



**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. 2024.  
hal-04156441v2

**HAL Id: hal-04156441**

**<https://hal.science/hal-04156441v2>**

Preprint submitted on 18 Mar 2024

**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 i.i.d. S.D.E. paths

Eddy Ella-Mintsa  
LAMA, Université Gustave Eiffel.

March 18, 2024

## 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  $X$  that are obtained by minimization of the least squares contrast. We prove that the estimators are consistent and derive rates of convergence as the number of 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:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0 \in \mathbb{R} \quad (1)$$

where  $(W_t)_{t \geq 0}$  is a standard Brownian motion. The drift function  $b$  and the diffusion coefficient  $\sigma$  are assumed to be unknown.

For a number of years now, there has been a great deal of interest in stochastic processes that are solutions of stochastic differential equations, also known as diffusion processes. This mathematical tool has been the subject of many studies in the modelling of complex random phenomena in a number of fields such as finance, with the modelling of the price of a financial asset (see *e.g.* Jaillet *et al.* (1990), Lambertson & Lapeyre (2011), El Karoui *et al.* (1997)), or the modelling of interest rates (see *e.g.* Kijima *et al.* (2009), Duffie & Kan (1996)), in cell biology (see *e.g.* Sbalzarini (2006), Erban & Chapman (2009)), population genetics or population dynamics (see *e.g.* Etheridge (2011), Ethier & Norman (1977)), or in physics (see *e.g.* Domingo *et al.* (2020)). The estimation of coefficients of stochastic differential equations plays a key role in the study of these processes. It is the case for example for the construction of some classification procedures for diffusion paths (see Gadat *et al.* (2020), Denis *et al.* (2020)), the automation of which is actively contributing to the development of several scientific and economic sectors. The analysis of the diffusion coefficient and its estimation plays a significant role in finance, with the volatility study of a financial asset.

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  $Nn$  observations  $\left\{X_{k\Delta_n}^j, k = 0, \dots, n-1, j = 1, \dots, N\right\}$  coming from  $N$  independent discrete time copies  $\{\bar{X}^1, \dots, \bar{X}^N\}$  of  $X$  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  $\Delta_n$  tends to zero as  $n$  tends to infinity. Furthermore, we consider estimators of  $\sigma^2(\cdot)$  built from a single diffusion path and the  $n$  observations  $(X_{k\Delta_n}, k = 0, \dots, n)$  ( $N = 1$ ), and those built when  $N$  tends to infinity. In this paper, we first propose a ridge estimator of  $\sigma^2(\cdot)$  on a compact interval. Secondly, we focus on the nonparametric estimation of  $\sigma^2(\cdot)$  on a non-compact interval, especially the real line  $\mathbb{R}$ . We measure the risk of any estimator  $\hat{\sigma}^2$  of the square of the diffusion coefficient  $\sigma^2$  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  $Nn$  observations.

**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  $\sigma^2$ .

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  $Nn$  observations of the diffusion process, with  $N, n \rightarrow \infty$ . In Denis *et al.* (2022), a nonparametric estimator of  $\sigma^2$  on the real line  $\mathbb{R}$  is proposed 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  $\sigma^2$  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  $\sigma^2$  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 paper, we assume to have at our disposal the following  $Nn$  observations

$$\left\{ X_{k\Delta_n}^j, j = 1, \dots, N; k = 0, \dots, n-1 \right\},$$

gathered in the  $N$  independent sample paths  $\left\{ \bar{X}^j = (X_{k\Delta_n}^j)_{0 \leq k \leq n-1}, j = 1, \dots, N \right\}$  which are discrete-time copies of the diffusion process  $X$  solution of Equation (1). The main objectives of this paper are the following.

1. The construction of a consistent ridge estimator of  $\sigma^2$  on a compact interval from a single diffusion path ( $N = 1$ ), using the least squares approach. In the literature and in the same framework, Hoffmann (1999b) proposed a projection estimator of  $\sigma^2$  on the compact interval  $[0, 1]$ , over an approximation space spanned by the wavelet basis. The author established an optimal rate of order  $n^{-ps/2(2s+1)}$  on a Besov space of parameter  $s$ , defined from the space  $L^p([0, 1])$ . To obtain this result, it is necessary to be sure that the data points are uniformly distributed over the compact interval  $[0, 1]$  using the local time of the diffusion process  $X$ , and working conditionally on the event "the local time is lower bounded on  $[0, 1]$ ". In this paper, we do not establish our results conditionally on the same event, we rather take advantage of the assumptions made on the coefficients  $b$  and  $\sigma$ , the properties of the local time, and on the approximation of the transition density by Gaussian densities proposed in Gobet (2002), to prove that the data points are uniformly scattered in the estimation interval. We establish, over a Hölder space, a rate of the same order as the one obtained in Hoffmann (1999b) with  $p = 2$ . Moreover, we propose an adaptive estimator of  $\sigma^2$  based on a data-driven selection of the dimension through the minimization of a penalized least squares contrast.
2. The construction of a nonparametric estimator of  $\sigma^2$  on a non-compact interval and from  $n + 1$  observations  $(X_{k\Delta_n}, k = 0, \dots, n)$  of a single diffusion path ( $N = 1$ ), particularly on the real line  $\mathbb{R}$ . In the literature, only Florens-Zmirou (1993) studied the nonparametric estimation of  $\sigma^2$  on a non-compact interval from a single diffusion path. The author proposed a kernel estimator of  $\sigma^2$  and proved its consistency. No rate of convergence has been established. In this article, we propose a projection estimator  $\sigma^2$  using the least squares approach. We establish a rate of convergence, and propose an adaptive estimator through a data-driven selection of the dimension.
3. For the case of repeated observations of the diffusion process ( $N \rightarrow \infty$ ), only Denis *et al.* (2022) tackled the problem of estimation of  $\sigma^2$  and established a rate of order  $N^{-1/5}$  over a space of Lipschitz functions. In this paper, we propose projection estimators of  $\sigma^2$  both on a compact interval and on a non-compact interval, particularly on the real line  $\mathbb{R}$ . We establish rates of convergence and propose adaptive estimators for both the compact and the non-compact cases.
4. 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  $\sigma^2$  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  $\mathbb{R}$ .

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	$\log^{\beta/2}(N)(Nn)^{-\beta/(2\beta+1)}, \quad N \propto n$
$I = \mathbb{R}$	$\log^{\beta}(n)n^{-\beta/(4\beta+1)}$	$\log^{\beta}(N)(Nn)^{-\beta/(4\beta+1)} + n^{-1}$

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 discrete observation of the diffusion process both on a compact interval and on the real line  $\mathbb{R}$ . In Section 4, we study the non-adaptive estimation of the diffusion coefficient from repeated discrete observations of the diffusion process. We propose in Section 5, adaptive estimators of the diffusion coefficient, considering the nature of the estimation interval together with the number of discrete observations of the process, and Section 6 complete the study with numerical evaluation of the performance of estimators. A conclusion and some perspectives are given in Section 7, and 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  $\sigma$  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) (see Karatzas & Shreve (2014)), and this unique solution admits a transition density  $(t, x_0, x) \mapsto p_X(t, x_0, x)$  (see e.g. Gobet (2002)). Besides, we obtain from Assumption 2.1 the following result.

**Lemma 2.2.** *Under Assumption 2.1, it holds:*

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

The above result shows that under Assumption 2.1, the diffusion process solution of Equation (1) admits moments of any order  $q \in \mathbb{N}^*$ . The proof of Lemma 2.2 is provided in appendix.

## 2.1 Definitions and notations

We suppose to have at our disposal  $Nn$  observations  $D_{N,n} = \{X_{k\Delta}^j, k = 0, \dots, n, j = 1, \dots, N\}$  constituted of  $N$  independent copies of the discrete observation  $(X_{k\Delta_n}, k = 0, \dots, n)$  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  $\sigma^2$  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  $\sigma^2$  on the interval  $I$  from  $n+1$  observations  $D_{1,n} = \{X_{k\Delta_n}^1, k = 0, \dots, n\}$  ( $N = 1$ ), with  $n \rightarrow \infty$ . 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:

$$\|h\|_n^2 := \mathbb{E}_X \left[ \frac{1}{n} \sum_{k=0}^{n-1} h^2(X_{k\Delta_n}) \right], \quad \|h\|_{n,N}^2 := \frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} h^2(X_{k\Delta_n}^j), \quad \|h\|_X^2 := \int_0^1 h^2(X_s) ds \quad (2)$$

where  $\mathbb{E}_X$  is the corresponding expectation of the probability distribution of the diffusion process  $X$  solution of Equation (1). Besides, 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  $f_n : x \mapsto \frac{1}{n} \sum_{k=0}^{n-1} p_X(k\Delta_n, x_0, x)$  is a density function. For the case of non-adaptive estimators of  $\sigma^2$ , 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  $M$ .

In the sequel, without loss of generality and for the simplification of notations, we set  $x_0 = 0$ . The transition density is therefore defined by  $p_X : (t, x) \mapsto p_X(t, x)$ . Thus, for any solution  $X$  of Equation (1) such that  $X_0 = x_0 \in \mathbb{R} \setminus \{0\}$ , we consider the diffusion process  $Y = (X_t - x_0)_{t \in [0,1]}$  solution of

$$dY_t = \tilde{b}(Y_t)dt + \tilde{\sigma}(Y_t)dW_t, \quad Y_0 = 0,$$

with  $\tilde{b} : x \mapsto b(x + x_0)$  and  $\tilde{\sigma} : x \mapsto \sigma(x + x_0)$ . Besides, setting the time horizon to  $T = 1$  is without loss of generality.

For any real number  $x \in \mathbb{R}_{+*}$ , we denote by  $[x] \in \mathbb{N}$  the largest integer strictly smaller than  $x$ , and for all integer  $p, q \in \mathbb{N} \setminus \{0, 1\}$  such that  $p < q$ , we denote by  $\llbracket p, q \rrbracket$  the set of integers  $\{p, p+1, \dots, q\}$ .

## 2.2 Spaces of approximation

We propose projection estimators of  $\sigma^2$  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:

$$\mathcal{S}_m := \text{Span}(\phi_\ell, \ell = 0, \dots, m-1), \quad m \geq 1 \quad (3)$$

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

$$\mathcal{S}_{m,L} := \left\{ h = \sum_{\ell=0}^{m-1} a_\ell \phi_\ell, \quad \sum_{\ell=0}^{m-1} a_\ell^2 = \|\mathbf{a}\|_2^2 \leq mL, \quad \mathbf{a} = (a_0, \dots, a_{m-1}), \quad L > 0 \right\}. \quad (4)$$

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  $\sigma^2$  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 example 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 smooth 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 (3) and (4)) 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  $\sigma^2$  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 in the following bias term:

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

where the constant  $C > 0$  depends on  $\beta, R$  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

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

The above Fourier basis is defined on the compact interval  $[0, 1]$ . The definition can be extended to any compact interval  $I \subset \mathbb{R}$ , 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  $\sigma^2$  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  $[s]^{th}$  derivative  $f^{([s])}$  belongs to the space  $\mathcal{B}_{2,\infty}^{s-[s]}(I)$  given by

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

where for  $s - [s] \in (0, 1)$ ,  $w_{2,f}(t) = \sup_{|h| \leq t} \|\tau_h f - f\|_2$  with  $\tau_h f(x) = f(x - h)$ , and for  $s - [s] = 1$ ,

$w_{2,f}(t) = \sup_{|h| \leq t} \|\tau_h f + \tau_{-h} f - 2f\|_2$ . Thus, if we assume that the function  $\sigma_{|I}^2$  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  $\tau_1$  of Equation (9), the smoothness parameter  $s$  of the Besov space such that

$$\inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 \leq \tau_1 \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|^2 \leq C \left| \sigma_{|I}^2 \right|_\beta^2 m^{-2\beta}$$

where  $|\sigma_{|I}^2|_s$  is the semi-norm of  $\sigma_{|I}^2$  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:

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

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 each  $j \geq 1$ , the function  $h_j$  satisfies  $|h_j(x)| \leq c|x| \exp(-c_0 x^2)$  for all  $x \in \mathbb{R}$  such that  $x^2 \geq (3/2)(4j + 3)$ , 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  $\sigma^2$  on the real line  $\mathbb{R}$ .

