling scheme. For every $p\in[1,+\infty[$, there exists a positive constant C_1 depending on p and M such that

$$\inf_{\hat\sigma^2\in\mathscr{F}} R_m(\hat\sigma^2, V)\geqslant C_1 m^{-sp/(1+2s)}.$$

In other words, the classical rate for nonparametric models such as density estimation or nonparametric regression is a lower bound. Moreover, this bound cannot be improved by choosing any particular design of observation times.

3.2. Upper bounds

Construction of wavelet estimators. We propose three linear estimators of σ^2 according to the regular conditions B1–B3 developed in the previous section. They all coincide in the case of equispaced observation times and can be understood as a generalization of the estimator of Genon-Catalot et al. (1992). We consider a multiresolution analysis of $L^2([0,1])$, say $(V_j,\ j\in\mathbb{Z})$, and estimate the projection of σ^2 onto V_j. For this, we can rely on the construction of wavelets on the interval, following Cohen et al. (1994) – the abbreviation used will be CDV – which provide unconditional basis of $B^s_{p,\lambda}([0,1])$ if we calibrate properly their regularity with s. More precisely, let us be given a multiresolution analysis of $L^2(\mathbb{R})$, of regularity $r > s$, derived from a scaling function $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$. We assume that φ is compactly supported, with length support N_0.

1. We first choose an integer J such that $2^J \geqslant 2N_0$. We estimate σ^2 by $P_j^{[0,1]}\sigma^2$ for $j \geqslant J$:

$$P_j^{[0,1]}\sigma^2 = \sum_{k=0}^{N_0-1} \alpha_{j,k}^0 \varphi_{j,k}^0 + \sum_{k \in s_j} \alpha_{j,k}\varphi_{j,k} + \sum_{k=0}^{N_0-1} \alpha_{j,k}^1 \varphi_{j,k}^1,$$

where the $(\varphi_k^0, \varphi_k^1,\ k = 0,\dots,N_0-1)$ are the edges functions associated to φ, and $\varphi_{j,k}^i(x) = \varphi_k^i(2^j x), i = 0,1$. s_j is the set of indexes which define the interior functions of the preceding expansion, according to the CDV terminology, i.e. $k \in s_j$ if and only if supp $\varphi_{j,k} \subset]0,1[$.

2. We then estimate the wavelets coefficients $\alpha_{j,k}$, $\alpha_{j,k}^i, i = 0,1$, for $j \geqslant J$ which are the scalar product of σ^2 w.r.t. $\varphi_{j,k}$, $\varphi_{j,k}^i, i = 0,1$.

The method consists in approximating $\alpha_{j,k}$ by the Riemann sum $\sum_i \sigma^2(x_i)\varphi_{j,k}(x_i)(x_i - x_{i-1})$, where $(x_i, i = 1,\dots,2^n)$ is a suitable subdivision of $[0,1]$.

We have then to find a preliminary estimator of $\sigma^2(x_i)$, say $\hat{\sigma}_0^2(x_i)$, in order to get the corresponding $\hat{\alpha}_{j,k}$

$$\hat{\alpha}_{j,k} = \sum_i \hat{\sigma}_0^2(x_i)\varphi_{j,k}(x_i)(x_i - x_{i-1}). \tag{3.1}$$

The choice of $\hat{\sigma}_0^2(x_i)$ and x_i will depend on the regularity of the sampling scheme $t_{i,n}$'s. In other words, we first construct a rough preliminary estimator by means of the $\hat{\sigma}_0^2(x_i)$'s and then operate a smoothing procedure thanks to the projection onto $V_j([0,1])$. We shall lay the emphasis on the fact that the ideas we develop here are strongly related to classical nonparametric regression techniques.

Regular case B1: For the sake of simplicity, we put $b = 0$, as we will see in Section 4.2. that the addition of a drift does not interfer in the estimation problem. A natural choice is to set $x_i = t_i$. Then, $\sigma^2(t_i)$ can be estimated by

$$\hat{\sigma}_0^2(t_i) = \frac{1}{\Delta t_i}(\Delta X_{t_i})^2 = \frac{1}{\Delta t_i}\int_{t_{i-1}}^{t_i} \sigma^2(s)\mathrm{d}s + \eta_{t_i}. \tag{3.2}$$

with: $\Delta X_{t_i} = X_{t_i} - X_{t_{i-1}}$, $\Delta t_i = t_i - t_{i-1}$ and $\eta_{t_i} = \frac{1}{\Delta t_i}(\int_{t_{i-1}}^{t_i}\sigma(s)\mathrm{d}W_s)^2 - \frac{1}{\Delta t_i}\int_{t_{i-1}}^{t_i}\sigma^2(s)\mathrm{d}s$.

