Bernoulli **5**(3), 1999, 447–481

L_p estimation of the diffusion coefficient

MARC HOFFMANN

Unité de Recherche associée au CNRS 1321, Université Paris VII, Unité de Formation et de Recherche de Mathématiques, couloir 45–55, 5e étage, 2 place Jussieu, 75251 Paris Cédex 05, France. e-mail: hoffmann@mathp7.jussieu.fr

We study the functional estimation of the space-dependent diffusion coefficient in a one-dimensional framework. The sample path is observed at discrete times. We study global L_p-loss errors ($1 \leqslant p < +\infty$) over Besov spaces $B_{sp\infty}$. We show that, under suitable conditions, the minimax rate of convergence is the usual $n^{-s/(1+2s)}$. Linking our model to nonparametric regression, we provide an estimating procedure based on a linear wavelet method which is optimal in the minimax sense.

Keywords: Besov spaces; diffusion processes; local time; minimax estimation; nonparametric regression; wavelets on the interval; wavelet orthonormal bases

1. Introduction

1.1. Motivation

In recent years, much effort has been devoted to statistical inference in diffusion processes when only a discrete sampling of the trajectory is available. In particular, a growing interest in the diffusion coefficient has come from mathematical finance, where the diffusion coefficient represents volatility. Whereas parametric inference is quite well known when the discretization step is small (Donhal 1987; Genon-Catalot and Jacod 1993; 1994; Jacod 1993) only emerging results have been proposed for the functional estimation of the diffusion coefficient, when the parameter of interest is globally unknown, and subject only to a functional constraint (usually a smoothness property).

A relatively simple situation consists in studying the time-dependent diffusion coefficient $\sigma^2(\cdot)$ in a model governed by a diffusion process X of the type

$$dX_t = b(t, X_t)\,dt + \sigma(t)\,dW_t,\ X_0 = x_0,\ t \in [0, 1], \tag{1.1}$$

observed at times i/n, $i = 0, \ldots, n$, where $(W_t, 0 \leqslant t \leqslant 1)$ is a standard Wiener process and $x_0 \in \mathbb{R}$. Genon-Catalot *et al.* (1992) proposed a nonparametric estimator of $\sigma^2(t)$ based on orthonormal wavelets and studied its asymptotic properties in L_2 error. Soulier (1993) proposed some extensions in L_p. Hoffmann (1997) computed the minimax rate of convergence for both upper and lower bounds. The usual $n^{-s/(1+2s)}$ rate holds. The technicalities are close to nonparametric regression (see, for example, Korostelev and

Tsybakov (1993)); by a change in probability argument, i.e. setting $b = 0$ in (1.1) the process X is Gaussian with independent increments; therefore, standard nonparametric techniques apply.

A more elaborate model, which will be studied in this paper, consists of a space-dependent diffusion coefficient model, where now

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t,\ X_0 = x_0,\ t \in [x_0, 1]. \tag{1.2}$$

The situation becomes substantially more difficult. The identifiability domain (the interval where σ can be estimated) is random, and the achievable accuracy of estimation (the analogous of the design in a regression model) is given by the observed process $X^{(n)} = (X_{i/n},\ i = 0, \ldots, n)$ itself. The first researcher to study this model was Florens-Zmirou (1993), who proposed a consistent and asymptotically normal estimator. However, the results obtained could not be linked to the smoothness of the parameter $\sigma^2(x)$ and the minimax properties of the model were left open.

The aim of this paper is to fill in this gap from the minimax theory point of view. We define a suitable minimax framework to work with (Definition 2 below) by comparing the model driven by (1.2) with a regression framework, with random design. We exhibit the asymptotic minimax rate of convergence for a global L_p loss ($1 \leqslant p < +\infty$) for an unknown function σ lying in a Besov space. This choice of function spaces is motivated by the fact that, in regression or density estimation, the case of Besov spaces is optimal (Kerkyacharian and Picard 1993). Another interesting point is that wavelet bases offer unconditional bases for Besov spaces. We propose a linear estimating procedure based on wavelets which is optimal. The numerical properties offered by multiscale schemes (in particular, wavelets on the interval (Cohen *et al.* 1994)) suggest fast practical implementation.

1.2. Outline

We investigate the functional estimation of the diffusion coefficient $\sigma^2(x)$ in the one-dimensional model driven by the stochastic differential equation (1.2). The starting point x_0 is fixed and $\sigma(x)$ and $b(t, x)$ are unknown. The sample path $(X_t,\ 0 \leqslant t \leqslant 1)$ is discretely observed at equidistant times $i/n,\ i = 0, \ldots, n$.

Let us first describe the heuristics upon which our procedure relies. An underlying idea is that the diffusion coefficient can be recovered through the quadratic variation of the process X, which leads us to *nonparametric regression* (with random design), a paradigmatic example of well-known statistical models. More precisely, put temporarily $b = 0$ in (1.2) for simplicity and set

$$Y_{i/n} = n(X_{(i+1)/n} - X_{i/n})^2 = n\int_{i/n}^{(i+1)/n} \sigma^2(X_s)\,\mathrm{d}s + \epsilon_{i/n}, \quad i = 0, \ldots, n-1. \tag{1.3}$$

where

$$\epsilon_{i/n} = n\left(\int_{i/n}^{(i+1)/n} \sigma(X_s)\,\mathrm{d}W_s\right)^2 - n\int_{i/n}^{(i+1)/n} \sigma^2(X_s)\,\mathrm{d}s.$$

The $\epsilon_{i/n}$ are uncorrelated centred variables (in fact, martingale increments) which may be viewed as noise terms when estimating $n\int_{i/n}^{(i+1)/n}\sigma^2(X_s)\,\mathrm{d}s$ from the observation $Y_{i/n}$. Provided that there is some smoothness condition on σ^2, the quantity $\sigma^2(X_{i/n})$ can be recovered from

$$n\int_{i/n}^{(i+1)/n} \sigma^2(X_s)\,\mathrm{d}s$$

up to a negligible error. In other words, we can translate our problem in a nonparametric regression setting, i.e. try to estimate the whole function σ^2 from the observation of $(X_{i/n}, Y_{i/n}, 0 \leqslant i \leqslant n-1)$ in the model

$$Y_{i/n} \simeq \sigma^2(X_{i/n}) + \epsilon_{i/n}, \quad i = 0, \ldots, n-1. \tag{1.4}$$

However, the model suggested by (1.3) differs from usual regression frameworks as follows.

(1) The $X_{i/n}$ which play the role of the observation points in (1.4) are not independent and identically distributed (i.i.d.) variables. Moreover, the domain which is asymptotically covered by the observation points is random itself. In classical nonparametric regression with random design, whenever global rates are studied (Stone 1982; Hall 1984; Korostelev and Tsybakov 1993) the observation points are assumed to have a density f, bounded away from zero in some compact interval, say D, where the estimation is to be performed. In our case, the density $f(x)$ at some point $x \in D$ is given by the local time of the process X at x up to time 1, namely

$$L^x = \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^1 1_{|X_s - x|\leqslant\epsilon}\,\mathrm{d}s. \tag{1.5}$$

In order to obtain global rates of convergence, we shall assume that the local time of X is bounded from below on D (with $x_0 \in D$) and therefore study the risk of an estimator *conditionally on the event "L^x bounded away from zero on D"*.

(2) A second major feature is that the noise variables $\epsilon_{i/n}$ are not independent from the design points $X_{i/n}$ nor i.i.d. To face this, we enhance the martingale structure of the $\epsilon_{i/n}$ by considering an increasing sequence $T_0 \leqslant T_1 \leqslant \cdots \leqslant T_{k_n}$ of stopping times of the discrete time filtration $\mathscr{F}_i^n = \sigma(X_s, 0 \leqslant s \leqslant i/n)$, $i = 0, \ldots, n$ and we look for an estimator of the form

$$\hat{\sigma}_n^2(x) = \sum_{i=0}^{k_n} G_{n,i}(X_{T_i}, X_{T_i+1/n}, x). \tag{1.6}$$

Considering estimates of this form will enable us to get rid of the spatial inhomogeneity of the observation points $X_{i/n}$ and still treat the noise terms as in classical regression by means of a martingale version of the Rosenthal inequality (Hall and Heyde 1980). The precise form of the $G_{n,i}$ and k_n will be given in Section 3. Further, we construct an estimator on the whole domain D by means of wavelets on the interval, as developed by Cohen *et al*. (1994), which avoids the effects of boundaries.

Let us also mention that our model can be extended by a classical argument, derived from the Itô formula, to the more general equation

$$dX_t = b(t, X_t)\,dt + h(t)\sigma(X_t)\,dW_t, \quad X_0 = x_0,\ t \in [0, 1],$$

where h is a known function which is assumed to be smooth (see, for example, Genon-Catalot *et al*. (1992)).

1.3. Contents

Section 2 describes the model and hypotheses. Section 3 presents the construction of the estimator and the derived minimax results. We show that, for $s > 2$, the rate $n^{-s/(1+2s)}$ measured in some L_p norm ($1 \leqslant p < \infty$) is a minimax lower bound over Besov balls $B_{sp\infty}$ (see below) and that this rate is attained by our estimator for $s > 1 + 1/p$. Sections 4 and 5 are devoted to the proofs. Some additional results and comments are given in Section 6. Appendix 1 contains an auxiliary result on the rate of convergence in L_p of the empirical local time and recalls some definitions about wavelets and Besov spaces.

2. Statistical model

We consider the discrete observation $X^{(n)} = (X_0, X_{1/n}, \ldots, X_1)$ defined through the stochastic differential equation (1.2) defined in Section 1 and denote by $P_{\sigma,b}$ the law on the space of continuous functions under which the canonical process $(X_t, 0 \leqslant t \leqslant 1)$ is a solution of (1.2). We consider a compact interval D. We denote by $B_{sp\infty}(D)$ the Besov space on the interval D (the restriction of the functions of the space $B_{sp\infty}(\mathbb{R})$ to D) and by $\|.\|_{sp\infty}$ the Besov norm over D (see, for example, Peetre (1976) and Appendix 1 below). For $M > 0$ we write $B_{sp\infty}(D; M)$ for the ball of radius M of $B_{sp\infty}(D)$, i.e.

$$B_{sp\infty}(D; M) = \{f \in B_{sp\infty}(D): \|f\|_{sp\infty} \leqslant M\}.$$

For $s > 1 + 1/p$, $1 \leqslant p < \infty$ we make the following assumptions.

Assumption 1. *σ is positive, non-vanishing, and σ^2 belongs to*

$$V_{sp}(\mathbf{M}) = \{f \in \mathscr{C}^1(\mathbb{R}): 0 < M_0 \leqslant f(x) \leqslant M_1,\ \|f'\|_\infty \leqslant M_2,\ f_{|D} \in B_{sp\infty}(D, M_3)\}$$

where $\mathbf{M} = (M_0, \ldots, M_3)$ is a given (multivariate) constant.

Assumption 2. *The drift b is continuous and belongs to the class $\mathscr{H} = \mathscr{H}(\tilde{M})$ of functions of uniform linear growth, i.e. such that*

$$\forall (t, x) \in [0, 1] \times \mathbb{R}: b^2(t, x) \leqslant \tilde{M}^2(1 + x^2)$$

where $\tilde{M}$ is a given constant.

Remarks.

(1) Assumptions 1 and 2 imply the existence and uniqueness of a strong solution for (1.2) (Zvonkin 1974).

(2) The conditions $s > 1 + 1/p$ ensures that f is $\mathscr{C}^1$ over D.

(3) Assumption 2 is a rather technical condition. Because of Assumption 1, it implies that there exists $\tau > 0$ such that

$$\sup_{0 \leqslant t \leqslant 1} \mathrm{E}_{\sigma,b}[\mathrm{e}^{\tau X_t^2}] < \infty$$

(see, for example, Lipster and Shiryayev (1977, Theorem 4.7)). Actually, this condition holds uniformly over $V_{sp}(\mathbf{M}) \times \mathscr{H}$ and allows one to consider the drift as a nuisance parameter which does not interfer in the estimation problem (see the proof of Proposition 4).

We estimate σ^2 over D. Let L^x denote the local time of X at x up to time 1, as defined by (1.5). Set $L^D = \inf_{x \in D} L^x$. Considering the arguments given in Section 1.2, we define the following criterion for the accuracy of estimation.

Definition 1. *For $\nu > 0$ the L_p risk of an estimator $\hat{\sigma}_n^2$ under the constraint $V_{sp}(\mathbf{M})$ conditionally on the event $(L^D \geqslant \nu)$ is*

$$R_n(\hat{\sigma}_n^2, V_{sp}(\mathbf{M}), \nu) = \sup_{(\sigma^2, b) \in V_{sp}(\mathbf{M}) \times \mathscr{H}} \mathrm{E}_{\sigma,b}\left(\int_D |\sigma^2(x) - \hat{\sigma}_n^2(x)|^p \, \mathrm{d}x | L^D \geqslant \nu \right). \tag{2.1}$$

3. Main results

3.1. Lower bounds

Proposition 1. *Let $\mathscr{F}$ denote the set of all estimators constructed from the observation $X^{(n)}$. Suppose that Assumptions 1 and 2 hold. For $s > 2$ and every $p \in [1, \infty[$, there exists a constant $C_1 = C_1(s, p, \mathbf{M}, \nu)$ such that*

$$\inf_{\hat{\sigma}_n^2 \in \mathscr{F}} R_n(\hat{\sigma}_n^2, V_{sp}(\mathbf{M}), \nu) \geqslant C_1 n^{-sp/(1+2s)}.$$

Remarks.