If one assumes that  $\sigma^2$  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  $\ell \geq 1$ ,  $g_\ell$  is the  $L^2(\mathbb{R}, f_n(x)dx)$ -orthogonal projection of  $g$  on the  $\ell$ -dimensional vector space  $\mathcal{S}_\ell$  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  $\ell \geq 1$ ,  $g_\ell$  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  $W_{f_n}^s(I, R)$  where  $I = \mathbb{R}$  and  $(\phi_j)_{j \geq 0}$  is the Hermite basis.

**[L] The Laguerre basis.** The Laguerre basis  $(\ell_j)_{j \geq 1}$  is defined from the Laguerre functions defined on  $\mathbb{R}^+$  and given for each  $j \geq 1$  by

$$\ell_j(x) = \sqrt{2}L_j(2x)e^{-x}\mathbf{1}_{x \geq 0}, \quad \text{with} \quad L_j(x) = \sum_{k=0}^j (-1)^k \binom{j}{k} \frac{x^k}{k!}.$$

The collection [L] is a complete orthonormal basis of  $\mathbb{L}^2(\mathbb{R}^+)$ .

**[CS–OB] Other orthonormal bases.** We consider any orthonormal basis  $(\phi_0, \dots, \phi_{m-1})$  of compactly supported functions that satisfies the following condition:

$$\exists r \in \mathbb{N} : \mathfrak{L}(m) := \sum_{j=0}^{m-1} \|\phi_j\|_\infty^2 = O(m^r), \quad \text{and} \quad \exists r' \in \mathbb{N}^* : \mathfrak{R}(m) := \sum_{j=0}^{m-1} \|\phi'_j\|_\infty^2 = O(m^{r'}). \quad (5)$$

Thus, under Condition (5), any function  $h \in \mathcal{S}_m$  satisfies  $\|h\|_\infty = O(m^{r/2}\|\mathbf{a}\|_2)$  and  $\|h'\|_\infty = O(m^{r'/2}\|\mathbf{a}\|_2)$ , and, applying the finite increments theorem, there exists a constant  $C > 0$  such that for all  $x, y \in \text{Support}(h)$ ,

$$|h(x) - h(y)| \leq Cm^{r'/2}\|\mathbf{a}\|_2|x - y|.$$

Among these bases, we have the Legendre basis with  $r = r' = 2$ , the orthonormal basis of piecewise polynomials on  $[0, 1]$  with  $r = r' = 2$ , the Fourier basis with  $r = 1$  and  $r' = 3$ .

**Remark 2.3.** *The B-spline basis is used for the estimation of  $\sigma^2$  on a compact interval on one side ( $N = 1$  and  $N > 1$ ), and on the real line on the other side restricting  $\sigma^2$  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  $\sigma^2$  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  $\sigma^2$  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  $\sigma^2$  whose support is the real line  $\mathbb{R}$ .*

**Remark 2.4.** *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 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  $[\mathbf{CS} - \mathbf{OB}]$ ,  $[\mathbf{NCS} - \mathbf{OB}]$  and  $[\mathbf{B}]$ , the respective collections of subspaces spanned by the orthonormal bases of compactly supported functions, the orthonormal bases of non-compactly supported functions, and the spline basis. Note that on one side, the collection  $[\mathbf{CS} - \mathbf{OB}]$  includes the Fourier basis related to the collection  $[\mathbf{F}]$ , and on the other side, the collection  $[\mathbf{NCS} - \mathbf{OB}]$  includes the collection  $[\mathbf{H}]$  for the Hermite basis, and the collection  $[\mathbf{L}]$  for the Laguerre basis. Finally, we denote by  $[\mathbf{OB}]$ , the merger of the two collections  $[\mathbf{CS} - \mathbf{OB}]$  and  $[\mathbf{NCS} - \mathbf{OB}]$ .

### 2.3 Ridge estimators of the square of the diffusion coefficient

We establish from Equation (1) and the sample paths  $D_{N,n}$  the following regression model for the estimation of  $\sigma^2$ :

$$U_{k\Delta_n}^j = \sigma^2(X_{k\Delta_n}^j) + \zeta_{k\Delta_n}^j + R_{k\Delta_n}^j, \quad \forall (j, k) \in \llbracket 1, N \rrbracket \times \llbracket 0, n-1 \rrbracket \quad (6)$$

where for each pair  $(j, k) \in \llbracket 1, N \rrbracket \times \llbracket 0, n-1 \rrbracket$ , the response variable  $U_{k\Delta_n}^j$  is given by

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

and the error terms are respectively given by  $\zeta_{k\Delta}^j = \zeta_{k\Delta}^{j,1} + \zeta_{k\Delta}^{j,2} + \zeta_{k\Delta}^{j,3}$ , with:

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

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

$$\zeta_{k\Delta_n}^{j,3} = 2b(X_{k\Delta_n}^j) \int_{k\Delta_n}^{(k+1)\Delta_n} \sigma(X_s^j) dW_s^j,$$

and  $R_{k\Delta}^j = R_{k\Delta}^{j,1} + R_{k\Delta}^{j,2} + R_{k\Delta}^{j,3}$ , with:

$$R_{k\Delta_n}^{j,1} = \frac{1}{\Delta_n} \left( \int_{k\Delta_n}^{(k+1)\Delta_n} b(X_s^j) ds \right)^2, \quad R_{k\Delta_n}^{j,3} = \frac{1}{\Delta_n} \int_{k\Delta_n}^{(k+1)\Delta_n} ((k+1)\Delta_n - s) \Phi(X_s^j) ds$$

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

where

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

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  $d\langle X, X \rangle_t = \sigma^2(X_t) dt$ . The proof of the establishment of the regression model is given in appendix.

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

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

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

$$\hat{\sigma}_m^2 \in \arg \min_{h \in \mathcal{S}_{m,L}} \gamma_{n,N}(h). \quad (7)$$

Indeed, for each dimension  $m \in \mathcal{M}$ , the estimator  $\hat{\sigma}_m^2$  of  $\sigma^2$  given in Equation (7) satisfies  $\hat{\sigma}_m^2 = \sum_{\ell=0}^{m-1} \hat{a}_\ell \phi_\ell$ , where

$$\hat{\mathbf{a}} = (\hat{a}_0, \dots, \hat{a}_{m-1}) := \arg \min_{\|\mathbf{a}\|_2^2 \leq mL} \|\mathbf{U} - \mathbf{F}_m \mathbf{a}\|_2^2 \quad (8)$$

with  $\mathbf{U} = (U_0^1, \dots, U_{(n-1)\Delta_n}^1, \dots, U_0^N, \dots, U_{(n-1)\Delta_n}^N)$  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  $\ell^2$  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  $\sigma^2$  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  $\sigma^2$  on a compact interval  $I \subset \mathbb{R}$ , say for example  $I = [-1, 1]$ . Secondly, we extend the study to the estimation of  $\sigma^2$  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

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

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

$$\tau_0 \|h\|^2 \leq \|h\|_n^2 \leq \tau_1 \|h\|^2 \quad (9)$$

where  $\|\cdot\|$  is the  $\mathbb{L}^2$ -norm. Equation (9) allows establishing 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  $\mathbb{L}^2$ -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  $\|\cdot\|_X$  (see Equation (2)) on one side, and  $\|\cdot\|_X$  and the  $\mathbb{L}^2$ -norm  $\|\cdot\|$  on the other side. Define for  $x \in \mathbb{R}$ , the local time  $\mathcal{L}^x$  of the diffusion process  $X = (X_t)_{t \in [0,1]}$  by

$$\mathcal{L}^x = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^1 \mathbf{1}_{(x-\varepsilon, x+\varepsilon)}(X_s) ds. \quad (10)$$

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 continuous 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 holds,*

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 show that there is a link between the local time and the transition density of the diffusion process. The proof of the Lemma is provided in appendix. Thus, if we consider the pseudo-norm  $\|\cdot\|_X$  depending on the process  $X = (X_t)_{t \in [0,1]}$  and given in Equation (2), and using Lemma 3.1, we obtain that,

$$\mathbb{E} [\|h\|_X^2] = \int_{\mathbb{R}} h^2(x)\mathbb{E}[\mathcal{L}^x] dx = \int_{\mathbb{R}} h^2(x) \int_0^1 p_X(s, x)ds dx \geq \tau_0 \|h\|^2. \quad (11)$$

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  $L = \log(n)$  and  $n \rightarrow \infty$ . Suppose that  $\sigma^2$  is approximated in one of the collections [CS – OB] and [B]. Under Assumption 2.1, it holds*

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

where the numerical constant  $\gamma$  can take any value of the interval  $(1, +\infty)$ , with  $\alpha = 2$  and  $\alpha' = 1$  for the collection [B],  $\alpha = (r + r')/2$  and  $\alpha' = r$  for the collection [CS – OB]. The constant  $C > 0$  depends on  $\sigma_1$  and the constant  $C' > 0$  depends on  $\sigma_1, \tau_0$  and  $\tau_1$ .

We observe that the upper-bound of the risk of estimation of  $\widehat{\sigma}_m^2$  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  $O(m^{\alpha'}/n)$  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  $\left\| \widehat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{\infty}^2$  is bounded by the term  $O\left(\frac{m^{\alpha\gamma+1} \log(n)}{n^{\gamma/2}}\right)$  (see Lemma 8.6 and proof of Theorem 3.2).

Note that for the two collections [B] and [F], we have  $\alpha = 2$ , and  $\alpha' = 1$ . The next result proves that for the collections [B] and [F], the risk of estimation can reach a rate of convergence of the same order as 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 negligible with respect to the estimation error of order  $m/n$ . Let us also note that the risk  $\left\| \widehat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_n^2$  is random since

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

and the estimator  $\widehat{\sigma}_m^2$  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  $\widehat{\sigma}_m^2$ .

**Corollary 3.3.** *Suppose that  $\sigma^2 \in \Sigma_I(\beta, R)$  with  $\beta > 3/2$ , and  $\gamma = 4(2\beta + 1)/(2\beta - 3)$ . Assume that  $K_{\text{opt}} \propto n^{1/(2\beta+1)}$  for  $[\mathbf{B}]$  ( $m_{\text{opt}} = K_{\text{opt}} + M$ ),  $m_{\text{opt}} \propto n^{1/(2\beta+1)}$  for  $[\mathbf{F}]$ , and  $n \rightarrow \infty$ . Under Assumptions 2.1, it holds,*

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

Note that we obtain the exact same rates when considering the risk of  $\widehat{\sigma}_{m_{\text{opt}}}^2$  defined with the  $\mathbb{L}^2$ -norm equivalent to the empirical norm  $\|\cdot\|_n$ . Moreover, these rates of convergence are of the same order as 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 non-compact interval

In this section, we propose a ridge estimator of  $\sigma^2$  on a non-compact interval  $I = \mathcal{I}$ , 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 assumptions made in the previous sections. Consider the truncated estimator  $\widehat{\sigma}_{m,L}^2$  of  $\sigma^2$  given by

$$\widehat{\sigma}_{m,L}^2(x) = \widehat{\sigma}_m^2(x) \mathbf{1}_{\widehat{\sigma}_m^2(x) \leq \sqrt{L}} + \sqrt{L} \mathbf{1}_{\widehat{\sigma}_m^2(x) > \sqrt{L}}. \quad (12)$$

For the specific case of approximation spaces spanned by bases of compactly supported functions, we proceed to the dilation of the bases and consider the following decomposition of the risk of the ridge estimator  $\widehat{\sigma}_{m,L}^2$ :

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

The first term on the *r.h.s.* is equivalent to the risk of a ridge estimator of  $\sigma^2$  on the compact interval  $[-\log(n), \log(n)]$ . The second term on the *r.h.s.* is upper-bounded using Lemma 8.4. The estimation interval  $\mathcal{I}$  is set to  $\mathcal{I} = \mathbb{R}^+$  for the collection  $[\mathbf{L}]$ , and  $\mathcal{I} = \mathbb{R}$  for the collection  $[\mathbf{H}]$  and any chosen basis of compactly supported functions like those related to the collections  $[\mathbf{B}]$  and  $[\mathbf{F}]$ , and in this case,  $\sigma_{|\mathcal{I}}^2 = \sigma^2$ . We derive below, an upper-bound of the risk of estimation of  $\widehat{\sigma}_m^2$ .

