



# Nonparametric Estimation for I.I.D. Paths of Fractional SDE

Fabienne Comte, Nicolas Marie

## ► To cite this version:

Fabienne Comte, Nicolas Marie. Nonparametric Estimation for I.I.D. Paths of Fractional SDE. Statistical Inference for Stochastic Processes, 2021, 24 (3), pp.669-705. 10.1007/s11203-021-09246-4 . hal-02532339v2

**HAL Id: hal-02532339**

**<https://hal.science/hal-02532339v2>**

Submitted on 31 May 2021

**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 FOR I.I.D. PATHS OF FRACTIONAL SDE

FABIENNE COMTE\* AND NICOLAS MARIE†

**ABSTRACT.** This paper deals with nonparametric estimators of the drift function  $b$  computed from independent continuous observations, on a compact time interval, of the solution of a stochastic differential equation driven by the fractional Brownian motion (fSDE). First, a risk bound is established on a Skorokhod's integral based least squares oracle  $\hat{b}$  of  $b$ . Thanks to the relationship between the solution of the fSDE and its derivative with respect to the initial condition, a risk bound is deduced on a calculable approximation of  $\hat{b}$ . Another bound is directly established on an estimator of  $b'$  for comparison. The consistency and rates of convergence are established for these estimators in the case of the compactly supported trigonometric basis or the  $\mathbb{R}$ -supported Hermite basis.

**Keywords.** Fractional Brownian motion. Nonparametric projection estimator. Stochastic differential equation.

**MS classification 2020.** 62M09 - 62G08 - 60G22 - 60H07

## CONTENTS

1. Introduction	1
2. Stochastic integrals with respect to the fractional Brownian motion	3
2.1. The pathwise stochastic integral	3
2.2. Skorokhod's integral and density of the solution	4
3. Projection estimator of the drift function	7
3.1. A Skorokhod's integral based oracle	7
3.2. Rates in some usual bases	11
3.3. An approximate estimator	13
4. An alternative estimator	15
5. Concluding remarks	17
6. Proofs	18
6.1. Proof of Theorem 2.9	18
6.2. Proof of Proposition 2.11	18
6.3. Proof of Theorem 3.4	19
6.4. Proof of Corollary 3.5	22
6.5. Proof of Proposition 3.7	23
6.6. Proof of Proposition 3.11	24
6.7. Proof of Corollary 3.13	25
6.8. Proof of Proposition 4.2	26
6.9. Proof of Corollary 4.3	28
References	28

## 1. INTRODUCTION

Consider the stochastic differential equation

$$(1) \quad X(t) = x_0 + \int_0^t b(X(s))ds + \sigma B(t) ; t \in [0, T],$$

where  $\sigma, T > 0$ ,  $B$  is a fractional Brownian motion of Hurst index  $H \in ]1/2, 1[$ ,  $b : \mathbb{R} \rightarrow \mathbb{R}$  is a continuous map and  $x_0 \in \mathbb{R}^*$ .

In this work, we assume that we observe  $N$  i.i.d. paths of the solution  $X$  of Equation (1). For instance, this situation may occur in pharmacokinetics when a group of patients can be monitored: for each patient, a bolus of drug is injected and the "path" of its diffusion in the body can be observed (with a delay) (see details in Subsection 3.3.1). Our aim is to propose and study nonparametric estimators of the drift function  $b$  based on these observations. This problem is related to functional data analysis, and more specifically, there are various recent contributions about i.i.d. parametric models of (non fractional) stochastic differential equations with mixed effects (see, e.g., Ditlevsen and De Gaetano [24], Overgaard et al. [43], Picchini, De Gaetano and Ditlevsen [44], Picchini and Ditlevsen [45], Comte, Genon-Catalot and Samson [13], Delattre and Lavielle [20], Delattre, Genon-Catalot and Samson [17], Dion and Genon-Catalot [23], Delattre, Genon-Catalot and Larédo [18]). Also, i.i.d. samples of stochastic differential equations have been recently considered in the framework of multiclass classification of diffusions (see Denis, Dion and Martinez [21]). The need of flexibility to deal with the information contained in functional data analysis make it interesting to use a nonparametric approach.

Along the last two decades, many authors studied statistical inference from observations drawn from stochastic differential equations driven by fractional Brownian motion, considering the observation of one path, either in continuous time, or in discrete time with fixed or small step size.

Most references on the estimation of the trend component in Equation (1) deal with parametric estimators. Let us start by papers considering continuous time observations. In Kleptsyna and Le Breton [28] and Hu and Nualart [30], strategies to estimate the trend component in Langevin's equation are studied. Kleptsyna and Le Breton [28] provide a maximum likelihood estimator, where the stochastic integral with respect to the solution of Equation (1) returns to an Itô integral. In [50], Tudor and Viens extend this estimator to equations with a drift function depending linearly on the unknown parameter. Hu and Nualart [30] provide a least squares oracle, not an estimator, because the stochastic integral with respect to the solution of Equation (1) is taken in the sense of Skorokhod and is not computable. In [31], Hu, Nualart and Zhou extend this oracle to equations with a drift function depending linearly on the unknown parameter. Finally, in [37], Marie and Raynaud de Fitte extend this oracle to non-homogeneous semi-linear equations with almost periodic coefficients.

Now, considering discrete time observations, still in the parametric context, Tindel and Neuenkirch [40] study a least squares-type estimator defined by an objective function, tailor-made with respect to the main result of Tudor and Viens [51] on the rate of convergence of the quadratic variation of the fractional Brownian motion. In [47], Panloup, Tindel and Varvenne extend the results of [40] under much more flexible conditions. In [8], Chronopoulou and Tindel provide a likelihood based numerical procedure to estimate a parameter involved in both the drift and the volatility functions in a stochastic differential equation with multiplicative fractional noise.

About nonparametric methods for the estimation of the function  $b$  in Equation (1), there are only few references. Saussereau [48] and Comte and Marie [15] study the consistency of Nadaraya-Watson type estimators of the drift function  $b$  in Equation (1). In [38], Mishra and Prakasa Rao established the consistency and a rate of convergence of a nonparametric estimator of the whole trend of the solution to Equation (1) extending that of Kutoyants [32]. Marie [36] deals with the same estimator but for reflected fractional SDE. For nonparametric kernel-based estimators in Itô's calculus framework, the reader is referred to Kutoyants [32] and [33].

The present paper deals with nonparametric estimators of  $b$ , computed from  $N$  independent continuous time observations of the solution of Equation (1) on  $[0, T]$ . Let us mention that it became usual that such functional data are available and can be processed thanks to the improvements of computers. The question of nonparametric drift estimation in stochastic differential equations from such data has been studied in Comte and Genon-Catalot [12] who consider an Itô's calculus framework.

On the one hand, we extend the functional least squares strategy of Comte and Genon-Catalot [12] to

fractional SDE by replacing Itô's integral by Skorokhod's integral in the objective function defining their projection estimator. The Skorokhod integral is a Malliavin calculus based stochastic integral extending Itô's integral to nonsemimartingale Gaussian signals as the fractional Brownian motion, but which has the major drawback to be non computable in general. For this reason, the oracle  $\widehat{b}$  mentioned above is just an auxiliary *entity*, on which we are able to establish a satisfactory risk bound. Thanks to the relationship between the solution  $X_{x_0}$  of Equation (1) and its derivative with respect to the initial condition  $x_0$ , we define an approximate estimator  $\widehat{b}_\varepsilon$  of  $\widehat{b}$ , calculable from  $N$  i.i.d. paths of the couple  $(X_{x_0}, X_{x_0+\varepsilon})$ , where  $x_0$  and  $x_0 + \varepsilon$  are two close initial conditions. We deduce a risk bound on the estimator  $\widehat{b}_\varepsilon$  from the risk bound established on  $\widehat{b}$ .

On the other hand, by using the relationship between  $X_{x_0}$  and  $\partial_{x_0} X_{x_0}$  directly, we define an estimator  $\widehat{b}_\varepsilon^{\dagger, \prime}$  of the derivative  $b'$  of  $b$ , also calculable from  $N$  i.i.d. paths of the couple  $(X_{x_0}, X_{x_0+\varepsilon})$ . We prove a risk bound on  $\widehat{b}_\varepsilon^{\dagger, \prime}$  with a parametric rate of convergence, but then, we show that deducing a risk bound on a primitive estimator  $\widehat{b}_\varepsilon^\dagger$  of  $b$  is not straightforward unless the function is known at one point. In addition, a compactness condition on the support of  $b$  must be added. For this reason, the estimator  $\widehat{b}_\varepsilon$  is of interest to estimate  $b$  and  $\widehat{b}_\varepsilon^{\dagger, \prime}$  is of interest to estimate  $b'$ , both functions may be useful depending on the application context.

Note that almost all the references cited above on the statistical inference for fractional SDE are based on long-time behavior properties of the solution which are often difficult to check in practice, but not required here.

The oracle  $\widehat{b}$  and the estimator  $\widehat{b}_\varepsilon$  are respectively studied in Subsections 3.1 and 3.3 of Section 3. Subsection 3.2 provides examples of function bases well adapted in our situation. We can in these frameworks obtain convergence results and rates. The estimators  $\widehat{b}_\varepsilon^{\dagger, \prime}$  and  $\widehat{b}_\varepsilon^\dagger$  are studied in Section 4. Lastly, concluding remarks are gathered in Section 5 while most proofs are postponed in Section 6.

**Notations.** The vector space of Lipschitz continuous maps from  $\mathbb{R}$  into itself is denoted by  $\text{Lip}(\mathbb{R})$  and equipped with the usual Lipschitz semi-norm  $\|\cdot\|_{\text{Lip}}$ . Now, consider  $m \in \mathbb{N}^*$ . The Euclidean norm on  $\mathbb{R}^m$  is denoted by  $\|\cdot\|_{2,m}$ ,

$$C_{\mathbf{b}}^m(\mathbb{R}) := \left\{ \varphi \in C^m(\mathbb{R}) : \max_{k \in \llbracket 0, m \rrbracket} \|\varphi^{(k)}\|_\infty < \infty \right\}$$

and

$$\text{Lip}_{\mathbf{b}}^m(\mathbb{R}) := \left\{ \varphi \in C^m(\mathbb{R}) : \|\varphi\|_{\text{Lip}_{\mathbf{b}}^m} = \|\varphi\|_{\text{Lip}} \vee \max_{k \in \llbracket 1, m \rrbracket} \|\varphi^{(k)}\|_\infty < \infty \right\}.$$

Note that  $C_{\mathbf{b}}^m(\mathbb{R}) \subset \text{Lip}_{\mathbf{b}}^m(\mathbb{R})$ . Finally, for every  $n \in \mathbb{N}^*$ , the vector space of infinitely continuously differentiable maps  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $f$  and all its partial derivatives have polynomial growth is denoted by  $C_{\mathbf{p}}^\infty(\mathbb{R}^n, \mathbb{R})$ .

## 2. STOCHASTIC INTEGRALS WITH RESPECT TO THE FRACTIONAL BROWNIAN MOTION

This section presents two different methods to define a stochastic integral with respect to the fractional Brownian motion. The first one is based on the pathwise properties of the fractional Brownian motion. Another stochastic integral with respect to the fractional Brownian motion is defined via the Malliavin divergence operator. This stochastic integral is called Skorokhod's integral with respect to  $B$ . If  $H = 1/2$ , which means that  $B$  is a Brownian motion, the Skorokhod integral defined via the divergence operator coincides with Itô's integral on its domain. This integral is appropriate to define a suitable oracle of the drift function  $b$  in Equation (1), while the first one is used in Section 3.3 to propose a calculable approximation.

**2.1. The pathwise stochastic integral.** This subsection deals with some definitions and basic properties of the pathwise stochastic integral with respect to the fractional Brownian motion of Hurst index greater than  $1/2$ .

**Definition 2.1.** Consider  $x$  and  $w$  two continuous functions from  $[0, T]$  into  $\mathbb{R}$ . Consider a dissection  $D := (t_0, \dots, t_m)$  of  $[s, t]$  with  $m \in \mathbb{N}^*$  and  $s, t \in [0, T]$  such that  $s < t$ . The Riemann sum of  $x$  with respect to  $w$  on  $[s, t]$  for the dissection  $D$  is

$$J_{x,w,D}(s, t) := \sum_{k=0}^{m-1} x(t_k)(w(t_{k+1}) - w(t_k)).$$

**Notation.** With the notations of Definition 2.1, the mesh of the dissection  $D$  is

$$\pi(D) := \max_{k \in \llbracket 0, m-1 \rrbracket} |t_{k+1} - t_k|.$$

The following theorem ensures the existence and the uniqueness of Young's integral (see Friz and Victoir [26], Theorem 6.8).

**Theorem 2.2.** Let  $x$  (resp.  $w$ ) be a  $\alpha$ -Hölder (resp.  $\beta$ -Hölder) continuous map from  $[0, T]$  into  $\mathbb{R}$  with  $\alpha, \beta \in ]0, 1]$  such that  $\alpha + \beta > 1$ . There exists a unique continuous map  $J_{x,w} : [0, T] \rightarrow \mathbb{R}$  such that for every  $s, t \in [0, T]$  satisfying  $s < t$  and any sequence  $(D_n)_{n \in \mathbb{N}}$  of dissections of  $[s, t]$  such that  $\pi(D_n) \rightarrow 0$  as  $n \rightarrow \infty$ ,

$$\lim_{n \rightarrow \infty} |J_{x,w}(t) - J_{x,w}(s) - J_{x,w,D_n}(s, t)| = 0.$$

The map  $J_{x,w}$  is the Young integral of  $x$  with respect to  $w$  and  $J_{x,w}(t) - J_{x,w}(s)$  is denoted by

$$\int_s^t x(u)dw(u)$$

for every  $s, t \in [0, T]$  such that  $s < t$ .

For any  $\alpha \in ]1/2, H[$ , the paths of  $B$  are  $\alpha$ -Hölder continuous (see Nualart [42], Section 5.1). So, for every process  $Y = (Y(t))_{t \in [0, T]}$  with  $\beta$ -Hölder continuous paths from  $[0, T]$  into  $\mathbb{R}$  such that  $\alpha + \beta > 1$ , by Theorem 2.2, it is natural to define the pathwise stochastic integral of  $Y$  with respect to  $B$  by

$$\left( \int_0^t Y(s)dB(s) \right) (\omega) := \int_0^t Y(\omega, s)dB(\omega, s)$$

for every  $\omega \in \Omega$  and  $t \in [0, T]$ .

**2.2. Skorokhod's integral and density of the solution.** This subsection deals with some definitions and results on Malliavin calculus.

Let  $\mathbb{L}^0([0, T], \mathbb{R})$  be the space of measurable functions from  $[0, T]$  into  $\mathbb{R}$ , and consider the reproducing kernel Hilbert space

$$\mathcal{H} := \{h \in \mathbb{L}^0([0, T], \mathbb{R}) : \langle h, h \rangle_{\mathcal{H}} < \infty\}$$

of  $B$ , where  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  is the inner product defined by

$$\langle h, \eta \rangle_{\mathcal{H}} := H(2H - 1) \int_0^T \int_0^T |t - s|^{2H-2} h(s)\eta(s)dsdt$$

for every  $h, \eta \in \mathbb{L}^0([0, T], \mathbb{R})$ . Let  $(\mathbf{B}(h))_{h \in \mathcal{H}}$  be the isonormal Gaussian process defined by

$$\mathbf{B}(h) := \int_0^\cdot h(s)dB(s),$$

which is the Wiener integral of  $h \in \mathcal{H}$  with respect to  $B$ .

**Definition 2.3.** The Malliavin derivative of a smooth functional

$$F = f(\mathbf{B}(h_1), \dots, \mathbf{B}(h_n))$$

where  $n \in \mathbb{N}^*$ ,  $f \in C_{\mathbf{p}}^\infty(\mathbb{R}^n, \mathbb{R})$  and  $h_1, \dots, h_n \in \mathcal{H}$ , is the  $\mathcal{H}$ -valued random variable

$$\mathbf{D}F := \sum_{k=1}^n \partial_k f(\mathbf{B}(h_1), \dots, \mathbf{B}(h_n)) h_k.$$

The key property of the operator  $\mathbf{D}$  is the following one.

**Proposition 2.4.** *The map  $\mathbf{D}$  is closable from  $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$  into  $\mathbb{L}^2(\Omega; \mathcal{H})$ . Its domain in  $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ , denoted by  $\mathbb{D}^{1,2}$ , is the closure of the smooth functionals space for the seminorm  $\|\cdot\|_{1,2}$  defined by*

$$\|F\|_{1,2}^2 := \mathbb{E}(|F|^2) + \mathbb{E}(\|\mathbf{D}F\|_{\mathcal{H}}^2) < \infty$$

for every  $F \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ . The Malliavin derivative of  $F \in \mathbb{D}^{1,2}$  at time  $s \in [0, T]$  is denoted by  $\mathbf{D}_s F$ .

For a proof, see Nualart [42], Proposition 1.2.1.