The η_{t_i}'s are independent and centered random variables. $\sigma^2(t_i)$ can be recovered from $\frac{1}{\Delta t_i}\int_{t_{i-1}}^{t_i}\sigma^2(s)\mathrm{d}s$ up to an (optimal) error of order Δt_i. Here appears the relevance of choosing σ^2 instead of σ in our estimating problem.

Let $S_{j,k}$ denote the support of $\varphi_{j,k}$. Following (3.2) we set

$$\hat{\alpha}_{j,k} = \sum_{t_i \in S_{j,k}} \varphi_{j,k}(t_i)(\Delta X_{t_i})^2,$$

putting $\hat{\alpha}_{j,k} = 0$ whenever $\{t_i : t_i \in S_{j,k}\}$ is empty. This also could have been obtained directly from the definition of the quadratic variation of X if one recalls that for a continuous semimartingale X and for any test function f

$$\sum_{t_{i,n}} f(t_i)(\Delta X_{t_i})^2 \to \int_0^1 f(s)\mathrm{d}\langle X\rangle_s = \int_0^1 f(s)\sigma^2(s)\mathrm{d}s$$

in probability, as n tends to infinity, for any subdivision $t_{i,n}$ of $[0,1]$ with step tending to 0.

Regular cases B2 *and* B3: Another choice consists in setting $x_i = i2^{-n}$, trying then to find a preliminary estimator of $\sigma^2(i2^{-n})$. Although surprising, this approach is motivated by the fact that, following (3.1), the estimator

$$\hat{\alpha}_{j,k} = \sum_i \hat{\sigma}_0^2(t_i)\varphi_{j,k}(i2^{-n})2^{-n} \tag{3.3}$$

would attain the minimax rate of convergence $m^{-sp/(1+2s)}$ if the t_i's were equispaced on $[0,1]$ (q.v. Genon-Catalot et al., 1992). We shall therefore be concerned with finding an estimator of $\sigma^2(i2^{-n})$, for instance by means of the (rescaled) observations $\hat{\sigma}_0^2(t_i) = 1/\Delta t_i(\Delta X_{t_i})^2$. An underlying condition on the accuracy of this procedure is the distance between the regular grid design and the t_i's.

- A first naive estimator is obtained by taking $\hat{\sigma}_0^2(t_i)$ for the estimation of $\sigma^2(i2^{-n})$. The wavelet coefficient estimate is then

$$\hat{\alpha}_{j,k} = \sum_i (\Delta X_{t_i})^2 \, (2^{-n}/\Delta t_i) \, \varphi_{j,k}(i2^{-n}). \tag{3.4}$$

The regularity condition B2 on the size control of $\sum_i |t_i - i2^{-n}|$ will ensure that the estimator attains the $m^{-sp/(1+2s)}$ rate of convergence.

- A second and more sophisticated approach consists in estimating $\sigma^2(i2^{-n})$ by $\hat{\sigma}_0^2(t_i^*)$, where t_i^* is defined following sequel (2.2), whenever it exists. The wavelet coefficient estimate is then

$$\hat{\alpha}_{j,k} = \sum_i (\Delta X_{t_i^*})^2 \, (2^{-n}/\Delta t_i^*)\varphi_{j,k}(i2^{-n}) \, 1_{t_i^* < \infty}. \tag{3.5}$$

The reliability of this estimate will be linked to the threshold h_n and the number $\sum_i 1_{t_i^* = \infty}$ of rejected observations.

We have exhibited three different procedures to estimate $\alpha_{j,k}$, depending on the regularity of the sampling scheme. We define $\hat{\alpha}_{j,k}{}^i$, $i = 0, 1$ in an analogous way. We thus get our estimators $\hat{\sigma}^2$ of σ^2:

$$\hat{\sigma}^2 = \sum_{k=0}^{N_0-1} \hat{\alpha}_{j,k}{}^0 \varphi_{j,k}^0 + \sum_{k \in s_j} \hat{\alpha}_{j,k}\varphi_{j,k} + \sum_{k=0}^{N_0-1} \hat{\alpha}_{j,k}{}^1 \varphi_{j,k}^1.$$

The order of magnitude of j is to be given by the behaviour of $\hat{\sigma}^2$ when n tends to infinity. The regularity of the sampling scheme will ensure that the rate of the approximation is optimal.

For the sake of simplicity, we will hereafter denote by $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ the estimators associated to B1-, B2- and B3-*regular schemes*, respectively, following the foregoing construction, and will simply write $\hat{\sigma}^2$ when no confusion is possible.