(1) The classical rate for nonparametric models such as density estimation or regression is a minimax lower bound.

(2) The method that we shall employ in the proof requires the smoothness condition $s > 2$, which is more restrictive than the initial assumption $s > 1 + 1/p$ for $p > 1$.

3.2. Upper bounds

We begin this section by recalling some classical facts about the use of wavelets in nonparametric regression. We then give the construction of our estimator and a bound for the risk of Definition 1.

3.2.1. Preliminary results

Consider the model given by

$$Y_i = f(x_i) + \epsilon_i, \quad i = 0, \ldots, n, \tag{3.1}$$

where f is a smooth function defined on the interval D (for instance f belongs to some Besov ball $\mathscr{V}$ of the space $B_{sp\infty}(D)$, with $s > 1$, $p \in [1, \infty[$) and $\epsilon_i \sim_{iid} \mathcal{N}(0, 1)$ is a standard Gaussian noise. The x_i are equally spaced on D. Without loss of generality, we may assume that $D = [0, 1]$, hence $x_i = i/n$.

Let $(V_j, j \in \mathbf{Z})$ be a multiresolution analysis of $L_2(D)$, generated by a smooth, compactly supported orthonormal scaling function φ (see Appendix 1 for a precise definition of V_j and φ) with correction on the boundaries of D (Cohen *et al*. 1994). We assume that the length support of φ is an integer N_0 and we denote by φ^l_k, φ^r_k, $k = 0, \ldots, N_0 - 1$ the left and right edge scaling functions on the boundary of D respectively. Set also $\varphi_{jk} = 2^{j/2}\varphi(2^j x - k)$ and $\varphi^{\#}_{jk}(x) = \varphi^{\#}_k(2^j x)$, for $\# = l, r$.

The linear wavelet estimator of f on D is constructed as follows.

(1) We first choose an integer J such that $2^J \geqslant 2N_0$. According to the Cohen–Daubechies–Vial algorithm, we approximate f by $P^{[0,1]}_j f$ for $j \geqslant J$, where

$$P^{[0,1]}_j f(x) = \sum_{k=0}^{N_0-1} \alpha^l_{jk}\varphi^l_{jk}(x) + \sum_{k \in s_j} \alpha_{jk}\varphi_{jk}(x) + \sum_{k=0}^{N_0-1} \alpha^r_{jk}\varphi^r_{jk}(x).$$

The set of indices s_j defines the interior functions on D: $k \in s_j$ if and only if $\mathrm{supp}\varphi_{jk} \subset D$. We take $2^J \geqslant 2N_0$ so that the left and right edge functions φ^l_{jk} and φ^r_{jk} do not interact.

(2) We then estimate the wavelets coefficients α_{jk}, α^i_{jk}, $i = l, r$, for $j \geqslant J$, where

$$\alpha_{jk} = \int_D f(x)\varphi_{jk}(x)\,\mathrm{d}x, \; \alpha^{\#}_{jk} = \int_D f(x)\varphi^{\#}_{jk}(x)\,\mathrm{d}x.$$

A standard procedure consists in estimating α_{jk} by its empirical wavelet coefficient

$$\hat{\alpha}_{jk} = \frac{1}{n}\sum_i Y_i\varphi_{jk}\left(\frac{i}{n}\right) \tag{3.2}$$

and proceed analogously for $\alpha^{\#}_{jk}$, $\# = l, r$.

The asymptotic results are obtained when letting j and n tend to ∞. Choosing $j = j_n$ such that $2^{j_n} \asymp n^{1/(1+2s)}$ ($a_n \asymp b_n$ means that there exist two positive constants A and B independent of n such that $Aa_n \leqslant b_n \leqslant Ba_n$) the wavelet estimator

$$\hat{f}_n(x) = \sum_{k=0}^{N_0-1} \hat{\alpha}^l_{j_nk}\varphi^l_{j_nk}(x) + \sum_{k\in s_j} \hat{\alpha}_{jnk}\varphi_{j_nk}(x) + \sum_{k=0}^{N_0-1} \hat{\alpha}^r_{j_nk}\varphi^r_{j_nk}(x)$$

is optimal in the minimax sense:

$$\sup_{f\in V} \mathrm{E}\left(\int_D |\hat{f}(x) - f_n(x)|^p \,\mathrm{d}x\right) \leqslant Cn^{-sp/(1+2s)} \tag{3.3}$$

for some constant $C = C(\mathscr{V}, \varphi)$.

Suppose that the x_i are no longer equispaced on D. There is no reason why $\hat{f}_n$ should converge to f in any sense. Of course, one may wish to alter the procedure in (3.2) by considering the empirical wavelet coefficient along the design generated by the x_i, namely

$$\hat{\alpha}_{jk} = \sum_i Y_i\varphi_{jk}(x_i)(x_{i+1} - x_i). \tag{3.4}$$

Such an estimator will still have good minimax properties provided that some accurate control on $\sup_i(x_{i+1} - x_i)$ is ensured, merely $\sup_i(x_{i+1} - x_i) = O(n^{-1})$. If such a condition is out of reach, say for technical reasons (and this will be the case for diffusion processes) one may remark that, if $f(x_i) - f(i/n)$ is small, or equivalently (as f is smooth) if x_i is close to i/n, the estimator given by (3.2) will still enjoy good convergence properties. More precisely, one can check that, if

$$\sup_{i\leqslant n}\left|x_i - \frac{i}{n}\right| \leqslant Kn^{-s/(1+2s)} \tag{3.5}$$

for an absolute constant K, (3.3) still holds. This means that we do not need to bound the distance between two successive observation points if we have insight into the number of points located asymptotically around any given level. Hence, under condition (3.5), we can simply take the estimate

$$\hat{\alpha}_{jk} = \frac{1}{n}\sum_i Y_i\varphi_{jk}\left(\frac{i}{n}\right)$$

although the considered data are non-equally spaced. Further data on irregular samplings following this approach has been used for instance by Hoffmann (1997).

3.2.2. Construction of an estimator for the diffusion coefficient

3.2.2.1. Preliminaries. We shall henceforth assume that $D = [0, 1]$. The general case is obtained by dilating and translating the unit interval, the difficulty being merely notational.

Let us be given $\nu \in\,]0, 1[$ and choose a threshold $h_n > 0$. We divide D into $\lfloor h_n^{-1}\rfloor$ identical boxes of size h_n, denoted by C_λ, $\lambda = 1, \ldots, \lfloor h_n^{-1}\rfloor$. The convergence of the empirical sampling measure to the local time, namely

$$\frac{1}{nh_n}\sum_{i=0}^{n} 1_{|X_{i/n}-x|\leqslant h_n/2} \to L^x$$

if $h_n \to 0$ and $nh_n \to \infty$ will ensure that on the event $(L^D \geqslant \nu)$, $\lfloor nh_n\nu \rfloor$ observation points $X_{i/n}$ will lie (at least) in each C_λ with high probability.

We keep in the $\lfloor nh_n\nu \rfloor$ first observation points $X_{i/n}$ hitting each box C_λ and apply the empirical wavelet transform to the considered $(X_{i/n}, Y_{i/n})$, with

$$Y_{i/n} = n(X_{(i+1)/n} - X_{i/n})^2 \simeq \sigma^2(X_{i/n}) + \epsilon_{i/n}. \tag{3.6}$$

The error of location due to the non-equally spaced design generated by the subsampled $X_{i/n}$ is controlled by the threshold h_n (recall 3.5). The admissible i/n lie in a random set.

3.2.2.2. The algorithm. Define

$$N_i^\lambda = \left(\sum_{j\leqslant i} 1_{x_{j/n}\in C_\lambda}\right) \wedge \lfloor nh_n\nu \rfloor \tag{3.7}$$

as the C_λ *counter* stopped when exactly $\lfloor nh_n\nu \rfloor$ observation points lie in C_λ. Set

$$T_1 = 0, \text{ and for } i \geqslant 2\colon T_i = \inf\left\{\frac{j}{n} > T_{i-1}\colon \sum_\lambda (N_j^\lambda - N_{T_{i-1}}^\lambda) \geqslant 1\right\} \wedge 1. \tag{3.8}$$

The T_i, $i = 1, \ldots, \lfloor n\nu \rfloor$ are increasing $(\mathscr{F}_i^n)$-stopping times which correspond to the times when the boxes are filled up to $\lfloor nh_n\nu \rfloor$ points each. The major point of interest is that (recall (1.3)) the process $\sum_{j\leqslant i}\epsilon_{T_j}$ remains a $(\mathscr{F}_{T_{i+1}}^n)$ martingale.

We extract from $X^{(n)}$ the subsampling $(X_{T_1}, \ldots, X_{T_{\lfloor n\nu \rfloor}})$ and apply the empirical wavelet transform to the (X_{T_i}, Y_{T_i}) suggested by (3.2) on the regular grid $(i/\lfloor n\nu \rfloor, i = 0, \ldots, \lfloor n\nu \rfloor)$, with

$$Y_{T_i} = n(X_{T_i+1/n} - X_{T_i})^2, \; i = 1, \ldots, \lfloor n\nu \rfloor. \tag{3.9}$$

A last technical difficulty is that the $X_{i/n}$ do not appear of course in an increasing order. We cannot link (X_{T_i}, Y_{T_i}) to $\lfloor i/n\nu \rfloor$ directly. For a given box C_λ, the points on the regular design are the

$$(\lambda - 1)h_n + \frac{l}{\lfloor n\nu \rfloor}, \; l = 1, \ldots, \lfloor nh_n\nu \rfloor.$$

We set λ_{T_i} for the index of the box C_λ in which X_{T_i} falls. For a given (X_{T_i}, Y_{nT_i}), the corresponding point on the regular grid will be defined as

$$x_{T_i} = (\lambda_{T_i} - 1)h_n + \frac{l_{T_i}}{\lfloor n\nu \rfloor}, \tag{3.10}$$

where $l_{T_i} = \#\{X_{T_j} \in C_{\lambda_{T_i}}, j \leqslant i\}$.

Note that this choice of sampling points x_{T_i} on the uniform grid at courser level $1/\nu$ provides us with $(\mathscr{F}_{T_i}^n)$-measurable x_{T_i}.

Definition 2. *The wavelet coefficient estimator at level ν is*

$$\hat{\alpha}_{jk} = \frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} Y_{T_i} \varphi_{jk}(x_{T_i}). \tag{3.11}$$

We define the edge wavelet coefficients estimate analogously. The estimator of σ^2 on D is then

$$\hat{\sigma}_n^2(x) = \sum_{k=0}^{N_0-1} \hat{\alpha}_{jk}^l \varphi_{jk}^l(x) + \sum_{k \in s_j} \hat{\alpha}_{jk} \varphi_{jk}(x) + \sum_{k=0}^{N_0-1} \hat{\alpha}_{jk}^r \varphi_{jk}^r(x).$$

We may now state our result on upper bounds. For technical convenience, we work with the wavelet estimator at level $\nu/2$.

Proposition 2. *Let $p \in [1, \infty[$. Suppose that Assumptions 1 and 2 hold. Let $\hat{\sigma}_n^2$ be the estimator given by Definition 2 at level $\nu/2$. If $2^{j_n} \asymp n^{1/(1+2s)}$ and $h_n \asymp n^{-s/(1+2s)}$, then there exists $C_2 = C_2(\varphi, s, p, \mathbf{M}, \tilde{M}, \nu)$ such that*

$$R_n(\hat{\sigma}_n^2, V_{sp}(\mathbf{M}), \nu) \leqslant C_2 n^{-sp/(1+2s)}.$$

Corollary 1. *The minimax rate of convergence for the minimax risk defined by* (2.1) *over Besov balls for $s > 2$ and $p \in [1, +\infty[$ is the classical $n^{-sp/(1+2s)}$ and is attained by our estimator.*

4. Lower bounds; proof of Proposition 1

4.1. Sketch of the proof

We follow a classical method in nonparametric estimation, restricting ourselves to a hypercube of $V_{sp}(\mathbf{M})$. We refer to Korostelev and Tsybakov (1993) for general results on proving lower bounds and to Kerkyacharian and Picard (1992; 1993) for the specific use of wavelets and Besov spaces in this context. We outline the difficulties encountered when considering the case of diffusion processes. Let P^ν denote the probability measure conditioned on the event $(L^D \geqslant \nu)$ and let E^ν denote the corresponding expectation. Note that

$$R_n(\hat{\sigma}_n^2, V_{sp}(\mathbf{M}), \nu) \geqslant \sup_{\sigma^2 \in C_{jn}, b=0} \mathrm{E}^\nu \{\|\hat{\sigma}^2 - \sigma^2\|_p^p\},$$

where C_{j_n} is some parametric set included in $V_{sp}(\mathbf{M})$ of size 2^{j_n}, with j_n increasing as $n \to \infty$. We shall henceforth consider the model without drift.

