Nadaraya-Watson Estimator for I.I.D. Paths of Diffusion Processes

Nicolas Marie, Amélie Rosier

NADARAYA-WATSON ESTIMATOR FOR I.I.D. PATHS OF DIFFUSION PROCESSES

NICOLAS MARIE[†] AND AMÉLIE ROSIER[◇]

Abstract. This paper deals with a nonparametric Nadaraya-Watson estimator $\widehat{b}$ of the drift function computed from independent continuous observations of a diffusion process. Risk bounds on $\widehat{b}$ and its discrete-time approximation are established. The paper also deals with extensions of the PCO and leave-one-out cross validation bandwidth selection methods for $\widehat{b}$. Finally, some numerical experiments are provided.

MSC2020 subject classifications. 62G05 ; 62M05.

1. Introduction

Consider the stochastic differential equation

$$X_t = x_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s, \quad (1)$$

where $b, \sigma : \mathbb{R} \to \mathbb{R}$ are two continuous functions and $W = (W_t)_{t\in[0,T]}$ is a Brownian motion.

Since the 1980's, the statistical inference for stochastic differential equations (SDE) has been widely investigated by many authors in the parametric and in the nonparametric frameworks. Classically (see Hoffmann [20], Kessler [22], Kutoyants [24], Dalalyan [9], Comte et al. [6], etc.), the estimators of the drift function are computed from one path of the solution to Equation (1) and converge when T goes to infinity. The existence and the uniqueness of the stationary solution to Equation (1) are then required, and obtained thanks to restrictive conditions on b.

Let $\mathcal{I} : (x, w) \mapsto \mathcal{I}(x, w)$ be the Itô map for Equation (1) and, for $N \in \mathbb{N}^*$ copies $W^1, \dots, W^N$ of W, consider $X^i = \mathcal{I}(x_0, W^i)$ for every $i \in \{1, \dots, N\}$. The estimation of the drift function b from continuous-time and discrete-time observations of $(X^1, \dots, X^N)$ is a functional data analysis problem already investigated in the parametric framework (see Ditlevsen and De Gaetano [16], Overgaard et al. [27], Picchini, De Gaetano and Ditlevsen [28], Picchini and Ditlevsen [29], Comte, Genon-Catalot and Samson [7], Delattre and Lavielle [12], Delattre, Genon-Catalot and Samson [11], Dion and Genon-Catalot [15], Delattre, Genon-Catalot and Larédo [10], etc.) and more recently in the nonparametric framework (see Comte and Genon-Catalot [4, 5], Della Maestra and Hoffmann [13] and Denis et al. [14]). In [13], the authors also study a Nadaraya-Watson type estimator, presented bellow, of the drift function in McKean-Vlasov models.

Under the appropriate conditions on b and σ recalled at Section 2, the distribution of X_t has a density $p_t(x_0, .)$ for every $t \in (0, T]$, and then one can define

$$f(x) := \frac{1}{T - t_0} \int_{t_0}^T p_t(x_0, x)dt \ ; \ x \in \mathbb{R}$$

Key words and phrases. Diffusion processes ; Nonparametric drift estimation ; Nadaraya-Watson estimator ; PCO method ; Cross validation.

for any $t_0 > 0$. Clearly, f is a density function:

$$\int_{-\infty}^{\infty} f(x)dx = \frac{1}{T-t_0}\int_{t_0}^{T}\int_{-\infty}^{\infty} p_t(x_0,x)dxdt = 1.$$

Let $K : \mathbb{R} \to \mathbb{R}$ be a kernel (i.e. an integrable function such that $\int K = 1$) and consider $K_h(x) := h^{-1}K(h^{-1}x)$ with $h \in (0,1]$. In the spirit of Comte and Genon-Catalot [4, 5], our paper deals first with the continuous-time Nadaraya-Watson estimator

$$\widehat{b}_{N,h}(x) := \frac{\widehat{bf}_{N,h}(x)}{\widehat{f}_{N,h}(x)} \tag{2}$$

of the drift function b, where

$$\widehat{f}_{N,h}(x) := \frac{1}{N(T-t_0)}\sum_{i=1}^{N}\int_{t_0}^{T} K_h(X_t^i - x)dt \tag{3}$$

is an estimator of f and

$$\widehat{bf}_{N,h}(x) := \frac{1}{N(T-t_0)}\sum_{i=1}^{N}\int_{t_0}^{T} K_h(X_t^i - x)dX_t^i \tag{4}$$

is an estimator of bf. From independent copies of X continuously observed on $[0,T]$, $\widehat{b}_{N,h}$ is a natural extension of the Nadaraya-Watson estimator already well-studied in the regression framework (see Comte [2], Chapter 4 or Györfi et al. [18], Chapter 5). The paper also deals with a discrete-time approximate of the previous Nadaraya-Watson estimator:

$$\widehat{b}_{n,N,h}(x) := \frac{\widehat{bf}_{n,N,h}(x)}{\widehat{f}_{n,N,h}(x)}, \tag{5}$$

where

$$\widehat{f}_{n,N,h}(x) := \frac{1}{nN}\sum_{i=1}^{N}\sum_{j=0}^{n-1} K_h(X_{t_j}^i - x) \tag{6}$$

is an estimator of f,

$$\widehat{bf}_{n,N,h}(x) := \frac{1}{N(T-t_0)}\sum_{i=1}^{N}\sum_{j=0}^{n-1} K_h(X_{t_j}^i - x)(X_{t_{j+1}}^i - X_{t_j}^i) \tag{7}$$

is an estimator of bf, and $(t_0, t_1, \dots, t_n)$ is the dissection of $[t_0, T]$ such that $t_j = t_0 + (T-t_0)j/n$ for every $j \in \{1, \dots, n\}$. Finally, our paper deals with a risk bound on the adaptive double bandwidths Nadaraya-Watson's estimator

$$\widehat{b}_{N,\widehat{h},\widehat{h}'}(x) := \frac{\widehat{bf}_{N,\widehat{h}}(x)}{\widehat{f}_{N,\widehat{h}'}(x)},$$

where $\widehat{h}$ (resp. $\widehat{h}'$) is selected via a penalized comparison to overfitting (PCO) type criterion for its numerator (resp. denominator). However, in the nonparametric regression framework, it is established in Comte and Marie [8] that the leave-one-out cross-validation (looCV) bandwidth selection method for Nadaraya-Watson's estimator is numerically more satisfactory than two alternative procedures based on Goldenshluger-Lepski's method and on the PCO method. For this reason, an extension of the looCV method to $\widehat{b}_{n,N,h}$ is also provided, with numerical experiments, even if it seems difficult to establish a risk bound on the associated adaptive estimator.

Now, let us compare $\widehat{b}_{N,h}$ with the estimator of Della Maestra and Hoffmann [13] restricted to our framework:

$$\widehat{b}_{N,\mathbf{h}}(x) := \frac{\displaystyle\sum_{i=1}^{N}\int_0^T L_{h_1}(\tau - t)K_{h_2}(X_t^i - x)dX_t^i}{\displaystyle\sum_{i=1}^{N} K_{h_3}(X_\tau^i - x)},$$

where $L : \mathbb{R} \to \mathbb{R}$ is another kernel, $h_1, h_2, h_3 \in (0, 1]$, $\mathbf{h} = (h_1, h_2, h_3)$ and $\tau \in (0, T)$. In [13], the authors provide a nice risk bound on the adaptive estimator obtained by selecting (h_1, h_2) (resp. h_3) via a Goldenshluger-Lepski type procedure on the numerator (resp. the denominator) of $\widehat{b}_{N,\mathbf{h}}$. As mentioned above, in Comte and Marie [8], it has been established that in the nonparametric regression framework, this approach is numerically less satisfactory than the looCV method. However, the looCV method provided in our paper doesn't extend to $\widehat{b}_{N,\mathbf{h}}$ because it cannot be written easily as a linear combination. Note also that even if it is numerically less satisfactory than the looCV method provided at Subsection 5.2, the PCO type method provided in our paper at Subsection 5.1 is easier to implement and numerically faster than a Goldenshluger-Lepski type method because, as in the nonparametric regression framework, the criterion to minimize depends on one variable instead of two, and because there is no constant to calibrate. For technical reasons explained at Section 6, the condition $(Nh^3)^{-1} \leqslant 1$ is required on the bandwidths collection to establish a risk bound on our PCO based adaptive estimator of bf, when Della Maestra and Hoffmann only need the condition $\log(N)^2(Nh)^{-1} \leqslant 1$ to establish a risk bound on their Goldenshluger-Lepski based adaptive estimator of $bp_\tau(x_0, .)$ in [13]. However, Remark 5.4 explains why the condition $(Nh^3)^{-1} \leqslant 1$ on the bandwidths collection is not that uncomfortable. Finally, under similar conditions on b, σ and K, the rate of convergence of our continuous-time Nadaraya-Watson estimator is of same order than the rate of convergence of the estimator of Della Maestra and Hoffmann [13] in the nonadaptive case. There is no discrete-time approximate of $\widehat{b}_{N,\mathbf{h}}$ studied in [13].

Finally, even if they deal with a different type of nonparametric estimators, let us say few words on the recent papers of Comte and Genon-Catalot [4] and Denis et al. [14]. On the one hand, in [4], the authors extend to the diffusion processes framework, for continuous-time observations, the least squares projection estimator already well studied in the regression framework (see Cohen et al. [1], Comte and Genon-Catalot [3], etc.). In particular, they provide a model selection procedure and establish a risk bound on the associated adaptive estimator. As explained at Section 3, in the nonadaptive case, the variance term of their estimator is comparable with the variance term of $\widehat{b}_{N,h}$, but as in the nonparametric regression framework, the rate of convergence of the least squares projection estimator depends on the regularity space associated to the projection basis. On the other hand, in [14], the authors focus on a projection least squares estimator computed from discrete-time observations and on a B-spline space. They provide a model selection procedure and prove both upper and lower bounds on the associated adaptive estimator.

Section 2 deals with the existence and the regularity of the density $p_t(x_0, .)$ of X_t for every $t \in (0, T]$, and with a Nikol'skii type condition fulfilled by f. Section 3 deals with a risk bound on the continuous-time Nadaraya-Watson estimator and Section 4 with a risk bound on its discrete-time approximate. Finally, Section 5 provides extensions of the PCO and looCV methods for the Nadaraya-Watson estimator studied in this paper. Some numerical experiments on the looCV based adaptive Nadaraya-Watson estimator are also provided. The proofs are postponed to Appendix A.

Notations and basic definitions:

- For every $A, B \in \mathbb{R}$ such that $A < B$, $C^0([A, B]; \mathbb{R})$ is equipped with the uniform norm $\|.\|_{\infty,A,B}$, and $C^0(\mathbb{R})$ is equipped with the uniform (semi-)norm $\|.\|_\infty$.
- For every $p \in \overline{\mathbb{N}}$, $C_{\rm b}^p(\mathbb{R}) := \cap_{j=0}^p\{\varphi \in C^p(\mathbb{R}) : \varphi^{(j)} \text{ is bounded}\}$.

- For every $p \geqslant 1$, $\mathbb{L}^p(\mathbb{R}, dx)$ is equipped with its usual norm $\|.\|_p$ such that

$$\|\varphi\|_p := \left(\int_{-\infty}^{\infty} \varphi(x)^p dx\right)^{1/p} \; ; \; \forall \varphi \in \mathbb{L}^p(\mathbb{R}, dx).$$

- $\mathbb{H}^2$ is the space of the processes $(Y_t)_{t\in[0,T]}$, adapted to the filtration generated by W, such that

$$\int_0^T \mathbb{E}(Y_t^2)dt < \infty.$$

- For a given kernel δ, the usual scalar product on $\mathbb{L}^2(\mathbb{R}, \delta(x)dx)$ is denoted by $\langle .,. \rangle_{2,\delta}$ and the associated norm by $\|.\|_{2,\delta}$.

2. Preliminaries: regularity of the density and estimates

This section deals with the existence and the regularity of the density $p_t(x_0, .)$ of X_t for every $t \in (0, T]$, with the Kusuoka-Stroock bounds on $(t, x) \mapsto p_t(x_0, x)$ and its derivatives, and then with a Nikol'skii type condition fulfilled by f.

In the sequel, in order to ensure the existence and the uniqueness of the (strong) solution to Equation (1), b and σ fulfill the following regularity assumption.

Assumption 2.1. *The functions b and σ are Lipschitz continuous.*

Now, assume that the solution X to Equation (1) fulfills the following assumption.

Assumption 2.2. *There exists $\beta \in \mathbb{N}^*$ such that, for any $t \in (0, T]$, the distribution of X_t has a β times continuously derivable density $p_t(x_0, .)$. Moreover, for every $x \in \mathbb{R}$,*

$$0 < p_t(x_0, x) \leqslant \frac{\mathfrak{c}_{2.2,1}}{t^{1/2}} \exp\left[-\mathfrak{m}_{2.2,1}\frac{(x - x_0)^2}{t}\right]$$

and

$$|\partial_x^\ell p_t(x_0, x)| \leqslant \frac{\mathfrak{c}_{2.2,2}(\ell)}{t^{q_2(\ell)}} \exp\left[-\mathfrak{m}_{2.2,2}(\ell)\frac{(x - x_0)^2}{t}\right] \; ; \; \forall \ell \in \{1, \dots, \beta\},$$

where all the constants are positive, depend on T, but not on t and x.