Remark. We have chosen $2^J \geqslant 2N_0$ so that the left and right edge functions $\varphi_{j,k}^0$ and $\varphi_{j,k}^1$ do not interact; for $j \geqslant J$ the CDV algorithm requires exactly 2^j coefficients for the analysis of σ^2 at level 2^{-j}.

We may now state our result on upper bounds:

Proposition 2. *Let $p \in [1, +\infty[$. Assume the sampling scheme is regular, and let $\hat{\sigma}^2$ be the corresponding estimator as defined above. If $j - n/2$ tends to $-\infty$ as n tends to $+\infty$, then*

$$R_m(\hat{\sigma}^2, V) \leqslant C_2 m^{-sp/(1+2s)},$$

where C_2 is a positive constant, depending only on φ, p, M and v.

As a consequence, according to Proposition 1, the minimax rate of convergence for regular samplings is the classical $m^{-sp/(1+2s)}$ and is attained by our estimators.

4. Proofs

4.1. Proof of Proposition 1

We give a sketch of the proof, following a classical method (see Kerkyacharian and Picard, 1992). We first remark that

$$R_m(\hat{\sigma}^2, V) \geqslant \sup_{\sigma^2 \in C_j} E_{\sigma,0} \|\hat{\sigma}^2 - \sigma^2\|_p^p,$$

where C_j is some parametric set included in V. We will henceforth consider the model without drift.

• *Constructing a hypercube of V.*

We start with a function g of the ball of radius $M/2$ of the space $B^s_{p,\lambda}(\mathbb{R})$, satisfying $g(x) = c \geqslant 1$, for $x \in [0,1]$. Let ψ be a wavelet of regularity $r > s$, with compact support included in $[-A, A]$, where A is a fixed integer. We set

$$C_j(\gamma) = \left\{ \sigma_\varepsilon^2 = g + \sum_{k \in K_j} \gamma_k \varepsilon_k \psi_{j,k}, \varepsilon = (\varepsilon_1, \ldots, \varepsilon_{2^j}), \varepsilon = \pm 1 \right\},$$

where $K_j = \{A + 2kA, k = 0, \ldots, 2^j - 1\}$, and $\psi_{j,k} = 2^{j/2}\psi(2^j x - k)$ so that $\psi_{j,k}$ and $\psi_{j,k'}$ have disjoint supports for $k \neq k'$.

We look for conditions in order to have $C_j(\gamma) \subset V$. This is released if

$$\left(2^{-j} \sum_k \gamma_k^p \right)^{1/p} \leqslant C 2^{-j(s+1/2)} \quad \text{and} \quad \|\gamma\|_\infty \leqslant 2^{-j/2}. \tag{4.1}$$

• *Bounds on the Hellinger distance.* Let P_+(resp. P_-) denote the law of a sample of observation, derived from a model with a diffusion coefficient $\sigma^2_{\varepsilon,+} = g + \sum_{k' \neq k} \gamma_{k'} \varepsilon_{k'} \varphi_{j,k'} + \gamma_k \psi_{j,k}$, (resp. $\sigma^2_{\varepsilon,-} = g + \sum_{k' \neq k} \gamma_{k'} \varepsilon_{k'} \varphi_{j,k'} - \gamma_k \psi_{j,k}$), for some fixed $k \in K_j$.

Let $h^2(P_+, P_-)$ denote the Hellinger distance between P_+ and P_-. We look for conditions in order to have $h^2(P_+, P_-) \leqslant \frac{1}{2}$, and hence apply Assouad's lemma. Our method relies on the following lemma:

Lemma 1.

$$h^2(P_+, P_-) = 1 - \prod_{t_i \in S_{j,k}} \left(1 - \left(\frac{\gamma_k}{c \Delta t_i} \int_{t_{i-1}}^{t_i} \psi_{j,k} \right)^2 \right)^{1/4}.$$

The proof, given in the appendix, is based upon the fact that the solution of (1.1) with $b = 0$ is a Brownian motion, up to a (deterministic) time change, so that computation of the transition density is easily performed.

Assuming $\{t_i : t_i \in S_{j,k}\}$ is nonempty, we may define

$$i_0(\gamma) = \arg\inf_{t_i \in S_{j,k}} \left| 1 - \left(\frac{\gamma_k}{c \Delta t_i} \int_{t_{i-1}}^{t_i} \psi_{j,k} \right)^2 \right| \tag{4.2}$$

Remark. If $\{t_i : t_i \in S_{j,k}\}$ is empty, we obviously have $h^2(P_+, P_-) = 0$.

