

ON ESTIMATING THE DIFFUSION COEFFICIENT: PARAMETRIC VERSUS NONPARAMETRIC

Marc HOFFMANN

*Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599,
Université Paris 6 et 7, 16, rue Clisson, 75013 Paris, France*

Received 2 September 1999, revised 29 August 2000

ABSTRACT. – We consider the following problem: estimate the Lipschitz continuous diffusion coefficient σ^2 from the path of a 1-dimensional diffusion process sampled at times i/n , $i = 0, \dots, n$, when we believe that σ^2 actually belongs to a smaller regular parametric set Σ_0 . By introducing random normalizing factors in the risk function, we obtain confidence sets which can be essentially better than the minimax rate $n^{-1/3}$ of estimation for Lipschitz functions in diffusion models. With a prescribed confidence level α_n , we show that the best possible attainable (random) rate is $(\sqrt{\log \alpha_n^{-1}/n})^{2/5}$. We construct an optimal estimator and an optimal random normalizing factor in the sense of Lepski (1999).

This has some consequences for classical estimation: our procedure is adaptive w.r.t. Σ_0 and enables us to test the hypothesis that σ^2 is parametric against a family of local alternatives with prescribed 1st and 2nd-type error probabilities. © 2001 Éditions scientifiques et médicales Elsevier SAS

AMS classification: 62G05; 62M05

1. Introduction

In this paper, we study the statistical estimation of the diffusion coefficient, when one observes a 1-dimensional diffusion process at times i/n , $i = 0, \dots, n$, and asymptotics are taken as $n \rightarrow \infty$. The sample size increases not because of a longer observation period but, rather, because of more frequent observations. This setting has been addressed by several authors, both from a parametric or a nonparametric point of view. A brief summary of the state of the art yields the following conclusions:

- (1) In regular parametric models, the LAMN property holds with rate $1/\sqrt{n}$ (see Donhal [1], or more recently Gobet [3]), but the MLE is not tractable in general. Computationally fast methods based on contrasts are known and possess good optimality properties as far as rates of convergence are concerned (Genon-Catalot and Jacod [2]).

- (2) For nonparametric models, if the diffusion coefficient has smoothness of order s (in a Sobolev or Hölder sense for instance but this can easily be embedded in a Besov space framework), estimators based on kernels (see Jacod [8]) or linear wavelets techniques (see [6]) achieve the rate $n^{-s/(1+2s)}$. This, of course, under some restriction which are specific to diffusion processes. Moreover, the rate $n^{-s/(1+2s)}$ has been proved to be optimal in the minimax sense when the diffusion coefficient possesses at least bounded derivatives up to order 2 (see [6]). However, from a practical point of view, the methods proposed have drawbacks and are not always easily implementable on numerical data.
- (3) But, for nonparametric models with low order of smoothness (precisely: with diffusion coefficient no more regular than Lipschitz continuous), the Nadaraya–Watson estimator, introduced in this context by Florens in [4] – which historically is the first nonparametric estimator of the diffusion coefficient – has good convergence properties and is easy to implement in practice. (In this paper, we also complete Florens’ results by showing that the Nadaraya–Watson estimator achieves the rate $n^{-1/3}$ if the diffusion coefficient is Lipschitz continuous and that this rate is optimal in the minimax sense.)

A caricatural synthesis could be the following: theoretically optimal and computationally fast methods are known when the underlying model is either parametric and regular (take then the contrast estimators of Genon-Catalot and Jacod and the – optimal – rate $1/\sqrt{n}$ is achievable) or nonparametric but the diffusion coefficient is Lipschitz continuous (take the Nadaraya–Watson estimator of Florens and the – optimal – rate $n^{-1/3}$ is achievable).

In this paper, we address the following problem: how can we combine both technologies and what precise mathematical consequences can we derive? We believe that such a question has some importance in practice: given two different methods, a practitioner – motivated by a specific experiment in e.g. finance, biology or physics, say – would legitimately ask which one to choose. Of course, a *prior knowledge*, intuition, suspicion or guess about the underlying structure of the model (parametric) usually exists, and this should be taken into account, even at a mathematical level. Also, the answer we want to give must be numerically feasible and must quantify precisely the consequences of the choice (parametric versus nonparametric), especially when the initial suspicion turns out to be wrong. Our angle is thus the following: we believe that the diffusion coefficient has a given regular parametric structure, but we wish to take into account the possibility that this prior intuition is wrong, in which case the diffusion coefficient could be any Lipschitz continuous function within a certain nonparametric class.

To formulate and solve this problem mathematically, we use the notion of *minimax risk with random normalizing factors*, which is based on *adaptive estimation* and *nonparametric testing* (Theorems 1 and 2 below). The method is easily tractable on numerical data. The ideas developed here heavily rely on the work of Lepski [11]. However, Lepski considers in [11] a slightly different problem in the white noise model context. Therefore, both techniques and answers given here differ a bit from his paper, and we borrow his formalism and mathematical devices rather than complete his theory. Note also that our approach is different from robustness, where misspecified models

are allowed. In general, such models are defined around tubular neighbourhoods of the original parametric model, at a distance vanishing as $n \rightarrow \infty$, an assumption we do not have to make here.

A by-product of our approach is that we complete Florens' paper [4] (and also [6]) by showing that her estimator is optimal in the minimax sense under squared-error loss for Lipschitz continuous diffusion coefficients (Proposition 2), but we know from [9] that this is no longer true for a diffusion coefficient with a higher degree of smoothness.

1.1. Statistical setting

We observe $X^n = (X_{i/n}, i = 0, \dots, n)$ where $(X_t)_{t \in [0,1]}$ is a 1-dimensional diffusion process of the form

$$X_t = x_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(X_s) dW_s, \quad t \in [0, 1] \quad (1.1)$$

with $x_0 \in \mathbf{R}$, W a standard Brownian motion, b smooth, σ Lipschitz continuous and nonvanishing. Our aim is to estimate the function $(\sigma^2(x), x \in I)$, for an arbitrary compact interval I . In this setting, the drift b cannot be identified from the data and is a nuisance parameter.

Formally, we take X as the canonical process on the space $\Omega = \mathcal{C}([0, 1], \mathbf{R})$ of continuous functions equipped with the norm of uniform convergence, endowed with its Borel σ -field \mathcal{F} . We assume that the drift b has linear growth, therefore (1.1) has a unique solution. We further denote by P_{σ^2} the probability measure on (Ω, \mathcal{F}) under which X solves (1.1).

There are several ways of assessing the quality of an estimation procedure. First, the estimation of $\sigma^2(x)$ at a point $x \in I$ is meaningful only if the process X hits the point x before time 1, or if $L_1^x(X) > 0$, where $L_1^x(X) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_0^1 \mathbf{1}_{|X_s - x| \leq \epsilon} ds$ denotes the local time of X at level x and time 1. So if

$$D(x, \nu) = \{\omega \in \Omega: L_1^x(X(\omega)) \geq \nu\},$$

we shall restrict our attention to the set $D(x, \nu)$ for a given $\nu > 0$, fixed throughout the paper. However, the set $D(x, \nu)$ is not observable, therefore it is better for practical purposes to replace – like in Jacod, [8] – the set $D(x, \nu)$ by a set $D^n(x, \nu)$ measurable w.r.t. the σ -field \mathcal{G}_n generated by the $X_{i/n}$, $i = 0, \dots, n$, at stage n . To do so, we introduce the following empirical local time

$$\mathcal{L}_n^x(X^n) = \frac{1}{2n^{2/3}} \sum_{i=1}^n \mathbf{1}_{X_{(i-1)/n} \in [x - n^{-1/3}, x + n^{-1/3}]} \quad (1.2)$$

which converges to $L_1^x(X)$ as $n \rightarrow \infty$ (see, e.g., [9]). The choice of the bandwidth $n^{-1/3}$ will prove to be technically useful. Define

$$D^n(x, \nu) = \{\omega \in \Omega: \mathcal{L}_n^x(X^n(\omega)) \geq \nu\}$$

accordingly. We will further restrict our attention to $D^n(x, \nu)$. For $c > 1$, let $\Sigma_c = \{f : \mathbf{R} \rightarrow \mathbf{R}; c^{-1} \leq f(x) \leq c\}$ and define the Lipschitz class

$$\Sigma = \Sigma_c(L) = \{f : \mathbf{R} \rightarrow \mathbf{R}; |f(x) - f(y)| \leq L|x - y|\} \cap \Sigma_c.$$

The space Σ describes the minimal smoothness properties we require for the unknown parameter σ^2 .

An estimator $T_n = (T_n(X^n, x), x \in I)$ of $(\sigma^2(x), x \in I)$ is a function which is $\mathcal{G}_n \otimes \mathcal{B}(I)$ measurable; we evaluate its performance uniformly over Σ by means of its *minimax risk*

$$\mathbf{R}_n(T_n, \Sigma, \varphi_n) = \sup_{\sigma^2 \in \Sigma} E_{\sigma^2} \left\{ \varphi_n^{-2} \int_I (T_n(X^n, x) - \sigma^2(x))^2 dx \mid D_I^n(\nu) \right\}$$

where $D_I^n(\nu) = \bigcap_{x \in I} D^n(x, \nu)$ and $\varphi_n > 0$ is a normalizing factor. Of course, the finiteness of \mathbf{R}_n will only be meaningful if $\varphi_n \rightarrow 0$ as $n \rightarrow \infty$. Here, E_{σ^2} means integration with respect to the probability P_{σ^2} . Thus we measure the quality of estimation in integrated quadratic loss, conditional on the event $D_I^n(\nu)$.

1.2. Statement of the problem and objectives

An estimator T_n^* is said to attain an optimal rate of convergence $\varphi_n(\Sigma)$ if

$$\limsup_{n \rightarrow \infty} \mathbf{R}_n(T_n^*, \Sigma, \varphi_n(\Sigma)) < +\infty \tag{1.3}$$

and no estimator can attain a better rate:

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \mathbf{R}_n(T_n, \Sigma, \varphi_n(\Sigma)) > 0 \tag{1.4}$$

where the infimum is taken over all estimators. In Section 2, we show that $\varphi_n(\Sigma) = n^{-1/3}$ is an optimal rate of convergence and prove that the Nadaraya–Watson estimator, introduced in this context by Florens in [4] attains the optimal rate (see also [8]). We understand $\varphi_n(\Sigma)$ as an *accuracy of estimation*: for any confidence level $\alpha > 0$, we guarantee from (1.3) the existence of (an explicitly computable) $\gamma_\alpha > 0$ s.t.

$$\sup_{\sigma^2 \in \Sigma} P_{\sigma^2}^{n,\nu} \{ \|T_n^* - \sigma^2\|_I \geq \gamma_\alpha \varphi_n(\Sigma) \} \leq \alpha, \tag{1.5}$$

where $P_{\sigma^2}^{n,\nu}(\cdot) = P_{\sigma^2} \{ \cdot \mid D_I^n(\nu) \}$ and $\|f\|_I = (\int_I f^2(x) dx)^{1/2}$. Furthermore, in the optimality sense described by (1.4), this accuracy is the best one achievable uniformly over Σ .

However, suppose we suspect σ^2 to actually lie in a smaller parametric set, namely $\sigma^2 \in \Sigma_0 = \Sigma_0(I)$ given by

$$\Sigma_0(I) = \{f \in \Sigma: f(x) = \sigma_0^2(x, \theta), \theta \in \Theta, x \in I\},$$

where $\Theta \subset \mathbf{R}^s, s \geq 1$ is given and the function $\sigma_0^2(\cdot, \theta)$ is known up to θ .

Under some regularity assumptions on Θ and σ_0^2 (see Section 2 below), an optimal rate of convergence over Σ_0 is $\varphi_n(\Sigma_0) = n^{-1/2}$ and is attained by the least-square estimator $T_n^{(0)} = \sigma_0^2(\cdot, \hat{\theta}_n)$, where

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \sum_{i=1}^n ((\Delta_i^n X)^2 - \sigma_0^2(X_{(i-1)/n}, \theta))^2 \tag{1.6}$$

and where we denote $\Delta_i^n X = \sqrt{n}(X_{i/n} - X_{(i-1)/n})$ the normalized increments of the observed process. Based on the hypothesis: $\sigma^2 \in \Sigma_0 \subset \Sigma$, we can hope to improve the accuracy of estimation. A traditional way of improvement is the so-called *adaptive approach*.

1.2.1. The adaptive approach

Intuitively, a practitioner would presumably: (1) test the hypothesis $\sigma^2 \in \Sigma_0$, (2) based on the acceptance of the test, choose the parametric estimator $T_n^{(0)}$, (3) keep the nonparametric estimator T_n^* otherwise. From a mathematical point of view, such a procedure – call it temporarily $T_n^{(a)} = T_n^{(a)}(x, X^n)$ – is admissible if it adapts to the sets (Σ_0, Σ) in the following sense: define the *adaptive rate*

$$\psi_n(\sigma^2) = \begin{cases} n^{-1/2} & \text{if } \sigma^2 \in \Sigma_0, \\ n^{-1/3} & \text{if } \sigma^2 \in \Sigma \setminus \Sigma_0. \end{cases} \tag{1.7}$$

Then $T_n^{(a)}$ should verify

$$\limsup_{n \rightarrow \infty} \mathbf{R}_n(T_n^{(a)}, \Sigma, \psi_n(\cdot)) < +\infty. \tag{1.8}$$

However, even if we have satisfied the adaptive criterion (1.8), we are unable to state any accuracy of the method since $\psi_n = \psi_n(\sigma^2)$ depends on the unknown, we cannot provide any confidence set of the type (1.5).

In this paper, we propose an alternative approach by introducing a procedure based on random normalizing factors (r.n.f. for abbreviation), following Lepski in [11] and [7]. This will enable us to improve the accuracy of estimation in the sense of (1.5). We will even show that a procedure based on r.n.f. can simultaneously give an improvement of accuracy and be adaptive in the sense of (1.8).

1.2.2. Random normalizing factors

We introduce the class of observable normalizing factors (r.n.f.)

$$\Omega_n = \{ \rho_n \in (0, \varphi_n(\Sigma)]: \rho_n \text{ is } \mathcal{G}_n\text{-measurable} \},$$

where \mathcal{G}_n is the σ -field generated by the observation $X_{i/n}$, $i = 0, \dots, n$. Clearly, any estimator T_n satisfying

$$\limsup_{n \rightarrow \infty} \mathbf{R}_n(T_n, \Sigma, \rho_n) < \infty \quad \text{for some } \rho_n \in \Omega_n \tag{1.9}$$

attains the optimal rate of convergence over Σ . But in contrast to an adaptive estimator, we now guarantee the existence of an explicitly computable γ_α from (1.9) such that for

any $\alpha > 0$:

$$\sup_{\sigma^2 \in \Sigma} P_{\sigma^2}^{n,v} \{ \|T_n^* - \sigma^2\|_I \geq \gamma_\alpha \rho_n \} \leq \alpha,$$

where we have set – recall (1.5) – $P_{\sigma^2}^{n,v}(\cdot) = P_{\sigma^2}\{\cdot \mid D_I^n(v)\}$. This provides us with a new (possibly random) accuracy of estimation. The possibility that σ^2 belongs to Σ_0 may give a value to ρ_n essentially better than $\varphi_n(\Sigma) = n^{-1/3}$ with some probability, while still ensuring a confidence set uniformly over Σ .