**Definition 2.5.** *The adjoint  $\delta$  of the Malliavin derivative  $\mathbf{D}$  is the divergence operator. The domain of  $\delta$  is denoted by  $\text{dom}(\delta)$ , and  $u \in \text{dom}(\delta)$  if and only if there exists a deterministic constant  $\mathbf{c}_u > 0$  such that for every  $F \in \mathbb{D}^{1,2}$ ,*

$$|\mathbb{E}(\langle \mathbf{D}F, u \rangle_{\mathcal{H}})| \leq \mathbf{c}_u \mathbb{E}(|F|^2)^{1/2}.$$

For any process  $Y = (Y(s))_{s \in [0, T]}$  and every  $t \in [0, T]$ , if  $Y\mathbf{1}_{[0, t]} \in \text{dom}(\delta)$ , then its Skorokhod integral with respect to  $B$  is defined on  $[0, t]$  by

$$\int_0^t Y(s) \delta B(s) := \delta(Y\mathbf{1}_{[0, t]}),$$

and its Skorokhod integral with respect to  $X$  is defined by

$$\int_0^t Y(s) \delta X(s) := \int_0^t Y(s) b(X(s)) ds + \sigma \int_0^t Y(s) \delta B(s).$$

Note that since  $\delta$  is the adjoint of the Malliavin derivative  $\mathbf{D}$ , the Skorokhod integral of  $Y$  with respect to  $B$  on  $[0, t]$  is a centered random variable. Indeed,

$$(2) \quad \mathbb{E} \left( \int_0^t Y(s) \delta B(s) \right) = \mathbb{E}(1 \cdot \delta(Y\mathbf{1}_{[0, t]})) = \mathbb{E}(\langle \mathbf{D}(1), Y\mathbf{1}_{[0, t]} \rangle_{\mathcal{H}}) = 0.$$

Let  $\mathcal{S}$  be the space of the smooth functionals presented in Definition 2.3 and consider  $\mathbb{D}^{1,2}(\mathcal{H})$ , the closure of

$$\mathcal{S}_{\mathcal{H}} := \left\{ \sum_{j=1}^n F_j h_j ; h_1, \dots, h_n \in \mathcal{H}, F_1, \dots, F_n \in \mathcal{S} \right\}$$

for the seminorm  $\|\cdot\|_{1,2,\mathcal{H}}$  defined by

$$\|u\|_{1,2,\mathcal{H}}^2 := \mathbb{E}(\|u\|_{\mathcal{H}}^2) + \mathbb{E}(\|\mathbf{D}u\|_{\mathcal{H} \otimes \mathcal{H}}^2) < \infty$$

for every  $u \in \mathbb{L}^2(\Omega \times [0, T])$  (see Nualart [42], p. 31, Remark 2). The following proposition provides an isometry type property for the Skorokhod integral with respect to  $B$  on  $\mathbb{D}^{1,2}(\mathcal{H})$ , which is a subspace of  $\text{dom}(\delta)$  by Nualart [42], Proposition 1.3.1. This result is useful for our purpose and is proved in Biagini et al. [3] (see Theorem 3.11.1).

**Proposition 2.6.** *For every  $Y, Z \in \mathbb{D}^{1,2}(\mathcal{H})$ ,*

$$\begin{aligned} \mathbb{E}(\delta(Y)\delta(Z)) &= \alpha_H \int_0^T \int_0^T \mathbb{E}(Y(u)Z(v)) |v - u|^{2H-2} dv du \\ &\quad + \alpha_H^2 \int_{[0, T]^4} \mathbb{E}(\mathbf{D}_u Y(v) \mathbf{D}_v Z(u)) |u - u'|^{2H-2} |v - v'|^{2H-2} du du' dv dv'. \end{aligned}$$

In the sequel, the function  $b$  fulfills the following assumption.

**Assumption 2.7.** *The function  $b$  belongs to  $C^1(\mathbb{R})$  and there exist  $m, M \in \mathbb{R}$  such that*

$$m \leq b'(x) \leq M ; \forall x \in \mathbb{R}.$$

Under Assumption 2.7, the following result is a straightforward application of Proposition 2.6 to functionals of the solution  $X$  of Equation (1).

**Corollary 2.8.** *Let  $X$  be the solution of Equation (1). Under Assumption 2.7,  $X \in \mathbb{D}^{1,2}(\mathcal{H})$  and for every  $\varphi, \psi \in \text{Lip}_b^1(\mathbb{R})$ ,*

$$\mathbb{E}(\delta(\varphi(X))\delta(\psi(X))) = \alpha_H \int_0^T \int_0^T \mathbb{E}(\varphi(X(u))\psi(X(v))|v-u|^{2H-2}dvdu) + R_{\varphi,\psi}$$

where

$$\begin{aligned} R_{\varphi,\psi} &:= \alpha_H^2 \int_{[0,T]^4} \mathbb{E}(\varphi'(X(v))\psi'(X(v'))\mathbf{D}_u X(v)\mathbf{D}_u X(v'))|u-u'|^{2H-2}|v-v'|^{2H-2}dudv'dv' \\ &= \alpha_H^2 \sigma^2 \int_{[0,T]^2} \int_0^v \int_0^{v'} |u-u'|^{2H-2}|v-v'|^{2H-2} \\ &\quad \times \mathbb{E} \left( \varphi'(X(v))\psi'(X(v')) \exp \left( \int_{u'}^v b'(X(s))ds + \int_u^{v'} b'(X(s))ds \right) \right) dudv'dv'. \end{aligned}$$

The following theorem provides suitable controls of the moments of Skorokhod's integral.

**Theorem 2.9.** *Under Assumption 2.7, for every  $p > 1/H$ , there exists a deterministic constant  $\mathbf{c}_{p,H,\sigma} > 0$ , only depending on  $p, H$  and  $\sigma$ , such that for every  $\varphi \in \text{Lip}_b^1(\mathbb{R})$ ,*

$$\begin{aligned} \mathbb{E} \left( \left| \int_0^T \varphi(X(s))\delta B(s) \right|^p \right) &\leq \mathbf{c}_{p,H,\sigma} \bar{\mathbf{m}}_{p,H,M}(T) \left[ \left( \int_0^T \mathbb{E}(|\varphi(X(s))|^{1/H})ds \right)^{pH} \right. \\ &\quad \left. + \left( \int_0^T \mathbb{E}(|\varphi'(X(s))|^p)^{1/(pH)}ds \right)^{pH} \right] < \infty \end{aligned}$$

where  $\bar{\mathbf{m}}_{p,H,M}(T) := \mathbf{m}_{p,H,M}(T) \vee 1$  and

$$\mathbf{m}_{p,H,M}(T) := \left( -\frac{H}{M} \right)^{pH} \mathbf{1}_{M < 0} + T^{pH} \mathbf{1}_{M=0} + \left( \frac{H}{M} \right)^{pH} e^{pMT} \mathbf{1}_{M > 0}.$$

Note that if  $M < 0$ , then Theorem 2.9 has been already proved in Hu, Nualart and Zhou [31] (see Proposition 4.4.(2)).

**Remark 2.10.** *On the one hand, note that the control of the variance of Skorokhod's integral provided in Theorem 2.9 is a straightforward consequence of Corollary 2.8. On the other hand, with the notations of Corollary 2.8, note that for  $H = 1/2$ , the solution  $X$  of Equation (1) is adapted and then*

$$R_{\varphi,\psi} = \int_0^T \int_0^T \mathbb{E}(\mathbf{D}_u X(v)\mathbf{D}_v X(u))dudv = 0.$$

*This reduces importantly the order of the variance of Skorokhod's integral with respect to the case  $H > 1/2$ .*

Lastly, the following proposition provides an expression and a bound for the density of the solution to Equation (1).

**Proposition 2.11.** *Under Assumption 2.7, for any  $t \in ]0, T]$ , the probability distribution of  $X^*(t) := X(t) - \mathbb{E}(X(t))$  has a  $\mathbb{R}$ -supported density with respect to Lebesgue's measure  $p_t^*(x_0, \cdot)$  such that, for every  $x \in \mathbb{R}$ ,*

$$p_t^*(x_0, x) = \frac{\mathbb{E}(|X^*(t)|)}{2g_t^*(x_0, x)} \exp \left( - \int_0^x \frac{zdz}{g_t^*(x_0, z)} \right)$$

where

$$\begin{aligned} g_t^*(x_0, x) &:= \mathbb{E}(\langle \mathbf{D}X(t), -\mathbf{D}\mathbf{L}^{-1}X(t) \rangle_{\mathcal{H}} | X^*(t) = x) \\ &\in [\sigma(m, t), \sigma(M, t)] \subset ]0, \infty[, \end{aligned}$$

$\mathbf{L}$  is the Ornstein-Uhlenbeck operator and

$$\sigma(\mu, t)^2 := \alpha_H \sigma^2 \int_0^t \int_0^t |v-u|^{2H-2} e^{\mu(2t-v-u)} dudv > 0 ; \forall \mu \in \mathbb{R}.$$

In particular, for every  $x \in \mathbb{R}$ ,

$$\frac{\mathbb{E}(|X^*(t)|)}{2\sigma(M, t)^2} \exp\left(-\frac{x^2}{2\sigma(m, t)^2}\right) \leq p_t^*(x_0, x) \leq \frac{\mathbb{E}(|X^*(t)|)}{2\sigma(m, t)^2} \exp\left(-\frac{x^2}{2\sigma(M, t)^2}\right).$$

A straightforward consequence of Proposition 2.11 is that for any  $t \in ]0, T]$ , the probability distribution of  $X(t)$  has a  $\mathbb{R}$ -supported density with respect to Lebesgue's measure  $p_t(x_0, \cdot)$  such that, for every  $x \in \mathbb{R}$ ,

$$p_t(x_0, x) = p_t^*(x_0, x - \mathbb{E}(|X(t)|))$$

and

$$(3) \quad \mathfrak{m}_t(x_0, x, M, m) \leq p_t(x_0, x) \leq \mathfrak{m}_t(x_0, x, m, M),$$

where

$$\mathfrak{m}_t(x_0, x, \mu_1, \mu_2) := \frac{\mathbb{E}(|X^*(t)|)}{2\sigma(\mu_1, t)^2} \exp\left[-\frac{(x - \mathbb{E}(|X(t)|))^2}{2\sigma(\mu_2, t)^2}\right] ; \forall \mu_1, \mu_2 \in \mathbb{R}.$$

Since the paths of  $X$  are  $\alpha$ -Hölder continuous for any  $\alpha \in ]0, H[$ ,

$$\begin{aligned} \mathbb{E}(|X^*(t)|) &= \mathbb{E}(|X(t) - x_0 - \mathbb{E}(X(t) - x_0)|) \\ &\leq 2\mathbb{E}(|X(t) - X(0)|) \leq 2\mathbb{E}(\|X\|_{\alpha\text{-Hölder}, T})t^\alpha \end{aligned}$$

where

$$\|X\|_{\alpha\text{-Hölder}, T} := \sup_{0 \leq s < t \leq T} \frac{|X(t) - X(s)|}{|t - s|^\alpha},$$

which has a finite first order moment because  $\mathbb{E}(\|B\|_{\alpha\text{-Hölder}, T}) < \infty$  and  $b$  is Lipschitz continuous. Then, since  $\sigma(m, t)^2 \geq \sigma^2 e^{-2\|b'\|_\infty T} t^{2H}$ ,

$$\mathfrak{m}_t(x_0, x, m, M) \leq \mathfrak{c}_T t^{\alpha-2H}$$

with

$$\mathfrak{c}_T := \frac{\mathbb{E}(\|X\|_{\alpha\text{-Hölder}, T})}{\sigma^2 e^{-2\|b'\|_\infty T}}.$$

Therefore, by taking  $\alpha \in ]2H - 1, H[$ , Inequality (3) implies that for every  $x \in \mathbb{R}$ ,  $p_\cdot(x_0, x) \in \mathbb{L}^1([0, T], dt)$ .

### 3. PROJECTION ESTIMATOR OF THE DRIFT FUNCTION

Under Assumption 2.7,  $b$  is Lipschitz continuous on  $\mathbb{R}$  and its derivative is bounded. So, Equation (1) has a unique solution  $X$  and the associated Itô map  $\mathcal{I}$  is continuously differentiable from  $\mathbb{R} \times C^0([0, T], \mathbb{R})$  into  $C^0([0, T], \mathbb{R})$ .

**3.1. A Skorokhod's integral based oracle.** This subsection deals with an oracle of  $b$ , constructed as the estimator of Comte and Genon-Catalot [12] by replacing Itô's integral by Skorokhod's one. The risk bound on this oracle will allow us to establish a risk bound on an estimator at Subsection 3.3.

**3.1.1. The objective function.** Let  $f_T$  be the function defined by

$$f_T(x) := \frac{1}{T} \int_0^T p_s(x_0, x) ds ; \forall x \in \mathbb{R},$$

where  $p_s(x_0, \cdot)$  is the density with respect to Lebesgue's measure of the probability distribution of  $X(s)$  for any  $s \in ]0, T]$ . For any  $x \in \mathbb{R}$ ,  $p_s(x_0, x)$  is well defined and belongs to  $]0, \infty[$  because  $\text{supp}(p_s(x_0, \cdot)) = \mathbb{R}$  (see Proposition 2.11). So, since  $p_\cdot(x_0, x) \in \mathbb{L}^1([0, T], dt)$  as established at the end of Section 2,  $f_T(x)$  is well defined, belongs to  $]0, \infty[$ , and then  $\text{supp}(f_T) = \mathbb{R}$ . Moreover, by Fubini-Tonelli's theorem,  $f_T$  is measurable and

$$\int_{-\infty}^{\infty} f_T(x) dx = \frac{1}{T} \int_0^T \int_{-\infty}^{\infty} p_s(x_0, x) dx ds = 1.$$

So,  $f_T$  is a  $\mathbb{R}$ -supported density function.



Now, consider  $N \in \mathbb{N}^*$  independent copies  $B_1, \dots, B_N$  of  $B$ ,  $X^i := \mathcal{I}(x_0, B^i)$  for every  $i \in \{1, \dots, N\}$ , and the objective function  $\gamma_N$  defined by

$$\gamma_N(\tau) := \frac{1}{NT} \sum_{i=1}^N \left( \int_0^T \tau(X^i(s))^2 ds - 2 \int_0^T \tau(X^i(s)) \delta X^i(s) \right)$$

for every function  $\tau : \mathbb{R} \rightarrow \mathbb{R}$ .

Note that for any bounded function  $\tau$  from  $\mathbb{R}$  into itself, thanks to Equality (2),

$$\begin{aligned} \mathbb{E}(\gamma_N(\tau)) &= \frac{1}{T} \int_0^T \mathbb{E}(\tau(X(s))^2 - 2\tau(X(s))b(X(s)))ds + \frac{\sigma}{T} \mathbb{E} \left( \int_0^T \tau(X(s)) \delta B(s) \right) \\ &= \frac{1}{T} \int_0^T \mathbb{E}((\tau(X(s)) - b(X(s)))^2)ds - \frac{1}{T} \int_0^T \mathbb{E}(b(X(s))^2)ds. \end{aligned}$$

Then, the definition of  $f_T$  gives

$$(4) \quad \mathbb{E}(\gamma_N(\tau)) = \int_{-\infty}^{\infty} (\tau(x) - b(x))^2 f_T(x) dx - \int_{-\infty}^{\infty} b(x)^2 f_T(x) dx.$$

Equality (4) shows that  $\mathbb{E}(\gamma_N(\tau))$  is the smallest for  $\tau$  the nearest of  $b$ . Therefore, to minimize its empirical version  $\gamma_N(\tau)$  should provide a function near of  $b$ .

**Remark 3.1.** *The pathwise stochastic integral with respect to  $B$ , defined in Subsection 2.1, is not centered in general, and not even for  $H = 1/2$ . Indeed, if  $H = 1/2$ , then it coincides with Stratonovich's integral. This is the main reason why the objective function above is defined via Skorokhod's integral. Moreover, the pathwise stochastic integral doesn't satisfy an isometry type property as Skorokhod's integral (see Proposition 2.6), which is crucial in the sequel.*

**3.1.2. The oracle.** Consider  $A \in \mathcal{B}(\mathbb{R})$  and assume that  $\mathbb{L}^2(A, dx)$  (resp.  $\mathbb{L}^2(A, f_T(x)dx)$ ) is equipped with its usual inner product  $\langle \cdot, \cdot \rangle$  (resp.  $\langle \cdot, \cdot \rangle_{f_T}$ ). For any  $m \in \mathbb{N}^*$ , consider also

$$\mathcal{S}_m := \text{span}\{\varphi_0, \dots, \varphi_{m-1}\},$$

where  $(\varphi_0, \dots, \varphi_{m-1})$  is an orthonormal family of  $\mathbb{L}^2(A, dx)$ . Moreover, assume that the functions  $\varphi_j$ ,  $j \in \mathbb{N}$  are bounded. So,  $\mathcal{S}_m \subset \mathbb{L}^2(A, f_T(x)dx)$ .

Consider

$$\widehat{\Psi}(m) := \left( \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X^i(s)) \varphi_k(X^i(s)) ds \right)_{j,k=0,\dots,m-1},$$

assume that this matrix is invertible, and let

$$\widehat{b}_m = \arg \min_{\tau \in \mathcal{S}_m} \gamma_N(\tau)$$

be the Skorokhod's integral based projection oracle of  $b_A := b|_A$  on  $\mathcal{S}_m$ . As in Comte and Genon-Catalot [12], Section 2.2,

$$\widehat{b}_m = \sum_{j=0}^{m-1} \widehat{\theta}_j \varphi_j$$

where

$$\widehat{\theta}(m) := (\widehat{\theta}_0, \dots, \widehat{\theta}_{m-1})^* = \widehat{\Psi}(m)^{-1} \widehat{\mathbf{x}}(m)$$

with

$$\widehat{\mathbf{x}}(m) := \left( \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X^i(s)) \delta X^i(s) \right)_{j=0,\dots,m-1}^*.$$

Note that

$$\widehat{\mathbf{x}}(m) = (\langle \varphi_j, b \rangle_N)_{j=0,\dots,m-1}^* + \mathbf{e}(m)$$

and

$$\widehat{\Psi}(m) = (\langle \varphi_j, \varphi_k \rangle_N)_{j,k=0,\dots,m-1},$$

where

$$\langle \tau_1, \tau_2 \rangle_N := \frac{1}{NT} \sum_{i=1}^N \int_0^T \tau_1(X^i(s)) \tau_2(X^i(s)) ds$$

for every measurable functions  $\tau_1, \tau_2 : \mathbb{R} \rightarrow \mathbb{R}$  (the associated norm is denoted by  $\|\cdot\|_N$ ), and

$$\mathbf{e}(m) := \left( \frac{\sigma}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X^i(s)) \delta B^i(s) \right)_{j=0,\dots,m-1}^*.$$

By Equality (2),  $\mathbf{e}(m)$  is centered, as expected for an error term in regression.

**3.1.3. Risk bounds on the oracle.** Throughout this subsection,  $f_T$  and the functions  $\varphi_j$ ,  $j \in \mathbb{N}$  fulfill the following assumption.

