



HAL
open science

Nonparametric estimation of the diffusion coefficient from S.D.E. paths

Eddy ELLA MINTSA

► **To cite this version:**

Eddy ELLA MINTSA. Nonparametric estimation of the diffusion coefficient from S.D.E. paths. 2023.
hal-04156441v1

HAL Id: hal-04156441

<https://hal.science/hal-04156441v1>

Preprint submitted on 10 Jul 2023 (v1), last revised 18 Mar 2024 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nonparametric estimation of the diffusion coefficient from S.D.E. paths

Eddy Ella-Mintsa
LAMA, Université Gustave Eiffel.

July 10, 2023

Abstract

Consider a diffusion process $X = (X_t)_{t \in [0,1]}$ observed at discrete times and high frequency, solution of a stochastic differential equation whose drift and diffusion coefficients are assumed to be unknown. In this article, we focus on the nonparametric estimation of the diffusion coefficient. We propose ridge estimators of the square of the diffusion coefficient from discrete observations of X and that are obtained by minimization of the least squares contrast. We prove that the estimators are consistent and derive rates of convergence as the size of the sample paths tends to infinity, and the discretization step of the time interval $[0, 1]$ tend to zero. The theoretical results are completed with a numerical study over synthetic data.

Keywords. Nonparametric estimation, diffusion process, diffusion coefficient, least squares contrast, repeated observations.

MSC: 62G05; 62M05; 60J60

1 Introduction

Let $X = (X_t)_{t \in [0,1]}$ be a one dimensional diffusion process with finite horizon time, solution of the following stochastic differential equation:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = 0 \quad (1)$$

where $(W_t)_{t \geq 0}$ is a standard Brownian motion. The drift function b and the diffusion coefficient σ are assumed to be unknown Lipschitz functions. We denote by $(\mathcal{F}_t)_{t \in [0,1]}$ the natural filtration of the diffusion process X . The goal of the article is to construct, from N discrete observations $\bar{X}^j = (X_{k\Delta_n}^j)_{0 \leq k \leq n}$, $1 \leq j \leq N$ with time step $\Delta_n = 1/n$, a nonparametric estimator of the square of the diffusion coefficient $\sigma^2(\cdot)$. We are in the framework of high frequency data since the time step Δ_n tends to zero as n tends to infinity. Furthermore, we consider estimators of $\sigma^2(\cdot)$ built from a single diffusion path ($N = 1$), and those built on N paths when $N \rightarrow \infty$. In this paper, we first propose a ridge estimator of $\sigma^2(\cdot)$ on a compact interval. Secondly, we focus on a nonparametric estimation of $\sigma^2(\cdot)$ on the real line \mathbb{R} . We measure the risk of any estimator $\hat{\sigma}^2$ of the square of the diffusion coefficient σ^2 by $\mathbb{E} \left[\|\hat{\sigma}^2 - \sigma^2\|_{n,N}^2 \right]$, where $\|\hat{\sigma}^2 - \sigma^2\|_{n,N}^2 := (Nn)^{-1} \sum_{j=1}^N \sum_{k=0}^{n-1} \left(\hat{\sigma}^2(X_{k\Delta}^j) - \sigma^2(X_{k\Delta}^j) \right)^2$ is an empirical norm defined from the sample paths.

Related works. There is a large literature on the estimation of coefficients of diffusion processes, and we focus on the papers studying the estimation of σ^2 .

Estimation of the diffusion coefficient has been considered in the parametric case (see e.g. Genon-Catalot & Jacod (1993), Jacod (1993), Genon-Catalot & Jacod (1994), Clement (1997), Genon-Catalot *et al.* (1999), Gloter (2000), Sørensen (2002)). In the nonparametric case, estimators of the diffusion coefficient from discrete observations are proposed under various frameworks.

First, the diffusion coefficient is constructed from one discrete observation of the diffusion process ($N = 1$) in long time ($T \rightarrow \infty$) (see e.g. Hoffmann (1999a), Comte *et al.* (2007), Schmisser (2012), Schmisser (2019)), or in short time ($T = 1$) (see e.g. Genon-Catalot *et al.* (1992), Soulier (1993), Hoffmann (1997), Florens (1998), Hoffmann (1999b)). Note that in short time ($T < \infty$), only the diffusion coefficient can be estimated consistently from a single discrete path contrary to the drift function whose consistent estimation relies on repeated discrete observations of the diffusion process (see e.g. Comte & Genon-Catalot (2020a), Denis *et al.* (2021)). For the case of short time diffusion processes (for instance $T = 1$), estimators of a time-dependent diffusion coefficients $t \mapsto \sigma^2(t)$ have been proposed. In this context, Genon-Catalot *et al.* (1992) built a nonparametric estimator of $t \mapsto \sigma^2(t)$ and studied its L_2 risk using wavelets methods, Soulier (1993) studies the L_p risk of a kernel estimator of $\sigma^2(t)$, and Hoffmann (1997) derived a minimax rate of convergence of order $n^{-ps/(1+2s)}$ where $s > 1$ is the smoothness parameter of the Besov space $\mathcal{B}_{p,\infty}^s([0,1])$ (see later in the paper). For the space-dependent diffusion coefficient $x \mapsto \sigma^2(x)$, a first estimator based on kernels and built from a single discrete observation of the diffusion process with $T = 1$ is proposed in Florens-Zmirou (1993). The estimator has been proved to be consistent under a condition on the bandwidth, but a rate of convergence of its risk of estimation has not been established.

Secondly, the diffusion coefficient is built in short time ($T < \infty$) from N repeated discrete observations with $N \rightarrow \infty$. In Denis *et al.* (2022), a nonparametric estimator of σ^2 is proposed from repeated discrete observations on the real line \mathbb{R} when the time horizon $T = 1$. The estimator has been proved to be consistent with a rate of order $N^{-1/5}$ over the space of Lipschitz functions.

Two main methods are used to build consistent nonparametric estimators of $x \mapsto \sigma^2(x)$. The first method is the one using kernels (see e.g. Florens (1998), Bandi & Phillips (2003), Renò (2006), Gouriéroux *et al.* (2017), Schmisser (2019), Park & Wang (2021)), the other method consists in estimating σ^2 as solution of a nonparametric regression model using the least squares approach. Since the diffusion coefficient is assumed to belong to an infinite dimensional space, the method consists in projecting σ^2 into a finite dimensional subspace, estimating the projection and making a data-driven selection of the dimension by minimizing a penalized least squares contrast (see e.g. Hoffmann (1999b), Hoffmann (1999a), Schmisser (2012), Comte (2017), Schmisser (2019), Denis *et al.* (2022)).

Main contribution. In this article, we assume to have at our disposal N i.i.d. discrete observations of length n of the diffusion process X . The main objectives of this paper are the following.

1. Construct a consistent and implementable ridge estimator of σ^2 from a single diffusion path ($N = 1$) using the least squares approach. We derive rates of convergence of the risk of estimation of the ridge estimators built on a compact interval and on the real line \mathbb{R} over a Hölder space, taking advantage of the properties of the local time of the diffusion process, and its link with the transition density.
2. We extend the result to the estimation of σ^2 on repeated observations of the diffusion process ($N \rightarrow \infty$). We prove that the estimators built on a compact interval and on \mathbb{R} are more efficient considering their respective rates compared to nonparametric estimators built from a single diffusion path.

3. Focusing on the support of the diffusion coefficient, we consider an intermediate case between a compact interval and \mathbb{R} by proposing a ridge estimator of σ^2 restricted to the compact interval $[-A_N, A_N]$ where $A_N \rightarrow \infty$ as $N \rightarrow \infty$. The benefit of this approach is that the resulting projection estimator can reach a faster rate of convergence compared to the rate obtained on the real line \mathbb{R} .
4. Finally, we propose adaptive estimators of σ^2 based on a data-driven selection of the dimension through the minimization of the penalized least squares contrast in different settings.

We sum up below the rates of convergence (up to a log-factor) of the ridge estimators of $\sigma_{|I}^2$ with $I \subseteq \mathbb{R}$ over a Hölder space defined in the next section with a smoothness parameter $\beta \geq 1$.

Estimation interval	$N = 1$ and $n \rightarrow +\infty$	$N \rightarrow +\infty$ and $n \rightarrow +\infty$
$I = [-A, A], \quad A > 0$	$n^{-\beta/(2\beta+1)}$	$(Nn)^{-\beta/(2\beta+1)}$
$I = [-A_N, A_N], \quad A_N \xrightarrow{N \rightarrow +\infty} +\infty$	xxxx	$(Nn)^{-\beta/(2\beta+1)}, \quad N \propto n$
$I = \mathbb{R}$	$n^{-\beta/(4\beta+1)}$	$(Nn)^{-\beta/(4\beta+1)} + n^{-2}$

Table 1: Rates of convergence of the square root of the risk of estimation $\mathbb{E} \left[\left\| \hat{\sigma}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \right]$ of the non-adaptive estimators $\hat{\sigma}^2$ of the square of the diffusion coefficient $\sigma_{|I}^2$ built from one diffusion path ($N = 1$) on the left column, and from repeated observations of the diffusion process ($N \rightarrow \infty$) on the right column. For the precise results, see Sections 3 and 4.

Outline of the paper. In Section 2, we define our framework with the key assumptions on the coefficients of the diffusion process ensuring for instance that Equation (1) admits a unique strong solution. Section 3 is devoted to the non-adaptive estimation of the diffusion coefficient from one diffusion path both on a compact interval and on the real line \mathbb{R} . In Section 4, we extend the study to the non-adaptive estimation of the diffusion coefficient from repeated observations of the diffusion process. We propose in Section 5, adaptive estimators of the diffusion coefficient, and Section 6 complete the study with numerical evaluation of the performance of estimators. We prove our theoretical results in Section 8.

2 Framework and assumptions

Consider a diffusion process $X = (X_t)_{t \in [0,1]}$, solution of Equation (1) whose drift and diffusion coefficient satisfy the following assumption.

Assumption 2.1. 1. *There exists a constant $L_0 > 0$ such that b and σ are L_0 -Lipschitz functions on \mathbb{R} .*

2. *There exist constants $\sigma_0, \sigma_1 > 0$ such that : $\sigma_0 \leq \sigma(x) \leq \sigma_1, \quad \forall x \in \mathbb{R}$.*

3. *$\sigma \in \mathcal{C}^2(\mathbb{R})$ and there exist $C > 0$ and $\alpha \geq 0$ such that:*

$$|\sigma'(x)| + |\sigma''(x)| \leq C(1 + |x|^\alpha), \quad \forall x \in \mathbb{R}.$$

Under Assumption 2.1, $X = (X_t)_{t \in [0,1]}$ is the unique strong solution of Equation (1), and this unique solution admits a transition density $(t, x) \mapsto p_X(t, x)$. Besides, we draw from Assumption 2.1 that

$$\forall q \geq 1, \quad \mathbb{E} \left[\sup_{t \in [0,1]} |X_t|^q \right] < \infty. \quad (2)$$

2.1 Definitions and notations

We suppose to have at our disposal, a sample $D_{N,n} = \{\bar{X}^j, j = 1, \dots, N\}$ constituted of N independent copies of the discrete observation $\bar{X} = (X_{k\Delta_n})_{0 \leq k \leq n}$ of the diffusion process X where $\Delta_n = 1/n$ is the time-step. The objective is to construct, from the sample $D_{N,n}$, a nonparametric estimator of the square σ^2 of the diffusion coefficient on an interval $I \subseteq \mathbb{R}$. In the sequel, we consider two main cases, the first one being the estimation of σ^2 on the interval I from a single path ($N = 1$ and $n \rightarrow \infty$). For the second case, we assume that both N and n tend to infinity.

For each measurable function h , such that $\mathbb{E}[h^2(X_t)] < \infty$ for all $t \in [0, 1]$, we define the following empirical norms:

$$\|h\|_n^2 := \mathbb{E}_X \left[\frac{1}{n} \sum_{k=0}^{n-1} h^2(X_{k\Delta_n}) \right], \quad \|h\|_{n,N}^2 := \frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} h^2(X_{k\Delta_n}^j). \quad (3)$$

For all $h \in \mathbb{L}^2(I)$, we have

$$\|h\|_n^2 = \int_I h^2(x) \frac{1}{n} \sum_{k=0}^{n-1} p_X(k\Delta_n, x_0, x) dx = \int_I h^2(x) f_n(x) dx,$$

where $f_n : x \mapsto \frac{1}{n} \sum_{k=0}^{n-1} p_X(k\Delta_n, x)$ is a density function. For the case of non-adaptive estimators of σ^2 , we also establish bounds of the risks of the estimators based on the empirical norm $\|\cdot\|_n$ or the \mathbb{L}^2 -norm $\|\cdot\|$ when the estimation interval I is compact.

For any integers $p, q \geq 2$ and any matrix $M \in \mathbb{R}^{p \times q}$, we denote by ${}^t M$, the transpose of M .

2.2 Spaces of approximation

We propose projection estimators of σ^2 on a finite-dimensional subspace. To this end, we consider for each $m \geq 1$, a m -dimensional subspace \mathcal{S}_m given as follows:

$$\mathcal{S}_m := \text{Span}(\phi_\ell, \ell = 0, \dots, m-1), \quad m \geq 1 \quad (4)$$

where the functions $(\phi_\ell, \ell \in \mathbb{N})$ are continuous, linearly independent and bounded on I . Furthermore, we need to control the ℓ^2 -norm of the coordinate vectors of elements of \mathcal{S}_m , which leads to the following constrained subspace,

$$\mathcal{S}_{m,L} := \left\{ h = \sum_{\ell=0}^{m-1} a_\ell \phi_\ell, \quad \sum_{\ell=0}^{m-1} a_\ell^2 = \|\mathbf{a}\|_2^2 \leq mL, \quad \mathbf{a} = (a_0, \dots, a_{m-1}), \quad L > 0 \right\}. \quad (5)$$

Note that $\mathcal{S}_{m,L} \subset \mathcal{S}_m$ and $\mathcal{S}_{m,L}$ is no longer a vector space. The control of the coordinate vectors allows to establish an upper bound of the estimation error that tends to zero as $n \rightarrow \infty$ or $N, n \rightarrow \infty$. In fact, we prove in the next sections that the construction of consistent estimators of σ^2 requires the functions $h = \sum_{\ell=0}^{m-1} a_\ell \phi_\ell$ to be bounded, such that

$$\|h\|_\infty \leq \max_{\ell=0, \dots, m-1} \|\phi_\ell\|_\infty \|\mathbf{a}\|_2.$$

This condition is satisfied for the functions of the constrained subspaces $\mathcal{S}_{m,L}$ with $m \geq 1$. In this article, we work with the following bases.

[B] The B-spline basis This is an exemple of a non-orthonormal basis defined on a compact interval. Let $A > 0$ be a real number, and suppose (without restriction) that $I = [-A, A]$. Let $K, M \in \mathbb{N}^*$, and consider $\mathbf{u} = (u_{-M}, \dots, u_{K+M})$ a knots vector such that $u_{-M} = \dots = u_{-1} = u_0 = -A$, $u_{K+1} = \dots = u_{K+M} = A$, and for all $i = 0, \dots, K$,

$$u_i = -A + i \frac{2A}{K}.$$

One calls **B-spline functions**, the piecewise polynomial functions $(B_\ell)_{\ell=-M, \dots, K-1}$ of degree M , associated with the knots vector \mathbf{u} (see Györfi *et al.* (2006), *Chapter 14*). The **B-spline functions** are linearly independent smooths functions returning zero for all $x \notin [-A, A]$, and satisfying some smoothness conditions established in Györfi *et al.* (2006). Thus, we consider approximation subspaces \mathcal{S}_{K+M} defined by

$$\mathcal{S}_{K+M} = \text{Span} \{B_\ell, \ell = -M, \dots, K-1\}$$

of dimension $\dim(\mathcal{S}_{K+M}) = K+M$, and in which, each function $h = \sum_{\ell=-M}^{K-1} a_\ell B_\ell$ is $M-1$ times continuously differentiable thanks to the properties of the spline functions (see Györfi *et al.* (2006)). Besides, the spline basis is included in the definition of both the subspace \mathcal{S}_m and the constrained subspace $\mathcal{S}_{m,L}$ (see Equations (4) and (5)) with $m = K+M$ and for any coordinates vector $(a_{-M}, \dots, a_{K-1}) \in \mathbb{R}^{K+M}$,

$$\sum_{\ell=-M}^{K-1} a_\ell B_\ell = \sum_{\ell=0}^{m-1} a_{\ell-M} B_{\ell-M}.$$

The integer $M \in \mathbb{N}^*$ is fixed, while K varies in the set of integers \mathbb{N}^* . If we assume that σ^2 belongs to the Hölder space $\Sigma_I(\beta, R)$ given as follows:

$$\Sigma_I(\beta, R) := \left\{ h \in \mathcal{C}^{\lfloor \beta \rfloor + 1}(I), \left| h^{(\ell)}(x) - h^{(\ell)}(y) \right| \leq R|x - y|^{\beta - \ell}, \quad x, y \in I \right\},$$

where $\beta \geq 1$, $\ell = \lfloor \beta \rfloor$ and $R > 0$, then the unknown function $\sigma^2|_I$ restricted to the compact interval I can be approximated in the constrained subspace $\mathcal{S}_{K+M,L}$ spanned by the spline basis. This approximation results to the following bias term:

$$\inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2|_I\|_n^2 \leq C|I|^{2\beta} K^{-2\beta} \quad (6)$$

where the constant $C > 0$ depends on β, R and M , and $|I| = \sup I - \inf I$. The above result is a modification of *Lemma D.2* in Denis *et al.* (2021).

[F] The Fourier basis The subspace \mathcal{S}_m can be spanned by the Fourier basis

$$\{f_\ell, \ell = 0, \dots, m-1\} = \{1, \sqrt{2} \cos(2\pi jx), \sqrt{2} \sin(2\pi jx), j = 1, \dots, d\} \quad \text{with } m = 2d + 1.$$

The above Fourier basis is defined on the compact interval $[0, 1]$. The definition can be extended to any compact interval, replacing the bases functions $x \mapsto f_\ell(x)$ by $x \mapsto 1/(\max I - \min I) f_\ell(\frac{x - \min I}{\max I - \min I})$. We use this basis to build the estimators of σ^2 on a compact interval $I \subset \mathbb{R}$.

Define for all $s \geq 1$ and for any compact interval $I \subset \mathbb{R}$, the Besov space $\mathcal{B}_{2,\infty}^s(I)$ which is a space of functions $f \in L^2(I)$ such that the $\lfloor s \rfloor^{\text{th}}$ derivative $f^{(\lfloor s \rfloor)}$ belongs to the space $\mathcal{B}_{2,\infty}^{s-\lfloor s \rfloor}(I)$ given by

$$\mathcal{B}_{2,\infty}^{s-\lfloor s \rfloor}(I) = \left\{ f \in L^2(I) \text{ and } \frac{w_{2,f}(t)}{t^{s-\lfloor s \rfloor}} \in L^\infty(I \cap \mathbb{R}^+) \right\}$$

where for $s - \lfloor s \rfloor \in (0, 1)$, $w_{2,f}(t) = \sup_{|h| \leq t} \|\tau_h f - f\|_2$ with $\tau_h f(x) = f(x - h)$, and for $s - \lfloor s \rfloor = 1$, $w_{2,f}(t) = \sup_{|h| \leq t} \|\tau_h f + \tau_{-h} f - 2f\|_2$. Thus, if we assume that the function $\sigma^2|_I$ belongs to the Besov space $\mathcal{B}_{2,\infty}^s$, then it can be approximated in a constrained subspace $\mathcal{S}_{m,L}$ spanned by the Fourier basis. Moreover, under Assumption 2.1 and from *Lemma 12* in Barron *et al.* (1999), there exists a constant $C > 0$ depending on the constant τ_1 of Equation (12), the smoothness parameter s of the Besov space such that

$$\inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma^2|_I \right\|_n^2 \leq \tau_1 \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma^2|_I \right\|^2 \leq C \left| \sigma^2|_I \right|_\beta^2 m^{-2\beta} \quad (7)$$

where $|\sigma^2|_I|_s$ is the semi-norm of $\sigma^2|_I$ in the Besov space $\mathcal{B}_{2,\infty}^s(I)$.

Note that for all $\beta \geq 1$, the Hölder space $\Sigma_I(\beta, R)$ and the Besov space $\mathcal{B}_{2,\infty}^\beta$ satisfy:

$$L^\infty(\mathbb{R}) \cap \Sigma_I(\beta, R) \subset \mathcal{B}_{\infty,\infty}^\beta(I) \subset \mathcal{B}_{2,\infty}^\beta(I)$$

(see DeVore & Lorentz (1993), *Chap. 2 page 16*). As a result, we rather consider in the sequel the Hölder space $\Sigma_I(\beta, R)$ which can also be approximated by the Fourier basis.

[H] The Hermite basis The basis is defined from the Hermite functions $(h_j, j \geq 0)$ defined on \mathbb{R} and given for all $j \geq 0$ and for all $x \in \mathbb{R}$ by:

$$h_j(x) = c_j H_j(x), \quad \text{where } H_j(x) = (-1)^j \exp\left(\frac{x^2}{2}\right) \frac{d^j}{dx^j} \left(e^{-x^2/2}\right) \quad \text{and } c_j = (2^j j! \sqrt{\pi})^{-1/2}.$$

The polynomials $H_j(x)$, $j \geq 0$ are the Hermite polynomials, and $(h_j, j \geq 0)$ is an orthonormal basis of $L^2(\mathbb{R})$. Furthermore, for all $j \geq 1$ and $x \in \mathbb{R}$, $|h_j(x)| \leq c|x| \exp(-c_0 x^2)$ for $x^2 \geq (3/2)(4j + 3)$ where $c, c_0 > 0$ are constants independent of j (see Comte & Genon-Catalot (2020b), Proof of Proposition 3.5). We use the Hermite basis in the sequel for the estimation of σ^2 on the real line \mathbb{R} .

If one assumes that σ^2 belongs to the Sobolev space $W_{f_n}^s(\mathbb{R}, R)$ given for all $s \geq 1$ by

$$W_{f_n}^s(\mathbb{R}, R) := \{g \in L^2(\mathbb{R}, f_n(x)dx), \forall \ell \geq 1, \|g - g_\ell\|_n^2 \leq R\ell^{-s}\}$$

where for each $\ell \geq 1$, g_ℓ is the $L^2(\mathbb{R}, f_n(x)dx)$ -orthogonal projection of g on the ℓ -dimensional vector space \mathcal{S}_ℓ spanned by the Hermite basis. Consider a compact interval $I \subset \mathbb{R}$ and the following spaces:

$$W^s(I, R) := \left\{ g \in L^2(I), \sum_{j=0}^{\infty} j^s \langle g, \phi_j \rangle^2 \leq R \right\},$$

$$W_{f_n}^s(I, R) := \{g \in L^2(I, f_n(x)dx), \forall \ell \geq 1, \|g - g_\ell\|_n^2 \leq R\ell^{-s}\}$$

where $(\phi_j)_{j \geq 0}$ is an orthonormal basis defined on I and for all $\ell \geq 1$, g_ℓ is the orthogonal projection of g onto $\mathcal{S}_\ell = \text{Span}(h_j, j \leq \ell)$ of dimension $\ell \geq 1$ (see *e.g.* Comte & Genon-Catalot (2021)). Then, for all $g \in W^s(I, R)$, we have

$$g = \sum_{j=0}^{\infty} \langle g, \phi_j \rangle \phi_j \quad \text{and} \quad \|g - g_\ell\|^2 = \sum_{j=\ell+1}^{\infty} \langle g, \phi_j \rangle^2 \leq \ell^{-s} \sum_{j=\ell+1}^{\infty} j^s \langle g, \phi_j \rangle^2 \leq R\ell^{-s}.$$

We have $W_{f_n}^s(I, R) = W^s(I, R)$ as the empirical norm $\|\cdot\|_n$ and the L^2 -norm $\|\cdot\|$ are equivalent. The space $W_{f_n}^s(\mathbb{R}, R)$ is an extension of the space $W_{f_n}^s(I, R)$ when $I = \mathbb{R}$ and $(\phi_j)_{j \geq 0}$ is the Hermite basis.

Remark 2.2. The \mathbf{B} -spline basis is used for the estimation of σ^2 on a compact interval on one side ($N = 1$ and $N > 1$), and on the real line on the other side restricting σ^2 on the compact interval $[-\log(n), \log(n)]$ for $N = 1$, or $[-\log(N), \log(N)]$ for $N > 1$, and bounding the exit probability of the process X from the interval $[-\log(N), \log(N)]$ (or $[-\log(n), \log(n)]$) by a negligible term with respect to the estimation error. In a similar context, the Fourier basis is used as an orthonormal basis to build nonparametric estimators of σ^2 on a compact interval and on \mathbb{R} , both for $N = 1$ and for $N > 1$. The main goal is to show that, in addition to the spline basis which is not orthogonal, we can build projection estimators of σ^2 on orthonormal bases that are consistent. The advantage of the Hermite basis compared to the Fourier basis is its definition on the real line \mathbb{R} . As a result, we use the Hermite basis to propose for $N > 1$, a projection estimator of σ^2 whose support is the real line \mathbb{R} .

Remark 2.3. Denote by \mathcal{M} , the set of possible values of the dimension $m \geq 1$ of the approximation subspace \mathcal{S}_m . If $(\phi_0, \dots, \phi_{m-1})$ is an orthonormal basis, then for all $m, m' \in \mathcal{M}$ such that $m < m'$, we have $\mathcal{S}_m \subset \mathcal{S}_{m'}$. For the case of the \mathbf{B} -spline basis, one can find a subset $\mathcal{K} \subset \mathcal{M}$ of the form

$$\mathcal{K} = \{2^q, q = 0, \dots, q_{\max}\}$$

such that for all $K, K' \in \mathcal{K}$, $K < K'$ implies $\mathcal{S}_{K+M} \subset \mathcal{S}_{K'+M}$ (see for example Denis et al. (2021)). The nesting of subspaces \mathcal{S}_m , $m \in \mathcal{M}$ is of great importance in the context of adaptive estimation of the diffusion coefficient and the establishment of upper-bounds for the risk of adaptive estimators.

In the sequel, we denote by $[\mathbf{F}]$, $[\mathbf{H}]$ and $[\mathbf{B}]$ the respective collection of subspaces spanned by the Fourier basis, the Hermite basis and the \mathbf{B} -spline basis.

2.3 Ridge estimators of the square of the diffusion coefficient

We establish from Equation (1) and the sample $D_{N,n}$ the regression model for the estimation of σ^2 . For all $j \in \llbracket 1, N \rrbracket$ and $k \in \llbracket 0, n-1 \rrbracket$, define

$$U_{k\Delta_n}^j := \frac{\left(X_{(k+1)\Delta_n}^j - X_{k\Delta_n}^j\right)^2}{\Delta_n}.$$

The increments $U_{k\Delta_n}^j$ are approximations in discrete times of $\frac{d\langle X, X \rangle_t}{dt}$ since, from Equation (1), one has $d\langle X, X \rangle_t = \sigma^2(X_t)dt$. From Equation (1), we obtain the following regression model,

$$U_{k\Delta_n}^j = \sigma^2(X_{k\Delta_n}^j) + \zeta_{k\Delta_n}^j + R_{k\Delta_n}^j, \quad \forall (j, k) \in \llbracket 1, N \rrbracket \times \llbracket 0, n-1 \rrbracket$$

where $U_{k\Delta_n}^j$ is the response variable, $\zeta_{k\Delta_n}^j$ and $R_{k\Delta_n}^j$ are respectively the error term and a negligible residual whose explicit formulas are given in Section 8.