Next, we need a consistent way to compare 2 r.n.f. in order to define an optimality criterion. Since a r.n.f. is random, we introduce the following (deterministic) characteristic:

DEFINITION 1. – For a given confidence level $0 < \alpha_n < 1$, the characteristic of $\rho_n \in \Omega_n$ is

$$\chi_n(\rho_n) = \inf\{t \in (0, \varphi_n(\Sigma))\}: \inf_{\sigma^2 \in \Sigma_0} P_{\sigma^2}^{n,v}(\rho_n \leq t) \geq 1 - \alpha_n\}.$$

Note that $\chi_n(\rho_n)$ depends on α_n and on Σ_0 .

Remark. – A heuristic approach to understand this definition can be the following: let us fix $t > 0$ “small”, i.e. at least smaller than $\varphi_n(\Sigma)$. What we require is that a “good” ρ_n will provide improvement of accuracy if the guess ($\sigma^2 \in \Sigma_0$) turns out to be true. This means that under $P_{\sigma^2}^{n,v}$, for $\sigma^2 \in \Sigma_0$, the event “ $\rho_n \leq t$ ” has a controlled probability. Mathematically, we translate this idea by saying that for a given confidence level α_n , we guarantee that

$$\inf_{\sigma^2 \in \Sigma_0} P_{\sigma^2}^{n,v}(\rho_n \leq t) \geq 1 - \alpha_n. \tag{1.10}$$

Next, the smaller t we can find such that (1.10) holds, the better ρ_n hence $\chi_n(\rho_n)$ is defined as infimum of t providing (1.10).

We now have a canonical way to compare r.n.f. We naturally derive the following optimality criterion:

DEFINITION 2. – $\rho_n^* \in \Omega_n$ is optimal (or α -optimal) w.r.t. (Σ, Σ_0) if

(i) There exists an estimator T_n^{**} such that

$$\limsup_{n \rightarrow \infty} \mathbf{R}_n(T_n^{**}, \Sigma, \rho_n^*) < +\infty.$$

(ii) For any $\rho_n \in \Omega_n$ such that

$$\frac{\chi_n(\rho_n)}{\chi_n(\rho_n^*)} \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

we have

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \mathbf{R}_n(T_n, \Sigma, \rho_n) = +\infty$$

where the infimum is taken over all estimators.

Remark 2. – Following Lepski, we call T_n^{**} an α -adaptive estimator. Note that by definition, we always have $\rho_n^* \leq \varphi_n(\Sigma)$. Thus an α -adaptive estimator is optimal with respect to Σ . Also, note that an α -optimal random normalizing factor may depend in general on the quantity α_n .

The aim of this paper is to construct an optimal random normalizing factor w.r.t. (Σ, Σ_0) following Definition 2 and an α -optimal estimator accordingly.

1.3. Organization of the paper

In Section 2, we recall and adapt some facts about statistical estimation of the diffusion coefficient from discrete observations. The nonparametric kernel estimator of Section 2.2 was introduced in [4] and later generalized to our setting in [8]. However, we give a self-contained proof of the upper bound. The lower bound is new, and follows the same strategy as Proposition 1 in [6], with new technicalities.

Section 3 is devoted to the construction of an optimal r.n.f. for the diffusion coefficient under a parametric hypothesis. We discuss the link to adaptive estimation and show how a slight modification of the optimal (α -adaptive) estimator enables to obtain simultaneously an optimal accuracy of estimation and an adaptive estimator. Links to testing are also mentioned. The proofs are delayed until Section 4.

2. Preliminary results

2.1. Parametric estimation

We need the following regularity assumptions on Θ and σ_0^2 .

Assumption A. – We have

- (1) The set Θ is compact in $\mathbf{R}^s = \mathbf{R}$.
- (2) For some $M = M_I > 0$,

$$\sup_{x \in I} |\sigma_0^2(x, \theta_1) - \sigma_0^2(x, \theta_2)| \leq M |\theta_1 - \theta_2| \quad \text{for } \theta_1, \theta_2 \in \Theta.$$

- (3) For some $\eta = \eta_I > 0$, $\inf_{(x, \theta) \in I \times \Theta} \left| \frac{d}{d\theta} \sigma_0^2(x, \theta) \right| \geq \eta$.
- (4) The functions $\frac{d^2}{dx^2} \sigma_0^2$ and $\frac{d}{dx} \frac{d}{d\theta} \sigma_0^2$ are well defined and continuous on $I \times \Theta$.
- (5) The equality $\sigma_0^2(x, \theta_1) = \sigma_0^2(x, \theta_2)$ for all $x \in I$ implies $\theta_1 = \theta_2$.

Remark 1. – The above assumptions are standard in parametric estimation but do not claim to be minimal. For instance, A5 can be relaxed but known extensions are usually difficult to check (see, e.g., [2]).

By Ito formula, the model given by (1.1) can be recast in a regression setting, having

$$(\Delta_i^n X)^2 = n \int_{(i-1)/n}^{i/n} \sigma^2(X_s) ds + \varepsilon_i^n + \text{a higher order term}, \quad (2.11)$$

where $\varepsilon_i^n = 2n \int_{(i-1)/n}^{i/n} (X_s - X_{(i-1)/n}) dW_s$ is a martingale increment. Thus we observe on the non-uniform random grid $(X_{(i-1)/n}, i = 1, \dots, n)$ the value $n \int_{(i-1)/n}^{i/n} \sigma^2(X_s) ds \sim \sigma^2(X_{(i-1)/n})$, contaminated by the noise ε_i^n , plus a negligible drift effect. From this formulation, we readily obtain the least-square estimator $T_n^{(0)} = \sigma_0^2(\cdot, \hat{\theta}_n)$, where $\hat{\theta}_n$ is defined by (1.6).

PROPOSITION 1. – *Grant Assumption A. Then $\varphi_n(\Sigma_0) = n^{-1/2}$ is an optimal rate of convergence and is attained by $T_n^{(0)}$.*

The proof of the upper bound is readily obtained from Theorem 1 in [2]. We simply added *ad hoc* assumptions in order to obtain the uniformity in $\theta \in \Theta$ for the integrated risk. The proof of the lower bound follows from the LAMN property of the parametric model (see, e.g., [1]).

Remark 2. – Note that $\hat{\theta}_n$ is not the best available parametric estimator of θ – it is not equivalent to the MLE – but since we focus on rates of convergence only, this intuitively simple choice is sufficient.

2.2. Nonparametric estimation

We assume (with no loss of generality as far as practical considerations are concerned) that I is on a dyadic scale, namely

$$I = [k_0 2^{-j_0}, k_1 2^{-j_0}], \quad \text{for some integers } k_0, k_1, j_0.$$

For integers (k, j) , let $I_{jk} = [k 2^{-j}, (k + 1) 2^{-j})$ and define for $j \geq j_0$

$$V_I^j = \{f \in L_2: f \text{ is constant on } I_{jk}, I_{jk} \subset I\}$$

the finite element space of functions which are piecewise constant on I over a grid of mesh 2^{-j} . Indeed an orthogonal basis for V_I^j is given by the family

$$\phi_{jk} = 2^{j/2} \phi(2^j \cdot -k), \quad k \text{ such that } I_{jk} \subset I,$$

where $\phi = 1_{[0,1)}$. We estimate σ^2 by an element of $V_I^{j_n}$, for some projection level j_n chosen in accordance with the asymptotics in the following way. Let

$$\hat{c}_{j_n k} = \frac{1}{n \ell_{j_n k}} \sum_{i=1}^n (\Delta_i^n X)^2 \phi_{j_n k}(X_{(i-1)/n}),$$

(with $0/0 = 0$), where $\ell_{j_n k} = \frac{2^{j_n}}{n} \sum_{i=1}^n 1_{X_{(i-1)/n} \in I_{j_n k}}$. Informally, we use the regression analogy

$$(\Delta_i^n X)^2 \sim \sigma^2(X_{(i-1)/n}) + \varepsilon_i^n$$

defined in (2.11) and we weight the local average by an approximation $\ell_{j_n k}$ of the time spent by the process X in $I_{j_n k}$. Finally, the nonparametric estimator $T_n^*(x)$ of $\sigma^2(x)$ is

defined by

$$T_n^*(x) = \sum_{k: I_{j_n k} \subset I} \hat{c}_{j_n k} \phi_{j_n k}(x)$$

and is specified by the projection level j_n . The performances of T_n^* are summarized in the following result. For $x \in \mathbf{R}$, we denote by $\lfloor x \rfloor$ the integer part of x .

PROPOSITION 2. – *An optimal rate of convergence over Σ is $\varphi_n(\Sigma) = n^{-1/3}$. Moreover, T_n^* , calibrated by $j_n = \lfloor \frac{1}{3} \log n \vee j_0 \rfloor$ is optimal for the criterion given by (1.3) and (1.4).*

3. Main result

This section is devoted to the construction of an optimal r.n.f. ρ_n^* in the sense of Definition 2. Accordingly, we construct an α -adaptive estimator w.r.t. (Σ_0, Σ) . Our algorithm can be described as follows:

- (1) Estimate the distance \hat{d}_n (for the $\|\cdot\|_I$ seminorm) between σ^2 and Σ_0 ,
- (2) Take $\rho_n^* = n^{-1/3}$ if \hat{d}_n is above some threshold level (possibly depending on the confidence level α_n).
- (3) Take $\rho_n^* = \varphi_{n, \alpha_n} > 0$ for some normalizing factor φ_{n, α_n} (tuned with the asymptotics) otherwise.

We will show that we can take φ_{n, α_n} converging to 0 faster than $n^{-1/3}$ by a polynomial power. The value φ_{n, α_n} corresponds to the acceptance that $\sigma^2 \in \Sigma_0$ and measures the improvement of the accuracy of estimation. The assumptions on the parametric family Σ_0 are less stringent than in Section 2.

Assumption B. – We have

- (1) The set Θ is compact in \mathbf{R}^s , $s \geq 1$.
- (2) There exist $\mu > 0$ and $M = M_{I, \mu} > 0$ such that

$$\sup_{x \in I} |\sigma_0^2(x, \theta_1) - \sigma_0^2(x, \theta_2)| \leq M \|\theta_1 - \theta_2\|^\mu \quad \text{for } \theta_1, \theta_2 \in \Theta.$$

- (3) For all $\theta \in \Theta$, $\sigma_0^2(\cdot, \theta) \in \Sigma_0$.
- (4) There exist $\theta_0 \in \Theta$ and $\tilde{L} < L$ such that $\sigma_0^2(\cdot, \theta_0) \in \Sigma_c(\tilde{L})$.

3.1. Construction of ρ_n^* and main result

For $J_n \geq j_0$ and $\theta \in \Theta$, define

$$d_n(\theta) = \sum_{k: I_{J_n k} \subset I} \{ \hat{c}_{J_n k} - c_{J_n k}(\sigma_0^2(\cdot, \theta)) \}^2,$$

where

$$c_{J_n k}(\sigma_0^2(\cdot, \theta)) = 2^{J_n/2} \int_{I_{J_n k}} \sigma_0^2(x, \theta) dx.$$

By Assumption B2, there exists $C = C(\mu, M_{I,\mu})$ such that

$$|d_n(\theta_1) - d_n(\theta_2)| \leq C \|\theta_1 - \theta_2\|^\mu \quad \text{for } \theta_1, \theta_2 \in \Theta.$$

This, together with Assumption B1, ensures the existence of, $\theta_n^* \in \Theta$, measurable w.r.t. \mathcal{G}_n -measurable and solution to

$$d_n(\theta_n^*) = \inf_{\theta \in \Theta} d_n(\theta).$$

For technical reasons, we need to compensate the variance of d_n as follows. Let

$$C_{J_n k} = \frac{2}{3n^2 \ell_{J_n k}^2} \sum_{i=1}^n (\Delta_i^n X)^4 \phi_{J_n, k}^2(X_{(i-1)/n})$$

and

$$\hat{d}_n(\theta_n^*) = d_n(\theta_n^*) - \sum_{k: I_{J_n k} \subset I} C_{J_n k}.$$

The estimate $\hat{d}_n(\theta_n^*)$ will determine the following decision rule. For $\alpha_n > 0$, let $J_n = \lfloor \frac{2}{5} \log \frac{n}{\sqrt{\log \alpha_n^{-1}}} \vee j_0 \rfloor$ and

$$\varphi_{n, \alpha_n} = \left(\frac{\sqrt{\log \alpha_n^{-1}}}{n} \right)^{2/5}.$$

Note that 2^{-J_n} and φ_{n, α_n} coincide up to a constant. Define now the threshold

$$\lambda = A(10, c, \nu)^{-1/4},$$

where $A(t, c, \nu)$ is specified in (4.17), see the proof of Lemma 3. (The choice of λ will become transparent in the proof of Theorem 1 in Section 4 below when Lemma 3 is used.) The decision rule then takes the form

$$\rho_n^* = \varphi_{n, \alpha_n} 1_{\{\hat{d}_n(\theta_n^*) \leq \lambda^2 \varphi_{n, \alpha_n}^2\}} + n^{-1/3} 1_{\{\hat{d}_n(\theta_n^*) > \lambda^2 \varphi_{n, \alpha_n}^2\}}$$

and

$$\bar{T}_n(x) = \sum_{k: I_{J_n k} \subset I} c_{J_n k} (\sigma_0^2(\cdot, \theta_n^*)) \phi_{J_n k}(x).$$

Finally, our estimator of $\sigma^2(x)$ is

$$T_n^{**}(x) = \bar{T}_n(x) 1_{\{\rho_n^* = \varphi_{n, \alpha_n}\}} + T_n^*(x) 1_{\{\rho_n^* = n^{-1/3}\}},$$

where T_n^* is the nonparametric estimator of Section 2.2 specified with $j_n = \lfloor \frac{1}{3} \log n \vee j_0 \rfloor$.

THEOREM 1. – *Grant Assumption B. Assume that $n^{-a} \leq \alpha_n < 2^{-4}$ for some arbitrary $a > 0$. Then ρ_n^* is an optimal random normalizing factor w.r.t. (Σ_0, Σ) and T_n^{**} is α -adaptive.*

Remark 1. – In particular, we see that if our parametric assumption is correct, with prescribed confidence $1 - \alpha_n$, we are able to improve (asymptotically) the accuracy of estimation, i.e., the size of the confidence band we construct, by a factor $\varphi_{n,\alpha_n} = (\sqrt{\log \alpha_n^{-1}/n})^{2/5}$.

Remark 2. – The improvement of the random confidence band is of a polynomial order, but is lowered down by the size of α_n . However, the restriction $\alpha_n \geq n^{-a}$ ensures that it is at least of order $(\sqrt{\log n/n})^{2/5}$.

Remark 3. – For practical purposes, it seems more clever to replace ν in the definition of λ by $\inf_{k: I_{J_n k} \subset I} \ell_{J_n k}$. For technical reasons, we are unable however to prove Theorem 1 in this setting. Note also that the practical implementation of ρ_n^* is easy: the computation of $\hat{c}_{J_n k}$ is reasonably fast and the cardinality (in k) of such $\hat{c}_{J_n k}$ is of order $\log n$. Likewise for the $C_{J_n k}$. Eventually, the minimization problem $\arg \min_{\theta} d_n(\theta)$ has the same complexity as the computation of a standard parametric estimator. Once ρ_n^* is computed, one readily computes either \bar{T}_n (which is no more difficult to obtain than T_n^*) or T_n^* itself, the standard Nadaraya–Watson estimator.