**Assumption 3.2.**  $\lambda(A) > 0$  and, for  $m \leq NT$ ,

- (1)  $(\varphi_0, \dots, \varphi_{m-1})$  is an orthonormal family of  $\mathbb{L}^2(A, dx)$ .
- (2) The functions  $\varphi_j$ ,  $j = 0, \dots, m-1$ , are bounded and belong to  $C_b^1(A)$ .
- (3) There exist  $x_0, \dots, x_{m-1} \in A$  such that

$$\det[(\varphi_j(x_k))_{j,k=0,\dots,m-1}] \neq 0.$$

By Comte and Genon-Catalot [12], Lemma 1, which remains true for  $H > 1/2$  without additional arguments,

$$\Psi(m) := \mathbb{E}(\widehat{\Psi}(m)) = \left( \int_A \varphi_j(x) \varphi_k(x) f_T(x) dx \right)_{0 \leq j,k \leq m-1}$$

is invertible under Assumption 3.2. In addition, we impose that

$$L(m) := \sup_{x \in A} \sum_{j=0}^{m-1} \varphi_j(x)^2 \text{ and } R(m) := \sup_{x \in A} \sum_{j=0}^{m-1} \varphi_j'(x)^2$$

fulfill the following assumption.

**Assumption 3.3.** There exists  $\rho > 0$  and  $\kappa \geq 1$  such that  $R(m) \leq \rho L(m)^\kappa$  and

$$L(m)(\|\Psi(m)^{-1}\|_{\text{op}} \vee 1) \leq \frac{\mathbf{c}_{\kappa,T}}{2} \cdot \frac{NT}{\log(NT)} \quad \text{with} \quad \mathbf{c}_{\kappa,T} := \frac{3 \log(3/2) - 1}{(7 + \kappa)T}.$$

The above condition is a generalization of the so-called *stability condition* introduced for the standard regression by Cohen et al. [9, 10], also considered in Comte and Genon-Catalot [12].

**Convention.** When  $\mathbf{M}$  is a symmetric nonnegative and noninvertible matrix,  $\|\mathbf{M}^{-1}\|_{\text{op}} := \infty$ . This is a coherent convention because if  $\mathbf{M}$  is invertible, then  $\|\mathbf{M}^{-1}\|_{\text{op}} = 1/\inf\{\text{sp}(\mathbf{M})\}$ .

In order to ensure the existence of the oracle and to be able to bound its integrated risk,  $\widehat{b}_m$  is replaced by

$$\widetilde{b}_m := \widehat{b}_m \mathbf{1}_{\widehat{\Lambda}_\kappa(m)},$$

where

$$(5) \quad \widehat{\Lambda}_\kappa(m) := \left\{ L(m)(\|\widehat{\Psi}(m)^{-1}\|_{\text{op}} \vee 1) \leq \mathbf{c}_{\kappa,T} \frac{NT}{\log(NT)} \right\}.$$

Note that with the previous convention, on the event  $\widehat{\Lambda}_\kappa(m)$ ,  $\widehat{\Psi}(m)$  is invertible because

$$\inf\{\text{sp}(\widehat{\Psi}(m))\} \geq \frac{L(m)}{\mathbf{c}_{\kappa,T}} \cdot \frac{\log(NT)}{NT}.$$

Then,  $\widetilde{b}_m$  is well-defined. Moreover, necessarily,  $m \leq NT/\log(NT)$  on  $\widehat{\Lambda}_\kappa(m)$ .

The two following results provide controls of the empirical risk and of the  $f_T$ -weighted integrated risk of  $\tilde{b}_m$  respectively.

**Theorem 3.4.** *Under Assumptions 2.7, 3.2 and 3.3,*

$$\mathbb{E}(\|\tilde{b}_m - b_A\|_N^2) \leq \inf_{\tau \in \mathcal{S}_m} \|\tau - b_A\|_{f_T}^2 + \frac{2}{NT} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) + \mathfrak{c}_{\rho, \kappa, \sigma} (1 + \mathfrak{b}_T) \frac{\bar{\mathfrak{m}}_{2, H, M}(T)}{NT}$$

where  $\mathfrak{c}_{\rho, \kappa, \sigma} > 0$  is a deterministic constant depending only on  $\rho$ ,  $\kappa$  and  $\sigma$ ,  $\bar{\mathfrak{m}}_{2, H, M}(T)$  is a constant defined in Theorem 2.9,

$$\Psi(m, \sigma) := \left( \frac{\sigma^2}{T} \mathbb{E} \left( \left( \int_0^T \varphi_j(X(s)) \delta B(s) \right) \left( \int_0^T \varphi_k(X(s)) \delta B(s) \right) \right) \right)_{j, k=0, \dots, m-1}$$

and  $\mathfrak{b}_T := \|b_A^2\|_{f_T}$ .

**Corollary 3.5.** *Under Assumptions 2.7, 3.2 and 3.3,*

$$\begin{aligned} \mathbb{E}(\|\tilde{b}_m - b_A\|_{f_T}^2) &\leq \left( 1 + \frac{4T\mathfrak{c}_{\kappa, T}}{\log(NT)} \right) \inf_{\tau \in \mathcal{S}_m} \|\tau - b_A\|_{f_T}^2 \\ &\quad + \frac{8}{NT} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) + \bar{\mathfrak{c}}_{\rho, \kappa, \sigma} (1 + \mathfrak{b}_T) \frac{\bar{\mathfrak{m}}_{2, H, M}(T)}{NT} \end{aligned}$$

where  $\bar{\mathfrak{c}}_{\rho, \kappa, \sigma} > 0$  is a deterministic constant depending only on  $\rho$ ,  $\kappa$  and  $\sigma$ .

**Remark 3.6.** *Note that*

$$\begin{aligned} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) &= \text{trace}(\Psi(m)^{-1/2} \Psi(m, \sigma) \Psi(m)^{-1/2}) \leq m \|\Psi(m)^{-1/2} \Psi(m, \sigma) \Psi(m)^{-1/2}\|_{\text{op}} \\ &= m \sup_{\tau \in \mathcal{S}_m : \|\tau\|_{f_T} = 1} \mathbb{E} \left[ \left( \int_0^T \tau(X(s)) \delta B(s) \right)^2 \right]. \end{aligned}$$

The risk decompositions given in Theorem 3.4 and Corollary 3.5 both involve the same types of terms:

- The first one is equal or proportional to  $\inf_{\tau \in \mathcal{S}_m} \|\tau - b_A\|_{f_T}^2$  and is a squared bias term due to the projection strategy. It is decreasing when  $m$  increases, because then the projection space grows.
- The second one,  $\text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) / (NT)$ , is a variance term. From the remark above, it is bounded by  $m \|\Psi(m)^{-1} \Psi(m, \sigma)\|_{\text{op}} / (NT)$  which is increasing with  $m$ .
- The last one is a residual negligible term, which is small when  $N$  is large. Note that if the upper-bound  $M$  on  $b'$  is nonnegative, then  $\bar{\mathfrak{m}}_{2, H, M}(T)$  explodes for large values of  $T$ .

The order of the bias generally depends on the regularity of the function, and the order of the trace term is discussed below. Both quantities imply that a choice of  $m$  ensuring a compromise between the bias and the variance is required to obtain the convergence of  $\tilde{b}_m$  and a rate.

Finally, let us provide a control for  $\text{trace}(\Psi(m)^{-1} \Psi(m, \sigma))$  which allows comparison with non fractional results.

**Proposition 3.7.** *Under Assumptions 2.7 and 3.2,*

$$\begin{aligned} \frac{\text{trace}(\Psi(m)^{-1} \Psi(m, \sigma))}{NT} &\leq \mathfrak{c}_{2, H, \sigma} \sigma^2 \frac{\bar{\mathfrak{m}}_{2, H, M}(T)}{NT^{2-2H}} \\ &\quad \times \min\{(L(m) + R(m)) \|\Psi(m)^{-1}\|_{\text{op}}, m(1 + R(m) \|\Psi(m)^{-1}\|_{\text{op}})\}. \end{aligned}$$

In the standard case, with  $H = 1/2$  and a constant volatility function  $\sigma$ , it holds that

$$\frac{1}{NT} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) = \sigma^2 \frac{m}{NT}$$

as established in Comte and Genon-Catalot [12]. Here, for  $M < 0$ , the constant  $\bar{\mathfrak{m}}_{2, H, M}(T)$  does not depend on  $T$  and thus,  $NT$  becomes  $NT^{2-2H}$  which is coherent. However, the additional term  $R(m) \|\Psi(m)^{-1}\|_{\text{op}}$  may have an important order in  $m$  and substantially increases the variance. Thus, it will deteriorate the rate of the estimators. So, there is a *discontinuity* between the cases  $H = 1/2$  and

$H > 1/2$ , which is explained in Remark 2.10. However, note that this discontinuity is specific to the estimation strategy investigated in this paper.

**3.2. Rates in some usual bases.** Now, for projection estimators, different bases can be considered. In the present setting, the bases have to be differentiable. Let us present two examples.

**3.2.1. Rates on Fourier-Sobolev spaces for the trigonometric basis.** A first example is the compactly supported trigonometric basis. For  $A = [\ell, \mathbf{r}]$ , it is defined by

$$\begin{aligned}\varphi_0(x) &:= \frac{1}{\sqrt{\mathbf{r} - \ell}} \mathbf{1}_{[\ell, \mathbf{r}]}(x), \\ \varphi_{2j+1}(x) &:= \sqrt{\frac{2}{\mathbf{r} - \ell}} \cos\left(2\pi j \frac{x - \ell}{\mathbf{r} - \ell}\right) \mathbf{1}_{[\ell, \mathbf{r}]}(x) \text{ and} \\ \varphi_{2j}(x) &:= \sqrt{\frac{2}{\mathbf{r} - \ell}} \sin\left(2\pi j \frac{x - \ell}{\mathbf{r} - \ell}\right) \mathbf{1}_{[a, b]}(x)\end{aligned}$$

for every  $x \in \mathbb{R}$  and  $j \geq 1$ . This basis satisfies, for  $m$  odd and any  $x \in [\ell, \mathbf{r}]$ ,

$$\sum_{j=0}^{m-1} \varphi_j^2(x) = m \text{ and } \sup_{x \in [\ell, \mathbf{r}]} \sum_{j=0}^{m-1} \varphi_j'(x)^2 \leq \frac{(2\pi)^2}{(\mathbf{r} - \ell)^3} m^3.$$

So,

$$L(m) = m \text{ and } R(m) = \rho(\ell, \mathbf{r}) m^3$$

where  $\rho(\ell, \mathbf{r}) = (2\pi)^2 / (\mathbf{r} - \ell)^3$ .

In the Brownian setting, where  $H = 1/2$ , for a constant volatility function  $\sigma(x) \equiv \sigma$ , as recalled above, the variance term is  $\sigma^2 m / (NT)$  (see Comte and Genon-Catalot [12]). Here, if we assume that  $f_T$  is lower bounded on  $A$  by  $f_0 > 0$ , then  $\|\Psi(m)^{-1}\|_{\text{op}} \leq 1/f_0$  and the bound of Proposition 3.7 becomes

$$\frac{1}{NT} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) \leq \mathbf{c}_{2, H, \sigma} \sigma^2 \frac{\overline{\mathbf{m}}_{2, H, M}(T)}{NT^{2-2H}} \cdot \frac{m}{f_0} [1 + \rho(\ell, \mathbf{r}) m^2].$$

The additional term  $R(m) \|\Psi(m)^{-1}\|_{\text{op}}$  discussed after Proposition 3.7 is here of order  $m^3$ .

Now, let us evaluate the bias term. Consider  $\beta \in \mathbb{N}^*$  and the Sobolev space

$$W_2^\beta([\ell, \mathbf{r}]) := \left\{ \varphi : [\ell, \mathbf{r}] \rightarrow \mathbb{R} : \int_\ell^\mathbf{r} |\varphi^{(\beta)}(x)|^2 dx < \infty \right\}.$$

If  $b_A \in W_2^\beta([\ell, \mathbf{r}])$ , by DeVore and Lorentz [22], Theorem 2.3 p. 205, then there exists a deterministic constant  $\mathbf{c}_{\beta, \ell, \mathbf{r}} > 0$ , not depending on  $m$ , such that

$$\|p_{\mathcal{S}_m}^\perp(b_A) - b_A\|^2 \leq \mathbf{c}_{\beta, \ell, \mathbf{r}} m^{-2\beta},$$

where  $p_{\mathcal{S}_m}^\perp$  is the orthogonal projection from  $\mathbb{L}^2(A, dx)$  onto  $\mathcal{S}_m$ . If in addition  $f_T$  is upper bounded on  $A$  by  $f_1$ , then

$$\inf_{\tau \in \mathcal{S}_m} \|\tau - b_A\|_{f_T}^2 \leq f_1 \|p_{\mathcal{S}_m}^\perp(b_A) - b_A\|^2 \leq \mathbf{c}_{\beta, \ell, \mathbf{r}} f_1 m^{-2\beta}.$$

As a consequence, the inequality of Theorem 3.4 can be written

$$\mathbb{E}(\|\tilde{b}_m - b_A\|_N^2) \leq \mathbf{c}_{\beta, \ell, \mathbf{r}} f_1 m^{-2\beta} + \mathbf{c}_{2, H, \sigma} \sigma^2 \frac{\overline{\mathbf{m}}_{2, H, M}(T)}{NT^{2-2H}} \cdot \frac{m}{f_0} [1 + \rho(\ell, \mathbf{r}) m^2] + \mathbf{c}_{\rho, \kappa, \sigma} (1 + \mathbf{b}_T) \frac{\overline{\mathbf{m}}_{2, H, M}(T)}{NT}.$$

We obtain the following result.

**Proposition 3.8.** *Under Assumption 2.7, if  $f_0 \leq f_T(x) \leq f_1$  for every  $x \in [\ell, \mathbf{r}]$ ,  $b_A \in W_2^\beta([\ell, \mathbf{r}])$  and  $\tilde{b}_m$  is computed in the trigonometric basis on  $[\ell, \mathbf{r}]$ , then there exists a deterministic constant  $\mathbf{c}_{\beta, \ell, \mathbf{r}, f_0, f_1} > 0$ , not depending on  $N$  and  $T$ , such that with  $m = m_{\text{opt}} := [(NT^{2-2H})^{1/(2\beta+3)}]$ ,*

$$\mathbb{E}(\|\tilde{b}_{m_{\text{opt}}} - b_A\|_N^2) \leq \mathbf{c}_{\beta, \ell, \mathbf{r}, f_0, f_1} \overline{\mathbf{m}}_{2, H, M}(T) (NT^{2-2H})^{-2\beta/(2\beta+3)} + \mathbf{c}_{\rho, \kappa, \sigma} (1 + \mathbf{b}_T) \frac{\overline{\mathbf{m}}_{2, H, M}(T)}{NT}.$$

We obtain the convergence of the oracle with respect to the empirical risk for a fixed  $T$  and  $N \rightarrow \infty$ , and a rate of convergence which degrades from the rate  $N^{-2\beta/(2\beta+1)}$  found in Comte and Genon-Catalot [12] for  $H = 1/2$  and  $\sigma$  constant, to the rate  $N^{-2\beta/(2\beta+3)}$ .

The choice of  $m_{\text{opt}}$  above has the interest to provide a rate, but it is not possible in practice, as it depends on  $\beta$  which is unknown.

Finally, note that the function  $b : x \mapsto \mu x$  with  $\mu \in \mathbb{R}^*$  fulfills the conditions of Proposition 3.8. Indeed, since  $b' = \mu$ , the function  $b$  satisfies Assumption 2.7 with  $m = M = \mu$ , and  $b_A \in W_2^1([\ell, \mathbf{r}])$  for every  $\mathbf{r} > \ell$ . Moreover, since the solution of Equation (1) in this case is the fractional Ornstein-Uhlenbeck process, which is a Gaussian process, then for every  $\mathbf{r} > \ell$ , there exist  $f_0, f_1 > 0$  such that  $f_0 \leq f_T \leq f_1$ . In fact, under Assumption 2.7, thanks to Inequality (3),  $f_T$  is still upper-bounded for nonlinear drift functions.

**3.2.2. Discussion on the Hermite example.** The second example is the non-compactly supported Hermite basis. Here,  $A = \mathbb{R}$ , and the Hermite polynomial and the Hermite function of order  $j \geq 0$  are given by

$$(6) \quad H_j(x) := (-1)^j e^{x^2} \frac{d^j}{dx^j} (e^{-x^2}) \text{ and } h_j(x) := \mathbf{c}_j H_j(x) e^{-x^2/2}; \forall x \in \mathbb{R},$$

where  $\mathbf{c}_j = (2^j j! \sqrt{\pi})^{-1/2}$ .

The sequence  $(h_j)_{j \geq 0}$  is an orthonormal basis of  $\mathbb{L}^2(\mathbb{R}, dx)$ . By Abramowitz and Stegun [1], and In-dritz [27],

$$\|h_j\|_\infty \leq \Phi_0 \text{ with } \Phi_0 = 1/\pi^{1/4}.$$

So that, for  $\varphi_j = h_j$ ,  $L(m) \leq \Phi_0^2 m$ . In fact, we can prove that  $L(m) \leq K\sqrt{m}$  for a constant  $K > 0$  (see Comte and Lacour [14]). Moreover, as

$$h'_j(x) = \sqrt{\frac{j}{2}} h_{j-1}(x) - \frac{\sqrt{j+1}}{2} h_{j+1}(x),$$

we find

$$\sup_{x \in \mathbb{R}} \sum_{j=0}^{m-1} h'_j(x)^2 \leq 2Km^{3/2}.$$

Thus,  $R(m) \leq 2Km^{3/2}$ . Here, the bound of Proposition 3.7 becomes

$$\frac{1}{NT} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) \leq \mathbf{c}_{2,H,\sigma} \sigma^2 \frac{\overline{m}_{2,H,M}(T)}{NT^{2-2H}} K\sqrt{m}(1+2m) \|\Psi(m)^{-1}\|_{\text{op}},$$

where  $\mathbf{c} > 0$  is a universal constant.