4.1.1. *Constructing a hypercube of $V_{sp}(\mathbf{M})$*

For technical convenience, we suppose that $M_0 < 1 < M_1$ without loss of generality. Let ψ be a wavelet (Meyer 1990) of regularity $r > s$, with compact support included in $[-A, A]$, where A is a fixed integer. We set

$$C_{j_n}(\gamma_n) = \left\{ \sigma_\epsilon^2(x) = 1 + \gamma_n \sum_{k \in K_{j_n}} \epsilon_k \psi_{j_n k}(x),\ \epsilon_k = \pm 1,\ k = 1, \ldots, 2^{j_n} \right\},$$

where γ_n is a positive number which measures the size of oscillations of the cube. We define $K_{j_n} = \{A + 2kA,\ k = 0, \ldots, 2^{j_n} - 1\}$ and $\psi_{j_n k} = 2^{j_n/2}\psi(2^{j_n}x - k)$ so that $\psi_{j_n k}$ and $\psi_{j_n k'}$ have disjoint supports for $k \neq k'$. Thus we disturb an original function identically equal to 1 by adding $\gamma_n \psi_{j_n k}$, for $k = 1, \ldots, 2^{j_n}$.

We look for conditions in order to have $C_{j_n}(\gamma_n) \subset V_{sp}(\mathbf{M})$. From the definition of Besov spaces in terms of wavelet sequences (see Appendix 1), this is satisfied if

$$\gamma_n \leqslant \frac{M_1}{2} 2^{-j_n(s+1/2)} \quad \text{and} \quad \gamma_n \leqslant \frac{1 - M_0}{\|\psi\|_\infty} 2^{-j_n/2}. \tag{4.1}$$

4.1.2. Bounds on the minimax risk

The crucial point is to find a condition on (the order of magnitude of) γ_n in order to bound the likelihood ratio induced by two generic points of $C_{j_n}(\gamma_n)$. More precisely, let P_+ (or P_-) denote the law of a sample of observation, derived from a model with a diffusion coefficient $\sigma_+^2 = 1 + \gamma_n \sum_{k' \neq k} \epsilon_{k'} \psi_{j_n k'} + \gamma_n \psi_{j_n k}$ (or $\sigma_-^2 = 1 + \gamma_n \sum_{k' \neq k} \epsilon_{k'} \psi_{j_n k'} - \gamma_n \psi_{j_n k}$), for some fixed $k \in K_{j_n}$ and $\epsilon_{k'}$, $k' \neq k$. Let $\Lambda(\sigma_+, \sigma_-, X^{(n)})$ denote the likelihood ratio $(\mathrm{d}P_+/\mathrm{d}P_-)(X^{(n)})$. If we prove that there exist positive λ and p_0, independent of n such that for sufficiently large n

$$P_-^\nu(\Lambda(\sigma_+, \sigma_-, X^{(n)}) > \mathrm{e}^{-\lambda}) \geqslant p_0 > 0, \tag{4.2}$$

then (Korostelev and Tsybakov 1993) we can derive the following bound:

$$\inf_{\hat\sigma_n^2 \in \mathscr{F}} R_n(\hat\sigma_n^2, V_{sp}(\mathbf{M}), \nu) \geqslant 2^{j_n p/2} \gamma_n^p \|\psi\|_p^p \, \mathrm{e}^{-\lambda} \, \frac{p_0}{2}. \tag{4.3}$$

The conditions on j_n and γ_n exhibited in (4.1) and (4.2) will therefore provide a lower bound.

4.1.3. Control of the likelihood ratio

For $\delta > 0$, let $\tilde\Lambda$ denote the likelihood ratio associated with the Markov process of transition semigroup

$$\tilde p_\delta(x, \mathrm{d}y) = \frac{1}{(2\pi\delta)^{1/2}} \frac{1}{\sigma(x)} \exp\left(-\frac{1}{2\delta} \frac{(y - x)^2}{\sigma^2(x)} \right) \mathrm{d}y. \tag{4.4}$$

In our time equispaced design we take $\delta = 1/n$. We first prove an intermediate result for $\tilde\Lambda(\sigma_+, \sigma_-, X^{(n)})$, when the observation $X^{(n)}$ is taken under P_- and j, γ are fixed (independent of n).

Lemma 1. *Assume that j and γ are fixed. The following expansion holds:*

$$\log \tilde\Lambda(\sigma_+, \sigma_-, X^{(n)}) = n^{1/2}\gamma U_{n,j} - \tfrac{1}{2} n \gamma^2 V_{n,j} + n^{3/2} \gamma^3 R_{n,j}(\gamma), \tag{4.5}$$

where $U_{n,j}$ converges in distribution under $P_- = P_{\sigma_-,0}$ to U_j as n goes to $+\infty$, U_j is a centred mixed normal variable, with conditional variance $V_j = 2\int_0^1 \psi_{jk}^2(X_s)\,\mathrm{d}s$, $V_{n,j}$ converges in P_--probability to V_j and $R_{n,j} \to 0$ in P_- measure (uniformly in γ) as n goes to infinity.

Remark. This result is not surprising if one recalls that model (1.2) is parametrically local asymptotic mixed normality (Donhal 1987; Genon-Catalot and Jacod 1993) since, for fixed j, we are in a parametric submodel. We shall not use directly Lemma 1 for the proof of Proposition 1, but we emphasize expansion (4.5) to provide our intuition for the model and for the proof of Lemma 2.

Lemma 2. *Let $j = j_n$ and $\gamma = \gamma_n$ and assume that $\gamma_n \asymp 1/n^{1/2}$ and $2^{5j_n/2}/n^{1/2} \to 0$ as $n \to \infty$. Then*

(i) *there exist positive λ and p_0 such that, for sufficiently large n,*

$$P_-^{\nu}(\tilde{\Lambda}(\sigma_+, \sigma_-, X^{(n)}) > \mathrm{e}^{-\lambda}) \geqslant p_0 > 0 \tag{4.6}$$

and

(ii) (4.6) *remains true when replacing $\tilde{\Lambda}$ by Λ, for a modification of the constants λ and p_0.*

4.1.4. Completion of proof of Proposition 1

We take $\gamma_n \asymp 1/n^{1/2}$ in (4.1). This leads to

$$2^{j_n} \asymp n^{1/(1+2s)}. \tag{4.7}$$

The condition $2^{5j_n/2}/n^{1/2} \to 0$ as $n \to \infty$ is satisfied since $s > 2$. From Lemma 2 and (4.3), we deduce that

$$\inf_{\hat{\sigma}_n^2 \in \mathscr{F}} R_{n,\nu}(\hat{\sigma}_n^2, V) \geqslant C_1 n^{-sp/(1+2s)} \tag{4.8}$$

where $C_1 = C_1(s, p, \mathbf{M}, \nu)$. The proof is complete. □

4.2. Proof of Lemma 1

For notational simplicity, we shall write X_i instead of $X_{i/n}$, P (or E) instead of P_- (or E_-) and set $\Delta X_i = X_{(i+1)/n} - X_{i/n}$. The quantity C will denote an generic constant depending on s, p, $\mathbf{M}$ and ψ which may vary at each occurrence. We shall also write $\tilde{\Lambda}$ for $\tilde{\Lambda}(\sigma_+, \sigma_-, X^{(n)})$ when no confusion is possible. One has, under P,

$$(\Delta X_i)^2 = \left(\int_{i/n}^{(i+1)/n} \sigma_-(X_s)\,\mathrm{d}W_s\right)^2 = \sigma_-^2(X_i)(\Delta W_i)^2 + R_{i,n}^{(1)}. \tag{4.9}$$

Consequently

$$\log \tilde{\Lambda} = \sum_{i=0}^{n-1} \log \frac{\sigma_-}{\sigma_+}(X_i) - \frac{1}{2\delta} \sum_{i=0}^{n-1} \left\{ \left(\frac{\sigma_-}{\sigma_+} \right)^2 (X_i) - 1 \right\} (\Delta W_i)^2 + R_n^{(2)}$$

with

$$R_n^{(2)} = \frac{1}{2\delta} \sum_{i=0}^{n-1} \left(\frac{1}{\sigma_+^2} - \frac{1}{\sigma_-^2} \right) (X_i) R_{i,n}^{(1)}.$$

A second-order Taylor expansion yields

$$\log \frac{\sigma_-}{\sigma_+}(X_i) = -\gamma \psi_{jk}(X_i) + R_{i,n}^{(3)}$$

$$-\frac{1}{2} \left\{ \left(\frac{\sigma_-}{\sigma_+} \right)^2 (X_i) - 1 \right\} = \gamma \psi_{jk}(X_i) - \gamma^2 \psi_{jk}^2(X_i) + R_{i,n}^{(4)}.$$

Since $\delta = 1/n$

$$\log \tilde{\Lambda} = \sum_{i=0}^{n-1} \gamma \psi_{jk}(X_i) \{ n(\Delta W_i)^2 - 1 \} - \sum_{i=0}^{n-1} n \gamma^2 \psi_{jk}^2(X_i) (\Delta W_i)^2 + R_n^{(5)} \tag{4.10}$$

where

$$R_n^{(5)} = R_n^{(2)} + \sum_{i=0}^{n-1} (R_{i,n}^{(3)} + R_{i,n}^{(4)}).$$

Define now

$$\epsilon_i = n(\Delta W_i)^2 - 1,$$

$$U_{n,j} = \frac{1}{n^{1/2}} \sum_{i=0}^{n-1} \psi_{jk}(X_i) \epsilon_i,$$

$$V_{n,j} = 2 \sum_{i=0}^{n-1} \psi_{jk}^2(X_i) (\Delta W_i)^2,$$

$$R_{n,j}(\gamma) = (n^{3/2} \gamma^3)^{-1} R_n^{(5)}.$$

Equation (4.10) can therefore be written as (4.5). Thus Lemma 1 is proved provided that the following convergences hold:

$$V_{n,j} \xrightarrow{P} 2 \int_0^1 \psi_{jk}^2(X_s) \, \mathrm{d}s \quad \text{as} \quad n \to +\infty, \tag{4.11}$$

$$U_{n,j} \to U_j \quad \text{as} \quad n \to +\infty, \tag{4.12}$$

in the P distribution, where U_j is a mixed normal centred variable with conditional variance V_j, and

$$R_{n,j}(\gamma) \xrightarrow{P} 0 \quad \text{as} \quad n \to +\infty. \tag{4.13}$$

□

Proof of (4.13). Recall that

$$R_{i,n}^{(1)} = \left(\int_{i/n}^{(i+1)/n} \sigma_-(X_s)\,\mathrm{d}W_s\right)^2 - \sigma_-^2(X_i)(\Delta W_i)^2,$$

$$R_n^{(2)} = \frac{1}{2\delta}\sum_{i=0}^{n-1} T_{n,i}R_{i,n}^{(1)}, \quad \text{with } T_{n,i} = \left(\frac{1}{\sigma_+^2} - \frac{1}{\sigma_-^2}\right)(X_i).$$

$$R_n^{(5)} = R_n^{(2)} + \sum_{i=0}^{n-1}(R_{i,n}^{(3)} + R_{i,n}^{(4)}).$$

Thus, we shall successively prove that

$$\frac{1}{n^{3/2}\gamma^3} R_n^{(2)} \xrightarrow{P} 0, \tag{4.14}$$

$$\frac{1}{n^{3/2}\gamma^3}\sum_{i=0}^{n-1}(R_{i,n}^{(3)} + R_{i,n}^{(4)}) \xrightarrow{P} 0. \tag{4.15}$$

□

Proof of (4.14). We use the following lemma from Genon-Catalot and Jacod (1993).