3.2. Discussion

3.2.1. Links to adaptive estimation

We show in this paragraph how a simple modification of T_n^{**} provides us with an adaptive estimator – in the usual sense of (1.8) – without losing the optimality in terms of r.n.f. Assumption A is in force here. We consider the estimator $T_n^{(0)} = \sigma_0^2(\cdot, \hat{\theta}_n)$ introduced in Section 1.2, where $\hat{\theta}_n$ solves (1.6). Note that $T_n^{(0)}$ is well defined thanks to Assumption A. Let

$$T_n^{(a)}(x) = T_n^{(0)}(x)1_{\{\hat{d}_n(\hat{\theta}_n) \leq \lambda^2 \varphi_{n,\alpha_n}^2\}} + T_n^*(x)1_{\{\hat{d}_n(\hat{\theta}_n) > \lambda^2 \varphi_{n,\alpha_n}^2\}}.$$

Define the random normalizing factor $\rho_n^{(a)}$ accordingly

$$\rho_n^{(a)} = \varphi_{n,\alpha_n} 1_{\{\hat{d}_n(\hat{\theta}_n) \leq \lambda^2 \varphi_{n,\alpha_n}^2\}} + n^{-1/3} 1_{\{\hat{d}_n(\hat{\theta}_n) > \lambda^2 \varphi_{n,\alpha_n}^2\}}.$$

THEOREM 2. – *Grant Assumption A. We have*

- (i) *The estimator $T_n^{(a)}$ is α -adaptive w.r.t. (Σ_0, Σ) .*
- (ii) *If moreover $\alpha_n = \mathcal{O}(n^{-1})$, the estimator $T_n^{(a)}$ is adaptive in the usual sense w.r.t. (Σ_0, Σ) :*

$$\limsup_{n \rightarrow \infty} \mathbf{R}_n(T_n^{(a)}, \Sigma, \psi_n(\cdot)) < \infty,$$

where $\psi_n(\sigma^2)$ denotes the adaptive rate defined by (1.7).

The proof of (i) is delayed until Section 4. The proof of (ii) is a direct consequence of (i) and Proposition 2 in [11].

Remark 1. – The decision rule $1_{\{\hat{d}_n(\hat{\theta}_n) \leq \lambda^2 \varphi_{n,\alpha_n}^2\}}$ answers to our original question: given a parametric procedure $T_n^{(0)}$ versus a nonparametric one T_n^* , which one shall we use in practice? The answer 1 to our test yields the choice $T_n^{(0)}$, whereas the answer 0 yields

T_n^* . The precise mathematical consequences of this choice are described by the r.n.f. $\rho_n^{(a)}$. Moreover, this choice is optimal in the sense of Definition 2.

Remark 2. – Again – see Remark 3 in Section 3.1 – we can see that the practical implementation of $T_n^{(a)}$ is fast and has complexity no worse than that of T_n^* .

3.2.2. Links to testing

We explore in this section another virtue of ρ_n^* , namely the possibility to build a test for the hypothesis $\sigma^2 \in \Sigma_0$ against a family of local alternatives. More precisely, given $h > 0$ and $0 < \alpha < 1$, define

$$W_n(h, I, \alpha) = \{f \in \Sigma_c(L) : \inf_{\theta \in \Theta} \|f - \sigma_0^2(\cdot, \theta)\|_I \geq h \cdot \varphi_{n,\alpha}\}.$$

THEOREM 3. – *Grant Assumption A.* For any $0 < \alpha, \beta < 1$, we have, for large enough n

- (i) $\sup_{\sigma^2 \in \Sigma_0(I)} P_{\sigma^2}^{n,v}(\rho_n^* = n^{-1/3}) \leq \alpha.$
- (ii) *There exists $h(\beta) > 0$ such that*

$$\sup_{\sigma^2 \in W_n(h(\beta), I, \alpha)} P_{\sigma^2}^{n,v}(\rho_n^* = \varphi_{n,\alpha}) \leq \beta.$$

In words, the hypothesis: $\sigma^2 \in \Sigma_0(I)$ can be tested against the family of local alternatives: $\sigma^2 \in W_n(h(\beta), I, \alpha)$ with prescribed first and second type error probability. The proof of (i) readily follows from (4.27) in the proof of Theorem 1 below. The proof of (ii) follows from Theorem 1 together with Proposition 3 in [11].

4. Proofs

The proof of Proposition 2 can be read independently from the that of Theorems 1 and 2. The reader is however invited to first scan the *notation and preliminaries* section.

4.1. Notation and preliminaries

For $\theta \in \Theta$, we abbreviate $P_{\sigma_0^2(\cdot, \theta)}$ by P_θ when no confusion is possible. For $\sigma^2 \in \Sigma$, let \tilde{P}_{σ^2} denote the law of the process Y such that $dY_t = \sigma(Y_t) dW_t$, $Y_0 = x_0$. Define $\tilde{P}_{\sigma^2}^{n,v}$ accordingly. We denote by C a generic constant, possibly varying from line to line and which may depend on I . Any other dependence will be explicitly mentioned.

4.1.1. Preliminary decompositions

- (a) For $\sigma^2 \in \Sigma$, define

$$a_k^n(\sigma^2) = \frac{2^{3J_n/2}}{n \ell_{J_{nk}}} \sum_{i=1}^n 1_{X_{(i-1)/n} \in I_{J_{nk}}} \int_{I_{J_{nk}}} \tau_i^n(x, X) dx,$$

$$b_k^n(\sigma^2) = \frac{2^{J_n/2}}{n \ell_{J_{nk}}} \sum_{i=1}^n 1_{X_{(i-1)/n} \in I_{J_{nk}}} \varepsilon_i^n,$$

where

$$\tau_i^n(x, X) = \sigma^2(x) - n \int_{(i-1)/n}^{i/n} \sigma^2(X_s) ds. \tag{4.12}$$

Define the random variable

$$\Delta_n(\sigma^2) = \sum_{k: I_{Jnk} \subset I} \{c_{Jnk}(\sigma_0^2(\cdot, \theta_n^*) - \sigma^2) - a_k^n(\sigma^2)\}^2, \tag{4.13}$$

where – recall Section 2 – we denote $c_{jk}(\sigma^2) = 2^{j/2} \int_{I_{jk}} \sigma^2(x) dx$. Using that $(\Delta_i^n X)^2 = n \int_{(i-1)/n}^{i/n} \sigma^2(X_s) ds + \varepsilon_i^n$ under \tilde{P}_{σ^2} we have

$$\hat{c}_{Jnk} - c_{Jnk} = a_k^n(\sigma^2) + b_k^n(\sigma^2).$$

We thus obtain the following decomposition

$$\hat{d}_n(\theta_n^*) = \Delta_n(\sigma^2) + M_n(\sigma^2) + N_n(\sigma^2),$$

having

$$M_n(\sigma^2) = 2 \sum_{k: I_{Jnk} \subset I} [c_{Jnk}(\sigma^2 - \sigma_0^2(\cdot, \theta_n^*) - a_k^n(\sigma^2))] b_k^n(\sigma^2),$$

$$N_n(\sigma^2) = \sum_{k: I_{Jnk} \subset I} [(b_k^n)^2 - C_{Jnk}].$$

(b) Define $\psi(x) = 1_{[0,1/2)}(x) - 1_{[1/2,1)}(x)$. If $d_{jk}(\sigma^2) = \int \sigma^2(x) \psi_{jk}(x) dx$ is the wavelet coefficient of σ^2 in the Haar basis, from the multiscale decomposition of σ^2 , we have by Parseval’s identity

$$\begin{aligned} \|\bar{T}_n - \sigma^2\|_I^2 &= \sum_{k: I_{Jnk} \subset I} c_{Jnk}^2(\sigma^2 - \sigma_0^2(\cdot, \theta_n^*)) + \epsilon_{J_n}(\sigma^2) \\ &\leq 2\Delta_n(\sigma^2) + 2 \sum_{k: I_{Jnk} \subset I} (a_k^n(\sigma^2))^2 + \epsilon_{J_n}(\sigma^2), \end{aligned}$$

where the remainder term $\epsilon_{J_n}(\sigma^2) = \sum_{j \geq J_n} \sum_{k: I_{jk} \subset I} d_{jk}^2(\sigma^2)$ satisfies

$$\epsilon_{J_n}(\sigma^2) \leq C(c, L)2^{-2J_n} = C'(c, L)\varphi_{n, \alpha_n}^2$$

since $\sigma^2 \in \Sigma$. Note that

$$|\tau_i^n(x, X)| 1_{X_{(i-1)/n} \in I_{Jnk}} \leq L \left(2^{-J_n} + n \int_{(i-1)/n}^{i/n} |X_s - X_{(i-1)/n}| ds \right), \tag{4.14}$$

therefore $|\tau_i^n(x, X)| 1_{X_{(i-1)/n} \in I_{Jnk}} \leq L(2^{-J-n} + n^{-1/2} r_i^n(\sigma^2))$, and since we have $\sup_{\sigma^2 \in \Sigma} \tilde{E}_{\sigma^2}^{n, \nu} (|r_i^n(\sigma^2)|^p) \leq C_p$ for all $p \geq 1$ by the Burkholder–Davis–Gundy inequality

– further abbreviated by BDG – it follows that

$$|a_k^n(\sigma^2)| \leq L(2^{-3J_n/2} + 2^{-J_n/2}n^{-1/2}r_i^n)$$

and finally

$$\tilde{E}_{\sigma^2}^{n,v} \left\{ \sum_k (a_k^n(\sigma^2))^2 \right\} \leq C(c, L)2^{-2J_n} = C'(c, L)\varphi_{n,\alpha_n}^2. \tag{4.15}$$

(c) The random function $\sigma_0^2(\cdot, \theta_n^*) - \sigma^2$ belongs to $\Sigma_{2c}(L)$. Moreover, since $\sup_x |\tau_i^n(x, X)| \leq 2c$, we always have $|a_k^n(\sigma^2)| \leq 2^{-J_n/2+1}c$. Therefore $|\Delta_n(\sigma^2)| \leq 8c^2$.

(d) We further decompose $N_n = N_n^{(1)} + N_n^{(2)}$, having

$$N_n^{(1)} = \sum_{k: I_{J_n k} < l} \frac{1}{n^2 \ell_{J_n k}^2} \sum_{i=1}^n \left[(\varepsilon_i^n)^2 - \frac{2}{3}(\Delta_i^n X)^2 \right] \phi_{J_n k}^2(X_{(i-1)/n})$$

$$N_n^{(2)} = \sum_{k: I_{J_n k} < l} \frac{1}{n^2 \ell_{J_n k}^2} \sum_{1 \leq i < j \leq n} \varepsilon_i^n \varepsilon_j^n \phi_{J_n k}(X_{(i-1)/n}) \phi_{J_n k}(X_{(j-1)/n}).$$

4.1.2. Technical lemmas

LEMMA 1. – Let $x > 0$. For all $t \geq 2$, we have

$$\tilde{P}_{\sigma^2}^{n,v} \{ N_n^{(1)} \geq x \} \leq C(t, c, L, v)n^{-3t/2}2^{J_n t} x^{-t}.$$

Proof. – It is easily seen that $(\varepsilon_i^n)^2 - \frac{2}{3}(\Delta_i^n X)^2 = U_i^n + V_i^n$, with

$$U_i^n = \sigma^4(X_{(i-1)/n}) [((\Delta_i^n W)^2 - 1)^2 - \frac{2}{3}(\Delta_i^n W)^4]$$

and where the remainder term V_i^n satisfies (apply the BDG inequality)

$$\sup_{\sigma^2 \in \Sigma} \tilde{E}_{\sigma^2}^{n,v} \{ |V_i^n|^p \} \leq C(p, c, v)n^{-p/2}, \quad \text{for all } p > 0.$$

Since under $\tilde{P}_{\sigma^2}^{n,v}$, we have $\inf_k \ell_{J_n k}^2 \geq v^2$ and noting that $\sum_k \phi_{J_n k}^2(X_{(i-1)/n}) \leq 2^{J_n}$, it suffices then to bound

$$\tilde{P}_{\sigma^2}^{n,v} \left\{ \left| \frac{1}{n^2} \sum_{i=1}^n Z_i^n \right| \geq \frac{1}{2} 2^{-J_n} x v \right\}, \quad Z_i^n = U_i^n, V_i^n.$$

By Chebyshev’s inequality, we readily see that the term involving $Z_i^n = V_i^n$ has the right order. Likewise, using that $E_{\sigma^2}^{n,v} \{ |\frac{1}{\sqrt{n}} \sum_{i=1}^n U_i^n|^p \} \leq C(p, c, v)$ since the U_i^n are zero-mean and have bounded moments of any order, we obtain the desired result. \square

LEMMA 2 (Fuk and Nagaev [5]). – Let $S_n = \sum_{i=1}^n M_i$ be a \mathcal{F}_n -martingale. For $x > 0$ and $t \geq 2$, we have

$$P(|S_n| \geq x) \leq C_1(t)A_{t,n}x^{-t} + \exp \left\{ C_2(t) \frac{x^2}{B_n^2} \right\},$$

where $A_{t,n} = \sum_{i=1}^n E(|M_i|^t)$, $B_n^2 = \sum_{i=1}^n E(M_i^2)$ and $C_1(t) = (1 + 2/t)^t$, $C_2(t) = 2/(t + 2)^2 e^{-t}$.

LEMMA 3. – Let $x > 0$. For all $t \geq 2$, we have

$$\tilde{P}_{\sigma^2}^{n,v} \{N_n^{(2)} \geq x\} \leq C(t, c, L, v) n^{1-3t/2} 2^{J_n t} x^{-t} + \exp\{-A(t, c, v) x^2 n^2 2^{-J_n}\},$$

where $A(t, c, v)$ is specified in (4.17) below.

Proof. – Define, for $j = 2, \dots, n$

$$\chi_j^n = 2^{-J_n} \sum_{k: I_{J_n k} \subset I} \varepsilon_j^n \phi_{J_n k}(X_{(j-1)/n}) \sum_{i=1}^{j-1} \varepsilon_i^n \phi_{J_n k}(X_{(i-1)/n}).$$

(a) Since under $\tilde{P}_{\sigma^2}^{n,v}$, we have $\inf_k \ell_{J_n k}^2 \geq v^2$, it suffices to bound

$$\tilde{P}_{\sigma^2}^{n,v} \left\{ \left| \sum_{j=1}^n \chi_j^n \right| \geq n^2 2^{-J_n} v^2 x \right\}.$$

By Lemma 2, this last quantity is less than

$$C_1(t) \sum_{j=1}^n \tilde{E}_{\sigma^2}^{n,v} \{|\chi_j^n|^t\} + \exp\left\{-C_2(t) v^2 x^2 n^2 2^{-J_n} \frac{n^2 2^{-J_n}}{B_n^2}\right\}, \tag{4.16}$$

where $B_n^2 = \sum_{j=1}^n \tilde{E}_{\sigma^2}^{n,v} \{(\chi_j^n)^2\}$. We readily have

$$\begin{aligned} \tilde{E}_{\sigma^2}^{n,v} \{|\chi_j^n|^t\} &\leq 2^{-J_n t} E_{\sigma^2}^{n,v} \left\{ |\varepsilon_j^n|^t \sup_k \phi_{J_n k}^t(X_{(j-1)/n}) \left| \sum_{i=1}^{j-1} \varepsilon_i^n \sum_k \phi_{J_n k}(X_{(i-1)/n}) \right|^t \right\} \\ &\leq C(c, L, v) 2^{-J_n t/2} \left(E_{\sigma^2}^{n,v} \left\{ \left| \sum_{i=1}^{j-1} \varepsilon_i^n \sum_k \phi_{J_n k}(X_{(i-1)/n}) \right|^{2t} \right\} \right)^{1/2}, \end{aligned}$$

where we used Cauchy–Schwarz inequality and $\sup_k \phi_{J_n k}(X_{(j-1)/n}) \leq 2^{J_n/2}$. Using $\sum_k \phi_{J_n k}(X_{(i-1)/n}) \leq 2^{J_n/2}$ and a martingale argument, the last quantity is less than $C(t, c, L, v) j^{t/2}$. Using $\sum_{j=1}^n j^{t/2} \leq C n^{t/2+1}$, we obtain the desired bound for the first term.