This case is more complicated since  $f_T$  can no longer be assumed lower bounded on  $\mathbb{R}$ , otherwise it would not be integrable. Therefore, the order of the variance and specifically of  $\|\Psi(m)^{-1}\|_{\text{op}}$  is more difficult to evaluate in general contexts. What is known is that it is growing with  $m$  and with order larger than order  $\sqrt{m}$  (see [11]). However, we can still assume that  $f_T$  is upper-bounded by a constant  $f_1 > 0$ , and thus, we can evaluate the bias in a similar way as previously by considering Sobolev-Hermite spaces (see Bongioanni and Torrea [5] or Belomestny et al. [4]) and balls. The Sobolev-Hermite space with regularity  $s > 0$  is given by

$$(7) \quad W_H^s := \left\{ \theta \in \mathbb{L}^2(\mathbb{R}) : \sum_{k \geq 0} k^s a_k(\theta)^2 < \infty \right\},$$

where  $a_k(\theta) := \langle \theta, h_k \rangle$ ,  $k \in \mathbb{N}$ . The Sobolev-Hermite ball is given by

$$W_H^s(D) := \left\{ \theta \in \mathbb{L}^2(\mathbb{R}) : \sum_{k \geq 0} k^s a_k(\theta)^2 \leq D \right\}; D > 0.$$

Thus, if  $b$  belongs to  $W_H^s(D)$ , then we have

$$\|p_{\tilde{S}_m}^\perp(b) - b\|^2 \leq Dm^{-s}.$$

For details, and especially for regularity properties of functions in these spaces, we refer to Section 4.1 of Belomestny et al. [4].

**Proposition 3.9.** *Under Assumption 2.7, if  $f_T(x) \leq f_1$  for every  $x \in \mathbb{R}$ ,  $b \in W_H^s(D)$ ,  $\tilde{b}_m$  is computed in the Hermite basis on  $\mathbb{R}$ , and  $\|\Psi(m)^{-1}\|_{\text{op}}^2 \leq m^\kappa$  with  $\kappa \geq 1$ , then there exists a deterministic constant  $\mathbf{c}_{s,D,f_1} > 0$ , not depending on  $N$  and  $T$ , such that with  $m = m_{\text{opt}} := [(NT^{2-2H})^{1/(s+3/2+\kappa/2)}]$ ,*

$$\mathbb{E}(\|\tilde{b}_{m_{\text{opt}}} - b\|_N^2) \leq \mathbf{c}_{s,D,f_1} \bar{\mathbf{m}}_{2,H,M}(T) (NT^{2-2H})^{-2s/(2s+3+\kappa)} + \mathbf{c}_{\rho,\kappa,\sigma} (1 + \mathbf{b}_T) \frac{\bar{\mathbf{m}}_{2,H,M}(T)}{NT}.$$

The rate depends on the unknown  $\kappa$ , and we mention that if  $\|\Psi(m)^{-1}\|_{\text{op}}$  grows exponentially with  $m$ , then the rate will become logarithmic, except if the bias also decreases exponentially.

**3.3. An approximate estimator.** In this subsection, assume that the process  $X$  has been observed  $N$  times for two close initial conditions  $x_0$  and  $x_0 + \varepsilon$  with  $\varepsilon > 0$ , a situation which can occur in pharmacology. We first present a possible application context, and then build an approximate estimator.

**3.3.1. Application context.** Let us give details about the application field we have in mind. If  $t \mapsto X_{x_0}(t)$  denotes the concentration of a drug along time during its elimination by a patient with initial dose  $x_0 > 0$ ,  $t \mapsto X_{x_0+\varepsilon}(t)$  could be approximated by replicating the exact same protocol on the same patient, but with the initial dose  $x_0 + \varepsilon$  after the complete elimination of the previous dose.

This is an interesting perspective because differential equations driven by the fractional Brownian motion with  $H > 1/2$  are well adapted to model the concentration process in pharmacokinetics. Indeed, D'Argenio and Park [16] showed that the elimination process has both a deterministic and a random components. A natural way to take into account these two components is to add a stochastic noise in the linear differential equation which classically models the concentration. It has been studied in the Itô calculus framework by many authors (see e.g. Donnet and Samson [25]). However, as mentioned in Delattre and Lavielle [19], the extension of the deterministic concentration model as a diffusion process is not realistic on the biological side because its paths are too rough.

So, as mentioned in Marie [35], Section 5, a way to increase the regularity of the paths of the concentration process is to replace the Brownian motion by a fractional Brownian motion of Hurst index close to 1 as driving signal.

**3.3.2. Risk bounds on the approximate estimator.** Throughout this subsection,  $b$  fulfills the following reinforced assumption.

**Assumption 3.10.** *The function  $b$  belongs to  $\text{Lip}_b^2(\mathbb{R})$  and fulfills Assumption 2.7.*

Under Assumption 3.10, the following proposition allows to approximate the Skorokhod integral of the solution  $X$  to Equation (1) with respect to  $B$  if two paths of  $X$  can be observed with different but close initial conditions.

**Proposition 3.11.** *Under Assumption 3.10, for every  $\varphi \in \text{Lip}_b^1(\mathbb{R})$ ,  $\varepsilon > 0$  and  $t \in ]0, T]$ ,*

$$\left| \int_0^t \varphi(X_{x_0}(u)) \delta X_{x_0}(u) - S_\varphi(x_0, \varepsilon, t) \right| \leq \alpha_H \sigma^2 \frac{\|b''\|_\infty \|\varphi'\|_\infty}{2} \varepsilon t^{2H-1} \mathbf{m}_{H,M}(t)$$

where

$$\begin{aligned} S_\varphi(x_0, \varepsilon, t) &:= \int_0^t \varphi(X_{x_0}(u)) dX_{x_0}(u) \\ &\quad - \alpha_H \sigma^2 \int_0^t \int_0^u \varphi'(X_{x_0}(u)) \frac{X_{x_0+\varepsilon}(u) - X_{x_0}(u)}{X_{x_0+\varepsilon}(v) - X_{x_0}(v)} |u - v|^{2H-2} dv du, \end{aligned}$$

$X_x$  is the solution to Equation (1) with initial condition  $x \in \mathbb{R}$ , and

$$\mathbf{m}_{H,M}(t) := \frac{1}{M^2(2H-1)} \mathbf{1}_{M < 0} + \frac{t^2}{2H(2H+1)} \mathbf{1}_{M=0} + \frac{e^{2Mt}}{M^2(2H-1)} \mathbf{1}_{M > 0}.$$

Note that if  $M < 0$ , then Theorem 3.11 has been already established in Comte and Marie [15] (see Corollary 2.8).

By Proposition 3.11, for every  $j \in \mathbb{N}$  and  $i \in \{1, \dots, N\}$ ,

$$S_{\varphi_j}^i(x_0, \varepsilon, T) := \int_0^T \varphi_j(X_{x_0}^i(u)) dX_{x_0}^i(u) - \alpha_H \sigma^2 \int_0^T \int_0^u \varphi_j'(X_{x_0}^i(u)) \frac{X_{x_0+\varepsilon}^i(u) - X_{x_0}^i(u)}{X_{x_0+\varepsilon}^i(v) - X_{x_0}^i(v)} |u - v|^{2H-2} dv du$$

provides a good approximation of

$$\int_0^T \varphi_j(X_{x_0}^i(s)) \delta X_{x_0}^i(s).$$

This legitimates to consider the approximate estimator

$$\widehat{b}_{m,\varepsilon} := \sum_{j=0}^{m-1} \widehat{\theta}_{j,\varepsilon} \varphi_j,$$

where

$$\widehat{\theta}(m, \varepsilon) := (\widehat{\theta}_{0,\varepsilon}, \dots, \widehat{\theta}_{m-1,\varepsilon})^* = \widehat{\Psi}(m)^{-1} \widehat{\mathbf{x}}(m, \varepsilon)$$

and

$$\widehat{\mathbf{x}}(m, \varepsilon) := \left( \frac{1}{NT} \sum_{i=1}^N S_{\varphi_j}^i(x_0, \varepsilon, T) \right)_{j=0, \dots, m-1}^*.$$

Contrary to the oracle  $\widehat{b}_m$ ,  $\widehat{b}_{m,\varepsilon}$  is computable.

**Remark 3.12.** For any  $i \in \{1, \dots, N\}$ , since  $\mathcal{I}(\cdot, B^i)$  is locally Lipschitz continuous on  $\mathbb{R}$  and  $X_{x_0}^i$  is  $\alpha$ -Hölder continuous on  $[0, T]$  with  $\alpha \in ]1/2, H[$ , there exists a square integrable random variable  $C : \Omega \rightarrow ]0, \infty[$  such that for any  $\eta, \varepsilon \in ]0, 1[$ ,

$$|X_{x_0}^i(t + \eta) - X_{x_0+\varepsilon}^i(t)| \leq |X_{x_0}^i(t + \eta) - X_{x_0}^i(t)| + |X_{x_0}^i(t) - X_{x_0+\varepsilon}^i(t)| \leq C(\eta^\alpha + \varepsilon).$$

By taking  $\eta = \varepsilon^{1/\alpha}$ ,

$$|X_{x_0}^i(t + \varepsilon^{1/\alpha}) - X_{x_0+\varepsilon}^i(t)| \leq 2C\varepsilon.$$

So, despite the lack of information available on the behavior of the quotient involved in  $S_{\varphi_j}^i(x_0, \varepsilon, T)$ , one could replace it by

$$\widetilde{S}_{\varphi_j}^i(x_0, \varepsilon, T) := \int_0^T \varphi_j(X_{x_0}^i(u)) dX_{x_0}^i(u) - \alpha_H \sigma^2 \int_0^T \int_0^u \varphi_j'(X_{x_0}^i(u)) \frac{X_{x_0}^i(u + \varepsilon^{1/\alpha}) - X_{x_0}^i(u)}{X_{x_0}^i(v + \varepsilon^{1/\alpha}) - X_{x_0}^i(v)} |u - v|^{2H-2} dv du$$

in the expression of  $\widehat{\theta}(m, \varepsilon)$ . This would avoid the requirement of the paths  $X_{x_0+\varepsilon}^i$  and  $X_{x_0}^i$  for each individual  $i$ .

Thus, to be coherent and realistic, we consider the estimator modified as follows:

$$\widetilde{b}_{m,\varepsilon} := \widehat{b}_{m,\varepsilon} \mathbf{1}_{\widehat{\Lambda}_\kappa(m)}.$$

Then, we can prove the following result as a consequence of Corollary 3.5.

**Corollary 3.13.** Under Assumptions 3.10, 3.2 and 3.3,

$$\begin{aligned} \mathbb{E}(\|\widetilde{b}_{m,\varepsilon} - b_A\|_{f_T}^2) &\leq \mathfrak{d}_{\kappa, NT} \inf_{\tau \in \mathcal{S}_m} \|\tau - b_A\|_{f_T}^2 + \frac{16}{NT} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) + \bar{\mathfrak{c}}_{\rho, \kappa, \sigma} (1 + \mathfrak{b}_T) \frac{\overline{\mathfrak{m}}_{2,H,M}(T)}{NT} \\ &\quad + \mathfrak{c}_{\rho, \kappa, \sigma, H, \mathfrak{b}''} L(m)^{\kappa-1} \left( \frac{NT}{\log(NT)} \right)^2 \frac{\mathfrak{m}_{H,M}(T)^2}{T^{4-4H}} \varepsilon^2 \end{aligned}$$

where  $\mathfrak{c}_{\rho, \kappa, \sigma, H, \mathfrak{b}''} > 0$  is a deterministic constant depending only on  $\rho, \kappa, \sigma, H$  and  $\mathfrak{b}'' := \|b''\|_\infty^2$ , and

$$\mathfrak{d}_{\kappa, NT} := 2 \left( 1 + 4 \frac{3 \log(3/2) - 1}{(7 + \kappa) \log(NT)} \right).$$

In order to keep the rate of convergence obtained for  $\tilde{b}_m$ , one can assume that  $\varepsilon$  depends on  $N$  and  $T$ , and take

$$\varepsilon_{N, T} = \frac{1}{(NT)^{1/2}} \left( \frac{NT}{\log(NT)} \right)^{-1}.$$

#### 4. AN ALTERNATIVE ESTIMATOR

The cornerstone of the proof of Proposition 3.11 is that, for every  $t \in [0, T]$ ,

$$Y_x(t) = \int_0^t b'(X_x(s)) ds \text{ with } Y_x(t) := \log(\partial_x X_x(t)).$$

By assuming in the first place that  $X_x$  and  $\partial_x X_x$  can be observed  $N$  times on  $[0, T]$ , the previous equality suggests the noiseless model

$$(8) \quad Y_x^i(T) = \int_0^T b'(X_x^i(s)) ds ; i \in \{1, \dots, N\},$$

where  $X_x^1, \dots, X_x^N$  (resp.  $Y_x^1, \dots, Y_x^N$ ) are  $N$  independent copies of  $X_x$  (resp.  $Y_x$ ).

Consider the objective function  $\gamma_N^\dagger$  defined by

$$\gamma_N^\dagger(\tau) := \frac{1}{NT} \sum_{i=1}^N \left( \int_0^T \tau(X_x^i(s))^2 ds - 2 \int_0^T \tau(X_x^i(s)) dY_x^i(s) \right)$$

for every function  $\tau : \mathbb{R} \rightarrow \mathbb{R}$ .

Note that for any bounded function  $\tau$  from  $\mathbb{R}$  into itself,

$$\begin{aligned} \mathbb{E}(\gamma_N^\dagger(\tau)) &= \frac{1}{T} \int_0^T \mathbb{E}(\tau(X_x(s))^2) ds - \frac{2}{T} \int_0^T \mathbb{E}(\tau(X_x(s)) b'(X_x(s))) ds \\ &= \frac{1}{T} \int_0^T \mathbb{E}((\tau(X_x(s)) - b'(X_x(s)))^2) ds - \frac{1}{T} \int_0^T \mathbb{E}(b'(X_x(s))^2) ds. \end{aligned}$$

Then, the definition of  $f_T$  gives

$$(9) \quad \mathbb{E}(\gamma_N^\dagger(\tau)) = \int_{-\infty}^{\infty} (\tau(x) - b'(x))^2 f_T(x) dx - \int_{-\infty}^{\infty} b'(x)^2 f_T(x) dx.$$

Equality (9) shows that  $\mathbb{E}(\gamma_N^\dagger(\tau))$  is the smallest for  $\tau$  the nearest of  $b'$ . Therefore, to minimize its empirical version  $\gamma_N^\dagger(\tau)$  should provide a functional estimator near of  $b'$ .

Let

$$\hat{b}_m^{\dagger, \prime} = \arg \min_{\tau \in \mathcal{S}_m} \gamma_N^\dagger(\tau)$$

be the projection estimator of  $b'_A := b'|_A$  on  $\mathcal{S}_m$ . Then,

$$\hat{b}_m^{\dagger, \prime} = \sum_{j=0}^{m-1} \hat{\theta}_j' \varphi_j$$

where

$$\hat{\theta}'(m) := (\hat{\theta}_0', \dots, \hat{\theta}_{m-1}')^* = \hat{\Psi}(m)^{-1} \hat{\mathbf{y}}(m)$$



with

$$\widehat{\mathbf{y}}(m) := \left( \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X_x^i(s)) dY_x^i(s) \right)_{j=0, \dots, m-1}^* = (\langle \varphi_j, b' \rangle_N)_{j=0, \dots, m-1}^*.$$

As for  $\widehat{b}_m$ , in order to ensure the existence and the stability of the estimator,  $\widehat{b}_m^{\dagger, \prime}$  is replaced by

$$\widetilde{b}_m^{\dagger, \prime} := \widehat{b}_m^{\dagger, \prime} \mathbf{1}_{\widehat{\Lambda}_0(m)},$$

where  $\widehat{\Lambda}_0(m)$  is defined by (5) with  $\kappa = 0$ , and  $L(m)$  fulfills the following assumption. In the continuous-time processes framework, this estimator is similar to the least square estimator for the noiseless regression model of Cohen et al. [9].

**Assumption 4.1.** *The quantity  $L(m)$  satisfies*

$$L(m)(\|\Psi(m)^{-1}\|_{\text{op}} \vee 1) \leq \frac{\mathfrak{c}_{0,T}}{2} \cdot \frac{NT}{\log(NT)} \quad \text{with} \quad \mathfrak{c}_{0,T} = \frac{3 \log(3/2) - 1}{7T}.$$

The following result provides controls of the empirical risk and of the  $f_T$ -weighted integrated risk of  $\widetilde{b}_m^{\dagger, \prime}$  respectively.

**Proposition 4.2.** *Under Assumptions 2.7, 3.2 and 4.1,*

$$(10) \quad \mathbb{E}(\|\widetilde{b}_m^{\dagger, \prime} - b'_A\|_N^2) \leq \inf_{\tau \in \mathcal{S}_m} \|\tau - b'_A\|_{f_T}^2 + \frac{\mathfrak{c}_{4,2}}{NT}$$

where  $\mathfrak{c}_{4,2} := \mathfrak{c}_\Omega^{1/2} \|b'\|_\infty^2$ . In addition,

$$(11) \quad \mathbb{E}(\|\widetilde{b}_m^{\dagger, \prime} - b'_A\|_{f_T}^2) \leq \left(1 + \frac{2T \mathfrak{c}_{0,T}}{\log(NT)}\right) \inf_{\tau \in \mathcal{S}_m} \|\tau - b'_A\|_{f_T}^2 + \bar{\mathfrak{c}}_{b'} \frac{1 + \mathfrak{c}_{0,T}}{NT}$$

where  $\bar{\mathfrak{c}}_{b'} > 0$  is a deterministic constant depending only on  $\mathfrak{b}' := \|b'\|_\infty$ .

A consequence of Proposition 4.2 is that the estimator can reach a parametric rate and a risk of order  $1/(NT)$ . No model selection step is required, one has only to choose the largest  $m$  of the collection. In practice, this leads to a good but non parsimonious estimator. A theoretical stopping rule leading to keep only an adequate number of projection coefficients would be interesting but is not available yet.