Lemma 3. *Let χ_i^n, U be random variables, the χ_i^n being $\mathscr{F}_{i+1}^n$ measurable. The following two conditions imply that $\sum_{i=0}^{n-1}\chi_i^n \xrightarrow{P} U$:*

$$\sum_{i=0}^{n-1} \mathrm{E}(\chi_i^n|\mathscr{F}_i^n) \xrightarrow{P} U,$$

$$\sum_{i=0}^{n-1} \mathrm{E}(|\chi_i^n|^2|\mathscr{F}_i^n) \xrightarrow{P} 0.$$

We have to check that

$$\frac{1}{n^{3/2}\gamma^3}\frac{n}{2}\sum_{i=0}^{n-1} \mathrm{E}(T_{n,i}R_{n,i}^{(1)}|\mathscr{F}_i^n) \xrightarrow{P} 0$$

$$\frac{1}{n^3\gamma^6}\frac{n^2}{4}\sum_{i=0}^{n-1} \mathrm{E}(|T_{n,i}R_{n,i}^{(1)}|^2|\mathscr{F}_i^n) \xrightarrow{P} 0.$$

Elementary computation yields

$$\mathrm{E}(T_{n,i}R^{(1)}_{n,i}|\mathscr{F}^n_i) = -\frac{2\gamma\psi_{jk}(X_i)}{\sigma^2_-(X_i)\sigma^2_+(X_i)}\mathrm{E}\left(\int_{i/n}^{(i+1)/n}\{\sigma^2_-(X_s)-\sigma^2_-(X_i)\}\,\mathrm{d}s|\mathscr{F}^n_i\right). \tag{4.16}$$

Since the choice of the wavelet ψ is free, we may assume that ψ is twice differentiable, hence

$$\sigma^2_-(X_s)-\sigma^2_-(X_i) = (\sigma^2_-)'(X_i)(X_s-X_i)+Z_{n,i},$$

the remainder term $Z_{n,i}$ satisfying

$$|Z_{n,i}| \leqslant C\gamma^2 2^{5j/2}|X_s-X_i|^2.$$

We apply the Fubini theorem and we use the fact that $(X_t,\, 0 \leqslant t \leqslant 1)$ is a martingale under P_- to obtain

$$\left|\mathrm{E}\left(\int_{i/n}^{(i+1)/n}\{\sigma^2_-(X_s)-\sigma^2_-(X_i)\}\,\mathrm{d}s|\mathscr{F}^n_i\right)\right| \leqslant C\gamma^2 2^{5j/2}\mathrm{E}\left(\int_{i/n}^{(i+1)/n}(X_s-X_i)^2\,\mathrm{d}s|\mathscr{F}^n_i\right).$$

We again apply the Fubini theorem and the Doob inequality to get

$$\mathrm{E}\left(\int_{i/n}^{(i+1)/n}(X_s-X_i)^2\,\mathrm{d}s|\mathscr{F}^n_i\right) \leqslant C\Delta^2_n.$$

Since σ^2_- is bounded from below (Assumption 1), we deduce that

$$|\mathrm{E}(T_{n,i}R^{(1)}_{n,i}|\mathscr{F}^n_i)| \leqslant Cn^{-2}\gamma^3 2^{3j}.$$

It follows that

$$\frac{1}{n^{1/2}\gamma^3}\left|\frac{1}{2}\sum_{i=0}^{n-1}\mathrm{E}(T_{n,i}R^{(1)}_{n,i}|\mathscr{F}^n_i)\right| \leqslant C\left(\frac{2^j}{n^{1/2}}\right)^3.$$

The last quantity converges to 0. The second follows likewise; so we omit it. The proof of (4.14) is finished. □

Proof of (4.15). From the definition of $R^{(l)}_{n,i}$ for $l = 3, 4$

$$R^{(l)}_{n,i} \leqslant C\gamma^3|\psi_{jk}(X_i)|^3. \tag{4.17}$$

This entails

$$\frac{1}{\gamma^3 n^{3/2}}|R^{(5)}_n| \leqslant C\frac{2^{3j/2}}{n^{1/2}}. \tag{4.18}$$

The last quantity converges to 0 from the hypothesis. Equation (4.15) is established, and (4.13) follows. □

Proof of (4.11). From the definition of the quadratic variation of a continuous semimartingale we immediately deduce (4.11). □

Proof of (4.12). We first set $\xi_i^n = (1/n^{1/2})\psi_{jk}(X_i)\epsilon_i$. We have successively

$$\max_{i\leqslant n}|\xi_i^n| \xrightarrow{P} 0, \tag{4.19}$$

$$\mathrm{E}\Big(\max_{i\leqslant n}|\xi_i^n|^2\Big) \text{ bounded,} \tag{4.20}$$

$$\sum_{i=0}^{n-1}|\xi_i^n|^2 \xrightarrow{P} 2\int_0^1 \psi_{jk}^2(X_s)\,\mathrm{d}s. \tag{4.21}$$

Equations (4.19) and (4.20) are straightforward. □

Proof of (4.21). We apply again the lemma of Genon-Catalot and Jacod (1993) (Lemma 3). We must check that

$$\sum_{i=0}^{n-1}\mathrm{E}(|\xi_i^n|^2|\mathscr{F}_i^n) \xrightarrow{P} 2\int_0^1 \psi_{jk}^2(X_s)\,\mathrm{d}s, \tag{4.22}$$

$$\sum_{i=0}^{n-1}\mathrm{E}(|\xi_i^n|^4|\mathscr{F}_i^n) \xrightarrow{P} 0. \tag{4.23}$$

It is easily seen that

$$\sum_{i=0}^{n-1}\mathrm{E}(|\xi_i^n|^2|\mathscr{F}_i^n) = \frac{2}{n}\sum_{i=0}^{n-1}\psi_{jk}^2(X_i) \to 2\int_0^1 \psi_{jk}^2(X_s)\,\mathrm{d}s \text{ almost surely.} \tag{4.24}$$

On the other hand

$$\mathrm{E}(|\xi_i^n|^4|\mathscr{F}_i^n) \leqslant \frac{C2^{2j}}{n^2}. \tag{4.25}$$

So (4.23) follows and (4.21) is proved. □

We are now ready to turn to (4.12) itself. We are in fact under the conditions for the convergence of $U_{n,j}$ to a mixed normal variable with conditional variance V_j (Hall and Heyde 1980, p. 58, Theorem 3.2). Unfortunately, we do not have the nesting condition on the filtrations $(\mathscr{F}_i^n)$ which is necessary to accommodate the random limit. This difficulty has already been encountered in the paper by Genon-Catalot and Jacod (1993) and was solved by a martingale characterization limit theorem. The same arguments can be used in our setting and we refer to their paper for the method. The proof of Lemma 1 is complete.

4.3. Proof of Lemma 2

4.3.1. Proof of (i)

For simplicity, we take $\gamma = 1/n^{1/2}$. Recall that now P_- depends on n through $j = j_n$ and γ_n. Note that

$$(\tilde{\Lambda}(\sigma_+, \sigma_-, X^{(n)}) > \mathrm{e}^{-\lambda} \cap L^D \geqslant \nu) \supseteq (|\log \tilde{\Lambda}(\sigma_+, \sigma_-, X^{(n)})| \leqslant \lambda \cap L^D \geqslant \nu).$$

Using the Chebyshev inequality

$$\begin{aligned} P_-^\nu(\tilde{\Lambda}(\sigma_+, \sigma_-, X^{(n)}) > \mathrm{e}^{-\lambda}) &\geqslant P_-(|\log \tilde{\Lambda}(\sigma_+, \sigma_-, X^{(n)})| \leqslant \lambda) + P_-(L^D \geqslant \nu) - 1 \\ &\geqslant P_-(L^D \geqslant \nu) - \frac{1}{\lambda}\mathrm{E}_-\{|\log \tilde{\Lambda}(\sigma_-, \sigma_+, X^{(n)})|\}. \end{aligned}$$

The assumption that σ^2 is bounded from below (Assumption 1) implies that

$$C_3 = \inf_{\sigma^2 \in V_{sp}} P_-(L^D \geqslant \nu) > 0. \tag{4.26}$$

For a proper choice of λ (specified after (4.30) below), (i) will follow from

$$\mathrm{E}_-\{|\log \tilde{\Lambda}(\sigma_+, \sigma_-, X^{(n)})|\} \leqslant C_4 < \infty. \tag{4.27}$$

With the notation of Lemma 1, (4.27) is a consequence of the three following bounds:

$$\mathrm{E}_-(|U_{n,j_n})| \leqslant \frac{C_4}{3}, \tag{4.28}$$

$$\mathrm{E}_-(|V_{n,j_n}|) \leqslant \frac{C_4}{3}, \tag{4.29}$$

$$\mathrm{E}_-(|R_{n,j_n})| \leqslant \frac{C_4}{3} \tag{4.30}$$

for some constant C_4. One completes the proof by taking $\lambda > C_4/C_3$. □

Proof of (4.28), (4.29) and (4.30). Since $\sigma > 0$, the random variables $X_{i/n}$ admit a density with respect to the Lebesgue measure, say $p_{i/n}$, which is given (see for instance (4.31) below) by

$$p_{i/n}(x) = \frac{1}{\sigma(x)(2\pi i/n)^{1/2}} r_{i/n}(x_0, x) \exp\left(H(x) - H(x_0) - \frac{\{S(x) - S(x_0)\}^2}{2i/n} \right),$$

where the functions S, H and r are described in the proof of (ii) below. One readily checks that the following bound holds for every real number x:

$$p_{i/n}(x) \leqslant C_5 \left(\frac{n}{i}\right)^{1/2},$$

where C_5 depends on $\mathbf{M}$, and s, p.

For (4.28), we note that the discrete time process $(M_i = \sum_{l=0}^{i} \psi_{j_n k}(X_l)\epsilon_l,\ i = 0, \ldots, n-1)$ is a $(\mathscr{F}_i^n)$ martingale. Hence

$$\mathrm{E}_-\{(U_{n,j_n})^2\} = \sum_{i=0}^{n-1} \mathrm{E}_-\{\psi_{j_n k}^2(X_i)\} \frac{2}{n}$$

after conditioning with respect to $\mathscr{F}_i^n$. On the other side, for $i \geqslant 1$

$$\mathrm{E}\{\psi_{j_n k}^2(X_i)\} = \int_{-\infty}^{+\infty} \psi_{j_n k}^2(x) p_{i/n}(x)\,\mathrm{d}x \leqslant C_5 \left(\frac{n}{i}\right)^{1/2}$$

since ψ is orthonormal in L_2. It follows that

$$\mathrm{E}_-\{(U_{n,j_n})^2\} \leqslant 2C_5 \frac{C}{n^{1/2}} \sum_{i=1}^{n-1} \frac{1}{i^{1/2}}.$$

Since this last quantity is bounded, the proof for U_{n,j_n} is complete.

We now turn to (4.29). Recall that

$$V_{n,j_n} = \sum_{i=0}^{n-1} \psi_{j_n k}^2(X_i)(\Delta W_i)^2.$$

Conditioning with respect to $\mathscr{F}_i^n$, one has

$$\mathrm{E}_-(V_{n,j_n}) = \sum_{i=0}^{n-1} \mathrm{E}_-\{\psi_{j_n k}^2(X_i)\} \frac{1}{n}$$

and we conclude as for (4.28).

Finally, let us prove (4.30). One has

$$R_n^{(2)} = -\sum_{i=0}^{n-1} \frac{2\gamma\psi_{j_n k}(X_i)}{\sigma_-^2(X_i)\sigma_+^2(X_i)} \left\{ \left(\int_{i/n}^{(i+1)/n} \sigma_-(X_s)\,\mathrm{d}W_s \right)^2 - \sigma_-^2(X_i)(\Delta W_i)^2 \right\}.$$

Hence

$$\mathrm{E}_-(|R_n^{(2)}|) \leqslant 2\sum_i \mathrm{E}_- \left[\gamma \frac{|\psi_{j_n k}(X_i)|}{M_1^2} \mathrm{E}_- \left\{ \left| \left(\int_{i/n}^{(i+1)/n} \sigma_-(X_s)\,\mathrm{d}W_s \right)^2 - \sigma_-^2(X_i)(\Delta W_i)^2 \right| \mathscr{F}_i^n \right\} \right].$$

Writing

$$\left(\int_{i/n}^{(i+1)/n} \sigma_-(X_s)\,\mathrm{d}W_s) \right)^2 - \sigma_-^2(X_i)(\Delta W_i)^2 = \int_{i/n}^{(i+1)/n} \{\sigma_-(X_s) - \sigma_-(X_i)\}\,\mathrm{d}W_s \times \int_{i/n}^{(i+1)/n} \{\sigma_-(X_s) + \sigma_-(X_i)\}\,\mathrm{d}W_s,$$

applying the Schwarz and the Doob inequalities and using $\gamma = 1/n^{1/2}$, one easily obtains

$$\mathrm{E}_-(|R_n^{(2)}|) \leqslant C \sum_{i=0}^{n-1} [\mathrm{E}_-\{\psi_{j_n k}^2(X_i)\}]^{1/2} \frac{1}{n}.$$

The conclusion follows from the same arguments as for U_{n,j_n} using the bound on the density of the $X_{i/n}$. □

4.3.2. Proof of (ii)

To approximate Λ, we need an explicit form for the transition semigroup of the process X under P_-. We first recall an expansion of the transition density p_δ which may be found for instance in Dacunha-Castelle and Florens-Zmirou (1986).

If X is a solution of (1.2) with coefficients $b = 0$ and if σ is twice differentiable (for $\sigma^2 \in V_{jn}$, this is obtained by taking the regularity r of ψ greater than 2) then

$$p_\delta(x, y) = \frac{1}{\sigma(y)(2\pi\delta)^{1/2}} r_\delta(x, y) \exp\left(H(y) - H(x) - \frac{\{S(y) - S(x)\}^2}{2\delta} \right), \tag{4.31}$$

where $S(x) = \int_0^x \{1/\sigma(t)\}\,\mathrm{d}t$, $\mathrm{e}^{H(y)-H(x)} = \{\sigma(x)/\sigma(y)\}^{1/2}$ (in the case when $b = 0$ this term reduces to a very simple formula) and

$$r_\delta(x, y) = \mathrm{E}\left\{ \exp\left(\delta \int_0^1 c\{(1-u)S(x) + uS(y) + \delta^{1/2} B_u\}\,\mathrm{d}u \right) \right\},$$

where $(B_t, 0 \leqslant t \leqslant 1)$ is a standard Brownian bridge and c is a function which is bounded if $\gamma_n 2^{3/2 j_n}$ is bounded. In fact (Dacunha-Castelle and Florens-Zmirou 1986)

$$c = \tfrac{1}{4}\{\sigma'' - \tfrac{1}{2}(\sigma')^2\} \circ S^{-1}.$$

Therefore there exists C_6 depending on $\mathbf{M}$ and ψ such that

$$\mathrm{e}^{-C_6\delta} \leqslant r_\delta(x, y) \leqslant \mathrm{e}^{C_6\delta}. \tag{4.32}$$

We first need some technical results. Let us write $R^+(X^{(n)})$ for $\prod_{i=0}^{n-1} r^+_{1/n}(X_{(i+1)/n}, X_{i/n})$ where r^+_δ is the function of (4.31) and (4.32) associated with the diffusion coefficient σ_+ and define R^- analogously.