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

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

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

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

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

Secondly, the estimation error is not as fine as the one established in Theorem 3.2 where  $\sigma^2$  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  $\mathbb{L}^2$ -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  $\mathbb{L}^2$ -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  $\sigma^2$  for each of the collections **[B]** and **[F]**.

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

**For [B].** *Assume that  $K \propto n^{1/(4\beta+1)}$  and  $n \rightarrow \infty$ . Under Assumptions 2.1, there exists a constant  $C > 0$  depending on  $\beta$  and  $\sigma_1$  such that*

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

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

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

where the constant  $C > 0$  depends on  $\beta$  and  $\sigma_1$ .

As we can remark, the obtained rates are slower than the ones established in Section 3.1 where  $\sigma^2$  is estimated on a compact interval. This result is the immediate consequence of the result of Theorem 3.4. At this stage, we cannot position ourselves on the optimality or not of the obtained rates of convergence, as a lower-bound on the risk of estimation is not provided in this paper. The establishment of a lower-bound could be obtained by adapting the method provided in Hoffmann (1999b) to the case of a non-compactly supported diffusion coefficient. This study is beyond the scope of this paper, and will be subject to further investigations.

## 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  $L = \log(Nn)$ ,  $n, N \rightarrow \infty$  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  $\sigma_1$  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 generates 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 third term on the right-hand side is negligible with respect to the estimation error, and is derived from the result of Lemma 8.8. Note that Theorem 4.1 does not extend Theorem 3.2 since Theorem 4.1, contrary to Theorem 3.2, holds for any compactly supported basis, takes advantage of the independence of the sample paths  $\bar{X}^j$ ,  $j = 1, \dots, N$ , and requires the condition  $m \leq \sqrt{\min(n, N)}/\log(Nn)$  on the dimensions  $m$  of the approximation spaces, which reduces the set of dimension  $\mathcal{M}$  to the empty set  $\emptyset$  for  $N = 1$ .

We derive below and from Theorem 4.1, some rates of convergence related to the collections  $[\mathbf{B}]$  and  $[\mathbf{F}]$ .

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

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

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  $[\mathbf{B}]$  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

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

Under the condition  $n \propto \sqrt{N}/\log^4(N)$ , 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).

It remains to prove the optimality of the obtained rate of convergence, especially when  $N \propto n$ . Unfortunately, the method proposed in Tsybakov (2008) does not work for the establishment of a lower-bound of the risk of estimation of the diffusion coefficient. In fact, the third condition of Theorem 2.5 in Tsybakov (2008) states that there exists a numerical constant  $\alpha \in (0, 1/8)$  such that

$$\frac{1}{M} \sum_{j=1}^M K(P_0^{\otimes N}, P_j^{\otimes N}) \leq \alpha \log(M),$$

where  $M$  is the number of hypotheses satisfying  $M \geq 2^{m/8}$  with, in our framework,  $m$  defined such that  $m^{2\beta} \propto (Nn)^{2\beta/(2\beta+1)}$ , and  $\mathcal{K}(P_j^{\otimes N}, P_0^{\otimes N})$  is the Kullback divergence between the two joint probability distributions  $P_j^{\otimes N}$  and  $P_0^{\otimes N}$  characterized by their respective diffusion coefficients  $\sigma^{(j)}$  and  $\sigma^{(0)}$ . In our framework, the computation of the Kullback divergence is based on the ratio of their respective transition densities, and one obtains:

$$\frac{1}{M} \sum_{j=1}^M K(P_0^{\otimes N}, P_j^{\otimes N}) = O(Nn).$$

Consequently, since  $m$  is negligible with respect to  $Nn$ , it is not possible to satisfy this third condition using the inequality  $M \geq 2^{m/8}$ . A way to solve this problem may be to reconsider the construction of the set  $\{\sigma^{(0)}, \dots, \sigma^{(M)}\}$  of hypotheses, and use the technique developed in Hoffmann (1999b). This study is beyond the scope of this paper.

## 4.2 Non-adaptive estimation of the diffusion coefficient on a non-compact interval

Consider a ridge estimator of  $\sigma^2$  on a non-compact interval  $\mathcal{I}$  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  $\sigma^2$  and  $\hat{\sigma}_{m,L}^2$  the truncated estimators of  $\sigma^2$  given in Equation (12). We establish, through the following theorem, the first risk bound that highlights the main error terms.

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

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

where  $C > 0$  is a constant depending on the upper bound  $\sigma_1$  of the diffusion coefficient. Moreover,  $q = 1$  for the collection **[B]** and  $q = 2$  for the collection **[OB]**.

The result of Theorem 4.3 is clearly an extension of Theorem 3.4, it suffices to set  $N = 1$  in Theorem 4.3 to obtain Theorem 3.4 with the same hypotheses. Moreover, the respective proofs of the two theorems are similar, the only difference being the number of sample paths.

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

$$\mathbb{E} \left[ \left\| \hat{\sigma}_{m,L}^2 - \sigma_{|\mathcal{I}}^2 \right\|_n^2 \right] \leq 2 \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma_{|\mathcal{I}}^2\|_n^2 + C \left( \sqrt{\frac{m^q \log^2(Nn)}{Nn}} + \frac{m^2 \log^3(Nn)}{N} + \Delta_n^2 \right) \quad (13)$$

The risk bound given in Equation (13) 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. In this case, Theorem 4.3 can no longer be an extension of Theorem 3.4 since the establishment of Equation (13), which is derived from Lemma 8.7, requires a large number  $N > 1$  of independent sample paths (see Denis *et al.* (2021), *proof of Theorem 3.3*).

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 **[B]** and **[H]** respectively.

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

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

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

where  $C > 0$  is a constant depending on  $\beta$  and  $\sigma_1$  for **[B]**, or  $s$  and  $\sigma_1$  for **[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 (13) with  $n \propto N$  and assuming that  $m \propto N^{1/4(s+1)}$  for **[H]** or  $K \propto N^{1/4(\beta+1)}$  for **[B]**, we obtain

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

where  $C > 0$  is a constant depending on  $\sigma_1$  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  $N \rightarrow \infty$ .

Finally, the study of an optimal rate of convergence, which is beyond the scope of this paper, has to take into account the new problem related to the non-compact support of the square of the diffusion coefficient.

## 5 Adaptive estimation of the diffusion coefficient from repeated observations

In this section, we propose an adaptive ridge estimator of  $\sigma^2$  by selecting an optimal dimension from the sample  $D_{N,n}$ . In fact, consider the estimator  $\hat{\sigma}_{\hat{K}}^2$  where  $\hat{K}$  satisfies:

$$\hat{K} := \arg \min_{K \in \mathcal{K}} \{ \gamma_{n,N}(\hat{\sigma}_K^2) + \text{pen}(K) \} \quad (14)$$

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 number  $N$  of sample paths tends to infinity.

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

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

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

$$\mathbb{E} \left[ \left\| \hat{\sigma}_{\hat{K}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \right] \leq 34 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma_{|I}^2\|_n^2 + \text{pen}_1(K) \right\} + \frac{C}{Nn}$$

*where the penalty function  $\text{pen}_1$  is given by*

$$\text{pen}_1(K) := \kappa \frac{(K+M) \log(N)}{Nn} \quad (15)$$

*with  $\kappa > 0$  a numerical constant.*

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

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

This result is justified since the penalty term is of the same order (up to a log-factor) as the estimation error established in Theorem 4.1. If we consider the estimation interval  $[-A_N, A_N]$  where  $A_N \propto \sqrt{\log(N)}$ , then we obtain the following result:

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

where the penalty function  $K \mapsto \text{pen}(K)$  is given by

$$\text{pen}(K) := \kappa \frac{(K + M) \log^2(N)}{Nn}. \quad (16)$$

Considering the adaptive estimator of  $\sigma^2$  on the real line  $I = \mathbb{R}$  when  $N \geq 1$  and  $n \rightarrow \infty$ , we obtain the following result.

**Theorem 5.2.** *Suppose that  $N \geq 1$ ,  $n \rightarrow \infty$  and  $L = \log(Nn)$ , and consider the collection  $[\mathbf{B}]$  with*

$$K \in \mathcal{K} = \{2^q, q = 0, 1, \dots, q_{\max}\} \subset \mathcal{M} = \{1, \dots, Nn\}.$$

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

$$\mathbb{E} \left[ \left\| \hat{\sigma}_{\hat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 \right] \leq 3 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2\|_n^2 + \text{pen}_2(K) \right\} + \frac{C}{\sqrt{Nn}}$$

*where the penalty function  $\text{pen}_2$  is given by*

$$\text{pen}_2(K) := \kappa \sqrt{\frac{(K + M) \log^2(Nn)}{Nn}} \quad (17)$$

*with  $\kappa > 0$  a numerical constant.*

We have a penalty term of the same order as the one obtained in Theorem 5.1 where  $\sigma^2$  is estimated on a compact interval. One can deduce that the adaptive estimator reaches a rate of the same order as the rate of the non-adaptive estimator given in Corollary 4.4 for the collection  $[\mathbf{B}]$ . For the numerical study of the nonparametric estimation of the square of the diffusion coefficient, we rather consider the compact interval  $[-\log(Nn), \log(Nn)]$  on which is built the spline basis, and which tends to the real line as  $N, n \rightarrow \infty$  or  $N = 1$  and  $n \rightarrow \infty$ . Then, we use the penalty function  $K \mapsto \text{pen}(K)$  given in Equation (16).

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

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

$$K \in \mathcal{K} = \{2^q, q = 0, \dots, q_{\max}\} \subset \mathcal{M} = \{1, \dots, \lfloor \sqrt[5]{n} \rfloor\}.$$

*Under Assumption 2.1, it holds*

$$\mathbb{E} \left[ \left\| \hat{\sigma}_{\hat{K}}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] \leq 3 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma_{|I}^2\|_n^2 + \text{pen}_3(K) \right\} + \frac{C}{n}.$$

*where  $C > 0$  is a constant depending on  $\tau_0$ , and*

$$\text{pen}_3(K) = \kappa \frac{(K + M) \log(n)}{n}$$

*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  $\hat{\sigma}_{\hat{K}}^2$  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. Since the condition imposed on the dimension of the approximation subspace strongly restricts the model selection, we extend the set of possible values of the dimension to  $\mathcal{M} = \{1, \dots, \sqrt{n}/\log(n)\}$  for the numerical study of the estimation of the square  $\sigma^2$  of the diffusion coefficient.

## 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 **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  $\sigma^2$  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 number  $N$  of sample paths takes values in the set  $\{1, 10, 100, 1000\}$  where the number  $n$  of observations of each diffusion path 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 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  $\hat{\sigma}_{\hat{m}}^2$  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:

$$m^* := \arg \min_{m \in \mathcal{M}} \|\hat{\sigma}_m^2 - \sigma^2\|_{n,N}^2. \quad (18)$$

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  $\sigma^2$  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 observations  $D_{N,n}$  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 (7) and (8).
3. Select the optimal dimension  $\hat{K} \in \mathcal{K}$  using Equation (14) and compute  $K^*$  from Equation (18)
4. Using  $D_{N',n}$ , evaluate  $\left\| \hat{\sigma}_{\hat{K}}^2 - \sigma^2 \right\|_{n,N'}^2$  and  $\left\| \hat{\sigma}_{K^*}^2 - \sigma^2 \right\|_{n,N'}^2$ .

We deduce the risks of estimation considering the average values of  $\left\| \hat{\sigma}_{\hat{m}}^2 - \sigma^2 \right\|_{n,N'}^2$  and  $\left\| \hat{\sigma}_{m^*}^2 - \sigma^2 \right\|_{n,N'}^2$  over the 100 repetitions. Note that we consider in this section, the estimation of  $\sigma^2$  on the compact interval  $I = [-1, 1]$  and on the real line  $\mathbb{R}$ . The unknown parameter  $\kappa$  of the corresponding penalty functions given in Equations (15) and (17), are numerically calibrated (details are given in appendix). We choose  $\kappa = 4$  for the compact interval  $I = [-1, 1]$ , and  $\kappa = 5$  for the estimation of  $\sigma^2$  on the real line  $\mathbb{R}$ , more precisely, on the interval  $[-\log(N), \log(N)]$  which tends to  $\mathbb{R}$  as the number  $N$  of sample paths tends to infinity.