At Section 4, the following assumption on X is also required.

Assumption 2.3. *For any $x \in \mathbb{R}$, the function $t \in (0, T] \mapsto p_t(x_0, x)$ is continuously derivable. Moreover,*

$$|\partial_t p_t(x_0, x)| \leqslant \frac{\mathfrak{c}_{2.3,3}}{t^{q_3}} \exp\left[-\mathfrak{m}_{2.3,3}\frac{(x - x_0)^2}{t}\right] \; ; \; \forall t \in (0, T],$$

where $\mathfrak{c}_{2.3,3}$, $\mathfrak{m}_{2.3,3}$ and q_3 are three positive constants depending on T but not on t and x.

Let us provide some examples of diffusion processes categories satisfying Assumptions 2.2 and/or 2.3.

Examples:

(1) Assume that the functions b and σ belong to $C_{\mathrm{b}}^{\infty}(\mathbb{R})$, and that there exists $\alpha > 0$ such that

$$|\sigma(x)| > \alpha \; ; \; \forall x \in \mathbb{R}. \tag{8}$$

Then, by Kusuoka and Stroock [23], Corollary 3.25, X fulfills Assumptions 2.2 and 2.3.

(2) Assume that b is Lipschitz continuous (but not bounded) and that $\sigma \in C_{\mathrm{b}}^1(\mathbb{R})$. Assume also that σ satisfies the non-degeneracy condition (8) and that σ' is Hölder continuous. Then, by Menozzi et al. [26], Theorem 1.2, X fulfills Assumption 2.2 with $\beta = 1$ (but not necessarily Assumption 2.3). Note that the conditions required to apply Menozzi et al. [26], Theorem 1.2 are fulfilled by the so-called Ornstein-Uhlenbeck process, that is the solution to the Langevin equation:

$$X_t = x_0 - \theta\int_0^t X_s ds + \sigma W_t \; ; \; t \in \mathbb{R}_+, \tag{9}$$

where $\theta, \sigma > 0$ and $x_0 \in \mathbb{R}_+$. In this special case, since it is well-known that the solution to Equation (9) is a Gaussian process such that

$$\mathbb{E}(X_t) = x_0 e^{-\theta t} \quad \text{and} \quad \text{var}(X_t) = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t}) \ ; \forall t \in [0, T],$$

one can show that X also fulfills Assumption 2.3.

Remark 2.4. *Under Assumptions 2.1 and 2.2, for any $p \geqslant 1$ and any continuous function $\varphi : \mathbb{R} \to \mathbb{R}$ having polynomial growth, $t \in [0, T] \mapsto \mathbb{E}(|\varphi(X_t)|^p)$ is bounded. Indeed, for any $t \in [0, T]$,*

$$\begin{aligned}\mathbb{E}(|\varphi(X_t)|^p) &\leqslant \mathfrak{c}_1(1 + \mathbb{E}(|X_t|^{pq})) = \mathfrak{c}_1 \int_{-\infty}^{\infty}(1 + |x|^{pq})p_t(x_0, x)dx \\ &\leqslant \mathfrak{c}_1\mathfrak{c}_{2.2,1}\int_{-\infty}^{\infty}(1 + |t^{1/2}x + x_0|^{pq})e^{-\mathfrak{m}_{2.2,1}x^2}dx \leqslant \mathfrak{c}_2(1 \vee T^{pq/2})\end{aligned}$$

where

$$\mathfrak{c}_2 = \mathfrak{c}_1\mathfrak{c}_{2.2,1}\int_{-\infty}^{\infty}[1 + (|x| + |x_0|)^{pq}]e^{-\mathfrak{m}_{2.2,1}x^2}dx$$

and the constants $\mathfrak{c}_1, q > 0$ only depend on φ. Moreover,

$$\begin{aligned}\|\varphi\|_{p,f} &:= \int_{-\infty}^{\infty}|\varphi(x)|^p f(x)dx \\ &= \frac{1}{T - t_0}\int_{t_0}^{T}\mathbb{E}(|\varphi(X_t)|^p)dt \leqslant \mathfrak{c}_2(1 \vee T^{pq/2}).\end{aligned}$$

Then, $\varphi \in \mathbb{L}^p(\mathbb{R}, f(x)dx)$ and $\|\varphi\|_{p,f}$ is bounded by a constant which doesn't depend on t_0. In particular, the remark applies to b and σ with $q = 1$ by Assumption 2.1.

Finally, let us show that f fulfills a Nikol'skii type condition.

Corollary 2.5. *Under Assumption 2.1, $f(x) > 0$ for every $x \in \mathbb{R}$. Moreover, under Assumptions 2.1 and 2.2, there exists $\mathfrak{c}_{2.5} > 0$, depending on T but not on t_0, such that for every $\ell \in \{0, \dots, \beta - 1\}$ and $\theta \in \mathbb{R}$,*

$$\int_{-\infty}^{\infty}[f^{(\ell)}(x + \theta) - f^{(\ell)}(x)]^2 dx \leqslant \frac{\mathfrak{c}_{2.5}}{t_0^{2q_2(\ell+1)}}(\theta^2 + |\theta|^3).$$

Remark 2.6. *Assumption 2.2, Assumption 2.3 and Corollary 2.5 are crucial in the sequel, but t_0 has to be chosen carefully to get reasonable risk bounds on the estimators $\widehat{b}_{N,h}$ and $\widehat{b}_{n,N,h}$. Indeed, the behavior of the Kusuoka-Stroock bounds on $(t, x) \mapsto p_t(x_0, x)$ and its derivatives is singular at point $(0, x_0)$. This is due to the fact that the distribution of X at time 0 is a Dirac measure while that it has a smooth density with respect to Lebesgue's measure for every $t \in (0, T]$. Moreover, since X is not a stationary process in general, the Kusuoka-Stroock bounds on $(t, x) \mapsto p_t(x_0, x)$ and its derivatives explode when $T \to \infty$. The same difficulty appears with the estimators studied in Comte and Genon-Catalot [4] and in Della Maestra and Hoffmann [13]. So, it is recommended to take T as small as possible in practice. In the sequel, only the dependence in t_0 is tracked in the risk bounds derived from Assumption 2.2, Assumption 2.3 and Corollary 2.5 because it is specific to our approach. Finally, these risk bounds only depend on t_0 through a multiplicative constant of order $1/\min\{t_0^{\alpha}, T - t_0\}$ ($\alpha > 0$). So, to take $t_0 \in [1, T - 1]$ when $T > 1$ gives constants not depending on t_0.*

3. Risk bound on the continuous-time Nadaraya-Watson estimator

This section deals with risk bounds on $\widehat{f}_{N,h}$, $\widehat{bf}_{N,h}$, and then on the Nadaraya-Watson estimator $\widehat{b}_{N,h}$.

In the sequel, the kernel K fulfills the following usual assumptions.

Assumption 3.1. *The kernel K is symmetric, continuous and belongs to $\mathbb{L}^2(\mathbb{R}, dx)$.*

Assumption 3.2. *There exists $\upsilon \in \mathbb{N}^*$ such that*

$$\int_{-\infty}^{\infty} |z^{\upsilon+1}K(z)|dz < \infty \quad \textit{and} \quad \int_{-\infty}^{\infty} z^{\ell}K(z)dz = 0 \ ; \forall \ell \in \{1,\dots,\upsilon\}.$$

About the construction of kernels fulfilling both Assumptions 3.1 and 3.2, the reader can refer to Comte [2], Proposition 2.10. The following proposition provides a risk bound on $\widehat{f}_{N,h}$ (see (3)).

Proposition 3.3. *Under Assumptions 2.1, 2.2, 3.1 and 3.2 with $\upsilon = \beta$,*

$$\mathbb{E}(\|\widehat{f}_{N,h} - f\|_2^2) \leqslant \mathfrak{c}_{3.3}(t_0)h^{2\beta} + \frac{\|K\|_2^2}{Nh}$$

with

$$\mathfrak{c}_{3.3}(t_0) = \frac{\mathfrak{c}_{2.5}}{|(\beta-2)!|^2 t_0^{2q_2(\beta)}} \left(\int_{-\infty}^{\infty} |z|^{\beta}(1+|z|^{1/2})|K(z)|dz\right)^2.$$

Note that thanks to Proposition 3.3, the bias-variance tradeoff is reached by (the risk bound on) $\widehat{f}_{N,h}$ when h is of order $N^{-1/(2\beta+1)}$, leading to a rate of order $N^{-2\beta/(2\beta+1)}$. Moreover, by Remark 2.6, to take $t_0 \geqslant 1$ when $T > 1$ gives

$$\mathbb{E}(\|\widehat{f}_{N,h} - f\|_2^2) \leqslant \mathfrak{c}_{3.3}h^{2\beta} + \frac{\|K\|_2^2}{Nh} \quad \text{with} \quad \mathfrak{c}_{3.3} = \frac{\mathfrak{c}_{2.5}}{|(\beta-2)!|^2} \left(\int_{-\infty}^{\infty} |z|^{\beta}(1+|z|^{1/2})|K(z)|dz\right)^2.$$

Note also that in the risk bound on $\widehat{f}_{N,h}$ of Proposition 3.3, only the control of the bias term depends on T, trough the constant $\mathfrak{c}_{2.5}$, depending itself on the constants $\mathfrak{c}_{2.2,2}(\ell)$, $\ell \in \{1,\dots,\beta\}$, involved in the Kusuoka-Strook bounds (see Assumption 2.2). Indeed, except in the special case of the Ornstein-Uhlenbeck process which is stationary, for all the examples of diffusion processes fulfilling Assumption 2.2 (see Kusuoka and Stroock [23] and Menozzi et al. [26]), the constants $\mathfrak{c}_{2.2,2}(\ell)$, $\ell \in \{1,\dots,\beta\}$, depend on T. The variance term doesn't depend on time at all.

In the sequel, $\mathbb{L}^2(\mathbb{R}, f(x)dx)$ is equipped with the f-weighted norm $\|.\|_{2,f}$ defined at the end of the introduction section. Let us recall that by Remark 2.4, for every $\varphi \in \mathbb{L}^2(\mathbb{R}, f(x)dx)$, $\|\varphi\|_{2,f}$ is bounded by a constant which doesn't depend on t_0.

The following proposition provides a risk bound on $\widehat{bf}_{N,h}$ (see (4)).

Proposition 3.4. *Under Assumptions 2.1 and 3.1,*

$$\mathbb{E}(\|\widehat{bf}_{N,h} - bf\|_2^2) \leqslant \|(bf)_h - bf\|_2^2 + \frac{\mathfrak{c}_{3.4}(t_0)}{Nh}$$

with $(bf)_h := K_h * (bf)$ and

$$\mathfrak{c}_{3.4}(t_0) = 2\|K\|_2^2 \left(\|b\|_{2,f}^2 + \frac{1}{T-t_0}\|\sigma\|_{2,f}^2\right).$$

Assume that bf is $\gamma \in \mathbb{N}^*$ times continuously derivable and that there exists $\varphi \in \mathbb{L}^1(\mathbb{R}, |z|^{\gamma-1}K(z)dz)$ such that, for every $\theta \in \mathbb{R}$ and $h \in (0,1]$,

$$\int_{-\infty}^{\infty} [(bf)^{(\gamma-1)}(x+h\theta) - (bf)^{(\gamma-1)}(x)]^2 dx \leqslant \varphi(\theta)h^2. \tag{10}$$

If in addition K fulfills Assumption 3.2 with $\upsilon = \gamma$, then $\|(bf)_h - bf\|_2^2$ is of order $h^{2\gamma}$, and by Proposition 3.4, the bias-variance tradeoff is reached by $\widehat{bf}_{N,h}$ when h is of order $N^{-1/(2\gamma+1)}$, leading to the rate $N^{-2\gamma/(2\gamma+1)}$. Moreover, by Remark 2.6, to take $t_0 \leqslant T-1$ when $T > 1$ gives

$$\mathbb{E}(\|\widehat{bf}_{N,h} - bf\|_2^2) \leqslant \|(bf)_h - bf\|_2^2 + \frac{\mathfrak{c}_{3.4}}{Nh} \quad \text{with} \quad \mathfrak{c}_{3.4} = 2\|K\|_2^2(\|b\|_{2,f}^2 + \|\sigma\|_{2,f}^2).$$

Note also that the variance term in this risk bound doesn't depend on T.

Finally, Propositions 3.3 and 3.4 allow to provide a risk bound on a truncated version of the Nadaraya-Watson estimator $\widehat{b}_{N,h}$ (see (2)).