We define H^+, S^+ (or H^-, S^-) using the same convention. More generally, for any function f, we set

$$Df(x) = f^+(x) - f^-(x) \quad \text{and: } \Delta f(X_i) = f(X_{(i+1)/n}) - f(X_{i/n}).$$

We have the following.

Lemma 4.

$$0 < C_7 \leqslant \exp\left(\sum_{i=0}^{n-1} D\Delta H(X_i) \right) \leqslant C_8 < \infty, \tag{4.33}$$

$$\sum_{i=0}^{n-1} D(\Delta S(X_i))^2 = \sum_{i=0}^{n-1} \left(\frac{1}{\sigma_+^2(X_i)} - \frac{1}{\sigma_-^2(X_i)} \right)(\Delta X_i)^2 + R_n^{(6)}, \tag{4.34}$$

where C_7 and C_8 only depend on $\mathbf{M}$ and if $2^{5j_n/2} n^{1/2} \to 0$ as $n \to \infty$; then

$$n\mathrm{E}(|R_n^{(6)}|) \to 0.$$

Completion of proof of (ii). We have

$$\Lambda(\sigma_-, \sigma_+, X^{(n)}) = \frac{R^+}{R^-}(X^{(n)}) \prod_{i=0}^{n-1} \frac{\sigma_-}{\sigma_+}(X_i) \exp \sum_{i=0}^{n-1} \left(D\,\Delta H(X_i) - \frac{n}{2} D(\Delta S(X_i))^2 \right). \tag{4.35}$$

Therefore

$$P_-^\nu(\Lambda(\sigma_-, \sigma_+, X^{(n)}) > \mathrm{e}^{-\lambda}) = P_-^\nu\Bigg(\log\tilde{\Lambda}(\sigma_-, \sigma_+, X^{(n)}) > -\lambda - \left(\sum_{i=0}^{n-1} D\,\Delta H(X_i) + \frac{n}{2} R_n^{(6)} + \frac{R^+}{R^-}(X^{(n)}) \right)\Bigg).$$

We then apply Lemma 4 and the Chebyshev inequality and we derive

$$P_-^\nu(\Lambda(\sigma_-, \sigma_+, X^{(n)}) > \mathrm{e}^{-\lambda}) \geqslant P_-^\nu(\log\tilde{\Lambda}(\sigma_-, \sigma_+, X^{(n)}) > -(\lambda + \eta)) - \frac{C_9}{\eta}, \tag{4.36}$$

for any $\eta > 0$. The constant C_9 depends on C_6 and $\mathbf{M}$. Since the choice of λ is free and η can be chosen arbitrarily large, we apply Lemma 2(i) and we can conclude that there exists $p_0 > 0$ depending on λ, η such that

$$P_-^\nu(\Lambda(\sigma_-, \sigma_+, X^{(n)}) > \mathrm{e}^{-\lambda}) \geqslant p_0 > 0.$$

The proof of Lemma 2 is complete. □

4.3.3. Proof of Lemma 4

Equation (4.33) is a direct consequence of the definition of H. To prove (4.34), we write

$$\left(\int_x^y \frac{\mathrm{d}t}{\sigma(t)} \right)^2 = \frac{(y-x)^2}{\sigma^2(x)} + R^{(7)}(x, y). \tag{4.37}$$

We therefore need a bound for

$$R_n^{(6)} = \sum_{i=0}^{n-1} \{ R_+^{(7)}(X_{t_i}, X_{t_{i+1}}) - R_-^{(7)}(X_{t_i}, X_{t_{i+1}}) \}.$$

We first introduce some notation. R will denote a real function vanishing at the origin, possibly varying at each occurrence and for which there exists a constant (possibly depending on $\mathbf{M}$ and s, p) such that

$$|R(x)| \leqslant C|x|.$$

For $\sigma^2 \in V_{j_n}$, we denote by $g_{j_n k} = \sum_k \epsilon_k \psi_{j_n k}$ the added function to the initial condition. Thus

$$\sigma^2 = 1 + \gamma_n g_{j_n,k}.$$

Using a Taylor argument, it is easily seen that

$$\frac{1}{\sigma(t)} - \frac{1}{\sigma(x)} = -\tfrac{1}{2}\gamma_n\left((t-x)(g_{j_n,k})'(x) + \frac{(t-\xi)^2}{2}(g_{j_n,k})''(\xi)\right)$$
$$+ \tfrac{3}{8}\gamma_n^2(t-\xi)(g_{j_n,k})'(\xi)\{g_{j_n,k}(t) + g_{j_n,k}(x)\}$$
$$+ R(\gamma_n^3 g_{j_n,k}^3(t)) + R(\gamma_n^3 g_{j_n,k}^3(x)),$$

$$\frac{1}{\sigma(t)} + \frac{1}{\sigma(x)} = 2 - \frac{\gamma_n}{2}\{g_{j_n,k}(x) + g_{j_n,k}(t)\} + R(\gamma_n^2 g_{j_n,k}^2(t)) + R(\gamma_n^2 g_{j_n,k}^2(x)).$$

We use the following notation:

$$\frac{1}{\sigma(t)} - \frac{1}{\sigma(x)} = -\tfrac{1}{2}\gamma_n(t-x)(g_{j_n,k})'(x) + L(j_n, \gamma_n, t, x),$$
$$\frac{1}{\sigma(t)} + \frac{1}{\sigma(x)} = 2 + M(j_n, \gamma_n, t, x),$$

the definition of L and M being given by the foregoing expansion. Next, we have

$$R^{(7)}(x, y) = \left(\int_x^y \frac{\mathrm{d}t}{\sigma(t)}\right)^2 - \frac{(y-x)^2}{\sigma^2(x)}$$
$$= \int_x^y \left(\frac{1}{\sigma(t)} - \frac{1}{\sigma(x)}\right) \int_x^y \left(\frac{1}{\sigma(t)} + \frac{1}{\sigma(x)}\right)$$
$$= B_1(x, y) + B_2(x, y) + B_3(x, y) + B_4(x, y),$$

with

$$B_1(x, y) = -\gamma_n(g_{j_n,k})'(x)(y-x)^3,$$
$$B_2(x, y) = 2(y-x)\int_x^y L(j_n, \gamma_n, t, x)\,\mathrm{d}t,$$
$$B_3(x, y) = -\tfrac{1}{2}\gamma_n(g_{j_n,k})'(x)(y-x)^2\int_x^y M(j_n, \gamma_n, t, x)\,\mathrm{d}t,$$
$$B_4(x, y) = \int_x^y L(j_n, \gamma_n, t, x)\,\mathrm{d}t \int_x^y M(j_n, \gamma_n, t, x)\,\mathrm{d}t.$$

We successively prove that

$$nE\left(\left|\sum_{i=0}^{n-1} B_l(X_i, X_{i+1})\right|\right) \to 0, \quad l = 1, \ldots, 4.$$

under the condition $2^{5j_n/2}/n^{1/2} \to 0$.

4.3.3.1. Convergence of B_1. The process

$$M_i = n \sum_{l=0}^{i-1} B_1(X_l, X_{l+1}) = -n \sum_{l=0}^{i} (g_{j_n,k})'(X_l)(X_{l+1} - X_l)^3, \quad i = 0, \dots n,$$

is a $(\mathscr{F}_i^n)$ martingale. Hence

$$\begin{aligned} \mathrm{E}(M_n^2) &= n^2 \sum_{i=0}^{n-1} \mathrm{E}\{B_1^2(X_i, X_{i+1})\} \\ &= n^2 \gamma_n^2 \sum_{i=0}^{n-1} \mathrm{E}[\{(g_{j_n,k})'(X_i)\}^2 \mathrm{E}\{(X_{i+1} - X_i)^6 | \mathscr{F}_i^n\}] \\ &\leqslant C\gamma_n^2 \sum_{i=0}^{n-1} \mathrm{E}[\{(g_{j_n,k})'(X_i)\}^2] \frac{1}{n}, \end{aligned}$$

the last inequality coming from the Burckholder–Davis–Gundy inequality.

From $|(g_{j_n,k})'(X_i)| \leqslant C2^{3j_n/2}$, we deduce that

$$\mathrm{E}(M_n^2) \leqslant C\left(\frac{2^{3j_n/2}}{n^{1/2}}\right)^2,$$

which converges to 0 from the hypothesis. The conclusion follows from the Schwarz inequality.

4.3.2.2. Convergence of B_2. We have

$$|L(j_n, \gamma_n, t, x)| \leqslant C(\gamma_n(t-x)^2 2^{5j_n/2} + \gamma_n^2 |t-x| 2^{2j_n} + \gamma_n^3 2^{3j_n/2}).$$

Therefore

$$\begin{aligned} & n\left|\sum_{i=0}^{n-1} B_2(X_i, X_{i+1})\right| \\ & \leqslant C\left(n\gamma_n 2^{5j_n/2} \sum_i (X_{i+1} - X_i)^4 + n\gamma_n^2 2^{2j_n} \sum_i |X_{i+1} - X_i|^3 + n\gamma_n^3 2^{3j_n/2} \sum_i (X_{i+1} - X_i)^2\right). \end{aligned}$$

Taking the expectation and applying the Burckholder–Davis–Gundy inequality yields

$$n\mathrm{E}\left|\sum_{i=0}^{n-1} B_2(X_i, X_{i+1})\right| \leqslant C\left(\frac{2^{5j_n/2}}{n^{1/2}} + \frac{2^{2j_n}}{n^{1/2}} + \frac{2^{3j_n/2}}{n^{3/2}}\right).$$

The conclusion follows.

4.3.3.3. Convergence of B_3. Likewise, one readily checks that

$$|M(j_n, \gamma_n, t, x)| \leqslant C\gamma_n 2^{j_n/2}.$$

Hence

$$n\left|\sum_{i=0}^{n-1} B_2(X_i, X_{i+1})\right| \leq Cn\gamma_n^2 2^{3j_n/2} \sum_i |X_{i+1} - X_i|^3.$$

The same arguments as for B_2 lead to

$$n\mathrm{E}\left|\sum_{i=0}^{n-1} B_3(X_i, X_{i+1})\right| \leq C\frac{2^{2j_n}}{n^{1/2}}.$$

4.3.3.4. Convergence of B_4. This is straightforward from the results of B_2 and B_3. This ends the proof of Lemma 4. □

5. Upper bounds; proof of Proposition 2

We shall prove general approximation results from which we can deduce Proposition 2. We first state a result on the rate of convergence in L_p of the empirical local time, proved in Appendix 1.

Proposition 3. *Let ϕ be a compactly supported positive function of class $\mathscr{C}^3$ such that $\int \phi = 1$. For $h_n > 0$, define*

$$L_n^x = \frac{1}{nh_n}\sum_{i=0}^{n} \phi(h_n^{-1}(X_{i/n} - x))$$

as the ϕ-empirical local time of X at x. For $\gamma \in [2, +\infty[$ if $nh_n^2 \to \infty$ as $n \to \infty$, then there exists a constant $C_{10} = C_{10}(s, p, \mathbf{M}, \gamma)$ such that

$$\sup_{x \in D} \mathrm{E}_{\sigma,b}(|L_n^x - L^x|^\gamma) \leq C_{10}\left\{h_n^{\gamma/2} + \left(\frac{1}{nh_n^2}\right)^\gamma\right\}.$$

Remark. In our framework, we shall assume that $h_n \asymp n^{-s/(1+2s)}$; hence the condition $nh_n^2 \to +\infty$ is fulfilled. In general, optimizing the bound given in Proposition 3 leads to the rate $n^{-1/5}$. As a consequence, our rate of convergence is suboptimal (but sufficient for our purpose) compared with the classical nonparametric rate for a function of Besov regularity $\frac{1}{2}$ which is $n^{-1/4}$. In fact, the Brownian local paths belong almost surely to the space $B_{sp\infty}$, for $1 \leq p < \infty$ (Boufoussi and Roynette 1993).

The next result gives an upper bound for the minimax risk penalized by the empirical local time. We first need a definition.

Definition 3. *If x_λ is the midpoint of the box C_λ, define, for $\nu > 0$,*

$$g_{n,\nu}(x) = \sum_l \prod_{\lambda \in C_{j,l}} 1_{L_n^{x_\lambda} \geq \nu} 1_{[l2^{-j},(l+1)2^{-j}[}(x),$$

where $C_{j,l} = \{\lambda\colon C_\lambda \cap [l2^{-j}, (l+1)2^{-j}[\neq \emptyset\}$ *as the* ϕ*-empirical penalization function at level* ν. *(Recall that* L_n^x *depends on* ϕ *from construction.)*

In the following, we shall assume that $\phi(x) = (1/\int\tilde{\phi})\tilde{\phi}(x)$, where

$$\tilde{\phi}(x) \leqslant 1_{[-1/2,1/2]}(x),$$

$$\int\tilde{\phi} > \tfrac{1}{2}$$

This will enable us to have a control with the empirical local time on the number of observation points lying in each C_λ.