### 6.3 Numerical results

We present in this section the numerical results of the performance of the spline-based adaptive estimators of  $\sigma_I^2$  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  $\sigma_I^2$  from simulated data following the steps given in Section 6.2.

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  $n = 250$ .

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 the number  $Nn$  of observations is 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 number  $n$  of each diffusion path 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  $\sigma^2$  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  $\sigma^2$  is consistent as the sample size  $N$  and the length  $n$  paths take larger values.

**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

$$\hat{K} = \arg \min_{K \in \mathcal{K}} \gamma_n(\hat{\sigma}_K^2) + \text{pen}(K) \quad (19)$$

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.

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 (19). For the numerical assessment, we use the interval  $I = [-10^6, 10^6]$  to approximate the real line  $\mathbb{R}$ , and then, use Equation (19) 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  $\sigma^2$  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  $\sigma^2$  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  $\sigma^2$  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  $\sigma^2$  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  $\sigma^2$  and derive a rate of convergence of the same order than the optimal rate established in Hoffmann (1999b). We also propose an estimator of  $\sigma^2$  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  $\sigma^2$  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  $\sigma^2$  on a compact interval for  $N = 1$  and  $N \rightarrow \infty$ , and on the real line  $\mathbb{R}$  for  $N \rightarrow \infty$ .

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).

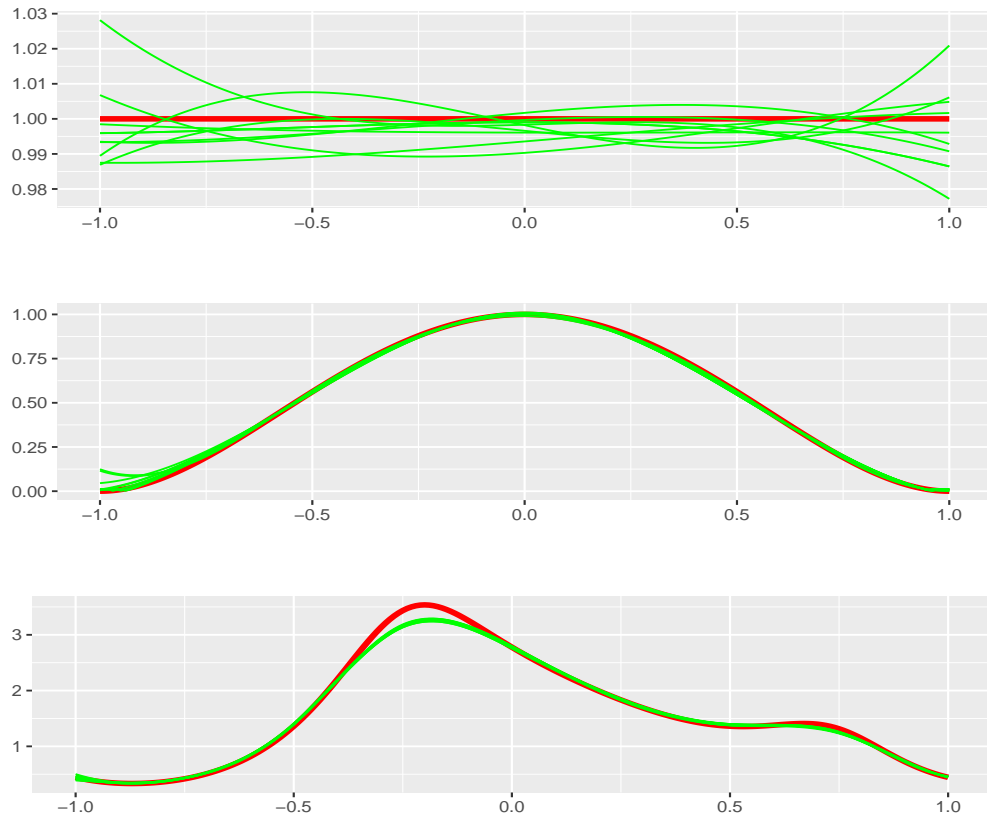


Figure 1: Bundles of 10 estimators  $\hat{\sigma}_{\hat{K},L}^2$  (in green) of the true diffusion coefficient  $\sigma|_I^2$  restricted on the compact interval  $I = [-1, 1]$  (in red) of each of Models 1, 2, 3 (from top to bottom) using samples of size  $N = 1000$  with diffusion paths of length 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



## 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  $C_\alpha$  in case we need to specify the dependency of the constant  $C$  on a parameter  $\alpha$ .

### 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 (10). The proof of Lemma 8.5 is provided in appendix.

## 8.2 Proofs of Section 3

### 8.2.1 Proof of Theorem 3.2

Let  $\Omega_{n,m}$  be a random event given by

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

On the event  $\Omega_{n,m}$ , the two pseudo-norms  $\|\cdot\|_{n,1}$  and  $\|\cdot\|_X$  are equivalent with  $(1/2)\|g\|_X^2 \leq \|g\|_{n,1}^2 \leq (3/2)\|g\|_X^2$  for each function  $g \in \mathcal{S}_m \setminus \{0\}$ . 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*

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

where  $C > 0$  is a constant depending on  $\gamma$ , with  $\alpha = 2$  for the collection  $[\mathbf{B}]$ , and  $\alpha = (r + r')/2$  for the collection  $[\mathbf{CS-OB}]$ .

The parameter  $\gamma > 1$  has to be chosen appropriately so that we obtain a variance term of the risk of the estimator  $\hat{\sigma}_m^2$  that is negligible with respect to  $m^{\alpha'}/n$  as  $n$  tends to infinity (see Theorem 3.2 and Corollary 3.3). The proof of Lemma 8.6 is provided in appendix.

**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:

$$\zeta_{k\Delta}^{1,1} = \frac{1}{\Delta} \left[ \left( \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s^1 \right)^2 - \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^1) ds \right], \quad (20)$$

$$\zeta_{k\Delta}^{1,2} = \frac{2}{\Delta} \int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s) \sigma'(X_s^1) \sigma^2(X_s^1) dW_s^1, \quad (21)$$

$$\zeta_{k\Delta}^{1,3} = 2b(X_{k\Delta}^1) \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s^1. \quad (22)$$

Besides,  $R_{k\Delta}^1 = R_{k\Delta}^{1,1} + R_{k\Delta}^{1,2} + R_{k\Delta}^{1,3}$ , with:

$$R_{k\Delta}^{1,1} = \frac{1}{\Delta} \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s^1) ds \right)^2, \quad R_{k\Delta}^{1,2} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s) \Phi(X_s^1) ds \quad (23)$$

$$R_{k\Delta}^{1,3} = \frac{2}{\Delta} \left( \int_{k\Delta}^{(k+1)\Delta} (b(X_s^1) - b(X_{k\Delta}^1)) ds \right) \left( \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^1) dW_s^1 \right) \quad (24)$$

where

$$\Phi := 2b\sigma'\sigma + [\sigma''\sigma + (\sigma')^2] \sigma^2. \quad (25)$$

By definition of the projection estimator  $\hat{\sigma}_m^2$  for each  $m \in \mathcal{M}$  (see Equation (7)), for all  $h \in \mathcal{S}_{m,L}$ , we have:

$$\gamma_{n,1}(\hat{\sigma}_m^2) - \gamma_{n,1}(\sigma_{[I]}^2) \leq \gamma_{n,1}(h) - \gamma_{n,1}(\sigma_{[I]}^2). \quad (26)$$

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,

$$\nu_i(h) = \frac{1}{n} \sum_{k=0}^{n-1} h(X_{k\Delta}^1) \zeta_{k\Delta}^{1,i}, \quad i \in \{1, 2, 3\}, \quad \mu(h) = \frac{1}{n} \sum_{k=0}^{n-1} h(X_{k\Delta}^1) R_{k\Delta}^1, \quad (27)$$

and  $\zeta_{k\Delta}^{1,1}$ ,  $\zeta_{k\Delta}^{1,2}$ ,  $\zeta_{k\Delta}^{1,3}$  are given in Equations (20), (21), (22), and finally,  $R_{k\Delta}^1 = R_{k\Delta}^{1,1} + R_{k\Delta}^{1,2}$  given in Equations (23) and (24). Then, for all  $m \in \mathcal{M}$ , and for all  $h \in \mathcal{S}_{m,L}$ , we obtain from Equation (26) that

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

Then, it comes,

$$\mathbb{E} \left[ \left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + 2\mathbb{E}[\nu(\hat{\sigma}_m^2 - h)] + 2\mathbb{E}[\mu(\hat{\sigma}_m^2 - h)]. \quad (28)$$

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

$$\begin{cases} 2\nu(\hat{\sigma}_m^2 - h) \leq \frac{2}{a} \left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_X^2 + \frac{2}{a} \left\| h - \sigma_{|I}^2 \right\|_X^2 + a \sup_{h \in \mathcal{S}_m, \|h\|_X=1} \nu^2(h), \\ 2\mu(\hat{\sigma}_m^2 - h) \leq \frac{2}{d} \left\| \hat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,1}^2 + \frac{2}{d} \left\| h - \sigma_{|I}^2 \right\|_{n,1}^2 + \frac{d}{n} \sum_{k=1}^n (R_{k\Delta}^1)^2. \end{cases} \quad (29)$$

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

We have:

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

For all  $k \in \llbracket 1, n \rrbracket$ , using the Cauchy-Schwarz inequality and Lemma (2.2),

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

Consider now the term  $R_{k\Delta}^{1,2}$ . From Equation (25), 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  $\sigma_1$  and  $\alpha$  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 Lemma (2.2) 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 Lemma (2.2) 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,

$$\mathbb{E} \left[ \frac{1}{n} \sum_{k=1}^n (R_{k\Delta}^1)^2 \right] \leq C\Delta^2. \quad (30)$$

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 (28), (29) and (30) that,

$$\mathbb{E} \left[ \left\| \widehat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \mathbf{1}_{\Omega_{n,m}} \right] \leq 3 \inf_{h \in \mathcal{S}_m} \left\| h - \sigma_{|I}^2 \right\|_n^2 + C \mathbb{E} \left( \sup_{h \in \mathcal{S}_m, \|h\|_X=1} \nu^2(h) \right) + C\Delta^2 \quad (31)$$

where  $C > 0$  is a constant depending on  $\sigma_1$ .

**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 (11)) and the coordinate vector  $\mathbf{a} = (a_{-M}, \dots, a_{K-1})$  satisfies:

- $\|\mathbf{a}\|_2^2 \leq Cm$  ( $m = K + M$ ) for the collection **[B]** (see Denis *et al.* (2021), Lemma 2.6)
- $\|\mathbf{a}\|_2^2 \leq 1/\tau_0$  for the collection **[CS-OB]** since  $\|h\|^2 = \|\mathbf{a}\|_2^2$ .

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

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

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  $k \in \llbracket 0, n-1 \rrbracket$ ,

$$\mathbb{E} \left[ \zeta_{k\Delta}^{1,1} | \mathcal{F}_{k\Delta} \right] = 0, \quad \mathbb{E} \left[ \left( \zeta_{k\Delta}^{1,1} \right)^2 | \mathcal{F}_{k\Delta} \right] \leq \frac{C}{\Delta^2} \mathbb{E} \left[ \left( \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_u^1) du \right)^2 \right] \leq C\sigma_1^4. \quad (32)$$

Then, using Equation (32) 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:

$$\begin{cases} \sum_{\ell=-M}^{K-1} B_\ell^2(X_{\eta(s)}^1) \leq 1 & \text{for the Collection } [\mathbf{B}] \quad (m = K + M), \\ \sum_{\ell=0}^{m-1} \phi_\ell^2(X_{\eta(s)}^1) \leq Cm^r & \text{for the collection } [\mathbf{CS} - \mathbf{OB}]. \end{cases}$$

Thus, it follows that

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

and,

$$\mathbb{E} \left( \sup_{h \in \mathcal{S}_m, \|h\|_X^2=1} \nu_1^2(h) \right) \leq C \frac{m^{\alpha'}}{n} \quad (33)$$

where  $C > 0$  is a constant depending on  $\sigma_1$  and the basis, with  $\alpha' = 1$  for the collection  $[\mathbf{B}]$ , and  $\alpha' = r$  for the collection  $[\mathbf{CS} - \mathbf{OB}]$ .