We consider the least squares contrast $\gamma_{n,N}$ defined for all $m \in \mathcal{M}$ and for all function $h \in \mathcal{S}_{m,L}$ by

$$\gamma_{n,N}(h) := \frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \left(U_{k\Delta}^j - h(X_{k\Delta_n}^j)\right)^2. \quad (8)$$

For each dimension $m \in \mathcal{M}$, the projection estimator $\hat{\sigma}_m^2$ of σ^2 over the subspace $\mathcal{S}_{m,L}$ satisfies:

$$\hat{\sigma}_m^2 \in \arg \min_{h \in \mathcal{S}_{m,L}} \gamma_{n,N}(h). \quad (9)$$

Indeed, for each dimension $m \in \mathcal{M}$, the estimator $\hat{\sigma}_m^2$ of σ^2 given in Equation (9) satisfies $\hat{\sigma}_m^2 = \sum_{\ell=0}^{m-1} \hat{a}_\ell \phi_\ell$, where

$$\hat{\mathbf{a}} = (\hat{a}_0, \dots, \hat{a}_{m-1}) := \arg \min_{\|\mathbf{a}\|_2^2 \leq mL} \|\mathbf{U} - \mathbf{F}_m \mathbf{a}\|_2^2 \quad (10)$$

with $\mathbf{t}\mathbf{U} = \left(U_0^1, \dots, U_{(n-1)\Delta_n}^1, \dots, U_0^N, \dots, U_{(n-1)\Delta_n}^N \right)$ and the matrix \mathbf{F}_m is defined as follows

$$\mathbf{F}_m := \left({}^t(\phi_\ell(X_0^j), \dots, \phi_\ell(X_{(n-1)\Delta_n}^j)) \right)_{\substack{0 \leq \ell \leq m-1 \\ 1 \leq j \leq N}} \in \mathbb{R}^{Nn \times m}.$$

The vector of coefficients $\hat{\mathbf{a}}$ is unique and called the ridge estimator of \mathbf{a} because of the ℓ^2 constraint on the coordinate vectors (see Hastie *et al.* (2001) Chap. 3 page 61).

3 Estimation of the diffusion coefficient from a single diffusion path

This section focuses on the nonparametric estimation of the square of the diffusion coefficient σ^2 on an interval $I \subseteq \mathbb{R}$ when only a single diffusion path is observed at discrete times ($N = 1$). It is proved in the literature that one can construct consistent estimators of the diffusion coefficient from one path when the time horizon T is finite (see *e.g.* Hoffmann (1999b)). Two cases are considered. First, we propose a ridge estimator of σ^2 on a compact interval $I \subset \mathbb{R}$, say for example $I = [-1, 1]$. Secondly, we extend the study to the estimation of σ^2 on the real line $I = \mathbb{R}$.

3.1 Non-adaptive estimation of the diffusion coefficient on a compact interval

In this section, we consider the estimator $\hat{\sigma}_m^2$ of the compactly supported square of the diffusion coefficient $\sigma_{|I}^2$ on the constrained subspaces $\mathcal{S}_{m,L}$ from the observation of a single diffusion path.

Since the interval $I \subset \mathbb{R}$ is compact, the immediate benefit is that the density function f_n defined from the transition density of the diffusion process $\bar{X} = (X_{k\Delta})$ is bounded from below. In fact, there exist constants $\tau_0, \tau_1 \in (0, 1]$ such that

$$\forall x \in I, \quad \tau_0 \leq f_n(x) \leq \tau_1, \quad (11)$$

(see Denis *et al.* (2021)). Thus, for each function $h \in \mathbb{L}^2(I)$,

$$\tau_0 \|h\|^2 \leq \|h\|_n^2 \leq \tau_1 \|h\|^2 \quad (12)$$

where $\|\cdot\|$ is the \mathbb{L}^2 -norm. Equation (12) allows to establish global rates of convergence of the risk of the ridge estimators $\hat{\sigma}_m^2$ of $\sigma_{|I}^2$ with $m \in \mathcal{M}$ using the L^2 -norm $\|\cdot\|$ which is, in this case, equivalent with the empirical norm $\|\cdot\|_n$.

To establish an upper-bound of the risk of estimation that tends to zero as n tends to infinity, we need to establish equivalence relations between the pseudo-norms $\|\cdot\|_{n,1}$ ($N = 1$) and $\|\cdot\|_X$ on one side, and $\|\cdot\|_X$ and the L^2 -norm $\|\cdot\|$ on the other side, where the random pseudo-norm $\|\cdot\|_X$ is defined for each function $h \in \mathbb{L}^2(I)$ by

$$\|h\|_X^2 := \int_0^1 h^2(X_s) ds. \quad (13)$$

Define for $x \in \mathbb{R}$, the local time \mathcal{L}^x of the diffusion process $X = (X_t)_{t \in [0,1]}$ by

$$\mathcal{L}^x = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^1 \mathbf{1}_{(x-\varepsilon, x+\varepsilon)}(X_s) ds. \quad (14)$$

In general, the local time of a continuous semimartingale is a.s. *càdlàg* (see *e.g.* Revuz & Yor (2013)). But, for diffusion processes and under Assumption 2.1, the local time \mathcal{L}^x is bicontinuous at any point $x \in \mathbb{R}$ (see Lemma 8.5 in Section 8). Furthermore, we obtain the following result.

Lemma 3.1. *Under Assumption 2.1, and for any continuous and integrable function h , it yields,*

1. $\int_0^1 h(X_s) ds = \int_{\mathbb{R}} h(x) \mathcal{L}^x dx$.
2. For all $x \in \mathbb{R}$, $\mathbb{E}(\mathcal{L}^x) = \int_0^1 p_X(s, x) ds$.

In Lemma 3.1, we remark that there is a link between the local time and the transition density of the diffusion process. Thus, if we consider the pseudo-norm $\|\cdot\|_X$ depending on the process $X = (X_t)_{t \in [0,1]}$ and given in Equation (13), and using Lemma 3.1, we obtain that,

$$\mathbb{E} [\|h\|_X^2] = \int_{\mathbb{R}} h^2(x) \mathbb{E}[\mathcal{L}^x] dx = \int_{\mathbb{R}} h^2(x) \int_0^1 p_X(s, x) ds dx \geq \tau_0 \|h\|^2. \quad (15)$$

where $\int_0^1 p_X(s, x) ds \geq \tau_0 > 0$ (see Denis *et al.* (2021), Lemma 4.3), and $\|h\|^2$ is the \mathbb{L}^2 -norm of h .

Theorem 3.2. *Set $L = \log(n)$. Suppose that σ^2 is approximated in one of the collections $[\mathbf{B}]$ and $[\mathbf{F}]$. Under Assumption 2.1, it yields*

$$\begin{aligned} \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] &\leq 3 \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + C \left(\frac{m}{n} + \frac{m^{2\gamma+1} \log(n)}{n^{\gamma/2}} + \Delta_n^2 \right) \\ \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2 \right] &\leq \frac{34\tau_1}{\tau_0} \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + C' \left(\frac{m}{n} + \frac{m^{2\gamma+1} \log(n)}{n^{\gamma/2}} + \Delta_n^2 \right) \end{aligned}$$

where the number $\gamma > 1$ comes from the use of the Hölder inequality. The constant $C > 0$ depends on σ_1 and the constant $C' > 0$ depends on σ_1, τ_0 and τ_1 .

We observe that the upper-bound of the risk of estimation of $\hat{\sigma}_m^2$ is composed of the bias term, which quantifies the cost of approximation of $\sigma_{|I}^2$ in the constrained space $\mathcal{S}_{m,L}$, the estimation error $O(m/n)$ and the cost of the time discretization $O(\Delta_n^2)$ are established on a random event in which the pseudo-norms $\|\cdot\|_{n,1}$ and $\|\cdot\|_X$ are equivalent, and whose probability of the complementary times $\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{\infty}^2$ is bounded by the term $O\left(\frac{m^{2\gamma+1} \log(n)}{n^{\gamma/2}}\right)$ (see Lemma 8.9 and proof of Theorem 3.2).

The next result proves that the risk of estimation can reach a rate of convergence of the same order than the rate established in Hoffmann (1999b) if the parameter $\gamma > 1$ is chosen such that the term $O(m^{2\gamma+1} \log(n)/n^{\gamma/2})$ is of the same order than the estimation error of order m/n . Note that the risk $\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2$ is random since

$$\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2 = \mathbb{E}_X \left[\frac{1}{n} \sum_{k=0}^{n-1} (\hat{\sigma}_m^2 - \sigma_{|I}^2)(X_{k\Delta}) \right]$$

and the estimator $\hat{\sigma}_m^2$ is built from an independent copy \bar{X}^1 of the discrete times process \bar{X} . Thus, the expectation \mathbb{E} relates to the estimator $\hat{\sigma}_m^2$.

Corollary 3.3. *Suppose that $\sigma^2 \in \Sigma_I(\beta, R)$ with $\beta > 3/2$, and $\gamma = 2(2\beta + 1)/(2\beta - 3)$. Assume that $K_{\text{opt}} \propto n^{1/(2\beta+1)}$ for $[\mathbf{B}]$ ($m_{\text{opt}} = K_{\text{opt}} + M$), and $m_{\text{opt}} \propto n^{1/(2\beta+1)}$ for $[\mathbf{F}]$. Under Assumptions 2.1, it yields,*

$$\begin{aligned} \mathbb{E} \left[\left\| \hat{\sigma}_{m_{\text{opt}}}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] &= O\left(\log(n) n^{-2\beta/(2\beta+1)}\right) \\ \mathbb{E} \left[\left\| \hat{\sigma}_{m_{\text{opt}}}^2 - \sigma_{|I}^2 \right\|_n^2 \right] &= O\left(\log(n) n^{-2\beta/(2\beta+1)}\right). \end{aligned}$$

Note that we obtain the exact same rates when considering the risk of $\hat{\sigma}_{m_{\text{opt}}}^2$ defined with the \mathbb{L}^2 -norm equivalent to the empirical norm $\|\cdot\|_n$. Moreover, these rates of convergence are of the same order than the optimal rate $n^{-s/(2s+1)}$ established in Hoffmann (1999b) over a Besov ball.

3.2 Non-adaptive estimation of the diffusion coefficient on the real line

In this section, we propose a ridge estimator of σ^2 on the real line \mathbb{R} , built from one diffusion path. In this context, the main drawback is that the density function $f_n : x \mapsto \frac{1}{n} \sum_{k=0}^{n-1} p_X(k\Delta, x)$ is no longer lower bounded. Consequently, the empirical norm $\|\cdot\|_n$ is no longer equivalent to the L_2 -norm $\|\cdot\|$ and the consistency of the estimation error is no longer ensured under the only assumptions made in the previous sections. Consider the truncated estimator $\hat{\sigma}_{m,L}^2$ of σ^2 given by

$$\hat{\sigma}_{m,L}^2(x) = \hat{\sigma}_m^2(x) \mathbf{1}_{\hat{\sigma}_m^2(x) \leq \sqrt{L}} + \sqrt{L} \mathbf{1}_{\hat{\sigma}_m^2(x) > \sqrt{L}}. \quad (16)$$

Thus, the risk of the ridge estimator $\hat{\sigma}_{m,L}^2$ is upper-bounded as follows:

$$\begin{aligned} \mathbb{E} \left[\|\hat{\sigma}_{m,L}^2 - \sigma^2\|_{n,1}^2 \right] &\leq \mathbb{E} \left[\left\| (\hat{\sigma}_{m,L}^2 - \sigma^2) \mathbf{1}_{[-\log(n), \log(n)]} \right\|_{n,1}^2 \right] + \mathbb{E} \left[\left\| (\hat{\sigma}_{m,L}^2 - \sigma^2) \mathbf{1}_{[-\log(n), \log(n)]^c} \right\|_{n,1}^2 \right] \\ &\leq \mathbb{E} \left[\left\| (\hat{\sigma}_{m,L}^2 - \sigma^2) \mathbf{1}_{[-\log(n), \log(n)]} \right\|_{n,1}^2 \right] + 4 \log^2(n) \sup_{t \in [0,1]} \mathbb{P}(|X_t| > \log(n)). \end{aligned}$$

The first term on the *r.h.s.* is equivalent to the risk of a ridge estimator of σ^2 on the compact interval $[-\log(n), \log(n)]$. The second term on the *r.h.s.* is upper-bounded using Lemma 8.4. We derive below, an upper-bound of the risk of estimation of $\hat{\sigma}_m^2$.

Theorem 3.4. *Suppose that $L = \log^2(n)$. Under Assumption 2.1, it yields,*

$$\mathbb{E} \left[\|\hat{\sigma}_{m,L}^2 - \sigma^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + C \sqrt{\frac{m^q \log^2(n)}{n}}$$

where $C > 0$ is a constant, $q = 1$ for the collection $[\mathbf{B}]$, and $q = 2$ for the collection $[\mathbf{F}]$.

We first remark that the upper-bound of the risk of the truncated estimator of σ^2 differs with respect to each of the chosen bases. This contrast comes from the fact that the Fourier basis $\{f_\ell, \ell = 0, \dots, m-1\}$ and the spline basis $\{B_{\ell-M}, \ell = 0, \dots, m-1\}$ satisfy

$$\sum_{\ell=0}^{m-1} f_\ell(x) \leq C_f m, \quad \text{and} \quad \sum_{\ell=0}^{m-1} B_{\ell-M}(x) = 1.$$

Secondly, the estimation error is not as fine as the one established in Theorem 3.2 where σ^2 is estimated on a compact interval. In fact, on the real line \mathbb{R} , the pseudo-norm $\|\cdot\|_X$ can no longer be equivalent to the L^2 -norm since the transition density is not bounded from below on \mathbb{R} . Consequently, we cannot take advantage of the exact method used to establish the risk bound obtained in Theorem 3.2 which uses the equivalence relation between the pseudo-norms $\|\cdot\|_{n,1}$ and $\|\cdot\|_X$ on one side, and $\|\cdot\|_X$ and the L^2 -norm $\|\cdot\|$ on the other side. Moreover, we can also notice that the term of order $1/n^2$ does not appear since it is dominated by the estimation error.

We obtain below rates of convergence of the ridge estimator of σ^2 for each of the collections $[\mathbf{B}]$ and $[\mathbf{F}]$.

Corollary 3.5. *Suppose that $\sigma^2 \in \Sigma_I(\beta, R)$ with $\beta \geq 1$*

For $[\mathbf{B}]$. *Assume that $K \propto n^{1/(4\beta+1)}$. Under Assumptions 2.1, there exists a constant $C > 0$ depending on β and σ_1 such that*

$$\mathbb{E} \left[\|\hat{\sigma}_{m,L}^2 - \sigma^2\|_{n,1}^2 \right] \leq C \log^{2\beta}(n) n^{-2\beta/(4\beta+1)}.$$

For [F]. Assume that $m \propto n^{1/2(2\beta+1)}$. Under Assumptions 2.1, it yields,

$$\mathbb{E} \left[\left\| \hat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,1}^2 \right] \leq C \log(n) n^{-\beta/(2\beta+1)}$$

where the constant $C > 0$ depends on β and σ_1 .

As we can remark, the obtained rates are slower than the ones established in Section 3.1 where σ^2 is estimated on a compact interval. This result is the immediate consequence of the result of Theorem 3.4.

4 Estimation of the diffusion coefficient from repeated diffusion paths

We now focus on the estimation of the (square) of the diffusion coefficient from i.i.d. discrete observations of the diffusion process ($N \rightarrow \infty$).

4.1 Non-adaptive estimation of the diffusion coefficient on a compact interval

We study the rate of convergence of the ridge estimators $\hat{\sigma}_m^2$ of $\sigma_{|I}^2$ from $D_{N,n}$ when I is a compact interval. The next theorem gives an upper-bound of the risk of our estimators $\hat{\sigma}_m^2$, $m \in \mathcal{M}$.

Theorem 4.1. Suppose that $L = \log(Nn)$ and $\mathcal{M} = \{1, \dots, \sqrt{\min(n, N)}/\log(Nn)\}$. Under Assumption 2.1 and for all $m \in \mathcal{M}$, there exist constants $C > 0$ and $C' > 0$ depending on σ_1 such that,

$$\begin{aligned} \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \right] &\leq 3 \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + C \left(\frac{m}{Nn} + m \log(Nn) \exp \left(-C \sqrt{\min(n, N)} \right) + \Delta_n^2 \right) \\ \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2 \right] &\leq 34 \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + C' \left(\frac{m}{Nn} + m \log(Nn) \exp \left(-C \sqrt{\min(n, N)} \right) + \Delta_n^2 \right). \end{aligned}$$

Note that the result of Theorem 4.1 is independent of the choice of the basis that generate the approximation space \mathcal{S}_m . The first term on the right-hand side represents the approximation error of the initial space, the second term $O(m/(Nn))$ is the estimation error, and the last term characterizes the cost of the time discretization. The next result is derived from Theorem 4.1.

Corollary 4.2. Suppose that $\sigma^2 \in \Sigma_I(\beta, R)$ with $\beta > 3/2$. Moreover, assume that $K_{\text{opt}} \propto (Nn)^{1/(2\beta+1)}$ for [B] ($m_{\text{opt}} = K_{\text{opt}} + M$), and $m_{\text{opt}} \propto (Nn)^{1/(2\beta+1)}$ for [F]. Under Assumptions 2.1, it yields,

$$\begin{aligned} \mathbb{E} \left[\left\| \hat{\sigma}_{m_{\text{opt}}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \right] &= O \left((Nn)^{-2\beta/(2\beta+1)} \right) \\ \mathbb{E} \left[\left\| \hat{\sigma}_{m_{\text{opt}}}^2 - \sigma_{|I}^2 \right\|_n^2 \right] &= O \left((Nn)^{-2\beta/(2\beta+1)} \right). \end{aligned}$$

The obtained result shows that the nonparametric estimators of $\sigma_{|I}^2$ based on repeated observations of the diffusion process are more efficient when $N, n \rightarrow \infty$. Note that the same rate is obtained if the risk of $\hat{\sigma}_{m_{\text{opt}}}^2$ is defined with the \mathbb{L}^2 -norm $\|\cdot\|$ equivalent to the empirical norm $\|\cdot\|_n$.

The rate obtained in Corollary 4.2 is established for $\beta > 3/2$. If we consider for example the collection [B] and assume that $\beta \in [1, 3/2]$, then $K_{\text{opt}} \propto (Nn)^{1/(2\beta+1)}$ belongs to \mathcal{M} for $n \propto \sqrt{N}/\log^4(N)$ and we have

$$\mathbb{E} \left[\left\| \hat{\sigma}_{m_{\text{opt}}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \right] \leq C (Nn)^{-2\beta/(2\beta+1)}.$$

Under the condition $n \propto \sqrt{N}/\log^4(N)$ imposed on the length of diffusion paths, the obtained rate is of order $n^{-3\beta/(2\beta+1)}$ (up to a log-factor) which is equivalent to $N^{-3\beta/2(2\beta+1)}$ (up to a log-factor).

4.2 Non-adaptive estimation of the diffusion coefficient on the real line

Consider a ridge estimator of σ^2 on \mathbb{R} built from N independent copies of the diffusion process X observed in discrete times, where both N and n tend to infinity. For each $m \in \mathcal{M}$, we still denote by $\hat{\sigma}_m^2$ the ridge estimators of σ^2 and $\hat{\sigma}_{m,L}^2$ the truncated estimators of σ^2 given in Equation (16). We establish, through the following theorem, the first risk bound that highlights the main error terms.

Theorem 4.3. *Suppose that $L = \log^2(N)$. Under Assumptions 2.1 and for any dimension $m \in \mathcal{M}$, the following holds:*

$$\mathbb{E} \left[\|\hat{\sigma}_{m,L}^2 - \sigma^2\|_{n,N}^2 \right] \leq 2 \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + C \left(\sqrt{\frac{m^q \log^2(N)}{Nn}} + \Delta_n^2 \right)$$

where $C > 0$ is a constant depending on the upper bound σ_1 of the diffusion coefficient. Moreover, $q = 1$ for the collection $[\mathbf{B}]$ and $q = 2$ for the collection $[\mathbf{H}]$.

If we consider the risk of $\hat{\sigma}_{m,L}^2$ using the empirical norm $\|\cdot\|_n$, then we obtain

$$\mathbb{E} \left[\|\hat{\sigma}_{m,L}^2 - \sigma^2\|_n^2 \right] \leq 2 \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + C \left(\sqrt{\frac{m^q \log^2(N)}{Nn}} + \frac{m^2 \log^3(N)}{N} + \Delta_n^2 \right) \quad (17)$$

The risk bound given in Equation (17) is a sum of four error terms. The first term is the approximation error linked to the choice of the basis, the second term is the estimation error given in Theorem 4.3, the third term $m^2 \log^3(N)/N$ comes from the relation linking the empirical norm $\|\cdot\|_n$ to the pseudo-norm $\|\cdot\|_{n,N}$ (see Lemma 8.7), and the last term is the cost of the time-discretization.

We derive, in the next result, rates of convergence of the risk bound of the truncated ridge estimators $\hat{\sigma}_{m,L}^2$ based on the collections $[\mathbf{B}]$ and $[\mathbf{H}]$ respectively.

Corollary 4.4. *Suppose that $\sigma^2 \in \Sigma_I(\beta, R)$ with $\beta \geq 1$, $I = [-\log(N), \log(N)]$, and $K \propto (Nn)^{1/(4\beta+1)}$ for $[\mathbf{B}]$, and $\sigma^2 \in W_{f_n}^s(\mathbb{R}, R)$ with $s \geq 1$ and $m \propto (Nn)^{1/2(2s+1)}$ for $[\mathbf{H}]$. Under Assumption 2.1, the following holds:*

$$\text{For } [\mathbf{B}] \quad \mathbb{E} \left[\|\hat{\sigma}_{m,L}^2 - \sigma^2\|_{n,N}^2 \right] \leq C \left(\log^{2\beta}(N)(Nn)^{-2\beta/(4\beta+1)} + \frac{1}{n^2} \right),$$

$$\text{For } [\mathbf{H}] \quad \mathbb{E} \left[\|\hat{\sigma}_{m,L}^2 - \sigma^2\|_{n,N}^2 \right] \leq C \left(\log^3(N)(Nn)^{-s/(2s+1)} + \frac{1}{n^2} \right).$$

where $C > 0$ is a constant depending on β and σ_1 for $[\mathbf{B}]$, or s and σ_1 for $[\mathbf{H}]$.

The obtained rates are slower compared to the rates established in Section 4.1 for the estimation of $\sigma_{|I}^2$ where the interval $I \subset \mathbb{R}$ is compact. In fact, the method used to establish the rates of Theorem 4.3 from which the rates of Corollary 4.4 are obtained, does not allow us to derive rates of order $(Nn)^{-\alpha/(2\alpha+1)}$ (up to a log-factor) with $\alpha \geq 1$ (e.g. $\alpha = \beta, s$). Finally, if we consider the risk defined with the empirical norm $\|\cdot\|_n$, then from Equation (17) with $n \propto N$ and assuming that $m \propto N^{1/4(s+1)}$ for $[\mathbf{H}]$ or $K \propto N^{1/4(\beta+1)}$ for $[\mathbf{B}]$, we obtain

$$[\mathbf{B}] : \quad \mathbb{E} \left[\|\hat{\sigma}_{m,L}^2 - \sigma^2\|_n^2 \right] \leq C \log^{2\beta}(N)(Nn)^{-\beta/2(\beta+1)},$$

$$[\mathbf{H}] : \quad \mathbb{E} \left[\|\hat{\sigma}_{m,L}^2 - \sigma^2\|_n^2 \right] \leq C \log^3(N)(Nn)^{-s/2(s+1)},$$

where $C > 0$ is a constant depending on σ_1 and on the smoothness parameter. We can see that the obtained rates are slower compared to the results of Corollary 4.4 for $n \propto N$. The deterioration of the rates comes from the additional term of order $m^2 \log^3(N)/N$ which is now regarded as the new estimation error since it dominates the other term in each case as $N \rightarrow \infty$.

4.3 Non-adaptive estimation of the diffusion coefficient on a compact interval depending on the sample size

This section combines the two first sections 4.1 and 4.2 focusing on the estimation of σ^2 on the compact interval $[-A_N, A_N]$ where (A_N) is a strictly positive sequence such that $A_N \rightarrow \infty$ as $N \rightarrow \infty$. Consequently, we obtain that the estimation interval tends to \mathbb{R} as the sample size N tends to infinity.

Define from the observations and for each dimension $m \in \mathcal{M}$, the following matrices:

$$\widehat{\Psi}_m := \left(\frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \phi_\ell \left(X_{k\Delta}^j \right) \phi_{\ell'} \left(X_{k\Delta}^j \right) \right)_{0 \leq \ell, \ell' \leq m-1}, \quad (18)$$

$$\Psi_m := \mathbb{E} \left(\widehat{\Psi}_m \right) = \left(\mathbb{E} \left[\frac{1}{n} \sum_{k=0}^{n-1} \phi_\ell \left(X_{k\Delta} \right) \phi_{\ell'} \left(X_{k\Delta} \right) \right] \right)_{0 \leq \ell, \ell' \leq m-1}. \quad (19)$$

These two matrices play an essential role in the construction of a consistent projection estimator of σ^2 over any approximation subspace \mathcal{S}_m spanned by the basis $(\phi_0, \dots, \phi_{m-1})$. Furthermore, for all $h = \sum_{\ell=0}^{m-1} a_\ell \phi_\ell \in \mathcal{S}_m$, we have:

$$\|h\|_{n,N}^2 = {}^t \mathbf{a} \widehat{\Psi}_m \mathbf{a}, \quad \|h\|_n^2 = \mathbb{E} (\|h\|_{n,N}^2) = {}^t \mathbf{a} \Psi_m \mathbf{a},$$

where $\mathbf{a} = (a_0, \dots, a_{m-1})$. The Gram matrix Ψ_m is invertible under the spline basis (see Denis *et al.* (2022)) and the Hermite basis (see Comte & Genon-Catalot (2021)). We define for any invertible matrix M , the operator norm $\|M^{-1}\|_{\text{op}}$ of M^{-1} given by $\|M^{-1}\|_{\text{op}} = 1/\inf\{\lambda_j\}$ where the λ_j are eigenvalues of M .