Now, assume that the process  $X$  has been observed  $N$  times for two close initial conditions  $x_0$  and  $x_0 + \varepsilon$  with  $\varepsilon > 0$ . Let us consider the estimator

$$\widehat{b}_{m,\varepsilon}^{\dagger, \prime} := \sum_{j=0}^{m-1} \widehat{\theta}_{j,\varepsilon}' \varphi_j$$

with

$$\widehat{\theta}'(m, \varepsilon) := (\widehat{\theta}_{0,\varepsilon}', \dots, \widehat{\theta}_{m-1,\varepsilon}') = \widehat{\Psi}(m)^{-1} \widehat{\mathbf{y}}(m, \varepsilon)$$

and

$$\widehat{\mathbf{y}}(m, \varepsilon) := \left( \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X_{x_0}^i(s)) dY_{x_0,\varepsilon}^i(s) \right)_{j=0, \dots, m-1}^*,$$

where  $Y_{x_0,\varepsilon}^1, \dots, Y_{x_0,\varepsilon}^N$  are  $N$  independent copies of the process  $Y_{x_0,\varepsilon}$  defined by

$$Y_{x_0,\varepsilon}(t) := \log \left( \frac{X_{x_0+\varepsilon}(t) - X_{x_0}(t)}{\varepsilon} \right); \quad \forall t \in [0, T].$$

**Corollary 4.3.** *Under Assumptions 3.10, 3.2 and 4.1,*

$$\begin{aligned} \mathbb{E}(\|\widetilde{b}_{m,\varepsilon}^{\dagger, \prime} - b'_A\|_{f_T}^2) &\leq \mathfrak{d}_{0,NT} \inf_{\tau \in \mathcal{S}_m} \|\tau - b'_A\|_{f_T}^2 + \bar{\mathfrak{c}}_{b'} \frac{1 + \mathfrak{c}_{0,T}}{NT} \\ &\quad + \frac{\|b''\|_\infty^2}{2} \mathfrak{c}_{0,T}^2 e^{2MT} \frac{1}{L(m)} \left( \frac{NT}{\log(NT)} \right)^2 \varepsilon^2. \end{aligned}$$

Finally, we have obtained a good estimator of  $b'_A$ , but there remains difficulties to deduce an estimator of  $b_A$ . Assume that  $A = [\mathbf{r}, \ell]$  with  $\mathbf{r}, \ell \in \mathbb{R}$  and  $\mathbf{r} < \ell$ , and that there exists  $f_0 > 0$  such that  $f_T(x) \geq f_0$  for every  $x \in A$ . For every  $x \in A$ , consider the estimator

$$\widehat{b}_{m,\varepsilon}^\dagger(x) := \widehat{b}_A(\mathbf{r}) + \int_{\mathbf{r}}^x \widehat{b}_{m,\varepsilon}^{\dagger,\prime}(y) dy$$

of

$$b_A(x) := b_A(\mathbf{r}) + \int_{\mathbf{r}}^x b'_A(y) dy ; x \in A.$$

Then,

$$\begin{aligned} \|\widehat{b}_{m,\varepsilon}^\dagger - b_A\|_{f_T}^2 &= \int_{\mathbf{r}}^{\ell} \left[ (\widehat{b}_A(\mathbf{r}) - b_A(\mathbf{r})) + \int_{\mathbf{r}}^x (\widehat{b}_{m,\varepsilon}^{\dagger,\prime}(y) - b'_A(y)) dy \right]^2 f_T(x) dx \\ &\leq 2(\widehat{b}_A(\mathbf{r}) - b_A(\mathbf{r}))^2 + 2(\ell - \mathbf{r}) \left( \int_{\mathbf{r}}^{\ell} f_T(x) dx \right) \left( \int_{\mathbf{r}}^{\ell} |\widehat{b}_{m,\varepsilon}^{\dagger,\prime}(y) - b'_A(y)|^2 dy \right) \\ &\leq 2(\widehat{b}_A(\mathbf{r}) - b_A(\mathbf{r}))^2 + 2 \frac{\ell - \mathbf{r}}{f_0} \|\widehat{b}_{m,\varepsilon}^{\dagger,\prime} - b'_A\|_{f_T}^2. \end{aligned}$$

So, thanks to Corollary 4.3,

$$\begin{aligned} \mathbb{E}(\|\widehat{b}_{m,\varepsilon}^\dagger - b_A\|_{f_T}^2) &\leq 2(\widehat{b}_A(\mathbf{r}) - b_A(\mathbf{r}))^2 + 2 \frac{\ell - \mathbf{r}}{f_0} \left[ \mathfrak{d}_{0,NT} \inf_{\tau \in \mathcal{S}_m} \|\tau - b'_A\|_{f_T}^2 + \mathfrak{c}_{b'} \frac{1 + \mathfrak{c}_{0,T}}{NT} \right. \\ &\quad \left. + \frac{\|b''\|_{\infty}^2}{2} \mathfrak{c}_{0,T}^2 e^{2MT} \frac{1}{L(m)} \left( \frac{NT}{\log(NT)} \right)^2 \varepsilon^2 \right]. \end{aligned}$$

Clearly,  $\widehat{b}_{m,\varepsilon}^\dagger$  keeps a good rate of convergence only if the function  $b_A$  is known at one point, or if  $b_A$  can be estimated at one point with a good rate. No pointwise procedure which would grant to keep the parametric feature is available. Note also that, even if  $b_A(\mathbf{r})$  is known, another drawback of  $\widehat{b}_{m,\varepsilon}^\dagger$  is that, since it is obtained by integrating an estimator of  $b'_A$ , there are restrictive conditions:  $A$  must be a compact interval with  $f_T \geq f_0 > 0$  on  $A$ .

## 5. CONCLUDING REMARKS

In this paper, we propose two estimation strategies for nonparametric reconstruction of the drift function  $b$  from  $N$  i.i.d. continuous observations drawn in the fractional SDE given by (1). On the one hand, we bound the empirical and  $f_T$ -weighted  $\mathbb{L}^2$ -risks of the auxiliary oracle  $\widehat{b}_m$ , and then of the approximate estimator  $\widehat{b}_{m,\varepsilon}$ . We compare the result with non fractional case. In the case of a specific trigonometric basis, we can, under additional assumptions, prove the consistency of the estimator and evaluate its rate of convergence. However, the choice of  $m_{\text{opt}}$  which is proposed to obtain this result is not possible in practice, as it depends on unknown parameters. Therefore, a model selection step in the spirit of Comte and Genon-Catalot [12] would have to be settled. On the other hand, we bound the empirical and  $f_T$ -weighted  $\mathbb{L}^2$ -risks of the alternative calculable estimator  $\widehat{b}_{m,\varepsilon}^{\dagger,\prime}$ , get a parametric rate of convergence, and show that we can not deduce a  $f_T$ -weighted  $\mathbb{L}^2$ -risk bound on the primitive estimator  $\widehat{b}_{m,\varepsilon}^\dagger$ , except if the function is known at one point. Even so, there will be the restrictive conditions requiring that  $A$  is a compact interval and that  $f_T$  is lower-bounded by a strictly positive constant on  $A$ . For this reason,  $\widehat{b}_{m,\varepsilon}$  is adequate to estimate  $b$  and the other strategy to estimate its derivative, which has interest also in many applications.

For both estimators  $\widehat{b}_{m,\varepsilon}$  and  $\widehat{b}_{m,\varepsilon}^\dagger$ , two paths with slightly different initial conditions have to be available for each individual  $i$ , but this context may be realistic in pharmacokinetics experiments. We also propose another idea of transformation of  $\widehat{b}_m$  into a computable version, which seems intuitive and does not require two paths per individual, but would require a deeper study (see Remark 3.12).

Lastly, the tedious question of discretization may be investigated in the future, to take into account the fact that, for each individual  $i$ , the observation is  $(X_i(k\Delta))_{1 \leq k \leq n}$  where  $n\Delta = T$ .

## 6. PROOFS

6.1. **Proof of Theorem 2.9.** On the one hand, for any  $s, t \in [0, T]$ ,

$$\mathbf{D}_s X(t) = \sigma \mathbf{1}_{[0,t]}(s) \exp \left( \int_s^t b'(X(u)) du \right)$$

and, by Assumption 2.7,

$$|\mathbf{D}_s X(t)| \leq \sigma \mathbf{1}_{[0,t]}(s) e^{M(t-s)}.$$

Then, by the chain rule for Malliavin's derivative and Jensen's inequality,

$$\begin{aligned} \mathbb{E} \left( \int_0^T \int_0^T |\mathbf{D}_s[\varphi(X(t))]|^{1/H} ds dt \right)^{pH} &= \mathbb{E} \left( \int_0^T \int_0^T |\varphi'(X(t)) \mathbf{D}_s X(t)|^{1/H} ds dt \right)^{pH} \\ &\leq \sigma^p \left( \int_0^T \mathbb{E}(|\varphi'(X(t))|^{1/H}) \int_0^t e^{M/H(t-s)} ds dt \right)^{pH} \\ (12) \quad &\leq \sigma^p \mathfrak{m}_{p,H,M}(T) \left( \int_0^T \mathbb{E}(|\varphi'(X(t))|^p)^{1/(pH)} dt \right)^{pH} \end{aligned}$$

where

$$\mathfrak{m}_{p,H,M}(T) = \left( -\frac{H}{M} \right)^{pH} \mathbf{1}_{M < 0} + T^{pH} \mathbf{1}_{M=0} + \left( \frac{H}{M} \right)^{pH} e^{pMT} \mathbf{1}_{M > 0}.$$

On the other hand, by Hu et al. [31], Lemma 3.1, there exists a deterministic constant  $\mathfrak{c}_{p,H} > 0$ , depending only on  $p$  and  $H$ , such that for any  $\varphi \in \text{Lip}_b^1(\mathbb{R})$ ,

$$\begin{aligned} \mathbb{E} \left( \left| \int_0^T \varphi(X(t)) \delta B(t) \right|^p \right) &\leq \mathfrak{c}_{p,H} \left[ \left( \int_0^T \mathbb{E}(|\varphi(X(t))|^{1/H}) dt \right)^{pH} \right. \\ (13) \quad &\left. + \mathbb{E} \left( \int_0^T \int_0^T |\mathbf{D}_s[\varphi(X(t))]|^{1/H} ds dt \right)^{pH} \right]. \end{aligned}$$

Inequalities (12) and (13) together allow to conclude.

6.2. **Proof of Proposition 2.11.** Consider  $t \in ]0, T]$ . On the one hand, for any  $s \in \mathbb{R}_+$ ,

$$\mathbf{D}_s X(t) = \sigma \mathbf{1}_{[0,t]}(s) \exp \left( \int_s^t b'(X(u)) du \right).$$

So, since  $b'$  is a  $[m, M]$ -valued function,

$$\sigma \mathbf{1}_{[0,t]}(s) e^{m(t-s)} \leq \mathbf{D}_s X(t) \leq \sigma \mathbf{1}_{[0,t]}(s) e^{M(t-s)}.$$

On the other hand, by Nourdin and Viens [41], Proposition 3.7,

$$-\mathbf{D}_s \mathbf{L}^{-1} X(t) = \int_0^\infty e^{-u} \mathbf{T}_u(\mathbf{D}_s X(t)) du$$

where  $(\mathbf{T}_u)_{u \in \mathbb{R}_+}$  is the Ornstein-Uhlenbeck semigroup (see Nualart [42], Section 1.4). Moreover, for any  $u \in \mathbb{R}_+$ ,  $(\mathbf{T}_u)|_{\mathbb{R}} = \text{Id}_{\mathbb{R}}$  by Mehler's formula (see Nualart [42], Equation (1.67)), and for every  $U_1, U_2 \in \mathbb{L}^2(\Omega, \mathbb{P})$ ,

$$U_1 \geq U_2 \implies \mathbf{T}_u(U_1) \geq \mathbf{T}_u(U_2)$$

by Nualart [42], Property (i) page 55. Then,

$$\sigma \mathbf{1}_{[0,t]}(s) e^{m(t-s)} \leq -\mathbf{D}_s \mathbf{L}^{-1} X(t) \leq \sigma \mathbf{1}_{[0,t]}(s) e^{M(t-s)}.$$

Therefore,

$$\sigma(m, t)^2 \leq g_t^*(x_0, X^*(t)) \leq \sigma(M, t)^2$$

where

$$\sigma(\mu, t)^2 = \alpha_H \sigma^2 \int_0^t \int_0^t |v - u|^{2H-2} e^{\mu(2t-v-u)} du dv > 0 ; \forall \mu \in \mathbb{R}.$$

Nourdin and Viens [41], Theorem 3.1 and Corollary 3.5 allow to conclude.

**6.3. Proof of Theorem 3.4.** The proof relies on two lemmas which are stated first.

**Lemma 6.1.** *There exists a deterministic constant  $\mathbf{c}_e > 0$ , not depending on  $m, N$  and  $T$ , such that*

$$\mathbb{E}(|\mathbf{e}(m)^* \mathbf{e}(m)|^2) \leq \mathbf{c}_e \sigma^4 \frac{\overline{\mathbf{m}}_{4,H,M}(T)}{N^2 T^{4-4H}} m(L(m)^2 + R(m)^2).$$

**Lemma 6.2.** *Consider*

$$\Omega(m) := \left\{ \sup_{\tau \in \mathcal{S}_m} \left| \frac{\|\tau\|_N^2}{\|\tau\|_{f_T}^2} - 1 \right| \leq \frac{1}{2} \right\}.$$

*Under Assumptions 2.7, 3.2 and 3.3, there exists a deterministic constant  $\mathbf{c}_\Omega > 0$ , not depending on  $m, N$  and  $T$ , such that*

$$\mathbb{P}(\widehat{\Lambda}_\kappa(m)^c) \leq \frac{\mathbf{c}_\Omega}{(NT)^{6+\kappa}} \text{ and } \mathbb{P}(\Omega(m)^c) \leq \frac{\mathbf{c}_\Omega}{(NT)^{6+\kappa}}.$$

**6.3.1. Steps of the proof.** First of all,

$$(14) \quad \|\widetilde{b}_m - b_A\|_N^2 = \|\widehat{b}_m - b_A\|_N^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m)} + \|b_A\|_N^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m)^c} = U_1 + U_2 + U_3$$

where

$$\begin{aligned} U_1 &:= \|b_A\|_N^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m)^c}, \\ U_2 &:= \|\widehat{b}_m - b_A\|_N^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m) \cap \Omega(m)} \text{ and} \\ U_3 &:= \|\widehat{b}_m - b_A\|_N^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m) \cap \Omega(m)^c}. \end{aligned}$$

Let us find suitable bounds on  $\mathbb{E}(U_1)$ ,  $\mathbb{E}(U_2)$  and  $\mathbb{E}(U_3)$ .

- **Bound for  $\mathbb{E}(U_1)$ .** By Lemma 6.2,

$$\mathbb{E}(U_1) \leq \mathbb{E}(\|b_A\|_N^4)^{1/2} \mathbb{P}(\widehat{\Lambda}_\kappa(m)^c)^{1/2} \leq \frac{\mathbf{c}_1}{NT}$$

where

$$\mathbf{c}_1 := \mathbf{c}_\Omega^{1/2} \left( \int_{-\infty}^{\infty} b_A(x)^4 f_T(x) dx \right)^{1/2} = \mathbf{c}_\Omega^{1/2} \|b_A\|_{f_T}^2.$$

- **Bound for  $\mathbb{E}(U_2)$ .** By denoting  $p_{N, \mathcal{S}_m}^\perp$  the orthogonal projection from  $\mathbb{L}^2(A, f_T(x)dx)$  onto  $\mathcal{S}_m$  with respect to the empirical scalar product  $\langle \cdot, \cdot \rangle_N$ ,

$$(15) \quad \|\widehat{b}_m - b_A\|_N^2 = \|\widehat{b}_m - p_{N, \mathcal{S}_m}^\perp(b_A)\|_N^2 + \inf_{\tau \in \mathcal{S}_m} \|b_A - \tau\|_N^2.$$

As in the proof of Comte and Genon-Catalot [12], Proposition 2.1, on  $\Omega(m)$ ,

$$\|\widehat{b}_m - p_{N, \mathcal{S}_m}^\perp(b_A)\|_N^2 = \mathbf{e}(m)^* \widehat{\Psi}(m)^{-1} \mathbf{e}(m) \leq 2\mathbf{e}(m)^* \Psi(m)^{-1} \mathbf{e}(m).$$

So,

$$\begin{aligned} \mathbb{E}(\|\widehat{b}_m - p_{N, \mathcal{S}_m}^\perp(b_A)\|_N^2 \mathbf{1}_{\Omega(m) \cap \widehat{\Lambda}_\kappa(m)}) &\leq 2\mathbb{E} \left( \sum_{j,k=0}^{m-1} \mathbf{e}(m)_j \mathbf{e}(m)_k \Psi(m)_{j,k}^{-1} \right) \\ &= \frac{2\sigma^2}{NT^2} \sum_{j,k=0}^{m-1} \Psi(m)_{j,k}^{-1} \\ &\quad \times \mathbb{E} \left( \left( \int_0^T \varphi_j(X(s)) \delta B(s) \right) \left( \int_0^T \varphi_k(X(s)) \delta B(s) \right) \right) \\ &= \frac{2}{NT} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)). \end{aligned}$$

Therefore, by Inequality (15),

$$\mathbb{E}(U_2) \leq \inf_{\tau \in \mathcal{S}_m} \|b_A - \tau\|_{f_T}^2 + \frac{2}{NT} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)).$$

- **Bound for  $\mathbb{E}(U_3)$ .** On the one hand,

$$\|\widehat{b}_m - p_{N, \mathcal{S}_m}^\perp(b_A)\|_N^2 = \mathbf{e}(m)^* \widehat{\Psi}(m)^{-1} \mathbf{e}(m).$$

On the other hand, on  $\widehat{\Lambda}_\kappa(m)$ , for  $N, T$  large enough,

$$\begin{aligned} (L(m) + R(m)) \|\widehat{\Psi}(m)^{-1}\|_{\text{op}} &\leq L(m) (\|\widehat{\Psi}(m)^{-1}\|_{\text{op}} \vee 1) + \rho L(m)^\kappa (\|\widehat{\Psi}(m)^{-1}\|_{\text{op}} \vee 1)^\kappa \\ &\leq (1 + \rho) \mathbf{c}_{\kappa, T} \left( \frac{NT}{\log(NT)} \right)^\kappa. \end{aligned}$$

Then, by Lemmas 6.1 and 6.2, there exists a deterministic constant  $\mathbf{c}_2 > 0$ , not depending on  $m$  (satisfying  $m \leq NT$ ),  $N$  and  $T$ , such that

$$\begin{aligned} \mathbb{E}(U_3) &\leq \mathbb{E}(\|\widehat{b}_m - p_{N, \mathcal{S}_m}^\perp(b_A)\|_N^2 + \|b_A\|_N^2) \mathbf{1}_{\widehat{\Lambda}_\kappa(m) \cap \Omega(m)^c} \\ &\leq \left[ \frac{(1 + \rho) \mathbf{c}_{\kappa, T}}{L(m) + R(m)} \left( \frac{NT}{\log(NT)} \right)^\kappa \mathbb{E}(|\mathbf{e}(m)^* \mathbf{e}(m)|^2)^{1/2} + \mathbb{E}(\|b_A\|_N^4)^{1/2} \right] \mathbb{P}(\Omega(m)^c)^{1/2} \\ &\leq \mathbf{c}_2 \frac{\overline{\mathbf{m}}_{2, H, M}(T)}{NT}. \end{aligned}$$

These bounds together with Inequality (14) allow to conclude.