(b) It remains to bound B_n^2 . The process $v_{k,j}^n = \frac{1}{\sqrt{n}} \sum_{i=1}^{j-1} \varepsilon_i^n \phi_{J_n k}(X_{(i-1)/n})$ indexed by j is a martingale, therefore

$$B_n^2 = n 2^{-2J_n} \sum_{j=1}^n \tilde{E}_{\sigma^2}^{n,v} \left\{ \left(\sum_{k: I_{J_n k} \subset I} \phi_{J_n k}(X_{(j-1)/n}) v_{k,j}^n \right)^2 \right\}.$$

Now

$$\left| \sum_{k: I_{J_n k} \subset I} \phi_{J_n k}(X_{(j-1)/n}) v_{k,j}^n \right| \leq 2^{J_n/2} |v_{k(X_{(j-1)/n}, j)}^n|,$$

where $k(X_{(j-1)/n}) = \inf\{k: X_{(j-1)/n} \leq (k + 1)2^{-J_n}\}$. Using

$$(v_{k(X_{(j-1)/n}),j}^n)^2 \leq \sum_k \sup_{1 \leq l \leq n} (v_{k,l}^n)^2 1_{\{k(X_{(j-1)/n})=k\}}$$

and $\{k(X_{(j-1)/n}) = k\} = \{X_{(j-1)/n} \in I_{J_n k}\}$, we successively have

$$\begin{aligned} B_n^2 &\leq n2^{-J_n} \sum_{j=1}^n \tilde{E}_{\sigma^2}^{n,v} \{(v_{k,j}^n)^2\} \\ &\leq n2^{-J_n} \sum_k \tilde{E}_{\sigma^2}^{n,v} \{(\sup_l v_{k,l}^n)^4\}^{1/2} \tilde{E}_{\sigma^2}^{n,v} \left\{ \left(\sum_{j=1}^n 1_{X_{(j-1)/n} \in I_{J_n k}} \right)^2 \right\}^{1/2} \end{aligned}$$

by Cauchy–Schwarz inequality. Using Doob’s inequality, we get

$$\tilde{E}_{\sigma^2}^{n,v} \{(\sup_l v_{k,l}^n)^4\} \leq C(c, v).$$

Likewise, we have

$$\tilde{E}_{\sigma^2}^{n,v} \left\{ \left(\frac{2^{J_n}}{n} \sum_{j=1}^n 1_{X_{(j-1)/n} \in I_{J_n k}} \right)^2 \right\} \leq \tilde{E}_{\sigma^2}^{n,v} \{(\sup_x \mathcal{L}_n^x(X^n))^2\} \leq C(c, L, v)$$

by approximation of the local time (see, e.g., [9]). Therefore $B_n^2 \leq C(c, v)n^2 2^{-J_n}$. A more detailed examination of B_n^2 shows that

$$B_n^2 \leq n^2 2^{-J_n} Z(c) := n^2 2^{-J_n} 2c^2 \inf_q C(2p)K(c, q),$$

where $C(2p)$ is the optimal constant in Rosenthal’s inequality, and $1/p + 1/q = 1$. The constant $K(c, q)$ is defined by

$$K(c, q) \geq E \{(\sup_x L_{c^2}^x)^2\} [E \{(\sup_x L_{c^2}^x)^q\}]^{1/q},$$

where L_t^x is the Brownian local time. (The proof is rather technical so we omit it.) Back to (4.16), we see that the choice

$$A(t, c, v) := C_2(t)v^2/Z(c) \tag{4.17}$$

yields the desired bound. \square

4.2. Proof of Proposition 2: upper bound

(a) Note that on I_{jnk} , $T_n^*(x)$ is identically equal to

$$\frac{\frac{2^{jn}}{n} \sum_{i=1}^n (\Delta_i^n X)^2 1_{X_{(i-1)/n} \in I_{jnk}}}{\frac{2^{jn}}{n} \sum_{i=1}^n 1_{X_{(i-1)/n} \in I_{jnk}}}.$$

Therefore, from the definition of \mathcal{L}_n^x given in (1.2) and the choice of $j_n = \lfloor \frac{1}{3} \log n \vee j_0 \rfloor$, we have under $\tilde{P}_{\sigma^2}^{n,v}$:

$$\tilde{E}_{\sigma^2}^{v,n} \int_{I_{j_n k}} (\sigma^2(x) - T_n^*(x))^2 dx \leq v^{-1} \int_{I_{j_n k}} \tilde{E}_{\sigma^2}^{v,n} \left(\frac{2^{j_n}}{n} \sum_{i=1}^n 1_{X_{(i-1)/n} \in I_{j_n k}} [\sigma^2(x) - (\Delta_i^n X)^2] \right)^2.$$

Recall from Section 2.2 and Eq. (2.11) that under $\tilde{P}_{\sigma^2}^{n,v}$, we have the following decomposition: $(\Delta_i^n X)^2 = n \int_{(i-1)/n}^{i/n} \sigma^2(X_s) ds + \varepsilon_i^n$, where the martingale term is defined by $\varepsilon_i^n = 2n \int_{(i-1)/n}^{i/n} (X_s - X_{(i-1)/n}) \sigma(X_s) dW_s$. Recall from (4.12) and (4.14) that for $x \in I_{j_n k}$:

$$|\tau_i^n(x, X)| 1_{X_{(i-1)/n} \in I_{j_n k}} \leq L \left(2^{-j_n} + n \int_{(i-1)/n}^{i/n} |X_s - X_{(i-1)/n}| ds \right),$$

where $\tau_i^n(x, X)$ is defined in (4.12). It follows that

$$\tilde{E}_{\sigma^2}^{v,n} \left\{ \int_I (\sigma^2(x) - T_n^*(x))^2 dx \right\} \leq C(L, v) (A_1^n + A_2^n + A_3^n),$$

where

$$\begin{aligned} A_1^n &= \sum_{k: I_{j_n k} \subset I_{j_n}} \int_{I_{j_n k}} \tilde{E}_{\sigma^2}^{n,v} \left\{ \left(\frac{1}{n} \sum_{i=1}^n 1_{X_{(i-1)/n} \in I_{j_n k}} \right)^2 \right\} dx, \\ A_2^n &= \sum_{k: I_{j_n k} \subset I_{j_n}} \int_{I_{j_n k}} 2^{2j_n} \tilde{E}_{\sigma^2}^{n,v} \left\{ \left(\sum_{i=1}^n \int_{(i-1)/n}^{i/n} |X_s - X_{(i-1)/n}| 1_{X_{(i-1)/n} \in I_{j_n k}} \right)^2 \right\} dx, \\ A_3^n &= \sum_{k: I_{j_n k} \subset I_{j_n}} \int_{I_{j_n k}} \tilde{E}_{\sigma^2}^{n,v} \left\{ \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i^n 1_{X_{(i-1)/n} \in I_{j_n k}} \right)^2 \right\} dx. \end{aligned}$$

(b) Let $2 \leq i < j \leq n$. The random variable $X_{(i-1)/n}$ has a density $p_{(i-1)/n}^\sigma(x)$ w.r.t. the Lebesgue measure under \tilde{P}_{σ^2} which satisfies $p_{(i-1)/n}^\sigma(x) \leq C(L, c) \sqrt{n/(i-1)}$. (See (4.21) in Section 4.3.2. below for an explicit form of p_i^σ .) Thus, by use of the Markov property, we have

$$\tilde{P}_{\sigma^2}^{n,v} \{ X_{(i-1)/n} \in I_{j_n k}, X_{(j-1)/n} \in I_{j_n k} \} \leq C(c, L, v) 2^{-2j_n} \frac{n}{\sqrt{ij}}.$$

Expanding the sum within the expectation, it follows that A_1^n is less than

$$|I| n^{-1} + C(c, L, v) 2^{-2j_n} \sum_{2 \leq i < j \leq n} [n/(i-1)(j-1)]^{1/2} \leq C(c, L, v) n^{-2/3}$$

from the choice of j_n , thus A_1^n has the right order. (One easily checks that expanding the sum from 2 to n instead of 1 to n does not alter the order of magnitude of A_1^n .)

(c) Let $2 \leq i < j \leq n$ and $(u, s) \in [\frac{i-1}{n}, \frac{i}{n}] \times [\frac{j-1}{n}, \frac{j}{n}]$. By repeated use of the Markov property, using the same argument on the density of $X_{(i-1)/n}$ as in (b) and the BDG inequality, we obtain

$$\begin{aligned} & \tilde{E}_{\sigma^2}^{n,v} \{ |X_s - X_{(i-1)/n}| 1_{X_{(i-1)/n} \in I_{jk}} |X_u - X_{(j-1)/n}| 1_{X_{(j-1)/n} \in I_{jk}} \} \\ & \leq C(c, L, v) \frac{2^{-2j_n}}{\sqrt{(i-1)(j-1)}}. \end{aligned}$$

Thus, by expanding the sum within the expectation in A_2^n , we readily get $A_2^n \leq C(c, L, v)n^{-1}$ which is asymptotically negligible.

(d) Finally, we have $\tilde{E}_{\sigma^2}^{n,v} \{(\varepsilon_i^n)^2\} \leq C(c, L)$ from the BDG inequality; using that the ε_i^n are martingale increments, it is easily seen that $A_3^n \leq C(c, L, v)2^{j_n}n^{-1} \leq C(c, L, v)n^{-2/3}$ from the choice of j_n . The proof of the upper bound is complete.

4.3. Proof of Proposition 2: lower bound

For technical convenience and with no loss of generality, we prove the lower bound for the new parametrization $\sigma^2 \mapsto 1/\sigma^2$. Indeed, for any normalizing factor $z_n \rightarrow 0$, the infimum in $\inf_{T_n} \mathbf{R}_n(T_n, \Sigma, z_n)$ is clearly attained among the estimators T_n such that $T_n \in \Sigma$. Thus:

$$\forall x \in \mathbf{R}: |T_n(x) - \sigma^2(x)| \geq \frac{1}{c^2} |T_n^{-1}(x) - \sigma(x)^{-2}|$$

and

$$\inf_{T_n} \mathbf{R}_n(T_n, \Sigma, z_n) \geq c^{-4} \inf_{T_n} \sup_{\sigma^2 \in \Sigma} E_{\sigma^2}^{n,v} \left\{ z_n^{-2} \left\| T_n - \frac{1}{\sigma^2} \right\|_I^2 \right\}. \tag{4.18}$$

Therefore, it is sufficient to prove a lower bound for the RHS of (4.18). We further assume for simplicity that $I = [0, 1]$. Let $\psi : \mathbf{R} \rightarrow \mathbf{R}$ be of class \mathcal{C}^4 , with $\sup_x |\psi'(x)| \leq L/c^2$ and bounded derivative up to order 4, with support in $[0, 1]$. For j_n and $\gamma_n > 0$, we consider the following parametric subfamily:

$$\mathcal{C}_{j_n, \gamma_n} = \left\{ \sigma^2 \in \Sigma_c(L): \frac{1}{\sigma^2(x)} = \frac{2}{c} + \gamma_n \sum_{k=0}^{2j_n} v_k \psi_{j_n k}(x), v_k = \pm 1 \right\}$$

so that $\# \mathcal{C}_{j_n, \gamma_n} = 2^{2j_n}$ and the functions $\psi_{j_n k}$ and $\psi_{j_n k'}$ have disjoint support for $k \neq k'$. We impose that $\tau = \sup_x |\psi(x)|$ is such that $2c - \tau \geq c$ and $\frac{2}{c} + \tau \leq c^{-1}$. We take $j_n = \lfloor \frac{1}{3} \log_2 n \rfloor$ and $\gamma_n = n^{-1/2}$. These conditions imply

$$\mathcal{C}_{j_n, \gamma_n} \subset \Sigma_c(L).$$

4.3.1. First reduction

For $\sigma^2 \in \mathcal{C}_{j_n, \gamma_n}$, denote by Q_{σ^2} the law of

$$dX_t = \frac{1}{2} \sigma \sigma'(X_t) dt + \sigma(X_t) dW_t \tag{4.19}$$

and define $Q_{\sigma^2}^{n,v}$ analogously. We will also use the notation $Q_{\sigma^2}^n$ for the law of $X^n = (X_{i/n}, i = 1, \dots, n)$ on \mathbf{R}^n where X solves (4.25). For $k = 1, \dots, 2^{j_n}$, denote by $\sigma_{k,+}^2$ and $\sigma_{k,-}^2$ any pair of functions in $\mathcal{C}_{j_n, \gamma_n}$ such that $\sigma_{k,+}^{-2}(x) - \sigma_{k,-}^{-2}(x) = 2\gamma_n \psi_{j_n k}(x)$. Set

$$\lambda(\sigma_{k,+}^2, \sigma_{k,-}^2, X^{(n)}) = \frac{dQ_{\sigma_{k,+}^2}^n}{dQ_{\sigma_{k,-}^2}^n}.$$

LEMMA 4. – *Under the assumptions of Proposition 2, the following condition implies the lower bound of Proposition 2:*

$$\limsup_{n \rightarrow \infty} \sup_k E_{Q_{\sigma_{k,-}^2}^n} (|\log \lambda(\sigma_{k,+}^2, \sigma_{k,-}^2, X^{(n)})|) < \infty. \tag{4.20}$$

Proof. – We abbreviate $\lambda(\sigma_{k,+}^2, \sigma_{k,-}^2, X^{(n)})$ by λ and set $u = P_{\sigma_{k,-}^2}^n (\inf_{x \in I} \mathcal{L}_n^x \geq v)$. Let $t > 0$. Clearly

$$P_{\sigma_{k,-}^2}^{n,v} (\lambda \geq e^{-t}) \geq P_{\sigma_{k,-}^2}^n (\lambda \geq e^{-t}) + u - 1 \geq u - \frac{1}{t} E_{Q_{\sigma_{k,-}^2}^n} (|\log \lambda(\sigma_{k,+}^2, \sigma_{k,-}^2, X^{(n)})|),$$

where we used Chebyshev’s inequality. By taking t large enough, (4.20) implies for large enough n

$$P_{\sigma_{k,-}^2}^{n,v} (\lambda \geq e^{-t}) \geq s > 0,$$

where s does not depend on n . This implies the lower bound $2^{j_n/2} \gamma_n = n^{-1/3}$ for the risk \mathbf{R}_n , as follows from [10, Chapter 2]. □