Proposition 4. *Set* $\overline{\nu} = \nu/2\int\tilde{\phi}$. *Suppose that Assumptions 1 and 2 hold. Let* $\hat{\sigma}_n^2$ *be the estimator of Definition 2, constructed at level* $\nu/2$. *If* $2^{j_n} \asymp n^{1/(1+2s)}$ *and* $h_n \asymp n^{-s/(1+2s)}$ *then*

$$\sup_{(\sigma^2,b)\in V\times\mathscr{H}} \mathrm{E}_{\sigma,b}\left(\int_D |\sigma^2(x) - \hat{\sigma}_n^2(x)|^p g_{n,\overline{\nu}}(x)\,\mathrm{d}x\right) \leqslant C_{11} n^{-sp/(1+2s)}$$

for some constant $C_{11} = C_{11}(\varphi, s, p_1, \mathbf{M}, \nu)$.

5.1. Proof of Proposition 2

As for the lower bounds, we shall use the notation C for a generic constant. Clearly

$$\mathrm{E}_{\sigma,b}\left(\int_D |\sigma^2(x) - \hat{\sigma}_n^2(x)|^p\,\mathrm{d}x\, 1_{L^D\geqslant\nu}\right) \leqslant \mathrm{E}_{\sigma,b}\left(\int_D |\sigma^2(x) - \hat{\sigma}_n^2(x)|^p g_{n,\overline{\nu}}(x)\,\mathrm{d}x\right)$$

$$+ \mathrm{E}_{\sigma,b}\left(\int_D |\sigma^2(x) - \hat{\sigma}_n^2(x)|^p \{1 - g_{n,\overline{\nu}}(x)\}\, dx 1_{L^D\geqslant\nu}\right).$$

The first term on the right-hand side of the last inequality is of order $n^{-sp/(1+2s)}$ because of Proposition 4. For the second term, we use the fact that it is unlikely that L_n^x is small and L^x large simultaneously. More precisely, it is easily seen that

$$\limsup_{n\to\infty} \sup_{(\sigma^2,b)\in V_{sp}(\mathbf{M})\times\mathscr{H}} \mathrm{E}_{\sigma,b}\left(\int_D |\sigma^2(x) - \hat{\sigma}_n^2(x)|^{2p}\,\mathrm{d}x\right) < \infty.$$

Hence, by the Schwarz inequality and using the fact that $g_{n,\nu}$ is a step function, it is enough to prove

$$\left\{\mathrm{E}_{\sigma,b}\left(\int_D (1 - g_{n,\overline{\nu}}(x))1_{L^D\geqslant\nu}\,\mathrm{d}x\right)\right\}^{1/2} \leqslant Cn^{-sp/(1+2s)}. \tag{5.1}$$

We write

$$(1-g_{n,\overline{\nu}}(x))1_{L^D\geqslant\nu}=\sum_l\left(1-\prod_{\lambda\in C_{j_n,l}}1_{L_n^{x_\lambda}\geqslant\overline{\nu}}\right)1_{L^D\geqslant\nu}1_{[l2^{-j_n},(l+1)2^{-j}[}(x)$$

$$\leqslant\sum_l\left(\sum_\lambda 1_{L_n^{x_\lambda}\leqslant\overline{\nu}}1_{L^D\geqslant\nu}\right)1_{[l2^{-j},(l+1)2^{-j}[}(x).$$

Consequently, because $\overline{\nu}=1/(2\int\tilde{\phi})$, we derive

$$\left\{\mathrm{E}_{\sigma,b}\left(\int_D(1-g_{n,\overline{\nu})(x)}1_{L^D\geqslant\nu}\,\mathrm{d}x\right)\right\}^{1/2}\leqslant C\sup_{x\in D}P_{\sigma,b}\left(|L_n^x-L^x|\geqslant\nu\left(1-\frac{1}{2\int\tilde{\phi}}\right)\right)^{1/2}.\quad(5.2)$$

The choice of $\tilde{\phi}$ ensures that $1-1/(2\int\tilde{\phi})>0$.

Now, using the Chebyshev inequality and applying Proposition 3, it suffices to pick a $\gamma>0$ large enough in Proposition 3 so that (5.1) holds. The proof of Proposition 2 is complete. □

5.2. Proof of Proposition 4

We denote by $\|.\|_p$ the L_p norm on the interval $D=[0,1]$. We write $\mathrm{E}_{\sigma,b}\|(\hat{\sigma}_n^2-\sigma^2)g_{n,\overline{\nu}}\|_p^p$ as a sum of a stochastic term and an approximating term linked to the wavelets method of projection. More precisely

$$\mathrm{E}_{\sigma,b}\{\|(\hat{\sigma}_n^2-\sigma^2)g_{n,\overline{\nu}}\|_p^p\}\leqslant 2^{p-1}(S_n+A_n),$$

with

$$S_n=\mathrm{E}_{\sigma,b}\{\|(\hat{\sigma}_n^2-P_{j_n}^{[0,1]}\sigma^2)g_{n,\overline{\nu}}\|_p^p\}$$

and

$$A_n=\|\sigma^2-P_{j_n}^{[0,1]}\sigma^2\|_p^p,$$

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

Let us first study A_n. Resulting from the approximation of Besov spaces by wavelets sequences, for any $f\in B_{sp\infty}([0,1];M_3)$, the following inequality holds:

$$\|f-P_{j_n}^{[0,1]}f\|_{L_p}\leqslant 2^{-j_n(s\wedge r)}\epsilon_{j_n}\quad(5.3)$$

where ϵ_{j_n} is bounded by a constant depending only on M_3. The constant r in (5.3) is the regularity of φ. Assuming $r>s$ and $\sigma^2\in V_{sp}(\mathbf{M})$, we get

$$A_n\leqslant C2^{-j_nsp}.\quad(5.4)$$

We now consider the stochastic term. We use the localization property of φ and $g_{n,\nu}$ (note that $g_{n,\nu}$ is a step function expanded in the Haar basis at the same resolution level as φ_{j_nk}). We apply the lemma of Meyer (1990, p. 30) to obtain

$$S_n \leqslant C 2^{j_n(p/2-1)}\left(\sum_{k=0}^{N_0-1} \mathrm{E}|(\hat{\alpha}^l_{j_n k} - \alpha^l_{j_n k})c_{j_n,k}|^p + \sum_{k\in s_j} \mathrm{E}|(\hat{\alpha}_{j_n k - \alpha jk})c_{j_n,k}|^p \right.$$
$$\left. + \sum_{k=0}^{N_0-1} \mathrm{E}|(\hat{\alpha}^r_{j_n k} - \alpha^r_{j_n k})c_{j_{n,k}}|^p\right). \tag{5.5}$$

$c_{j_n,k} = \prod_{\lambda \in C_{j_n,k}} 1_{L_n^{x_\lambda} \geqslant \nu}$ are the coefficients of the penalization function. From now on, we shall no longer distinguish in our notation the edge and interior components involved in expansion (5.5). Recall that the number of coefficients is eactly 2^{j_n}.

We next show that the drift b can be regarded as a nuisance term which does not interefere in the rate of convergence of our estimator.

Lemma 5. *There exists $q \in [1, \infty[$ and a constant K_5 depending only on p, M_1 and M_2 such that*

$$\mathrm{E}_{\sigma,b}(|\hat{\alpha}_{j_n k} - \alpha_{j_n k}|^p) \leqslant K_2 \mathrm{E}_{\sigma,0}(|\hat{\alpha}_{j_n k} - \alpha_{j_n k}|^{qp})^{1/q}.$$

Proof. From the Girsanov theorem, the two measures $P_{\sigma,b}$ and $P_{\sigma,0}$ are equivalent on $\mathscr{F}_1 = \sigma(X_s, 0 \leqslant s \leqslant 1)$, with density

$$\mathscr{D} = \frac{\mathrm{d}P_{\sigma,b}}{\mathrm{d}P_{\sigma,0}} = \exp\left(\int_0^1 \frac{b(s, X_s)}{\sigma^2(X_s)}\,\mathrm{d}X_s - \frac{1}{2}\int_0^1 \frac{b^2(s, X_s)}{\sigma^2(X_s)}\,\mathrm{d}s\right).$$

Using the Hölder inequality, we have

$$\mathrm{E}_{\sigma,b}(|\hat{\alpha}_{j_n k} - \alpha_{j_n k}|^p) \leqslant \{\mathrm{E}_{\sigma,0}(\mathscr{D}^{q'})\}^{1/q'}\{\mathrm{E}_{\sigma,0}(|\hat{\alpha}_{j_n k} - \alpha_{j_n k}|^{qp})\}^{1/q}$$

with $1/q + 1/q' = 1$. It suffices then to show that

$$\sup_{(\sigma^2,b)\in V\times\mathscr{H}} \mathrm{E}_{\sigma,0}(\mathscr{D}^{q'}) < \infty$$

for some $q' \in]1, +\infty[$. The conclusion follows from the uniform linear growth hypothesis and the fact that there exists a $\tau > 0$ such that $\sup_{0\leqslant t\leqslant 1} \sup(\sigma^2, b) \in V \times \mathscr{H} E_{\sigma,b}(\mathrm{e}^{\tau X_t^2}) < \infty$.

With a modification of the constants, it is enough to concentrate on $P_{\sigma,0}$. We shall now work under $P_{\sigma,0}$.

Recall that the wavelet estimator is constructed at level $\nu/2$. Coming back to Definition 2, we write

$$\hat{\alpha}_{j_n k} - \alpha_{j_n k} = Q_1 + Q_2 + Q_3,$$

with

$$Q_1 = \frac{1}{\lfloor n\nu/2 \rfloor} \sum_{i=1}^{\lfloor n\nu/2 \rfloor} \sigma^2(X_{T_i})\varphi_{j_n k}(x_{T_i}) - \int_{S_{j_n,k}} \sigma^2(x)\varphi_{j_n k}(x)\,\mathrm{d}x,$$

$$Q_2 = \frac{1}{\lfloor n\nu/2 \rfloor} \sum_{i=1}^{\lfloor n\nu/2 \rfloor} \frac{1}{\Delta_n} \int_{T_i}^{T_i+\Delta_n} \{\sigma^2(X_{T_i}) - \sigma^2(X_s)\} \varphi_{j_n k}(x_{T_i}) \, \mathrm{d}s,$$

$$Q_3 = \frac{1}{\lfloor n\nu/2 \rfloor} \sum_{i=1}^{\lfloor n\nu/2 \rfloor} \epsilon_{T_i} \varphi_{j_n k}(x_{T_i}),$$

where $S_{j_n,k}$ denotes the support of $\varphi_{j_n k}$. We then have

$$\mathrm{E}_{\sigma,0}|(\hat{\alpha}_{j_n k} - \alpha_{j_n k})c_{j_n,k}|^p \leqslant 3^{p-1}\{\mathrm{E}_{\sigma,0}(|Q_1|^p c_{j_n,k}) + \mathrm{E}_{\sigma,0}|Q_2|^p + \mathrm{E}_{\sigma,0}|Q_3|^p\}. \tag{5.6}$$

Considering Q_1, we write

$$Q_1 = Q_{1,1} + Q_{1,2}, \tag{5.7}$$

with

$$Q_{1,1} = \frac{1}{\lfloor n\nu/2 \rfloor} \sum_{(\lambda: C_\lambda \subset S_{j_n,k})} \sum_{(i: i/(n\nu) \in C_\lambda)} \{\sigma^2(X_{T_i}) - \sigma^2(x_{T_i})\} \varphi_{j_n k}(x_{T_i}),$$

$$Q_{1,2} = \frac{1}{\lfloor n\nu/2 \rfloor} \sum_{i=1}^{\lfloor n\nu/2 \rfloor} \sigma^2(x_{T_i}) \varphi_{j_n k}(x_{T_i}) - \int_{S_{j_n,k}} \sigma^2 \varphi_{j_n k}.$$

We are ready to use the ϕ-penalization $c_{j_n,k}$. Let us assume that $\phi = (1/\int \tilde{\phi})\tilde{\phi}$, with $\tilde{\phi}(x) \leqslant 1_{[-1/2,1/2]}$. Since $\overline{\nu} = \nu/(2\int \tilde{\phi})$ the following inclusion holds for any $x \in D$:

$$(L_n^x \geqslant \overline{\nu}) = \left(\frac{1}{nh_n} \sum_{i=0}^{n} \tilde{\phi}(h_n^{-1}(X_{t_i} - x)) \geqslant \frac{\nu}{2} \right)$$

$$\subseteq \left(\sum_{i=0}^{n} 1_{|X_{t_i} - x| \leqslant h_n/2} \geqslant \lfloor nh_n \nu/2 \rfloor \right).$$