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  $\sigma_1$  such that:

$$\mathbb{E} \left( \sup_{h \in \mathcal{S}_m, \|h\|_X^2=1} \nu_2^2(h) \right) \leq C \frac{m^{\alpha'}}{n^2}. \quad (34)$$

### 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  $\sigma_1$  such that:

$$\mathbb{E} \left( \sup_{h \in \mathcal{S}_m, \|h\|_X^2=1} \nu_3^2(h) \right) \leq C \frac{m^{\alpha'}}{n^2}. \quad (35)$$

We finally obtain from Equations (33), (34) and (35) that there exists a constant  $C > 0$  depending on  $\sigma_1$  such that:

$$\mathbb{E} \left( \sup_{h \in \mathcal{S}_m, \|h\|_X^2=1} \nu^2(h) \right) \leq C \frac{m^{\alpha'}}{n}. \quad (36)$$

We deduce from Equations (31) and (36) that there exists a constant  $C > 0$  depending on  $\sigma_1$  such that,

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

For  $n$  large enough, we have  $\|\widehat{\sigma}_m^2 - \sigma_{|I}^2\|_\infty^2 \leq 2mL$  since  $\|\widehat{\sigma}_m^2\|_\infty \leq \sqrt{mL}$ . Then, from Lemma 8.6 and for all  $m \in \mathcal{M}$ , there exists a constant  $C > 0$  depending on  $\sigma_1$  such that

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

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  $\sigma_1$  such that

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

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 (11) and (9), 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  $\sigma_1, \tau_0$  and  $\tau_1$  such that

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

□

## 8.2.2 Proof of Theorem 3.4

*Proof.* Since  $L = \log^2(n)$ . From Equation (28) (Proof of Theorem 3.2), for all  $h \in \mathcal{S}_{m,L}$ ,

$$\mathbb{E} \left[ \left\| (\hat{\sigma}_{m,L}^2 - \sigma_{|I}^2) \right\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + 2 \sum_{i=1}^3 \mathbb{E} [\nu_i (\hat{\sigma}_m^2 - h)] + 2\mathbb{E} [\mu (\hat{\sigma}_m^2 - h)] \quad (37)$$

where  $\nu_i$ ,  $i = 1, 2, 3$  and  $\mu$  are given in Equation (27). For all  $i \in \{1, 2, 3\}$  and for all  $h \in \mathcal{S}_{m,L}$ , one has

$$\mathbb{E} [\nu_i (\hat{\sigma}_{m,L}^2 - h)] \leq \sqrt{2m \log^2(n)} \sqrt{\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_i^2(\phi_\ell)]}. \quad (38)$$

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

According to Equation (27), 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 $\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_3^2(\phi_\ell)]$

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 (37) and (38) that for all  $h \in \mathcal{S}_{m,L}$ ,

$$\begin{cases} \mathbb{E} \left[ \|\widehat{\sigma}_{m,L}^2 - \sigma_{|\mathcal{I}}^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma_{|\mathcal{I}}^2\|_n^2 + C \sqrt{\frac{m \log^2(n)}{n}} + 2\mathbb{E} \left[ \mu(\widehat{\sigma}_{m,L}^2 - h) \right] & \mathbf{[B]} \\ \mathbb{E} \left[ \|\widehat{\sigma}_{m,L}^2 - \sigma_{|\mathcal{I}}^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma_{|\mathcal{I}}^2\|_n^2 + C \sqrt{\frac{m^2 \log^2(n)}{n}} + 2\mathbb{E} \left[ \mu(\widehat{\sigma}_{m,L}^2 - h) \right] & \mathbf{[OB]} \end{cases} \quad (39)$$

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

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

Using Equations (30), (39) and setting  $a = 4$ , we deduce that there exists constant  $C > 0$  depending on  $\sigma_1$  such that,

$$\begin{cases} \mathbb{E} \left[ \|\widehat{\sigma}_{m,L}^2 - \sigma_{|\mathcal{I}}^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma_{|\mathcal{I}}^2\|_n^2 + C \sqrt{\frac{m \log^2(n)}{n}} & \mathbf{[B]} \\ \mathbb{E} \left[ \|\widehat{\sigma}_{m,L}^2 - \sigma_{|\mathcal{I}}^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma_{|\mathcal{I}}^2\|_n^2 + C \sqrt{\frac{m^2 \log^2(n)}{n}} & \mathbf{[OB]}. \end{cases} \quad (40)$$

For the case of bases of compactly supported functions, we proceed to the dilation of the bases so that  $\sigma^2$  can be estimated on the compact interval  $[-\log(n), \log(n)]$ . Then, we use the following decomposition:

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

and we deduce from Equation (40) that

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

and

$$\left\{ \begin{array}{l} \mathbb{E} \left[ \|\hat{\sigma}_{m,L}^2 - \sigma_{|\mathcal{I}}^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma_{|\mathcal{I}}^2\|_n^2 + C\sqrt{\frac{m \log^2(n)}{n}} + 2 \log^2(n) \sup_{t \in (0,1)} \mathbb{P}(|X_t| > A_n) \quad [\mathbf{B}] \\ \mathbb{E} \left[ \|\hat{\sigma}_{m,L}^2 - \sigma_{|\mathcal{I}}^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma_{|\mathcal{I}}^2\|_n^2 + C\sqrt{\frac{m^2 \log^2(n)}{n}} + 2 \log^2(n) \sup_{t \in (0,1)} \mathbb{P}(|X_t| > A_n) \quad [\mathbf{OB}]. \end{array} \right. \quad (41)$$

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 (41) that

$$\left\{ \begin{array}{l} \mathbb{E} \left[ \|\hat{\sigma}_{m,L}^2 - \sigma_{|\mathcal{I}}^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma_{|\mathcal{I}}^2\|_n^2 + C\sqrt{\frac{m \log^2(n)}{n}} \quad [\mathbf{B}] \\ \mathbb{E} \left[ \|\hat{\sigma}_{m,L}^2 - \sigma_{|\mathcal{I}}^2\|_{n,1}^2 \right] \leq \inf_{h \in \mathcal{S}_m} \|h - \sigma_{|\mathcal{I}}^2\|_n^2 + C\sqrt{\frac{m^2 \log^2(n)}{n}} \quad [\mathbf{OB}]. \end{array} \right.$$

□

### 8.3 Proof of Section 4

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

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

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

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  $m \in \mathcal{M}$ ,

$$\Omega_{n,N,m} := \bigcap_{h \in \mathcal{S}_m \setminus \{0\}} \left\{ \left| \frac{\|h\|_{n,N}^2}{\|h\|_n^2} - 1 \right| \leq \frac{1}{2} \right\}. \quad (42)$$

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  $m \in \mathcal{M} = \left\{ 1, \dots, \left\lfloor \sqrt{N} / \log(Nn) \right\rfloor \right\}$  and,

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

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

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

where  $C > 0$  is a constant.

The proof of Lemma 8.8 is provided in appendix.

**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 (31) that

$$\mathbb{E} \left[ \left\| \widehat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,m}} \right] \leq 3 \inf_{h \in \mathcal{S}_{m,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + C \mathbb{E} \left( \sup_{h \in \mathcal{S}_m, \|h\|_n=1} \nu^2(h) \right) + C\Delta^2 \quad (43)$$

where  $C > 0$  is a constant depending on  $\sigma_1$ , 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–Schwarz 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 (33), (34) and (35) that there exists a constant  $C > 0$  depending on  $\sigma_1$  such that:

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

We deduce from (43) and (44) that there exists a constant  $C > 0$  such that,

$$\mathbb{E} \left[ \left\| \widehat{\sigma}_m^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,m}} \right] \leq 3 \inf_{h \in \mathcal{S}_{m,L}} \left\| \sigma_{|I}^2 - h \right\|_n^2 + C \left( \frac{m}{Nn} + \Delta^2 \right). \quad (45)$$

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 (45), Lemma 8.8 and for  $m \in \mathcal{M} = \{1, \dots, \sqrt{\min(n, N)}/\sqrt{\log(Nn)}\}$ , 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.* For each dimension  $m \in \mathcal{M}$ , 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

$$\mathbb{E} \left[ \left\| \hat{\sigma}_{m,L}^2 - \sigma^2 \right\|_n^2 \right] = \mathbb{E} \left[ \left\| (\hat{\sigma}_{m,L}^2 - \sigma^2) \mathbf{1}_{[-\log(N), \log(N)]} \right\|_n^2 \right] + \frac{C}{\log(N)} \exp(-c \log^2(N)). \quad (46)$$

It remains to upper-bound the first term on the right hand side of Equation (46).

**Upper bound of**  $\mathbb{E} \left[ \left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_n^2 \mathbf{1}_{[-\log(N), \log(N)]} \right]$ . For all  $h \in \mathcal{S}_{m,L}$ , we obtain from Equation (7),

$$\gamma_{n,N}(\widehat{\sigma}_{m,L}^2) - \gamma_{n,N}(\sigma^2) \leq \gamma_{n,N}(h) - \gamma_{n,N}(\sigma^2). \quad (47)$$

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

$$\nu_i(h) = \frac{1}{nN} \sum_{j=1}^N \sum_{k=0}^{n-1} h(X_{k\Delta}^j) \zeta_{k\Delta}^{j,i}, \quad i \in \{1, 2, 3\}, \quad \mu(h) = \frac{1}{nN} \sum_{j=1}^N \sum_{k=0}^{n-1} h(X_{k\Delta}^j) R_{k\Delta}^j, \quad (48)$$

we deduce from Equation (47) that for all  $h \in \mathcal{S}_{m,L}$ ,

$$\mathbb{E} \left[ \left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,N}^2 \mathbf{1}_{[-\log(N), \log(N)]} \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + 2 \sum_{i=1}^3 \mathbb{E} [\nu_i(\widehat{\sigma}_{m,L}^2 - h)] + 2\mathbb{E} [\mu(\widehat{\sigma}_{m,L}^2 - h)]. \quad (49)$$

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

$$\mathbb{E} [\nu_i(\widehat{\sigma}_{m,L}^2 - h)] \leq \sqrt{2mL} \sqrt{\sum_{\ell=0}^{m-1} \mathbb{E} [\nu_i^2(\phi_\ell)]}. \quad (50)$$

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

According to Equation (48), 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:

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

Since for all  $x \in \mathbb{R}$ ,  $b(x) \leq C_0(1+x^2)$  and  $\sup_{t \in [0,1]} \mathbb{E} (|X_t|^2) < \infty$ , 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 (49) and (50) that for all  $h \in \mathcal{S}_{m,L}$ ,

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

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

(51)

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

$$2\mu(\widehat{\sigma}_{m,L}^2 - h) \leq \frac{2}{a} \|\widehat{\sigma}_{m,L}^2 - \sigma^2\|_{n,N}^2 + \frac{2}{a} \|h - \sigma^2\|_{n,N}^2 + \frac{a}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} (R_{k\Delta}^j)^2$$

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

Using Equations (30), (51) and setting  $a = 4$ , we deduce that there exists a constant  $C > 0$  depending on  $\sigma_1$  such that,

$$\begin{cases} \mathbb{E} \left[ \left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,N}^2 \mathbb{1}_{[-\log(N), \log(N)]} \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + C \left( \sqrt{\frac{mL}{Nn}} + \Delta^2 \right) & \text{[B]} \\ \mathbb{E} \left[ \left\| \widehat{\sigma}_{m,L}^2 - \sigma^2 \right\|_{n,N}^2 \mathbb{1}_{[-\log(N), \log(N)]} \right] \leq \inf_{h \in \mathcal{S}_{m,L}} \|h - \sigma^2\|_n^2 + C \left( \sqrt{\frac{m^2L}{Nn}} + \Delta^2 \right) & \text{[H]}. \end{cases} \quad (52)$$