4.3.2. Proof of (4.20)

Preliminary decomposition. Let $S_\sigma(x) = \int_0^x \frac{du}{\sigma(u)}$. If $\sigma^2 \in \mathcal{C}_{j_n, \gamma_n}$, we have by Ito’s formula

$$X_t = S_\sigma^{-1}(S_\sigma(x_0) + W_t),$$

therefore, the transition density of X_t under Q_{σ^2} reads for $t > 0$

$$p_t^\sigma(x, y) = \frac{1}{\sqrt{2\pi\sigma(y)t}} \exp -\frac{1}{2t} [S_\sigma(y) - S_\sigma(x)]^2. \tag{4.21}$$

Elementary computations yield, under $Q_{\sigma_{k,-}^2}$

$$\log \lambda = - \sum_{i=1}^n \left\{ \log \frac{\sigma_{k,-}}{\sigma_{k,+}}(X_{i/n}) + \frac{n}{2} [(\delta_i^n S_{\sigma_{k,+}}(X))^2 - (\delta_i^n S_{\sigma_{k,-}}(X))^2] \right\},$$

where we denote $\delta_i^n f(X) = f(X_{i/n}) - f(X_{(i-1)/n})$. For a generic σ^2 in $\mathcal{C}_{j_n, \gamma_n}$, define $H_{j_n k}$ by the formula

$$\frac{1}{\sigma^2(x)} = \frac{2}{c} + \gamma_n v_k \psi_{j_n k}(x) + \gamma_n H_{j_n k}(x).$$

We thus have $\frac{1}{\sigma_{k,+}^2(x)} - \frac{1}{\sigma_{k,-}^2(x)} = 2\gamma_n \psi_{j_n k}(x)$ and

$$\frac{1}{\sigma_{k,+}^2(x)} + \frac{1}{\sigma_{k,-}^2(x)} = \frac{4}{c} + \gamma_n H_{j_n k}(x).$$

It follows that

$$\begin{aligned} & (\delta_i^n S_{\sigma_{k,+}}(X))^2 - (\delta_i^n S_{\sigma_{k,-}}(X))^2 \\ &= \frac{8\gamma_n}{c} (\delta_i^n X) \int_{X_{(i-1)/n}}^{X_{i/n}} \psi_{j_n k}(u) du + 2\gamma_n^2 \int_{X_{(i-1)/n}}^{X_{i/n}} \psi_{j_n k}(u) du \int_{X_{(i-1)/n}}^{X_{i/n}} H_{j_n k}(u) du. \end{aligned} \tag{4.22}$$

By Taylor’s formula, summing in n , we have from the decomposition of $\log \lambda$:

$$\begin{aligned} & \frac{n}{2} \sum_{i=1}^n \frac{8\gamma_n}{c} (\delta_i^n X) \int_{X_{(i-1)/n}}^{X_{i/n}} \psi_{j_n k}(u) du \\ &= 4\sqrt{n} \sum_{p=1}^4 \frac{2^{(p-\frac{1}{2})j_n}}{(p+1)!} \sum_{i=1}^n \psi^{(p-1)}(2^{j_n} X_{(i-1)/n} - k) (\delta_i^n X)^{p+1} \\ & \quad + \frac{\sqrt{n}}{30} 2^{9j_n/2} \sum_{i=1}^n \psi^{(4)}(2^{j_n} \xi_{X_{(i-1)/n}, X_{i/n}} - k) (\delta_i^n X)^6, \end{aligned} \tag{4.23}$$

where $\xi_{X_{(i-1)/n}, X_{i/n}} \in [X_{(i-1)/n}, X_{i/n}]$.

Remainder terms. Let us first show that the terms of order 2 and more in (4.23) have finite moments, uniformly in n and k . Up to a drift translation, we may assume that X is a local martingale with diffusion coefficient $\sigma_{k,-}$. (This only amounts to a change of probability – further denoted by $\tilde{P}_{\sigma_{k,-}}$ – which is sensitive to constants only.)

(a) We first study the term corresponding to $p = 2$. By a standard martingale argument its variance is less than $C(c, L) 2^{3j_n} n \sup_x |\psi'(x)| \times \sum_{i=1}^n \tilde{E}_{\sigma_{k,-}} \{(\delta_i^n X)^6\}$. Using the BDG inequality the properties of ψ and the fact that $\sigma^2 \in \Sigma_c(L)$, we see that the term within the expectation is of order n^{-3} . This, together with the choice of j_n shows that the above variance is less than $C(c, L)$.

(b) Let us denote by A_n the term corresponding to $p = 3$ in (4.23). We have

$$\tilde{E}_{\sigma_{k,-}} \{|A_n| \mid \mathcal{F}_i^n\} \leq C(c, L) n \sum_{i=1}^n 2^{j_n} |\psi''| (2^{j_n} X_{(i-1)/n} - k) \tilde{E}_{\sigma_{k,-}} \{(\delta_i^n X)^4 \mid \mathcal{F}_i^n\}.$$

Applying again the BDG inequality, the term within the conditional expectation is of order n^{-2} . Therefore

$$\tilde{E}_{\sigma_{k,-}}^2 (|A_n|) \leq C(c, L) n^{-1} \sum_{i=1}^n 2^{j_n} \tilde{P}_{\sigma_{k,-}} \{X_{(i-1)/n} \in I_{j_n k}\}.$$

Next, we remark that if $x(I_{jnk})$ denotes the midpoint of I_{jnk} , then the above term is equal to $C(c, L) \tilde{E}_{\sigma_{k,-}}(\mathcal{L}_n^{x(I_{jnk})})$, which is uniformly bounded (see, e.g., [8]).

(c) In the same line as in (a), the variance of the term corresponding to $p = 4$ is less than $C(c, L) 2^{7j_n} n \sum_{i=1}^n \tilde{E}_{\sigma_{k,-}}^2 \{(\delta_i^n X)^{10}\}$ and is asymptotically negligible from the choice of j_n .

(d) The remainder term has a first moment bounded by $C(c, L) \sqrt{n} 2^{9j_n/2} \sum_{i=1}^n \tilde{E}_{\sigma_{k,-}}^2 \{(\delta_i^n X)^6\}$, which has the right order.

(e) We now bound the second term in the decomposition (4.22), that is, after summing in n ,

$$Q_n = \sum_{i=1}^n \int_{X_{(i-1)/n}}^{X_{i/n}} \psi_{jnk}(u) du \int_{X_{(i-1)/n}}^{X_{i/n}} H_{jnk}(u) du.$$

By Taylor’s formula, we successively have

$$\int_{X_{(i-1)/n}}^{X_{i/n}} \psi_{jnk}(u) du = \psi_{jnk}(X_{(i-1)/n}) \delta_i^n X + \frac{1}{2} 2^{3j_n/2} (\psi)'(2^{j_n} \xi_{X_{(i-1)/n}, X_{i/n}} - k) (\delta_i^n X)^2$$

and

$$\int_{X_{(i-1)/n}}^{X_{i/n}} H_{jnk}(u) du = H_{jnk}(X_{(i-1)/n}) \delta_i^n X + \frac{1}{2} (H_{jnk})'(\zeta_{X_{(i-1)/n}, X_{i/n}}) (\delta_i^n X)^2$$

with $\xi_{X_{(i-1)/n}, X_{i/n}}$ and $\zeta_{X_{(i-1)/n}, X_{i/n}}$ in $[X_{(i-1)/n}, X_{i/n}]$, therefore, since ψ_{jnk} and H_{jnk} have disjoint support, it suffices to bound the cross-terms in the product of the two above expansions. We first need an auxiliary result.

LEMMA 5. – *Let g be a Borel function with compact support in $[0, 1]$. For $\xi_{X_{(i-1)/n}, X_{i/n}} \in [X_{(i-1)/n}, X_{i/n}]$, we have*

$$|g(2^{j_n} \xi_{X_{(i-1)/n}, X_{i/n}} - k)| \leq \sup_x |g(x)| (1_{\{X_{(i-1)/n} \in I_{jnk}\}} + 1_{\{X_{(i-1)/n} \in I_{jnk}\}} + 1_{\kappa_n}),$$

where $\tilde{P}_{\sigma_{k,-}}(\kappa_n) \leq C_k(c, L) n^{-k/6}$ for all $k \geq 0$.

Proof. – Clearly

$$(\xi_{X_{(i-1)/n}, X_{i/n}} \in I_{jnk}) \subset (X_{(i-1)/n} \in I_{jnk}) \cup (X_{i/n} \in I_{jnk}) \cup (|\delta_i^n X| \geq 2^{-j_n}).$$

Thus we have the announced inequality with $\kappa_n = (|\delta_i^n X| \geq 2^{-j_n})$. By Chebyshev’s inequality, for $r \geq 0$

$$\tilde{P}_{\sigma_{k,-}}(\tau_n) \leq 2^{jnk} \tilde{E}_{\sigma_{k,-}}(|\delta_i^n X|^k) \leq C(k, c, L) n^{k(\frac{1}{3} - \frac{1}{2})},$$

where we used the BDG inequality. \square

(e1) We first bound

$$\sum_{i=1}^n H_{jnk}(X_{i-1}/n) 2^{3jn/2} \psi'(2^{jn} \xi_{X_{(i-1)/n}, X_{i/n}} - k) (\delta_i^n X)^3.$$

Applying Lemma 5 with $g = \psi'$, by taking k large enough, it suffices to bound

$$2^{3jn/2} \sum_{i=1}^n (\delta_i^n X)^3 1_{X_{(l/n)} \in I_{jnk}}, \quad l = i - 1, i.$$

By combining Hölder and BDG inequalities, the expectation of the last term is less than

$$C(\beta, c, L) n^{\frac{1}{6}-1} 2^{-jn/\beta} \sum_{i=2}^n (n/l)^{1/2\beta}$$

for $\beta > 1$, where we used that $\tilde{P}_{\sigma_{k,-}}(X_{l/n} \in I_{jnk}) \leq \sqrt{\frac{n}{l}} 2^{-jn}$. For $\beta < 2$, this last quantity is asymptotically negligible.

(e2) The two other terms

$$\sum_{i=1}^n (H_{jnk})'(\zeta_{X_{(i-1)/n}, X_{i/n}}) \psi_{jnk}(X_{(i-1)/n}) (\delta_i^n X)^3$$

and $\sum_{i=1}^n (H_{jnk})'(\zeta_{X_{(i-1)/n}, X_{i/n}}) 2^{3jn/2} \psi'(2^{jn} \xi_{X_{(i-1)/n}, X_{i/n}} - k) (\delta_i^n X)^4$ can be bound in the same line as in (e1) so we omit them.

Completion of proof. It remains to bound the main term, namely

$$\sum_{i=1}^n \left\{ \log \frac{\sigma_{k,-}(X_{i/n})}{\sigma_{k,+}} + n \gamma_n \psi_{jnk}(X_{(i-1)/n}) (\delta_i^n X)^2 \right\}.$$

Clearly, the order of magnitude is not altered if we replace $\log \frac{\sigma_{k,-}(X_{i/n})}{\sigma_{k,+}}$ by $\log \frac{\sigma_{k,-}(X_{(i-1)/n})}{\sigma_{k,+}}$, as we shall do for technical convenience.

LEMMA 6. – *The following expansion holds*

$$\log \frac{\sigma_{k,-}(x) - \frac{2}{c} + 1}{\sigma_{k,+}(x) - \frac{2}{c} + 1} = -2\gamma_n \frac{\psi_{jnk}(x)}{1 + \gamma_n H_{jnk}(x)} - \frac{2}{3} \frac{\gamma_n^3 \psi_{jnk}^3(\zeta_x)}{[1 + \gamma_n H_{jnk}(\zeta_x)]^3}$$

for some $\zeta_x \in [0, 1]$.

Proof. – Elementary by a Taylor’s expansion. \square

For technical convenience, we further assume that $c = 2$, with no loss of generality. The general case is obtained by a modification of the constants. We thus need a bound for

$$\sum_{i=1}^n \left\{ -2\gamma_n \frac{\psi_{jnk}(X_{(i-1)/n})}{1 + \gamma_n H_{jnk}(X_{(i-1)/n})} + 2n\gamma_n \psi_{jnk}(X_{(i-1)/n}) (\Delta_i^n X)^2 \right\} - \frac{2\gamma_n^3}{3} \sum_{i=1}^n \frac{\psi_{jnk}^3(\xi_{X_{(i-1)/n}})}{[1 + \gamma_n H_{jnk}(\xi_{X_{(i-1)/n}})]^3}. \tag{4.24}$$

Clearly, the second term has its first order moment bounded thanks to the choice of j_n and γ_n .

Up to a change of probability, we may again assume that X is a local martingale with diffusion coefficient $\sigma_{k,-}^2$. Therefore, by a Taylor’s expansion, the first term in (4.24) splits into four terms:

$$\begin{aligned} A_1^n &= -4\gamma_n \sum_{i=1}^n \psi_{jnk}(X_{(i-1)/n}) n \int_{(i-1)/n}^{i/n} (X_s - X_{(i-1)/n}) \sigma_{k,-}(X_s) dW_s, \\ A_2^n &= -2\gamma_n^2 \sum_{i=1}^n \psi_{jnk}(X_{(i-1)/n}) \left[H_{jnk}(X_{(i-1)/n}) - n \int_{(i-1)/n}^{i/n} H_{jnk}(X_s) ds \right], \\ A_3^n &= -2\gamma_n^3 \sum_{i=1}^n \psi_{jnk}(X_{(i-1)/n}) \mathcal{O}(\sup_x H_{jnk}^2(x)), \\ A_4^n &= \gamma_n^2 \sum_{i=1}^n \left[\psi_{jnk}(X_{(i-1)/n}) - n \int_{(i-1)/n}^{i/n} \psi_{jnk}(X_s) ds \right]. \end{aligned}$$

The expectation of the four terms is bounded using the same techniques: A_1^n is bounded by a martingale argument; for A_2^n , we use the fact that ψ_{jnk} and H_{jnk} have disjoint supports; A_3^n and A_4^n are bounded using the same straightforward arguments. The proof of Proposition 2 is complete.