Applying Lemma 1 entails

$$h^2(P_+,P_-) \leqslant 1 - \left|1 - \left(\frac{\gamma_k}{c\Delta t_{i_0}} \int_{t_{i_0-1}}^{t_{i_0}} \psi_{j,k}\right)^2\right|^{2^n \mu_n(S_{j,k})/4},$$

where μ_n is the sampling measure associated to the $t_{i,n}$'s, so that $2^n \mu_n(S_{j,k}) = \#\{i : t_i \in S_{j,k}\}$. Condition (4.1) on γ_k implies

$$\left(\frac{\gamma_k}{c\Delta t_{i_0}} \int_{t_{i_0-1}}^{t_{i_0}} \psi_{j,k}\right)^2 \leqslant 2^j \|\psi\|_\infty^2 \frac{\gamma_k^2}{c^2} \leqslant \left(1 - \frac{1}{c}\right)^2 \leqslant 1.$$

Now, using $1 - u^n \leqslant n(1-u)$ for $0 \leqslant u \leqslant 1$, we get

$$h^2(P_+,P_-) \leqslant \frac{1}{4} 2^n \mu_n(S_{j,k}) \frac{\gamma_k^2}{c^2} \left(\frac{1}{\Delta t_{i_0}} \int_{t_{i_0-1}}^{t_{i_0}} \psi_{j,k}\right)^2.$$

As a consequence, if we assume that

$$\gamma_k \leqslant \frac{\sqrt{2}}{c} 2^{-j/2} (2^n \mu_n(S_{j,k}))^{-1/2} \tag{4.3}$$

we will have $h^2(P_+,P_-) \leqslant \frac{1}{2}$. Combining (4.1) and (4.3), we get optimality for

$$2^{-n(p/2)} \left(2^{-j} \sum_k \mu_n(S_{j,k})^{-p/2}\right) \simeq 2^{-jsp}.$$

Using then Assouad's lemma provides a lower bound. We shall prove that the classical bound $m^{-sp/(1+2s)}$ cannot be improved. This happens if

$$2^{-n(p/2)} \left(2^{-j} \sum_k \mu_n(S_{j,k})^{-p/2}\right) = o(2^{(j-n)(p/2)}). \tag{4.4}$$

Indeed, $2^{(j-n)p/2}$ is the quantity which leads to the classical bound when balanced with 2^{-jsp}. On the other side, we always have

$$\sum_k \mu_n(S_{j,k}) = 1. \tag{4.5}$$

Our task is then to find a sequence μ_n which minimizes $\sum_k \mu_n(S_{j,k})^{-p/2}$. By an easy argument of convexity, under the constraint (4.5)

$$\inf_{\mu_n} \sum_k \mu_n(S_{j,k})^{-p/2} \geqslant 2^{j[(p/2)-1]} \tag{4.6}$$

so that (4.4) can never be obtained, whatever the sampling μ_n we choose. The proof of Proposition 1 is complete.

4.2. Proof of Proposition 2

We adopt the following convention: C will stand for an absolute constant (depending only on φ, p, M and ν) which may differ at any occurrence. Let us mention that the following bound on constants will be implicitly used throughout the proof:

$$\|\sigma^2\|_\infty \leqslant M(1-2^{-(s-1/p)\lambda'})^{1/\lambda'}$$

with $1/\lambda' + 1/\lambda = 1$. This inequality is a consequence of the characterization of Besov spaces in terms of wavelets sequences and may be found in Donoho et al. (1995).

• *Eliminating the drift.* We first show that the drift b can be regarded as a nuisance term which does not interfers in the estimation poblem. We will use a change of probability method. Let $P_{\sigma,b}$ (resp. $P_{\sigma,0}$) denote the law of the canonical process $(X_t, 0 \leqslant t \leqslant 1)$ on the space of continuous functions, which is a solution of (1.1) driven by σ and b (resp. σ and $b=0$). Following Girsanov's theorem (Revuz–Yor, 1991, p. 344), one has $P_{\sigma,b} = D.P_{\sigma,0}$, the density D being given by

$$D = \exp\left(\int_0^1 \frac{b_s(X_.)}{\sigma^2(s)}\mathrm{d}X_s - \frac{1}{2}\int_0^1 \frac{b_s^2(X_.)}{\sigma^2(s)}\mathrm{d}s\right).$$

Using Hölder's inequality, we may write

$$E_{\sigma,b}\|\sigma^2 - \hat{\sigma}^2\|_p^p \leqslant C_{q'}(E_{\sigma,0}\|\sigma^2 - \hat{\sigma}^2\|_p^{qp})^{1/q}$$

with $1/q + 1/q' = 1$ and $C_q' = (E_{\sigma,0}D^{q'})^{1/q'}$. By the increasing property of the L^p norm in $[0,1]$

$$E_{\sigma,0}\|\hat{\sigma}^2 - \sigma^2\|_p^{qp} \leqslant E_{\sigma,0}\|\hat{\sigma}^2 - \sigma^2\|_{qp}^{qp}.$$

As a consequence, the $E_{\sigma,b}$-risk is controlled by the $E_{\sigma,0}$-risk, provided $C_{q'} < \infty$, uniformly in σ^2 and b. For this, we rely on Lemma 2, proved in the appendix:

Lemma 2. *There exists a* $q' \in]1, +\infty[$ *such that*

$$\sup_{(\sigma^2,b)\in\mathscr{H}} E_{\sigma,0}(D^{q'}) < +\infty.$$

We will henceforth consider the model derived from (1.1) with $b=0$.