The final result is obtained from Equations (46) and (52).  $\square$

## 8.4 Proof of Section 5

### 8.4.1 Proof of Theorem 5.1

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

$$\mathcal{T}_{K,K'} = \left\{ g \in \mathcal{S}_{K+M} + \mathcal{S}_{K'+M}, \|g\|_n = 1, \|g\|_\infty \leq \sqrt{L} \right\}. \quad (53)$$

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 8.9.** *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 (14), we have

$$\widehat{K} := \arg \min_{K \in \mathcal{K}} \{ \gamma_{n,N}(\widehat{\sigma}_K^2) + \text{pen}_1(K) \}.$$

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

$$\gamma_{n,N}(\widehat{\sigma}_{\widehat{K}}^2) + \text{pen}_1(\widehat{K}) \leq \gamma_{n,N}(h) + \text{pen}_1(K),$$

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

$$\begin{aligned} \gamma_{n,N}(\widehat{\sigma}_{\widehat{K}}^2) - \gamma_{n,N}(\sigma_{|I}^2) &\leq \gamma_{n,N}(h) - \gamma_{n,N}(\sigma_{|I}^2) + \text{pen}_1(K) - \text{pen}_1(\widehat{K}) \\ \left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 &\leq \left\| h - \sigma_{|I}^2 \right\|_{n,N}^2 + 2\nu \left( \widehat{\sigma}_{\widehat{K}}^2 - h \right) + 2\mu \left( \widehat{\sigma}_{\widehat{K}}^2 - h \right) + \text{pen}_1(K) - \text{pen}_1(\widehat{K}) \\ &\leq \left\| h - \sigma_{|I}^2 \right\|_{n,N}^2 + \frac{1}{d} \left\| \widehat{\sigma}_{\widehat{K}}^2 - t \right\|_n^2 + d \sup_{g \in \mathcal{T}_{K,\widehat{K}}} \nu^2(g) + \frac{1}{d} \left\| \widehat{\sigma}_{\widehat{K}}^2 - h \right\|_{n,N}^2 \\ &\quad + \frac{d}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} (R_{k\Delta}^j)^2 + \text{pen}_1(K) - \text{pen}_1(\widehat{K}) \end{aligned}$$

where  $d > 1$  and the space  $\mathcal{T}_{K,\widehat{K}}$  is given in Equation (53). On the set  $\Omega_{n,N,K_{\max}}$  (given in Equation (42)):  $\forall h \in \mathcal{S}_{K+M}$ ,  $\frac{1}{2} \|h\|_n^2 \leq \|h\|_{n,N}^2 \leq \frac{3}{2} \|h\|_n^2$ . Then on  $\Omega_{n,N,K_{\max}}$ , for all  $d > 1$  and for all  $h \in \mathcal{S}_{K+M}$  with  $K \in \mathcal{K}$ ,

$$\begin{aligned} \left(1 - \frac{10}{d}\right) \left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 &\leq \left(1 + \frac{10}{d}\right) \left\| h - \sigma_{|I}^2 \right\|_{n,N}^2 + d \sup_{h \in \mathcal{T}_{K,\widehat{K}}} \nu^2(h) + \frac{d}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} (R_{k\Delta}^j)^2 \\ &\quad + \text{pen}_1(K) - \text{pen}_1(\widehat{K}). \end{aligned}$$

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

$$\left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \leq 3 \left\| h - \sigma_{|I}^2 \right\|_{n,N}^2 + 20 \sup_{h \in \mathcal{T}_{K,\widehat{K}}} \nu^2(h) + \frac{20}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} (R_{k\Delta}^j)^2 + 2 \left( \text{pen}_1(K) - \text{pen}_1(\widehat{K}) \right). \quad (54)$$

Let  $q_1 : \mathcal{K}^2 \rightarrow \mathbb{R}_+$  such that  $160q_1(K, K') \leq 18\text{pen}_1(K) + 16\text{pen}_1(K')$ . Thus, on the set  $\Omega_{n,N,K_{\max}}$ , there exists a constant  $C > 0$  such that for all  $h \in \mathcal{S}_{K+M}$

$$\begin{aligned} \mathbb{E} \left[ \left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right] &\leq 34 \left( \inf_{h \in \mathcal{S}_{K+M,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + \text{pen}_1(K) \right) \\ &\quad + 160 \mathbb{E} \left( \sup_{h \in \mathcal{T}_{K,\widehat{K}}} \nu_1^2(h) - q_1(K, \widehat{K}) \right) + C\Delta^2 \end{aligned}$$

where

$$\nu_1(h) := \frac{1}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} h(X_{k\Delta}^j) \zeta_{k\Delta}^{j,1} \quad (55)$$

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

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



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,

$$\begin{aligned} \mathbb{E} \left[ \left\|\widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2\right\|_{n,N}^2 \right] &\leq \mathbb{E} \left[ \left\|\widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2\right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right] + \mathbb{E} \left[ \left\|\widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2\right\|_{n,N}^2 \mathbf{1}_{\Omega_{n,N,K_{\max}}^c} \right] \\ &\leq 34 \inf_{K \in \mathcal{K}} \left( \inf_{h \in \mathcal{S}_{K+M}} \left\|h - \sigma_{|I}^2\right\|_n^2 + \text{pen}_1(K) \right) \\ &\quad + C\Delta^2 + 4(K+M)L\mathbb{P}(\Omega_{n,N,K_{\max}}^c) \\ &\quad + 160\mathbb{E} \left[ \left( G_K(\widehat{K}) - q_1(K, \widehat{K}) \right)_+ \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right]. \end{aligned}$$

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:

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

admits an  $\varepsilon$ -net  $E_\varepsilon$  such that for each  $\varepsilon \in (0,1]$ ,  $|E_\varepsilon| \leq \left(\frac{3}{\varepsilon\tau_0}\right)^{K+M}$  (see Lorentz *et al.* (1996), Proposition 1.3, page 487).

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  $\varepsilon$ -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

$$\tau := \sigma_1^2 \sqrt{6x_0^{n,N}} + \sigma_1^2 \sqrt{Lx_0^{n,N}} + \sum_{k \geq 1} \varepsilon_{k-1} \left\{ \sigma_1^2 \sqrt{6x_k^{n,N}} + 2\sigma_1^2 \sqrt{Lx_k^{n,N}} \right\} = y_0^{n,N} + \sum_{k \geq 0} y_k^{n,N}.$$

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 (55) and Lemma 8.9, there exists a constant  $C > 0$  such that

$$\begin{aligned} \tilde{\mathbb{P}} \left( \nu_1(h_0) > y_0^{n,N} \right) &\leq \tilde{\mathbb{P}} \left( \nu_1(h_0) > \sigma_1 \sqrt{6x_0^{n,N}} + \sigma_1^2 \|h_0\|_\infty x_0^{n,N} \right) \\ &\leq \exp(-CNnx_0^{n,N}), \\ \forall k \geq 1, \tilde{\mathbb{P}} \left( \nu_1(h_k - h_{k-1}) > y_0^{n,N} \right) &\leq \tilde{\mathbb{P}} \left( \nu_1(h_k - h_{k-1}) > \sigma_1 \sqrt{6x_k^{n,N}} + \sigma_1^2 \|h_k - h_{k-1}\|_\infty x_k^{n,N} \right) \\ &\leq \exp(-CNnx_k^{n,N}). \end{aligned}$$

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

$$x_k^{n,N} \leq C_0(1+k) \frac{K + K' + 2M}{Nn} \quad \text{and} \quad \tau \leq C_1 \sigma_1^2 \sqrt{\sqrt{L} \frac{K + K' + 2M}{Nn}} \quad (56)$$

with  $C_0 > 0$  and  $C_1$  two constants depending on  $a$  and  $b$ . It follows 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 (56), we set

$$q_1(K, K') = \kappa^* \sigma_1^2 \sqrt{L} \frac{K + K' + 2M}{Nn}$$

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

$$\mathbb{P} \left( \left\{ \sup_{h \in \mathcal{T}_{K,K'}} \nu^2(h) > q_1(K, K') \right\} \cap \Omega_{n,N,K_{\max}} \right) \leq \frac{e^{-b+1}}{e+1} \exp \{ -a(K + K' + 2M) \}$$

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

$$\begin{aligned} \mathbb{E} \left[ (G_K(K') - q_1(K, K'))_+ \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right] \\ \leq \frac{c(K + K')}{Nn} \mathbb{P} \left( \left\{ \sup_{t \in \mathcal{T}_{K,K'}} \nu^2(t) > q_1(K, K') \right\} \cap \Omega_{n,N,K_{\max}} \right) \\ \leq \frac{C}{Nn} \exp \left\{ -\frac{a}{2}(K + K') \right\}. \end{aligned}$$

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

$$\mathbb{E} \left[ (G_K(\hat{K}) - q_1(K, \hat{K}))_+ \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right] \leq \sum_{K' \in \mathcal{K}} \mathbb{E} \left[ (G_K(K') - q_1(K, K'))_+ \mathbf{1}_{\Omega_{n,N,K_{\max}}} \right] \leq \frac{C}{Nn}.$$

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

$$\text{pen}_1(K) \geq \kappa \sigma_1^2 \sqrt{L} \frac{K + M}{Nn}.$$

For  $N$  large enough, one has  $\sigma_1^2 \leq \sqrt{L}$ . Thus, we finally set  $\text{pen}_1(K) = \kappa \frac{(K+M) \log(N)}{Nn}$  with  $L = \log(N)$ . Then, there exists a constant  $C > 0$  such that,

$$\mathbb{E} \left[ \left\| \hat{\sigma}_{\hat{K}}^2 - \sigma_{|I}^2 \right\|_{n,N}^2 \right] \leq 34 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + \text{pen}_1(K) \right\} + \frac{C}{Nn}.$$

□

### 8.4.2 Proof of Theorem 5.2

*Proof.* From Equation (14), we have

$$\widehat{K} := \arg \min_{K \in \mathcal{K}} \gamma_{n,N}(\widehat{\sigma}_{K+M,L}^2) + \text{pen}_2(K).$$

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

$$\gamma_{n,N}(\widehat{\sigma}_{\widehat{K},L}^2) + \text{pen}_2(\widehat{K}) \leq \gamma_{n,N}(h) + \text{pen}_2(K).$$

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

$$\begin{aligned} \gamma_{n,N}(\widehat{\sigma}_{\widehat{K},L}^2) - \gamma_{n,N}(\sigma^2) &\leq \gamma_{n,N}(h) - \gamma_{n,N}(\sigma^2) + \text{pen}_2(K) - \text{pen}_2(\widehat{K}) \\ \left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 &\leq \|h - \sigma^2\|_{n,N}^2 + 2\nu(\widehat{\sigma}_{\widehat{K},L}^2 - h) + 2\mu(\widehat{\sigma}_{\widehat{K},L}^2 - h) + \text{pen}_2(K) - \text{pen}_2(\widehat{K}). \end{aligned}$$

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,

$$\begin{aligned} \left( 1 - \frac{2}{a} \right) \mathbb{E} \left[ \left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 \right] &\leq \left( 1 + \frac{2}{a} \right) \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2\|_n^2 + 2c \mathbb{E} \left[ \nu_1(\widehat{\sigma}_{\widehat{K},L}^2 - h) \right] \\ &\quad + \text{pen}_2(K) - \text{pen}_2(\widehat{K}) + \frac{a}{Nn} \sum_{j=1}^N \sum_{k=0}^{n-1} \mathbb{E} \left[ \left( R_{k\Delta}^j \right)^2 \right] \end{aligned}$$

From Equation (30) and for  $a = 4$ , there exists a constant  $C > 0$  such that

$$\mathbb{E} \left[ \left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 \right] \leq 3 \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2\|_n^2 + 4c \mathbb{E} \left[ \nu_1(\widehat{\sigma}_{\widehat{K},L}^2 - h) \right] + 2 \left( \text{pen}_2(K) - \text{pen}_2(\widehat{K}) \right) + C\Delta^2. \quad (57)$$