For all dimension $m \in \mathcal{M}$, the matrices $\widehat{\Psi}_m$ and \mathbf{F}_m satisfy:

$$\widehat{\Psi}_m = {}^t \mathbf{F}_m \mathbf{F}_m.$$

Consider the ridge estimator $\widehat{\sigma}_m^2$ of $\sigma_{A_N}^2 = \sigma^2 \mathbf{1}_{[-A_N, A_N]}$, with $m \in \mathcal{M}$ and $A_N \rightarrow \infty$ as $N \rightarrow \infty$. The estimator $\widehat{\sigma}_m^2$ can reach a faster rate of convergence if the Gram matrix Ψ_m given in Equation (19) satisfies the following condition,

$$\mathcal{L}(m) \left(\|\Psi_m^{-1}\|_{\text{op}} \vee 1 \right) \leq C \frac{N}{\log^2(N)}, \quad \text{where } \mathcal{L}(m) := \sup_{x \in \mathbb{R}} \sum_{\ell=0}^{m-1} \phi_\ell^2(x) < \infty \quad (20)$$

where $C > 0$ is a constant. In fact, the optimal rate of convergence is achieved on a random event $\Omega_{n,N,m}$ in which the two empirical norms $\|\cdot\|_{n,N}$ and $\|\cdot\|_n$ are equivalent (see Comte & Genon-Catalot (2020a), Denis *et al.* (2021)). Then, Condition (20) is used to upper-bound $\mathbb{P} \left(\Omega_{n,N,m}^c \right)$ by a negligible term with respect to the considered rate (see Comte & Genon-Catalot (2020a)). Note that in Equation (20), the square on $\log(N)$ is justified by the fact that the value of constant $C > 0$ is unknown, and that the spline basis is not orthonormal (see Denis *et al.* (2022), proof of Lemma 7.8). The assumption of Equation (20) is also made in Comte & Genon-Catalot (2020a) on the operator norm of Ψ_m^{-1} based on an orthonormal basis with the bound $\mathbf{c}N/\log(N)$ where the value of \mathbf{c} is known, and chosen and such that the upper-bound of $\mathbb{P}(\Omega_{n,N,m}^c)$ is negligible with respect to the estimation error. In our framework, since the transition density is approximated by Gaussian densities, we derive the following result.

Lemma 4.5. *Suppose that $n \propto N$ and that the spline basis is constructed on the interval $[-A_N, A_N]$ with $A_N > 0$. Under Assumption 2.1, for all $m \in \mathcal{M}$ and for all $w \in \mathbb{R}^m$ such that $\|w\|_{2,m} = 1$, there exists a constant $C > 0$ such that*

$$\begin{aligned} \text{For } [\mathbf{H}] : \quad w' \Psi_m w &\geq \frac{C}{\log(N)} \exp\left(-\frac{3c_\sigma(4m+3)}{2(1-\log^{-1}(N))}\right), \\ \text{For } [\mathbf{B}] : \quad w' \Psi_m w &\geq \frac{CA_N}{m \log(N)} \exp(-c_\sigma A_N^2), \end{aligned}$$

where the constant $c_\sigma > 1$ that comes from the approximation of the transition density, depends on the diffusion coefficient σ .

The result of Lemma 4.5 implies for the Hermite basis that

$$\left(\|\Psi_m^{-1}\|_{\text{op}} \vee 1\right) \leq \frac{\log(N)}{C} \exp\left(\frac{3c_\sigma(4m+3)}{2(1-\log^{-1}(N))}\right)$$

where the upper-bound is an exponentially increasing sequence of N since the dimension $m \in \mathcal{M}$ has a polynomial growth with respect to N . Thus, Condition (20) cannot be satisfied for the Hermite basis in our framework. Considering the spline basis, one has $\mathcal{L}(m) = \mathcal{L}(K+M) \leq 1$ and there exists a constant $C > 0$ such that

$$\|\Psi_m^{-1}\|_{\text{op}} \leq C \frac{m \log(N)}{A_N} \exp(c_\sigma A_N^2). \quad (21)$$

For $K \propto (N^{2/(2\beta+1)} A_N)$, Condition (20) is satisfied if the estimation interval $[-A_N, A_N]$ is chosen such that $A_N = o\left(\sqrt{\log(N)}\right)$. In the next theorem, we prove that the spline-based ridge estimator of $\sigma_{A_N}^2$ reaches a faster rate of convergence compared to the result of Corollary 4.4 for the collection $[\mathbf{B}]$.

Theorem 4.6. *Suppose that $N \propto n$ and consider the ridge estimator $\hat{\sigma}_{A_N, m}^2$ of $\sigma_{A_N}^2$ based on the spline basis. Furthermore, suppose that $L = \log(N)$, $A_N = o\left(\sqrt{\log(N)}\right)$ and $K \propto (Nn)^{1/(2\beta+1)} A_N$ ($m = K + M$). Under Assumptions 2.1 and for $\sigma^2 \in \Sigma_I(\beta, R)$ with $I = [-A_N, A_N]$, the following holds:*

$$\mathbb{E} \left[\|\hat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \right] \leq C \log^\beta(N) (Nn)^{-2\beta/(2\beta+1)}$$

where $C > 0$ is a constant depending on β .

The above result shows that the risk of the ridge estimator of $\sigma_{A_N}^2$ on $[-A_N, A_N]$ reaches a rate of order $(Nn)^{-\beta/(2\beta+1)}$ (up to a log-factor) thanks to Condition (20) which allows us to take advantage of the equivalence relation between the empirical norms $\|\cdot\|_n$ and $\|\cdot\|_{n, N}$ given in Equation (3) to derive a finer estimation error (see proof of Theorem 4.6). Note that the obtained result depends on an appropriate choice of the estimation interval $[-A_N, A_N]$ which tends to \mathbb{R} as N tends to infinity. Therefore, any choice of A_N such that $A_N/\sqrt{\log(N)} \rightarrow +\infty$ cannot lead to a consistent estimation error since Equation (20) is no longer satisfied for the upper-bounding of $\mathbb{P}\left(\Omega_{n, N, m}^c\right)$ by a term that tends to zero as $N \rightarrow \infty$. Thus, the assumption $A_N = o\left(\sqrt{\log(N)}\right)$ is a necessary and sufficient condition for the validation of Condition (20) which leads, together with Assumption 2.1, to the result of Theorem 4.6. Finally, under the assumptions of Theorem 4.6 and considering the risk of $\hat{\sigma}_{A_N, m}^2$ based on the empirical norm $\|\cdot\|_n$, we also obtain

$$\mathbb{E} \left[\|\hat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_n^2 \right] = \mathcal{O}\left(\log^\beta(N) (Nn)^{-2\beta/(2\beta+1)}\right).$$

In fact, under Condition (20), the estimator $\widehat{\sigma}_{A_N, m}^2$ satisfies the results of Theorem 4.1 with $I = [-A_N, A_N]$ and $A_N = o\left(\sqrt{\log(N)}\right)$, which implies rates of the same order for the two empirical norms.

5 Adaptive estimation of the diffusion coefficient from repeated observations

In this section, we suppose that $n \propto N$ and we propose a adaptive ridge estimator of σ^2 by selecting an optimal dimension from the sample D_N . In fact, consider the estimator $\widehat{\sigma}_{\widehat{K}, L}^2$ where \widehat{K} satisfies:

$$\widehat{K} := \arg \min_{K \in \mathcal{K}} \{ \gamma_{n, N}(\widehat{\sigma}_K^2) + \text{pen}(K) \} \quad (22)$$

and the penalty function $\text{pen} : K \mapsto \text{pen}(K)$ is established using the chaining technique of Baraud *et al.* (2001). We derive below the risk of the adaptive estimator of $\sigma_{|I}^2$ when the interval $I \subset \mathbb{R}$ is compact and the sample size $N \rightarrow \infty$.

Theorem 5.1. *Suppose that $N \propto n$, $L = \log(N)$ and consider the collection $[\mathbf{B}]$ with*

$$K \in \mathcal{K} = \{2^q, q = 0, 1, \dots, q_{\max}\} \subset \mathcal{M} = \{1, \dots, \sqrt{N}/\log(N)\}.$$

Under Assumption 2.1, there exists a constant $C > 0$ such that,

$$\mathbb{E} \left[\left\| \widehat{\sigma}_{\widehat{K}, L}^2 - \sigma_{|I}^2 \right\|_{n, N}^2 \right] \leq 34 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M, L}} \|h - \sigma_{|I}^2\|_n^2 + \text{pen}(K) \right\} + \frac{C}{Nn}$$

where $\text{pen}(K) = \kappa(K+M)\log(N)/Nn$ with $\kappa > 0$ a numerical constant.

We deduce from Corollary 4.2 and its assumptions that the adaptive estimator $\widehat{\sigma}_{\widehat{K}, L}^2$ satisfies:

$$\mathbb{E} \left[\left\| \widehat{\sigma}_{\widehat{K}, L}^2 - \sigma_{|I}^2 \right\|_n^2 \right] = \mathcal{O} \left((Nn)^{-2\beta/(2\beta+1)} \right).$$

This result is justified since the penalty term is of the same order (up to a log-factor) than the estimation error established in Theorem 4.1.

Considering the adaptive estimator of σ^2 on the real line $I = \mathbb{R}$ when the sample size $N \rightarrow \infty$, we obtain the following result.

Theorem 5.2. *Suppose that $N \propto n$ and $L = \log(N)$, and consider the collection $[\mathbf{B}]$ with*

$$K \in \mathcal{K} = \{2^q, q = 0, 1, \dots, q_{\max}\} \subset \mathcal{M} = \{1, \dots, \sqrt{N}/\log(N)\}.$$

Under Assumption 2.1 and for N large enough, there exists a constant $C > 0$ such that,

$$\mathbb{E} \left[\left\| \widehat{\sigma}_{\widehat{K}, L}^2 - \sigma^2 \right\|_{n, N}^2 \right] \leq 3 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M, L}} \|h - \sigma^2\|_n^2 + \text{pen}(K) \right\} + \frac{C}{Nn}.$$

where $\text{pen}(K) = \kappa' \frac{(K+M)\log(N)}{Nn}$ with $\kappa' > 0$ a numerical constant.

We have a penalty term of the same order than the one obtained in Theorem 5.1 where σ^2 is estimated on a compact interval. One can deduce that the adaptive estimator reaches a rate of the same order than the rate of the non-adaptive estimator given in Corollary 4.4 for the collection $[\mathbf{B}]$.

If we consider the adaptive estimator of the compactly supported diffusion coefficient built from a single diffusion path, we obtain below an upper-bound of its risk of estimation.

Theorem 5.3. *Suppose that $N = 1$, $L = \sqrt{\log(n)}$ and consider the collection $[\mathbf{B}]$ with*

$$K \in \mathcal{K} = \{2^q, q = 0, \dots, q_{\max}\} \subset \mathcal{M} = \{1, \dots, \sqrt{n}/\log(n)\}.$$

Under Assumption 2.1, it yields

$$\mathbb{E} \left[\left\| \widehat{\sigma}_{\widehat{K}, L}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] \leq 3 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M, L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + \text{pen}(K) \right\} + \frac{C}{n}.$$

where $C > 0$ is a constant depending on τ_0 , and $\text{pen}(K) = \kappa \frac{(K+M)\log(n)}{n}$ with $\kappa > 0$ a numerical constant.

We deduce from Theorem 5.3 that if we assume that $\sigma^2 \in \Sigma_I(\beta, R)$, then the adaptive estimator $\widehat{\sigma}_{\widehat{K}, L}^2$ reaches a rate of order $n^{-\beta/(2\beta+1)}$ (up to a log-factor). The result of this theorem is almost a deduction of the result of Theorem 5.1, the slight difference being the use, in the proofs, of the local time of the process and the equivalence relation between the pseudo-norm $\|\cdot\|_{n,1}$ with the pseudo-norm $\|\cdot\|_X$ instead of the empirical norm $\|\cdot\|_n$ considered in the proof of Theorem 5.1.

6 Numerical study

This section is devoted to the numerical study on a simulation scheme. Section 6.1 focuses on the presentation of the chosen diffusion models. In Section 6.2, we describe the scheme for the implementation of the ridge estimators. We mainly focus on the \mathbf{B} -spline basis for the numerical study, and in Section 6.3, we add a numerical study on the performance of the Hermite-based ridge estimator of σ^2 on \mathbb{R} . Finally, we compare the efficiency of our estimator built on the real line \mathbb{R} from a single path with that of the Nadaraya-Watson estimator proposed in Florens-Zmirou (1993).

6.1 Models and simulations

Recall that the time horizon is $T = 1$ and $X_0 = 0$. Consider the following diffusion models:

Model 1 Ornstein-Uhlenbeck: $b(x) = 1 - x$, $\sigma(x) = 1$

Model 2: $b(x) = 1 - x$, $\sigma(x) = 1 - x^2$

Model 3: $b(x) = 1 - x$, $\sigma(x) = \frac{1}{3 + \sin(2\pi x)} + \cos^2\left(\frac{\pi}{2}x\right)$

Model 1 is the commonly used Ornstein-Uhlenbeck model, known to be a simple diffusion model satisfying Assumption 2.1. Model 2 does not satisfy Assumption 2.1. Model 3 satisfies Assumption 2.1 with a multimodal diffusion coefficient.

The size N of the sample D_N takes values in the set $\{1, 10, 100, 1000\}$ where the length n of paths varies in the set $\{100, 250, 500, 1000\}$. As we work with the spline basis, the dimension $m = K + M$ of the approximation space is chosen such that $M = 3$ and K takes values in $\mathcal{K} = \{2^p, p = 0, \dots, 5\}$ so that the subspaces are nested inside each other. We are using \mathbf{R} for the simulation of diffusion paths via the function `sde.sim` of `sde` package, (see Iacus (2009) for more details on the simulation of SDEs).

6.2 Implementation of the ridge estimators

In this section, we assess the quality of estimation of the adaptive estimator $\hat{\sigma}_m^2$ in each of the 3 models through the computation of its risk of estimation. We compare the performance of the adaptive estimator with that of the oracle estimator $\hat{\sigma}_{m^*}^2$ where m^* is given by:

$$m^* := \arg \min_{m \in \mathcal{M}} \|\hat{\sigma}_m^2 - \sigma^2\|_{n,N}^2. \quad (23)$$

For the spline basis, we have $m^* = K^* + M$ with $M = 3$. Finally, we complete the numerical study with a representation of a set of 10 estimators of σ^2 for each of the 3 models.

We evaluate the MISE of the spline-based adaptive estimators $\hat{\sigma}_{\hat{K}}^2$ by repeating 100 times the following steps:

1. Simulate samples $D_{N,n}$ and $D_{N',n}$ with $N \in \{1, 10, 100, 1000\}$, $N' = 100$ and $n \in \{100, 250, 1000\}$.
2. For each $K \in \mathcal{K}$, and from $D_{N,n}$, compute estimators $\hat{\sigma}_K^2$ given in Equations (9) and (10).
3. Select the optimal dimension $\hat{K} \in \mathcal{K}$ using Equation (22) and compute K^* from Equation (23)
4. Using $D_{N',n}$, evaluate $\|\hat{\sigma}_{\hat{K}}^2 - \sigma^2\|_{n,N'}^2$ and $\|\hat{\sigma}_{K^*}^2 - \sigma^2\|_{n,N'}^2$.

We deduce the risks of estimation considering the average values of $\|\hat{\sigma}_{\hat{m}}^2 - \sigma^2\|_{n,N'}^2$ and $\|\hat{\sigma}_{m^*}^2 - \sigma^2\|_{n,N'}^2$ over the 100 repetitions. Note that we consider in this section, the estimation of σ^2 on the compact interval $I = [-1, 1]$ and on the real line \mathbb{R} . The unknown parameters κ and κ' in the penalty functions given in Theorem 5.1 and Theorem 5.2 respectively, are numerically calibrated (details are given in Appendix 8.4.3), and we choose $\kappa = 4$ and $\kappa' = 5$ as their respective values.

6.3 Numerical results

We present in this section the numerical results of the performance of the spline-based adaptive estimators of $\sigma^2|_I$ with $I \subseteq \mathbb{R}$ together with the performance of the oracle estimators. We consider the case $I = [-1, 1]$ for the compactly supported diffusion coefficient, and the case $I = \mathbb{R}$.

Tables 2 and 3 present the numerical results of estimation of $\sigma^2|_I$ from simulated data following the steps given in Section 6.2.

The results of Table 2 and Table 3 show that the adapted estimator $\hat{\sigma}_{\hat{K}}^2$ is consistent, since its MISE tends to zero as both the size N of the sample $D_{N,n}$ and the length n of paths are larger. Moreover, note that in most cases, the ridge estimators of the compactly supported diffusion coefficients perform better than those of the non-compactly supported diffusion functions. As expected, we observe that the oracle estimator has generally a better performance compared to the adaptive estimator. Nonetheless, we can remark that the performances are very close in several cases, highlighting the efficiency of the data-driven selection of the dimension.

An additional important remark is the significant influence of the length n of paths on the performance of $\hat{\sigma}_{\hat{K}}^2$ and $\hat{\sigma}_{K^*,L}^2$ (by comparison of Table 2 with Table 3), which means that estimators built from higher frequency data are more efficient. A similar remark is made for theoretical results obtained in Sections 4.2 and 4.1.

Performance of the Hermite-based estimator of the diffusion coefficient We focus on the estimation of σ^2 on \mathbb{R} and assess the performance of its Hermite-based estimator (see Section 4.2). We present in Table 4, the performance of the oracle estimator $\hat{\sigma}_{m^*,L}^2$.

From the numerical results of Table 4, we observe that the Hermite-based estimator of σ^2 is consistent as the sample size N and the length n paths take larger values.

Models	Intervals	Estimators	$N = 10$	$N = 100$	$N = 1000$
Model 1	[-1, 1]	$\hat{\sigma}_{\hat{K},L}^2$	0.0102 (0.0083)	0.0009 (0.0009)	0.0002 (0.0001)
		$\hat{\sigma}_{K^*,L}^2$	0.0094 (0.0065)	0.0009 (0.0009)	0.0002 (0.0001)
	\mathbb{R}	$\hat{\sigma}_{\hat{K},L}^2$	0.0096 (0.0062)	0.0009 (0.0008)	0.0003 (0.0002)
		$\hat{\sigma}_{K^*,L}^2$	0.0093 (0.0057)	0.0009 (0.0008)	0.0003 (0.0002)
Model 2	[-1, 1]	$\hat{\sigma}_{\hat{K},L}^2$	0.0048 (0.0052)	0.0019 (0.0008)	0.0005 (0.0002)
		$\hat{\sigma}_{K^*,L}^2$	0.0039 (0.0043)	0.0009 (0.0005)	0.0005 (0.0002)
	\mathbb{R}	$\hat{\sigma}_{\hat{K},L}^2$	0.0195 (0.0140)	0.0057 (0.0006)	0.0012 (0.0002)
		$\hat{\sigma}_{K^*,L}^2$	0.0048 (0.0064)	0.0025 (0.0021)	0.0010 (0.0003)
Model 3	[-1, 1]	$\hat{\sigma}_{\hat{K},L}^2$	0.0521 (0.0191)	0.0176 (0.0070)	0.0073 (0.0021)
		$\hat{\sigma}_{K^*,L}^2$	0.0260 (0.0081)	0.0073 (0.0030)	0.0048 (0.0009)
	\mathbb{R}	$\hat{\sigma}_{\hat{K},L}^2$	0.1132 (0.0595)	0.0319 (0.0031)	0.0179 (0.0054)
		$\hat{\sigma}_{K^*,L}^2$	0.0351 (0.0169)	0.0319 (0.0031)	0.0116 (0.0051)

Table 2: Assessment of MISEs (mean and standard deviation between brackets) of both the adaptive estimator $\hat{\sigma}_{\hat{K},L}^2$ and the oracle estimator $\hat{\sigma}_{K^*,L}^2$ from diffusion paths of size $n = 100$.

Models	Intervals	Estimators	$N = 10$	$N = 100$	$N = 1000$
Model 1	[-1, 1]	$\hat{\sigma}_{\hat{K},L}^2$	0.0047 (0.0037)	0.0003 (0.0002)	0.0001 (0.00003)
		$\hat{\sigma}_{K^*,L}^2$	0.0042 (0.0030)	0.0003 (0.0002)	0.0001 (0.00003)
	\mathbb{R}	$\hat{\sigma}_{\hat{K},L}^2$	0.0053 (0.0037)	0.0003 (0.0002)	0.0001 (0.00004)
		$\hat{\sigma}_{K^*,L}^2$	0.0050 (0.0031)	0.0003 (0.0002)	0.0001 (0.00004)
Model 2	[-1, 1]	$\hat{\sigma}_{\hat{K},L}^2$	0.0027 (0.0019)	0.0003 (0.0002)	0.0002 (0.00004)
		$\hat{\sigma}_{K^*,L}^2$	0.0018 (0.0019)	0.0002 (0.0001)	0.0001 (0.00004)
	\mathbb{R}	$\hat{\sigma}_{\hat{K},L}^2$	0.0091 (0.0077)	0.0028 (0.0025)	0.0008 (0.0002)
		$\hat{\sigma}_{K^*,L}^2$	0.0020 (0.0023)	0.0021 (0.0023)	0.0002 (0.00004)
Model 3	[-1, 1]	$\hat{\sigma}_{\hat{K},L}^2$	0.0306 (0.0150)	0.0058 (0.0012)	0.0010 (0.0003)
		$\hat{\sigma}_{K^*,L}^2$	0.0216 (0.0067)	0.0023 (0.0020)	0.0010 (0.0003)
	\mathbb{R}	$\hat{\sigma}_{\hat{K},L}^2$	0.0560 (0.0313)	0.0275 (0.0049)	0.0069 (0.0049)
		$\hat{\sigma}_{K^*,L}^2$	0.0261 (0.0127)	0.0096 (0.0051)	0.0065 (0.0041)

Table 3: Assessment of MISEs of both the adaptive estimator $\hat{\sigma}_{\hat{K},L}^2$ and the oracle estimator $\hat{\sigma}_{K^*,L}^2$ from diffusion paths of size $n = 250$.

Models	Intervals	Estimators	$N = 10, n = 100$	$N = 100, n = 100$	$N = 100, n = 250$
Model 1	\mathbb{R}	$\hat{\sigma}_{K^*,L}^2$	0.0082 (0.0059)	0.0015 (0.0008)	0.0006 (0.0004)
Model 2	\mathbb{R}	$\hat{\sigma}_{K^*,L}^2$	0.0058 (0.0111)	0.0007 (0.0004)	0.0003 (0.0002)
Model 3	\mathbb{R}	$\hat{\sigma}_{K^*,L}^2$	0.0188 (0.0151)	0.0077 (0.0037)	0.0040 (0.0036)

Table 4: Assessment of MISEs of the Hermite-based oracle estimator $\hat{\sigma}_{K^*,L}^2$ of the square of the diffusion coefficient.

Estimation of the diffusion coefficient from one path Consider ridge estimators of $\sigma_{|I}^2$ with $I = [-1, 1]$. For the case of the adaptive estimators of $\sigma_{|I}^2$, the dimension \hat{K} is selected such that

$$\hat{K} = \arg \min_{K \in \mathcal{K}} \gamma_n(\hat{\sigma}_K^2) + \text{pen}(K) \quad (24)$$

where $\text{pen}(K) = \kappa(K + M) \log(n)/n$ with $\kappa > 0$. We choose the numerical constant $\kappa = 4$ and we derive the numerical performance of the adaptive estimator of $\sigma_{|I}^2$.

Models	Intervals	Estimators	$n = 100$	$n = 1000$
Model 1	[-1, 1]	$\hat{\sigma}_{\hat{K},L}^2$	0.1751 (0.1921)	0.0915 (0.1925)
		$\hat{\sigma}_{K^*,L}^2$	0.1563 (0.1776)	0.0783 (0.1699)
Model 2	[-1, 1]	$\hat{\sigma}_{\hat{K},L}^2$	0.1721 (0.3483)	0.1365 (0.5905)
		$\hat{\sigma}_{K^*,L}^2$	0.0987 (0.1644)	0.0552 (0.2409)
Model 3	[-1, 1]	$\hat{\sigma}_{\hat{K},L}^2$	0.2184 (0.2780)	0.2106 (0.5790)
		$\hat{\sigma}_{K^*,L}^2$	0.1263 (0.1486)	0.0751 (0.1469)

Table 5: Evaluation of MISEs of adaptive estimators $\hat{\sigma}_{\hat{K},L}^2$ built from a single diffusion path ($N = 1$) for each of the three models.

Table 5 gives the numerical performances of both the adaptive estimator and the oracle estimator of $\sigma_{|I}^2$ on the compact interval $I = [-1, 1]$ and from a single diffusion path. From the obtained results, we see that the estimators are numerically consistent. However, we note that the convergence is slow (increasing n from 100 to 1000), which highlights the significant impact of the number N of paths on the efficiency of the ridge estimator.

Comparison of the efficiency of the ridge estimator of the diffusion coefficient with its Nadaraya-Watson estimator. Consider the adaptive estimator $\hat{\sigma}_{\hat{K}}^2$ of the square of the diffusion coefficient built on the real line \mathbb{R} from a single diffusion path ($N = 1$), where the dimension \hat{K} is selected using Equation (24). For the numerical assessment, we use the interval $I = [-10^6, 10^6]$ to approximate the real line \mathbb{R} , and then, use Equation (24) for the data-driven selection of the dimension.

We want to compare the efficiency of $\hat{\sigma}_{\hat{K}}^2$ with that of the Nadaraya-Watson estimator of σ^2 given

from a diffusion path $\bar{X} = (X_{k/n})_{1 \leq k \leq n}$ and for all $x \in \mathbb{R}$ by

$$S_n(x) = \frac{\sum_{k=1}^{n-1} K\left(\frac{X_{k/n} - x}{h_n}\right) [X_{(k+1)/n} - X_{k/n}]^2/n}{\sum_{k=1}^n K\left(\frac{X_{k/n} - x}{h_n}\right)}$$

where K is a positive kernel function, and h_n is the bandwidth. Thus, the estimator $S_n(x)$ is consistent under the condition $nh_n^4 \rightarrow 0$ as n tends to infinity (see Florens-Zmirou (1993)). We use the function `ksdiff()` of the R-package `sde` to compute the Nadaraya-Watson estimator S_n .

Models	Ridge estimator	Nadaraya-Watson estimator
Model 1	0.0020 (0.0023)	0.9377 (0.0017)
Model 2	0.1323 (0.0794)	0.5086 (0.0885)
Model 3	0.4077 (0.1178)	1.3175 (0.3039)

Table 6: This table shows the loss errors of the ridge estimator $\hat{\sigma}_{\hat{K},L}^2$ on \mathbb{R} and the Nadaraya-Watson estimator S_n of σ^2 built from a diffusion path ($N = 1$) of length $n = 1000$.

We remark from the results of Table 6 that our ridge estimator is more efficient. Note that for the kernel estimator S_n , the bandwidth is computed using the rule of thumb of Scott (see Odell-Scott (1992)). The bandwidth is proportional to $n^{-1/(d+4)}$ where n is the number of points, and d is the number of spatial dimensions.

6.4 Concluding remarks

The results of our numerical study show that our ridge estimators built both on a compact interval and on the real line are consistent as N and n take larger values, or as only n takes larger values when the estimators are built from a single path. These results are in accordance with the theoretical results established in the previous sections. Moreover, as expected, we obtained the consistency of the Hermite-based estimators of σ^2 on the real line \mathbb{R} . Nonetheless, we only focus on the Hermite-based oracle estimator since we did not establish a risk bound of the corresponding adaptive estimator. Finally, we remark that the ridge estimator of σ^2 built from a single path performs better than its Nadaraya-Watson kernel estimator proposed in Florens-Zmirou (1993) and implemented in the R-package `sde`.

7 Conclusion

In this article, we have proposed ridge-type estimators of the diffusion coefficient on a compact interval from a single diffusion path. We took advantage of the local time of the diffusion process to prove the consistency of non-adaptive estimators of σ^2 and derive a rate of convergence of the same order than the optimal rate established in Hoffmann (1999b). We also propose an estimator of σ^2 on the real line from a single path. We proved its consistency using the method described in Section 3.2, and derive a rate of convergence order $n^{-\beta/(4\beta+1)}$ over a Hölder space for the collection $[\mathbf{B}]$. Then, we extended the study to the estimation of σ^2 from repeated discrete observations of the diffusion process. We establish rates of convergence of the ridge estimators both on a compact interval and on \mathbb{R} . We complete the study proposing adaptive estimators of σ^2 on a compact interval for $N = 1$ and $N \rightarrow \infty$, and on the real line \mathbb{R} for $N \rightarrow \infty$.