6.3.2. *Proof of Lemma 6.1.* On the one hand, for any bounded and measurable function  $\tau : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$(16) \quad \int_0^T \mathbb{E}(\tau(X(s)) ds = \int_{-\infty}^\infty \int_0^T \tau(x) p_s(x_0, x) ds dx = T \int_{-\infty}^\infty \tau(x) f_T(x) dx$$

and, by Jensen's inequality, for every  $p > 1/H$ ,

$$\begin{aligned} \int_0^T \mathbb{E}(\tau(X(s)))^{1/(pH)} ds &= T \int_0^T \left( \int_{-\infty}^\infty \tau(x) p_s(x_0, x) dx \right)^{1/(pH)} \frac{ds}{T} \\ &\leq T \left( \int_{-\infty}^\infty \int_0^T \tau(x) p_s(x_0, x) \frac{ds}{T} dx \right)^{1/(pH)} \\ (17) \quad &= T \left( \int_{-\infty}^\infty \tau(x) f_T(x) dx \right)^{1/(pH)}. \end{aligned}$$

On the other hand, by Jensen's inequality, since  $(B_1, X_1), \dots, (B_N, X_N)$  are i.i.d and by Equality (2),

$$\begin{aligned} \mathbb{E}(|\mathbf{e}(m)^* \mathbf{e}(m)|^2) &= \mathbb{E} \left( \left| \sum_{j=0}^{m-1} \mathbf{e}(m)_j^2 \right|^2 \right) \\ &\leq \frac{\sigma^4 m}{(NT)^4} \sum_{j=0}^{m-1} \mathbb{E} \left[ \left( \sum_{i=1}^N \int_0^T \varphi_j(X^i(s)) \delta B^i(s) \right)^4 \right] \\ &\leq \frac{\sigma^4 m}{N^3 T^4} \sum_{j=0}^{m-1} \mathbb{E}(Y_j^4) + \frac{\sigma^4 m}{N^2 T^4} \sum_{j=0}^{m-1} \mathbb{E}(Y_j^2)^2 \end{aligned}$$

where

$$Y_j := \int_0^T \varphi_j(X(s)) \delta B(s) ; \forall j \in \{0, \dots, m-1\}.$$

Moreover, by Theorem 2.9, Equality (16) and Inequality (17), for  $j = 0, \dots, m-1$ ,

$$\begin{aligned} \mathbb{E}(Y_j^4) &\leq \mathbf{c}_{4,H,\sigma} \bar{\mathbf{m}}_{4,H,M}(T) \left[ \left( \int_0^T \mathbb{E}(|\varphi_j(X(s))|^{1/H}) ds \right)^{4H} + \left( \int_0^T \mathbb{E}(|\varphi'_j(X(s))|^4)^{1/(4H)} ds \right)^{4H} \right] \\ &\leq \mathbf{c}_{4,H,\sigma} \bar{\mathbf{m}}_{4,H,M}(T) T^{4H} \left[ \left( \int_{-\infty}^{\infty} |\varphi_j(x)|^{1/H} f_T(x) dx \right)^{4H} + \int_{-\infty}^{\infty} |\varphi'_j(x)|^4 f_T(x) dx \right] \\ &\leq \mathbf{c}_{4,H,\sigma} \bar{\mathbf{m}}_{4,H,M}(T) T^{4H} \sup_{x \in A} \{|\varphi_j(x)|^4 + |\varphi'_j(x)|^4\}. \end{aligned}$$

Therefore, there exists a deterministic constant  $\mathbf{c}_e > 0$ , not depending on  $m$ ,  $N$  and  $T$ , such that

$$\mathbb{E}(|\mathbf{e}(m)^* \mathbf{e}(m)|^2) \leq \mathbf{c}_e \sigma^4 \frac{\bar{\mathbf{m}}_{4,H,M}(T)}{N^2 T^{4-4H}} m(L(m)^2 + R(m)^2).$$

6.3.3. *Proof of Lemma 6.2.* The proof of Lemma 6.2 is close to the proof of Comte and Genon-Catalot [12], Lemma 3.

On the one hand, by the beginning of Comte and Genon-Catalot [12], Subsection 6.1 which remains true for  $H > 1/2$  without additional arguments,

$$\Omega(m) = \left\{ \|\Psi(m)^{-1/2} \hat{\Psi}(m) \Psi(m)^{-1/2} - \mathbf{I}_m\|_{\text{op}} \leq \frac{1}{2} \right\}.$$

Then, as in the proof of Comte and Genon-Catalot [11], Proposition 3, by Cohen et al. [9], Theorem 1,

$$\mathbb{P}(\Omega(m)^c) \leq 2m \exp \left( -\mathbf{c}_1 \frac{NT}{L(m)(\|\Psi(m)^{-1}\|_{\text{op}} \vee 1)} \right)$$

with  $\mathbf{c}_1 := 1/2(3 \log(3/2) - 1)$ . Under Assumption 3.3,

$$L(m)(\|\Psi(m)^{-1}\|_{\text{op}} \vee 1) \leq \frac{\mathbf{c}_{\kappa,T}}{2} \cdot \frac{NT}{\log(NT)}.$$

Then, there exists a deterministic constant  $\mathbf{c}_\Omega > 0$ , not depending on  $m$ ,  $N$  and  $T$ , such that

$$\mathbb{P}(\Omega(m)^c) \leq \frac{\mathbf{c}_\Omega}{(NT)^{6+\kappa}}.$$

On the other hand, as in the proof of Comte and Genon-Catalot [11], Lemma 5, under Assumption 3.3 and on  $\hat{\Lambda}_\kappa(m)^c$ ,

$$L(m)\|\Psi(m)^{-1}\|_{\text{op}} \leq \frac{\mathbf{c}_{\kappa,T}}{2} \cdot \frac{NT}{\log(NT)}$$

and

$$L(m)\|\hat{\Psi}(m)^{-1}\|_{\text{op}} > \mathbf{c}_{\kappa,T} \frac{NT}{\log(NT)},$$

and then

$$\begin{aligned} \hat{\Lambda}_\kappa(m)^c &\subset \left\{ L(m)\|\hat{\Psi}(m)^{-1} - \Psi(m)^{-1}\|_{\text{op}} > \frac{\mathbf{c}_{\kappa,T}}{2} \cdot \frac{NT}{\log(NT)} \right\} \\ &\subset \{ \|\hat{\Psi}(m)^{-1} - \Psi(m)^{-1}\|_{\text{op}} > \|\Psi(m)^{-1}\|_{\text{op}} \}. \end{aligned}$$

Finally, by Comte and Genon-Catalot [11], Proposition 4.(ii) which remains true for  $H > 1/2$  without additional arguments,

$$\{ \|\hat{\Psi}(m)^{-1} - \Psi(m)^{-1}\|_{\text{op}} > \|\Psi(m)^{-1}\|_{\text{op}} \} \subset \Omega(m)^c.$$

Therefore,

$$\mathbb{P}(\hat{\Lambda}_\kappa(m)^c) \leq \mathbb{P}(\Omega(m)^c) \leq \frac{\mathbf{c}_\Omega}{(NT)^{6+\kappa}}.$$

6.4. **Proof of Corollary 3.5.** First of all,

$$(18) \quad \mathbb{E}(\|\widehat{b}_m - b_A\|_{f_T}^2) = \mathfrak{m}_1 + \mathfrak{m}_2 + \mathfrak{m}_3$$

where

$$\begin{aligned} \mathfrak{m}_1 &:= \mathbb{E}(\|b_A\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m)^c}), \\ \mathfrak{m}_2 &:= \mathbb{E}(\|\widehat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m) \cap \Omega(m)}) \text{ and} \\ \mathfrak{m}_3 &:= \mathbb{E}(\|\widehat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m) \cap \Omega(m)^c}). \end{aligned}$$

On the one hand, by Lemma 6.2, there exists a deterministic constant  $\mathfrak{c}_1 > 0$ , not depending on  $m$ ,  $N$  and  $T$ , such that

$$(19) \quad \mathfrak{m}_1 \leq \frac{\mathfrak{c}_1}{NT},$$

and as in the proof of Comte and Genon-Catalot [12], Proposition 1, Theorem 3.4 allows to get

$$(20) \quad \mathfrak{m}_2 \leq \left(1 + \frac{4\mathfrak{c}_{\kappa,T}}{\log(NT)}\right) \inf_{\tau \in \mathcal{S}_m} \|\tau - b_A\|_{f_T}^2 + \frac{8}{NT} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)).$$

On the other hand, by Lemma 6.1,

$$\begin{aligned} \mathbb{E}(|\widehat{\mathbf{x}}(m)^* \widehat{\mathbf{x}}(m)|^2) &\leq 4m \sum_{j=0}^{m-1} \mathbb{E}(\langle \varphi_j, b \rangle_N^4) + 4\mathbb{E}(|\mathbf{e}(m)^* \mathbf{e}(m)|^2) \\ &\leq 4mL(m)^2 \int_{-\infty}^{\infty} b_A(x)^4 f_T(x) dx \\ &\quad + 4\mathfrak{c}_e \sigma^4 \frac{\overline{\mathfrak{m}}_{4,H,M}(T)}{N^2 T^{4-4H}} m(L(m)^2 + R(m)^2) \\ &\leq \mathfrak{c}_2 \left( mL(m)^2 + \frac{\overline{\mathfrak{m}}_{4,H,M}(T)}{N^2 T^{4-4H}} m(L(m)^2 + \rho^2 L(m)^{2\kappa}) \right) \end{aligned}$$

where

$$\mathfrak{c}_2 := \left( 4 \int_{-\infty}^{\infty} b_A(x)^4 f_T(x) dx \right) \vee (4\mathfrak{c}_e \sigma^4).$$

As in the proof of Comte and Genon-Catalot [12], Proposition 1,  $\|\Psi(m)\|_{\text{op}} \leq L(m)$ , and by the definition of  $\widehat{\Lambda}_\kappa(m)$ ,

$$\begin{aligned} \mathbb{E}(\|\widehat{b}_m\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m) \cap \Omega(m)^c}) &\leq \|\Psi(m)\|_{\text{op}} \mathbb{E}(\|\widehat{\Psi}(m)^{-1}\|_{\text{op}}^2 \widehat{\mathbf{x}}(m)^* \widehat{\mathbf{x}}(m) \mathbf{1}_{\widehat{\Lambda}_\kappa(m) \cap \Omega(m)^c}) \\ &\leq \frac{\mathfrak{c}_{\kappa,T}^2}{L(m)} \left( \frac{NT}{\log(NT)} \right)^2 \mathbb{P}(\Omega(m)^c)^{1/2} \mathbb{E}(|\widehat{\mathbf{x}}(m)^* \widehat{\mathbf{x}}(m)|^2)^{1/2} \\ &\leq \frac{\mathfrak{c}_2^{1/2} \mathfrak{c}_{\kappa,T}^2}{L(m)} \left( \frac{NT}{\log(NT)} \right)^2 \frac{\mathfrak{c}_\Omega^{1/2}}{(NT)^{3+\kappa/2}} \\ &\quad \times \left( m^{1/2} L(m) + \frac{\overline{\mathfrak{m}}_{2,H,M}(T)}{NT^{2-2H}} m^{1/2} (L(m) + \rho L(m)^\kappa) \right). \end{aligned}$$

Then, there exists a deterministic constant  $\mathfrak{c}_3 > 0$ , not depending on  $m$ ,  $N$  and  $T$ , such that

$$(21) \quad \mathfrak{m}_3 \leq 2\mathbb{E}(\|\widehat{b}_m\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m) \cap \Omega(m)^c}) + 2\|b_A\|_{f_T}^2 \mathbb{P}(\Omega(m)^c) \leq \mathfrak{c}_3 \frac{\overline{\mathfrak{m}}_{2,H,M}(T)}{NT}.$$

Inequalities (19), (20) and (21) together with Inequality (18) allow to conclude.

**6.5. Proof of Proposition 3.7.** First note that  $\Psi(m)$  and  $\Psi(m, \sigma)$  are symmetric and nonnegative, as for any vector  $y$ ,  $y^* \Psi(m) y \geq 0$  and  $y^* \Psi(m, \sigma) y \geq 0$ . Indeed for  $y \in \mathbb{R}^m$ ,  $y^* \Psi_m y = \int \tau^2(x) f_T(x) dx \geq 0$  and

$$\begin{aligned} y^* \Psi(m, \sigma) y &= \frac{\sigma^2}{T} \sum_{j,k=0}^{m-1} y^j y^k \mathbb{E} \left( \left( \int_0^T \varphi_j(X(s)) \delta B(s) \right) \left( \int_0^T \varphi_k(X(s)) \delta B(s) \right) \right) \\ &= \frac{\sigma^2}{T} \mathbb{E} \left[ \left( \int_0^T \tau(X(s)) \delta B(s) \right)^2 \right] \geq 0, \quad \text{where } \tau := \sum_{j=0}^{m-1} y^j \varphi_j. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) &\leq \|\Psi(m)^{-1}\|_{\text{op}} \text{trace}(\Psi(m, \sigma)) \\ &= \frac{\sigma^2}{T} \|\Psi(m)^{-1}\|_{\text{op}} \sum_{j=0}^{m-1} \mathbb{E} \left[ \left( \int_0^T \varphi_j(X(s)) \delta B(s) \right)^2 \right] \end{aligned}$$

By Theorem 2.9, Equality (16) and Inequality (17),

$$\begin{aligned} \mathbb{E} \left[ \left( \int_0^T \varphi_j(X(s)) \delta B(s) \right)^2 \right] &\leq \mathfrak{c}_{2,H,\sigma} \bar{\mathfrak{m}}_{2,H,M}(T) \left[ \left( \int_0^T \mathbb{E}(|\varphi_j(X(s))|^{1/H}) ds \right)^{2H} \right. \\ &\quad \left. + \left( \int_0^T \mathbb{E}(|\varphi_j'(X(s))|^2)^{1/(2H)} ds \right)^{2H} \right] \\ (22) \quad &\leq \mathfrak{c}_{2,H,\sigma} \bar{\mathfrak{m}}_{2,H,M}(T) T^{2H} \left( \int_{-\infty}^{\infty} \varphi_j(x)^2 f_T(x) dx + \int_{-\infty}^{\infty} \varphi_j'(x)^2 f_T(x) dx \right). \end{aligned}$$

Finally we get

$$\text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) \leq \mathfrak{c}_{2,H,\sigma} \bar{\mathfrak{m}}_{2,H,M}(T) \frac{\sigma^2}{T^{1-2H}} \|\Psi(m)^{-1}\|_{\text{op}} (L(m) + R(m)).$$

For the second bound, let us denote by  $\Psi(m)^{-1/2}$  a symmetric square root of  $\Psi(m)^{-1}$ , and write

$$\begin{aligned} \text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) &= \text{trace}(\Psi(m)^{-1/2} \Psi(m, \sigma) \Psi(m)^{-1/2}) \\ &\leq m \|\Psi(m)^{-1/2} \Psi(m, \sigma) \Psi(m)^{-1/2}\|_{\text{op}} \\ (23) \quad &= m \sup_{\|x\|_{2,m}=1} x^* \Psi(m)^{-1/2} \Psi(m, \sigma) \Psi(m)^{-1/2} x = m \sup_{\|\Psi(m)^{1/2} y\|_{2,m}=1} y^* \Psi(m, \sigma) y. \end{aligned}$$

As already noticed, for any  $y \in \mathbb{R}^m$  such that  $\|\Psi(m)^{1/2} y\|_{2,m} = 1$ ,

$$y^* \Psi(m, \sigma) y = \frac{\sigma^2}{T} \mathbb{E} \left[ \left( \int_0^T \tau(X(s)) \delta B(s) \right)^2 \right] \quad \text{with } \tau := \sum_{j=0}^{m-1} y^j \varphi_j.$$

By Theorem 2.9, Equality (16) and Inequality (17) again,

$$(24) \quad \mathbb{E} \left[ \left( \int_0^T \tau(X(s)) \delta B(s) \right)^2 \right] \leq \mathfrak{c}_{2,H,\sigma} \bar{\mathfrak{m}}_{2,H,M}(T) T^{2H} \left( \int_{-\infty}^{\infty} \tau(x)^2 f_T(x) dx + \int_{-\infty}^{\infty} \tau'(x)^2 f_T(x) dx \right).$$

Since

$$\int_{-\infty}^{\infty} \tau(x)^2 f_T(x) dx = y^* \Psi(m) y = 1,$$

by Inequalities (23) and (24),

$$\text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) \leq \mathfrak{c}_{2,H,\sigma} \sigma^2 \frac{\bar{\mathfrak{m}}_{2,H,M}(T)}{T^{1-2H}} m (1 + \|\Psi(m)^{-1/2} \Psi(m, \sigma) \Psi(m)^{-1/2}\|_{\text{op}})$$



where

$$\Psi^*(m, \sigma) := \left( \int_{-\infty}^{\infty} \varphi'_j(x) \varphi'_k(x) f_T(x) dx \right)_{j,k=0,\dots,m-1}.$$

Moreover,

$$\|\Psi(m)^{-1/2} \Psi^*(m, \sigma) \Psi(m)^{-1/2}\|_{\text{op}} \leq \|\Psi(m)^{-1/2}\|_{\text{op}}^2 R(m).$$

Therefore,

$$\text{trace}(\Psi(m)^{-1} \Psi(m, \sigma)) \leq c_{2,H,\sigma} \sigma^2 \frac{\bar{m}_{2,H,M}(T)}{T^{1-2H}} m (1 + \|\Psi(m)^{-1}\|_{\text{op}} R(m)).$$

This concludes the proof.