Since for all  $K \in \mathcal{K}$ ,  $\text{pen}_2(K) \geq 2\kappa^* \sigma_1^2 \sqrt{(K+M)L/(Nn)}$ , define the function  $q_2 : (K, K') \mapsto q_2(K, K')$  such that

$$q_2(K, K') = 2C^* \sigma_1^2 \sqrt{\frac{(K+K'+2M)L}{Nn}} \geq 2\sigma_1^2 v \sqrt{x^{n,N}} + \sigma_1^2 v x^{n,N} \quad (58)$$

where

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

The constant  $C^* > 0$  depends on constants  $\kappa^* > 0$  and  $c > 0$  of Equation (57) such that

$$4cq_2(K, K') \leq \text{pen}_2(K) + 2\text{pen}_2(K').$$

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

$$\mathbb{E} \left[ \left\| \hat{\sigma}_{\hat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 \right] \leq 3 \left( \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2\|_n^2 + \text{pen}_2(K) \right) + 4c \mathbb{E} \left[ \left( \nu_1(\hat{\sigma}_{\hat{m},L}^2 - h) - q_2(K, \hat{K}) \right)_+ \right] + C\Delta^2. \quad (59)$$

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 (58) and Lemma 8.9, there exists a constant  $C > 0$  such that for all  $K, K' \in \mathcal{K}$  and for all  $h \in \mathcal{S}_{K+M,L}$ ,

$$\mathbb{P} \left( \nu_1(\hat{\sigma}_{K',L}^2 - h) \geq q(K, K'), \left\| \hat{\sigma}_{\hat{K},L}^2 - h \right\|_{n,N}^2 \leq v^2 \right) \leq \exp(-CNnx^{n,N}). \quad (60)$$

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

$$\text{pen}_2(K) = \kappa \sqrt{\frac{(K+M) \log^2(Nn)}{Nn}}$$

where  $\kappa > 0$  is a new constant. Since  $\mathbb{E} \left[ \nu_1(\hat{\sigma}_{\hat{K},L}^2 - h) \right] \leq \mathcal{O} \left( \sqrt{\frac{(K_{\max}+M) \log^2(N)}{Nn}} \right)$  (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

$$\begin{aligned} \mathbb{E} \left[ \left( \nu_1(\hat{\sigma}_{\hat{K},L}^2 - h) - q_2(K, \hat{K}) \right)_+ \right] &\leq \max_{K' \in \mathcal{K}} \left\{ \mathbb{E} \left[ \left( \nu_1(\hat{\sigma}_{K',L}^2 - h) - q_2(K, K') \right)_+ \right] \right\} \\ &\leq cq_2(K, K_{\max}) \max_{K' \in \mathcal{K}} \left\{ \mathbb{P} \left( \nu_1(\hat{\sigma}_{K',L}^2 - h) \geq q_2(K, K') \right) \right\}. \end{aligned}$$

From Equation (60), we obtain that

$$\mathbb{E} \left[ \left( \nu_1(\hat{\sigma}_{\hat{K},L}^2 - h) - q_2(K, \hat{K}) \right)_+ \right] \leq cq_2(K, K_{\max}) \exp(-C(K + K_{\max})) \leq \frac{C}{\sqrt{Nn}} \quad (61)$$

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

$$c\sqrt{Nn}q_2(K, K_{\max}) \exp(-C(K + K_{\max})) \rightarrow 0 \text{ as } N, n \rightarrow \infty \text{ (or } N = 1 \text{ and } n \rightarrow \infty).$$

Then, from Equations (61) and (59), there exists a constant  $C > 0$  such that

$$\mathbb{E} \left[ \left\| \hat{\sigma}_{\hat{K},L}^2 - \sigma^2 \right\|_{n,N}^2 \right] \leq 3 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M,L}} \|h - \sigma^2\|_n^2 + \text{pen}_2(K) \right\} + \frac{C}{\sqrt{Nn}}.$$

□

### 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 (31), for all  $h \in \mathcal{S}_{K+M}$ ,

$$\begin{aligned} \mathbb{E} \left[ \left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \mathbf{1}_{\Omega_{n,K_{\max}}} \right] &\leq 3 \inf_{h \in \mathcal{S}_{K+M}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + 20 \mathbb{E} \left( \sup_{h \in \mathcal{S}_{K+M}, \|h\|_X=1} \nu^2(h) \mathbf{1}_{\Omega_{n,K_{\max}}} \right) + C\Delta^2 \\ &\quad + 2 \left( \text{pen}(K) - \text{pen}(\widehat{K}) \right), \end{aligned}$$

where  $\mathcal{T}_{K,K'} = \{h \in \mathcal{S}_{K+M} + \mathcal{S}_{K'+M}, \|h\|_X = 1, \|h\|_\infty \leq \sqrt{L}\}$ . Let  $q_3 : \mathcal{K}^2 \rightarrow \mathbb{R}_+$  such that  $160q_3(K, K') \leq 18\text{pen}_3(K) + 16\text{pen}_3(K')$ .

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

$$\begin{aligned} \mathbb{E} \left[ \left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] &\leq 3 \left( \inf_{h \in \mathcal{S}_{K+M,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + \text{pen}_3(K) \right) + 20 \mathbb{E} \left[ \sup_{h \in \mathcal{T}_{K,\widehat{K}}} \nu_1^2(h) - q_3(K, \widehat{K}) \mathbf{1}_{\Omega_{n,K_{\max}}} \right] \\ &\quad + C\Delta^2 + CL \frac{K_{\max}^{2\gamma}}{n^{\gamma/2}}, \end{aligned}$$

where the numerical constant  $\gamma$  can take any value in  $(1, +\infty)$ , and

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

with  $\zeta_{k\Delta}^{1,1}$  the error term. We set for all  $K, K' \in \mathcal{K}$ ,  $G_K(K') := \sup_{h \in \mathcal{T}_{K,K'}} \nu_1^2(h)$  and  $\gamma = 20$ . Then, for  $K_{\max} = n^{1/5}$  and  $L = \log(n)$ , there exists  $C > 0$  such that

$$\begin{aligned} \mathbb{E} \left[ \left\| \widehat{\sigma}_{\widehat{K}}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] &\leq 3 \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + \text{pen}_3(K) \right\} + C\Delta^2 + C \frac{\log(n)}{n^2} \\ &\quad + 20 \sum_{K' \in \mathcal{K}} \mathbb{E} \left[ (G_K(K') - q_3(K, K'))_+ \mathbf{1}_{\Omega_{n,K_{\max}}} \right]. \end{aligned}$$

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

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

Then, using a similar reasoning as in Theorem 5.1, we obtain that,

$$\sum_{K' \in \mathcal{K}} \mathbb{E} \left[ (G_K(K') - q_3(K, K'))_+ \right] \leq \frac{C}{n},$$

where  $C > 0$  is a constant,  $q_3(K, K') \propto \sigma_1^4 \frac{(K+K'+2M)\sqrt{\log(n)}}{n}$  and  $\text{pen}_3(K) \propto \frac{(K+M)\log(n)}{n}$ . Then we obtain

$$\mathbb{E} \left[ \left\| \widehat{\sigma}_{\widehat{K},L}^2 - \sigma_{|I}^2 \right\|_{n,1}^2 \right] \leq \frac{3}{\tau_0} \inf_{K \in \mathcal{K}} \left\{ \inf_{h \in \mathcal{S}_{K+M,L}} \left\| h - \sigma_{|I}^2 \right\|_n^2 + \text{pen}_3(K) \right\} + \frac{C}{n}.$$

□

## 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* .
- Denis, C., Dion-Blanc, C. & Martinez, M. (2020). Consistent procedures for multiclass classification of discrete diffusion paths. *Scandinavian Journal of Statistics* **47**, 516–554.
- 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., Mintsa, 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.
- Domingo, D., d’Onofrio, A. & Flandoli, F. (2020). Properties of bounded stochastic processes employed in biophysics. *Stochastic Analysis and Applications* **38**, 277–306.
- Duffie, D. & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance* **6**, 379–406.
- El Karoui, N., Peng, S. & Quenez, M. C. (1997). Backward stochastic differential equations in finance. *Mathematical finance* **7**, 1–71.
- Erban, R. & Chapman, S. J. (2009). Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions. *Physical biology* **6**, 046001.
- Etheridge, A. (2011). *Some Mathematical Models from Population Genetics: École D’Été de Probabilités de Saint-Flour XXXIX-2009*, vol. 2012. Springer Science & Business Media.

- Ethier, S. N. & Norman, M. F. (1977). Error estimate for the diffusion approximation of the wright–fisher model. *Proceedings of the National Academy of Sciences* **74**, 5096–5098.
- 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.
- Gadat, S., Gerchinovitz, S. & Marteau, C. (2020). Optimal functional supervised classification with separation condition. *Bernoulli* **26**, 1797–1831.
- 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).  $l_p$  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.
- Jaillet, P., Lambertson, D. & Lapeyre, B. (1990). Variational inequalities and the pricing of american options. *Acta Applicandae Mathematica* **21**, 263–289.

- Karatzas, I. & Shreve, S. (2014). *Brownian motion and stochastic calculus*, vol. 113. Springer.
- Kijima, M., Tanaka, K. & Wong, T. (2009). A multi-quality model of interest rates. *Quantitative Finance* **9**, 133–145.
- Lamberton, D. & Lapeyre, B. (2011). *Introduction to stochastic calculus applied to finance*. CRC press.
- 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.
- Sbalzarini, I. F. (2006). Analysis, modeling, and simulation of diffusion processes in cell biology. Ph.D. thesis, ETH Zurich.
- 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.
- Tsybakov, A.-B. (2008). *Introduction to nonparametric estimation*. Springer Science & Business Media.



# 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  $\kappa$  of each penalty function.

As we already know, the adaptive estimator of  $\sigma^2$  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 (14). 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  $\kappa$  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  $\hat{\sigma}_{d_N, L_N}^2$  given in Equations (7) and (8).
  - (b) Select the optimal dimension  $\hat{K}_N \in \mathcal{K}$  using Equation (14)
  - (c) Using the learning sample  $D_{N'}$ , evaluate  $\left\| \hat{\sigma}_{\hat{d}_N, L_{N'}}^2 - \sigma_A^2 \right\|_{n, N'}^2$  where  $\hat{d}_N = \hat{K}_N + M$ .

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

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

**Proof of Lemma 3.1.** 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  $s \in [0, 1]$ ,

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

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

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

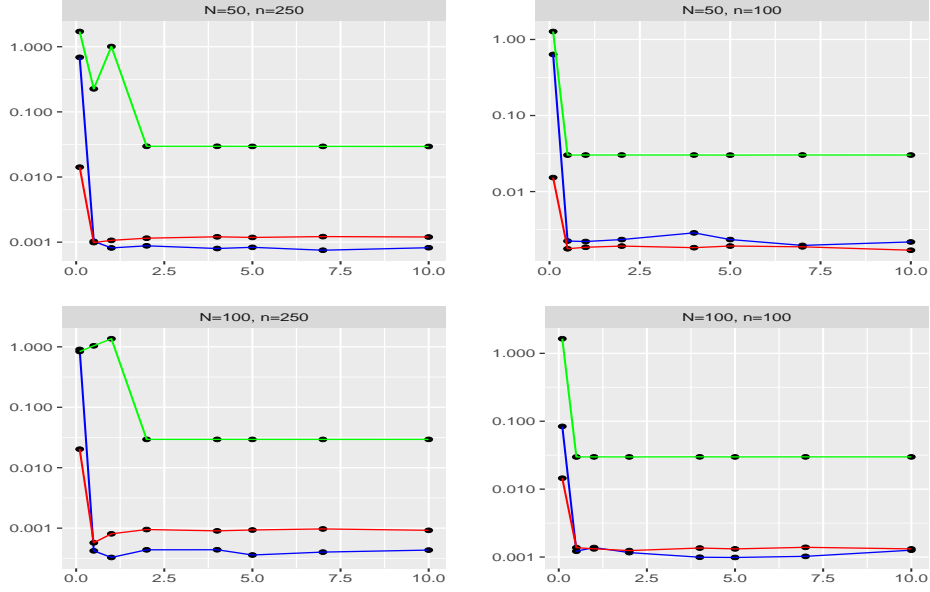


Figure 2: Calibration of the constant  $\kappa \in \mathcal{V}$  of the penalty function

**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:

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

□

**Proof of Lemma 8.5.** 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 each  $s \in [0, 1]$ , let  $F_s$  be the cumulative density function of the random variable  $X_s$ . 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

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

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  $x \in \mathbb{R}$ ,

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

□

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

$$|\|g\|_{n,1}^2 - \|g\|_X^2| \leq \int_0^1 |g^2(X_{\eta(s)}) - g^2(X_s)| ds \leq 2\|g\|_\infty \int_0^1 |g(X_{\eta(s)}) - g(X_s)| ds. \quad (62)$$

From Equation (11), 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

- **Collection [B]**:  $\|g\|_\infty \leq \|a\|_2 \leq C\sqrt{m}$  (see Denis *et al.* (2021))
- **Collection [CS–OB]**:  $\|g\|_\infty \leq Cm^{r'/2}$  since  $\|g\| = \|a\|_2$  and  $\sum_{\ell=0}^{m-1} \phi_\ell^2 = O(m^r)$ .