• *Bounds on the minimax risk.* For the sake of simplicity, we will hereafter write E instead of $E_{\sigma,0}$. Following the classical method, we write $E\|\hat{\sigma}^2 - \sigma^2\|_p^p$ as a sum of a stochastic term and an approximation term (linked to the wavelets method of projection). More precisely,

$$E\|\hat{\sigma}^2 - \sigma^2\|_p^p \leqslant 2^{p-1}(S_n + A_n)$$

with

$$S_n = E\|\hat{\sigma}^2 - P_j^{[0,1]}\sigma^2\|_p^p \quad \text{and} \quad A_n = \|\sigma^2 - P_j^{[0,1]}\sigma^2\|_p^p,$$

where $P_j^{[0,1]}$ denotes the projection operator onto $V_j([0,1])$, as defined in Section 3.2.

Let us first study the approximation term A_n. Resulting from the approximation of Besov spaces by wavelets sequences, for any $f \in B_{p,\lambda}^s([0,1])$, the following inequality holds

$$\|f - P_j^{[0,1]}f\|_{L^p} \leqslant 2^{-j(s\wedge r)}\varepsilon_j, \tag{4.7}$$

where $\varepsilon_j \to 0$ as j tends to infinity, r being the regularity of the scaling function φ. Assuming $r > s$ and $\sigma^2 \in V$ we get

$$A_n \leqslant C2^{-jsp}.$$

We now consider the stochastic term. Using the localization property of φ leads to

$$S_n \leqslant C2^{j[(p/2)-1]}\left(\sum_0^{N_0-1} E|\hat{\alpha}_{j,k}{}^0 - \alpha^0_{j,k}|^p + \sum_{k\in s_j} E|\hat{\alpha}_{j,k} - \alpha_{j,k}|^p + \sum_0^{N_0-1} E|\hat{\alpha}_{j,k}{}^1 - \alpha^1_{j,k}|^p\right) \tag{4.8}$$

this property being also known as Meyer's lemma (Meyer, 1990, p. 30). From now on, we will no more distinguish in our notations the edge and interior components involved in expansion (4.8). Formally, they are handled the same way. The only thing to keep in mind is that the number of coefficients involved is exactly 2^j.

Convergence of $\hat{\sigma}^2_1$. We first write

$$\int_{S_{j,k}} \varphi_{j,k}(s)\sigma^2(s)\mathrm{d}s = \sum_{t_i\in S_{j,k}} \int_{t_{i-1}}^{t_i} \varphi_{j,k}(s)\sigma^2(s)\mathrm{d}s + R_n^{(1)} \tag{4.9}$$

the remainder term $R_n^{(1)}$ coming from the overlap between $S_{j,k}$ and the right and left edges of $\{t_i : t_i \in S_{j,k}\}$. Nevertheless, using the regularity of the sampling and the assumption $j - n/2 \to -\infty$, one can easily check that the error of such an approximation is less than $2^{(j-n)p/2}$ for $p \geqslant 2$. For $1 \leqslant p < 2$, by Hölder's inequality, one is able to control the L^p-risk by the $L^{p'}$-risk, for some $p' \in [2, +\infty[$, which leads to the same rate of convergence.

Setting $\varepsilon_{t_i} = (\int_{t_{i-1}}^{t_i} \sigma(s)\mathrm{d}W_s)^2 - \int_{t_{i-1}}^{t_i} \sigma^2(s)\mathrm{d}s$, we have

$$\alpha_{j,k} - \hat{\alpha}_{j,k} = \sum_{t_i\in S_{j,k}} \varphi_{j,k}(t_i)\varepsilon_{t_i} + R_n^{(2)} \tag{4.10}$$

with

$$R_n^{(2)} = \sum_{t_i\in S_{j,k}} \int_{t_{i-1}}^{t_i} \sigma^2(s)\left(\varphi_{j,k}(t_i) - \varphi_{j,k}(s)\right)\mathrm{d}s + R_n^{(1)}. \tag{4.11}$$

Once again, one can easily see that $2^{j[(p/2)-1]}\sum_k (R_n^{(2)})^p = o(2^{(j-n)p/2})$, using a Taylor's expansion on φ. The subsequent arguments are then straightforward. We turn now to the first term involved in the right-hand side of (4.10). We first recall Rosenthal's inequality, which may be found in Petrov (1992).