**6.6. Proof of Proposition 3.11.** Consider  $x \in \mathbb{R}$  and  $\varepsilon, t > 0$ . For every  $s \in [0, t]$ ,

$$\partial_x X_x(s) = 1 + \int_0^s b'(X_x(r)) \partial_x X_x(r) dr$$

and, by Taylor's formula,

$$X_{x+\varepsilon}(s) - X_x(s) = \varepsilon + \int_0^s (X_{x+\varepsilon}(r) - X_x(r)) \int_0^1 b'(X_x(r) + \theta(X_{x+\varepsilon}(r) - X_x(r))) d\theta dr.$$

So, for every  $(u, v) \in [0, t]^2$  such that  $v < u$ ,

$$\frac{\partial_x X_x(u)}{\partial_x X_x(v)} = \exp \left( \int_v^u b'(X_x(r)) dr \right)$$

and

$$(25) \quad \frac{X_{x+\varepsilon}(u) - X_x(u)}{X_{x+\varepsilon}(v) - X_x(v)} = \exp \left( \int_v^u \int_0^1 b'(X_x(r) + \theta(X_{x+\varepsilon}(r) - X_x(r))) d\theta dr \right).$$

For a given  $\varphi \in \text{Lip}_b^1(\mathbb{R})$ , by Comte and Marie [15], Proposition 2.7,

$$\Delta_\varphi^S(x, \varepsilon, t) \leq \alpha_H \sigma^2 \int_0^t \int_0^u |\varphi'(X_x(u))| \Delta_\varphi(x, \varepsilon, u, v) (u - v)^{2H-2} dv du,$$

where

$$\Delta_\varphi^S(x, \varepsilon, t) := \left| \int_0^t \varphi(X_x(u)) \delta X_x(u) - S_\varphi(x, \varepsilon, t) \right|$$

and, for every  $(u, v) \in [0, t]^2$  such that  $v < u$ ,

$$\Delta_\varphi(x, \varepsilon, u, v) := \left| \frac{\partial_x X_x(u)}{\partial_x X_x(v)} - \frac{X_{x+\varepsilon}(u) - X_x(u)}{X_{x+\varepsilon}(v) - X_x(v)} \right|.$$

Since  $b'(\mathbb{R}) \subset [m, M]$  and  $b$  is two times continuously differentiable,

$$\begin{aligned} \Delta_\varphi(x, \varepsilon, u, v) &= \left| \exp \left( \int_v^u b'(X_x(r)) dr \right) \right. \\ &\quad \left. - \exp \left( \int_v^u \int_0^1 b'(X_x(r) + \theta(X_{x+\varepsilon}(r) - X_x(r))) d\theta dr \right) \right| \\ &\leq \sup_{z \in [m(u-v), M(u-v)]} e^z \\ &\quad \times \int_v^u \left| b'(X_x(r)) - \int_0^1 b'(X_x(r) + \theta(X_{x+\varepsilon}(r) - X_x(r))) d\theta \right| dr \\ &\leq (1 \vee e^{Mt}) \int_v^u \int_0^1 |b'(X_x(r)) - b'(X_x(r) + \theta(X_{x+\varepsilon}(r) - X_x(r)))| d\theta dr \\ &\leq \frac{\|b''\|_\infty}{2} (1 \vee e^{Mt}) \int_v^u |X_{x+\varepsilon}(r) - X_x(r)| dr. \end{aligned}$$

Consider  $s \in \mathbb{R}_+$ . By Equation (1),

$$\begin{aligned} (X_{x+\varepsilon}(s) - X_x(s))^2 &= \varepsilon^2 + 2 \int_0^s (X_{x+\varepsilon}(r) - X_x(r)) d(X_{x+\varepsilon} - X_x)(r) \\ &= \varepsilon^2 + 2 \int_0^s (X_{x+\varepsilon}(r) - X_x(r))(b(X_{x+\varepsilon}(r)) - b(X_x(r))) dr. \end{aligned}$$

By the mean value theorem, there exists  $x_s \in \mathbb{R}$  such that

$$\begin{aligned} \partial_s (X_{x+\varepsilon}(s) - X_x(s))^2 &= 2(X_{x+\varepsilon}(s) - X_x(s))^2 \frac{b(X_{x+\varepsilon}(s)) - b(X_x(s))}{X_{x+\varepsilon}(s) - X_x(s)} \\ &= 2(X_{x+\varepsilon}(s) - X_x(s))^2 b'(x_s) \leq 2M(X_{x+\varepsilon}(s) - X_x(s))^2 \end{aligned}$$

and then,

$$(26) \quad |X_{x+\varepsilon}(s) - X_x(s)| \leq \varepsilon e^{Ms}.$$

Therefore,

$$\begin{aligned} \Delta_\varphi(x, \varepsilon, u, v) &\leq \frac{\|b''\|_\infty}{2} \varepsilon (1 \vee e^{Mt}) \int_v^u e^{Mr} dr \\ &\leq \frac{\|b''\|_\infty}{2} \varepsilon (1 \vee e^{Mt}) \mathbf{m}_M(u, v) \end{aligned}$$

where

$$\mathbf{m}_M(u, v) := -\frac{1}{M} e^{Mv} \mathbf{1}_{M < 0} + (u - v) \mathbf{1}_{M=0} + \frac{1}{M} e^{Mu} \mathbf{1}_{M > 0}.$$

Finally, using the above bounds, and in a second stage, the integration by parts formula, we get

$$\begin{aligned} \Delta_\varphi^S(x, \varepsilon, t) &\leq \alpha_H \sigma^2 \frac{\|b''\|_\infty}{2} \varepsilon (1 \vee e^{Mt}) \int_0^t \int_0^u |\varphi'(X_x(u))| \mathbf{m}_M(u, v) (u - v)^{2H-2} dv du \\ &\leq \alpha_H \sigma^2 \frac{\|b''\|_\infty \|\varphi'\|_\infty}{2} \varepsilon (1 \vee e^{Mt}) \left( -\frac{1}{M} \mathbf{1}_{M < 0} \int_0^t e^{Mv} \int_v^t (u - v)^{2H-2} du dv \right. \\ &\quad \left. + \mathbf{1}_{M=0} \int_0^t \int_0^u (u - v)^{2H-2} dv du + \frac{1}{M} \mathbf{1}_{M > 0} \int_0^t e^{Mu} \int_0^u (u - v)^{2H-2} dv du \right) \\ &\leq \alpha_H \sigma^2 \frac{\|b''\|_\infty \|\varphi'\|_\infty}{2} \varepsilon (1 \vee e^{Mt}) \left( \frac{1}{M^2} \mathbf{1}_{M < 0} \left( \frac{t^{2H-1}}{2H-1} - \int_0^t e^{Mv} (t - v)^{2H-2} dv \right) \right. \\ &\quad \left. + \frac{1}{2H} \mathbf{1}_{M=0} \int_0^t u^{2H} du + \frac{1}{M^2} \mathbf{1}_{M > 0} \left( \frac{e^{Mt} t^{2H-1}}{2H-1} - \int_0^t e^{Mu} u^{2H-2} du \right) \right) \\ &\leq \alpha_H \sigma^2 \frac{\|b''\|_\infty \|\varphi'\|_\infty}{2} \varepsilon t^{2H-1} \mathbf{m}_{H,M}(t) \end{aligned}$$

where

$$\mathbf{m}_{H,M}(t) = \frac{1}{M^2(2H-1)} \mathbf{1}_{M < 0} + \frac{t^2}{2H(2H+1)} \mathbf{1}_{M=0} + \frac{e^{2Mt}}{M^2(2H-1)} \mathbf{1}_{M > 0}.$$

**6.7. Proof of Corollary 3.13.** First of all,

$$\mathbb{E}(\|\tilde{b}_{m,\varepsilon} - b_A\|_{f_T}^2) \leq 2\mathbb{E}(\|\hat{b}_{m,\varepsilon} - \hat{b}_m\|_{f_T}^2 \mathbf{1}_{\hat{\Lambda}_\kappa(m)}) + 2\mathbb{E}(\|\tilde{b}_m - b_A\|_{f_T}^2)$$

and

$$\begin{aligned} \mathbb{E}(\|\tilde{b}_m - b_A\|_{f_T}^2) &\leq \left( 1 + 4 \frac{3 \log(3/2) - 1}{(7 + \kappa) \log(NT)} \right) \inf_{\tau \in \mathcal{S}_m} \|\tau - b_A\|_{f_T}^2 \\ &\quad + \frac{8}{NT} \text{trace}(\mathbf{\Psi}(m)^{-1} \mathbf{\Psi}(m, \sigma)) + \bar{\mathbf{c}}_{\rho, \kappa, \sigma} (1 + \mathbf{b}_T) \frac{\bar{\mathbf{m}}_{2,H,M}(T)}{NT} \end{aligned}$$

by Corollary 3.5. So, let us find a suitable control for  $\mathbb{E}(\|\widehat{b}_{m,\varepsilon} - \widehat{b}_m\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m)})$ . On the one hand,

$$\begin{aligned} \|\widehat{b}_{m,\varepsilon} - \widehat{b}_m\|_{f_T}^2 &= \|\langle \widehat{\theta}(m, \varepsilon) - \widehat{\theta}(m), (\varphi_0, \dots, \varphi_{m-1})(\cdot) \rangle_{2,m}\|_{f_T}^2 \\ &\leq \|\widehat{\theta}(m, \varepsilon) - \widehat{\theta}(m)\|_{2,m}^2 \int_A \|(\varphi_0, \dots, \varphi_{m-1})(x)\|_{2,m}^2 f_T(x) dx \\ &\leq \|\widehat{\theta}(m, \varepsilon) - \widehat{\theta}(m)\|_{2,m}^2 L(m). \end{aligned}$$

On the other hand, by Proposition 3.11, for every  $i \in \{1, \dots, N\}$  and  $j \in \{0, \dots, m-1\}$ ,

$$\left| S_{\varphi_j}^i(x_0, \varepsilon, T) - \int_0^T \varphi_j(X_{x_0}^i(s)) \delta X_{x_0}^i(s) \right| \leq \alpha_H \sigma^2 \frac{\|b''\|_\infty \|\varphi'\|_\infty}{2} \varepsilon T^{2H-1} \mathbf{m}_{H,M}(T).$$

Then,

$$\begin{aligned} \|\widehat{\theta}(m, \varepsilon) - \widehat{\theta}(m)\|_{2,m}^2 &\leq \|\widehat{\Psi}(m)^{-1}\|_{\text{op}}^2 \|\widehat{\mathbf{x}}(m, \varepsilon) - \widehat{\mathbf{x}}(m)\|_{2,m}^2 \\ &= \|\widehat{\Psi}(m)^{-1}\|_{\text{op}}^2 \sum_{j=0}^{m-1} \left| \frac{1}{NT} \sum_{i=1}^N \left( S_{\varphi_j}^i(x_0, \varepsilon, T) - \int_0^T \varphi_j(X_{x_0}^i(s)) \delta X_{x_0}^i(s) \right) \right|^2 \\ &\leq \mathbf{c}_1 \|\widehat{\Psi}(m)^{-1}\|_{\text{op}}^2 \varepsilon^2 T^{4H-4} \mathbf{m}_{H,M}(T)^2 R(m) \end{aligned}$$

with

$$\mathbf{c}_1 := \alpha_H^2 \sigma^4 \frac{\|b''\|_\infty^2}{4}.$$

Therefore, by the definition of  $\widehat{\Lambda}_\kappa(m)$ , there exists a deterministic constant  $\mathbf{c}_{\rho, \kappa, \sigma, H, \mathbf{b}''} > 0$ , depending only on  $\rho, \kappa, \sigma, H$  and  $\mathbf{b}''$ , such that

$$\begin{aligned} 2\mathbb{E}(\|\widehat{b}_{m,\varepsilon} - \widehat{b}_m\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_\kappa(m)}) &\leq 2\mathbf{c}_1 \mathbf{c}_{\kappa,T}^2 \left( \frac{NT}{\log(NT)} \right)^2 L(m)^{-1} \varepsilon^2 T^{4H-4} \mathbf{m}_{H,M}(T)^2 R(m) \\ &\leq \mathbf{c}_{\rho, \kappa, \sigma, H, \mathbf{b}''} L(m)^{\kappa-1} \left( \frac{NT}{\log(NT)} \right)^2 \frac{\mathbf{m}_{H,M}(T)^2}{T^{4-4H}} \varepsilon^2. \end{aligned}$$

**6.8. Proof of Proposition 4.2.** First, let us prove (10). We have

$$\|\widetilde{b}_m^{\dagger, \prime} - b'_A\|_N^2 = \|\widehat{b}_m^{\dagger, \prime} - b'_A\|_N^2 \mathbf{1}_{\widehat{\Lambda}_0(m)} + \|b'_A\|_N^2 \mathbf{1}_{\widehat{\Lambda}_0(m)^c}.$$

On the one hand, by Lemma 6.2,

$$\mathbb{E}(\|b'_A\|_N^2 \mathbf{1}_{\widehat{\Lambda}_0(m)^c}) \leq \mathbb{E}(\|b'_A\|_N^4)^{1/2} \mathbb{P}(\widehat{\Lambda}_0(m)^c)^{1/2} \leq \frac{\mathbf{c}_1}{NT}$$

where

$$\mathbf{c}_1 := \mathbf{c}_\Omega^{1/2} \left( \int_{-\infty}^{\infty} b'_A(x)^4 f_T(x) dx \right)^{1/2} \leq \mathbf{c}_\Omega^{1/2} \|b'\|_\infty^2 = \mathbf{c}_{4.2}.$$

On the other hand, since Model (8) is noiseless,  $\widehat{b}_m^{\dagger, \prime} = p_{N, \mathcal{S}_m}^\perp(b'_A)$ , and then

$$\|\widehat{b}_m^{\dagger, \prime} - b'_A\|_N^2 = \inf_{\tau \in \mathcal{S}_m} \|\tau - b'_A\|_N^2.$$

Therefore,

$$\mathbb{E}(\|\widetilde{b}_m^{\dagger, \prime} - b'_A\|_N^2) \leq \inf_{\tau \in \mathcal{S}_m} \|\tau - b'_A\|_{f_T}^2 + \frac{\mathbf{c}_{4.2}}{NT}.$$

Now let us prove (11). We start by writing that

$$(27) \quad \mathbb{E}(\|\widetilde{b}_m^{\dagger, \prime} - b'_A\|_{f_T}^2) = \mathbf{m}_1 + \mathbf{m}_2 + \mathbf{m}_3$$

where

$$\begin{aligned} \mathbf{m}_1 &:= \mathbb{E}(\|b'_A\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_0(m)^c}), \\ \mathbf{m}_2 &:= \mathbb{E}(\|\widehat{b}_m^{\dagger, \prime} - b'_A\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_0(m) \cap \Omega(m)}) \text{ and} \\ \mathbf{m}_3 &:= \mathbb{E}(\|\widehat{b}_m^{\dagger, \prime} - b'_A\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_0(m) \cap \Omega(m)^c}). \end{aligned}$$

Let us find suitable bounds on  $\mathbf{m}_1$ ,  $\mathbf{m}_2$  and  $\mathbf{m}_3$ .