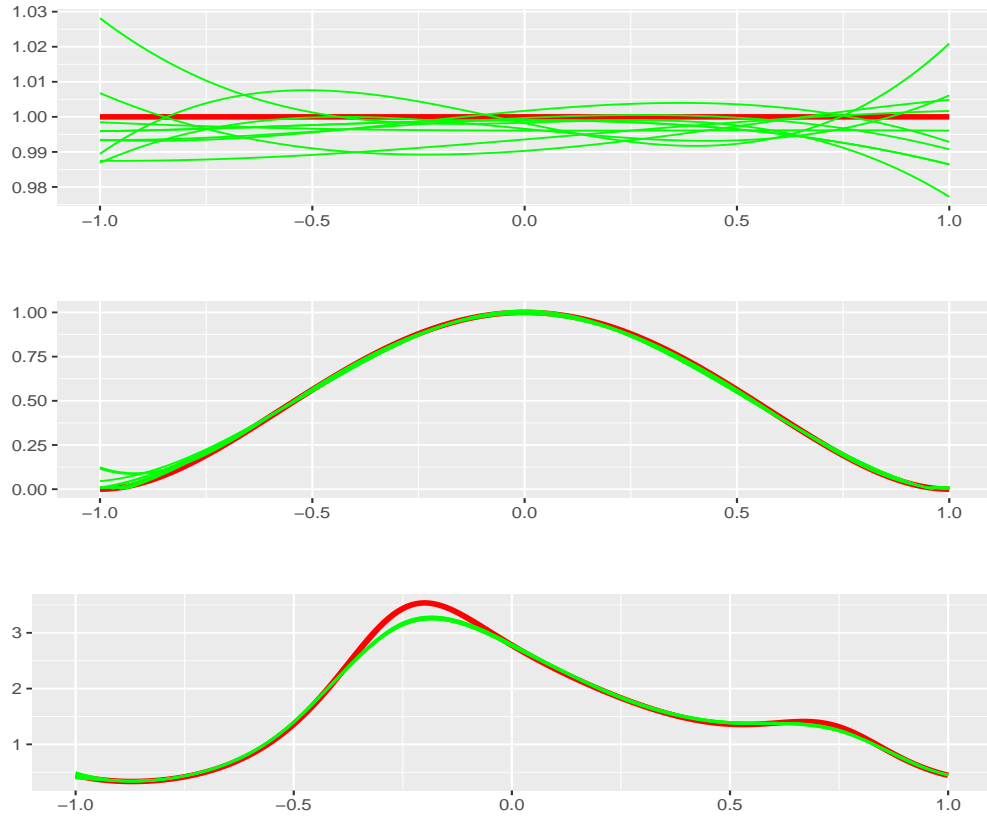


Figure 1: Bundles of 10 estimators $\hat{\sigma}_{\hat{K},L}^2$ (in green) of the true diffusion coefficient $\sigma^2|_I$ restricted on the compact interval $I = [-1, 1]$ (in red) of each of Models 1, 2, 3 (from top to bottom) using samples of size $N = 1000$ with diffusion paths of length $n = 500$.

A perspective on the estimation of the diffusion coefficient could be the establishment of a minimax rate of convergence of the compactly supported (square of the) diffusion coefficient from repeated discrete observations of the diffusion process. The case of the non-compactly supported diffusion coefficient may be a lot more challenging, since the transition density of the diffusion process is no longer lower-bounded. This new fact can lead to different rates of convergence depending on the considered method (see Section 4).

Acknowledgements

I would like to thank my supervisors, Christophe Denis, Charlotte Dion-Blanc, and Viet-Chi Tran, for their sound advice, guidance and support throughout this research project.

8 Proofs

In this section, we prove our main results of Sections 3, 4 and 5. To simplify our notations, we set $\Delta_n = \Delta (= 1/n)$ and constants are generally denoted by $C > 0$ or $c > 0$ whose values can change from a line to another. Moreover, we use the notation C_α in case we need to specify the dependency of the constant C on a parameter α .

8.1 Technical results

Recall first some useful results on the local time and estimates of the transition density of diffusion processes.

Lemma 8.1. *For all integer $q \geq 1$, there exists $C^* > 0$ depending on q such that for all $0 \leq s < t \leq 1$,*

$$\mathbb{E} \left[|X_t - X_s|^{2q} \right] \leq C^* (t - s)^q.$$

The proof of Lemma 8.1 is provided in Denis *et al.* (2022).

Proposition 8.2. *Under Assumptions 2.1, there exist constants $c_\sigma > 1$, $C > 1$ such that for all $t \in (0, 1]$, $x \in \mathbb{R}$,*

$$\frac{1}{C\sqrt{t}} \exp\left(-c_\sigma \frac{x^2}{t}\right) \leq p_X(t, x) \leq \frac{C}{\sqrt{t}} \exp\left(-\frac{x^2}{c_\sigma t}\right).$$

The proof of Proposition 8.2 is provided in Gobet (2002), *Proposition 1.2*.

Proposition 8.3. *Let h be a L_0 -lipschitz function. Then there exists $\tilde{h} \in \mathcal{S}_{K_N, M}$, such that*

$$|\tilde{h}(x) - h(x)| \leq C \frac{\log(N)}{K_N}, \quad \forall x \in (-\log(N), \log(N)),$$

where $C > 0$ depends on L_0 , and M .

The proof of Proposition 8.3 is provided in Denis *et al.* (2022). The finite-dimensional vector space $\mathcal{S}_{K_N, M} = \mathcal{S}_{K_N + M}$ is introduced in Section 2.

Lemma 8.4. *Under Assumption 2.1, there exist $C_1, C_2 > 0$ such that for all $A > 0$,*

$$\sup_{t \in [0, 1]} \mathbb{P}(|X_t| \geq A) \leq \frac{C_1}{A} \exp(-C_2 A^2).$$

The proof of Lemma 8.4 is provided in Denis *et al.* (2022), Lemma 7.3.

Lemma 8.5. *Under Assumption 2.1, the following holds:*

$$\forall x \in \mathbb{R}, \quad \mathcal{L}^x = \mathcal{L}^{x-} \quad a.s.$$

where $\mathcal{L}^{x-} = \lim_{\varepsilon \rightarrow 0} \mathcal{L}^{x-\varepsilon}$.

The result of Lemma 8.5 justifies the definition of the local time \mathcal{L}^x , for $x \in \mathbb{R}$, given in Equation (14).

Proof. From Revuz & Yor (2013), Theorem 1.7, we have

$$\forall x \in \mathbb{R}, \quad \mathcal{L}^x - \mathcal{L}^{x-} = 2 \int_0^1 \mathbb{1}_{X_s=x} dX_s = 2 \int_0^1 \mathbb{1}_{X_s=x} b(X_s) ds + 2 \int_0^1 \mathbb{1}_{X_s=x} \sigma(X_s) dW_s.$$

For all $x \in \mathbb{R}$ and for all $s \in [0, 1]$, we have for all $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}(X_s = x) &= \lim_{\varepsilon \rightarrow 0} \mathbb{P}(X_s \leq x + \varepsilon) - \lim_{\varepsilon \rightarrow 0} \mathbb{P}(X_s \leq x - \varepsilon) = \lim_{\varepsilon \rightarrow 0} F_s(x + \varepsilon) - \lim_{\varepsilon \rightarrow 0} F_s(x - \varepsilon) \\ &= F_s(x) - F_s(x^-) \\ &= 0 \end{aligned}$$

Thus, for all $x \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E} [|\mathcal{L}^x - \mathcal{L}^{x-}|] &\leq 2 \int_0^1 |b(x)| \mathbb{P}(X_s = x) ds + 2 \mathbb{E} \left[\left| \int_0^1 \mathbb{1}_{X_s=x} \sigma(X_s) dW_s \right| \right] \\ &= 2 \mathbb{E} \left[\left| \int_0^1 \mathbb{1}_{X_s=x} \sigma(X_s) dW_s \right| \right]. \end{aligned}$$

Using the Cauchy Schwartz inequality, we conclude that

$$\mathbb{E} [|\mathcal{L}^x - \mathcal{L}^{x-}|] \leq 2 \sqrt{\mathbb{E} \left(\int_0^1 \mathbb{1}_{X_s=x} \sigma^2(X_s) ds \right)} = 2\sigma(x) \int_0^1 \mathbb{P}(X_s = x) ds = 0.$$

Using the Markov inequality, we have

$$\forall \varepsilon > 0, \quad \mathbb{P}(|\mathcal{L}^x - \mathcal{L}^{x-}| > \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E} [|\mathcal{L}^x - \mathcal{L}^{x-}|] = 0.$$

We finally conclude that for all $x \in \mathbb{R}$,

$$\mathbb{P}(\mathcal{L}^x \neq \mathcal{L}^{x-}) = \mathbb{P}(|\mathcal{L}^x - \mathcal{L}^{x-}| > 0) = 0.$$

□

8.2 Proofs of Section 3

8.2.1 Proof of Lemma 3.1

Proof. The proof is divided into two parts for each of the two results to be proven.

First result. Since the function h is continuous on \mathbb{R} , let H be a primitive of h on \mathbb{R} . We deduce that for all $s \in [0, 1]$,

$$h(X_s) = \lim_{\varepsilon \rightarrow 0} \frac{H(X_s + \varepsilon) - H(X_s - \varepsilon)}{2\varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_{X_s - \varepsilon}^{X_s + \varepsilon} h(x) dx = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_{-\infty}^{+\infty} h(x) \mathbb{1}_{(x-\varepsilon, x+\varepsilon)}(X_s) dx.$$

Finally, since h is integrable on \mathbb{R} and using the theorem of dominated convergence, we obtain

$$\int_0^1 h(X_s) ds = \int_{-\infty}^{+\infty} h(x) \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^1 \mathbb{1}_{(x-\varepsilon, x+\varepsilon)}(X_s) ds dx = \int_{-\infty}^{+\infty} h(x) \mathcal{L}^x dx.$$

Second result. Fix $t \in (0, 1]$ and consider $P_X : (t, x) \mapsto \int_{-\infty}^x p_X(t, y) dy$ the cumulative density function of the random variable X_t of the density function $x \mapsto p_X(t, x)$. We have:

$$\begin{aligned} \forall x \in \mathbb{R}, \quad \mathbb{E}(\mathcal{L}^x) &= \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^1 \mathbb{E} [\mathbb{1}_{(x-\varepsilon, x+\varepsilon)}(X_s)] ds = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^1 \mathbb{P}(x - \varepsilon \leq X_s \leq x + \varepsilon) ds \\ &= \int_0^1 \lim_{\varepsilon \rightarrow 0} \frac{P_X(s, x + \varepsilon) - P_X(s, x - \varepsilon)}{2\varepsilon} ds \\ &= \int_0^1 p_X(s, x) ds. \end{aligned}$$

□

8.2.2 Proof of Theorem 3.2

Let $\Omega_{n,m}$ be the random event in which the two pseudo-norms $\|\cdot\|_{n,1}$ and $\|\cdot\|_X$ are equivalent and given by

$$\Omega_{n,m} := \bigcap_{g \in \mathcal{S}_m \setminus \{0\}} \left\{ \left| \frac{\|g\|_{n,1}^2}{\|g\|_X^2} - 1 \right| \leq \frac{1}{2} \right\}.$$

The proof of Theorem 3.2 relies on the following lemma.

Lemma 8.6. *Let $\gamma > 1$ be a real number. Under Assumption 2.1, the following holds*

$$\mathbb{P}(\Omega_{n,m}^c) \leq C \frac{m^{2\gamma}}{n^{\gamma/2}},$$

where $C > 0$ is a constant depending on γ .

The parameter $\gamma > 1$ has to be chosen appropriately (i.e. such that $m^{2\gamma}/n^{\gamma/2} = o(1/n)$) so that we obtain a variance term of the risk of the estimator $\hat{\sigma}_m^2$ of order $m \log(n)/n$ (see Theorem 3.2 and Corollary 3.3).

Proof of Theorem 3.2 . Recall that since $N = 1$, $\zeta_{k\Delta}^1 = \zeta_{k\Delta}^{1,1} + \zeta_{k\Delta}^{1,2} + \zeta_{k\Delta}^{1,3}$ is the error term of the regression model, with:

$$\zeta_{k\Delta}^{1,1} = \frac{1}{\Delta} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s^1 \right)^2 - \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^1) ds \right], \quad (25)$$

$$\zeta_{k\Delta}^{1,2} = \frac{2}{\Delta} \int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s) \sigma'(X_s^1) \sigma^2(X_s^1) dW_s^1, \quad (26)$$

$$\zeta_{k\Delta}^{1,3} = 2b(X_{k\Delta}^1) \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s^1. \quad (27)$$

Besides, $R_{k\Delta}^1 = R_{k\Delta}^{1,1} + R_{k\Delta}^{1,2}$, with:

$$R_{k\Delta}^{1,1} = \frac{1}{\Delta} \left(\int_{k\Delta}^{(k+1)\Delta} b(X_s^1) ds \right)^2 + \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s) \Phi(X_s^1) ds \quad (28)$$

$$R_{k\Delta}^{1,2} = \frac{2}{\Delta} \left(\int_{k\Delta}^{(k+1)\Delta} (b(X_s^1) - b(X_{k\Delta}^1)) ds \right) \left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s^1 \right) \quad (29)$$

where

$$\Phi := 2b\sigma'\sigma + [\sigma''\sigma + (\sigma')^2] \sigma^2. \quad (30)$$

By definition of the projection estimator $\hat{\sigma}_m^2$ for each $m \in \mathcal{M}$ (see Equation (9)), for all $h \in \mathcal{S}_{m,L}$, we have:

$$\gamma_{n,1}(\hat{\sigma}_m^2) - \gamma_{n,1}(\sigma_{|I}^2) \leq \gamma_{n,1}(h) - \gamma_{n,1}(\sigma_{|I}^2). \quad (31)$$

Furthermore, for all $h \in \mathcal{S}_{m,L}$,

$$\gamma_{n,1}(h) - \gamma_{n,1}(\sigma_{|I}^2) = \left\| \sigma_{|I}^2 - h \right\|_{n,1}^2 + 2\nu_1(\sigma_{|I}^2 - h) + 2\nu_2(\sigma_{|I}^2 - h) + 2\nu_3(\sigma_{|I}^2 - h) + 2\mu(\sigma_{|I}^2 - h),$$

where,

$$\nu_i(h) = \frac{1}{n} \sum_{k=0}^{n-1} h(X_{k\Delta}^1) \zeta_{k\Delta}^{1,i}, \quad i \in \{1, 2, 3\}, \quad \mu(h) = \frac{1}{n} \sum_{k=0}^{n-1} h(X_{k\Delta}^1) R_{k\Delta}^1, \quad (32)$$

and $\zeta_{k\Delta}^{1,1}$, $\zeta_{k\Delta}^{1,2}$, $\zeta_{k\Delta}^{1,3}$ are given in Equations (25), (26), (27), and finally, $R_{k\Delta}^1 = R_{k\Delta}^{1,1} + R_{k\Delta}^{1,2}$ given in Equations (28) and (29). Then, for all $m \in \mathcal{M}$, and for all $h \in \mathcal{S}_{m,L}$, we obtain from Equation (31) that

$$\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \leq \left\| h - \sigma_{|I}^2 \right\|_{n,1}^2 + 2\nu(\hat{\sigma}_m^2 - h) + 2\mu(\hat{\sigma}_m^2 - h), \quad \text{with } \nu = \nu_1 + \nu_2 + \nu_3. \quad (33)$$

Then, it comes,

$$\mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + 2\mathbb{E}[\nu(\hat{\sigma}_m^2 - h)] + 2\mathbb{E}[\mu(\hat{\sigma}_m^2 - h)]. \quad (34)$$

Besides, for any $a, d > 0$, using the inequality $xy \leq \eta x^2 + y^2/\eta$ with $\eta = a, d$, we have,

$$\begin{cases} 2\nu(\hat{\sigma}_m^2 - h) \leq \frac{2}{a} \left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_X^2 + \frac{2}{a} \left\| h - \sigma_{|I}^2 \right\|_X^2 + a \sup_{h \in \mathcal{S}_m, \|h\|_X=1} \nu^2(h), \\ 2\mu(\hat{\sigma}_m^2 - h) \leq \frac{2}{d} \left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,1}^2 + \frac{2}{d} \left\| h - \sigma_{|I}^2 \right\|_{n,1}^2 + \frac{d}{n} \sum_{k=1}^n (R_{k\Delta}^1)^2. \end{cases} \quad (35)$$

Upper bound of $\frac{1}{n} \sum_{k=1}^n (R_{k\Delta}^1)^2$

We have:

$$\begin{aligned} \forall k \in \llbracket 1, n \rrbracket, R_{k\Delta}^1 &= R_{k\Delta}^{1,1} + R_{k\Delta}^{1,2} + R_{k\Delta}^{1,3} \text{ with,} \\ R_{k\Delta}^{1,1} &= \frac{1}{\Delta} \left(\int_{k\Delta}^{(k+1)\Delta} b(X_s^1) ds \right)^2, \quad R_{k\Delta}^{1,2} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s) \Phi(X_s^1) ds \\ R_{k\Delta}^{1,3} &= \frac{2}{\Delta} \left(\int_{k\Delta}^{(k+1)\Delta} (b(X_s^1) - b(X_{k\Delta}^1)) ds \right) \left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s^1 \right). \end{aligned}$$

For all $k \in \llbracket 1, n \rrbracket$, using the Cauchy-Schwarz inequality and Equation (2),

$$\mathbb{E} \left[\left| R_{k\Delta}^{1,1} \right|^2 \right] \leq \mathbb{E} \left[\left(\int_{k\Delta}^{(k+1)\Delta} b^2(X_{k\Delta}^1) ds \right)^2 \right] \leq \Delta \mathbb{E} \left[\int_{k\Delta}^{(k+1)\Delta} b^4(X_{k\Delta}^1) ds \right] \leq C \Delta^2.$$

Consider now the term $R_{k\Delta}^{1,2}$. From Equation (30), we have $\Phi = 2b\sigma'\sigma + [\sigma''\sigma + (\sigma')^2]\sigma^2$ and according to Assumption 2.1, there exists a constant $C > 0$ depending on σ_1 and α such that

$$|\Phi(X_s^1)| \leq C [(2 + |X_s^1|)(1 + |X_s^1|^\alpha) + (1 + |X_s^1|^\alpha)^2].$$

Then, from Equation (2) and for all $s \in (0, 1]$,

$$\mathbb{E} [\Phi^2(X_s^1)] \leq C \sup_{s \in (0,1]} \mathbb{E} [(2 + |X_s^1|)^2 (1 + |X_s^1|^\alpha)^2 + (1 + |X_s^1|^\alpha)^4] < \infty$$

and

$$\mathbb{E} \left[\left| R_{k\Delta}^{1,2} \right|^2 \right] \leq \frac{1}{\Delta^2} \int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s)^2 ds \int_{k\Delta}^{(k+1)\Delta} \mathbb{E} [\Phi^2(X_s^1)] ds \leq C \Delta^2$$

Finally, under Assumption 2.1, from Equation (2) and using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \mathbb{E} \left[\left| R_{k\Delta}^{1,3} \right|^2 \right] &\leq \frac{4}{\Delta^2} \mathbb{E} \left[\Delta \int_{k\Delta}^{(k+1)\Delta} L_0^2 |X_s^1 - X_{k\Delta}^1|^2 ds \left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s \right)^2 \right] \\ &\leq \frac{4}{\Delta} \sqrt{\mathbb{E} \left[L_0^4 \Delta \int_{k\Delta}^{(k+1)\Delta} |X_s^1 - X_{k\Delta}^1|^4 ds \right] \mathbb{E} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s \right)^4 \right]} \\ &\leq C \Delta^2. \end{aligned}$$

As a result, there exists a constant $C > 0$ such that,

$$\mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n (R_{k\Delta}^1)^2 \right] \leq C \Delta^2. \quad (36)$$

We set $a = d = 8$ and considering the event $\Omega_{n,m}$ on which the empirical norms $\|\cdot\|_X$ and $\|\cdot\|_{n,1}$ are equivalent, we deduce from Equations (34), (35) and (36) that,

$$\mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \mathbf{1}_{\Omega_{n,m}} \right] \leq 3 \inf_{h \in \mathcal{S}_m} \left\| h - \sigma_{|I}^2 \right\|_n^2 + C \mathbb{E} \left(\sup_{h \in \mathcal{S}_m, \|h\|_X=1} \nu^2(h) \right) + C \Delta^2 \quad (37)$$

where $C > 0$ is a constant depending on σ_1 .

Upper bound of $\mathbb{E} \left(\sup_{h \in \mathcal{S}_m, \|h\|_X=1} \nu^2(h) \right)$

For all $h = \sum_{\ell=0}^{m-1} a_\ell \phi_\ell \in \mathcal{S}_m$ such that $\|h\|_X^2 = 1$, we have $\|h\|^2 \leq \frac{1}{\tau_0}$ (see Equation (15)) and the coordinate vector $\mathbf{a} = (a_{-M}, \dots, a_{K-1})$ satisfies:

- $\|\mathbf{a}\|_2^2 \leq Cm$ ($m = K + M$) for the spline basis (see Denis *et al.* (2021), Lemma 2.6)
- $\|\mathbf{a}\|_2^2 \leq 1/\tau_0$ for an orthonormal basis since $\|h\|^2 = \|\mathbf{a}\|_2^2$.

Furthermore, using the Cauchy-Schwarz inequality, we have:

$$\nu^2(h) = \left(\sum_{\ell=0}^{m-1} a_\ell \nu(\phi_\ell) \right)^2 \leq \|\mathbf{a}\|_2^2 \sum_{\ell=0}^{m-1} \nu^2(\phi_\ell). \quad (38)$$

Thus, since $\nu = \nu_1 + \nu_2 + \nu_3$, for all $\ell \in \llbracket -M, K-1 \rrbracket$ and for all $i \in \{1, 2, 3\}$,

$$\mathbb{E} [\nu_i^2(\phi_\ell)] = \frac{1}{n^2} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \phi_\ell(X_{k\Delta}^1) \zeta_{k\Delta}^{1,i} \right)^2 \right].$$

1. Case $i = 1$

Recall that $\zeta_{k\Delta}^{1,1} = \frac{1}{\Delta} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s \right)^2 - \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^1) ds \right]$ where $W = W^1$. We fix a initial time $s \in [0, 1)$ and set $M_t^s = \int_s^t \sigma(X_u^1) dW_u$, $\forall t \geq s$. $(M_t^s)_{t \geq s}$ is a martingale and for all $t \in [s, 1]$, we have:

$$\langle M^s, M^s \rangle_t = \int_s^t \sigma^2(X_u^1) du.$$

Then, $\zeta_{k\Delta}^{1,1} = \frac{1}{\Delta} \left(M_{(k+1)\Delta}^{k\Delta} \right)^2 - \langle M^{k\Delta}, M^{k\Delta} \rangle_{(k+1)\Delta}$ is also a $\mathcal{F}_{k\Delta}$ -martingale, and, using the Burkholder-Davis-Gundy inequality, we obtain for all $k \in \llbracket 0, n-1 \rrbracket$,

$$\mathbb{E} [\zeta_{k\Delta}^{1,1} | \mathcal{F}_{k\Delta}] = 0, \quad \mathbb{E} \left[\left(\zeta_{k\Delta}^{1,1} \right)^2 | \mathcal{F}_{k\Delta} \right] \leq \frac{C}{\Delta^2} \mathbb{E} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_u^1) du \right)^2 \right] \leq C\sigma_1^4. \quad (39)$$

Then, using Equation (39) we have:

$$\begin{aligned} \mathbb{E} [\nu_1^2(\phi_\ell)] &= \frac{1}{n^2} \mathbb{E} \left[\sum_{k=0}^{n-1} \phi_\ell^2(X_{k\Delta}^1) \left(\zeta_{k\Delta}^{1,1} \right)^2 \right] = \frac{1}{n^2} \mathbb{E} \left[\sum_{k=0}^{n-1} \phi_\ell^2(X_{k\Delta}^1) \mathbb{E} \left[\left(\zeta_{k\Delta}^{1,1} \right)^2 | \mathcal{F}_{k\Delta} \right] \right] \\ &\leq \frac{C\sigma_1^4}{n^2} \mathbb{E} \left[\sum_{k=0}^{n-1} \phi_\ell^2(X_{k\Delta}^1) \right] \end{aligned}$$

and,

$$\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_1^2(\phi_\ell)] \leq \frac{C\sigma_1^4}{n^2} \mathbb{E} \left[\sum_{k=0}^{n-1} \sum_{\ell=0}^{m-1} \phi_\ell^2(X_{k\Delta}^1) \right].$$

One has:

$$\begin{cases} \sum_{\ell=-M}^{K-1} B_\ell^2 (X_{\eta(s)}^1) \leq 1 \text{ for the Spline basis } (m = K + M), \\ \sum_{\ell=0}^{m-1} \phi_\ell^2 (X_{\eta(s)}^1) \leq Cm \text{ for an orthonormal basis with } C = \max_{0 \leq \ell \leq m-1} \|\phi_\ell\|_\infty^2. \end{cases} \quad (40)$$

Thus, it comes that

- $\sum_{\ell=-M}^{K-1} \mathbb{E} [\nu_1^2 (B_\ell)] \leq C/n$ for the Spline basis,
- $\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_1^2 (\phi_\ell)] \leq Cm/n$ for an orthonormal basis,

and,

$$\mathbb{E} \left(\sup_{h \in \mathcal{S}_m, \|h\|_X^2=1} \nu_1^2(h) \right) \leq C \frac{m}{n} \quad (41)$$

where $C > 0$ is a constant depending on σ_1 and the basis.