Proposition 3.5. *Consider the 2 bandwidths (truncated) Nadaraya-Watson (2bNW) estimator*

$$\widehat{b}_{N,h,h'}(x) := \frac{\widehat{bf}_{N,h}(x)}{\widehat{f}_{N,h'}(x)}\mathbf{1}_{\widehat{f}_{N,h'}(x)>m/2} \quad \text{with} \quad h,h'>0,$$

and assume that $f(x)>m>0$ for every $x\in[A,B]$ ($m\in(0,1]$ and $A,B\in\mathbb{R}$ such that $A<B$). Under Assumptions 2.1, 2.2, 3.1 and 3.2 with $\upsilon=\beta$,

$$\mathbb{E}(\|\widehat{b}_{N,h,h'}-b\|_{f,A,B}^2) \leqslant \frac{\mathfrak{c}_{3.5}}{m^2}\left[\|(bf)_h-bf\|_2^2+\frac{\mathfrak{c}_{3.4}(t_0)}{Nh}+2\|b\|_{2,f}^2\left(\mathfrak{c}_{3.3}(t_0)(h')^{2\beta}+\frac{\|K\|_2^2}{Nh'}\right)\right]$$

with $\mathfrak{c}_{3.5}:=8(\|f\|_\infty\vee\|b^2f\|_\infty)$ and $\|\varphi\|_{f,A,B}:=\|\varphi\mathbf{1}_{[A,B]}\|_{2,f}$ for every $\varphi\in\mathbb{L}^2(\mathbb{R},f(x)dx)$.

Proposition 3.5 says that the risk of $\widehat{b}_{N,h,h'}$ can be controlled by the sum of those of $\widehat{bf}_{N,h}$ and $\widehat{f}_{N,h'}$ up to a multiplicative constant. Now, if K fulfills Assumption 3.2 with $\upsilon=\beta\vee\gamma$, and if bf satisfies Condition (10), then the risk bound on $\widehat{b}_{N,h,h'}$ is of order $h^{2\gamma}+(h')^{2\beta}+1/(Nh)+1/(Nh')$, and the bias-variance tradeoff is reached when h (resp. h') is of order $N^{-1/(2\gamma+1)}$ (resp. $N^{-1/(2\beta+1)}$), leading to the rate

$$N^{-2[(\frac{\gamma}{2\gamma+1})\wedge(\frac{\beta}{2\beta+1})]}=N^{-\frac{2(\beta\wedge\gamma)}{2(\beta\wedge\gamma)+1}},$$

which is of same order than the rate of the nonadaptive version of the estimator of Della Maestra and Hoffmann [13] (see their Theorem 15). Note also that to consider the 2bNW estimator is crucial to extend the PCO method to our framework in the spirit of Comte and Marie [8] (see Subsection 5.1). However, by taking $h=h'$ of order $N^{-1/(2(\beta\wedge\gamma)+1)}$, the bias-variance tradeoff is reached by the 1 bandwidth (truncated) Nadaraya-Watson estimator with the same rate. Finally, if $h=h'$ and σ is bounded, then the variance term in the risk bound of Proposition 3.5 is comparable to the variance term in the risk bound obtained by Comte and Genon-Catalot in [4] for their least squares projection estimator (see [4], Propositions 2.1 and 2.2). Indeed, for a d-dimensional projection space, the variance term in the risk bound of Comte and Genon-Catalot [4], Proposition 2.1 is of order d/N which is comparable to $1/(Nh)$. The rate of convergence of their least squares projection estimator depends on the regularity space associated to the projection basis but, as in the nonparametric regression framework, not on the regularity of f.

The limitation of our Proposition 3.5 is that m is unknown in general and must be replaced by an estimator as well. Most of the time, as stated in Comte [2], Chapter 4, the minimum of an estimator of f is taken to choose m in practice:

$$\widehat{m}_{N,h'}=\min\{\widehat{f}_{N,h'}(x)\ ;\ x\in[A,B]\}$$

for instance. A more naive way to solve this difficulty in practice is to take

$$m=m_N=\mathfrak{c}N^{-\frac{\varepsilon}{2}\cdot\frac{2(\beta\wedge\gamma)}{2(\beta\wedge\gamma)+1}}\xrightarrow[N\to\infty]{}0,$$

where $\mathfrak{c}>0$ is a fixed constant and $\varepsilon\in(0,1)$ is chosen as close as possible to 0. Under Assumption 2.2, by Corollary 2.5,

$$\exists N_0\in\mathbb{N}:\forall N>N_0,\ \forall x\in[A,B],\ f(x)>m_N.$$

So, by Proposition 3.5, when h (resp. h') is of order $N^{-1/(2\gamma+1)}$ (resp. $N^{-1/(2\beta+1)}$), $\widehat{b}_{N,h,h'}$ converges with the slightly degraded rate

$$N^{-(1-\varepsilon)\frac{2(\beta\wedge\gamma)}{2(\beta\wedge\gamma)+1}}.$$

This last comment remains true for Proposition 4.5 and Corollary 5.3.

4. Risk bound on the discrete-time approximate Nadaraya-Watson estimator

This section deals with risk bounds on $\widehat{f}_{n,N,h}$, $\widehat{bf}_{n,N,h}$, and then on the approximate Nadaraya-Watson estimator $\widehat{b}_{n,N,h}$.

In the sequel, in addition to Assumptions 3.1 and 3.2, K fulfills the following one.

Assumption 4.1. *The kernel K is two times continuously derivable on $\mathbb{R}$ and $K', K'' \in \mathbb{L}^2(\mathbb{R}, dx)$.*

Compactly supported kernels belonging to $C^2(\mathbb{R})$ or Gaussian kernels fulfill Assumption 4.1. The following proposition provides a risk bound on $\widehat{f}_{n,N,h}$ (see (6)).

Proposition 4.2. *Under Assumptions 2.1, 2.2, 2.3, 3.1, 3.2 with $\upsilon = \beta$, and 4.1, if*

$$\frac{1}{nh^2} \leqslant 1,$$

then there exists a constant $\mathfrak{c}_{4.2} > 0$, not depending on h, N, n and t_0, such that

$$\mathbb{E}(\|\widehat{f}_{n,N,h} - f\|_2^2) \leqslant \frac{\mathfrak{c}_{4.2}}{\min\{t_0^{2q_2(\beta)}, t_0^{2q_3}\}}\left(h^{2\beta} + \frac{1}{Nh} + \frac{1}{n^2}\right) + \frac{1}{Nnh^3}.$$

Assume that $\beta = 1$ (extreme case) and h is of order $N^{-1/3}$. As mentioned at Section 3, under this condition, the bias-variance tradeoff is reached by the continuous-time estimator of f. Then, the approximation error of $\widehat{f}_{n,N,h}$ is of order $1/n$, which is the order of the variance of the Brownian motion increments along the dissection $(t_0, t_1, \dots, t_n)$ of $[t_0, T]$. For this reason, the risk bound established in Proposition 4.2 is satisfactory. Moreover, by Remark 2.6, to take $t_0 \geqslant 1$ when $T > 1$ gives

$$\mathbb{E}(\|\widehat{f}_{n,N,h} - f\|_2^2) \leqslant \mathfrak{c}_{4.2}\left(h^{2\beta} + \frac{1}{Nh} + \frac{1}{n^2}\right) + \frac{1}{Nnh^3}.$$

The following proposition provides a risk bound on $\widehat{bf}_{n,N,h}$ (see (7)).

Proposition 4.3. *Consider $\varepsilon \in (0, 1)$. Under Assumptions 2.1, 2.2, 2.3, 3.1 and 4.1, if*

$$\frac{1}{nh^{2-\varepsilon}} \leqslant 1,$$

the kernel K belongs to $\mathbb{L}^4(\mathbb{R}, dx)$ and $z \mapsto zK'(z)$ belongs to $\mathbb{L}^2(\mathbb{R}, dx)$, then there exist a constant $\mathfrak{c}_{4.3} > 0$, not depending on ε, h, N, n and t_0, and a constant $\mathfrak{c}_{4.3}(\varepsilon) > 0$, depending on ε but not on h, N, n and t_0, such that

$$\mathbb{E}(\|\widehat{bf}_{n,N,h} - bf\|_2^2) \leqslant \frac{\mathfrak{c}_{4.3}}{\min\{t_0^{1/2}, t_0^{2q_3}, T - t_0\}}\left(\|(bf)_h - bf\|_2^2 + \frac{1}{Nh} + \frac{1}{n}\right) + \frac{\mathfrak{c}_{4.3}(\varepsilon)}{\min\{1, t_0^{(1-\varepsilon)/2}\}} \cdot \frac{1}{Nnh^{3+\varepsilon}}.$$

Remark 4.4. *Note that if b and σ are bounded, Proposition 4.3 can be improved. Precisely, with $\varepsilon = 0$ and without the additional conditions $K \in \mathbb{L}^4(\mathbb{R}, dx)$ and $z \mapsto zK'(z)$ belongs to $\mathbb{L}^2(\mathbb{R}, dx)$, the risk bound on $\widehat{bf}_{n,N,h}$ is of same order than in Proposition 4.2 (see Remark A.3 for details).*

Assume that bf fulfills Condition (10) with $\gamma = 1$ (extreme case), and that h is of order $N^{-1/3}$. Then, for $\varepsilon > 0$ as close as possible to 0, the approximation error of $\widehat{bf}_{n,N,h}$ is of order $N^{\varepsilon/3}/n$. If in addition b and σ are bounded, thanks to Remark 4.4, with $\varepsilon = 0$ and without the additional conditions $K \in \mathbb{L}^4(\mathbb{R}, dx)$ and $z \mapsto zK'(z)$ belongs to $\mathbb{L}^2(\mathbb{R}, dx)$, then the approximation error of $\widehat{bf}_{n,N,h}$ is of order $1/n$ as the error of $\widehat{f}_{n,N,h}$. Moreover, by Remark 2.6, to take $t_0 \in [1, T-1]$ when $T > 1$ gives

$$\mathbb{E}(\|\widehat{bf}_{n,N,h} - bf\|_2^2) \leqslant \mathfrak{c}_{4.3}\left(\|(bf)_h - bf\|_2^2 + \frac{1}{Nh} + \frac{1}{n}\right) + \frac{\mathfrak{c}_{4.3}(\varepsilon)}{Nnh^{3+\varepsilon}}.$$

Finally, Propositions 4.2 and 4.3 allow to provide a risk bound on a truncated version the approximate Nadaraya-Watson estimator $\widehat{b}_{n,N,h}$ (see (5)).

Proposition 4.5. *Consider $\varepsilon > 0$, $m \in (0,1]$, and assume that $f(x) > m > 0$ for every $x \in [A,B]$ ($A, B \in \mathbb{R}$ such that $A < B$). Under the assumptions of Proposition 4.3 and, in addition, Assumptions 2.3 and 4.1, there exist a constant $\mathfrak{c}_{4.5} > 0$, not depending on ε, A, B, h, N, n and t_0, and a constant $\mathfrak{c}_{4.5}(\varepsilon) > 0$, depending on ε but not on A, B, h, N, n and t_0, such that*

$$\mathbb{E}(\|\widetilde{b}_{n,N,h} - b\|_{f,A,B}^2) \leqslant \frac{\mathfrak{c}_{3.5}}{m^2 \min\{1, t_0^{(1-\varepsilon)/2}, t_0^{1/2}, t_0^{2q_2(\beta)}, t_0^{2q_3}, T - t_0\}} \times \left[\mathfrak{c}_{4.5}\left(\|(bf)_h - bf\|_2^2 + h^{2\beta} + \frac{1}{Nh} + \frac{1}{n}\right) + \frac{\mathfrak{c}_{4.5}(\varepsilon)}{Nnh^{3+\varepsilon}}\right]$$

with $\widetilde{b}_{n,N,h}(.) := \widehat{b}_{n,N,h}(.)\mathbf{1}_{\widehat{f}_{n,N,h}(.)>m/2}$.

The proof of Proposition 4.5 given Propositions 4.2 and 4.3 is almost the same than the proof of Proposition 3.5 given Propositions 3.3 and 3.4. Of course one can establish a risk bound on the discrete-time approximate 2bNW estimator, but to focus on the 1 bandwidth estimator is clearer and sufficient to introduce the looCV selection method based on discrete-time observations of $X^1, \dots, X^N$ at Subsection 5.2. Now, assume that bf satisfies Condition (10) with $\gamma = \beta$, and that K fulfills Assumption 3.2 with $\upsilon = \beta$. Then, $\|(bf)_h - bf\|_2^2$ is of order $h^{2\beta}$. For the sake of simplicity, assume also that b is bounded, and then let's take $\varepsilon = 0$ in Proposition 4.5. First, note that the minimization problem

$$\min_{h\in(0,\infty)}\left\{h^{2\beta} + \frac{1}{Nh} + \frac{1}{n} + \frac{1}{Nnh^3}\right\}$$

has unfortunately no explicit solutions. However, let us provide an upper-bound on the rate of our discrete-time estimator. Since $(nh^2)^{-1} \leqslant 1$, Proposition 4.5 says that the risk of $\widetilde{b}_{n,N,h}$ is at most of order $h^{2\beta} + 1/(Nh) + 1/n$. So, the optimal bandwidth for this bound is of order $N^{-1/(2\beta+1)}$, leading to the rate

$$N^{-\frac{2\beta}{2\beta+1}} + \frac{1}{n}.$$

Moreover, by taking a bandwidth of order $N^{-1/(2\beta+1)}$ such that $(nh^2)^{-1} \leqslant 1$, N is at most of order $n^{(2\beta+1)/2}$. So, clearly, the more f and bf are regular, the more N can be chosen freely with respect to n, and if N is of order $n^{(2\beta+1)/2}$, then the risk of $\widetilde{b}_{n,N,h}$ is at most of order $1/n$. Finally, note that if $\beta = 1$, for a bandwidth of order $N^{-1/3}$ such that $(nh^2)^{-1} \leqslant 1$, then

$$1/n \leqslant h^2 \propto N^{-2/3},$$

and the rate of $\widetilde{b}_{n,N,h}$ is of order $N^{-2/3}$ (the optimal rate).