In other words, multiplying by the factor $c_{j_n,k}$ means that we retain the events $N^\lambda_{T_{\lfloor n\nu/2 \rfloor}} \geqslant \lfloor nh_n\nu/2 \rfloor$ (recall that $N^\lambda_{T_{\lfloor n\nu/2 \rfloor}}$ is the *counter* of the box C_λ), i.e. the T_i are all distinct so that we have sufficient points of observation to proceed to the approximation of the wavelet coefficient. Indeed, using the Jensen inequality

$$\mathrm{E}_{\sigma,0}(|Q_{1,1}|^p c_{j_n,k}) \leqslant C \frac{2^{-j_n(p-1)}}{n} \sum_{\lambda,i} \mathrm{E}_{\sigma,0}(|\sigma^2(X_{T_i}) - \sigma^2(x_{T_i})|^p |\varphi_{j_n k}(x_{T_i})|^p 1_{N^\lambda_{T_{\lfloor n\nu/2 \rfloor}} \geqslant nh_n\nu/2}). \tag{5.8}$$

For the event $(N^\lambda_{T_{\lfloor n\nu \rfloor}} \geqslant nh_n\nu/2)$, since σ^2 has a bounded derivative, the following inequality holds:

$$|\sigma^2(X_{T_i}) - \sigma^2(x_{T_i})| \leqslant C|X_{T_i} - x_{T_i}| \leqslant Ch_n, \tag{5.9}$$

the last inequality coming from the construction of the T_i and the x_{T_i}. Finally, from (5.8), (5.9) and using the fact that the sum in (λ, i) is of order $n2^{-j_n}$, we derive

$$E_{\sigma,0}(|Q_{1,1}|^p c_{j_n,k}) \leqslant C 2^{-j_n p/2} h_n^p. \tag{5.10}$$

Likewise

$$E_{\sigma,0}(|Q_{1,2}|^p c_{j_n,k}) \leqslant C\left(\frac{2^{j_n/2}}{n}\right)^p. \tag{5.11}$$

Now, putting together (5.10) and (5.11), we deduce that

$$E_{\sigma,0}(|Q_1|^p c_{j_n,k}) \leqslant C\left\{2^{-j_n p/2} h_n^p + \left(\frac{2^{j_n/2}}{n}\right)^p\right\}. \tag{5.12}$$

We turn to Q_2. Using the Jensen inequality and the same argument on the indices as for (5.10), one has

$$E_{\sigma,0}(|Q_2|^p) \leqslant \frac{2^{-j_n(p/2-1)}}{n} \sum_i E\left(\Delta_n^{-1} \int_{T_i}^{T_i+\Delta_n} |\sigma^2(X_s) - \sigma^2(X_{T_i})|^p \, ds\right). \tag{5.13}$$

From the Burckholder–Davis–Gundy inequality and the regularity of σ^2

$$E_{\sigma,0}\left(\Delta_n^{-1} \int_{T_i}^{T_i+\Delta_n} |\sigma^2(X_s) - \sigma^2(X_{T_i})|^p \, ds\right) \leqslant C\Delta_n^{p/2}.$$

Hence

$$E_{\sigma,0}(|Q_2|^p) \leqslant C 2^{-j_n p/2} n^{-p/2}. \tag{5.14}$$

We need a bound on Q_3 to complete our study. We first recall a martingale version of the Rosenthal inequality, which may be found in the book by Hall and Heyde (1980, p. 23).

Lemma 6 (The Rosenthal inequality for martingales). *Let* $S_i = \sum_{l=0}^{i} \chi_l$ *be a* $(\mathscr{F}_i)$ *martingale, for* $1 \leqslant i \leqslant n$. *For every* $p \in [1, \infty[$, *there exists a constant* C_p *depending only on* p *such that*

$$E(|S_n|^p) \leqslant C_p\left[E\left\{\left(\sum_{i=1}^{n} E(\chi_i^2 | \mathscr{F}_{i-1})\right)^{p/2}\right\} + \sum_{i=1}^{n} E(|\chi_i|^p)\right].$$

Going back to our study, we first remark that the random variables $\epsilon_{T_i}\varphi_{j_n k}(x_{T_i})$ are $(\mathscr{F}^n_{T_{i+1}})$-martingale increments. Hence, applying the Rosenthal inequality leads to

$$E_{\sigma,0}(|Q_3|^p) \leqslant \frac{C}{n^p}\left\{E\left(\left|\sum_{i=1}^{\lfloor n\nu/2\rfloor} E_{\sigma,0}\{\epsilon_{T_i}^2 \varphi_{j_n k}^2(x_{T_i}) | \mathscr{F}^n_{T_i}\}\right|^{p/2}\right) + \sum_{i=1}^{\lfloor n\nu/2\rfloor} E_{\sigma,0}(|\epsilon_{T_i}\varphi_{j_n k}(x_{T_i})|^p)\right\}.$$

With the same kind of arguments as for Q_2 we get

$$E_{\sigma,0}(|\epsilon_{T_i}\varphi_{j_n k}(x_{T_i})|^p | \mathscr{F}^n_{T_i}) \leqslant C 2^{j_n(p/2-1)}.$$

Hence

$$\mathrm{E}_{\sigma,0}(|Q_3|^p) \leqslant C(n^{-p/2} + n^{-(p-1)}2^{j_n(p/2-1)}). \tag{5.15}$$

Putting together (5.14), (5.12) and (5.15) in (5.6) and using that $2^{3j_n/2}/n^{1/2} \to 0$ since $2^{j_n} \asymp n^{1/(1+2s)}$ and $s > 1$ we finally obtain

$$S_n \leqslant C\left\{\left(\frac{2^{j_n}}{n}\right)^{p/2} + h_n^p\right\}. \tag{5.16}$$

The optimal rate is obtained when A_n and S_n are of the same order of magnitude, which leads to $2^{j_n} \simeq n^{1/(1+2s)}$. From (5.4) and (5.16), as h_n is chosen to be of order $n^{-s/(1+2s)}$ the conclusion follows. The proof of proposition 4 is complete. □

6. Further remarks

6.1. Note on a previous construction

Florens-Zmirou (1993) proposed an estimator of σ^2, namely

$$\tilde{\sigma}_n^2(x) = n\sum_{i=0}^{n} 1_{|X_{i/n}-x|\leqslant r_n}(X_{(i+1)/n} - X_{i/n})^2 \Big/ \sum_{i=0}^{n} 1_{|X_{i/n}-x|\leqslant r_n}. \tag{6.1}$$

This estimator is consistent and asymptotically normal. However, regarding minimax properties, this procedure is unlikely to be optimal. We propose the following heuristic explanation. Using the regression approximation approach described by (1.3) and (1.4), we can write (6.1) as

$$\tilde{\sigma}_n^2(x) = \sum_{i=1}^{n} K_{r_n}(X_{i/n} - x)Y_{i/n} \Big/ \sum_{i=1}^{n} K_{r_n}(X_{i/n} - x),$$

where K is the Haar kernel ($K_h(x) = h^{-1}K(h^{-1}x)$ and $K(x) = 1_{[-1/2,1/2]}(x)$). The factor r_n is a smoothing parameter which should be compared in our framework with 2^{-j_n}. The data $(X_{i/n}, Y_{i/n}, i = 0, \ldots, n-1)$ are obtained from the *regression approximation*

$$Y_{i/n} \simeq \sigma^2(X_{i/n}) + \epsilon_{i/n}, \quad i = 0, \ldots, n-1, \tag{6.2}$$

with $Y_{i/n} = n(X_{(i+1)/n} - X_{i/n})^2$. Thus $\tilde{\sigma}_n^2$ is the Nadaraya–Watson estimator of the regression model (6.2) with random design (Nadaraya 1964).

The properties of the Nadaraya–Watson estimator have been extensively discussed in the literature (see, for example, Chu and Marron (1991) and Wand and Jones (1995)). The minimax efficiency of the Nadaraya–Watson estimator is linked to the smoothness of the density of the design, say f. Poor smoothness of f leads to poor minimax results, whatever the kernel is chosen (Fan 1992; Wand and Jones 1995).

In our framework, the analogous of the density $x \to f(x)$ is the local time $x \to L^x$ of the process X up to time 1. Consequently, the low smoothness parameter of the local time ($x \to L^x$ does not belong almost surely to the Besov space $B_{sp\infty}$ for $s > \frac{1}{2}$, $1 \leqslant p < \infty$ (Boufoussi and Roynette 1993)) is likely to imply suboptimality for $\tilde{\sigma}^2_n$.

Indeed, we can emphasize this point noting that

$$\tilde{\sigma}^2_n(x) = \frac{\mathscr{L}^x_n}{L^x_n}, \tag{6.3}$$

where

$$\mathscr{L}^x_n = \frac{1}{h_n} \sum_{i=0}^{n} 1_{|X_{i/n} - x| \leqslant h_n} (X_{(i+1)/n} - X_{i/n})^2$$

is the empirical local time at its usual scale (see the remark in Appendix 1). In other words, $\tilde{\sigma}^2_n$ is obtained by estimating $\sigma^2(x) L^x$ and then dividing by an estimate of L^x.

6.2. A modification of the estimating procedure

Looking at the integrated risk, it appears that, the smaller the chosen ν, the better we can take into account different behaviours of the sample path. However, our procedure appears as "too cautious"; it operates well in the worst case when only a few observations are available. What about the regions of the domain D where more observation points lie? One may think of refining the method using the $\lfloor (1-\nu) n \rfloor$ left $X_{i/n}$ in order to improve locally the property of $\hat{\sigma}^2_n$, say for practical purpose (i.e. on constants since the rate is already optimal). We propose the following ongoing construction.

Following Section 3.2, we define the subsamplings

$$(T^2_1, \ldots, T^2_{\lfloor n\nu \rfloor}), \ldots, (T^l_1, \ldots, T^l_{\lfloor n\nu \rfloor}), \ldots$$

as the successive passage times in the boxes C_λ until saturation, which are all equal to 1 for i large enough. If $\overline{N}_\lambda$ is the number of layers saturated by the $X_{i/n}$ for the box C_λ, we write $k^*_j = \inf_{\lambda : C_\lambda \subset S_{j,k}} \overline{N}_\lambda$. One has $k^*_j \geqslant 1$ if $L_n \geqslant \nu$. The modified algorithm is then

$$\overline{\hat{\alpha}_{j,k}} = \frac{1}{\lfloor n\nu k^*_j \rfloor} \sum_{l=1}^{k^*_j} \sum_{i=1}^{\lfloor n\nu \rfloor} Y_{T^l_i} \varphi_{j,k}(x_{T^l_i}).$$

This estimator encompasses the same minimax properties as that from Proposition 2 and should become a better choice in practice. Unfortunately, there are still discarded observation points in this second estimate, because of the random character of k^*_j.

Indeed, we can construct a third estimator, still coming from that presented in Section 3 using now all the observation points. We work in the following way: we fit the regions with poor observation (i.e. low local time) with pseudo-data coming from the observed $X_{i/n}$. This procedure still depends on ν but discards no observation point. We intend to describe this method in a more practically oriented forthcoming work.

Appendix 1

A.1. Proof of Proposition 3

Remark. For the occupation times formula on the Lebesgue measure scale, we have, for every positive Borel function f,

$$\int_0^1 f(X_s)\,\mathrm{d}s = \int_{-\infty}^{+\infty} f(x)\mathscr{L}^x \frac{\mathrm{d}x}{\sigma^2(x)}, \tag{A.1}$$

where

$$\mathscr{L}^x = \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^1 1_{|X_s - x|\leqslant\epsilon}\,\mathrm{d}\langle X\rangle_s.$$

$\mathscr{L}^x$ is the usual local time of a continuous semimartingale (defined through the Tanaka formula), involving its quadratic variation.

For simplicity, we shall prove Proposition 3 under $P_{\sigma,0}$. The general case is obtained by a change in probability, exactly as in Lemma 5. The only requirement for the general case is a modification of the constants exhibited in the $P_{\sigma,0}$ case.

We use the following decomposition:

$$L_n^x - L^x = A_n + B_n,$$

where

$$A_n = \frac{1}{h_n}\int_0^1 \phi\{h_n^{-1}(X_s - x)\}\,\mathrm{d}s - L^x \tag{A.2}$$

and

$$B_n = L_n^x - \frac{1}{h_n}\int_0^1 \phi\{h_n^{-1}(X_s - x)\}\,\mathrm{d}s. \tag{A.3}$$

It is sufficient then to study the convergence of A_n and B_n respectively.

A.1.1. Convergence of A_n

By the occupation times formula

$$A_n = \frac{1}{h_n}\int_0^1 \left(\frac{1}{\sigma^2(y)}\mathscr{L}^y - \frac{1}{\sigma^2(x)}\mathscr{L}^x\right)\phi\{h_n^{-1}(y - x)\}\,\mathrm{d}y.$$

From the assumption that σ^2 is bounded below by M_0, since σ^2 is Lipschitz continuous we deduce that

$$|A_n| \leqslant \frac{1}{h_n}M_0^{-1}\int_0^1 |\mathscr{L}^y - \mathscr{L}^x|\phi\{h_n^{-1}(y - x)\}\,\mathrm{d}y + 2Ch_nM_0^{-2}\mathscr{L}^*, \tag{A.4}$$

where

$$\mathscr{L}^* = \sup_{x\in\mathbb{R}} \mathscr{L}^x$$

denotes the supremum of the local time. We shall need the Hölder property of the local time paths of a continuous martingale. This has been given by Revuz and Yor (1994, p. 227). In fact, under Assumption 1 and 2, for $\gamma \geqslant 2$

$$\mathrm{E}_{\sigma,0}(|\mathscr{L}^y - \mathscr{L}^x|^\gamma) \leqslant C|y-x|^{\gamma/2}. \tag{A.5}$$

The constant C depends only on γ, s, M_0 and M_1. Hence, by the Jensen inequality, from (A.4) and (A.5) we obtain, for $\gamma \geqslant 2$,

$$\mathrm{E}_{\sigma,0}(|A_n|^\gamma) \leqslant C h_n^{\gamma/2}. \tag{A.6}$$

Remark. We have implicitly used the property that the supremum of the local time is in L_γ, for all $\gamma > 0$ (Revuz and Yor 1994).