4.4. Proof of Theorem 1: upper bound

With no loss of generality, we prove the upper bound of Theorem 1 under the change of probability $P_{\sigma^2}^{n,\nu} \rightarrow \tilde{P}_{\sigma^2}^{n,\nu}$ under which X is a local martingale. This only amounts to a modification of the constants. Let

$$\begin{aligned} \mathbf{R}_n^{(1)}(\sigma^2) &= \tilde{E}_{\sigma^2}^{n,\nu} \{ n^{2/3} \|T_n^* - \sigma^2\|_I^2 1_{\rho_n^* = n^{-1/3}} \}, \\ \mathbf{R}_n^{(2)}(\sigma^2) &= \tilde{E}_{\sigma^2}^{n,\nu} \{ \varphi_{n,\alpha_n}^{-2} \|\bar{T}_n - \sigma^2\|_I^2 1_{\rho_n^* = \varphi_{n,\alpha_n}} \}. \end{aligned}$$

By Proposition 2, since T_n^* is optimal w.r.t. Σ , we have

$$\mathbf{R}_n(T_n^{**}, \Sigma, \rho_n^*) \leq C(c, L, \nu) + \sup_{\sigma^2 \in \Sigma} \mathbf{R}_n^{(2)}(\sigma^2).$$

Define

$$\mathbf{R}_n^{(3)}(\sigma^2) = \tilde{E}_{\sigma^2}^{n,\nu} \{ \varphi_{n,\alpha_n}^{-2} \Delta_n(\sigma^2) 1_{\{\rho_n^* = \varphi_{n,\alpha_n}\}} \}.$$

Using (a) and (b) of the preliminary decompositions, we have

$$\sup_{\sigma^2 \in \Sigma} \mathbf{R}_n^{(2)}(\sigma^2) \leq C(c, L, \nu) + 2 \sup_{\sigma^2 \in \Sigma} \mathbf{R}_n^{(3)}(\sigma^2).$$

Let $q, \delta > 0, 0 < u < 1$ and $k_n = \lfloor q \frac{\log n}{\log u} \rfloor$. Put $t_k^n = (\sqrt{\frac{k \log u}{\log \alpha_n}} + 1)$. We introduce the following partition of Ω :

$$\Gamma_0^n = \{(1 - \delta)\Delta_n(\sigma^2) \leq t_1^n\}, \quad \Gamma_{k_n}^n = \{(1 - \delta)\Delta_n(\sigma^2) \geq t_{k_n}^n\},$$

$$\Gamma_k^n = \{t_k^n \leq (1 - \delta)\Delta_n(\sigma^2) < t_{k+1}^n\}, \quad k = 1, \dots, k_n - 1.$$

Let also

$$D_n = \{|M_n(\sigma^2)| \leq \delta \Delta_n(\sigma^2)\}.$$

We thus obtain the decomposition $\mathbf{R}_n^{(3)}(\sigma^2) = \sum_{i=1}^4 \mathbf{R}_n^{(3,i)}(\sigma^2)$, with

$$\mathbf{R}_n^{(3,1)}(\sigma^2) = \tilde{E}_{\sigma^2}^{n,\nu} \{ \varphi_{n,\alpha_n}^{-2} \Delta_n(\sigma^2) 1_{\{\rho_n^* = \varphi_{n,\alpha_n}; \Gamma_0^n\}} \},$$

$$\mathbf{R}_n^{(3,2)}(\sigma^2) = \sum_{k=1}^{k_n-1} \tilde{E}_{\sigma^2}^{n,\nu} \{ \varphi_{n,\alpha_n}^{-2} \Delta_n(\sigma^2) 1_{\{\rho_n^* = \varphi_{n,\alpha_n}; \Gamma_k^n; D_n\}} \},$$

$$\mathbf{R}_n^{(3,3)}(\sigma^2) = \tilde{E}_{\sigma^2}^{n,\nu} \{ \varphi_{n,\alpha_n}^{-2} \Delta_n(\sigma^2) 1_{\{\rho_n^* = \varphi_{n,\alpha_n}; \Gamma_{k_n}^n; D_n\}} \},$$

$$\mathbf{R}_n^{(3,4)}(\sigma^2) = \tilde{E}_{\sigma^2}^{n,\nu} \{ \varphi_{n,\alpha_n}^{-2} \Delta_n(\sigma^2) 1_{\{\rho_n^* = \varphi_{n,\alpha_n}; \bigcup_{k=1}^{k_n} \Gamma_k^n; (D_n)^c\}} \}.$$

(a) Let us first study $\mathbf{R}_n^{(3,1)}(\sigma^2)$. Clearly $\mathbf{R}_n^{(3,1)}(\sigma^2) \leq (\frac{\log u}{\log \alpha_n})^{1/2} \frac{1}{1-\delta} \lambda^2$. Therefore, letting $\delta \rightarrow 0$, we obtain the bound

$$\lim_{n \rightarrow \infty} \sup_{\sigma^2 \in \Sigma} \mathbf{R}_n^{(3,1)}(\sigma^2) \leq \left(\frac{\log u}{\log \alpha} \right)^{1/2} \lambda^2,$$

where $\alpha = \liminf_{n \rightarrow \infty} \alpha_n$. ($1/\log 0 = 0$.)

(b) We next turn to $\mathbf{R}_n^{(3,4)}(\sigma^2)$. Since $|\Delta_n(\sigma^2)| \leq 8c^2$, by (c) of the preliminary decompositions and the definition of D_n , we have

$$\begin{aligned} \mathbf{R}_n^{(3,4)}(\sigma^2) &\leq 8c^2 \varphi_{n,\alpha_n}^{-2} \tilde{P}_{\sigma^2}^{n,\nu} \{ \Delta_n(\sigma^2) + M_n(\sigma^2) + N_n(\sigma^2) \leq \lambda^2 \varphi_{n,\alpha_n}^2; (\Gamma_0^n)^c; (D_n)^c \} \\ &\leq C(c, L, \nu) \varphi_{n,\alpha_n}^{-2} \tilde{P}_{\sigma^2}^{n,\nu} \left\{ |M_n(\sigma^2)| \geq \delta \Delta_n(\sigma^2); \Delta_n(\sigma^2) \geq \frac{t_1^n}{1-\delta} \right\} \\ &\leq C(c, L, \nu) \varphi_{n,\alpha_n}^{-2} \tilde{P}_{\sigma^2}^{n,\nu} \left\{ |M_n(\sigma^2)| \geq \delta \Delta_n(\sigma^2); \Delta_n(\sigma^2) \geq \frac{C(u)}{1-\delta} \lambda^2 \varphi_{n,\alpha_n}^2 \right\}, \end{aligned}$$

where $C(u) = \sqrt{\frac{\log u}{\log \alpha}}$. We now need the following fundamental technical result, proof of which we delay until Appendix A.

LEMMA 7. – Let $\delta_1, \delta_2 > 0$. We have

$$\tilde{P}_{\sigma^2}^{n,\nu} \{ |M_n(\sigma^2)| \geq \delta_1 \Delta_n(\sigma^2); \Delta_n(\sigma^2) \geq \delta_2 \varphi_{n,\alpha_n}^2 \} = o(\varphi_{n,\alpha_n}^2)$$

uniformly in $\sigma^2 \in \Sigma$ as $n \rightarrow \infty$.

Applying Lemma 7 with $\delta_1 = \delta$, $\delta_2 = \frac{C(u)}{1-\delta}\lambda^2$, we obtain

$$\lim_{n \rightarrow \infty} \sup_{\sigma^2 \in \Sigma} \mathbf{R}_n^{(3,4)}(\sigma^2) = 0.$$

(c) We now study $\mathbf{R}_n^{(3,3)}(\sigma^2)$. Analogously

$$\mathbf{R}_n^{(3,4)}(\sigma^2) \leq C(c, L, \nu) \varphi_{n,\alpha_n}^{-2} \tilde{P}_{\sigma^2}^{n,\nu} \{ \hat{d}_n(\theta_n^*) \leq \lambda^2 \varphi_{n,\alpha_n}^2; \Gamma_{k_n}^n; D_n \}.$$

The following inclusion holds

$$\begin{aligned} & \{ \hat{d}_n(\theta_n^*) \leq \lambda^2 \varphi_{n,\alpha_n}^2; \Gamma_{k_n}^n; D_n \} \\ &= \{ \Delta_n(\sigma^2) + M_n(\sigma^2) + N_n(\sigma^2) \leq \lambda^2 \varphi_{n,\alpha_n}^2 \} \cap \{ (1 - \delta) \Delta_n(\sigma^2) \geq t_{k_n}^n \} \\ & \quad \cap \{ |M_n(\sigma^2)| \leq \delta \Delta_n(\sigma^2) \} \\ & \subseteq \left\{ N_n(\sigma^2) \leq \sqrt{\frac{q \log n}{\log \alpha_n}} \lambda^2 \varphi_{n,\alpha_n}^2 \right\}. \end{aligned}$$

It remains to bound $P_{\sigma^2}^{n,\nu} \{ |N_n^{(i)}| \geq \frac{1}{2} \sqrt{\frac{q \log n}{\log \alpha_n}} \lambda^2 \varphi_{n,\alpha_n}^2 \}$, $i = 1, 2$.

We plan to use Lemma 1 to bound the first term. For $t \geq 2$, we have

$$\tilde{P}_{\sigma^2}^{n,\nu} \left\{ |N_n^{(1)}| \geq \frac{1}{2} \sqrt{\frac{q \log n}{\log \alpha_n}} \lambda^2 \varphi_{n,\alpha_n}^2 \right\} \leq C(t, c, L, \nu) n^{-3t/2} 2^{J_n t} \varphi_{n,\alpha_n}^{-2t}.$$

By taking t large enough, we see that this term is $o(\varphi_{n,\alpha_n}^2)$, uniformly in $\sigma^2 \in \Sigma$ as $n \rightarrow \infty$ and is thus asymptotically negligible.

Likewise, we plan to use Lemma 3 for the second term. For $t \geq 2$, we have

$$\begin{aligned} & \tilde{P}_{\sigma^2}^{n,\nu} \{ |N_n^{(2)}| \geq \frac{1}{2} \sqrt{\frac{q \log n}{\log \alpha_n}} \lambda^2 \varphi_{n,\alpha_n}^2 \} \\ & \leq C(t, c, L, \nu) n^{1-3t/2} 2^{J_n t} \varphi_{n,\alpha_n}^{-2t} + \exp \left\{ -A(t, c, \nu) n^2 2^{-J_n} \lambda^4 \varphi_{n,\alpha_n}^4 q \frac{\log n}{\log \alpha_n} \right\}. \end{aligned}$$

Taking $t = 10$ and using the definition of λ , the last quantity is less than

$$o(\varphi_{n,\alpha_n}^2) + n^{-q}$$

as $n \rightarrow \infty$, uniformly in $\sigma^2 \in \Sigma$. Since q is free, the above term is asymptotically negligible and

$$\lim_{n \rightarrow \infty} \sup_{\sigma^2 \in \Sigma} \mathbf{R}_n^{(3,3)}(\sigma^2) = 0.$$

(d) We eventually turn to $\mathbf{R}_n^{(3,2)}(\sigma^2)$. Let $k \in \{1, \dots, k_n - 1\}$. The following inclusion holds

$$\begin{aligned} & \{ \hat{d}_n(\theta_n^*) \leq \lambda^2 \varphi_{n,\alpha_n}^2; \Gamma_k^n; D_n \} \subseteq \{ N_n(\sigma^2) \leq -t_k^n + \lambda^2 \varphi_{n,\alpha_n}^2; (1 - \delta) \Delta_n(\sigma^2) \leq t_{k+1}^n \} \\ & = \left\{ N_n(\sigma^2) \leq -\lambda^2 \varphi_{n,\alpha_n}^2 \sqrt{\frac{k \log u}{\log \alpha_n}}; (1 - \delta) \Delta_n(\sigma^2) \leq t_{k+1}^n \right\}. \end{aligned}$$

It follows that

$$\mathbf{R}_n^{(3,2)}(\sigma^2) \leq \varphi_{n,\alpha_n}^{-2} \sum_{k=1}^{k_n-1} \frac{t_{k+1}^n}{1-\delta} \tilde{P}_{\sigma^2}^{n,v} \left\{ N_n(\sigma^2) \leq -\lambda^2 \varphi_{n,\alpha_n}^2 \sqrt{\frac{k \log u}{\log \alpha_n}}; (1-\delta)\Delta_n(\sigma^2) \leq t_{k+1}^n \right\}.$$

Again, we split $N_n(\sigma^2)$ into $N_n^{(1)} + N_n^{(2)}$ and we successively apply Lemmas 1 and 3. The calculation is done in the same way as for the term in (c) so we omit it. Eventually

$$\mathbf{R}_n^{(3,2)}(\sigma^2) \leq \frac{\lambda^2}{1-\delta} \sum_{k=1}^{k_n-1} \left[1 + \left(\frac{(k+1) \log u}{\log \alpha_n} \right)^{1/2} \right] u^k + r_n(\sigma^2),$$

where $r_n(\sigma^2) = o(1)$ uniformly in $\sigma^2 \in \Sigma$ as $n \rightarrow \infty$. Letting $\delta \rightarrow 0$, we obtain the bound

$$\limsup_{n \rightarrow \infty} \sup_{\sigma^2 \in \Sigma} \mathbf{R}_n^{(3,2)}(\sigma^2) \leq \frac{\lambda^2 \sqrt{\log u}}{\sqrt{\log \alpha}} \sum_{k=1}^{\infty} (k+1)^{1/2} u^k.$$

The proof of Theorem 1, upper bound, is complete.

4.5. Proof of Theorem 1: lower bound

4.5.1. Preliminaries

The same preliminary remark as for the proof of Proposition 2, lower bound, can be applied and we consider the parametrization $\sigma^2 \mapsto 1/\sigma^2$ with no loss of generality. We then follow the arguments developed by Lepski in [11]. For simplicity, we prove the lower bound under the additional restriction that $I = [0, 1]$ and $\sigma_0^2(\theta_0, x) \equiv 1$ in Assumption B4. Take an infinitely many times differentiable function $\psi : \mathbf{R} \rightarrow \mathbf{R}$ with support in $[0, 1]$ and such that $\sup_x |\psi'(x)| \leq L/c^2$ and $\int \psi^2(x) dx = 1$. Consider the parametric subfamily

$$\mathcal{C}_{J_n} = \left\{ \sigma^2 \in \Sigma_c(L) : \frac{1}{\sigma^2(x)} = 1 + 2^{-3J_n/2} \sum_{k=0}^{2^{J_n}} v_k \psi_{J_n k}(x), v_k = \pm 1 \right\}$$

indexed by $v = (v_1, \dots, v_{2^{J_n}})$, so that $\# \mathcal{C}_{J_n} = 2^{2^{J_n}}$ and the functions $\psi_{J_n k}$ and $\psi_{J_n k'}$ have disjoint support for $k \neq k'$. These conditions imply under assumption B4 that for sufficiently large n , we have

$$\mathcal{C}_{J_n} \subset \Sigma_c(L).$$

Recall that for $\sigma^2 \in \mathcal{C}_{J_n}$, we denote by $Q_{\sigma^2}^2$ the law of

$$dX_t = \frac{1}{2} \sigma \sigma'(X_t) dt + \sigma(X_t) dW_t \tag{4.25}$$

and that we define $Q_{\sigma^2}^{n,v}$ analogously. Likewise, we also use the notation $Q_{\sigma^2}^n$ for the law of $X^n = (X_{i/n}, i = 1, \dots, n)$ on \mathbf{R}^n where X solves (4.25). In the following, we

abbreviate $Q_{\sigma_0^2(\cdot, \theta_0)}$ by Q_0 and Q_{σ^2} by Q_v for a generic $\sigma^2 = \sigma^2(v) \in \mathcal{C}_{J_n}$ parametrized by v . If $\mathcal{V}_n = \{v = (v_1, \dots, v_{2^{J_n}}) : v_k = \pm 1\}$, we will use the following sets, for $k = 1, \dots, 2^{J_n}$

$$\mathcal{V}_{n,k}^{(\pm 1)} = \{v \in \mathcal{V}_n : v_k = \pm 1\}, \quad \mathcal{V}_{n,k}^{(0)} = \{v_{k'} = \pm 1, k' \neq k \text{ and } v_k = 0\}.$$

Define further for $v \in \mathcal{V}_n$ the vector $\tau_k(v) = (v_1, \dots, v_{k-1}, 0, v_{k+1}, \dots, v_{2^{J_n}})$ and

$$Z_{n,k}^{(\pm 1)} = \frac{dQ_v^{n,v}}{dQ_{\tau_k(v)}^{n,v}}, \quad \Lambda_n^k = \frac{1}{2^{J_n-1}} \sum_{v \in \mathcal{V}_{n,k}^{(0)}} \frac{dQ_v^{n,v}}{dQ_0^{n,v}}, \quad k = 1, \dots, 2^{J_n}.$$

The lower bound essentially relies on the following lemma.