5. Bandwidth selection and numerical experiments

This section deals with extensions of the PCO (see Lacour et al. [25]) and looCV methods to the Nadaraya-Watson estimator studied in this paper (see Subsections 5.2 and 5.1). Subsection 5.3 deals with some numerical experiments on the looCV based adaptive Nadaraya-Watson estimator which is, as explained in Comte and Marie [8] in the nonparametric regression framework, numerically more satisfactory than the PCO based one. However, and this is its main advantage, the PCO based adaptive Nadaraya-Watson estimator offers theoretical guarantees: an oracle inequality is established in Subsection 5.1. Note also that the PCO method is easier to implement and numerically faster than the Goldenshluger-Lepski method which has been extended by Della Maestra and Hoffmann in [13] for their estimator of the drift function in McKean-Vlasov models.

5.1. **An extension of the Penalized Comparison to Overfitting method.** Let $\mathcal{H}_N$ (resp. $\mathcal{H}'_N$) be a finite subset of $[h_0, 1]$ (resp. $[h'_0, 1]$), where $h_0 > 0$ and $(Nh_0^3)^{-1} \leqslant 1$ (resp. $h'_0 > 0$ and $(Nh'_0)^{-1} \leqslant 1$). Consider an additional kernel δ,

$$\widehat{h} \in \arg\min_{h\in\mathcal{H}_N}\{\|\widehat{bf}_{N,h} - \widehat{bf}_{N,h_0}\|_{2,\delta}^2 + \mathrm{pen}(h)\} \tag{11}$$

with

$$\text{pen}(h) := \frac{2}{(T-t_0)^2N^2}\sum_{i=1}^{N}\left\langle \int_{t_0}^{T}K_h(X_s^i-\cdot)dX_s^i, \int_{t_0}^{T}K_{h_0}(X_s^i-\cdot)dX_s^i\right\rangle_{2,\delta} \text{ ; } \forall h\in\mathcal{H}_N, \tag{12}$$

and

$$\widehat{h}'\in\arg\min_{h\in\mathcal{H}_N'}\{\|\widehat{f}_{N,h}-\widehat{f}_{N,h_0'}\|_2^2+\text{pen}'(h)\} \tag{13}$$

with

$$\text{pen}'(h) := \frac{2}{(T-t_0)^2N^2}\sum_{i=1}^{N}\left\langle \int_{t_0}^{T}K_h(X_s^i-\cdot)ds, \int_{t_0}^{T}K_{h_0'}(X_s^i-\cdot)ds\right\rangle_{2} \text{ ; } \forall h\in\mathcal{H}_N'.$$

This subsection deals with risk bounds on the adaptive estimators $\widehat{bf}_{N,\widehat{h}}(.)$ (see (3)), $\widehat{f}_{N,\widehat{h}'}(.)$ (see (4)) and

$$\widehat{b}_{N,\widehat{h},\widehat{h}'}(x)=\frac{\widehat{bf}_{N,\widehat{h}}(x)}{\widehat{f}_{N,\widehat{h}'}(x)}\mathbf{1}_{\widehat{f}_{N,\widehat{h}'}(x)>m/2} \text{ ; } x\in[A,B]$$

with the notations of Proposition 3.5. In the sequel, K, δ and σ fulfill the following technical assumption.

Assumption 5.1. *The kernels K and δ are continuously derivable on $\mathbb{R}$, the derivative of K belongs to $\mathbb{L}^2(\mathbb{R},dx)$, δ is positive and its derivative is bounded, and σ is bounded.*

Moreover, recall that under Assumptions 2.1 and 2.2, b^2 and σ^2 belong to $\mathbb{L}^1(\mathbb{R},f(x)dx)$ (see Remark 2.4).

Theorem 5.2. *Under Assumptions 2.1, 2.2, 3.1 and 5.1,*

(1) There exist two deterministic constants $\mathfrak{c}_{5.2,1},\mathfrak{c}_{5.2,2}>0$, not depending on N, such that for every $\vartheta\in(0,1)$ and $\lambda>0$, with probability larger than $1-\mathfrak{c}_{5.2,1}|\mathcal{H}_N|e^{-\lambda}$,

$$\|\widehat{bf}_{N,\widehat{h}}-bf\|_{2,\delta}^2\leqslant(1+\vartheta)\min_{h\in\mathcal{H}_N}\|\widehat{bf}_{N,h}-bf\|_{2,\delta}^2+\frac{\mathfrak{c}_{5.2,2}}{\vartheta}\left[\|(bf)_{h_0}-bf\|_{2,\delta}^2+\frac{(1+\lambda)^3}{N}\right].$$

(2) There exist two deterministic constants $\overline{\mathfrak{c}}_{5.2,1},\overline{\mathfrak{c}}_{5.2,2}>0$, not depending on N, such that for every $\vartheta\in(0,1)$ and $\lambda>0$, with probability larger than $1-\overline{\mathfrak{c}}_{5.2,1}|\mathcal{H}_N|e^{-\lambda}$,

$$\|\widehat{f}_{N,\widehat{h}'}-f\|_2^2\leqslant(1+\vartheta)\min_{h'\in\mathcal{H}_N}\|\widehat{f}_{N,h'}-f\|_2^2+\frac{\overline{\mathfrak{c}}_{5.2,2}}{\vartheta}\left[\|f_{h_0'}-f\|_2^2+\frac{(1+\lambda)^3}{N}\right].$$

Corollary 5.3. *Under Assumptions 2.1, 2.2, 3.1 and 5.1, if $f(x),\delta(x)>m>0$ for every $x\in[A,B]$ ($m\in(0,1]$ and $A,B\in\mathbb{R}$ such that $A<B$), then there exists a deterministic constant $\mathfrak{c}_{5.3}>0$, not depending on N, A and B, such that for every $\vartheta\in(0,1)$,*

$$\begin{aligned}\mathbb{E}(\|\widehat{b}_{N,\widehat{h},\widehat{h}'}-b\|_{f,A,B}^2)\leqslant\frac{2\mathfrak{c}_{3.5}(1\vee\|\delta\|_\infty)}{m^3}\Bigg[&(1+\vartheta)\min_{(h,h')\in\mathcal{H}_N\times\mathcal{H}_N'}\{\mathbb{E}(\|\widehat{bf}_{N,h}-bf\|_2^2)+\mathbb{E}(\|\widehat{f}_{N,h'}-f\|_2^2)\}\\&+\frac{\mathfrak{c}_{5.3}}{\vartheta}\left(\|(bf)_{h_0}-bf\|_2^2+\|f_{h_0'}-f\|_2^2+\frac{1}{N}\right)\Bigg].\end{aligned}$$

Corollary 5.3 says that the risk of the adaptive estimator $\widehat{b}_{N,\widehat{h},\widehat{h}'}$ is controlled by the sum of the minimal risks of

$$\widehat{bf}_{N,h} \quad\text{and}\quad \widehat{f}_{N,h'} \text{ ; } (h,h')\in\mathcal{H}_N,$$

up to a multiplicative constant and a negligible additive term.

Remark 5.4. *The condition $(Nh_0^3)^{-1}\leqslant 1$ on the bandwidths collection $\mathcal{H}_N$ is quite uncomfortable but not that much because if bf satisfies Condition (10) with $\gamma=\beta\geqslant 2$, then the (unknown) bandwidth h^* of order $N^{-1/(2\beta+1)}$ such that our estimator of bf reaches the bias-variance tradeoff (see Section 3) possibly belongs to $\mathcal{H}_N$. Indeed, there exists an unknown constant $\mathfrak{c}^*>0$ such that $h^*=\mathfrak{c}^*N^{-1/(2\beta+1)}$, and then*

$$\frac{1}{N(h^*)^3}=(\mathfrak{c}^*)^{-3}N^{\frac{2}{2\beta+1}(1-\beta)}\leqslant 1$$

for N large enough. Moreover, the proof of Proposition 5.2 remains true by replacing the condition $(Nh_0^3)^{-1} \leqslant 1$ by $(Nh_0^3)^{-1} \leqslant \mathfrak{m}$ with $\mathfrak{m} > 0$. So, even for $\beta = 1$, $(N(h^*)^3)^{-1} \leqslant (\mathfrak{c}^*)^{-3}$ and then h^* possibly belongs to $\mathcal{H}_N$ when $\mathfrak{m}$ is large enough.

Remark 5.5. *A nice choice for δ is the standard normal density:*

$$\delta(x) := \frac{e^{-x^2/2}}{\sqrt{2\pi}} \ ; \ \forall x \in \mathbb{R}.$$

First, δ obviously fulfills Assumption 5.1. Moreover, $\|\delta\|_\infty \leqslant 1$. Finally, by assuming that $f(x) > m_1$ for every $x \in [A, B]$ ($m_1 \in (0, 1]$ and $A, B \in \mathbb{R}$ such that $A < B$), since δ is continuous and positive on $\mathbb{R}$ ($\mathrm{supp}(\delta) = \mathbb{R}$), necessarily there exists $m_2 > 0$ such that $\delta(x) > m_2$ for every $x \in [A, B]$. So, $f(x), \delta(x) > m = m_1 \wedge m_2 > 0$ for every $x \in [A, B]$. Therefore, under Assumptions 2.1, 2.2, 3.1 and 5.1, by Corollary 5.3,

$$\mathbb{E}(\|\widehat{b}_{N,\widehat{h},\widehat{h}'} - b\|_{f,A,B}^2) \leqslant \frac{2\mathfrak{c}_{3.5}}{m^3}\left[(1+\vartheta)\min_{(h,h')\in\mathcal{H}_N\times\mathcal{H}'_N}\{\mathbb{E}(\|\widehat{bf}_{N,h} - bf\|_2^2) + \mathbb{E}(\|\widehat{f}_{N,h'} - f\|_2^2)\}\right.$$
$$\left. + \frac{\mathfrak{c}_{5.3}}{\vartheta}\left(\|(bf)_{h_0} - bf\|_2^2 + \|f_{h_0'} - f\|_2^2 + \frac{1}{N}\right)\right]$$

for every $\vartheta \in (0, 1)$.

5.2. **An extension of the leave-one-out cross-validation method.** First of all, note that the estimator $\widehat{b}_{n,N,h}$ (see (5)) can be written the following way:

$$\widehat{b}_{n,N,h}(x) = \sum_{i=1}^{N}\sum_{j=0}^{n-1}\omega_j^i(x)(X_{t_{j+1}}^i - X_{t_j^i})$$

with

$$\omega_j^i(x) := \frac{K_h(X_{t_j}^i - x)}{\displaystyle\sum_{k=1}^{N}\sum_{\ell=0}^{n-1}K_h(X_{t_\ell}^k - x)(t_{\ell+1} - t_\ell)} \ ; \ \forall (j, i) \in \{0, \dots, n-1\} \times \{1, \dots, N\},$$

satisfying

$$\sum_{i=1}^{N}\sum_{j=0}^{n-1}\omega_j^i(x)(t_{j+1} - t_j) = 1.$$

This nice (weighted) representation of $\widehat{b}_{n,N,h}(x)$ allows us to consider the following extension of the well-known looCV criterion in our framework:

$$\mathrm{CV}(h) := \sum_{i=1}^{N}\left[\sum_{j=0}^{n-1}\widehat{b}_{n,N,h}^{-i}(X_{t_j}^i)^2(t_{j+1} - t_j) - 2\sum_{j=0}^{n-1}\widehat{b}_{n,N,h}^{-i}(X_{t_j}^i)(X_{t_{j+1}}^i - X_{t_j}^i)\right]$$

with

$$\widehat{b}_{n,N,h}^{-i}(x) := \sum_{k\in\{1,\dots,N\}\setminus\{i\}}\sum_{j=0}^{n-1}\omega_j^k(x)(X_{t_{j+1}}^k - X_{t_j}^k) \ ; \ \forall i \in \{1, \dots, N\}.$$

Let us explain heuristically this extension of the looCV criterion. By assuming that $dX_t = Y_t dt$, Equation (1) leads to the regression model

$$Y_{t_j} = b(X_{t_j}) + \varepsilon_{t_j} \quad \text{with} \quad \int_0^{t_j}\varepsilon_s ds = \int_0^{t_j}\sigma(X_s)dW_s.$$

Then, a natural extension of the looCV criterion is

$$\begin{aligned}\mathrm{CV}^*(h) &:= \sum_{i=1}^{N}\sum_{j=0}^{n-1}(Y_{t_j}^i - \widehat{b}_{n,N,h}^{-i}(X_{t_j}^i))^2(t_{j+1}-t_j)\\ &\approx \mathrm{CV}(h) + \sum_{i=1}^{N}\sum_{j=0}^{n-1}(Y_{t_j}^i)^2(t_{j+1}-t_j)\end{aligned}$$

because $Y_{t_j}(t_{j+1}-t_j) \approx X_{t_{j+1}} - X_{t_j}$ thanks to the assumption $dX_t = Y_t dt$. Of course $\mathrm{CV}^*(h)$ is not satisfactory because the last term of its previous decomposition doesn't exist, but since this term doesn't depend on h, to minimize $\mathrm{CV}^*(.)$ is almost equivalent to minimize $\mathrm{CV}(.)$ which only involves quantities existing without the condition $dX_t = Y_t dt$.