Moreover, each  $g \in \mathcal{S}_m$  such that  $\|g\|_X^2 = 1$  is a Lipschitz function with a Lipschitz coefficient  $L_g = O(m^{3/2})$  for the collection [B], or  $L_g = O(m^{r'/2})$  for the collection [CS–OB]. 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 collection [CS–OB], for all  $x, y \in I$  and using the Cauchy–Schwarz inequality together with Equation (5) and the finite increments theorem, there exists a constant  $C > 0$  such that

$$\begin{aligned} |g(x) - g(y)| &\leq \sum_{\ell=0}^{m-1} |a_\ell| \cdot |\phi_\ell(x) - \phi_\ell(y)| \\ &\leq \|a\|_2 \sqrt{\sum_{\ell=0}^{m-1} \|\phi'_\ell\|_\infty^2} |x - y| \\ &\leq Cm^{r'/2} |x - y|. \end{aligned}$$

Back to Equation (62), there exists a constant  $C > 0$  such that

$$|\|g\|_{n,1}^2 - \|g\|_X^2| \leq Cm^\alpha \int_0^1 |X_{\eta(s)} - X_s| ds \quad (63)$$

with  $\alpha = 2$  for the collection [B], and  $\alpha = (r + r')/2$  for the collection [CS – OB]. 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 (63), we obtain

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

Finally, using the Markov inequality, the Hölder inequality, Lemma (2.2), and Lemma 8.1, we conclude that

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

with  $\gamma \in (1, +\infty)$ . □

**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  $\mathcal{H}_m = \{h \in \mathcal{S}_m, \|h\|_n = 1\}$ . 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  $\varepsilon$ -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

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

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

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

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

$$\mathbb{P}(\Omega_{n,N,m}^c) \leq 2\mathcal{N}_\infty(\varepsilon, \mathcal{H}_m) \exp\left(-c\frac{N}{m}\right) \quad (64)$$

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

$$\mathcal{N}_\infty(\varepsilon, \mathcal{H}_m) \leq \left(\kappa\frac{\sqrt{m}}{\varepsilon}\right)^m \quad (65)$$

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 (64) and (65) 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

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

- 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 Lemma 8.9

*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,

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

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  $-Nnu\varepsilon + Nnav^2u^2/(1-bu) = -Nn\varepsilon^2/2(\varepsilon b + 2av^2)$  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 2.2.** Let  $q \geq 1$ . Suppose that  $X_0 = 0$ . Under Assumption 2.1, we deduce from Equation (1) that for all  $t \in [0, 1]$ ,

$$\begin{aligned} X_t &= \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dW_s \\ |X_t| &\leq \int_0^t |b(X_s)| ds + \left| \int_0^t \sigma(X_s) dW_s \right| \\ |X_t| &\leq \int_0^t C_0(1 + |X_s|) ds + \left| \int_0^t \sigma(X_s) dW_s \right| \end{aligned}$$

where the constant  $C_0 > 0$  depends on the constant  $L_0 > 0$  given in Assumption 2.1. Let  $p > 1$  be a real number such that  $1/p + 1/q = 1$ . From the Hölder inequality, there exists a constant  $C_q$  depending on  $q$  such that

$$\forall t \in [0, 1], \quad |X_t|^q \leq C_q \left( \int_0^t (1 + |X_s|^q) ds + \left| \int_0^t \sigma(X_s) dW_s \right|^q \right).$$

Using the Cauchy Schwarz inequality, one deduces that

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in [0,1]} |X_t|^q \right] &\leq C_q \left( \int_0^1 (1 + \mathbb{E} [|X_s|^q]) ds + \mathbb{E} \left[ \sup_{t \in [0,1]} \left| \int_0^t \sigma(X_s) dW_s \right|^q \right] \right) \\ &\leq C_q \left( \int_0^1 (1 + \mathbb{E} [|X_s|^q]) ds + \sqrt{\mathbb{E} \left[ \sup_{t \in [0,1]} \left| \int_0^t \sigma(X_s) dW_s \right|^{2q} \right]} \right). \end{aligned}$$

From the Doob inequality (see Revuz & Yor (2013), corollary 1.6), one has

$$\mathbb{E} \left[ \sup_{t \in [0,1]} \left| \int_0^t \sigma(X_s) dW_s \right|^q \right] \leq \left( \frac{2q}{2q-1} \right)^q \sqrt{\mathbb{E} \left[ \left| \int_0^1 \sigma(X_s) dW_s \right|^{2q} \right]}.$$

moreover, using the Burkholder-Davis-Gundy inequality, there exists a constant  $C_{1,q}$  depending on  $q \geq 1$  such that

$$\sqrt{\mathbb{E} \left[ \left| \int_0^1 \sigma(X_s) dW_s \right|^{2q} \right]} \leq C_{1,q} \mathbb{E} \left[ \int_0^1 \sigma^q(X_s) ds \right] < C_{1,q} \sigma_1^q.$$

Then, we deduce that

$$\mathbb{E} \left[ \sup_{t \in [0,1]} |X_t|^q \right] \leq C_q \left( \int_0^1 (1 + \mathbb{E}[|X_s|^q]) ds + C_{1,q} \sigma_1^q \right).$$

Finally, from Proposition 8.2, we have for all  $q \geq 1$  and for all  $s \in [0, 1]$ ,  $\mathbb{E}[|X_s|^q] < \infty$ .  $\square$

**Lemma .10.** *Under Assumption 2.1, we obtain from Equation (1) and the sample paths  $D_{N,n}$  the following regression model for the estimation of  $\sigma^2$ :*

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

where for each pair  $(j, k) \in \llbracket 1, N \rrbracket \times \llbracket 0, n-1 \rrbracket$ , the response variable  $U_{k\Delta_n}^j$  is given by

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

and the error terms are respectively given by  $\zeta_{k\Delta}^j = \zeta_{k\Delta}^{j,1} + \zeta_{k\Delta}^{j,2} + \zeta_{k\Delta}^{j,3}$ , with:

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

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

$$\zeta_{k\Delta_n}^{j,3} = 2b(X_{k\Delta_n}^j) \int_{k\Delta_n}^{(k+1)\Delta_n} \sigma(X_s^j) dW_s^j,$$

and  $R_{k\Delta}^j = R_{k\Delta}^{j,1} + R_{k\Delta}^{j,2} + R_{k\Delta}^{j,3}$ , with:

$$R_{k\Delta_n}^{j,1} = \frac{1}{\Delta_n} \left( \int_{k\Delta_n}^{(k+1)\Delta_n} b(X_s^j) ds \right)^2, \quad R_{k\Delta_n}^{j,3} = \frac{1}{\Delta_n} \int_{k\Delta_n}^{(k+1)\Delta_n} ((k+1)\Delta_n - s) \Phi(X_s^j) ds$$

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

where

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

**Proof of Lemma .10.** From Equation (1), we have

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

and for all  $(j, k) \in \llbracket 1, N \rrbracket \times \llbracket 0, n-1 \rrbracket$ ,

$$\begin{aligned} \left( X_{(k+1)\Delta}^j - X_{k\Delta}^j \right)^2 &= \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s^j)ds + \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^j)dW_s^j \right)^2 \\ &= \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s^j)ds \right)^2 + 2 \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s^j)ds \right) \left( \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^j)dW_s^j \right) \\ &\quad + \left( \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^j)dW_s^j \right)^2 \\ &= 2 \left( \int_{k\Delta}^{(k+1)\Delta} (b(X_s^j) - b(X_{k\Delta}^j))ds \right) \left( \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^j)dW_s^j \right) \\ &\quad + \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s^j)ds \right)^2 + \left( \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^j)dW_s^j \right)^2 - \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^j)ds \\ &\quad + \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^j)ds + 2\Delta b(X_{k\Delta}^j) \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s^j)dW_s^j. \end{aligned}$$

Then, we deduce that

$$\left( X_{(k+1)\Delta}^j - X_{k\Delta}^j \right)^2 = \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^j)ds + \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s^j)ds \right)^2 + \Delta R_{k\Delta}^{j,2} + \Delta \zeta_{k\Delta}^{j,1} + \Delta \zeta_{k\Delta}^{j,3}. \quad (66)$$

As  $\sigma$  is a  $\mathcal{C}^2$  function on  $\mathbb{R}$ , from the Itô formula, one has for all  $j \in \llbracket 1, N \rrbracket$

$$\begin{aligned} d(\sigma^2(X_t^j)) &= 2\sigma'(X_t^j)\sigma(X_t^j)dX_t^j + 2[\sigma'' + (\sigma')^2](X_t^j)dt \\ &= \Phi(X_t^j)dt + 2\sigma'(X_t^j)\sigma^2(X_t^j)dW_t^j \end{aligned}$$

with  $\Phi = 2b\sigma'\sigma + [\sigma''\sigma + (\sigma')^2]\sigma^2$ . We deduce that for all  $(j, k) \in \llbracket 1, N \rrbracket \times \llbracket 0, n-1 \rrbracket$  and  $s \in [k\Delta, (k+1)\Delta]$  tels que,

$$\sigma^2(X_s^j) = \sigma^2(X_{k\Delta}^j) + \int_{k\Delta}^s \Phi(X_u^j)du + 2 \int_{k\Delta}^s \sigma'(X_u^j)\sigma^2(X_u^j)dW_u^j.$$



We then integrate on the interval  $[k\Delta, (k+1)\Delta)$  and we obtain

$$\begin{aligned}
\int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^j) ds &= \Delta \sigma^2(X_{k\Delta}^j) + \int_{k\Delta}^{(k+1)\Delta} \int_{k\Delta}^s \Phi(X_u^j) du ds \\
&\quad + 2 \int_{k\Delta}^{(k+1)\Delta} \int_{k\Delta}^s \sigma'(X_u^j) \sigma^2(X_u^j) dW_u^j ds \\
&= \Delta \sigma^2(X_{k\Delta}^j) + \int_{k\Delta}^{(k+1)\Delta} \Phi(X_u^j) \int_u^{(k+1)\Delta} ds du \\
&\quad + 2 \int_{k\Delta}^{(k+1)\Delta} \sigma'(X_u^j) \sigma^2(X_u^j) \int_u^{(k+1)\Delta} ds dW_u^j \\
&= \Delta \sigma^2(X_{k\Delta}^j) + \int_{k\Delta}^{(k+1)\Delta} [(k+1)\Delta - u] \Phi(X_u^j) du \\
&\quad + 2 \int_{k\Delta}^{(k+1)\Delta} [(k+1)\Delta - u] \sigma'(X_u^j) \sigma(X_u^j) dW_u^j.
\end{aligned}$$

Thus, we obtain for al  $(j, k) \in \llbracket 1, N \rrbracket \times \llbracket 0, n-1 \rrbracket$ ,

$$\int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s^j) ds = \Delta \sigma^2(X_{k\Delta}^j) + \int_{k\Delta}^{(k+1)\Delta} [(k+1)\Delta - u] \Phi(X_u^j) du + \zeta_{k\Delta}^{j,2} \quad (67)$$

We finally deduce from Equations (66) et (67) that

$$\frac{\left( X_{(k+1)\Delta}^j - X_{k\Delta}^j \right)^2}{\Delta} = \sigma^2(X_{k\Delta}^j) + \zeta_{k\Delta}^j + R_{k\Delta}^j.$$

□