LEMMA 8. – (i) *There exists $\delta_0 > 0$ such that for $0 < \delta \leq \delta_0$ and $j = 1, \dots, 2^{J_n}$, we have*

$$Q_v^{n,v}(\{Z_{n,j}^{(+1)} < 1 - \delta\} \cap \{Z_{n,j}^{(-1)} < 1 - \delta\}) \leq \delta.$$

(ii) *We have $\limsup_{n \rightarrow \infty} \sup_{1 \leq k \leq 2^{J_n}} \alpha_n^{1/2} Q_0^{n,v} \{(\Lambda_n^k)^2\} \leq 1$, where the notation $Q_{\sigma^2}^{n,v}$ is also used to denote expectation.*

The proof closely follows the techniques developed for the lower bound of Proposition 2, based on expansions of likelihood ratios $dQ_v^{n,v}/dQ_{\tau_k(v)}^{n,v}$ under Q_v^n and dQ_v^n/dQ_0^n under Q_0^n . To simplify the exposition, we omit the proof of Lemma 8 and refer the reader to Proposition 2, lower bounds in Appendix A.

4.5.2. Completion of proof

Let ρ_n be an arbitrary r.n.f. in Ω_n such that $\chi_n(\rho_n)/\chi_n(\rho_n^*) \rightarrow 0$ as $n \rightarrow \infty$. Let T_n be an arbitrary estimator of $1/\sigma^2$ and define

$$\mathcal{B}_n = \{\rho_n = \chi_n(\rho_n)\}.$$

We have

$$\begin{aligned} & \sup_{\sigma^2 \in \Sigma} E_{\sigma^2}^{n,v} \left\{ \rho_n^{-2} \left\| T_n - \frac{1}{\sigma^2} \right\|_I \right\} \\ & \geq \sup_{\sigma^2 \in \Sigma} Q_{\sigma^2}^{n,v} \left\{ \rho_n^{-2} \left\| T_n - \frac{1}{\sigma^2} \right\|_I^2 1_{\mathcal{B}_n} \right\} \\ & \geq \sup_{\sigma^2(v), v \in \mathcal{V}_n} Q_{\sigma^2}^{n,v} \left\{ \chi_n(\rho_n)^{-2} \left\| T_n - \frac{1}{\sigma^2} \right\|_I^2 1_{\mathcal{B}_n} \right\} \\ & \geq \frac{1}{2^{2^{J_n}}} \sum_{v \in \mathcal{V}_n} Q_v^{n,v} \left\{ \chi_n(\rho_n)^{-2} \left\| T_n - \frac{1}{\sigma^2} \right\|_I^2 1_{\mathcal{B}_n} \right\} \\ & \geq \frac{\chi_n(\rho_n)^{-2}}{2^{2^{J_n}}} \sum_k \sum_{v \in \mathcal{V}_n} Q_v^{n,v} \left\{ \int_{I_{n,k}} \left(T_n(x) - \frac{1}{\sigma_v^2(x)} \right)^2 1_{\mathcal{B}_n} dx \right\}. \end{aligned}$$

Setting $D_{n,k} = \{Z_{n,j}^{(+1)} < 1 - \delta\} \cap \{Z_{n,j}^{(-1)} < 1 - \delta\}$, the last quantity is greater than

$$\begin{aligned} & \frac{\chi_n(\rho_n)^{-2}}{2^{2J_n}} \sum_k \sum_{v \in \mathcal{V}_n} Q_{\tau_k(v)}^{n,v} \left\{ 1_{\mathcal{B}_n} \int_{I_{J_n k}} (T_n(x) - 2^{-3J_n/2} \psi_{J_n k}(x))^2 dx \right. \\ & \quad \left. + Z_{n,k}^{-1} \int_{I_{J_n k}} (T_n(x) + 2^{-3J_n/2} \psi_{J_n k}(x))^2 dx \right\} \\ & \geq \frac{(1 - \delta) \chi_n(\rho_n)^{-2}}{2^{2J_n - 1}} \sum_k \sum_{v \in \mathcal{V}_{n,k}^{(0)}} Q_v^{n,v} \{ \mathcal{B}_n \cap D_{n,k} \} \int_{I_{J_n k}} 2^{-3J_n} \psi_{J_n k}^2(x) dx \\ & \geq 2^{-3J_n} \frac{(1 - \delta) \chi_n(\rho_n)^{-2}}{2^{2J_n - 1}} \sum_k \sum_{v \in \mathcal{V}_{n,k}^{(0)}} (Q_v \{ \mathcal{B}_n \} - Q_v \{ (D_{n,k})^c \}). \end{aligned}$$

We claim that there exists $t > 0$ such that

$$P_{n,k} = \frac{1}{2^{2J_n - 1}} \sum_{v \in \mathcal{V}_{n,k}} Q_v^{n,v} \{ \mathcal{B}_n \} \geq t, \quad k = 1, \dots, 2^{J_n}. \tag{4.26}$$

Let us temporarily admit (4.26). Then, taking $\delta < t/2$, by (i) of Lemma 8, we obtain $Q_v^{n,v} \{ (D_{n,k})^c \} < t/2$ for all k and $v \in \mathcal{V}_{n,k}^{(0)}$. From the choice of J_n , it follows that

$$R_n(T_n, \Sigma, \rho_n) \geq \frac{t(1 - \delta)}{2} \chi_n(\rho_n)^{-2} 2^{-2J_n} \geq C \chi_n(\rho_n)^{-2} \varphi_{n,\alpha_n}^2.$$

Showing that $\chi_n(\rho_n)/\varphi_{n,\alpha_n} \rightarrow 0$ as $n \rightarrow \infty$ will complete the proof. For this, it is enough to show that $\chi_n(\rho_n^*) \leq \varphi_{n,\alpha_n}$, or, as follows from Definition 1,

$$\limsup_{n \rightarrow \infty} \alpha_n^{-1} \sup_{\theta \in \Theta} P_\theta^{n,v} \{ \rho_n^* = \varphi_{n,\alpha_n} \} \leq 1. \tag{4.27}$$

Let us prove (4.27). We have

$$P_\theta^{n,v} \{ \rho_n^* = \varphi_{n,\alpha_n} \} = P_\theta^{n,v} \left\{ \inf_{\theta \in \Theta} d_n(\theta) - \sum_k C_{J_n k} \geq \lambda^2 \varphi_{n,\alpha_n} \right\} \leq P_\theta^{n,v} \{ \hat{d}_n(\theta) \geq \lambda^2 \varphi_{n,\alpha_n} \}.$$

Let $\Delta_n(\sigma^2, \theta) = \sum_{k: I_{J_n k} \subset I} c_{J_n k}^2 (\sigma_0^2(\cdot, \theta) - \sigma^2) - a_k^n(\sigma^2)$. In the same line as (a) of the preliminary decompositions, we have $\hat{d}_n(\theta) = \Delta_n(\sigma^2, \theta) + M_n(\sigma_0^2(\cdot, \theta)) + N_n(\sigma_0^2(\cdot, \theta))$. First, $\Delta_n(\sigma^2, \theta) = 0$ under P_θ . Next, given $q > 0$, it suffices then to bound

$$P_\theta^{n,v} \{ M_n(\sigma_0^2(\cdot, \theta)) \geq q \lambda^2 \varphi_{n,\alpha_n} \} + P_\theta^{n,v} \{ N_n(\sigma_0^2(\cdot, \theta)) \geq (1 - q) \lambda^2 \varphi_{n,\alpha_n} \}.$$

By Lemma 7, the first term is asymptotically negligible. Applying Lemma 3 to the second term and letting $q \rightarrow 0$, one easily checks that the choice of λ yields (4.27).

It remains to prove (4.26). Since $\rho_n \in \Omega_n$, taking n large enough, we have $\alpha_n^{-1} Q_0^{n,v} \{ (\mathcal{B}_n)^c \} \leq 1 + \delta$. Therefore

$$P_{n,k} \geq Q_0^{n,v} \{ \Lambda_n^k 1_{\mathcal{B}_n} \} \geq Q_0^{n,v} \{ \Lambda_n^k 1_{\mathcal{B}_n} + c \alpha_n^{-1} 1_{(\mathcal{B}_n)^c} \} - c(1 + \delta)$$

for a given $c > 0$. Since $\inf_{\mathcal{A}}(\Lambda_n^k 1_{\mathcal{A}} + c\alpha_n^{-1} 1_{(\mathcal{A})^c}) = \Lambda_n^k 1_{\{\Lambda_n^k < c\alpha_n^{-1}\}} + c\alpha_n^{-1} 1_{\{\Lambda_n^k \geq c\alpha_n^{-1}\}}$, it follows that

$$\begin{aligned} \mathcal{P}_{n,k} &\geq \mathcal{Q}_0^{n,v} \{ \Lambda_n^k 1_{\{\Lambda_n^k < c\alpha_n^{-1}\}} \} - c(1 + \delta) \\ &= 1 - c(1 + \delta) - \frac{1}{2^{2^{J_n-1}}} \sum_{v \in \mathcal{V}_{n,k}^{(0)}} \mathcal{Q}_v^{n,v} \{ \Lambda_n^k \geq c\alpha_n^{-1} \} \\ &\geq 1 - c(1 + \delta) - \frac{\alpha_n}{c 2^{2^{J_n-1}}} \sum_{v \in \mathcal{V}_{n,k}^{(0)}} \mathcal{Q}_v^{n,v} \{ \Lambda_n^k \} \\ &= 1 - c(1 + \delta) - c^{-1} \alpha_n \mathcal{Q}_0^{n,v} \{ (\Lambda_n^k)^2 \}. \end{aligned}$$

Applying (ii) of Lemma 8, taking $c = \alpha_n^{1/4} / (1 + \delta)^{1/2}$ and using the assumption $\alpha_n < 2^{-4}$, we see that for δ small enough the last term in the above inequality is bounded below, uniformly in n and k . Inequality (4.26) follows and the proof of Theorem 1 is complete.

4.6. Proof of Theorem 2

Define

$$\tilde{\Delta}_n(\sigma^2) = \sum_{k: I_{J_n k} \subset I} \{ c_{J_n k}(\sigma_0^2(\cdot, \hat{\theta}_n) - \sigma^2) - a_k^n(\sigma^2) \}^2$$

which is defined like $\Delta_n(\sigma^2)$ of (4.13) in (a) of the preliminary decompositions in the proof of Theorem 1, replacing θ_n^* by $\hat{\theta}_n$. Next, in the same way as (b) of the preliminary decompositions, we have

$$\|T_n^{(0)} - \sigma^2\|_I^2 = \sum_{k: I_{J_n k} \subset I} c_{J_n k}^2(\sigma^2 - \sigma_0^2(\cdot, \hat{\theta}_n)) + \tilde{\epsilon}_{J_n}(\sigma^2),$$

where

$$\tilde{\epsilon}_{J_n}(\sigma^2) = \sum_{j \geq J_n} \sum_{k: I_{j k} \subset I} d_{j k}^2(\sigma_0^2(\cdot, \hat{\theta}_n) - \sigma^2) \leq C(c, L) 2^{-2J_n}$$

since $\sigma^2 - \sigma_0^2(\cdot, \hat{\theta}_n) \in \Sigma_{2c}(L)$. The proof of the upper bound then readily follows from that of Theorem 1, replacing $\Delta_n(\sigma^2)$ by $\tilde{\Delta}_n(\sigma^2)$. In view of the proof of Theorem 1, the lower bound will follow from

$$\chi_n(\rho_n^{(a)}) \leq \varphi_{n, \alpha_n}. \tag{4.28}$$

We first need an approximation result.

LEMMA 9. – We have, under P_θ

$$\hat{d}_n(\hat{\theta}_n) = \hat{d}_n(\theta) + v_n(\theta), \tag{4.29}$$

where $\lim_{n \rightarrow \infty} \alpha_n^{-1} \varphi_{n, \alpha_n}^{-2} \sup_{\theta \in \Theta} E_\theta^{n,v} \{|v_n(\theta)|\} = 0$.

Proof. – Straightforward computations yield (4.29) under P_θ , having

$$v_n(\theta) = \sum_k c_{J_n k}^2(\sigma_0^2(\cdot, \hat{\theta}_n) - \sigma_0^2(\cdot, \theta)) + 2 \sum_k c_{J_n k}(\sigma_0^2(\cdot, \hat{\theta}_n) - \sigma_0^2(\cdot, \theta)) a_k^n(\sigma_0^2(\cdot, \theta)).$$

Clearly

$$\sum_k c_{J_n k}^2 (\sigma_0^2(\cdot, \hat{\theta}_n) - \sigma_0^2(\cdot, \theta)) \leq \| \sigma_0^2(\cdot, \hat{\theta}_n) - \sigma_0^2(\cdot, \theta) \|_I^2 \leq M^2 |\theta - \hat{\theta}_n|^2$$

by Assumption A2. Integrating w.r.t. $P_\theta^{n,\nu}$ and recalling that $\hat{\theta}_n$ is optimal w.r.t. Θ , this term is of order $n^{-1/2} = o(\alpha_n \varphi_{n,\alpha_n})$. By Cauchy–Schwarz, the same argument together with (4.15) show that after integration w.r.t. $P_\theta^{n,\nu}$, the second term is less than $C(c, L, \nu)n^{-1/2}2^{-J_n}$ and thus has the right order. \square

We are ready to complete the proof. Let $q > 0$. We have

$$P_\theta^{n,\nu} \{ \hat{d}_n(\hat{\theta}_n) \geq \lambda^2 \varphi_{n,\alpha_n}^2 \} \leq P_\theta^{n,\nu} \{ \hat{d}_n(\theta) \geq (1 - q)\lambda^2 \varphi_{n,\alpha_n}^2 \} + P_\theta^{n,\nu} \{ v_n(\theta) \geq q\lambda^2 \varphi_{n,\alpha_n}^2 \}.$$

The second term is asymptotically negligible by Lemma 9. Letting $q \rightarrow 0$, we are back to (4.27) of the lower bound of Theorem 1 and (4.28) follows. The proof of Theorem 2 is complete.