5.3. **Numerical experiments.** Some numerical experiments on our estimation method are presented in this subsection. The discrete-time approximate Nadaraya-Watson (NW) (see (5)) estimator is computed on 4 datasets generated by SDEs with various types of vector fields. In each case, the bandwidth of the NW estimator is selected via the looCV method introduced at Subsection 5.2. On the one hand, two models with the same linear drift function are considered, but with an additive noise for the first one and a multiplicative noise for the second one:

1. The so-called Langevin equation, that is
$$X_t = x_0 - \int_0^t X_s ds + 0.1\cdot W_t.$$
2. The hyperbolic diffusion process, that is
$$X_t = x_0 - \int_0^t X_s ds + 0.1\int_0^t \sqrt{1+X_s^2}\,dW_s.$$

On the other hand, two models having the same non-linear drift function involving sin(.) are considered, but here again with an additive noise for the first one and a multiplicative noise for the second one:

3. The third model is defined by
$$X_t = x_0 - \int_0^t (X_s + \sin(4X_s))ds + 0.1\cdot W_t.$$
4. The fourth model is defined by
$$X_t = x_0 - \int_0^t (X_s + \sin(4X_s))ds + 0.1\int_0^t (2+\cos(X_s))dW_s.$$

The models and the estimator are implemented by taking $N = 200$, $n = 50$, $T = 5$, $x_0 = 2$, $t_0 = 1$ and K the Gaussian kernel $z \mapsto (2\pi)^{-1/2}e^{-z^2/2}$. For Models 1 and 2, the estimator of the drift function is computed for the bandwidths set

$$\mathcal{H}_1 := \{0.02k\ ;\ k = 1,\dots,10\},$$

and for Models 3 and 4, it is computed for the bandwidths set

$$\mathcal{H}_2 := \{0.01k\ ;\ k = 1,\dots,10\}.$$

Each set of bandwidths has been chosen after testing different values of h, to see with which ones the estimation performs better. To choose smaller values in the second set of bandwidths allows to check that the looCV method does not systematically select the smallest bandwidth for Models 3 and 4.

For each of the previous models, on Figures 1, 2, 3 and 4 respectively, the true drift function (in red) and the looCV adaptive NW estimator (in blue) are plotted on the left-hand side, and the beam of proposals is plotted in green on the right-hand side. On Figures 1 and 2, one can see that the drift function is well estimated by the looCV adaptive NW estimator, with a MSE equal to $2.95\cdot 10^{-4}$ for the Langevin equation and to $8.31\cdot 10^{-4}$ for the hyperbolic diffusion process. As presumed, the multiplicative noise in Model 2 slightly degrades the MSE. Note that when the bandwidth is too small, the estimation degrades, but the looCV method selects a higher value of h which performs better on the estimation. This means that, as expected, the looCV method selects a reasonable approximation of the bandwidth for which the NW estimator reaches the bias-variance tradeoff. On Figures 3 and 4, one can see that the drift function of Models 3 and 4 is still well estimated by our looCV adaptive NW estimator. However, note that there is a significant degradation of the MSE, which is equal to $2.89 \cdot 10^{-3}$ for Model 3 and to $9.26 \cdot 10^{-3}$ for Model 4. This is probably related to the *nonlinearity* of the drift function to estimate. Once again, to consider a multiplicative noise in Model 4 degrades the estimation quality with respect to Model 3. As for Models 1 and 2, note that the looCV does not systematically select the smallest bandwidth.

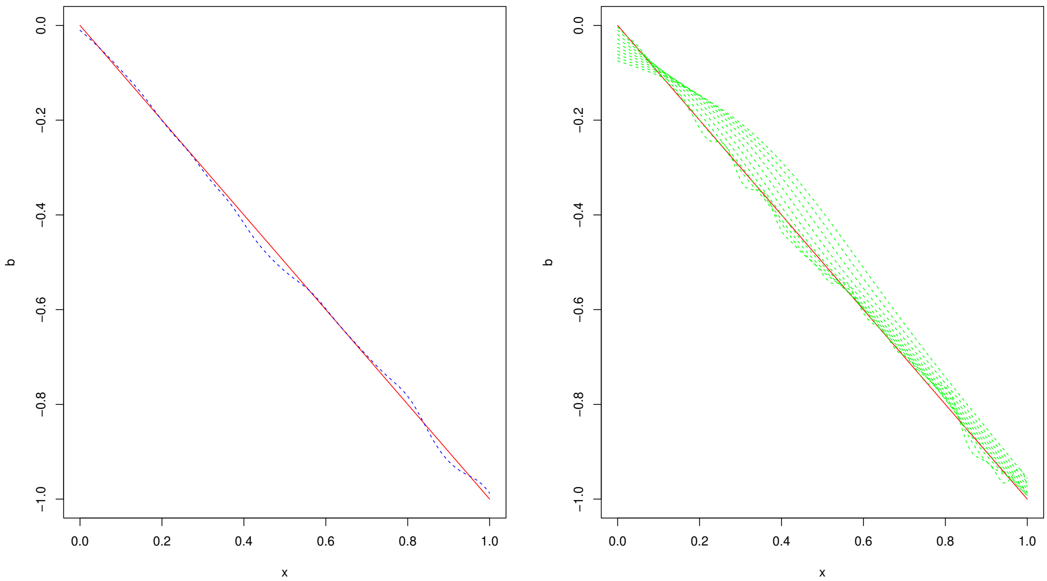

FIGURE 1. LooCV NW estimation for Model 1 (Langevin equation), $\widehat{h} = 0.04$.

For Model 1, at levels $n = 10, 20, \dots, 100$, Figure 5 shows the evolution of the MSE of the looCV adaptive NW estimator as a function of N. For this study, the value of N ranges from 20 to 200. Figure 5 shows that the MSE of our adaptive estimator remains low regardless to the value of (n, N) (from $4.50 \cdot 10^{-5}$ to $9.01 \cdot 10^{-3}$), decreases when N increases (for each n), and decreases when n increases (for a fixed N). This is consistent with the risk bounds of Section 4. Note also that for $N \geqslant 70$, there is no significant gain to take n larger than 30. For Model 3, Figure 6 shows the evolution of the MSE of the looCV adaptive NW estimator as a function of N and leads to the same conclusions than for Model 1. Note anyway that due to the *nonlinearity* of b, the MSE of our adaptive estimator reaches higher values (from $2.67 \cdot 10^{-4}$ to $4.49 \cdot 10^{-2}$) than for Model 1. Again, there is no significant gain to take n larger than 30, and above all larger than 70.

Finally, for each model, Table 1 gathers the mean MSE of 100 looCV NW estimations of the drift function as well as the mean MSE of the corresponding 100 oracle estimations. The mean MSEs are globally low, but significantly higher for the models with a nonlinear drift function (Models 3 and 4) than for the models with a linear one (Models 1 and 2). Moreover, for each drift function, the mean MSE is slightly degraded for the models with a multiplicative noise (Models 2 and 4) with respect to the models with an additive one (Models 1 and 3). Note also that for each model, the mean MSE of the looCV estimations is close to the mean MSE of the corresponding oracle estimations. This means that our looCV method performs well in practice.

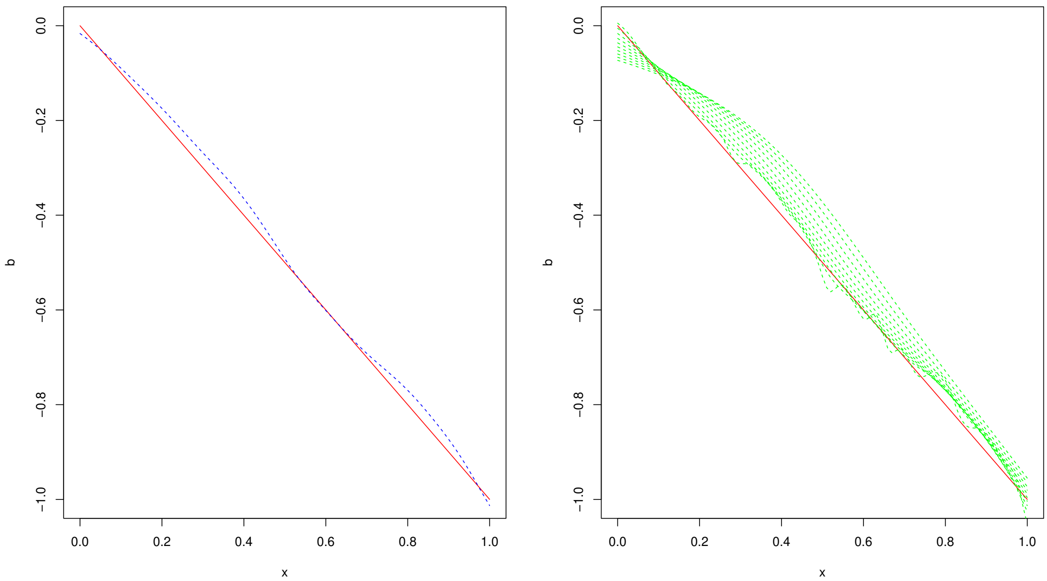

FIGURE 2. LooCV NW estimation for Model 2 (hyperbolic diffusion process), $\widehat{h} = 0.04$.

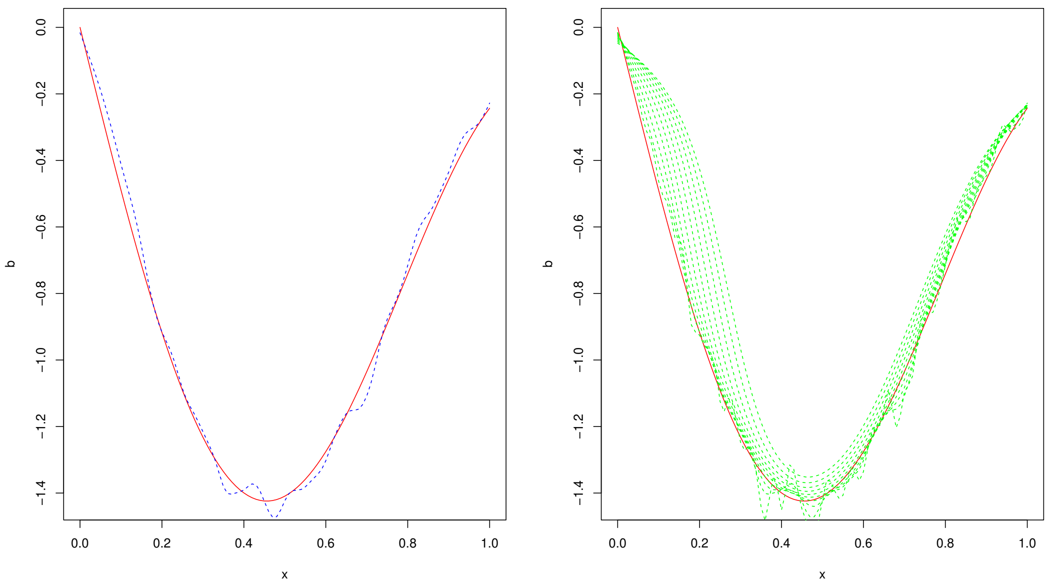

FIGURE 3. LooCV NW estimation for Model 3, $\widehat{h} = 0.02$.

Remark 5.6. *Note that to take $t_0 \geqslant 1$ (here $t_0 = 1$) is recommended even in numerical experiments. Indeed, for instance, the mean MSE of 10 looCV estimations for Model 1 is significantly lower with $t_0 = 1$ ($2.49 \cdot 10^{-4}$) than with $t_0 = 0$ ($3.82 \cdot 10^{-3}$).*

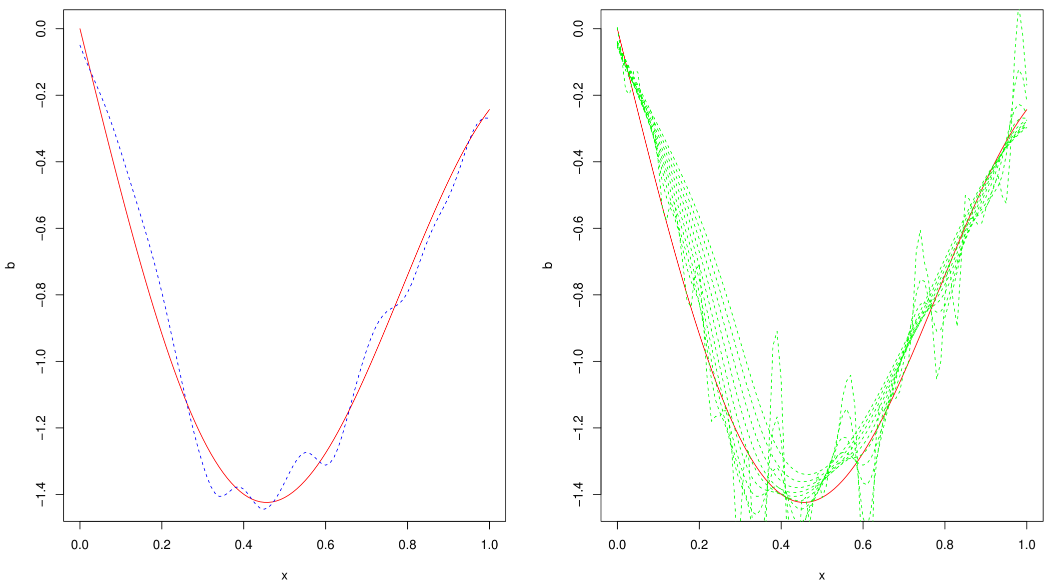

FIGURE 4. LooCV NW estimation for Model 4, $\widehat{h} = 0.06$.