2. Case $i = 2$

We have $\zeta_{k\Delta}^{1,2} = \frac{2}{\Delta} \int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s) \sigma' (X_s^1) \sigma^2 (X_s^1) dW_s$ and,

$$\begin{aligned} \mathbb{E} [\nu_2^2 (\phi_\ell)] &= 4\mathbb{E} \left[\left(\sum_{k=0}^{n-1} \phi_\ell (X_{k\Delta}^1) \int_{k\Delta}^{(k+1)\Delta} (k+1)\Delta - s) \sigma' (X_s^1) \sigma^2 (X_s^1) dW_s \right)^2 \right] \\ &= 4\mathbb{E} \left[\left(\int_0^1 \phi_\ell (X_{\eta(s)}^1) (\eta(s) + \Delta - s) \sigma' (X_s^1) \sigma^2 (X_s^1) dW_s \right)^2 \right] \\ &\leq C\sigma_1^4 \Delta^2 \mathbb{E} \left[\int_0^1 \phi_\ell^2 (X_{\eta(s)}^1) ds \right] \end{aligned}$$

where $C > 0$ is a constant. We deduce for both the spline basis and any orthonormal basis that there exists a constant $C > 0$ depending on σ_1 such that:

$$\mathbb{E} \left(\sup_{h \in \mathcal{S}_m, \|h\|_X^2=1} \nu_2^2(h) \right) \leq C \frac{m}{n^2}. \quad (42)$$

3. Case $i = 3$

We have $\zeta_{k\Delta}^{1,3} = 2b (X_{k\Delta}^1) \int_{k\Delta}^{(k+1)\Delta} \sigma (X_s^1) dW_s$ and,

$$\begin{aligned} \mathbb{E} [\nu_3^2 (\phi_\ell)] &= \frac{4}{n^2} \mathbb{E} \left[\left(\int_0^1 \phi_\ell (X_{\eta(s)}^1) b (X_{\eta(s)}^1) \sigma (X_s^1) dW_s \right)^2 \right] \\ &\leq \frac{4\sigma_1^2}{n^2} \mathbb{E} \left[\int_0^1 \phi_\ell^2 (X_{\eta(s)}^1) b^2 (X_{\eta(s)}^1) ds \right] \end{aligned}$$

Since for all $x \in \mathbb{R}$, $b^2(x) \leq C_0(1+x^2)$ and $\sup_{t \in [0,1]} \mathbb{E} (|X_t|^2) < \infty$, there exists a constant $C > 0$ depending on σ_1 such that:

$$\mathbb{E} \left(\sup_{h \in \mathcal{S}_m, \|h\|_X^2=1} \nu_3^2(h) \right) \leq C \frac{m}{n^2}. \quad (43)$$

We finally obtain from Equations (41), (42) and (43) that there exists a constant $C > 0$ depending on σ_1 such that:

$$\mathbb{E} \left(\sup_{h \in \mathcal{S}_m, \|h\|_X^2=1} \nu^2(h) \right) \leq C \frac{m}{n}. \quad (44)$$

We deduce from Equations (37) and (44) that there exists a constant $C > 0$ depending on σ_1 such that,

$$\mathbb{E} \left[\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_{n,1}^2 \mathbf{1}_{\Omega_{n,m}} \right] \leq 3 \inf_{h \in \mathcal{S}_{m,L}} \|\sigma_{|I}^2 - h\|_n^2 + C \left(\frac{m}{n} + \Delta^2 \right).$$

For n large enough, we have $\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_\infty^2 \leq 2mL$ since $\|\hat{\sigma}_m^2\|_\infty \leq \sqrt{mL}$. Then, from Lemma 8.6 and for all $m \in \mathcal{M}$, there exists a constant $C > 0$ depending on σ_1 such that

$$\begin{aligned} \mathbb{E} \left[\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_{n,1}^2 \right] &= \mathbb{E} \left[\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_{n,1}^2 \mathbf{1}_{\Omega_{n,m}} \right] + \mathbb{E} \left[\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_{n,1}^2 \mathbf{1}_{\Omega_{n,m}^c} \right] \\ &\leq \mathbb{E} \left[\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_{n,1}^2 \mathbf{1}_{\Omega_{n,m}} \right] + 2mL \mathbb{P}(\Omega_{n,m}^c) \\ &\leq 3 \inf_{h \in \mathcal{S}_{m,L}} \|\sigma_{|I}^2 - h\|_n^2 + C \left(\frac{m}{n} + \frac{m^{2\gamma+1}L}{n^{\gamma/2}} + \Delta^2 \right). \end{aligned}$$

Since the pseudo-norms $\|\cdot\|_{n,1}$ and $\|\cdot\|_X$ are equivalent on the event $\Omega_{n,m}$, then, using Lemma 8.6, there exists a constant $C > 0$ depending on σ_1 such that

$$\begin{aligned} \mathbb{E} \left[\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_X^2 \right] &= \mathbb{E} \left[\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_X^2 \mathbf{1}_{\Omega_{n,m}} \right] + \mathbb{E} \left[\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_X^2 \mathbf{1}_{\Omega_{n,m}^c} \right] \\ &\leq 8 \mathbb{E} \left[\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_{n,1}^2 \right] + 10 \inf_{h \in \mathcal{S}_m} \|\sigma_{|I}^2 - h\|_n^2 + 2mL \mathbb{P}(\Omega_{n,m}^c) \\ &\leq 34 \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma_{|I}^2\|_n^2 + C \left(\frac{m}{n} + \frac{m^{2\gamma+1}L}{n^{\gamma/2}} + \Delta^2 \right). \end{aligned}$$

Finally, since the estimator $\hat{\sigma}_m^2$ is built from a diffusion path \bar{X}^1 independent of the diffusion process X , and from Equations (15) and (12), the pseudo-norm $\|\cdot\|_X$ depending on the process X and the empirical norm $\|\cdot\|_n$ are equivalent ($\forall h \in \mathbb{L}^2(I)$, $\|h\|_n^2 \leq (\tau_1/\tau_0) \mathbb{E}[\|h\|_X^2]$), there exists a constant $C > 0$ depending on σ_1, τ_0 and τ_1 such that

$$\mathbb{E} \left[\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_n^2 \right] \leq \frac{34\tau_1}{\tau_0} \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma_{|I}^2\|_n^2 + C \left(\frac{m}{n} + \frac{m^{2\gamma+1}L}{n^{\gamma/2}} + \Delta^2 \right).$$

□

Proof of Lemma 8.6 . The proof of this Lemma mainly focus on the spline basis and the Fourier basis based on functions cos and sin which are Lipschitz functions. Thus, for all $g = \sum_{\ell=0}^{m-1} a_\ell \phi_\ell \in \mathcal{S}_m$,

$$\left| \|g\|_{n,1}^2 - \|g\|_X^2 \right| \leq \int_0^1 |g^2(X_{\eta(s)}) - g^2(X_s)| ds \leq 2\|g\|_\infty \int_0^1 |g(X_{\eta(s)}) - g(X_s)| ds. \quad (45)$$

From Equation (15), one has $\mathbb{E}[\|g\|_X^2] \geq \tau_0 \|g\|^2$. Thus, if $\|g\|_X^2 = 1$, then $\|g\|^2 \leq 1/\tau_0$, and we deduce for all $g = \sum_{\ell=0}^{m-1} a_\ell \phi_\ell$ that there exists a constant $C > 0$ such that

- **Spline basis:** $\|g\|_\infty \leq \|a\|_2 \leq C\sqrt{m}$ (see Denis *et al.* (2021))
- **Fourier basis:** $\|g\|_\infty \leq C\sqrt{m}$ since $\|g\| = \|a\|_2$ and $\sum_{\ell=0}^{m-1} \phi_\ell^2 = O(m)$.

Moreover, each $g \in \mathcal{S}_m$ such that $\|g\|_X^2 = 1$ is the Lipschitz function with a Lipschitz coefficient $L_g = O(m^{3/2})$. For the spline basis, this result is obtained in Denis *et al.* (2021), *proof of Lemma C.1* combined with *Lemma 2.6*. For the Fourier basis, for all $x, y \in I$ and using the Cauchy Schwarz inequality, we obtain

$$\begin{aligned} |g(x) - g(y)| &\leq \sum_{\ell=0}^{m-1} |a_\ell| \cdot |\phi_\ell(x) - \phi_\ell(y)| \\ &\leq 2\pi m \sqrt{m} \|\mathbf{a}\|_2 |x - y| \\ &\leq \frac{2\pi}{\tau_0} m \sqrt{m} |x - y|. \end{aligned}$$

Back to Equation (45), there exists a constant $C > 0$ such that

$$\left| \|g\|_{n,1}^2 - \|g\|_X^2 \right| \leq Cm^2 \int_0^1 |X_{\eta(s)} - X_s| ds \quad (46)$$

We have:

$$\Omega_{n,m}^c = \left\{ \omega \in \Omega, \exists g \in \mathcal{S}_m \setminus \{0\}, \left| \frac{\|g\|_{n,1}^2}{\|g\|_X^2} - 1 \right| > \frac{1}{2} \right\},$$

and, using Equation (46), we obtain

$$\sup_{g \in \mathcal{S}_m \setminus \{0\}} \left| \frac{\|g\|_{n,1}^2}{\|g\|_X^2} - 1 \right| = \sup_{g \in \mathcal{S}_m, \|g\|_X^2=1} \left| \|g\|_{n,1}^2 - \|g\|_X^2 \right| \leq Cm^2 \int_0^1 |X_{\eta(s)} - X_s| ds.$$

Finally, using the Markov inequality, the Hölder inequality, Equation (2), and Lemma 8.1, we conclude that

$$\begin{aligned} \mathbb{P}(\Omega_{n,m}^c) &\leq \mathbb{P}\left(Cm^2 \int_0^1 |X_{\eta(s)} - X_s| ds \geq \frac{1}{2}\right) \\ &\leq Cm^{2\gamma} \int_0^1 \mathbb{E}[|X_{\eta(s)} - X_s|^\gamma] ds \\ &\leq C \frac{m^{2\gamma}}{n^{\gamma/2}} \end{aligned}$$

with $\gamma \in (1, +\infty)$. □

8.2.3 Proof of Theorem 3.4

Proof. Since $L = \log^2(n)$, we have

$$\begin{aligned} \mathbb{E}\left[\|\hat{\sigma}_{m,L}^2 - \sigma^2\|_{n,1}^2\right] &= \mathbb{E}\left[\|(\hat{\sigma}_{m,L}^2 - \sigma^2)\mathbf{1}_{[-\log(n), \log(n)]}\|_{n,1}^2\right] + \mathbb{E}\left[\|(\hat{\sigma}_{m,L}^2 - \sigma^2)\mathbf{1}_{[-\log(n), \log(n)]^c}\|_{n,1}^2\right] \\ &\leq \mathbb{E}\left[\|(\hat{\sigma}_{m,L}^2 - \sigma^2)\mathbf{1}_{[-\log(n), \log(n)]}\|_{n,1}^2\right] + 2\log^2(n) \sup_{t \in (0,1]} \mathbb{P}(|X_t| > \log(n)). \end{aligned}$$

From Equation (34) (Proof of Theorem 3.2), for all $h \in \mathcal{S}_{m,L}$,

$$\mathbb{E}\left[\|(\hat{\sigma}_{m,L}^2 - \sigma^2)\mathbf{1}_{[-\log(n), \log(n)]}\|_{n,1}^2\right] \leq \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + 2 \sum_{i=1}^3 \mathbb{E}[\nu_i(\hat{\sigma}_m^2 - h)] + 2\mathbb{E}[\mu(\hat{\sigma}_m^2 - h)] \quad (47)$$

where ν_i , $i = 1, 2, 3$ and μ are given in Equation (32). For all $i \in \{1, 2, 3\}$ and for all $h \in \mathcal{S}_{m,L}$, one has

$$\mathbb{E} [\nu_i (\hat{\sigma}_{m,L}^2 - h)] \leq \sqrt{2m \log^2(n)} \sqrt{\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_i^2(\phi_\ell)]}. \quad (48)$$

1. Upper bound of $\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_1^2(\phi_\ell)]$

According to Equation (32), we have

$$\forall \ell \in \llbracket 0, m-1 \rrbracket, \nu_1(\phi_\ell) = \frac{1}{n} \sum_{k=0}^{n-1} \phi_\ell(X_{k\Delta}^1) \zeta_{k\Delta}^{1,1}$$

where $\zeta_{k\Delta}^{1,1} = \frac{1}{\Delta} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s \right)^2 - \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^1) ds \right]$ is a martingale satisfying

$$\mathbb{E} [\zeta_{k\Delta}^{1,1} | \mathcal{F}_{k\Delta}] = 0 \quad \text{and} \quad \mathbb{E} \left[\left(\zeta_{k\Delta}^{1,1} \right)^2 | \mathcal{F}_{k\Delta} \right] \leq \frac{1}{\Delta^2} \mathbb{E} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^1) ds \right)^2 \right] \leq C \sigma_1^4$$

with $C > 0$ a constant, $W = W^1$ and $(\mathcal{F}_t)_{t \geq 0}$ the natural filtration of the martingale $(M_t)_{t \in [0,1]}$ given for all $t \in [0, 1]$ by $M_t = \int_0^t \sigma(X_s^1) dW_s$. We derive that

$$\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_1^2(\phi_\ell)] = \frac{1}{n^2} \sum_{\ell=0}^{m-1} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \phi_\ell(X_{k\Delta}^1) \zeta_{k\Delta}^{1,1} \right)^2 \right] = \frac{1}{n^2} \mathbb{E} \left[\sum_{k=0}^{n-1} \sum_{\ell=0}^{m-1} \phi_\ell^2(X_{k\Delta}^1) \left(\zeta_{k\Delta}^{1,1} \right)^2 \right]$$

since for all integers k, k' such that $k > k' \geq 0$, we have

$$\mathbb{E} \left[\phi_\ell(X_{k\Delta}^1) \zeta_{k\Delta}^{1,1} \phi_\ell(X_{k'\Delta}^1) \zeta_{k'\Delta}^{1,1} | \mathcal{F}_{k\Delta} \right] = \phi_\ell(X_{k\Delta}^1) \zeta_{k\Delta}^{1,1} \phi_\ell(X_{k'\Delta}^1) \mathbb{E} \left[\zeta_{k'\Delta}^{1,1} | \mathcal{F}_{k\Delta} \right] = 0.$$

For each $k \in \llbracket 0, n-1 \rrbracket$, we have

$$\begin{cases} \sum_{\ell=0}^{m-1} \phi_\ell(X_{k\Delta}^1) = \sum_{\ell=-M}^{K-1} B_\ell(X_{k\Delta}^1) = 1 & \text{for the spline basis} \\ \sum_{\ell=0}^{m-1} \phi_\ell(X_{k\Delta}^1) \leq Cm & \text{For an orthonormal basis with } C = \max_{0 \leq \ell \leq m-1} \|\phi_\ell\|_\infty. \end{cases}$$

Finally, there exists a constant $C > 0$ such that

$$\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_1^2(\phi_\ell)] \leq \begin{cases} \frac{C}{n} & \text{for the spline basis} \\ C \frac{m}{n} & \text{for an orthonormal basis.} \end{cases}$$

2. Upper bound of $\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_2^2(\phi_\ell)]$

For all $k \in \llbracket 0, n-1 \rrbracket$ and for all $s \in [0, 1]$, set $\eta(s) = k\Delta$ if $s \in [k\Delta, (k+1)\Delta)$. We have:

$$\begin{aligned} \sum_{\ell=0}^{m-1} \mathbb{E} [\nu_2^2(\phi_\ell)] &= 4 \sum_{\ell=0}^{m-1} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \int_{k\Delta}^{(k+1)\Delta} \phi_\ell(X_{k\Delta}^1) ((k+1)\Delta - s) \sigma'(X_s^1) \sigma^2(X_s^1) dW_s \right)^2 \right] \\ &= 4 \sum_{\ell=0}^{m-1} \mathbb{E} \left[\left(\int_0^1 \phi_\ell(X_{\eta(s)}^1) (\eta(s) + \Delta - s) \sigma'(X_s^1) \sigma^2(X_s^1) dW_s \right)^2 \right]. \end{aligned}$$

We conclude that

$$\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_2^2(\phi_\ell)] \leq \begin{cases} \frac{C}{n^2} & \text{for the spline basis} \\ C \frac{m}{n^2} & \text{for an orthonormal basis.} \end{cases}$$

where the constant $C > 0$ depends on the diffusion coefficient and the upper bound of the basis functions.

3. Upper bound of $\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_2^3(\phi_\ell)]$

We have:

$$\begin{aligned} \sum_{\ell=0}^{m-1} \mathbb{E} [\nu_3^2(\phi_\ell)] &= \frac{4}{n^2} \sum_{\ell=0}^{m-1} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \int_{k\Delta}^{(k+1)\Delta} \phi_\ell(X_{k\Delta}^1) b(X_{k\Delta}^1) \sigma(X_s^1) dW_s \right)^2 \right] \\ &= \frac{4}{n^2} \sum_{\ell=0}^{m-1} \mathbb{E} \left[\left(\int_0^1 \phi_\ell(X_{\eta(s)}^1) b(X_{\eta(s)}^1) \sigma(X_s^1) dW_s \right)^2 \right] \\ &\leq \frac{4}{n^2} \mathbb{E} \left[\int_0^1 \sum_{\ell=0}^{m-1} \phi_\ell^2(X_{\eta(s)}^1) b^2(X_{\eta(s)}^1) \sigma^2(X_s^1) ds \right]. \end{aligned}$$

Since for all $x \in \mathbb{R}$, $b(x) \leq C_0(1+x^2)$ and $\sup_{t \in [0,1]} \mathbb{E} (|X_t|^4) < \infty$, there exists a constant $C > 0$ depending on the diffusion coefficient such that

$$\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_3^2(\phi_\ell)] \leq \begin{cases} \frac{C}{n^2} & \text{for the spline basis} \\ C \frac{m}{n^2} & \text{for an orthonormal basis.} \end{cases}$$

We finally deduce that from Equations (47) and (48) that for all $h \in \mathcal{S}_{m,L}$,

$$\begin{cases} \mathbb{E} \left[\left\| (\widehat{\sigma}_{m,L}^2 - \sigma^2) \mathbf{1}_{[-\log(n), \log(n)]} \right\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma^2\|_n^2 + C \sqrt{\frac{m \log^2(n)}{n}} + 2\mathbb{E} \left[\mu(\widehat{\sigma}_{m,L}^2 - h) \right] & \mathbf{[B]} \\ \mathbb{E} \left[\left\| (\widehat{\sigma}_{m,L}^2 - \sigma^2) \mathbf{1}_{[-\log(n), \log(n)]} \right\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma^2\|_n^2 + C \sqrt{\frac{m^2 \log^2(n)}{n}} + 2\mathbb{E} \left[\mu(\widehat{\sigma}_{m,L}^2 - h) \right] & \mathbf{[F]} \end{cases} \quad (49)$$

where $C > 0$ is a constant. It remains to obtain an upper bound of the term $\mu(\widehat{\sigma}_{m,L}^2 - h)$. For all $a > 0$ and for all $h \in \mathcal{S}_{m,L}$,

$$\begin{aligned} 2\mu(\widehat{\sigma}_{m,L}^2 - h) &\leq \frac{2}{a} \|\widehat{\sigma}_{m,L}^2 - \sigma^2\|_{n,1}^2 + \frac{2}{a} \|h - \sigma^2\|_{n,1}^2 + \frac{a}{n} \sum_{k=0}^{n-1} (R_{k\Delta}^1)^2 \\ 2\mathbb{E} [\mu(\widehat{\sigma}_{m,L}^2 - h)] &\leq \frac{2}{a} \mathbb{E} \|\widehat{\sigma}_{m,L}^2 - \sigma^2\|_{n,1}^2 + \frac{2}{a} \inf_{h \in \mathcal{S}_m} \|h - \sigma^2\|_n^2 + \frac{a}{n} \sum_{k=0}^{n-1} \mathbb{E} \left[(R_{k\Delta}^1)^2 \right]. \end{aligned}$$

Using Equations (36), (49) and setting $a = 4$, we deduce that there exists constant $C > 0$ depending on σ_1 such that,

$$\begin{cases} \mathbb{E} \left[\left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma^2\|_n^2 + C \sqrt{\frac{m \log^2(n)}{n}} + 2 \log^2(n) \sup_{t \in (0,1]} \mathbb{P}(|X_t| > A_n) & \mathbf{[B]} \\ \mathbb{E} \left[\left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma^2\|_n^2 + C \sqrt{\frac{m^2 \log^2(n)}{n}} + 2 \log^2(n) \sup_{t \in (0,1]} \mathbb{P}(|X_t| > A_n) & \mathbf{[F]}. \end{cases} \quad (50)$$

From Proposition 8.3, $\sup_{t \in (0,1]} \mathbb{P}(|X_t| > \log(n)) \leq \log^{-1}(n) \exp(-c \log^2(n))$ with $c > 0$ a constant. Then, we obtain from Equation (50) that

$$\begin{cases} \mathbb{E} \left[\|\widehat{\sigma}_{m,L}^2 - \sigma^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma^2\|_n^2 + C \sqrt{\frac{m \log^2(n)}{n}} & \text{[B]} \\ \mathbb{E} \left[\|\widehat{\sigma}_{m,L}^2 - \sigma^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma^2\|_n^2 + C \sqrt{\frac{m^2 \log^2(n)}{n}} & \text{[F]}. \end{cases} \quad (51)$$

□

8.3 Proof of Section 4

The following lemma allows us to obtain a risk bound of $\widehat{\sigma}_{m,L}^2$ defined with the empirical norm $\|\cdot\|_n$ from the risk bound defined from the pseudo norm $\|\cdot\|_{n,N}$.

Lemma 8.7. *Let $\widehat{\sigma}_{m,L}^2$ be the truncated projection estimator on \mathbb{R} of σ^2 over the subspace $\mathcal{S}_{m,L}$. Suppose that $L = \log^2(N)$, $N > 1$. Under Assumption 2.1, there exists a constant $C > 0$ independent of m and N such that*

$$\mathbb{E} \left[\|\widehat{\sigma}_{m,L}^2 - \sigma^2\|_{n,N}^2 \right] - 2\mathbb{E} \left[\|\widehat{\sigma}_{m,L}^2 - \sigma^2\|_n^2 \right] \leq C \frac{m^2 \log^3(N)}{N}.$$

The proof of Lemma 8.7 is provided in Denis *et al.* (2021), *Theorem 3.3*. The proof uses the independence of the copies $\bar{X}^1, \dots, \bar{X}^N$ of the process X at discrete times, and the Bernstein inequality.

8.3.1 Proof of Theorem 4.1

For fixed n and N in \mathbb{N}^* , we set for all $m \in \mathcal{M}$,

$$\Omega_{n,N,m} := \bigcap_{h \in \mathcal{S}_m \setminus \{0\}} \left\{ \left| \frac{\|h\|_{n,N}^2}{\|h\|_n^2} - 1 \right| \leq \frac{1}{2} \right\}. \quad (52)$$

As we can see, the empirical norms $\|h\|_{n,N}$ and $\|h\|_n$ of any function $h \in \mathcal{S}_m \setminus \{0\}$ are equivalent on $\Omega_{n,N,m}$. More precisely, on the set $\Omega_{n,N,m}$, for all $h \in \mathcal{S}_m \setminus \{0\}$, we have : $\frac{1}{2}\|h\|_n^2 \leq \|h\|_{n,N}^2 \leq \frac{3}{2}\|h\|_n^2$. We have the following result:

Lemma 8.8. *Under Assumption 2.1, the following holds:*

- If $n \geq N$ or $n \propto N$, then $m \in \mathcal{M} = \{1, \dots, \sqrt{N}/\log(Nn)\}$ and,

$$\mathbb{P}(\Omega_{n,N,m}^c) \leq 2 \exp(-C\sqrt{N}).$$

- If $n \leq N$, then $m \in \mathcal{M} = \{1, \dots, \sqrt{n}/\log(Nn)\}$ and

$$\mathbb{P}(\Omega_{n,N,m}^c) \leq 2 \exp(-C\sqrt{n})$$

where $C > 0$ is a constant.

Proof of Lemma 8.8 . We have:

$$\Omega_{n,N,m}^c = \left\{ \omega \in \Omega, \exists h_0 \in \mathcal{S}_m, \left| \frac{\|h\|_{n,N}^2}{\|h\|_n^2} - 1 \right| > \frac{1}{2} \right\},$$

Denote by $\mathcal{H}_m = \{h \in \mathcal{S}_m, \|h\|_n = 1\}$ and $\mathcal{H}_m^\varepsilon$ the ε -net of \mathcal{H}_m for any $\varepsilon > 0$. We have

$$\sup_{h \in \mathcal{H}_m} \left| \frac{\|h\|_{n,N}^2}{\|h\|_n^2} - 1 \right| = \sup_{h \in \mathcal{H}_m} \left| \|h\|_{n,N}^2 - 1 \right|.$$

Let $\varepsilon > 0$ and let $\mathcal{H}_m^\varepsilon$ be the ε -net of \mathcal{H}_m w.r.t. the supremum norm $\|\cdot\|_\infty$. Then, for each $h \in \mathcal{H}_m$, there exists $h_\varepsilon \in \mathcal{H}_m^\varepsilon$ such that $\|h - h_\varepsilon\|_\infty \leq \varepsilon$. Then

$$\left| \|h\|_{n,N}^2 - 1 \right| \leq \left| \|h\|_{n,N}^2 - \|h_\varepsilon\|_{n,N}^2 \right| + \left| \|h_\varepsilon\|_{n,N}^2 - 1 \right|$$

and,

$$\left| \|h\|_{n,N}^2 - \|h_\varepsilon\|_{n,N}^2 \right| \leq \frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \left| h(X_{k\Delta}^j) - h_\varepsilon(X_{k\Delta}^j) \right| (\|h\|_\infty + \|h_\varepsilon\|_\infty) \leq (\|h\|_\infty + \|h_\varepsilon\|_\infty) \varepsilon.$$

Moreover, we have $\|h\|^2, \|h_\varepsilon\|^2 \leq 1/\tau_0$. Then, there exists a constant $\mathbf{c} > 0$ such that

$$\begin{cases} \left| \|h\|_{n,N}^2 - \|h_\varepsilon\|_{n,N}^2 \right| \leq 2\sqrt{\frac{\mathbf{c}m}{\tau_0}} \varepsilon & \text{for the spline basis (see Lemma 2.6 in Denis et al.(2021))} \\ \left| \|h\|_{n,N}^2 - \|h_\varepsilon\|_{n,N}^2 \right| \leq 2\sqrt{\frac{\mathbf{c}m}{\tau_0}} \varepsilon & \text{for an orthonormal basis } (\|h\|_\infty^2 \leq (\max_{0 \leq \ell \leq m-1} \|\phi_\ell\|_\infty^2) m \|h\|^2). \end{cases}$$

Therefore, for all $\delta > 0$ and for both the spline basis and any orthonormal basis,

$$\mathbb{P} \left(\sup_{h \in \mathcal{H}_m} \left| \|h\|_{n,N}^2 - 1 \right| \geq \delta \right) \leq \mathbb{P} \left(\sup_{h \in \mathcal{H}_m^\varepsilon} \left| \|h\|_{n,N}^2 - 1 \right| \geq \delta/2 \right) + \mathbf{1}_{4\varepsilon\sqrt{\frac{\mathbf{c}m}{\tau_0}} \geq \delta}.$$

We set $\delta = 1/2$ and we choose $\varepsilon > 0$ such that $4\varepsilon\sqrt{\frac{\mathbf{c}m}{\tau_0}} < 1/2$. Then, using the Hoeffding inequality, there exists a constant $c > 0$ depending on \mathbf{c} and τ_0 such that

$$\mathbb{P}(\Omega_{n,N,m}^c) \leq 2\mathcal{N}_\infty(\varepsilon, \mathcal{H}_m) \exp\left(-c\frac{N}{m}\right) \quad (53)$$

where $\mathcal{N}_\infty(\varepsilon, \mathcal{H}_m)$ is the covering number of \mathcal{H}_m satisfying:

$$\mathcal{N}_\infty(\varepsilon, \mathcal{H}_m) \leq \left(\kappa \frac{\sqrt{m}}{\varepsilon} \right)^m \quad (54)$$

where the constant $\kappa > 0$ depends on $c > 0$ (see Denis et al. (2021), Proof of Lemma D.1). We set $\varepsilon = \frac{\kappa\sqrt{m^*}}{N}$ with $m^* = \max \mathcal{M}$ and we derive from Equations (53) and (54) that

$$\mathbb{P}(\Omega_{n,N,m}^c) \leq 2N^{m^*} \exp\left(-c\frac{N}{m^*}\right) = 2 \exp\left(-c\frac{N}{m^*} \left(1 - \frac{m^{*2} \log(N)}{cN}\right)\right).$$

- If $n \geq N$, then $m \in \mathcal{M} = \{1, \dots, \sqrt{N}/\log(Nn)\}$. Since $m^{*2} \log(N)/N \rightarrow 0$ as $N \rightarrow +\infty$, there exists a constant $C > 0$ such that

$$\mathbb{P}(\Omega_{n,N,m}^c) \leq 2 \exp\left(-C\sqrt{N}\right).$$

- If $n \leq N$, then $m \in \mathcal{M} = \{1, \dots, \sqrt{n}/\log(Nn)\}$, $m^{*2} \log(N)/N \leq \log(N)/\log^2(Nn) \rightarrow 0$ as $N, n \rightarrow \infty$, and

$$\mathbb{P}(\Omega_{n,N,m}^c) \leq 2 \exp(-C\sqrt{n}).$$

- If $n \propto N$, then $m \in \mathcal{M} = \{1, \dots, \sqrt{N}/\log(Nn)\}$. Since $m^{*2} \log(N)/N \rightarrow 0$ as $N \rightarrow +\infty$, there exists a constant $C > 0$ such that

$$\mathbb{P}(\Omega_{n,N,m}^c) \leq 2 \exp(-C\sqrt{N}).$$

□

Proof of Theorem 4.1 . The proof of Theorem 4.1 extends the proof of Theorem 3.2 when N tends to infinity. Then, we deduce from Equation (37) that

$$\mathbb{E} \left[\left\| \widehat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,m}} \right] \leq 3 \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + C \mathbb{E} \left(\sup_{h \in \mathcal{S}_m, \|h\|_n=1} \nu^2(h) \right) + C\Delta^2 \quad (55)$$

where $C > 0$ is a constant depending on σ_1 , and $\nu = \nu_1 + \nu_2 + \nu_3$ with

$$\nu_i(h) = \frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} h(X_{k\Delta}^j) \zeta_{k\Delta}^{j,i}, \quad i = 1, 2, 3$$

and the $\zeta_{k\Delta}^{j,i}$'s are the error terms depending on each path X^j , $j = 1, \dots, N$.