Appendix A. Proof of Lemma 7

We need some notation. For $\theta \in \Theta$ and $\sigma^2 \in \Sigma$, let

$$\xi(\sigma^2, \theta) = 2^{J_n} \sum_{k: I_{J_n k} \subset I} [c_{J_n k}(\sigma^2 - \sigma_0^2(\theta, \cdot)) + a_k^n(\sigma^2)] a_k^n(\sigma^2).$$

Note that since $\sup_x |\tau_i^n(x, X)| \leq 2c$ and $\sup_x |\sigma^2 - \sigma_0^2(\theta, \cdot)| \leq 2c$, we have $|\xi(\sigma^2, \theta)| \leq 4c^2$. For $\xi \in [-4c^2, 4c^2]$, define

$$\Delta_n(\sigma^2, \theta, \xi) = \sum_{k: I_{J_n k} \subset I} c_{J_n k}^2 (\sigma^2 - \sigma_0^2(\theta, \cdot)) + 2^{-J_n} \xi.$$

We thus have the following representation

$$\Delta_n(\sigma^2) = \Delta_n(\sigma^2, \theta_n^*, \xi(\sigma^2, \theta_n^*)). \tag{A.1}$$

By Assumption B, the set Θ is bounded, i.e., $\Theta \subseteq [-P, P]^s$ for some $P > 0$. Let $Q = \max(P, 4c^2)$. Fix some $b > 0$. We approximate $\Theta \times [-4c^2, 4c^2]$ by a n -net constructed as follows.

For $i = 0, \pm 1, \dots, \pm K_n$, $K_n = \lfloor Qn^b \rfloor$, put $p_i = in^{-b}$ and let

$$S_n = \{p_i, i = i = \pm 1, \dots, \pm K_n\}, \quad S_n^{s+1} = S_n \times \dots \times S_n \quad (s + 1 \text{ times.})$$

Finally, define

$$\mathcal{N}_n = \Theta \times [-Q, Q] \cap S_n.$$

Clearly, $\# \mathcal{N}_n \leq (2Qn^b)^{s+1}$. For $x \in [-Q, Q]$, let $u(x)$ be the unique element in S_n such that $u(x) < x \leq u(x) + n^{-b}$. Finally, if $\theta = (\theta_1, \dots, \theta_s)$ denotes a generic

element of Θ and $\xi \in [-Q, Q]$, define $u^{(s+1)}(\theta, \xi) = (u(\theta_1), \dots, u(\theta_s), u(\xi))$. Clearly, $u^{(s+1)}(\theta, \xi) \in \mathcal{N}_n$ and

$$\|u^{s+1}(\theta, \xi) - (\theta, \xi)\| \leq n^b \sqrt{s+1}.$$

Now, let $(\bar{\theta}_n, \bar{\xi}_n) = u^{(s+1)}(\theta_n^*, \xi(\sigma^2, \theta_n^*))$. We will first prove Lemma 4 replacing $(\theta_n^*, \xi(\sigma^2, \theta_n^*))$ by $(\bar{\theta}_n, \bar{\xi}_n)$ and then show that the approximation has the right order.

(a) More precisely, we claim that for $t \geq 2$

$$\begin{aligned} & \tilde{P}_{\sigma^2}^{v,n} \left\{ |M_n(\sigma^2, \bar{\theta}_n, \bar{\xi}_n)| \geq \frac{\delta_1}{2} \Delta_n(\sigma^2, \bar{\theta}_n, \bar{\xi}_n); \Delta_n(\sigma^2, \bar{\theta}_n, \bar{\xi}_n) \geq \frac{\delta_2}{2} \varphi_{n,\alpha_n}^2 \right\} \\ & \leq C(t, c, L, v) n^{-at} \end{aligned}$$

for some $a > 0$, uniformly in $\sigma^2 \in \Sigma$. Indeed, let

$$\mathcal{N}_n^+(\sigma^2) = \left\{ (\theta, \xi) \in \mathcal{N}_n: \Delta_n(\sigma^2, \theta, \xi) \geq \frac{1}{2} \varphi_{n,\alpha_n} \right\}.$$

We have

$$\begin{aligned} & \tilde{P}_{\sigma^2}^{v,n} \left\{ |M_n(\sigma^2, \bar{\theta}_n)| \geq \frac{\delta_1}{2} \Delta_n(\sigma^2, \bar{\theta}_n, \bar{\xi}_n); \Delta_n(\sigma^2, \bar{\theta}_n, \bar{\xi}_n) \geq \frac{\delta_2}{2} \varphi_{n,\alpha_n}^2 \right\} \\ & = \sum_{(\theta, \xi) \in \mathcal{N}_n^+(\sigma^2)} \tilde{P}_{\sigma^2}^{v,n} \left\{ |M_n(\sigma^2, \theta)| \geq \frac{\delta_1}{2} \Delta_n(\sigma^2, \theta, \xi); (\bar{\theta}_n, \bar{\xi}_n) = (\theta, \xi) \right\} \\ & \leq \sum_{(\theta, \xi) \in \mathcal{N}_n^+(\sigma^2)} \tilde{P}_{\sigma^2}^{v,n} \left\{ |M_n(\sigma^2, \theta)| \geq \frac{\delta_1 \delta_2}{4} \sqrt{\Delta_n(\sigma^2, \theta, \xi)} \varphi_{n,\alpha_n} \right\}. \end{aligned}$$

(b) Next, we need deviation bounds for $M_n(\sigma^2, \theta)$. Using (a) of the preliminary decompositions, we write $M_n(\sigma^2, \theta) = \frac{1}{\sqrt{n}} \tilde{M}_n(\sigma^2, \theta) + \sum_k a_k^n b_k^n$, where

$$\tilde{M}_n(\sigma^2, \theta) = \frac{2^{J_n/2}}{\sqrt{n}} \sum_{i=1}^n \left[\sum_k c_{J_n k}(\sigma^2 - \sigma_0^2(\theta, \cdot)) 1_{X_{(i-1)/n} \in I_{J_n k}} \right] \varepsilon_i^n.$$

Clearly, $\tilde{M}_n(\sigma^2, \theta)$ is a sum of martingale increments, with variance less than

$$C(c, L, v) \frac{2^{J_n}}{n} \sum_{i=1}^n \left(\tilde{E}_{\sigma^2}^{v,n} \left\{ \left[\sum_k c_{J_n k}(\sigma^2 - \sigma_0^2(\theta, \cdot)) 1_{X_{(i-1)/n} \in I_{J_n k}} \right]^4 \right\} \right)^{1/2}$$

where we used Cauchy–Schwarz inequality and the fact that $E_{\sigma^2}^{v,n} \{(\varepsilon_i^n)^4\} \leq C(c, L, v)$ thanks to the BDG inequality.

Let $i \geq 2$. Since $X_{(i-1)/n}$ has a density $p_{(i-1)/n}^\sigma(x)$ w.r.t. the Lebesgue measure under \tilde{P}_{σ^2} , which satisfies $p_{(i-1)/n}^\sigma(x) \leq C(c, L) \left(\frac{n}{i-1}\right)^{1/2}$ we successively have

$$\tilde{E}_{\sigma^2}^{v,n} \left\{ \left[\sum_k c_{J_n k}(\sigma^2 - \sigma_0^2(\theta, \cdot)) 1_{X_{(i-1)/n} \in I_{J_n k}} \right]^4 \right\}$$

$$\begin{aligned} &\leq C(c, L) \sqrt{\frac{n}{(i-1)}} \int_{\mathbf{R}} \left[\sum_k c_{J_n k} (\sigma^2 - \sigma_0^2(\theta, \cdot)) 1_{x \in I_{J_n k}} \right]^4 dx \\ &= C(c, L) \sqrt{\frac{n}{(i-1)}} \sum_{k_1, \dots, k_4} c_{J_n k_1} \cdots c_{J_n k_4} \int_{\mathbf{R}} 1_{x \in I_{J_n k_1}} \cdots 1_{x \in I_{J_n k_4}} dx. \end{aligned}$$

It is easily seen that restricting the sum to the 4-uplets $(k_1, \dots, k_4) \in \mathcal{I}_n$ defined by $\mathcal{I}_n = \{(k_1, \dots, k_4) : 1 \leq k_i \leq 2^{J_n}, k_1 < \dots < k_4\}$ does not alter the order (in n) of the above sum. Moreover, for $(k_1, \dots, k_4) \in \mathcal{I}_n$, we have $\int_{\mathbf{R}} 1_{x \in I_{J_n k_1}} \cdots 1_{x \in I_{J_n k_4}} dx = 2^{-4J_n}$.

Therefore, using the last bound and factorizing again, we have that the last term in the above inequality is of order

$$\sqrt{n/(i-1)} 2^{-4J_n} \left[\sum_k |c_{J_n k} (\sigma^2 - \sigma_0^2(\theta, \cdot))| \right]^4 \leq C(c, L) \sqrt{n/(i-1)} 2^{-2J_n} \Delta_n^2(\sigma^2, \theta, \xi)$$

for large enough n , where we used the decomposition (A.1) and Jensen’s inequality. We finally obtain

$$\begin{aligned} \tilde{E}_{\sigma^2}^{v,n} \{ \tilde{M}_n(\sigma^2, \theta)^2 \} &\leq C(c, L, v) \frac{1}{n} \sum_{i=1}^n (n/i)^{1/4} \Delta_n(\sigma^2, \theta_n, \xi) \\ &\leq C(c, L, v) \Delta_n(\sigma^2, \theta_n, \xi). \end{aligned}$$

Let $t \geq 2$. By Chebyshev inequality, it follows that

$$\tilde{P}_{\sigma^2}^{v,n} \left\{ \left| \frac{1}{\sqrt{n}} \tilde{M}_n(\sigma^2, \theta) \right| \geq \frac{1}{2} \sqrt{n} \varphi_{n, \alpha_n} \sqrt{\Delta_n(\sigma^2, \theta, \xi)} \right\} \leq C(t, c, L, v) n^{-t/2} \varphi_{n, \alpha_n}^{-t}.$$

From the choice of φ_{n, α_n} and since t is free, this term is arbitrarily small in power of n .

We now turn to $M_n(\sigma^2, \theta, \xi) - \frac{1}{\sqrt{n}} \tilde{M}_n(\sigma^2, \theta, \xi)$. By repeated use of Cauchy–Schwarz, we have

$$\begin{aligned} &\tilde{E}_{\sigma^2}^{v,n} \left\{ \left| \frac{1}{\sqrt{n}} M_n(\sigma^2, \theta, \xi) - \tilde{M}_n(\sigma^2, \theta, \xi) \right|^t \right\} \\ &\leq \left(\tilde{E}_{\sigma^2}^{v,n} \left\{ \left| \sum_k (a_k^n(\sigma^2))^2 \right|^t \right\} \right)^{1/2} \left(\tilde{E}_{\sigma^2}^{v,n} \left\{ \left| \sum_k (b_k^n(\sigma^2))^2 \right|^t \right\} \right)^{1/2}. \end{aligned}$$

The first term is of order $2^{-J_n t} \leq C \varphi_{n, \alpha_n}$, uniformly in $\sigma^2 \in \Sigma$, by (b) of the preliminary decompositions. Using the same technique as for \tilde{M}_n , the second term is less than $C(c, L, v) n^{-at} \sqrt{\Delta_n(\sigma^2, \theta, \xi)}$ for some $a > 0$. By Chebyshev inequality, we obtain

$$\begin{aligned} &\tilde{P}_{\sigma^2}^{v,n} \left\{ \left| M_n(\sigma^2, \theta, \xi) - \frac{1}{\sqrt{n}} \tilde{M}_n(\sigma^2, \theta) \right| \geq \frac{1}{2} \sqrt{n} \varphi_{n, \alpha_n} \sqrt{\Delta_n(\sigma^2, \theta, \xi)} \right\} \\ &\leq C(t, c, L, v) n^{-at}. \end{aligned}$$

(c) Putting together (a) and the deviation bounds of (b), we readily obtain, for $t \geq 2$

$$\begin{aligned} & \tilde{P}_{\sigma^2}^{v,n} \left\{ |M_n(\sigma^2, \bar{\theta}_n, \bar{\xi}_n)| \geq \frac{\delta_1}{2} \Delta_n(\sigma^2, \bar{\theta}_n, \bar{\xi}_n); \Delta_n(\sigma^2, \bar{\theta}_n, \bar{\xi}_n) \geq \frac{\delta_2}{4} \varphi_{n,\alpha_n}^2 \right\} \\ & \leq \# \mathcal{N}_n^+(\sigma^2) C(t, c, L, \nu) n^{-at} \leq C(t, c, L, \nu) n^{b(s+1)-at}. \end{aligned}$$

Since the choice of t is free, this term is arbitrarily small in power of n and claim (a) is proved.

(d) It remains to prove that Lemma 4 follows from claim (a). Indeed, it is easily checked that

$$|\Delta_n(\sigma^2, \bar{\theta}_n, \bar{\xi}_n) - \Delta_n(\sigma^2, \theta_n^*, \xi(\sigma^2, \theta_n^*))| \leq C(s, c, L) n^{-b\mu},$$

where μ is defined in Assumption B1. We further assume that $b \geq 2/\mu$. We thus obtain the following inclusion, for large enough n

$$\{|\Delta_n(\sigma^2)| \geq \delta_2 \varphi_{n,\alpha_n}^2\} \subseteq \left\{ \Delta_n(\sigma^2, \bar{\theta}_n, \bar{\xi}_n) \geq \frac{\delta_2}{2} \varphi_{n,\alpha_n}^2 \right\}. \quad (\text{A.2})$$

Likewise, for $t \geq 2$, we claim that

$$\tilde{P}_{\sigma^2}^{v,n} \{ |M_n(\sigma^2, \theta_n^*) - M_n(\sigma^2, \bar{\theta}_n)| \geq n \} \leq C(t, c, L, \nu) n^{-at}$$

for some $a > 0$. The proof is done using the same kind of arguments as for (b) so we omit it.

This last inequality, together with (A.2) and (a) completes the proof of Lemma 4.

Acknowledgements

I am indebted to Oleg Lepski, for helpful discussions and comments. Thanks to Sylvain Delattre and an anonymous referee for careful remarks and comments.

REFERENCES

- [1] Donhal G., On estimating the diffusion coefficient, *J. Appl. Probab.* 2 (1987) 105–114.
- [2] Genon-Catalot V., Jacod J., On the estimation of the diffusion coefficient for multidimensional processes, *Ann. Inst. H. Poincaré* 29 (1993) 119–151.
- [3] Gobet E., LAMN property for elliptic diffusions: a Malliavin calculus approach, Preprint 564, Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 6 et 7, 1999.
- [4] Florens-Zmirou D., On estimating the diffusion coefficient from discrete observations, *J. Appl. Probab.* 30 (1993) 790–804.
- [5] Fuk D.K., Nagaev S.V., Probability inequalities for sums of independent random variables, *Theory Probab. Appl.* 4 (1971) 643–660.
- [6] Hoffmann M., L_p estimation of the diffusion coefficient, *Bernoulli* 5 (3) (1999) 447–481.
- [7] Hoffmann M., Lepski O.V., Random rates in anisotropic regression, Preprint 568, Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 6 et 7, 2000.
- [8] Jacod J., Non-parametric kernel estimation of the diffusion coefficient of a diffusion, *Scand. J. Statist.* 27 (2000) 83–96.

- [9] Jacod J., Rates of convergence to the local time of a diffusion, Ann. I.H.P. 34 (1998) 505–544.
- [10] Korostelev A.P., Tsybakov A.D., Minimax Theory of Image Reconstruction, Lecture Notes in Statistics, Vol. 82, Springer, Berlin, 1993.
- [11] Lepski O.V., How to improve accuracy of estimation, Math. Methods Statist. 8 (1999) 441–486.
- [12] Revuz D., Yor M., Continuous Martingales and Brownian Motion, Springer, Berlin, 1991.