Lemma 3. (Rosenthal's inequality). *Let $\xi_0, \xi_1, \ldots, \xi_m$ be independent and centered variables. Then, for any $p \in [1, +\infty[$, there exists a constant C_p such that*

$$E\left|\sum_{i=0}^m \xi_i\right|^p \leqslant C_p\left(\sum_{i=0}^m E(|\xi_i|^p) + \left(\sum_{i=0}^m E(\xi_i^2)\right)^{p/2}\right).$$

Going back to our study, we first note that the random variables ε_{t_i} are independent and centered. Moreover, using bounds on Gaussian variables, we get

$$E|\varepsilon_{t_i}|^p \leqslant C_p\|\sigma^2\|_\infty^p(\Delta t_i)^p.$$

Then, using Rosenthal's inequality, we obtain:

$$E\left|\sum_{t_i\in S_{j,k}} \varphi_{j,k}(t_i)\varepsilon_{t_i}\right|^p \leqslant C2^{j(p/2)}\left(\sum_{t_i\in S_{j,k}}(\Delta t_i)^2\right)^{p/2}.$$

Hence, using the regularity of the design B1, we finally get $S_n \leqslant C2^{(j-n)p/2}$.The optimal rate is obtained when A_n and S_n are of the same order of magnitude, which leads to $2^j = 2^{n/(1+2s)}$. The proof of case B1 is complete.

Convergence of $\hat{\sigma}_2^2$. We write

$$(\Delta t_i)^{-1}\left(\int_{t_{i-1}}^{t_i}\sigma(s)\mathrm{d}W_s\right)^2 = \sigma^2(t_i) + R_n^{(3)} + \varepsilon_{t_i}, \tag{4.12}$$

and

$$\int \sigma^2(s)\varphi_{j,k}(s)\mathrm{d}s = \sum_i \sigma^2(i2^{-n})\varphi_{j,k}(i2^{-n})2^{-n} + R_n^{(4)} \tag{4.13}$$

The two remainder terms in (4.12) and (4.13) check the following inequalities: $R_n^{(3)} \leqslant C\Delta t_i$, and $R_n^{(4)} \leqslant C2^{-n+3/2j}$. This can be simply obtained by using a Taylor's expansion.

This leads to the following expansion

$$\alpha_{j,k} - \hat{\alpha}_{j,k} = A_1 + A_2 + A_3,$$

with

$$A_1 = \sum_i (\sigma^2(t_i) - \sigma^2(i2^{-n}))\varphi_{j,k}(i2^{-n})2^{-n},$$

$$A_2 = \sum_i \varepsilon_{t_i}\varphi_{j,k}(i2^{-n})2^{-n},$$

$$A_3 = \sum_i R_n^{(3)}\varphi_{j,k}(i2^{-n})2^{-n} + R_n^{(4)}.$$

One can check that the remainder term A_3 is of optimal order, provided $j - n/2$ tends to $-\infty$ when n tends to $+\infty$. A_2 is treated in the same way as in the previous sequel by means of Rosenthal's inequality.

Let us now concentrate on A_1. Note that $|\sigma^2(t_i) - \sigma^2(i2^{-n})| \leqslant C|t_i - i2^{-n}|$, which leads to

$$|A_1|^p \leqslant C2^{-np/2}\left(2^{j/2-n/2}\sum_i |t_i - i2^{-n}|\right)^p.$$

Noticing that the number of terms involved in the sum is of order 2^{n-j} and applying Hölder's inequality entails

$$\left(2^{j/2-n}\sum_i |t_i - i2^{-n}|\right)^p \leqslant 2^{(n-j)(1/2-1/p_1)p}\left(\sum_i |t_i - i2^{-n}|^{p_1}\right)^{p/p_1}, \tag{4.14}$$

for any $p_1 \in]1,+\infty[$. Assuming B2 is fullfilled and taking $p_1 = q$ and $2^j = 2^{n/(1+2s)}$ the following inequality follows from (4.14)

$$|A_1|^p \leqslant C2^{-np/2},$$

which ends the proof for *regular case* B2.