Upper bound of $\mathbb{E} \left(\sup_{h \in \mathcal{S}_m, \|h\|_n=1} \nu^2(h) \right)$

For all $h = \sum_{\ell=0}^{m-1} a_\ell \phi_\ell \in \mathcal{S}_m$ such that $\|h\|_n = 1$, we have $\|h\|^2 \leq \frac{1}{\tau_0}$ and the coordinate vector $a = (a_0, \dots, a_{m-1})$ satisfies:

- $\|a\|_2^2 \leq CK \leq Cm$ for the spline basis (see Denis *et al.* (2021), Lemma 2.6)
- $\|a\|_2^2 \leq 1/\tau_0$ for an orthonormal basis since $\|h\|^2 = \|a\|_2^2$.

Furthermore, using the Cauchy Schwartz inequality, we have:

$$\nu^2(h) = \left(\sum_{\ell=0}^{m-1} a_\ell \nu(\phi_\ell) \right)^2 \leq \|a\|_2^2 \sum_{\ell=0}^{m-1} \nu^2(\phi_\ell).$$

Thus, for all $\ell \in \llbracket 0, m-1 \rrbracket$, $\nu = \nu_1 + \nu_2 + \nu_3$ and for all $i \in \{1, 2, 3\}$

$$\mathbb{E} [\nu_i^2(\phi_\ell)] = \frac{1}{Nn^2} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \phi_\ell(X_{k\Delta}^1) \zeta_{k\Delta}^{1,i} \right)^2 \right].$$

We finally deduce from (41), (42) and (43) that there exists a constant $C > 0$ depending on σ_1 such that:

$$\mathbb{E} \left(\sup_{h \in \mathcal{S}_m, \|h\|_n=1} \nu^2(h) \right) \leq C \frac{m}{Nn}. \quad (56)$$

We deduce from (55) and (56) that there exists a constant $C > 0$ such that,

$$\mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,m}} \right] \leq 3 \inf_{h \in \mathcal{S}_{m,L}} \left\| \sigma_{|I}^2 - h \right\|_n^2 + C \left(\frac{m}{Nn} + \Delta^2 \right). \quad (57)$$

Since we have $\|\hat{\sigma}_m^2\|_\infty \leq \sqrt{mL}$, then for m and L large enough, $\|\hat{\sigma}_m^2 - \sigma_{|I}^2\|_\infty^2 \leq 2mL$. There exists a constant $C > 0$ such that for all $m \in \mathcal{M}$ and for m and L large enough,

$$\begin{aligned} \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \right] &= \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,m}} \right] + \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,m}^c} \right] \\ &\leq \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,m}} \right] + 2mL \mathbb{P}(\Omega_{n,N,m}^c). \end{aligned}$$

Then, from Equation (57), Lemma 8.8 and for $m \in \mathcal{M} = \left\{ 1, \dots, \sqrt{\min(n, N)} / \sqrt{\log(Nn)} \right\}$, we have:

$$\mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \right] \leq 3 \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + C \left(\frac{m}{Nn} + mL \exp \left(-C \sqrt{\min(n, N)} \right) + \Delta^2 \right)$$

where $C > 0$ is a constant. Recall that the empirical norms $\|\cdot\|_{n,N}$ and $\|\cdot\|_n$ are equivalent on $\Omega_{n,N,m}$, that is for all $h \in \mathcal{S}_m$, $\|h\|_n^2 \leq 2\|h\|_{n,N}^2$. Thus, we have

$$\begin{aligned} \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2 \right] &= \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2 \mathbf{1}_{\Omega_{n,N,m}} \right] + \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2 \mathbf{1}_{\Omega_{n,N,m}^c} \right] \\ &\leq \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2 \mathbf{1}_{\Omega_{n,N,m}} \right] + 2mL \mathbb{P}(\Omega_{n,N,m}^c). \end{aligned}$$

For all $h \in \mathcal{S}_{m,L} \subset \mathcal{S}_m$, we have:

$$\begin{aligned} \mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2 \mathbf{1}_{\Omega_{n,N,m}} \right] &\leq 2\mathbb{E} \left[\left\| \hat{\sigma}_m^2 - h \right\|_n^2 \mathbf{1}_{\Omega_{n,N,m}} \right] + 2 \left\| h - \sigma_{|I}^2 \right\|_n^2 \\ &\leq 4\mathbb{E} \left[\left\| \hat{\sigma}_m^2 - h \right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,m}} \right] + 2 \left\| h - \sigma_{|I}^2 \right\|_n^2 \\ &\leq 8\mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \right] + 10 \left\| h - \sigma_{|I}^2 \right\|_n^2. \end{aligned}$$

We finally conclude that

$$\mathbb{E} \left[\left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2 \right] \leq 34 \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + C \left(\frac{m}{Nn} + mL \exp \left(-C \sqrt{\min(n, N)} \right) + \Delta^2 \right).$$

□

8.3.2 Proof of Theorem 4.3

Proof. We consider the restriction $\sigma^2 \mathbf{1}_{[-\log(N), \log(N)]}$ of σ^2 on the compact interval $[-\log(N), \log(N)]$ on which the spline basis is built. Then we have:

$$\mathbb{E} \left[\left\| \hat{\sigma}_{m,L}^2 - \sigma^2 \right\|_n^2 \right] = \mathbb{E} \left[\left\| (\hat{\sigma}_{m,L}^2 - \sigma^2) \mathbf{1}_{[-\log(N), \log(N)]} \right\|_n^2 \right] + \mathbb{E} \left[\left\| (\hat{\sigma}_{m,L}^2 - \sigma^2) \mathbf{1}_{[-\log(N), \log(N)]^c} \right\|_n^2 \right]$$

and from Proposition 8.2, Lemma 8.4 and for N large enough, there exists constants $c, C > 0$ such that

$$\begin{aligned} \mathbb{E} \left[\left\| (\hat{\sigma}_{m,L}^2 - \sigma^2) \mathbf{1}_{[-\log(N), \log(N)]^c} \right\|_n^2 \right] &\leq \frac{2L}{n} \sum_{k=0}^{n-1} \mathbb{P}(|X_{k\Delta}| > \log(N)) \leq 2L \sup_{t \in [0,1]} \mathbb{P}(|X_t| \geq \log(N)) \\ &\leq \frac{C}{\log(N)} \exp(-c \log^2(N)). \end{aligned}$$

We deduce that

$$\mathbb{E} \left[\left\| \hat{\sigma}_{m,L}^2 - \sigma^2 \right\|_n^2 \right] = \mathbb{E} \left[\left\| (\hat{\sigma}_{m,L}^2 - \sigma^2) \mathbf{1}_{[-\log(N), \log(N)]} \right\|_n^2 \right] + \frac{C}{\log(N)} \exp(-c \log^2(N)). \quad (58)$$

It remains to upper-bound the first term on the right hand side of Equation (58).

Upper bound of $\mathbb{E} \left[\left\| \hat{\sigma}_{m,L}^2 - \sigma^2 \right\|_n^2 \mathbf{1}_{[-\log(N), \log(N)]} \right]$. For all $h \in \mathcal{S}_{m,L}$, we obtain from Equation (9),

$$\gamma_{n,N}(\hat{\sigma}_{m,L}^2) - \gamma_{n,N}(\sigma^2) \leq \gamma_{n,N}(h) - \gamma_{n,N}(\sigma^2). \quad (59)$$

For all $h \in \mathcal{S}_{m,L}$,

$$\gamma_{n,N}(h) - \gamma_{n,N}(\sigma^2) = \|h - \sigma^2\|_{n,N}^2 + 2\nu_1(\sigma^2 - h) + 2\nu_2(\sigma^2 - h) + 2\nu_3(\sigma^2 - h) + 2\mu(\sigma^2 - h)$$

where

$$\nu_i(h) = \frac{1}{nN} \sum_{j=1}^N \sum_{k=0}^{n-1} h(X_{k\Delta}^j) \zeta_{k\Delta}^{j,i}, \quad i \in \{1, 2, 3\}, \quad \mu(h) = \frac{1}{nN} \sum_{j=1}^N \sum_{k=0}^{n-1} h(X_{k\Delta}^j) R_{k\Delta}^j, \quad (60)$$

we deduce from Equation (59) that for all $h \in \mathcal{S}_{m,L}$,

$$\mathbb{E} \left[\left\| \hat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,N}^2 \mathbf{1}_{[-\log(N), \log(N)]} \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + 2 \sum_{i=1}^3 \mathbb{E} [\nu_i(\hat{\sigma}_{m,L}^2 - h)] + 2\mathbb{E} [\mu(\hat{\sigma}_{m,L}^2 - h)]. \quad (61)$$

For all $i \in \{1, 2, 3\}$ and for all $h \in \mathcal{S}_{m,L}$, one has

$$\mathbb{E} [\nu_i(\hat{\sigma}_{m,L}^2 - h)] \leq \sqrt{2mL} \sqrt{\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_i^2(\phi_\ell)]}. \quad (62)$$

1. Upper bound of $\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_1^2(\phi_\ell)]$

According to Equation (60), we have

$$\forall \ell \in \llbracket 0, m-1 \rrbracket, \quad \nu_1(\phi_\ell) = \frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \phi_\ell(X_{k\Delta}^j) \zeta_{k\Delta}^{j,1}$$

where $\zeta_{k\Delta}^{j,1} = \frac{1}{\Delta} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^j) dW_s^j \right)^2 - \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^j) ds \right]$ is a martingale satisfying

$$\mathbb{E} [\zeta_{k\Delta}^{1,1} | \mathcal{F}_{k\Delta}] = 0 \quad \text{and} \quad \mathbb{E} \left[\left(\zeta_{k\Delta}^{1,1} \right)^2 | \mathcal{F}_{k\Delta} \right] \leq \frac{1}{\Delta^2} \mathbb{E} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^1) ds \right)^2 \right] \leq C\sigma_1^4$$

with $C > 0$ a constant, $W = W^1$ and $(\mathcal{F}_t)_{t \geq 0}$ the natural filtration of the martingale $(M_t)_{t \in [0,1]}$ given for all $t \in [0, 1]$ by $M_t = \int_0^t \sigma(X_s^1) dW_s$. We derive that

$$\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_1^2(\phi_\ell)] = \frac{1}{Nn^2} \sum_{\ell=0}^{m-1} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \phi_\ell(X_{k\Delta}^j) \zeta_{k\Delta}^{1,1} \right)^2 \right] = \frac{1}{Nn^2} \mathbb{E} \left[\sum_{k=0}^{n-1} \sum_{\ell=0}^{m-1} \phi_\ell^2(X_{k\Delta}^1) \left(\zeta_{k\Delta}^{1,1} \right)^2 \right].$$

For each $k \in \llbracket 0, n-1 \rrbracket$, we have

$$\begin{cases} \sum_{\ell=0}^{m-1} \phi_\ell^2(X_{k\Delta}^1) = \sum_{\ell=-M}^{K-1} B_\ell^2(X_{k\Delta}^1) = 1 & \text{for the spline basis} \\ \sum_{\ell=0}^{m-1} \phi_\ell^2(X_{k\Delta}^1) \leq Cm & \text{For an orthonormal basis with } C = \max_{0 \leq \ell \leq m-1} \|\phi_\ell\|_\infty^2. \end{cases}$$

Finally, there exists a constant $C > 0$ such that

$$\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_1^2(\phi_\ell)] \leq \begin{cases} \frac{C}{Nn} & \text{for the spline basis} \\ C \frac{m}{Nn} & \text{for an orthonormal basis.} \end{cases}$$

2. Upper bound of $\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_2^2(\phi_\ell)]$

For all $k \in \llbracket 0, n-1 \rrbracket$ and for all $s \in [0, 1]$, set $\eta(s) = k\Delta$ if $s \in [k\Delta, (k+1)\Delta)$. We have:

$$\begin{aligned} \sum_{\ell=0}^{m-1} \mathbb{E} [\nu_2^2(\phi_\ell)] &= \frac{4}{N} \sum_{\ell=0}^{m-1} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \int_{k\Delta}^{(k+1)\Delta} \phi_\ell(X_{k\Delta}^1) ((k+1)\Delta - s) \sigma'(X_s^1) \sigma^2(X_s^1) dW_s \right)^2 \right] \\ &= \frac{4}{N} \sum_{\ell=0}^{m-1} \mathbb{E} \left[\left(\int_0^1 \phi_\ell(X_{\eta(s)}^1) (\eta(s) + \Delta - s) \sigma'(X_s^1) \sigma^2(X_s^1) dW_s \right)^2 \right]. \end{aligned}$$

We conclude that

$$\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_2^2(\phi_\ell)] \leq \begin{cases} \frac{C}{Nn^2} & \text{for the spline basis} \\ C \frac{m}{Nn^2} & \text{for an orthonormal basis.} \end{cases}$$

where the constant $C > 0$ depends on the diffusion coefficient and the upper bound of the basis functions.

3. Upper bound of $\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_3^2(\phi_\ell)]$

We have:

$$\begin{aligned} \sum_{\ell=0}^{m-1} \mathbb{E} [\nu_3^2(\phi_\ell)] &= \frac{4}{Nn^2} \sum_{\ell=0}^{m-1} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \int_{k\Delta}^{(k+1)\Delta} \phi_\ell(X_{k\Delta}^1) b(X_{k\Delta}^1) \sigma(X_s^1) dW_s \right)^2 \right] \\ &= \frac{4}{Nn^2} \sum_{\ell=0}^{m-1} \mathbb{E} \left[\left(\int_0^1 \phi_\ell(X_{\eta(s)}^1) b(X_{\eta(s)}^1) \sigma(X_s^1) dW_s \right)^2 \right] \\ &\leq \frac{4}{Nn^2} \mathbb{E} \left[\int_0^1 \sum_{\ell=0}^{m-1} \phi_\ell^2(X_{\eta(s)}^1) b(X_{\eta(s)}^1) \sigma^2(X_s^1) ds \right]. \end{aligned}$$

Since for all $x \in \mathbb{R}$, $b(x) \leq C_0(1+x^2)$ and $\sup_{t \in [0,1]} \mathbb{E}(|X_t|^2) < \infty$, there exists a constant $C > 0$ depending on the diffusion coefficient such that

$$\sum_{\ell=0}^{m-1} \mathbb{E}[\nu_3^2(\phi_\ell)] \leq \begin{cases} \frac{C}{Nn^2} & \text{for the spline basis} \\ C \frac{m}{Nn^2} & \text{for an orthonormal basis.} \end{cases}$$

We finally deduce that from Equations (61) and (62) that for all $h \in \mathcal{S}_{m,L}$,

$$\left\{ \begin{array}{l} \mathbb{E} \left[\left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,N}^2 \mathbb{1}_{[-\log(N), \log(N)]} \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + C \sqrt{\frac{mL}{Nn}} + 2\mathbb{E} \left[\mu(\widehat{\sigma}_{m,L}^2 - h) \right] \quad (1) \\ \mathbb{E} \left[\left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,N}^2 \mathbb{1}_{[-\log(N), \log(N)]} \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + C \sqrt{\frac{m^2L}{Nn}} + 2\mathbb{E} \left[\mu(\widehat{\sigma}_{m,L}^2 - h) \right] \quad (2) \end{array} \right. \quad (63)$$

where $C > 0$ is a constant, the result (1) corresponds to the spline basis, and the result (2) corresponds to any orthonormal basis. It remains to obtain an upper bound of the term $\mu(\widehat{\sigma}_{m,L}^2 - h)$. For all $a > 0$ and for all $h \in \mathcal{S}_{m,L}$,

$$\begin{aligned} 2\mu(\widehat{\sigma}_{m,L}^2 - h) &\leq \frac{2}{a} \left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,N}^2 + \frac{2}{a} \|h - \sigma^2\|_{n,N}^2 + \frac{a}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \left(R_{k\Delta}^j \right)^2 \\ 2\mathbb{E} \left[\mu(\widehat{\sigma}_{m,L}^2 - h) \right] &\leq \frac{2}{a} \mathbb{E} \left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,N}^2 + \frac{2}{a} \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + \frac{a}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \mathbb{E} \left[\left(R_{k\Delta}^j \right)^2 \right]. \end{aligned}$$

Using Equations (36), (63) and setting $a = 4$, we deduce that there exists constant $C > 0$ depending on σ_1 such that,

$$\left\{ \begin{array}{l} \mathbb{E} \left[\left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,N}^2 \mathbb{1}_{[-\log(N), \log(N)]} \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + C \left(\sqrt{\frac{mL}{Nn}} + \Delta^2 \right) \quad [\mathbf{B}] \\ \mathbb{E} \left[\left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,N}^2 \mathbb{1}_{[-\log(N), \log(N)]} \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + C \left(\sqrt{\frac{m^2L}{Nn}} + \Delta^2 \right) \quad [\mathbf{H}]. \end{array} \right. \quad (64)$$

The final result is obtained from Equations (58) and (64). \square

8.3.3 Proof of Lemma 4.5

Proof. It is proven in Comte & Genon-Catalot (2020a) that for each dimension $m \in \mathcal{M}$, the Gram matrix Ψ_m built from the Hermite basis is invertible. For the case of the \mathbf{B} -spline basis, let us consider a vector $(x_{-M}, \dots, x_{K-1}) \in \mathbb{R}^m$ such that $x_j \in [u_{j+M}, u_{j+M+1})$ and $B_j(x_j) \neq 0$. Since $[u_{j+M}, u_{j+M+1}) \cap [u_{j'+M}, u_{j'+M+1}) = \emptyset$ for all $j, j' \in \{-M, \dots, K-1\}$ such that $j \neq j'$, then for all $j, j' \in \{-M, \dots, K-1\}$ such that $j \neq j'$, $B_j(x_{j'}) = 0$. Consequently, we obtain:

$$\begin{aligned} \det \left((B_\ell(x_{\ell'}))_{-M \leq \ell, \ell' \leq K-1} \right) &= \det \left(\text{diag} (B_{-M}(x_M), \dots, B_{K-1}(x_{K-1})) \right) \\ &= \prod_{\ell=-M}^{K-1} B_\ell(x_\ell) \neq 0. \end{aligned}$$

Then, we deduce from Comte & Genon-Catalot (2020a), *Lemma 1* that the matrix Ψ_m is invertible for all $m \in \mathcal{M}$, where the function f_T are replaced by

$$f_n : x \mapsto \frac{1}{n} \sum_{k=0}^{n-1} p_X(k\Delta, x)$$

with $\lambda([-A_N, A_N] \cap \text{supp}(f_n)) > 0$, λ being the Lebesgue measure.

Case of the B-spline basis. For all $w \in \mathbb{R}^m$ such that $\|w\|_{2,m} = 1$, we have:

$$w' \Psi_m w = \|t_w\|_n^2 = \int_{-A_N}^{A_N} t_w^2(x) f_n(x) dx + \frac{t_w^2(x_0)}{n} \quad \text{with} \quad t_w = \sum_{\ell=-M}^{K-1} w_\ell B_\ell.$$

Under Assumption 2.1, the transition density $(t, x) \mapsto p_X(t, x)$ is approximated as follows

$$\forall (t, x) \in (0, 1] \times \mathbb{R}, \quad \frac{1}{K_* \sqrt{t}} \exp\left(-c_\sigma \frac{x^2}{t}\right) \leq p_X(t, x) \leq \frac{K_*}{\sqrt{t}} \exp\left(-\frac{x^2}{c_\sigma t}\right)$$

where $K_* > 1$ and $c_\sigma > 1$. Since $s \mapsto \exp(-c_\sigma x^2/s)$ is an increasing function, then for n large enough and for all $x \in [-A_N, A_N]$,

$$\begin{aligned} f_n(x) &\geq \frac{1}{K_* n} \sum_{k=1}^{n-1} \exp\left(-c \frac{x^2}{k\Delta}\right) \geq \frac{1}{K_*} \int_0^{1-\Delta} \exp\left(-c_\sigma \frac{x^2}{s}\right) ds \\ &\geq \frac{1}{K_*} \int_{1-(\log(N))^{-1}}^{1-(2\log(N))^{-1}} \exp\left(-c_\sigma \frac{x^2}{s}\right) ds \\ &\geq \frac{1}{2K_* \log(N)} \exp\left(-\frac{c_\sigma x^2}{1 - \log^{-1}(N)}\right). \end{aligned}$$

Thus, the density function satisfies

$$\forall x \in [-A_N, A_N], \quad f_n(x) \geq \frac{1}{2K_* \log(N)} \exp\left(-\frac{c_\sigma A_N^2}{1 - \log^{-1}(N)}\right) \geq \frac{1}{2K_* \log(N)} \exp(-c_\sigma A_N^2). \quad (65)$$

Finally, since there exists a constant $C_1 > 0$ such that $\|t_w\|^2 \geq C_1 A_N K_N^{-1}$ (see Denis *et al.* (2021), Lemma 2.6), for all $w \in \mathbb{R}^m$ ($m = K_N + M$) such that $\|w\|_{2,m} = 1$, there exists a constant $C > 0$ such that,

$$w' \Psi_m w \geq \frac{C A_N}{m \log(N)} \exp(-c_\sigma A_N^2).$$

Case of the Hermite basis. For all $w \in \mathbb{R}^m$ such that $\|w\|_{2,m} = 1$, we have

$$w' \Psi_m w = \|t_w\|_n^2 = \int_{-\infty}^{+\infty} t_w^2(x) f_n(x) dx + \frac{t_w^2(x_0)}{n} \quad \text{with} \quad t_w = \sum_{\ell=0}^{m-1} w_\ell h_\ell.$$

Recall that for all $x \in \mathbb{R}$ such that $|x| \geq \sqrt{(3/2)(4m+3)}$, $|h_\ell(x)| \leq c|x| \exp(-c_0x^2)$ for all $\ell \geq 0$. Then we have

$$\begin{aligned} w' \Psi_m w &\geq \int_{|x| \leq \sqrt{(3/2)(4m+3)}} \left(\sum_{\ell=0}^{m-1} w_\ell h_\ell(x) \right)^2 f_n(x) dx \\ &\geq \inf_{x \in [-\sqrt{(3/2)(4m+3)}, \sqrt{(3/2)(4m+3)}]} f_n(x) \int_{|x| \leq \sqrt{(3/2)(4m+3)}} \left(\sum_{\ell=0}^{m-1} w_\ell h_\ell(x) \right)^2 dx \\ &\geq \frac{1}{2K_* \log(N)} \exp\left(-\frac{3c_\sigma(4m+3)}{2(1-\log^{-1}(N))}\right) \int_{|x| \leq \sqrt{(3/2)(4m+3)}} \left(\sum_{\ell=0}^{m-1} w_\ell h_\ell(x) \right)^2 dx \end{aligned}$$

since for all $x \in \mathbb{R}$, $f_n(x) \geq (1/2K_* \log(N)) \exp\left(-\frac{c_\sigma x^2}{1-\log^{-1}(N)}\right)$. Set $a_N = \sqrt{(3/2)(4m+3)}$, then we obtain

$$\begin{aligned} w' \Psi_m w &\geq \frac{\exp\left(-\frac{3c_\sigma(4m+3)}{2(1-\log^{-1}(N))}\right)}{2K_* \log(N)} \left(\int_{-\infty}^{+\infty} \left(\sum_{\ell=0}^{m-1} w_\ell h_\ell(x) \right)^2 dx - \int_{|x| > a_N} \left(\sum_{\ell=0}^{m-1} w_\ell h_\ell(x) \right)^2 dx \right) \\ &\geq \frac{\exp\left(-\frac{3c_\sigma(4m+3)}{2(1-\log^{-1}(N))}\right)}{2K_* \log(N)} \left(1 - 2c^2 m \int_{a_N}^{+\infty} x^2 \exp(-8c_0x^2) dx \right) \\ &\geq \frac{\exp\left(-\frac{3c_\sigma(4m+3)}{2(1-\log^{-1}(N))}\right)}{2K_* \log(N)} \left(1 - \frac{c^2 m}{8c_0} \sqrt{\frac{3}{2}(4m+3)} \exp(-12c_0(4m+3)) \right) \end{aligned}$$

where $c, c_0 > 0$ are constants depending on the Hermite basis. Finally, for N large enough,

$$1 - \frac{c^2 m}{8c_0} \sqrt{\frac{3}{2}(4m+3)} \exp(-12c_0(4m+3)) \geq \frac{1}{2}.$$

Finally, there exists a constant $C > 0$ such that for all $w \in \mathbb{R}^m$ such that $\|w\|_{2,m}$,

$$w' \Psi_m w \geq \frac{C}{\log(N)} \exp\left(-\frac{3c_\sigma(4m+3)}{2(1-\log^{-1}(N))}\right).$$

□

8.3.4 Proof of Theorem 4.6

The proof of Theorem 4.6 relies on the following lemma:

Lemma 8.9. *Under Assumptions 2.1 and for $\sigma^2 \in \Sigma_I(\beta, R)$ with $\beta \geq 1$, $I = [-A_N, A_N]$ and*

$$N \propto n, \quad A_N = o\left(\sqrt{\log(N)}\right), \quad K \propto \left((Nn)^{1/(2\beta+1)} A_N\right) \quad (m = K + M),$$

the following holds:

$$\mathbb{P}(\Omega_{n,N,m}^c) \leq C \exp\left(-c \log^{3/2}(N)\right)$$

where $c, C > 0$ are constants independent of N .

Proof of Theorem 4.6. According to Equations (55) in the proof of Theorem 4.1, for all dimension $m = K + M$, with $K \in \mathcal{M}$, and for all $h \in \mathcal{S}_{K+M}$, there exists a constant $C > 0$ such that

$$\mathbb{E} \left[\|\widehat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \mathbf{1}_{\Omega_{n, N, m}} \right] \leq C \left[\inf_{h \in \mathcal{S}_{K+M, L}} \|h - \sigma_{A_N}^2\|_n^2 + \mathbb{E} \left(\sup_{h \in \mathcal{S}_{K+M}, \|h\|_n=1} \nu^2(h) \right) + \Delta^2 \right] \quad (66)$$

where $\Omega_{n, N, m}$ is given in Equation (52) and $\nu = \nu_1 + \nu_2 + \nu_3$ with the ν_i given in Equation (32).