A.1.2. Convergence of B_n

Using a second-order Taylor expansion, we have

$$B_n = B_{n,1} + B_{n,2} + B_{n,3},$$

with

$$B_{n,1} = \frac{1}{h_n^2}\sum_{i=0}^n \phi'\{h_n^{-1}(X_{i/n} - x)\}\int_{i/n}^{(i+1)/n}(X_s - X_{t_i})\,\mathrm{d}s,$$

$$B_{n,2} = \frac{1}{h_n^3}\sum_{i=0}^n \phi''\{h_n^{-1}(X_{i/n} - x)\}\int_{i/n}^{(i+1)/n}\frac{(X_s - X_{i/n})^2}{2}\,\mathrm{d}s,$$

$$B_{n,3} = \frac{1}{h_n^4}\sum_{i=0}^n \int_{i/n}^{(i+1)/n}\phi'''\{h_n^{-1}(\xi_{s,i/n} - x)\}\frac{(X_s - X_{i/n})^3}{6}\,\mathrm{d}s.$$

The term $B_{n,2}$ will give the order of magnitude. We focus on $B_{n,1}$. Set

$$Z_i = \phi'\{h_n^{-1}(X_{i/n} - x)\}\int_{i/n}^{(i+1)/n}(X_s - X_{i/n})\,\mathrm{d}s.$$

The discrete time process $(\sum_{i=0}^k Z_i,\ 1 \leqslant k \leqslant n)$ is a $(\mathscr{F}^n_{k+1})$-martingale. Hence we may apply the Rosenthal inequality to get

$$\mathrm{E}_{\sigma,0}\left(\left|\sum_{i=1}^n Z_i\right|^\gamma\right) \leqslant C_\gamma\left[\mathrm{E}_{\sigma,0}\left\{\left(\sum_{i=1}^n \mathrm{E}_{\sigma,0}(Z_i^2|\mathscr{F}^n_i)\right)^{\gamma/2}\right\} + \sum_{i=0}^n \mathrm{E}_{\sigma,0}|Z_i|^\gamma\right].$$

Using successively the Jensen inequality, the Fubini theorem and the Burckholder–Davis–Gundy inequality, we obtain

$$\mathrm{E}_{\sigma,0}(Z_i^2|\mathscr{F}_i^n) \leqslant (\phi')^2\{h_n^{-1}(X_{i/n}-x)\}M_1^2\Delta_n^3.$$

Hence

$$\mathrm{E}_{\sigma,0}\left\{\left(\sum_{i=1}^n \mathrm{E}_{\sigma,0}(Z_i^2|\mathscr{F}_i^n)\right)^{\gamma/2}\right\} \leqslant M_1^\gamma \Delta_n^\gamma \mathrm{E}_{\sigma,0}\left(\left|\sum_{i=0}^n (\phi')^2(h_n^{-1}(X_{i/n}-x))\Delta_n\right|^{\gamma/2}\right).$$

Using a Riemann approximation argument and the occupation times formula, because ϕ is compactly supported, we derive

$$\mathrm{E}_{\sigma,0}\left(\left|\sum_{i=0}^n (\phi')^2\{h_n^{-1}(X_{i/n}-x)\}\Delta_n\right|^{\gamma/2}\right) \leqslant C h_n^{\gamma/2}.$$

Likewise

$$h_n^{-2\gamma}\sum_{i=0}^n \mathrm{E}_{\sigma,0}|Z_i|^\gamma \leqslant C\Delta_n^{3/2\gamma-1}h_n^{-2\gamma+1}.$$

In conclusion, as $\gamma \geqslant 2$ and $nh_n \to +\infty$,

$$\mathrm{E}_{\sigma,0}(|B_{n,1}|^\gamma) \leqslant C\Delta_n^\gamma h_n^{-3/2}.$$

We now turn to $B_{n,2}$. From the Itô formula, one has under $P_{\sigma,0}$

$$(X_s - X_{i/n})^2 = \int_{i/n}^s (X_u - X_{i/n})\sigma(X_u)\,\mathrm{d}W_u + \int_{i/n}^s \sigma^2(X_u)\,\mathrm{d}u.$$

Hence

$$B_{n,1} = h_n^{-3}\sum_{i=0}^n (T_{i,1}+T_{i,2}),$$

with

$$T_{i,1} = \tfrac{1}{2}\phi''\{h_n^{-1}(X_{i/n}-x)\}\int_{i/n}^{(i+1)/n}\mathrm{d}s\int_{i/n}^s (X_u - X_{i/n})\sigma(X_u)\,\mathrm{d}W_u,$$

$$T_{i,2} = \tfrac{1}{2}\phi''\{h_n^{-1}(X_{i/n}-x)\}\int_{i/n}^{(i+1)/n}\left(\frac{i+1}{n}-s\right)\sigma^2(X_s)\,\mathrm{d}s.$$

For $T_{i,1}$, we apply the same martingale technique as for $B_{n,1}$ and we obtain

$$h_n^{-3\gamma}\mathrm{E}_{\sigma,0}\left(\left|\sum_{i=0}^n T_{i,1}\right|^\gamma\right) \leqslant C\Delta_n^{2\gamma}h_n^{-5/2\gamma}.$$

For $T_{i,2}$, we first remark that $|T_{i,2}| \leqslant M_3|\phi''\{h_n^{-1}(X_{i/n}-x)\}|\Delta_n^2$. Hence

$$h_n^{-3\gamma}\mathrm{E}_{\sigma,0}\left(\left|\sum_{i=0}^{n} T_{i,2}\right|^{\gamma}\right) \leqslant Ch_n^{-3\gamma}\Delta_n^{\gamma}\mathrm{E}_{\sigma,0}\left(\left|\sum_{i=0}^{n}|\phi''\{h_n^{-1}(X_{i/n}-x)\}|\Delta_n\right|^{\gamma}\right) \leqslant Ch_n^{-2\gamma}\Delta_n^{\gamma}.$$

Finally

$$\mathrm{E}_{\sigma,0}(|B_{n,i}|^{\gamma}) \leqslant Ch_n^{-2\gamma}\Delta_n^{\gamma}.$$

The bound for the third terms follows likewise using the same technique; so we omit it. In the same way

$$\mathrm{E}_{\sigma,0}(|B_{n,3}|^{\gamma}) = o(h_n^{-2\gamma}\Delta_n^{\gamma}).$$

The proof of Lemma 3 is complete. □

A.2. Wavelets and Besov spaces

We recall some well-known results from approximation theory. Some references are Bergh and Löfström (1976), Peetre (1976) and Meyer (1990).

For $f \in L_p(\mathbb{R})$, define $\omega_p(t) = \sup_{|h|\leqslant t}\|\tau_h f - f\|_p$, where $\tau_h f(x) = f(x) - f(x-h)$. Then

$$f \in B_{sp\infty}(\mathbb{R}) \Leftrightarrow f \in L_p(\mathbb{R}) \quad \text{and} \quad \frac{\omega_p(t)}{t^s} \in L_\infty(\mathbb{R}^+).$$

For $s = 1$, the same definition remains valid if we change $\tau_h f - f$ by $\tau_h f + \tau_{-h} f - 2f$. For $s = N + \alpha$, with $N \in \mathbf{N}$ and $0 < \alpha \leqslant 1$, $f \in B_{sp\infty}(\mathbb{R}) \Leftrightarrow f \in L_p(\mathbb{R})$ and $f^{(N)} \in B_{\alpha p\infty}$, where $f^{(N)}$ is an Nth weak derivative of f.

We now give the definition of Besov spaces in terms of wavelet coefficients. Further data may be found in Meyer (1990). We recall that one can construct a function φ such that the following are true.

(1) The sequence $\{\varphi(x-k),\ k \in \mathbf{Z}\}$ is an orthonormal family of $L_2(\mathbb{R})$. Let V_0 be the subspace spanned by this sequence.

(2) If $\varphi_{jk} = 2^{j/2}\varphi(2^j x - k)$, let V_j denote the subspace spanned by $\{\varphi_{jk},\ k \in \mathbf{Z}\}$. Then $\forall j \in \mathbf{Z}$: $V_j \subset V_{j+1}$. Consequently $\bigcap_{j\in\mathbf{Z}} V_j = \{0\}$. Furthermore $\bigcup_{j\in\mathbf{Z}} V_j$ is dense in L_2. The function φ is called the scaling function of the multiresolution analysis $(V_j,\ j \in \mathbf{Z})$. In addition, we may assume the following regularity condition.

(3) φ is of class $\mathscr{C}^r$, φ and every derivative up to order r has fast decay. In this case the multiresolution analysis is said to be r *regular*.

Under these conditions, define the space W_j by

$$V_{j+1} = V_j \oplus W_j.$$

Then, there exists a function ψ (called *the wavelet*) such that

(1) $\{\psi(x-k),\ k \in \mathbf{Z}\}$ is an orthonormal basis of W_0,

(2) $\{\psi_{j,k},\ k \in \mathbf{Z},\ j \in \mathbf{Z}\}$ is an orthonormal basis of $L_2(\mathbb{R})$, where $\psi_{jk} = 2^{j/2}\psi(2^j x - k)$, and

(3) ψ has the same regularity property as φ.

In addition, we have the following decomposition for any integer j_0:

$$L_2(\mathbb{R}) = V_{j_0} \oplus \oplus_{j \geqslant j_0} W_j.$$

We may now give the characterization of Besov spaces in terms of sequence spaces. Let P_j denote the projection operator onto V_j and $D_j = P_{j+1} - P_j$. A function f belongs to the space $B_{sp\infty}$ if and only if the norm

$$\|P_0 f\|_p + \sup_{j \geqslant 0} 2^{js} \|D_j f\|_p < +\infty.$$

Using now the decomposition of f according to

$$P_0 f = \sum_{k \in \mathbf{Z}} \alpha_{0,k} \varphi_{0,k},$$

$$D_j f = \sum_{k \in \mathbf{Z}} \beta_{j,k} \psi_{jk},$$

we may say equivalently that $f \in B_{sp\infty}$ if

$$\|\alpha_{0.}\|_{l^p} + \sup_{j \geqslant 0} 2^{s+1/2-1/p} \|\beta_{j.}\|_{l^p} < +\infty.$$

This second definition is equivalent to the previous one as a consequence of the lemma of Meyer (1990). This was also helpful to prove Proposition 4.

References

Berg, J. and Löfström, J. (1976) *Interpolation Spaces: an Introduction*. Berlin: Springer-Verlag.

Boufoussi, B. and Roynette, B. (1993) The Brownian local time belongs almost surely to the Besov space $B_p^{1/2,\infty}$. *C. R. Acad. Sci. Paris (I)*, **316**, 843–848.

Chu, C.K. and Marron, J.S. (1991) Choosing a kernel regression estimator. *Statist. Sci.*, **6**, 404–436.

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

Dacunha Castelle, D. and Florens Zmirou, D. (1986) Estimation of the coefficient of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.

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

Fan, J. (1992) Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998–1004.

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

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

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

Genon-Catalot, V., Laredo, C. and Picard, D. (1992) Nonparametric estimation of the diffusion coefficient by wavelets methods. *Scand. J. Statist.*, **19**, 317–335.

Hall, P. (1984) Integrated square error properties of kernel estimators of regression functions. *Ann. Statist*., **12**, 241–260.

Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and its Applications*. New York: Academic Press.

Hoffmann, M. (1997) Minimax estimation of the diffusion coefficient through irregular samplings. *Statist. Probab. Lett.*, **32**, 11–24.

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

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

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

Korostelev, A.P. and Tsybakov, A.D. (1993) *Minimax Theory of Image Reconstruction*. Lecture Notes Statist. **82**. Berlin: Springer-Verlag.

Lipster, R.S. and Shiryayev, A.N. (1977) *Statistics of Random Processes I: General Theory*. Berlin: Springer-Verlag.

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

Nadaraya, E.A. (1964) On estimating regression. *Theory Probab. Applic*., **10**, 186–190.

Peetre, N.C. (1976) *New Thoughts on Besov Spaces*. Math. Ser. I. Durham, NC: Duke University.

Revuz, D. and Yor, M. (1994) *Continuous Martingales and Brownian motion*, 2nd edn. Berlin: Springer-Verlag.

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

Stone, C.S. (1982) Optimal global rates of convergence for nonparametric estimators. *Ann. Statist.*, **10**, 1040–1053.

Wand, M.P. and Jones, M.C. (1995) *Kernal Smoothing*. London: Chapman & Hall.

Zvonkin, A.K. (1974) A transformation of the phase space of a diffusion process that removes the drift. *Math. USSR Sb*., **22**, 129–149.

Received July 1996 and revised October 1997