- **Bound on  $\mathbf{m}_1$ .** By Lemma 6.2, there exists a deterministic constant  $\mathbf{c}_1 > 0$ , not depending on  $m$ ,  $N$  and  $T$ , such that

$$(28) \quad \mathbf{m}_1 \leq \frac{\mathbf{c}_1}{NT}.$$

- **Bound on  $\mathbf{m}_2$ .** By denoting  $p_{f_T, \mathcal{S}_m}^\perp$  the orthogonal projection from  $\mathbb{L}^2(A, f_T(x)dx)$  onto  $\mathcal{S}_m$  with respect to the  $f_T$ -weighted scalar product  $\langle \cdot, \cdot \rangle_{f_T}$ ,

$$\begin{aligned} \|\widehat{b}_m^{\dagger, \prime} - b'_A\|_{f_T}^2 &= \|b'_A - p_{N, \mathcal{S}_m}^\perp(b'_A)\|_{f_T}^2 = \|b'_A - p_{f_T, \mathcal{S}_m}^\perp(b'_A) - p_{N, \mathcal{S}_m}^\perp(b'_A - p_{f_T, \mathcal{S}_m}^\perp(b'_A))\|_{f_T}^2 \\ &= \inf_{\tau \in \mathcal{S}_m} \|\tau - b'_A\|_{f_T}^2 + \|p_{N, \mathcal{S}_m}^\perp(g)\|_{f_T}^2 \end{aligned}$$

with  $g := b'_A - p_{f_T, \mathcal{S}_m}^\perp(b'_A)$ . By Comte and Genon-Catalot [12], Lemma 5 applied to this function  $g$ ,

$$\mathbb{E}(\|p_{N, \mathcal{S}_m}^\perp(g)\|_{f_T}^2 \mathbf{1}_{\widehat{\Lambda}_0(m) \cap \Omega(m)}) \leq \frac{2\mathbf{c}_{0,T}}{\log(NT)} \|g\|_{f_T}^2 = \frac{2\mathbf{c}_{0,T}}{\log(NT)} \inf_{\tau \in \mathcal{S}_m} \|\tau - b'_A\|_{f_T}^2.$$

- **Bound on  $\mathbf{m}_3$ .** First of all, by Lemma 6.2,

$$\mathbf{m}_3 \leq \frac{\mathbf{c}_\Omega^{1/2}}{(NT)^3} \mathbb{E}(\|\widehat{b}_m^{\dagger, \prime}\|_{f_T}^4 \mathbf{1}_{\widehat{\Lambda}_0(m)})^{1/2} + \|b'_A\|_{f_T}^2 \frac{\mathbf{c}_\Omega}{NT}.$$

So, it remains to find a bound on  $\mathbb{E}(\|\widehat{b}_m^{\dagger, \prime}\|_{f_T}^4 \mathbf{1}_{\widehat{\Lambda}_0(m)})$ . On the one hand,

$$\begin{aligned} \|\widehat{b}_m^{\dagger, \prime}\|_{f_T}^2 &= \int_{-\infty}^{\infty} \left[ \sum_{j=0}^{m-1} \widehat{\theta}'_j \varphi_j(x) \right]^2 f_T(x) dx \\ &= \widehat{\theta}'(m)^* \Psi(m) \widehat{\theta}'(m) = \widehat{\mathbf{y}}(m)^* \widehat{\Psi}(m)^{-1} \Psi(m) \widehat{\Psi}(m)^{-1} \widehat{\mathbf{y}}(m) = \|\Psi(m)^{1/2} \widehat{\Psi}(m)^{-1} \widehat{\mathbf{y}}(m)\|_{2,m}^2 \\ &\leq \|\Psi(m)\|_{\text{op}} \|\widehat{\Psi}(m)^{-1}\|_{\text{op}}^2 \sum_{j=0}^{m-1} \langle \varphi_j, b' \rangle_N^2. \end{aligned}$$

On the other hand,  $\|\Psi(m)\|_{\text{op}} \leq L(m)$ ,

$$\|\widehat{\Psi}(m)^{-1}\|_{\text{op}}^2 \leq \mathbf{c}_{0,T}^2 L(m)^{-2} \left( \frac{NT}{\log(NT)} \right)^2$$

on  $\widehat{\Lambda}_0(m)$ , and

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{j=0}^{m-1} \langle \varphi_j, b' \rangle_N^2 \right|^2 \right]^{1/2} &\leq \sqrt{m} \mathbb{E} \left[ \sum_{j=0}^{m-1} \left| \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X_{x_0}^i(s)) b'(X_{x_0}^i(s)) ds \right|^4 \right]^{1/2} \\ &\leq \sqrt{m} \left[ \frac{1}{T} \int_0^T \mathbb{E} \left[ \sum_{j=0}^{m-1} \varphi_j(X_{x_0}^1(s))^4 b'(X_{x_0}^1(s))^4 \right] ds \right]^{1/2} \\ &\leq \sqrt{m} L(m) \left( \frac{1}{T} \int_0^T \int_{-\infty}^{\infty} b'(x)^4 p_s(x_0, x) dx ds \right)^{1/2} = \sqrt{m} L(m) \left( \int_{-\infty}^{\infty} b'(x)^4 f_T(x) dx \right)^{1/2}. \end{aligned}$$

Then,

$$\mathbb{E}(\|\widehat{b}_m^{\dagger, \prime}\|_{f_T}^4 \mathbf{1}_{\widehat{\Lambda}_0(m)})^{1/2} \leq \mathbf{c}_{0,T} \|b'\|_\infty^2 \sqrt{m} \frac{NT}{\log(NT)}.$$

6.9. **Proof of Corollary 4.3.** First of all,

$$\mathbb{E}(\|\tilde{b}_{m,\varepsilon}^{\dagger,\prime} - b'_A\|_{f_T}^2) \leq 2\mathbb{E}(\|\tilde{b}_{m,\varepsilon}^{\dagger,\prime} - \hat{b}_m^{\dagger,\prime}\|_{f_T}^2 \mathbf{1}_{\hat{\Lambda}_0(m)}) + 2\mathbb{E}(\|\hat{b}_m^{\dagger,\prime} - b'_A\|_{f_T}^2)$$

and

$$\mathbb{E}(\|\tilde{b}_m^{\dagger,\prime} - b'_A\|_{f_T}^2) \leq \left(1 + \frac{2T\mathfrak{c}_{0,T}}{\log(NT)}\right) \inf_{\tau \in \mathcal{S}_m} \|\tau - b'_A\|_{f_T}^2 + \bar{\mathfrak{c}}_{b'} \frac{1 + \mathfrak{c}_{0,T}}{NT}$$

by (11) of Proposition 4.2. So, let us find a suitable control for  $\mathbb{E}(\|\tilde{b}_{m,\varepsilon}^{\dagger,\prime} - \hat{b}_m^{\dagger,\prime}\|_{f_T}^2 \mathbf{1}_{\hat{\Lambda}_0(m)})$ . On the one hand,

$$\begin{aligned} \|\hat{b}_{m,\varepsilon}^{\dagger,\prime} - \hat{b}_m^{\dagger,\prime}\|_{f_T}^2 &= \|\langle \hat{\theta}'(m, \varepsilon) - \hat{\theta}'(m), (\varphi_0, \dots, \varphi_{m-1})(\cdot) \rangle_{2,m}\|_{f_T}^2 \\ &\leq \|\hat{\theta}'(m, \varepsilon) - \hat{\theta}'(m)\|_{2,m}^2 \int_A \|(\varphi_0, \dots, \varphi_{m-1})(x)\|_{2,m}^2 f_T(x) dx \\ &\leq \|\hat{\theta}'(m, \varepsilon) - \hat{\theta}'(m)\|_{2,m}^2 L(m). \end{aligned}$$

On the other hand, for every  $s \in [0, T]$ , by (25) and (26),

$$\begin{aligned} |\dot{Y}_{x_0}(s) - \dot{Y}_{x_0+\varepsilon}(s)| &= \left| b'(X_{x_0}(s)) - \int_0^1 b'(X_{x_0}(s) + \theta(X_{x_0+\varepsilon}(s) - X_{x_0}(s))) d\theta \right| \\ &\leq \int_0^1 |b'(X_{x_0}(s)) - b'(X_{x_0}(s) + \theta(X_{x_0+\varepsilon}(s) - X_{x_0}(s)))| d\theta \\ &\leq \|b''\|_\infty |X_{x_0+\varepsilon}(s) - X_{x_0}(s)| \int_0^1 \theta d\theta \leq \frac{\|b''\|_\infty}{2} e^{Ms} \varepsilon. \end{aligned}$$

Therefore, thanks to the definition of  $\hat{\Lambda}_0(m)$ ,

$$\begin{aligned} &\|\hat{\theta}'(m, \varepsilon) - \hat{\theta}'(m)\|_{2,m}^2 \mathbf{1}_{\hat{\Lambda}_0(m)} \\ &\leq \|\hat{\Psi}(m)^{-1}\|_{\text{op}}^2 \mathbf{1}_{\hat{\Lambda}_0(m)} \sum_{j=0}^{m-1} \left| \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X_{x_0}^i(s)) d(Y_{x_0,\varepsilon}^i - Y_{x_0}^i)(s) \right|^2 \\ &\leq \frac{\|b''\|_\infty^2 e^{2MT}}{4} \varepsilon^2 \|\hat{\Psi}(m)^{-1}\|_{\text{op}}^2 \mathbf{1}_{\hat{\Lambda}_0(m)} \sum_{j=0}^{m-1} \left| \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X_{x_0}^i(s)) ds \right|^2 \\ &\leq \frac{\|b''\|_\infty^2 e^{2MT}}{4} \varepsilon^2 L(m) \|\hat{\Psi}(m)^{-1}\|_{\text{op}}^2 \mathbf{1}_{\hat{\Lambda}_0(m)} \leq \frac{\|b''\|_\infty^2}{4} \mathfrak{c}_{0,T}^2 e^{2MT} \frac{1}{L(m)} \left( \frac{NT}{\log(NT)} \right)^2 \varepsilon^2. \end{aligned}$$

**Conflict of interest statement.** The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

**Acknowledgement.** We thank an anonymous referee for interesting remarks and helpful suggestion.

#### REFERENCES

- [1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.
- [2] S. Bajja, K. Es-Sebaïy and L. Viitasaari. Least Square Estimator of Fractional Ornstein-Uhlenbeck Processes with Periodic Mean. *Journal of the Korean Statistical Society* 36, 4, 608-622, 2017.
- [3] F. Biagini, Y. Hu, B. Oksendal and T. Zhang. *Stochastic Calculus for Fractional Brownian Motion and Applications*. Springer, 2008.
- [4] D. Belomestny, F. Comte and V. Genon-Catalot. Sobolev-Hermite versus Sobolev Nonparametric Density Estimation on  $\mathbb{R}$ . *The Annals of the Institute of Mathematical Statist.* 71, 1, 29-62, 2019.
- [5] B. Bongioanni and J.L. Torrea. Sobolev Spaces Associated to the Harmonic Oscillator. *Proc. Indian Acad. Sci. (math. Sci.)* 116, 3, 337-360, 2006.
- [6] D. Bosq. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer, 1996.
- [7] P. Cheridito, H. Kawaguchi and M. Maejima. Fractional Ornstein-Uhlenbeck Processes. *Electronic Journal of Probability* 8, 3, 1-14, 2003.

- [8] A. Chronopoulou and S. Tindel. On Inference for Fractional Differential Equations. *Stat. Inference Stoch. Process.* 16, 1, 29-61, 2013.
- [9] A. Cohen, M.A. Davenport and D. Leviatan. On the Stability and Accuracy of Least Squares Approximations. *Foundations of Computational Mathematics* 13, 819-834, 2013.
- [10] A. Cohen, M.A. Davenport and D. Leviatan. Correction to: On the stability and accuracy of least squares approximations. *Foundations of Computational Mathematics* 19, 239-239, 2019.
- [11] F. Comte and V. Genon-Catalot. Regression Function Estimation on Non Compact Support as a Partly Inverse Problem. *Annals of the Institute of Statistical Mathematics*, DOI: 10.1007/s10463-019-00718-2, 2019.
- [12] F. Comte and V. Genon-Catalot. Nonparametric Drift Estimation for i.i.d. Paths of Stochastic Differential Equations. Accepted for publication in *The Annals of Statistics*, 2019.
- [13] F. Comte, V. Genon-Catalot and A. Samson. Nonparametric Estimation for Stochastic Differential Equations with Random Effects. *Stochastic Processes and their Applications* 123, 7, 2522-2551, 2013.
- [14] Comte, F. and Lacour, C. (2021). Adaptive estimation of the conditional density from direct or noisy data. Preprint.
- [15] F. Comte and N. Marie. Nonparametric Estimation in Fractional SDE. *Statistical Inference for Stochastic Processes* 22, 3, 359-382, 2019.
- [16] D. D'Argenio and K. Park. Uncertain Pharmacokinetics/Pharmacodynamics Systems: Design, Estimation and Control. *Control Engineering Practice* 5, 12, 1707-1716, 1997.
- [17] M. Delattre, V. Genon-Catalot and A. Samson. Maximum Likelihood Estimation for Stochastic Differential Equations with Random Effects. *Scandinavian Journal of Statistics* 40, 322-343, 2013.
- [18] M. Delattre, V. Genon-Catalot and C. Larédo. Parametric Inference for Discrete Observations of Diffusion Processes with Mixed Effects. *Stochastic Processes and their Applications* 128, 1929-1957, 2018.
- [19] M. Delattre and M. Lavielle. Pharmacokinetics and Stochastic Differential Equations: Model and Methodology. *Proceedings of the Annual Meeting of the Population Approach Group in Europe*, 2011.
- [20] M. Delattre and M. Lavielle. Coupling the SAEM Algorithm and the Extended Kalman Filter for Maximum Likelihood Estimation in Mixed-Effects Diffusion Models. *Stat. Interface* 6, 519-532, 2013.
- [21] C. Denis, C. Dion and M. Martinez. Procedures for Multiclass Classification of Discrete Diffusion Paths. To appear in *Scandinavian Journal of Statistics*, 2020.
- [22] R.A. DeVore and G.G. Lorentz. *Constructive Approximation*. Springer-Verlag, 1993.
- [23] C. Dion and V. Genon-Catalot. Bidimensional Random Effect Estimation in Mixed Stochastic Differential Equations. *Statistical Inference for Stochastic processes* 19, 131-158, 2016.
- [24] S. Ditlevsen and A. De Gaetano. Mixed Effects in Stochastic Differential Equation Models. *REVSTAT* 3, 137-153, 2005.
- [25] S. Donnet and A. Samson. A Review on Estimation of Stochastic Differential Equations for Pharmacokinetic/Pharmacodynamic Models. *Advanced Drug Delivery Reviews* 65, 7, 929-939, 2013.
- [26] P. Friz and N. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge Studies in Applied Mathematics 120, Cambridge University Press, 2010.
- [27] J. Indritz. An Inequality for Hermite Polynomials. *Proc. Amer. Math. Soc.* 12, 981-983, 1961.
- [28] M.L. Kleptsyna and A. Le Breton. Some Explicit Statistical Results about Elementary Fractional Type Models. *Nonlinear Analysis* 47, 4783-4794, 2001.
- [29] K. Kubilius and V. Skorniakov. On Some Estimators of the Hurst Index of the Solution of SDE Driven by a Fractional Brownian Motion. *Statistics and Probability Letters* 109, 159-167, 2016.
- [30] Y. Hu and D. Nualart. Parameter Estimation for Fractional Ornstein-Uhlenbeck Processes. *Statistics and Probability Letters* 80, 1030-1038, 2010.
- [31] Y. Hu, D. Nualart and H. Zhou. Drift Parameter Estimation for Nonlinear Stochastic Differential Equations Driven by Fractional Brownian Motion. *Stochastics* 91, 8, 1067-1091, 2019.
- [32] Y. Kutoyants. *Identification of Dynamical Systems with Small Noise*. Springer, 1994.
- [33] Y. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer, 2004.
- [34] A. Lejay. Controlled Differential Equations as Young Integrals: A Simple Approach. *Journal of Differential Equations* 249, 1777-1798, 2010.
- [35] N. Marie. A Generalized Mean-Reverting Equation and Applications. *ESAIM: Probability and Statistics* 18, 799-828, 2014.
- [36] N. Marie. Nonparametric Estimation of the Trend in Reflected Fractional SDE. *Statistics and Probability Letters*, DOI: 10.1016/j.spl.2019.108659, 2019.
- [37] N. Marie and P. Raynaud de Fitte. Almost Periodic and Periodic Solutions of Differential Equations Driven by the Fractional Brownian Motion with Statistical Application. *Working paper* arXiv: 2003.05800, 2020.
- [38] M.N. Mishra and B.L.S. Prakasa Rao. Nonparametric Estimation of Trend for Stochastic Differential Equations Driven by Fractional Brownian Motion. *Stat. Inference. Stoch. Process.* 14, 2, 101-109, 2011.
- [39] Y. Mishura and K. Ralchenko. On Drift Parameter Estimation in Models with Fractional Brownian Motion by Discrete Observations. *Austrian Journal of Statistics* 43, 3-4, 217-228, 2014.
- [40] A. Neuenkirch and S. Tindel. A Least Square-Type Procedure for Parameter Estimation in Stochastic Differential Equations with Additive Fractional Noise. *Stat. Inference Stoch. Process.* 17, 1, 99-120, 2014.
- [41] I. Nourdin and F. Viens. Density Formula and Concentration Inequalities with Malliavin Calculus. *Electronic Journal of Probability* 14, 78, 2287-2309, 2009.

- [42] D. Nualart. *The Malliavin Calculus and Related Topics*. Springer, 2006.
- [43] R. Overgaard, N. Jonsson, C. Tornøe and H. Madsen. Non-Linear Mixed Effects Models with Stochastic Differential Equations: Implementation of an Estimation Algorithm. *J. Pharmacokinet. Pharmacodyn.* 32, 85-107, 2005.
- [44] U. Picchini, A. De Gaetano and S. Ditlevsen. Stochastic Differential Mixed-Effects Models. *Scand. J. Statist.* 37, 67-90, 2010.
- [45] U. Picchini and S. Ditlevsen. Practical Estimation of High Dimensional Stochastic Differential Mixed-Effects Models. *Comput. Statist. Data Anal.* 55, 1426-1444, 2011.
- [46] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion. Third Edition*. A Series of Comprehensive Studies in Mathematics 293, Springer, 1999.
- [47] F. Panloup, S. Tindel and M. Varvenne. A General Drift Estimation Procedure for Stochastic Differential Equations with Additive Fractional Noise. *Electronic Journal of Statistics* 14, 1, 1075-1136, 2020.
- [48] B. Sausseureau. Nonparametric Inference for Fractional Diffusion. *Bernoulli* 20, 2, 878-918, 2014.
- [49] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer Series in Statistics. Springer, New York, 2009.
- [50] C.A. Tudor and F. Viens. Statistical Aspects of the Fractional Stochastic Calculus. *The Annals of Statistics* 35, 3, 1183-1212, 2007.
- [51] C.A. Tudor and F. Viens. Variations and Estimators for Self-Similarity Parameters via Malliavin Calculus. *The Annals of Probability* 37, 6, 2093-2134, 2009.

\*LABORATOIRE MAP5 UMR CNRS 8145, UNIVERSITÉ DE PARIS, PARIS, FRANCE  
*Email address:* `fabienne.comte@parisdescartes.fr`

†LABORATOIRE MODAL'X, UNIVERSITÉ PARIS NANTERRE, NANTERRE, FRANCE  
*Email address:* `nmarie@parisnanterre.fr`

†ESME SUDRIA, PARIS, FRANCE  
*Email address:* `nicolas.marie@esme.fr`