For all $h = \sum_{\ell=-M}^{K-1} a_\ell B_\ell \in \mathcal{S}_{K+M, L_N}$,

$$\|h\|_n^2 = \mathbb{E} \left[\frac{1}{n} \sum_{k=0}^{n-1} h^2(X_{k\Delta}) \right] = \sum_{\ell=-M}^{K-1} \sum_{\ell'=-M}^{K-1} a_\ell a_{\ell'} \mathbb{E} \left[\frac{1}{n} \sum_{k=0}^{n-1} B_\ell(X_{k\Delta}) B_{\ell'}(X_{k\Delta}) \right] = a' \Psi_m a.$$

The Gram matrix Ψ_m is invertible for each $K \in \mathcal{M}$ (see proof of Lemma 4.5). It follows that for all $h = \sum_{\ell=-M}^{K-1} a_\ell B_\ell$ such that $\|h\|_n^2 = a' \Psi_m a = 1$, one has $a = \Psi_m^{-1/2} u$ where $u \in \mathbb{R}^m$ and $\|u\|_{2, m} = 1$. Furthermore, we have:

$$h = \sum_{\ell=-M}^{K-1} a_\ell B_\ell = \sum_{\ell=-M}^{K-1} u_\ell \sum_{\ell'}^{K-1} [\Psi_m^{-1/2}]_{\ell', \ell} B_{\ell'}.$$

Then for all $h \in \mathcal{S}_{K+M}$, we have $\nu^2(h) \leq 3(\nu_1^2(h) + \nu_2^2(h) + \nu_3^2(h))$ where,

$$\forall i \in \{1, 2, 3\}, \nu_i^2(h) \leq \sum_{\ell=-M}^{K-1} \left(\frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \sum_{\ell'=-M}^{K-1} [\Psi_m^{-1/2}]_{\ell', \ell} B_{\ell'}(X_{k\Delta}^j) \zeta_{k\Delta}^{j, i} \right)^2.$$

So we obtain,

$$\forall i \in \{1, 2, 3\}, \mathbb{E} \left[\sup_{h \in \mathcal{S}_{K+M}, \|h\|_n=1} \nu_i^2(h) \right] \leq \frac{1}{Nn^2} \sum_{\ell=-M}^{K-1} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \sum_{\ell'=-M}^{K-1} [\Psi_m^{-1/2}]_{\ell', \ell} B_{\ell'}(X_{k\Delta}^j) \zeta_{k\Delta}^{1, i} \right)^2 \right]$$

For $i = 1$, we have $\zeta_{k\Delta}^{1, 1} = \frac{1}{\Delta} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s \right)^2 - \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^1) ds \right]$ and we obtained in the proof of Theorem 5.1 that there exists a constant $C > 0$ such that for all $k \in \llbracket 0, n-1 \rrbracket$,

$$\mathbb{E} \left[\zeta_{k\Delta}^{1, 1} | \mathcal{F}_{k\Delta} \right] = 0, \quad \mathbb{E} \left[\left(\zeta_{k\Delta}^{1, 1} \right)^2 | \mathcal{F}_{k\Delta} \right] \leq C \mathbb{E} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_u) du \right)^2 \right] \leq C \sigma_1^4 \Delta^2.$$

We deduce that

$$\begin{aligned} \mathbb{E} \left[\sup_{h \in \mathcal{S}_{K+M}, \|h\|_n=1} \nu_1^2(h) \right] &= \frac{1}{Nn^2 \Delta^2} \sum_{\ell=0}^{K-1} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \sum_{\ell'=-M}^{K-1} [\Psi_m^{-1/2}]_{\ell', \ell} B_{\ell'}(X_{k\Delta}^1) \zeta_{k\Delta}^{1, 1} \right)^2 \right] \\ &\leq \frac{1}{N} \sum_{\ell=-M}^{K-1} \sum_{k=0}^{n-1} \mathbb{E} \left\{ \left(\sum_{\ell'=-M}^{K-1} [\Psi_m^{-1/2}]_{\ell', \ell} B_{\ell'}(X_{k\Delta}^1) \right)^2 \left(\zeta_{k\Delta}^{1, 1} \right)^2 \right\} \\ &\leq \frac{4\sigma_1^2}{Nn} \sum_{\ell=-M}^{K-1} \sum_{\ell'=-M}^{K-1} \sum_{\ell''=-M}^{K-1} [\Psi_m^{-1/2}]_{\ell', \ell} [\Psi_m^{-1/2}]_{\ell'', \ell'} [\Psi_m^{-1/2}]_{\ell, \ell''}. \end{aligned}$$

We have:

$$\sum_{\ell=-M}^{K-1} \sum_{\ell'=-M}^{K-1} \sum_{\ell''=-M}^{K-1} \left[\Psi_m^{-1/2} \right]_{\ell',\ell} \left[\Psi_m^{-1/2} \right]_{\ell'',\ell} \left[\Psi_m^{-1/2} \right]_{\ell',\ell''} = \text{Tr} (\Psi_m^{-1} \Psi_m) = \text{Tr}(I_m) = m.$$

So we obtain

$$\mathbb{E} \left[\sup_{h \in \mathcal{S}_{K+M}} \nu_1^2(h) \right] \leq \frac{4\sigma_1^2 m}{Nn}.$$

For $i = 2$, we have $\zeta_{k\Delta}^{1,2} = \frac{2}{\Delta} \int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s) \sigma'(X_s^1) \sigma^2(X_s^1) dW_s$ and

$$\begin{aligned} \mathbb{E} \left[\sup_{h \in \mathcal{S}_{K+M}, \|h\|_n=1} \nu_2^2(h) \right] &\leq \frac{1}{Nn^2} \sum_{\ell=-M}^{K-1} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \sum_{\ell'=-M}^{K-1} \left[\Psi_m^{-1/2} \right]_{\ell',\ell} B_{\ell'}(X_{k\Delta}^j) \zeta_{k\Delta}^{1,2} \right)^2 \right] \\ &\leq \frac{4\sigma_1^4 \|\sigma'\|_\infty^2 \Delta}{Nn^2} \sum_{\ell=-M}^{K-1} \sum_{k=0}^{n-1} \mathbb{E} \left[\left(\sum_{\ell'=-M}^{K-1} \left[\Psi_m^{-1/2} \right]_{\ell',\ell} B_{\ell'}(X_{k\Delta}^1) \right)^2 \right] \\ &\leq \frac{4\sigma_1^2 \|\sigma'\|_\infty^2 m}{Nn^2}. \end{aligned}$$

For $i = 3$, we have $\zeta_{k\Delta}^{1,3} = 2b(X_{k\Delta}^1) \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s$ and there exists constants $C_1, C_2 > 0$ such that

$$\begin{aligned} \mathbb{E} \left[\sup_{h \in \mathcal{S}_{K+M}, \|h\|_n=1} \nu_3^2(h) \right] &\leq \frac{1}{Nn^2} \sum_{\ell=-M}^{K-1} \mathbb{E} \left[\left(\sum_{k=0}^{n-1} \sum_{\ell'=-M}^{K-1} \left[\Psi_m^{-1/2} \right]_{\ell',\ell} B_{\ell'}(X_{k\Delta}^j) \zeta_{k\Delta}^{1,3} \right)^2 \right] \\ &\leq C_1 \frac{\sigma_1^2 \Delta}{Nn^2} \sum_{\ell=-M}^{K-1} \sum_{k=0}^{n-1} \mathbb{E} \left[\left(\sum_{\ell'=-M}^{K-1} \left[\Psi_m^{-1/2} \right]_{\ell',\ell} B_{\ell'}(X_{k\Delta}^1) \right)^2 \right] \\ &\leq C_2 \frac{\sigma_1^2 m}{Nn^2}. \end{aligned}$$

Finally, there exists a constant $C > 0$ depending on σ_1 and M such that

$$\mathbb{E} \left[\sup_{h \in \mathcal{S}_{K+M}, \|h\|_n=1} \nu^2(h) \right] \leq C \frac{m}{Nn}. \quad (67)$$

From Equations (66) and (67), we deduce that

$$\mathbb{E} \left[\|\hat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \mathbf{1}_{\Omega_{n, N, m}} \right] \leq C \left(\inf_{h \in \mathcal{S}_{K+M, L}} \|h - \sigma_{A_N}^2\|_n^2 + \frac{m}{Nn} + \Delta^2 \right)$$

where $C > 0$ is a constant depending on σ_1 and M . We obtain

$$\mathbb{E} \left[\|\hat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \right] \leq C \left(\inf_{h \in \mathcal{S}_{K+M, L}} \|h - \sigma_{A_N}^2\|_n^2 + \frac{m}{Nn} + \Delta^2 \right) + \mathbb{E} \left[\|\hat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \mathbf{1}_{\Omega_{n, N, m}^c} \right]$$

and for N large enough, $\|\hat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \leq 4mL$, and according to Lemma 8.9,

$$\mathbb{E} \left[\|\hat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \mathbf{1}_{\Omega_{n, N, m}^c} \right] \leq 4mL \mathbb{P}(\Omega_{n, N, m}^c) \leq CmL \exp(-c \log^{3/2}(N))$$

where $c > 0$ is a constant. Thus, there exists a constant $C > 0$ such that

$$\begin{aligned} \mathbb{E} \left[\|\widehat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \right] &\leq \mathbb{E} \left[\|\widehat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \mathbf{1}_{\Omega_{n, N, m}} \right] + \mathbb{E} \left[\|\widehat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \mathbf{1}_{\Omega_{n, N, m}^c} \right] \\ &\leq C \left(\inf_{h \in \mathcal{S}_{K+M, L}} \|h - \sigma_{A_N}^2\|_n^2 + \frac{m}{Nn} + mL \exp \left(-c \log^{3/2}(N) \right) + \Delta^2 \right). \end{aligned}$$

Then, as $n \propto N$ and $L = \log(N)$, there exists a constant $C > 0$ such that

$$\mathbb{E} \left[\|\widehat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \right] \leq C \left(\inf_{h \in \mathcal{S}_{K+M, L}} \|h - \sigma_{A_N}^2\|_n^2 + \frac{m}{Nn} \right).$$

Finally, since $\sigma^2 \in \Sigma_I(\beta, R)$ with $\beta \geq 1$ and $I = [-A_N, A_N]$, one has

$$\inf_{h \in \mathcal{S}_{K+M, L}} \|h - \sigma_{A_N}^2\|_n^2 \leq CA_N^{2\beta} K^{-2\beta}$$

where the constant $C > 0$ depends on β, R and M . Furthermore, as we chose the interval $[-A_N, A_N]$ such that $A_N = o\left(\sqrt{\log(N)}\right)$ and for $K \propto ((Nn)^{1/(2\beta+1)} A_N)$, we obtain

$$\mathbb{E} \left[\|\widehat{\sigma}_{A_N, m}^2 - \sigma_{A_N}^2\|_{n, N}^2 \right] \leq C \log^\beta(N) (Nn)^{-2\beta/(2\beta+1)}.$$

□

8.4 Proof of Section 5

8.4.1 Proof of Theorem 5.1

Set for all $K, K' \in \mathcal{K} = \left\{ 2^q, q = 0, \dots, q_{\max}, 2^{q_{\max}} \leq \sqrt{N}/\log(N) \right\} \subset \mathcal{M}$,

$$\mathcal{T}_{K, K'} = \left\{ g \in \mathcal{S}_{K+M} + \mathcal{S}_{K'+M}, \|g\|_n = 1, \|g\|_\infty \leq \sqrt{L} \right\}. \quad (68)$$

Recall that for all $j \in \llbracket 1, N \rrbracket$ and for all $k \in \llbracket 0, n \rrbracket$,

$$\zeta_{k\Delta}^{j,1} = \frac{1}{\Delta} \left[\left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^j) dW_s^j \right)^2 - \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^j) ds \right].$$

The proof of Theorem 5.1 relies on the following lemma whose proof is in Appendix.

Lemma 8.10. *Under Assumption 2.1, for all $\varepsilon, v > 0$ and $g \in \mathcal{T}_{K, K'}$, there exists a real constant $C > 0$ such that,*

$$\mathbb{P} \left(\frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} g(X_{k\Delta}^j) \zeta_{k\Delta}^{j,1} \geq \varepsilon, \|g\|_{n, N}^2 \leq v^2 \right) \leq \exp \left(-C \frac{Nn\varepsilon^2}{\sigma_1^2 (\varepsilon \|g\|_\infty + 4\sigma_1^2 v^2)} \right)$$

and for all $x > 0$ such that $x \leq \varepsilon^2/\sigma_1^2 (\varepsilon \|g\|_\infty + 4\sigma_1^2 v^2)$,

$$\mathbb{P} \left(\frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} g(X_{k\Delta}^j) \zeta_{k\Delta}^{j,1} \geq 2\sigma_1^2 v \sqrt{x} + \sigma_1^2 \|g\|_\infty x, \|g\|_{n, N}^2 \leq v^2 \right) \leq \exp(-CNnx).$$

Proof of Theorem 5.1. From Equation (22), we have

$$\widehat{K} := \arg \min_{K \in \mathcal{K}} \{ \gamma_{n,N}(\widehat{\sigma}_K^2) + \text{pen}(K) \}.$$

For all $K \in \mathcal{K}$ and $h \in \mathcal{S}_{K+M,L}$,

$$\gamma_{n,N}(\widehat{\sigma}_{\widehat{K}}^2) + \text{pen}(\widehat{K}) \leq \gamma_{n,N}(h) + \text{pen}(K),$$

then, for all $K \in \mathcal{K}$ and for all $h \in \mathcal{S}_{K+M,L}$,

$$\begin{aligned} \gamma_{n,N}(\widehat{\sigma}_{\widehat{K}}^2) - \gamma_{n,N}(\sigma_{|I}^2) &\leq \gamma_{n,N}(h) - \gamma_{n,N}(\sigma_{|I}^2) + \text{pen}(K) - \text{pen}(\widehat{K}) \\ \left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 &\leq \left\| h - \sigma_{|I}^2 \right\|_{n,N}^2 + 2\nu \left(\widehat{\sigma}_{\widehat{K}}^2 - h \right) + 2\mu \left(\widehat{\sigma}_{\widehat{K}}^2 - h \right) + \text{pen}(K) - \text{pen}(\widehat{K}) \\ &\leq \left\| h - \sigma_{|I}^2 \right\|_{n,N}^2 + \frac{1}{d} \left\| \widehat{\sigma}_{\widehat{K}}^2 - t \right\|_n^2 + d \sup_{g \in \mathcal{T}_{K,\widehat{K}}} \nu^2(g) + \frac{1}{d} \left\| \widehat{\sigma}_{\widehat{K}}^2 - h \right\|_{n,N}^2 \\ &\quad + \frac{d}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \left(R_{k\Delta}^j \right)^2 + \text{pen}(K) - \text{pen}(\widehat{K}) \end{aligned}$$

where $d > 1$ and the space $\mathcal{T}_{K,\widehat{K}}$ is given in Equation (68). On the set $\Omega_{n,N,K_{\max}}$ (given in Equation (52)): $\forall h \in \mathcal{S}_{K+M}$, $\frac{1}{2} \|h\|_n^2 \leq \|h\|_{n,N}^2 \leq \frac{3}{2} \|h\|_n^2$. Then on $\Omega_{n,N,K_{\max}}$, for all $d > 1$ and for all $h \in \mathcal{S}_{K+M}$ with $K \in \mathcal{K}$,

$$\begin{aligned} \left(1 - \frac{10}{d} \right) \left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 &\leq \left(1 + \frac{10}{d} \right) \left\| h - \sigma_{|I}^2 \right\|_{n,N}^2 + d \sup_{h \in \mathcal{T}_{K,\widehat{K}}} \nu^2(h) + \frac{d}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \left(R_{k\Delta}^j \right)^2 \\ &\quad + \text{pen}(K) - \text{pen}(\widehat{K}). \end{aligned}$$

We set $d = 20$. Then, on $\Omega_{n,N,\max}$ and for all $h \in \mathcal{S}_{K+M,L}$,

$$\left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \leq 3 \left\| h - \sigma_{|I}^2 \right\|_{n,N}^2 + 20 \sup_{h \in \mathcal{T}_{K,\widehat{K}}} \nu^2(h) + \frac{20}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \left(R_{k\Delta}^j \right)^2 + 2 \left(\text{pen}(K) - \text{pen}(\widehat{K}) \right). \quad (69)$$

Let $q : \mathcal{K}^2 \rightarrow \mathbb{R}_+$ such that $160q(K, K') \leq 18\text{pen}(K) + 16\text{pen}(K')$. Thus, on the set $\Omega_{n,N,K_{\max}}$, there exists a constant $C > 0$ such that for all $h \in \mathcal{S}_{K+M}$

$$\begin{aligned} \mathbb{E} \left[\left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right] &\leq 34 \left(\inf_{h \in \mathcal{S}_{K+M,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + \text{pen}(K) \right) \\ &\quad + 160 \mathbb{E} \left(\sup_{h \in \mathcal{T}_{K,\widehat{K}}} \nu_1^2(h) - q(K, \widehat{K}) \right) + C\Delta^2 \end{aligned}$$

where

$$\nu_1(h) := \frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} h(X_{k\Delta}^j) \zeta_{k\Delta}^{j,1} \quad (70)$$

with $\zeta_{k\Delta}^{j,1}$ the error term. We set for all $K, K' \in \mathcal{K}$,

$$G_K(K') := \sup_{h \in \mathcal{T}_{K,K'}} \nu_1^2(h) \quad (71)$$

and for N and n large enough, $\left\|\widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2\right\|_{n,N}^2 \leq 4(K+M)L$. We deduce that,

$$\begin{aligned} \mathbb{E} \left[\left\|\widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2\right\|_{n,N}^2 \right] &\leq \mathbb{E} \left[\left\|\widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2\right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,K_{\max}}}\right] + \mathbb{E} \left[\left\|\widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2\right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,K_{\max}}^c}\right] \\ &\leq 34 \inf_{K \in \mathcal{K}} \left(\inf_{h \in \mathcal{S}_{K+M}} \left\|h - \sigma_{|I}^2\right\|_n^2 + \text{pen}(K) \right) \\ &\quad + C\Delta^2 + 4(K+M)L\mathbb{P}(\Omega_{n,N,K_{\max}}^c) \\ &\quad + 160\mathbb{E} \left[\left(G_K(\widehat{K}) - q(K, \widehat{K}) \right)_+ \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right]. \end{aligned}$$

In the sequel, we refer to the *proof of Proposition 6.1 in Baraud et al. (2001)*. We known from *Lorentz et al* (see *Lorentz et al. (1996)*) that given the unit ball $\overline{B}_{\|\cdot\|_n}(0,1)$ of the approximation subspace \mathcal{S}_{K+M} with respect to norm $\|\cdot\|_n$ defined as follows:

$$\overline{B}_{\|\cdot\|_n}(0,1) = \{h \in \mathcal{S}_{K+M} : \|h\|_n \leq 1\} = \left\{ h \in \mathcal{S}_{K+M} : \|h\| \leq \frac{1}{\tau_0} \right\} = \overline{B}_2(0,1/\tau_0),$$

we can find a ε -net E_ε such that for each $\varepsilon \in (0,1]$, $|E_\varepsilon| \leq \left(\frac{3}{\varepsilon\tau_0}\right)^{K+M}$.

Recall that $\mathcal{T}_{K,K'} = \left\{ g \in \mathcal{S}_{K+M} + \mathcal{S}_{K'+M}, \|g\|_n = 1, \|g\|_\infty \leq \sqrt{L} \right\}$ and consider the sequence $(E_{\varepsilon_k})_{k \geq 1}$ of ε -net with $\varepsilon_k = \varepsilon_0 2^{-k}$ and $\varepsilon_0 \in (0,1]$. Moreover, set $N_k = \log(|E_{\varepsilon_k}|)$ for each $k \geq 0$. Then for each $g \in \mathcal{S}_{K+M} + \mathcal{S}_{K'+M}$ such that $\|g\|_\infty \leq \sqrt{L}$, there exists a sequence $(g_k)_{k \geq 0}$ with $g_k \in E_{\varepsilon_k}$ such that $g = g_0 + \sum_{k=1}^{\infty} g_k - g_{k-1}$. Set $\tilde{\mathbb{P}} := \mathbb{P}(\cdot \cap \Omega_{n,N,K_{\max}})$ and

$$\tau := \sigma_1^2 \sqrt{6x_0^{n,N}} + \sigma_1^2 \sqrt{L}x_0^{n,N} + \sum_{k \geq 1} \varepsilon_{k-1} \left\{ \sigma_1^2 \sqrt{6x_k^{n,N}} + 2\sigma_1^2 \sqrt{L}x_k^{n,N} \right\} = y_0^{n,N} + \sum_{k \geq 0} y_k^{n,N}.$$

For all $h \in \mathcal{T}_{K,K'}$ and on the event $\Omega_{n,N,K_{\max}}$, one has $\|h\|_{n,N}^2 \leq \frac{3}{2}\|h\|_n^2 = \frac{3}{2}$. Then, using the chaining technique of *Baraud et al. (2001)*, we have

$$\begin{aligned} \tilde{\mathbb{P}} \left(\sup_{h \in \mathcal{T}_{K,K'}} \nu_1(h) > \tau \right) &= \tilde{\mathbb{P}} \left(\exists (h_k)_{k \geq 0} \in \prod_{k \geq 0} E_{\varepsilon_k} / \nu_1(h) = \nu_1(h_0) + \sum_{k=1}^{\infty} \nu_1(h_k - h_{k-1}) > \tau \right) \\ &\leq \sum_{h_0 \in E_0} \tilde{\mathbb{P}} \left(\nu_1(h_0) > y_0^{n,N} \right) + \sum_{k=1}^{\infty} \sum_{\substack{h_k \in E_{\varepsilon_k} \\ h_{k-1} \in E_{\varepsilon_{k-1}}}} \tilde{\mathbb{P}} \left(\nu_1(h_k - h_{k-1}) > y_k^{n,N} \right). \end{aligned}$$

According to Equation (70) and Lemma 8.10, there exists a constant $C > 0$ such that

$$\begin{aligned} \tilde{\mathbb{P}} \left(\nu_1(h_0) > y_0^{n,N} \right) &\leq \tilde{\mathbb{P}} \left(\nu_1(h_0) > \sigma_1 \sqrt{6x_0^{n,N}} + \sigma_1^2 \|h_0\|_\infty x_0^{n,N} \right) \\ &\leq \exp(-CNn x_0^{n,N}), \\ \forall k \geq 1, \tilde{\mathbb{P}} \left(\nu_1(h_k - h_{k-1}) > y_k^{n,N} \right) &\leq \tilde{\mathbb{P}} \left(\nu_1(h_k - h_{k-1}) > \sigma_1 \sqrt{6x_k^{n,N}} + \sigma_1^2 \|h_k - h_{k-1}\|_\infty x_k^{n,N} \right) \\ &\leq \exp(-CNn x_k^{n,N}). \end{aligned}$$

Finally, since $N_k = \log(|E_{\varepsilon_k}|)$ for all $k \geq 0$, we deduce that

$$\begin{aligned} \tilde{\mathbb{P}} \left(\sup_{h \in \mathcal{T}_{K,K'}} \nu_1(h) > \tau \right) &\leq |E_{\varepsilon_0}| \exp(-CNn x_0^{n,N}) + \sum_{k=1}^{\infty} (|E_{\varepsilon_k}| + |E_{\varepsilon_{k-1}}|) \exp(-CNn x_k^{n,N}) \\ &\leq \exp \left(N_0 - CNn x_0^{n,N} \right) + \sum_{k=1}^{\infty} \exp \left(N_k + N_{k-1} - CNn x_k^{n,N} \right). \end{aligned}$$

We choose $x_0^{n,N}$ and $x_k^{n,N}$, $k \geq 1$ such that,

$$\begin{aligned} N_0 - CNn x_0^{n,N} &= -a(K + K' + 2M) - b \\ N_k + N_{k-1} - CNn x_k^{n,N} &= -k(K + K' + 2M) - a(K + K' + 2M) - b \end{aligned}$$

where a and b are two positive real numbers. We deduce that

$$x_k^{n,N} \leq C_0(1+k) \frac{K + K' + 2M}{Nn} \quad \text{and} \quad \tau \leq C_1 \sigma_1^2 \sqrt{\sqrt{L} \frac{K + K' + 2M}{Nn}} \quad (72)$$

with $C_0 > 0$ and C_1 two constants depending on a and b . It comes that

$$\tilde{\mathbb{P}} \left(\sup_{t \in \mathcal{T}_{K,K'}} \nu(t) > \tau \right) \leq \frac{e}{e-1} e^{-b} \exp \{ -a(K + K' + 2M) \}.$$

From Equation (72), we set

$$q(K, K') = \kappa^* \sigma_1^2 \sqrt{L} \frac{K + K' + 2M}{Nn}$$

where $\kappa^* > 0$ depends on $C_1 > 0$. Thus, for all $K, K' \in \mathcal{K}$,

$$\mathbb{P} \left(\left\{ \sup_{h \in \mathcal{T}_{K,K'}} \nu^2(h) > q(K, K') \right\} \cap \Omega_{n,N,K_{\max}} \right) \leq \frac{e^{-b+1}}{e+1} \exp \{ -a(K + K' + 2M) \}$$

and there exists constants $c, C > 0$ such that

$$\begin{aligned} \mathbb{E} \left[(G_K(K') - q(K, K'))_+ \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right] \\ \leq \frac{c(K + K')}{Nn} \mathbb{P} \left(\left\{ \sup_{t \in \mathcal{T}_{K,K'}} \nu^2(t) > q(K, K') \right\} \cap \Omega_{n,N,K_{\max}} \right) \\ \leq \frac{C}{Nn} \exp \left\{ -\frac{a}{2}(K + K') \right\}. \end{aligned}$$

Finally, there exists a real constant $C > 0$ such that,

$$\mathbb{E} \left[(G_K(\hat{K}) - q(K, \hat{K}))_+ \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right] \leq \sum_{K' \in \mathcal{K}} \mathbb{E} \left[(G_K(K') - q(K, K'))_+ \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right] \leq \frac{C}{Nn}.$$

We choose the penalty function pen such that for each $K \in \mathcal{K}$,

$$\text{pen}(K) \geq \kappa \sigma_1^2 \sqrt{L} \frac{K + M}{Nn}.$$

For N large enough, one has $\sigma_1^2 \leq \sqrt{L}$. Thus, we finally set $\text{pen}(K) = \kappa \frac{(K+M) \log(N)}{Nn}$ with $L = \log(N)$. Then, there exists a constant $C > 0$ such that,

$$\mathbb{E} \left[\left\| \hat{\sigma}_{\hat{K}}^2 - \sigma_{|I|}^2 \right\|_{n,N}^2 \right] \leq 34 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M}} \left\| h - \sigma_{|I|}^2 \right\|_n^2 + \text{pen}(K) \right\} + \frac{C}{Nn}.$$

□

8.4.2 Proof of Theorem 5.2

Proof. From Equation (22), we have

$$\widehat{K} := \arg \min_{K \in \mathcal{K}} \gamma_{n,N}(\widehat{\sigma}_{K+M,L}^2) + \text{pen}(K).$$