Convergence of $\hat{\sigma}_3^2$. We again use (4.12) and (4.13) as in the previous sequel to get

$$\alpha_{j,k} - \hat{\alpha}_{j,k} = A_1 + A_2 + A_3 + A_4,$$

with

$$A_1 = \sum_i (\sigma^2(t_i^*) - \sigma^2(i2^{-n}))\varphi_{j,k}(i2^{-n})2^{-n}1_{t_i^* < \infty},$$

$$A_2 = \sum_i \varepsilon_{t_i^*}\varphi_{j,k}(i2^{-n})2^{-n}1_{t_i^* < \infty},$$

$$A_3 = \sum_i \sigma^2(i2^{-n})\varphi_{j,k}(i2^{-n})2^{-n}1_{t_i^* = \infty},$$

$$A_4 = \sum_i R_n^{(3)}\varphi_{j,k}(i2^{-n})2^{-n} + R_n^{(4)}.$$

Using the definition of the t_i^*'s as given in sequel 2.2 and a first-order Taylor's expansion, we get

$$|A_1|^p \leqslant h_n^p 2^{-jp/2}. \tag{4.16}$$

The terms A_2 and A_4 are treated in the same way as in *regular case* B2.

We now turn to A_3 to complete our study. One has

$$|A_3|^p \leqslant C2^{-np}2^{jp/2}\left|\sum_i 1_{t_i^* = \infty}\right|^p \leqslant C2^{-np}2^{jp/2}d_n^p. \tag{4.17}$$

The assumptions on d_n and the threshold h_n ensure $|A_3|^p \leqslant C2^{(j-n)p/2}$. The conclusion for $\hat{\sigma}_3^2$ follows easily. This ends the proof of Proposition 3.

Appendix A

A.1. Proof of Lemma 1

• *The case of Brownian motion.* Let $(W_t, t \geqslant 0)$ be a standard Brownian motion and $(s_1 < s_2 < \cdots < s_m)$ and $(t_1 < t_2 < \cdots < t_m)$ two increasing sequences of real numbers in $[0,1]$. We write Q_t and Q_s for the law of $(W_{t_1}, \ldots, W_{t_m})$ and $(W_{s_1}, \ldots, W_{s_m})$, respectively. Then, using the definition of the transition density of the Brownian motion:

$$\begin{aligned} h^2(Q_t, Q_s) &= 1 - (2\pi)^{-m/2}\prod_{i=1}^m (\Delta t_i \Delta s_i)^{-1/4}\int_{\mathbb{R}^m} \exp\left(-\frac{1}{2}\sum_{i=1}^m \frac{1}{2}\left(\frac{1}{\Delta t_i} + \frac{1}{\Delta s_i}\right)(x_i - x_{i-1})^2\right)\mathrm{d}x \\ &= 1 - 2^{m/2}\prod_{i=1}^m (\Delta t_i \Delta s_i)^{1/4}/(\Delta t_i + \Delta s_i)^{1/2}. \end{aligned}$$

• Under $P_{\sigma,0}$, the process X is a time-changed Brownian motion. More precisely, if $v(t) = \int_0^t \sigma^2(u)\mathrm{d}u$, we can write: $X_t = x + \beta_{v(t)}$. β is the Dubins–Schwarz Brownian motion of X (Revuz-Yor, 1991, p. 170).

Set, for σ in V: $F_\sigma(t_i) = \int_{t_{i-1}}^{t_i} \sigma^2(u)\mathrm{d}u$. The foregoing arguments lead to

$$h^2(P_{\sigma,0}, P_{\sigma',0}) = 1 - 2^{m/2}\prod_{i=1}^m \frac{(F_\sigma(t_i)F_{\sigma'}(t_i))^{1/4}}{(F_\sigma(t_i) + F_{\sigma'}(t_i))^{1/2}}.$$

• We may now procede to the computation of $h^2(P_+,P_-)$. We write Δt_i^+ for $F_{\sigma^2_{\varepsilon,+}}(t_i)$ and define Δt_i^- analogously.

For t_i lying in $S_{j,k}$ we have

$$\Delta t_i^+ + \Delta t_i^- = 2c\Delta t_i \quad \text{and} \quad \Delta t_i^+ \Delta t_i^- = (\Delta t_i)^2 c^2 - \gamma_k^2 \left(\int_{t_{i-1}}^{t_i} \psi_{j,k}\right)^2$$

Hence, setting $a_0 = \#\{i : t_i \in S_{j,k}\}$ and $b_0 = \#\{i : t_i \in S_{j,k}^c\}$,

$$h^2(P_+,P_-) = 1 - 2^m 2^{-b_0/2} 2^{-a_0/2} \prod_{t_i \in S_{j,k}} \left(1 - \left(\frac{\gamma_k}{c\Delta t_i} \int_{t_{i-1}}^{t_i} \psi_{j,k}\right)^2\right)^{1/4}.$$

Using the fact that $a_0 + b_0 = m$ completes the proof.