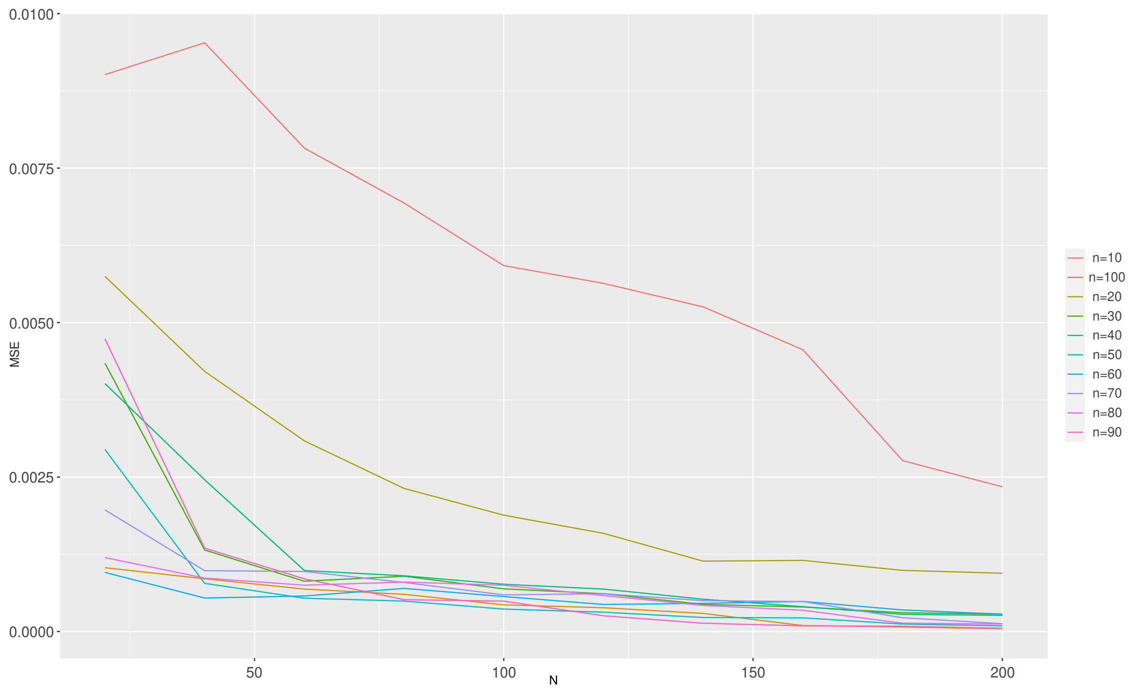

FIGURE 5. MSE of the looCV estimator with respect to N and n for Model 1.

6. CONCLUDING REMARKS

In this paper, first, a risk bound on our continuous-time Nadaraya-Watson estimator of b has been established. This bound is satisfactory because it leads to a rate of same order than in the classic non-parametric regression framework (see Comte [2], Chapter 4), and of same order than in Della Maestra

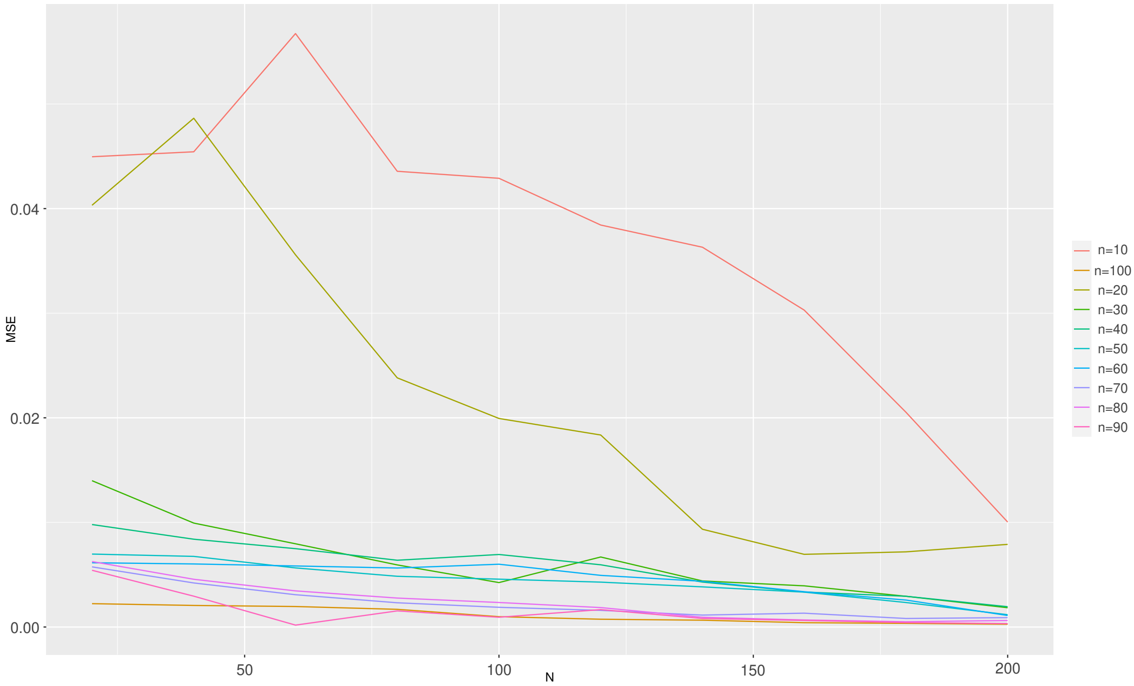

FIGURE 6. MSE of the looCV estimator with respect to N and n for Model 3.

	looCV	Oracle
Model 1	$3.03 \cdot 10^{-4}$	$2.67 \cdot 10^{-4}$
Model 2	$6.52 \cdot 10^{-4}$	$4.96 \cdot 10^{-4}$
Model 3	$2.45 \cdot 10^{-3}$	$1.99 \cdot 10^{-3}$
Model 4	$9.15 \cdot 10^{-3}$	$6.02 \cdot 10^{-3}$

TABLE 1. Mean MSEs of 100 looCV adaptive NW estimations compared to the oracle estimations.

and Hoffmann [13] for their estimator of the drift function in McKean-Vlasov models. Then, a risk bound on a discrete-time approximate estimator of b has been established too. The bound is satisfactory when b and σ are bounded, but a bit degraded when b is unbounded. To improve this bound will be the subject of future investigations.

In a second part, two bandwidth selection methods are provided. The first one is an extension of the PCO method to the 2bNW estimator of b in the spirit of Comte and Marie [8]. An oracle inequality is established but under the condition $(Nh^3)^{-1} \leqslant 1$ (instead of $(Nh)^{-1} \leqslant 1$) on the bandwidths collection. Unfortunately, it seems difficult to bypass this condition because of some constants involved in Bernstein's inequality and in the concentration inequality for U-statistics of Giné and Nickl [17] (see Subsection A.7), but as explained at Remark 5.4 this condition is not so bad. The second bandwidth selection method is an extension of the looCV procedure for the discrete-time approximate estimator written has a convex combination. As in the nonparametric regression framework, this method is numerically satisfactory but it seems difficult to establish a theoretical risk bound on the associated adaptive estimator.

Finally, the estimation of b has been only investigated in the case of one-dimensional diffusion processes because of its simplicity, but by following the same ideas than Halconruy and Marie used in the nonparametric regression framework in [19], the major part of the results of the present paper should be extendable to multidimensional diffusion processes.

References

[1] A. Cohen, M.A. Leviatan and D. Leviatan. *On the Stability and Accuracy of Least Squares Approximations.* Found. Comput. Math. 13, 819-834, 2013.

[2] F. Comte. *Estimation non-paramétrique. 2e édition.* Spartacus IDH, 2017.

[3] F. Comte and V. Genon-Catalot. *Regression Function Estimation on Non Compact Support as a Partly Inverse Problem.* The Annals of the Institute of Statistical Mathematics 72, 4, 1023-1054, 2020.

[4] F. Comte and V. Genon-Catalot. *Nonparametric Drift Estimation for i.i.d. Paths of Stochastic Differential Equations.* The Annals of Statistics 48, 6, 3336-3365, 2020.

[5] F. Comte and V. Genon-Catalot. *Drift Estimation on Non Compact Support for Diffusion Models.* Stoch. Proc. Appl. 134, 174-207, 2021.

[6] F. Comte, V. Genon-Catalot and Y. Rozenholc. *Penalized Nonparametric Mean Square Estimation of the Coefficients of Diffusion Processes.* Bernoulli 12, 2, 514-543, 2007.

[7] F. Comte, V. Genon-Catalot and A. Samson. *Nonparametric Estimation for Stochastic Differential Equations with Random Effects.* Stoch. Proc. Appl. 123, 2522-2551, 2013.

[8] F. Comte and N. Marie. *On a Nadaraya-Watson Estimator with Two Bandwidths.* Electronic Journal of Statistics 15, 1, 2566-2607, 2021.

[9] A. Dalalyan. *Sharp Adaptive Estimation of the Trend Coefficient for Ergodic Diffusion.* The Annals of Statistics 33, 6, 2507-2528, 2005.

[10] M. Delattre, V. Genon-Catalot and C. Larédo. *Parametric Inference for Discrete Observations of Diffusion Processes with Mixed Effects.* Stoch. Proc. Appl. 128, 1929-1957, 2018.

[11] M. Delattre, V. Genon-Catalot and A. Samson. *Maximum Likelihood Estimation for Stochastic Differential Equations with Random Effects.* Scand. J. Stat. 40, 322-343, 2013.

[12] 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.

[13] L. Della Maestra and M. Hoffmann. *Nonparametric Estimation for Interacting Particle Systems: McKean-Vlasov Models.* To appear in Probability Theory and Related Fields, 2021.

[14] C. Denis, C. Dion and M. Martinez. *A Ridge Estimator of the Drift from Discrete Repeated Observations of the Solutions of a Stochastic Differential Equation.* To appear in Bernoulli, 2021.

[15] C. Dion and V. Genon-Catalot. *Bidimensional Random Effect Estimation in Mixed Stochastic Differential Model.* Stat. Inference Stoch. Process. 19, 131-158, 2016.

[16] S. Ditlevsen and A. De Gaetano. *Mixed Effects in Stochastic Differential Equation Models.* REVSTAT 3, 137-153, 2005.

[17] E. Giné and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models.* Cambridge University Press, 2015.

[18] L. Györfi, M. Kohler, A. Krzyzak and H. Walk. *A Distribution-Free Theory of Nonparametric Regression.* Springer Series in Statistics. Springer-Verlag, New York, 2002.

[19] H. Halconruy and N. Marie. *Kernel Selection in Nonparametric Regression.* Mathematical Methods of Statistics 29, 1, 32-56, 2020.

[20] M. Hoffmann. *Adaptive Estimation in Diffusion Processes.* Stoch. Proc. Appl. 79, 135-163, 1999.

[21] C. Houdré and P. Reynaud-Bouret. *Exponential Inequalities, with Constants, for U-statistics of Order Two.* Stochastic Inequalities and Applications, vol. **56** of Progr. Proba, 55-69, Birkhauser, 2003.

[22] M. Kessler. *Simple and Explicit Estimating Functions for a Discretely Observed Diffusion Process.* Scandinavian Journal of Statistics 27, 1, 65-82, 2000.

[23] S. Kusuoka and D. Stroock. *Applications of the Malliavin Calculus, Part II.* J. Fac. Sci. Univ. Tokyo 32, 1-76, 1985.

[24] Y. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes.* Springer, 2004.

[25] C. Lacour, P. Massart and V. Rivoirard. *Estimator Selection: New Method with Applications to Kernel Density Estimation.* Sankhia A 79, 2, 298-335, 2017.

[26] S. Menozzi, A. Pesce and X. Zhang. *Density and Gradient Estimates for Non Degenerate Brownian SDEs with Unbounded Measurable Drift.* Journal of Differential Equations 272, 330-369, 2021.