Then, for all $K \in \mathcal{K}$ and for all $h \in \mathcal{S}_{K+M,L}$, we have

$$\gamma_{n,N}(\widehat{\sigma}_{\widehat{K},L}^2) + \text{pen}(\widehat{K}) \leq \gamma_{n,N}(h) + \text{pen}(K).$$

Then, for all $K \in \mathcal{K}$ and for all $h \in \mathcal{S}_{K+M,L}$,

$$\begin{aligned} \gamma_{n,N}(\widehat{\sigma}_{\widehat{K},L}^2) - \gamma_{n,N}(\sigma^2) &\leq \gamma_{n,N}(h) - \gamma_{n,N}(\sigma^2) + \text{pen}(K) - \text{pen}(\widehat{K}) \\ \left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 &\leq \|h - \sigma^2\|_{n,N}^2 + 2\nu(\widehat{\sigma}_{\widehat{K},L}^2 - h) + 2\mu(\widehat{\sigma}_{\widehat{K},L}^2 - h) + \text{pen}(K) - \text{pen}(\widehat{K}). \end{aligned}$$

We have for all $a > 0$,

$$2\mathbb{E} \left[\mu \left(\widehat{\sigma}_{\widehat{K},L}^2 - h \right) \right] \leq \frac{2}{a} \mathbb{E} \left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 + \frac{2}{a} \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2\|_n^2 + \frac{a}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \mathbb{E} \left[\left(R_{k\Delta}^j \right)^2 \right]$$

and since $\nu = \nu_1 + \nu_2 + \nu_3$, according to the proof of Theorem 4.3, there exists a constant $c > 0$ such that

$$\mathbb{E} \left[\nu(\widehat{\sigma}_{\widehat{K},L}^2 - h) \right] \leq c \mathbb{E} \left[\nu_1(\widehat{\sigma}_{\widehat{K},L}^2 - h) \right]$$

where the for $i \in \{1, 2, 3\}$ and for all $h \in \mathcal{S}_{K+M,L}$, $K \in \mathcal{K}$,

$$\nu_i(h) = \frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} h(X_{k\Delta}^j) \zeta_{k\Delta}^j,$$

and the $\zeta_{k\Delta}^j$ are given Then,

$$\begin{aligned} \left(1 - \frac{2}{a} \right) \mathbb{E} \left[\left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 \right] &\leq \left(1 + \frac{2}{a} \right) \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2\|_n^2 + 2c \mathbb{E} \left[\nu_1(\widehat{\sigma}_{\widehat{K},L}^2 - h) \right] \\ &\quad + \text{pen}(K) - \text{pen}(\widehat{K}) + \frac{a}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \mathbb{E} \left[\left(R_{k\Delta}^j \right)^2 \right] \end{aligned}$$

From Equation (36) and for $a = 4$, there exists a constant $C > 0$ such that

$$\mathbb{E} \left[\left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 \right] \leq 3 \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2\|_n^2 + 4c \mathbb{E} \left[\nu_1(\widehat{\sigma}_{\widehat{K},L}^2 - h) \right] + 2 \left(\text{pen}(K) - \text{pen}(\widehat{K}) \right) + C\Delta^2. \quad (73)$$

Since for all $K \in \mathcal{K}$, $\text{pen}(K) \geq 2\kappa^* \sigma_1^2(K+M) \sqrt{2L}/(Nn)$, define the function $q : (K, K') \mapsto q(K, K')$ such that

$$q(K, K') = 2C^* \sigma_1^2 \frac{(K + K' + 2M) \sqrt{2L}}{Nn} \geq 2\sigma_1^2 v \sqrt{x^{n,N}} + \sigma_1^2 v x^{n,N} \quad (74)$$

where

$$x^{n,N} \propto \left(\frac{K + K' + 2M}{Nn} \right)^2 \quad \text{and} \quad v = \sqrt{2L}.$$

The constant $C^* > 0$ depends on constants $\kappa^* > 0$ and $c > 0$ of Equation (73) such that

$$4cq(K, K') \leq \text{pen}(K) + 2\text{pen}(K').$$

Then for all $K \in \mathcal{K}$ and for all $h \in \mathcal{S}_{K+M,L}$,

$$\mathbb{E} \left[\left\| \hat{\sigma}_{\hat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 \right] \leq 3 \left(\inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2\|_n^2 + \text{pen}(K) \right) + 4c \mathbb{E} \left[\left(\nu_1(\hat{\sigma}_{\hat{m},L}^2 - h) - q(K, \hat{K}) \right)_+ \right] + C\Delta^2. \quad (75)$$

For all $K \in \mathcal{K}$ and for all $h \in \mathcal{S}_{K+M,L}$ such that $\|h\|_\infty \leq \sqrt{L}$, we have ,

$$\left\| \hat{\sigma}_{\hat{K},L}^2 - h \right\|_{n,N}^2 \leq \left\| \hat{\sigma}_{\hat{K},L}^2 - h \right\|_\infty^2 \leq 2L =: v^2.$$

Then, using Equation (74) and Lemma 8.10, there exists a constant $C > 0$ such that for all $K, K' \in \mathcal{K}$ and for all $h \in \mathcal{S}_{K+M,L}$,

$$\mathbb{P} \left(\nu_1(\hat{\sigma}_{\hat{K}',L}^2 - h) \geq q(K, K'), \left\| \hat{\sigma}_{\hat{K},L}^2 - h \right\|_{n,N}^2 \leq v^2 \right) \leq \exp(-CNn x^{n,N}). \quad (76)$$

Since $L = \log(N)$, then for N large enough, $\sigma_1^2 \leq \sqrt{\log(N)}$, we finally choose

$$\text{pen}(K) = \kappa \frac{(K+M) \log(N)}{Nn}$$

where $\kappa > 0$ is a new constant. Since $\mathbb{E} \left[\nu_1(\hat{\sigma}_{\hat{K},L}^2 - h) \right] \leq O \left(\sqrt{\frac{(K_{\max}+M) \log^2(N)}{Nn}} \right)$ (see proof of Theorem 4.3), for all $K \in \mathcal{K}$ and $h \in \mathcal{S}_{K+M,L}$, there exists a constant $c > 0$ such that

$$\begin{aligned} \mathbb{E} \left[\left(\nu_1(\hat{\sigma}_{\hat{K},L}^2 - h) - q(K, \hat{K}) \right)_+ \right] &\leq \max_{K' \in \mathcal{K}} \left\{ \mathbb{E} \left[\left(\nu_1(\hat{\sigma}_{K',L}^2 - h) - q(K, K') \right)_+ \right] \right\} \\ &\leq cq(K, K_{\max}) \max_{K' \in \mathcal{K}} \left\{ \mathbb{P} \left(\nu_1(\hat{\sigma}_{K',L}^2 - h) \geq q(K, K') \right) \right\}. \end{aligned}$$

From Equation (76), we obtain that

$$\mathbb{E} \left[\left(\nu_1(\hat{\sigma}_{\hat{K},L}^2 - h) - q(K, \hat{K}) \right)_+ \right] \leq cq(K, K_{\max}) \exp(-CNn) \leq \frac{C}{Nn} \quad (77)$$

since K and K_{\max} increase with the size N of the sample paths $D_{N,n}$, and

$$cNnq(K, K_{\max}) \exp(-CNn) \rightarrow 0 \text{ as } N \rightarrow \infty.$$

Then, from Equations (77) and (75), there exists a constant $C > 0$ such that

$$\mathbb{E} \left[\left\| \hat{\sigma}_{\hat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 \right] \leq 3 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2\|_n^2 + \text{pen}(K) \right\} + \frac{C}{Nn}. \quad (78)$$

□

8.4.3 Proof of Theorem 5.3

Proof. The proof of Theorem 5.3 is similar to the proof of Theorem 5.1. Then, from Equation (69), for all $h \in \mathcal{S}_{K+M}$,

$$\left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \leq 3 \left\| h - \sigma_{|I}^2 \right\|_{n,1}^2 + 20 \sup_{h \in \mathcal{T}_{K,\widehat{K}}} \nu^2(h) + \frac{20}{n} \sum_{k=0}^{n-1} (R_{k\Delta}^1)^2 + 2 \left(\text{pen}(K) - \text{pen}(\widehat{K}) \right),$$

where $\mathcal{T}_{K,K'} = \{h \in \mathcal{S}_{K+M} + \mathcal{S}_{K'+M}, \|h\|_X = 1, \|h\|_\infty \leq \sqrt{L}\}$. Let $q : \mathcal{K}^2 \rightarrow \mathbb{R}_+$ such that $160q(K, K') \leq 18\text{pen}(K) + 16\text{pen}(K')$.

Recall that the \mathbb{L}^2 -norm $\|\cdot\|$, the norm $\mathbb{E}[\|\cdot\|_X]$ and the empirical norm $\|\cdot\|_n$ are equivalent on $\mathbb{L}^2(I)$ since the transition density is bounded on the compact interval I . Then, for all $K \in \mathcal{K}$ and $h \in \mathcal{S}_{K+M,L}$, we have

$$\mathbb{E} \left[\left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] \leq 3 \left(\inf_{h \in \mathcal{S}_{K+M,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + \text{pen}(K) \right) + 20 \mathbb{E} \left(\sup_{h \in \mathcal{T}_{K,\widehat{K}}} \nu_1^2(h) - q(K, \widehat{K}) \right) + C\Delta^2$$

where

$$\nu_1(h) := \frac{1}{n} \sum_{k=0}^{n-1} h(X_{k\Delta}^1) \zeta_{k\Delta}^{1,1}$$

with $\zeta_{k\Delta}^{1,1}$ the error term. We set for all $K, K' \in \mathcal{K}$, $G_K(K') := \sup_{h \in \mathcal{T}_{K,K'}} \nu_1^2(h)$. Then, there exists $C > 0$

such that

$$\mathbb{E} \left[\left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] \leq 3 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + \text{pen}(K) \right\} + 20 \sum_{K' \in \mathcal{K}} \mathbb{E} \left[(G_K(K') - q(K, K'))_+ \right] + C\Delta^2.$$

Considering the unit ball $\overline{B}_{\|\cdot\|_X}(0, 1)$ of the approximation subspace given by

$$\overline{B}_{\|\cdot\|_X}(0, 1) = \{h \in \mathcal{S}_{K+M}, \|h\|_X^2 \leq 1\} = \left\{ h \in \mathcal{S}_{K+M}, \|h\|^2 \leq \frac{1}{\tau_0} \right\}.$$

We obtain from the proof of Theorem 5.1 with $N = 1$ that,

$$\sum_{K' \in \mathcal{K}} \mathbb{E} \left[(G_K(K') - q(K, K'))_+ \right] \leq \frac{C}{n},$$

where $C > 0$ is a constant, $q(K, K') \propto \sigma_1^4 \frac{(K+K'+2M)\sqrt{\log(n)}}{n}$ and $\text{pen}(K) \propto \frac{(K+M)\log(n)}{n}$. Then we obtain

$$\mathbb{E} \left[\left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] \leq \frac{3}{\tau_0} \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + \text{pen}(K) \right\} + \frac{C}{n}.$$

□

References

- Bandi, F.-M. & Phillips, P.-C.-B. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica* **71**, 241–283.
- Baraud, Y., Comte, F. & Viennet, G. (2001). Model selection for (auto-) regression with dependent data. *ESAIM: Probability and Statistics* **5**, 33–49.
- Barron, A., Lucien, B. & Pascal, M. (1999). Risk bounds for model selection via penalization. *Probability theory and related fields* **113**, 301–413.
- Clement, E. (1997). Estimation of diffusion processes by simulated moment methods. *Scandinavian Journal of Statistics* **24**, 353–369.
- Comte, F. (2017). *Nonparametric Estimation*. Spartacus IDH.
- Comte, F. & Genon-Catalot, V. (2020a). Nonparametric drift estimation for iid paths of stochastic differential equations. *The Annals of Statistics* **48**, 3336–3365.
- Comte, F. & Genon-Catalot, V. (2020b). Regression function estimation as a partly inverse problem. *Annals of the Institute of Statistical Mathematics* **72**, 1023–1054.
- Comte, F. & Genon-Catalot, V. (2021). Drift estimation on non compact support for diffusion models. *Stochastic Processes and their Applications* **134**, 174–207.
- Comte, F., Genon-Catalot, V., Rozenholc, Y. *et al.* (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* .
- Comte, F., Genon-Catalot, V. *et al.* (2018). Regression function estimation as a partly inverse problem. *Preprint Hal* .
- Denis, C., Dion-Blanc, C. & Martinez, M. (2021). A ridge estimator of the drift from discrete repeated observations of the solutions of a stochastic differential equation. *Bernoulli* .
- Denis, C., Dion-Blanc, C., Mintsu, E. E. & Tran, V.-C. (2022). Nonparametric plug-in classifier for multiclass classification of sde paths. *arXiv preprint arXiv:2212.10259* .
- DeVore, R. A. & Lorentz, G. G. (1993). *Constructive approximation*, vol. 303. Springer Science & Business Media.
- Florens, D. (1998). Estimation of the diffusion coefficient from crossings. *Statistical Inference for Stochastic Processes* **1**, 175–195.
- Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of applied probability* **30**, 790–804.
- Genon-Catalot, V. & Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. In *Annales de l’IHP Probabilités et statistiques*, vol. 29, pp. 119–151.
- Genon-Catalot, V. & Jacod, J. (1994). Estimation of the diffusion coefficient for diffusion processes: random sampling. *Scandinavian Journal of Statistics* pp. 193–221.
- Genon-Catalot, V., Jeantheau, T. & Laredo, C. (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* pp. 855–872.

- Genon-Catalot, V., Laredo, C. & Picard, D. (1992). Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics* pp. 317–335.
- Gloter, A. (2000). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM: Probability and Statistics* **4**, 205–227.
- Gobet, E. (2002). Lan property for ergodic diffusions with discrete observations. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics* **38**, 711–737.
- Gourieroux, C., Nguyen, H.-T. & Sriboonchitta, S. (2017). Nonparametric estimation of a scalar diffusion model from discrete time data: a survey. *Annals of Operations Research* **256**, 203–219.
- Györfi, L., Kohler, M., Krzyzak, A. & Walk, H. (2006). *A distribution-free theory of nonparametric regression*. Springer Science & Business Media.
- Hastie, T., Tibshirani, R. & Friedman, J. (2001). Springer series in statistics the elements of statistical learning data mining, inference. Tech. rep., and Prediction. Technical report.
- Hoffmann, M. (1997). Minimax estimation of the diffusion coefficient through irregular samplings. *Statistics & probability letters* **32**, 11–24.
- Hoffmann, M. (1999a). Adaptive estimation in diffusion processes. *Stochastic processes and their Applications* **79**, 135–163.
- Hoffmann, M. (1999b). Lp estimation of the diffusion coefficient. *Bernoulli* pp. 447–481.
- Iacus, S.-M. (2009). *Simulation and inference for stochastic differential equations: with R examples*. Springer Science & Business Media.
- Jacod, J. (1993). Random sampling in estimation problems for continuous gaussian processes with independent increments. *Stochastic processes and their applications* **44**, 181–204.
- Lorentz, G.-G., Golitschek, M.-V. & Makovoz, Y. (1996). *Constructive approximation: advanced problems*, vol. 304. Springer.
- Odell-Scott, D. W. (1992). Multivariate density estimation: theory, practice, and visualization.
- Park, J.-Y. & Wang, B. (2021). Nonparametric estimation of jump diffusion models. *Journal of Econometrics* **222**, 688–715.
- Renò, R. (2006). Nonparametric estimation of stochastic volatility models. *Economics Letters* **90**, 390–395.
- Revuz, D. & Yor, M. (2013). *Continuous martingales and Brownian motion*, vol. 293. Springer Science & Business Media.
- Schmisser, E. (2012). Non-parametric estimation of the diffusion coefficient from noisy data. *Statistical inference for stochastic processes* **15**, 193–223.
- Schmisser, E. (2019). Non parametric estimation of the diffusion coefficients of a diffusion with jumps. *Stochastic processes and their applications* **129**, 5364–5405.
- Sørensen, H. (2002). Estimation of diffusion parameters for discretely observed diffusion processes. *Bernoulli* pp. 491–508.
- Soulier, P. (1993). Estimation fonctionnelle dans divers cadres de dépendance. Ph.D. thesis, Paris 11.

Appendix

Calibration

Fix the drift function $b(x) = 1 - x$, the time-horizon $T = 1$ and at time $t = 0$, $x_0 = 0$. Consider the following three models:

Model 1: $\sigma(x) = 1$

Model 2: $\sigma(x) = 0.1 + 0.9/\sqrt{1 + x^2}$

Model 3: $\sigma(x) = 1/3 + \sin^2(2\pi x)/\pi + 1/(\pi + x^2)$.

The three diffusion models satisfy Assumption 2.1 and are used to calibrate the numerical constant κ of the penalty function given in Theorem 5.1

As we already know, the adaptive estimator of σ^2 on the interval $[-\sqrt{\log(N)}, \sqrt{\log(N)}]$ necessitate a data-driven selection of an optimal dimension through the minimization of the penalized least squares contrast given in Equation (22). Since the penalty function $\text{pen}(d_N) = \kappa(K_N + M) \log^2(N)/N^2$ depends on the unknown numerical constant $\kappa > 0$, the goal is to select an optimal value of κ in the set $\mathcal{V} = \{0.1, 0.5, 1, 2, 4, 5, 7, 10\}$ of its possible values. To this end, we repeat 100 times the following steps:

1. Simulate learning samples D_N and $D_{N'}$ with $N \in \{50, 100\}$, $N' = 100$ and $n \in \{100, 250\}$
2. For each $\kappa \in \mathcal{V}$:
 - (a) For each $K_N \in \mathcal{K}$ and from D_N , compute $\widehat{\sigma}_{d_N, L_N}^2$ given in Equations (9) and (10).
 - (b) Select the optimal dimension $\widehat{K}_N \in \mathcal{K}$ using Equation (22)
 - (c) Using the learning sample $D_{N'}$, evaluate $\left\| \widehat{\sigma}_{\widehat{d}_N, L_N}^2 - \sigma_A^2 \right\|_{n, N'}^2$ where $\widehat{d}_N = \widehat{K}_N + M$.

Then, we calculate average values of $\left\| \widehat{\sigma}_{d_N, L_N}^2 - \sigma_A^2 \right\|_{n, N'}^2$ for each $\kappa \in \mathcal{V}$ and obtain the following results:

We finally choose $5 \in \mathcal{V}$ as the optimal value of κ in reference to the results of Figure 2.

Proof of Lemma 8.10

Proof. We obtain from Comte, Genon-Catalot, Rozenholc (2007) proof of Lemma 3 that for each $j \in \llbracket 1, N \rrbracket$, $k \in \llbracket 0, n-1 \rrbracket$ and $p \in \mathbb{N} \setminus \{0, 1\}$

$$\mathbb{E} \left[\exp \left(ug(X_{k\Delta}^j) \xi_{k\Delta}^{j,1} - \frac{au^2 g^2(X_{k\Delta}^j)}{1 - bu} \right) \middle| \mathcal{F}_{k\Delta} \right] \leq 1$$

with $a = e(4\sigma_1^2 c^2)^2$, $b = 4\sigma_1^2 c^2 e \|g\|_\infty$, $u \in \mathbb{R}$ such that $bu < 1$ and $c > 0$ a real constant. Thus,

$$\begin{aligned} \mathbb{P} \left(\frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} g(X_{k\Delta}^j) \xi_{k\Delta}^{j,1} \geq \varepsilon, \|g\|_{N,n}^2 \leq v^2 \right) &= \mathbb{E} \left(\mathbf{1}_{\left\{ \sum_{j=1}^N \sum_{k=0}^{n-1} ug(X_{k\Delta}^j) \xi_{k\Delta}^{(j,1)} \geq Nnu\varepsilon \right\}} \mathbf{1}_{\|g\|_{N,n}^2 \leq v^2} \right) \\ &= \mathbb{E} \left(\mathbf{1}_{\left\{ \exp \left(\sum_{j=1}^N \sum_{k=0}^{n-1} ug(X_{k\Delta}^j) \xi_{k\Delta}^{(j,1)} \right) e^{-Nnu\varepsilon} \geq 1 \right\}} \mathbf{1}_{\|g\|_{N,n}^2 \leq v^2} \right) \\ &\leq e^{-Nnu\varepsilon} \mathbb{E} \left[\mathbf{1}_{\|g\|_{N,n}^2 \leq v^2} \exp \left\{ \sum_{j=1}^N \sum_{k=0}^{n-1} ug(X_{k\Delta}^j) \xi_{k\Delta}^{j,1} \right\} \right]. \end{aligned}$$

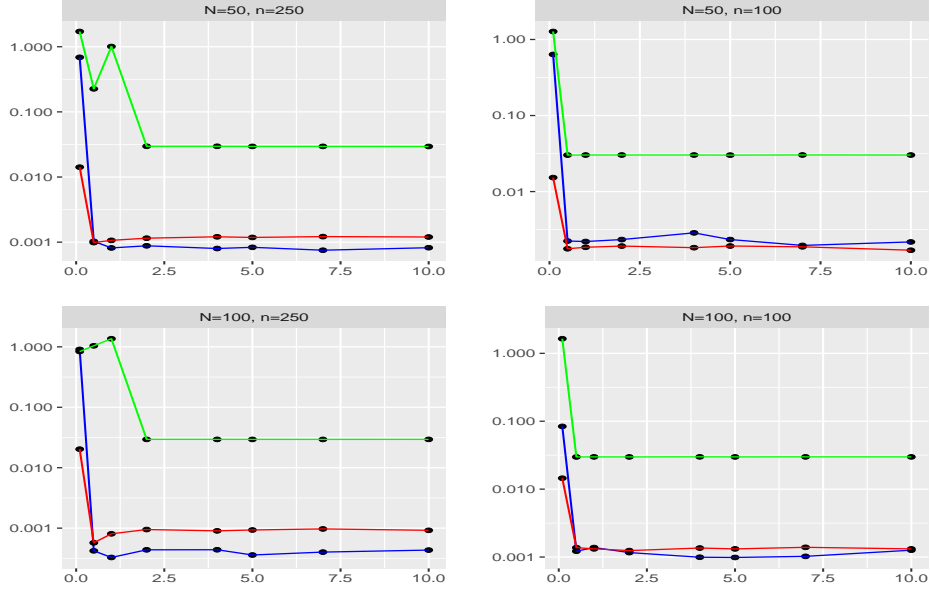


Figure 2: Calibration of the constant $\kappa \in \mathcal{V}$ of the penalty function

It follows that,

$$\mathbb{P} \left(\frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} g(X_{k\Delta}^j) \xi_{k\Delta}^{j,1} \geq \varepsilon, \|g\|_{n,N}^2 \leq v^2 \right) \leq \exp \left\{ -Nnu\varepsilon + \frac{Nnau^2v^2}{1-bu} \right\}.$$

We set $u = \frac{\varepsilon}{\varepsilon b + 2av^2}$. Then, we have $-Nnu\varepsilon + Nnau^2v^2/(1-bu) = -Nn\varepsilon^2/2(\varepsilon b + 2av^2)$ and,

$$\begin{aligned} \mathbb{P} \left(\frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} g(X_{k\Delta}^j) \xi_{k\Delta}^{j,1} \geq \varepsilon, \|g\|_{n,N}^2 \leq v^2 \right) &\leq \exp \left(-\frac{Nn\varepsilon^2}{2(\varepsilon b + av^2)} \right) \\ &\leq \exp \left(-C \frac{Nn\varepsilon^2}{\sigma_1^2 (\varepsilon \|g\|_\infty + 4\sigma_1^2 v^2)} \right) \end{aligned}$$

where $C > 0$ is a constant depending on $c > 0$. □

Proof of Lemma 8.9

Proof. Set $K_{n,N} = K_N$ since $N \propto n$. Let us remind the reader of the Gram matrix Ψ_{K_N} given in Equation (19),

$$\Psi_{K_N} = \mathbb{E} \left[\frac{1}{Nn} \mathbf{F}'_{K_N} \mathbf{F}_{K_N} \right] = \mathbb{E} \left(\widehat{\Psi}_{K_N} \right)$$

where,

$$\mathbf{F}_{K_N} := \left((B_\ell(X_0^j), \dots, (B_\ell(X_{(n-1)\Delta}^j))) \right)_{\substack{0 \leq \ell \leq K_N - 1 \\ 1 \leq j \leq N}} \in \mathbb{R}^{Nn \times (K_N + M)} \quad (79)$$

The empirical counterpart $\widehat{\Psi}$ is the random matrix given by $\widehat{\Psi}_{K_N}$ of size $(K_N + M) \times (K_N + M)$ is given by

$$\widehat{\Psi}_{K_N} := \frac{1}{Nn} \mathbf{F}'_{K_N} \mathbf{F}_{K_N} = \left(\frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} f_\ell(X_{k\Delta}^j) f_{\ell'}(X_{k\Delta}^j) \right)_{\ell, \ell' \in [-M, K_N - 1]} \quad (80)$$

For all $t = \sum_{\ell=-M}^{K_N-1} a_\ell B_{\ell,M}, \mathbf{u} \in S_{K_N,M}$ one has

$$\|t\|_{n,N}^2 = a' \widehat{\Psi}_{K_N} a \quad \text{and} \quad \|t\|_n^2 = a' \Psi_{K_N} a, \quad \text{with } a = (a_{-M}, \dots, a_{K_N-1})'.$$

Under Assumption 2.1, we follow the lines of Comte *et al.* (2018) *Proposition 2.3* and *Lemma 6.2*. Then,

$$\begin{aligned} \sup_{t \in S_{K_N,M}, \|t\|_n=1} \left| \|t\|_{n,N}^2 - \|t\|_n^2 \right| &= \sup_{w \in \mathbb{R}^{K_N+M}, \|\Phi_{K_N}^{1/2} w\|_{2,K_N+M}=1} \left| w' \left(\widehat{\Psi}_{K_N} - \Psi_{K_N} \right) w \right| \\ &= \sup_{u \in \mathbb{R}^{K_N+M}, \|u\|_{2,K_N+M}=1} \left| u' \Psi_{K_N}^{-1/2} \left(\widehat{\Psi}_{K_N} - \Psi_{K_N} \right) \Psi_{K_N}^{-1/2} u \right| \\ &= \left\| \Psi_{K_N}^{-1/2} \widehat{\Psi}_{K_N} \Psi_{K_N}^{-1/2} - \text{Id}_{K_N+M} \right\|_{\text{op}}. \end{aligned}$$

Therefore,

$$\Omega_{n,N,K_N}^c = \left\{ \left\| \Psi_{K_N}^{-1/2} \widehat{\Psi}_{K_N} \Psi_{K_N}^{-1/2} - \text{Id}_{K_N+M} \right\|_{\text{op}} > 1/2 \right\}.$$

Since $A_N = o\left(\sqrt{\log(N)}\right)$, we obtain from Denis *et al.* (2022), *proof of Lemma 7.8*, there exists a constant $C > 0$ such that

$$\mathbb{P}\left(\Omega_{n,N,K_N}^c\right) \leq 2(K_N + M) \exp\left(-C \log^{3/2}(N)\right). \quad (81)$$

Finally, since $2(K_N + M) \exp\left(-C/2 \log^{3/2}(N)\right) \rightarrow 0$ as $N \rightarrow +\infty$, one concludes from Equation (81) and for N large enough,

$$\mathbb{P}\left(\Omega_{n,N,K_N}^c\right) \leq C \exp\left(-c \log^{3/2}(N)\right)$$

where $c > 0$ and $C > 0$ are new constants. □