A.2. Proof of Lemma 2

Following the definition of D given in Section 4.2., there exists a $P_{\sigma,b}$-Brownian motion, say B, such that

$$\begin{aligned} E_{\sigma,0}(D^{q'}) &= E_{\sigma,b}\left[\exp\left((q'-1)\left(-\int_0^1 \frac{b_s(X_\cdot)}{\sigma(s)}\mathrm{d}B_s + \frac{1}{2}\int_0^1 \frac{b_s^2(X_\cdot)}{\sigma^2(s)}\mathrm{d}s\right)\right)\right] \\ &= E_{\sigma,b}\left[\exp\left((q'-1)\left(Y_1 + \frac{1}{2}\langle Y\rangle_1\right)\right)\right] \end{aligned}$$

with $Y_t = \int_0^t (b_s(X_\cdot)/\sigma(s))\mathrm{d}B_s$ and $\langle Y\rangle_t = \int_0^t (b_s^2(X_\cdot)/\sigma^2(s))\mathrm{d}s$. The process $(Y_t, 0 \leqslant t \leqslant 1)$ is a $P_{\sigma,b}$-local martingale. Using Cauchy–Schwarz inequality

$$E_{\sigma,0}(D^{q'}) \leqslant \left(E_{\sigma,b}\left[\exp\left(2(q'-1)Y_1 - \frac{1}{2}\langle 2(q'-1)Y\rangle_1\right)\right]\right)^{1/2} \left(E_{\sigma,b}\left[\exp\left(2(q'-1)(q'-\frac{1}{2})\langle Y\rangle_1\right)\right]\right)^{1/2}. \tag{5.1}$$

The first term in the right-hand side of (5.1) is equal to 1, since we take the expectation of an exponential martingale. The only thing we have to check is

$$\sup_{(\sigma^2,b)\in\mathscr{H}} E_{\sigma,b}\left[\exp\left(2(q'-1)(q'-\frac{1}{2})\langle Y\rangle_1\right)\right] < +\infty.$$

As q' is an arbitrary number in $]1,+\infty[$, we pick a q' such that $2(q'-1)(q'-\frac{1}{2}) \leqslant \nu$. Using assumption A2 completes the proof.

References

Cohen, A., I. Daubechies and P. Vial (1994), Wavelets on the interval and fast wavelet transform, *Appl. Comput. Harm. Anal.* **1**, 54–81.

Donhal, G. (1987), On estimating the diffusion coefficient, *J. Appl. Probab.* **24**, 105–114.

Donoho, D.L., I.M. Johnstone, G. Kerkyacharian and D. Picard (1995), Density estimation by wavelet thresholding, *Ann. Statist.*, to appear.

Dorogovcev, A.Ja. (1976), The consistency of an estimate of a parameter of a stochastic differential equation, *Theoret. Probab. Math. Statist.* **10**, 73–82.

Florens-Zmirou, D. (1993), On estimating the diffusion coefficient from discrete observations, *J. Appl. Probab.* **30**, 790–804.

Genon-Catalot, V. and J. Jacod (1993), On the estimation of the diffusion coefficient for multidimensional diffusion processes, *Ann. Inst. H. Poincaré* **29**, 119–151.

Genon-Catalot, V. and J. Jacod (1994), On the estimation of the diffusion coefficient for diffusion processes: random samplings, *Scand. J. Statist.* **21**, 193–221.

Genon-Catalot, V., C. Laredo and D. Picard (1992), Nonparametric estimation of the diffusion coefficient by wavelets methods, *Scand. J. Statist.* **19**, 319–335.

Ibragimov, I.A. and R.Z. Has'minski (1981), *Statistical Estimation, Asymptotic Theory* (Springer, Berlin).

Jacod, J. (1993), Random sampling in estimation problems for continuous Gaussian processes with independent increments, *Stochastic Process. Appl.* **44**, 181–204.

Kerkyacharian, G. and D. Picard (1992), Density estimation in Besov spaces, *Statist. Probab. Lett.* **13**, 15–24.

Kerkyacharian, G. and D. Picard (1993), Density estimation by kernel and wavelets methods: Optimality of Besov Spaces, *Statist. Probab. Lett.* **18**, 327–336.

Meyer, Y. (1990), *Ondelettes et Opérateurs I* (Hermann, Paris).

Revuz, D. and M. Yor (1991), *Continuous martingales and Brownian motion* (Springer, Berlin).

Petrov, V.V. (1992), Generalizations of Rosenthal's inequalities, *Ann. Acad. Sci. Fennicae* (A) **17**, 117–121.

Soulier, P. (1993), Thèse, Université Paris 11, Orsay.