[27] R. Overgaard, N. Jonsson, C. Tornoe, H. Madsen. *Non-Linear Mixed Effects Models with Stochastic Differential Equations: Implementation of an Estimation Algorithm.* J. Pharmacokinet. Pharmacodyn. 32, 85-107, 2005.

[28] U. Picchini, A. De Gaetano and S. Ditlevsen. *Stochastic Differential Mixed-Effects Models.* Scand. J. Stat. 37, 67-90, 2010.

[29] U. Picchini and S. Ditlevsen. *Practical Estimation of High Dimensional Stochastic Differential Mixed-Effects Models.* Comput. Statist. Data Anal. 55, 1426-1444, 2011.

Appendix A. Proofs

A.1. **Proof of Corollary 2.5.** First of all, since $p_t(x_0, x) > 0$ for every $(t, x) \in (0, T] \times \mathbb{R}$,

$$f(x) = \frac{1}{T - t_0} \int_{t_0}^{T} p_t(x_0, x) dt > 0$$

for every $x \in \mathbb{R}$. Consider $\ell \in \{0, \dots, \beta - 1\}$ and $\theta \in \mathbb{R}_+$. Thanks to the bound on $(t, x) \mapsto \partial_x^{\ell+1} p_t(x_0, x)$ given in Assumption 2.2,

$$\begin{aligned}
\|f^{(\ell)}(\cdot + \theta) - f^{(\ell)}\|_2^2 &= \int_{-\infty}^{\infty} [f^{(\ell)}(x + x_0 + \theta) - f^{(\ell)}(x + x_0)]^2 dx \\
&\leqslant \frac{1}{T - t_0} \int_{t_0}^{T} \int_{-\infty}^{\infty} (\partial_2^{\ell} p_t(x_0, x + x_0 + \theta) - \partial_2^{\ell} p_t(x_0, x + x_0))^2 dxdt \\
&\leqslant \frac{\theta^2}{T - t_0} \int_{t_0}^{T} \int_{-\infty}^{\infty} \sup_{z \in [x, x+\theta]} |\partial_2^{\ell+1} p_t(x_0, z + x_0)|^2 dxdt \\
&\leqslant \mathfrak{c}_{2.2,2}(\ell + 1)^2 \frac{\theta^2}{T - t_0} \int_{t_0}^{T} \frac{1}{t^{2q_2(\ell+1)}} \int_{-\infty}^{\infty} \sup_{z \in [x, x+\theta]} \exp\left(-2\mathfrak{m}_{2.2,2}(\ell + 1)\frac{z^2}{t}\right) dxdt \\
&= \mathfrak{c}_{2.2,2}(\ell + 1)^2 \frac{\theta^2}{T - t_0} \int_{t_0}^{T} \frac{1}{t^{2q_2(\ell+1)}} \times \\
&\qquad \left[\int_{-\infty}^{-\theta} \exp\left(-2\mathfrak{m}_{2.2,2}(\ell + 1)\frac{(x + \theta)^2}{t}\right) dx + \theta + \int_{0}^{\infty} \exp\left(-2\mathfrak{m}_{2.2,2}(\ell + 1)\frac{x^2}{t}\right) dx\right] dt \\
&\leqslant \frac{1}{t_0^{2q_2(\ell+1)}} \left[\mathfrak{c}_1 \theta^2 + \theta^3 \max_{k \in \{0, \dots, \beta-1\}} \mathfrak{c}_{2.2,2}(k + 1)^2\right]
\end{aligned}$$

with

$$\mathfrak{c}_1 = 2 \max_{k \in \{0, \dots, \beta-1\}} \left\{\mathfrak{c}_{2.2,2}(k + 1)^2 \int_{0}^{\infty} \exp\left(-2\mathfrak{m}_{2.2,2}(k + 1)\frac{x^2}{T}\right) dx\right\},$$

and the same way,

$$\begin{aligned}
\|f^{(\ell)}(\cdot - \theta) - f^{(\ell)}\|_2^2 &\leqslant \frac{\theta^2}{T - t_0} \int_{t_0}^{T} \int_{-\infty}^{\infty} \sup_{z \in [x-\theta, x]} |\partial_2^{\ell+1} p_t(x_0, z + x_0)|^2 dxdt \\
&\leqslant \mathfrak{c}_{2.2,2}(\ell + 1)^2 \frac{\theta^2}{T - t_0} \int_{t_0}^{T} \frac{1}{t^{2q_2(\ell+1)}} \times \\
&\qquad \left[\int_{-\infty}^{0} \exp\left(-2\mathfrak{m}_{2.2,2}(\ell + 1)\frac{x^2}{t}\right) dx + \theta + \int_{\theta}^{\infty} \exp\left(-2\mathfrak{m}_{2.2,2}(\ell + 1)\frac{(x - \theta)^2}{t}\right) dx\right] dt \\
&\leqslant \frac{1}{t_0^{2q_2(\ell+1)}} \left[\mathfrak{c}_1 \theta^2 + \theta^3 \max_{k \in \{0, \dots, \beta-1\}} \mathfrak{c}_{2.2,2}(k + 1)^2\right].
\end{aligned}$$

This concludes the proof.

A.2. **Proof of Proposition 3.3.** First of all, the bias of $\widehat{f}_{N,h}(x)$ is denoted by $\mathfrak{b}(x)$ and its variance by $\mathfrak{v}(x)$. Moreover, let us recall the bias-variance decomposition of the $\mathbb{L}^2$-risk of $\widehat{f}_{N,h}$:

$$\mathbb{E}(\|\widehat{f}_{N,h} - f\|_2^2) = \int_{-\infty}^{\infty} \mathfrak{b}(x)^2 dx + \int_{-\infty}^{\infty} \mathfrak{v}(x) dx.$$

On the one hand, let us find a suitable bound on the integrated variance of $\widehat{f}_{N,h}$. Since $X^1, \dots, X^N$ are i.i.d. copies of X, and thanks to Jensen's inequality,

$$\begin{aligned}
\mathfrak{v}(x) &= \operatorname{var}\left(\frac{1}{N(T - t_0)} \sum_{i=1}^{N} \int_{t_0}^{T} K_h(X_t^i - x) dt\right) \\
&= \frac{1}{N(T - t_0)^2} \operatorname{var}\left(\int_{t_0}^{T} K_h(X_t - x) dt\right) \leqslant \frac{1}{N} \mathbb{E}\left[\left(\int_{t_0}^{T} K_h(X_t - x) \frac{dt}{T - t_0}\right)^2\right] \\
&\leqslant \frac{1}{N(T - t_0)} \int_{t_0}^{T} \mathbb{E}(K_h(X_t - x)^2) dt = \frac{1}{N} \int_{-\infty}^{\infty} K_h(z - x)^2 f(z) dz.
\end{aligned}$$

Thus, since K is symmetric,

$$\begin{aligned}\int_{-\infty}^{\infty}\mathfrak{v}(x)dx &\leqslant \frac{1}{N}\int_{-\infty}^{\infty}f(z)\int_{-\infty}^{\infty}K_h(z-x)^2dxdz\\ &= \frac{1}{Nh}\left(\int_{-\infty}^{\infty}f(z)dz\right)\left(\int_{-\infty}^{\infty}K(x)^2dx\right) = \frac{\|K\|_2^2}{Nh}.\end{aligned}$$

On the other hand, let us find a suitable bound on the integrated squared-bias of $\widehat{f}_{N,h}(x)$. Since $X^1,\dots,X^N$ are i.i.d. copies of X,

$$\begin{aligned}\mathfrak{b}(x) &= \frac{1}{T-t_0}\int_{t_0}^{T}\mathbb{E}(K_h(X_t-x))dt - f(x)\\ &= \frac{1}{h}\int_{-\infty}^{\infty}K\left(\frac{z-x}{h}\right)f(z)dz - f(x)\\ &= \int_{-\infty}^{\infty}K(z)(f(hz+x)-f(x))dz.\end{aligned}$$

First, assume that $\beta = 1$. By Assumption 3.2, the generalized Minkowski inequality and Corollary 2.5,

$$\begin{aligned}\int_{-\infty}^{\infty}\mathfrak{b}(x)^2dx &\leqslant \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K(z)(f(hz+x)-f(x))dz\right)^2dx\\ &\leqslant \left[\int_{-\infty}^{\infty}K(z)\left(\int_{-\infty}^{\infty}(f(hz+x)-f(x))^2dx\right)^{1/2}dz\right]^2 \leqslant \mathfrak{c}_1(t_0)h^2\end{aligned}$$

with

$$\mathfrak{c}_1(t_0) = \frac{\mathfrak{c}_{2.5}}{t_0^{2q_2(1)}}\left(\int_{-\infty}^{\infty}|z|(1+|z|^{1/2})|K(z)|dz\right)^2.$$

Now, assume that $\beta \geqslant 2$. By the Taylor formula with integral remainder, for every $z \in \mathbb{R}$,

$$f(hz+x)-f(x) = \mathbf{1}_{\beta\geqslant 3}\sum_{\ell=1}^{\beta-2}\frac{(hz)^{\ell}}{\ell!}f^{(\ell)}(x) + \frac{(hz)^{\beta-1}}{(\beta-2)!}\int_0^1(1-\tau)^{\beta-2}f^{(\beta-1)}(\tau hz+x)d\tau.$$

Then, by Assumption 3.2, the generalized Minkowski inequality (two times) and Corollary 2.5,

$$\begin{aligned}\int_{-\infty}^{\infty}\mathfrak{b}(x)^2dx &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K(z)(f(hz+x)-f(x))dz\right)^2dx\\ &= \frac{h^{2(\beta-1)}}{|(\beta-2)!|^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}z^{\beta-1}K(z)\int_0^1(1-\tau)^{\beta-2}[f^{(\beta-1)}(\tau hz+x)-f^{(\beta-1)}(x)]d\tau dz\right)^2dx\\ &\leqslant \frac{h^{2(\beta-1)}}{|(\beta-2)!|^2}\times\\ &\qquad\left[\int_{-\infty}^{\infty}|z|^{\beta-1}|K(z)|\int_0^1(1-\tau)^{\beta-2}\left(\int_{-\infty}^{\infty}[f^{(\beta-1)}(\tau hz+x)-f^{(\beta-1)}(x)]^2dx\right)^{1/2}d\tau dz\right]^2\\ &\leqslant \frac{\mathfrak{c}_{2.5}h^{2\beta}}{|(\beta-2)!|^2t_0^{2q_2(\beta)}}\left(\int_{-\infty}^{\infty}|z|^{\beta-1}|K(z)|\int_0^1(1-\tau)^{\beta-2}[\tau|z|+(\tau|z|)^{3/2}]d\tau dz\right)^2 \leqslant \mathfrak{c}_2(t_0)h^{2\beta}\end{aligned}$$

with

$$\mathfrak{c}_2(t_0) = \frac{\mathfrak{c}_{2.5}}{|(\beta-2)!|^2t_0^{2q_2(\beta)}}\left(\int_{-\infty}^{\infty}|z|^{\beta}(1+|z|^{1/2})|K(z)|dz\right)^2.$$

This concludes the proof.

A.3. **Proof of Proposition 3.4.** First of all,

$$\mathbb{E}(\|\widehat{bf}_{N,h} - bf\|_2^2) = \int_{-\infty}^{\infty} \mathfrak{b}(x)^2 dx + \int_{-\infty}^{\infty} \mathfrak{v}(x) dx$$

where $\mathfrak{b}(x)$ (resp. $\mathfrak{v}(x)$) is the bias (resp. the variance) term of $\widehat{bf}_{N,h}(x)$ for any $x \in \mathbb{R}$. On the one hand, let us find a suitable bound on the integrated variance of $\widehat{bf}_{N,h}$. Since $X^1, \dots, X^N$ are i.i.d. copies of X,

$$\begin{aligned}
\mathfrak{v}(x) &= \text{var}\left(\frac{1}{N(T-t_0)}\sum_{i=1}^{N}\int_{t_0}^{T} K_h(X_t^i - x)dX_t^i\right) \leqslant \frac{1}{N(T-t_0)^2}\mathbb{E}\left[\left(\int_{t_0}^{T} K_h(X_t - x)dX_t\right)^2\right]\\
&\leqslant \frac{2}{N}\mathbb{E}\left[\left(\int_{t_0}^{T} K_h(X_t - x)b(X_t)\frac{dt}{T-t_0}\right)^2 + \frac{1}{(T-t_0)^2}\left(\int_{t_0}^{T} K_h(X_t - x)\sigma(X_t)dW_t\right)^2\right].
\end{aligned}$$

In the right-hand side of the previous inequality, Jensen's inequality on the first term and the isometry property for Itô's integral on the second one give

$$\begin{aligned}
\mathfrak{v}(x) &\leqslant \frac{2}{N(T-t_0)}\int_{t_0}^{T}\mathbb{E}[K_h(X_t - x)^2 b(X_t)^2]dt + \frac{2}{N(T-t_0)^2}\int_{t_0}^{T}\mathbb{E}[K_h(X_t - x)^2\sigma(X_t)^2]dt\\
&= \frac{2}{N}\int_{-\infty}^{\infty} K_h(z-x)^2 b(z)^2 f(z)dz + \frac{2}{N(T-t_0)}\int_{-\infty}^{\infty} K_h(z-x)^2\sigma(z)^2 f(z)dz.
\end{aligned}$$

Moreover, K is symmetric and $K \in \mathbb{L}^2(\mathbb{R}, dx)$ by Assumption 3.1, and $b, \sigma \in \mathbb{L}^2(\mathbb{R}, f(x)dx)$ by Remark 2.4. Then,

$$\begin{aligned}
\int_{-\infty}^{\infty}\mathfrak{v}(x)dx &\leqslant \frac{2}{N}\int_{\mathbb{R}^2} K_h(z-x)^2 b(z)^2 f(z)dzdx + \frac{2}{N(T-t_0)}\int_{\mathbb{R}^2} K_h(z-x)^2\sigma(z)^2 f(z)dzdx\\
&= \frac{2}{Nh}\int_{-\infty}^{\infty} b(z)^2 f(z)\int_{-\infty}^{\infty} K(x)^2 dxdz + \frac{2}{N(T-t_0)h}\int_{-\infty}^{\infty}\sigma(z)^2 f(z)\int_{-\infty}^{\infty} K(x)^2 dxdz\\
&\leqslant \frac{2\|K\|_2^2}{Nh}\left(\int_{-\infty}^{\infty} b(z)^2 f(z)dz + \frac{1}{T-t_0}\int_{-\infty}^{\infty}\sigma(z)^2 f(z)dz\right).
\end{aligned}$$

On the other hand, let us find a suitable bound on the integrated squared-bias of $\widehat{bf}_{N,h}(x)$. Again, since $X^1, \dots, X^N$ are i.i.d. copies of X, and since Itô's integral restricted to $\mathbb{H}^2$ is a martingale-valued map,

$$\begin{aligned}
\mathfrak{b}(x) &= \mathbb{E}\left[\frac{1}{N(T-t_0)}\sum_{i=1}^{N}\int_{t_0}^{T} K_h(X_t^i - x)dX_t^i\right] - b(x)f(x)\\
&= \frac{1}{T-t_0}\mathbb{E}\left(\int_{t_0}^{T} K_h(X_t - x)dX_t\right) - b(x)f(x)\\
&= \frac{1}{T-t_0}\left[\mathbb{E}\left(\int_{t_0}^{T} K_h(X_t - x)b(X_t)dt\right) + \mathbb{E}\left(\int_{t_0}^{T} K_h(X_t - x)\sigma(X_t)dW_t\right)\right] - b(x)f(x)\\
&= \frac{1}{T-t_0}\int_{t_0}^{T}\mathbb{E}(K_h(X_t - x)b(X_t))dt - b(x)f(x) = \int_{-\infty}^{\infty} K_h(z-x)b(z)f(z)dz - b(x)f(x).
\end{aligned}$$

Then,

$$\mathfrak{b}(x)^2 = ((bf)_h - bf)(x)^2 \quad \text{with} \quad (bf)_h = K_h * (bf).$$

Therefore, since f is bounded and b belongs to $\mathbb{L}^2(\mathbb{R}, f(x)dx)$ by Remark 2.4,

$$\int_{-\infty}^{\infty}\mathfrak{b}(x)^2 dx = \|bf - (bf)_h\|_2^2.$$

This concludes the proof.

A.4. **Proof of Proposition 3.5.** First of all,

$$\widehat{b}_{N,h,h'} - b = \left[\frac{\widehat{bf}_{N,h} - bf}{\widehat{f}_{N,h'}} + \left(\frac{1}{\widehat{f}_{N,h'}} - \frac{1}{f}\right) bf\right] \mathbf{1}_{\widehat{f}_{N,h'}(.)>m/2} - b\mathbf{1}_{\widehat{f}_{N,h'}(.)\leqslant m/2}.$$

Then,

$$\|\widehat{b}_{N,h,h'} - b\|_{f,A,B}^2 \leqslant 2\left\|\left[\frac{\widehat{bf}_{N,h} - bf}{\widehat{f}_{N,h'}} + \left(\frac{1}{\widehat{f}_{N,h'}} - \frac{1}{f}\right) bf\right] \mathbf{1}_{\widehat{f}_{N,h'}(.)>m/2}\right\|_{f,A,B}^2 + 2\|b\mathbf{1}_{\widehat{f}_{N,h'}(.)\leqslant m/2}\|_{f,A,B}^2.$$

Moreover, for any $x \in [A, B]$, since $f(x) > m$, for every $\omega \in \{\widehat{f}_{N,h'}(.) \leqslant m/2\}$,

$$|f(x) - \widehat{f}_{N,h'}(x,\omega)| \geqslant f(x) - \widehat{f}_{N,h'}(x,\omega) > m - \frac{m}{2} = \frac{m}{2}.$$

Thus,

$$\begin{aligned}
\|\widehat{b}_{N,h,h'} - b\|_{f,A,B}^2 &\leqslant \frac{8}{m^2}\|\widehat{bf}_{N,h} - bf\|_{2,f}^2 + \frac{8}{m^2}\|(f - \widehat{f}_{N,h'})b\|_{f,A,B}^2 + 2\|b\mathbf{1}_{|f(.)-\widehat{f}_{N,h'}(.)|>m/2}\|_{f,A,B}^2 \\
&\leqslant \frac{8}{m^2}\int_{-\infty}^{\infty}(\widehat{bf}_{N,h} - bf)(x)^2 f(x)dx \\
&\quad + \frac{8}{m^2}\int_A^B (f(x) - \widehat{f}_{N,h'}(x))^2 b(x)^2 f(x)dx \\
&\quad + 2\int_A^B b(x)^2 f(x)\mathbf{1}_{|f(x)-\widehat{f}_{N,h'}(x)|>m/2}dx.
\end{aligned}$$

Since f has a sub-Gaussian tail by Assumption 2.2, and since b has at most linear growth because it is Lipschitz continuous from $\mathbb{R}$ into itself (see Assumption 2.1), $b^2 f$ is bounded on $\mathbb{R}$. So,

$$\begin{aligned}
\|\widehat{b}_{N,h,h'} - b\|_{f,A,B}^2 &\leqslant \frac{8\|f\|_\infty}{m^2}\|\widehat{bf}_{N,h} - bf\|_2^2 \\
&\quad + \frac{8\|b^2 f\|_\infty}{m^2}\|\widehat{f}_{N,h'} - f\|_2^2 + 2\|b^2 f\|_\infty \int_{-\infty}^{\infty}\mathbf{1}_{|f(x)-\widehat{f}_{N,h'}(x)|>m/2}dx.
\end{aligned}$$

Therefore, thanks to Markov's inequality,

$$\begin{aligned}
\mathbb{E}(\|\widehat{b}_{N,h,h'} - b\|_{f,A,B}^2) &\leqslant \frac{8\|f\|_\infty}{m^2}\mathbb{E}(\|\widehat{bf}_{N,h} - bf\|_2^2) \\
&\quad + \frac{8\|b^2 f\|_\infty}{m^2}\mathbb{E}(\|\widehat{f}_{N,h'} - f\|_2^2) + \frac{8\|b^2 f\|_\infty}{m^2}\int_{-\infty}^{\infty}\mathbb{E}(|f(x) - \widehat{f}_{N,h'}(x)|^2)dx \\
&\leqslant \frac{8(\|f\|_\infty \vee \|b^2 f\|_\infty)}{m^2}[\mathbb{E}(\|\widehat{bf}_{N,h} - bf\|_2^2) + 2\mathbb{E}(\|\widehat{f}_{N,h'} - f\|_2^2)].
\end{aligned}$$

Propositions 3.4 and 3.3 allow to conclude.

A.5. **Proof of Proposition 4.2.** First of all, note that

$$\begin{aligned}
\mathbb{E}(\|\widehat{f}_{n,N,h} - f\|_2^2) &\leqslant 2\mathbb{E}(\|\widehat{f}_{N,h} - f\|_2^2) + 2\mathbb{E}(\|\widehat{f}_{N,h} - \widehat{f}_{n,N,h}\|_2^2) \\
&\leqslant 2\left[\mathfrak{c}_{3.3}(t_0)h^{2\beta} + \frac{1}{Nh} + \int_{-\infty}^{\infty}\mathbb{E}(\widehat{f}_{N,h}(x) - \widehat{f}_{n,N,h}(x))^2 dx\right. \\
&\quad \left. + \int_{-\infty}^{\infty}\mathrm{var}(\widehat{f}_{N,h}(x) - \widehat{f}_{n,N,h}(x))dx\right]
\end{aligned}$$

by Proposition 3.3, and note also that

$$\widehat{f}_{N,h}(x) - \widehat{f}_{n,N,h}(x) = \frac{1}{N(T - t_0)}\sum_{i=1}^{N}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}(K_{h,x}(X_t^i) - K_{h,x}(X_{t_j}^i))dt$$

with $K_{h,x}(.) := K_h(\cdot - x)$. On the one hand, for every $s, u \in [t_0, T]$ such that $s \leqslant u$, by Itô's formula, Jensen's inequality, the isometry property for Itô's integral and Remark 2.4,

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathbb{E}[(K_{h,x}(X_u) - K_{h,x}(X_s))^2]dx &= \int_{-\infty}^{\infty} \mathbb{E}\left[\left(\int_s^u K'_{h,x}(X_t)dX_t + \frac{1}{2}\int_s^u K''_{h,x}(X_t)d\langle X\rangle_t\right)^2\right]dx \\
&\leqslant \mathfrak{c}_1 \int_{-\infty}^{\infty}\left[\mathbb{E}\left[\left(\int_s^u K'_{h,x}(X_t)b(X_t)dt\right)^2\right]\right. \\
&\qquad + \mathbb{E}\left[\left(\int_s^u K''_{h,x}(X_t)\sigma(X_t)^2dt\right)^2\right] \\
&\qquad\qquad \left. + \mathbb{E}\left[\left(\int_s^u K'_{h,x}(X_t)\sigma(X_t)dW_t\right)^2\right]\right]dx \\
&\leqslant \mathfrak{c}_1\left[(u-s)\int_s^u \mathbb{E}\left(b(X_t)^2\int_{-\infty}^{\infty} K'_{h,x}(X_t)^2dx\right)dt\right. \\
&\qquad + (u-s)\int_s^u \mathbb{E}\left(\sigma(X_t)^4\int_{-\infty}^{\infty} K''_{h,x}(X_t)^2dx\right)dt \\
&\qquad\qquad \left. + \int_s^u \mathbb{E}\left(\sigma(X_t)^2\int_{-\infty}^{\infty} K'_{h,x}(X_t)^2dx\right)dt\right] \\
&\leqslant \mathfrak{c}_2\left[\frac{(u-s)^2}{h^3} + \frac{(u-s)^2}{h^5} + \frac{u-s}{h^3}\right]
\end{aligned}$$

where $\mathfrak{c}_1$ and $\mathfrak{c}_2$ are two positive constants not depending on s, u, h, N, n and t_0. Then,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \text{var}(\widehat{f}_{n,N,h}(x) - \widehat{f}_{N,h}(x))dx \\
&\qquad = \frac{1}{N(T-t_0)^2}\int_{-\infty}^{\infty} \text{var}\left[\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}(K_{h,x}(X_t) - K_{h,x}(X_{t_j}))dt\right]dx \\
&\qquad \leqslant \frac{1}{N(T-t_0)}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\mathbb{E}[(K_{h,x}(X_t) - K_{h,x}(X_{t_j}))^2]dxdt \leqslant \frac{\mathfrak{c}_3}{Nnh^3}
\end{aligned}$$

where the constant $\mathfrak{c}_3 > 0$ is not depending on h, N, n and t_0. On the other hand, by Assumption 2.3,

$$\begin{aligned}
|\mathbb{E}(\widehat{f}_{N,h}(x) - \widehat{f}_{n,N,h}(x))| &\leqslant \frac{1}{T-t_0}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}|\mathbb{E}(K_{h,x}(X_t)) - \mathbb{E}(K_{h,x}(X_{t_j}))|dt \\
&\leqslant \frac{1}{T-t_0}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}|K_h(z-x)|\cdot|p_t(x_0,z) - p_{t_j}(x_0,z)|dzdt \\
&\leqslant \frac{1}{T-t_0}\sum_{j=0}^{n-1}\left[\int_{t_j}^{t_{j+1}}(t-t_j)dt\right] \\
&\qquad\qquad \times\left[\int_{-\infty}^{\infty}|K_h(z-x)|\sup_{u\in[t_0,T]}|\partial_u p_u(x_0,z)|dz\right] \\
&\leqslant \mathfrak{c}_{2.3,3}\frac{T-t_0}{nt_0^{q_3}}\int_{-\infty}^{\infty}|K(z)|\exp\left[-\mathfrak{m}_{2.3,3}\frac{(hz+x-x_0)^2}{T}\right]dz.
\end{aligned}$$

Then, by Jensen's inequality,

$$\begin{aligned}\int_{-\infty}^{\infty}\mathbb{E}(\widehat{f}_{N,h}(x)-\widehat{f}_{n,N,h}(x))^2dx &\leqslant \frac{\mathfrak{c}_5}{n^2t_0^{2q_3}}\int_{-\infty}^{\infty}|K(z)|\int_{-\infty}^{\infty}\exp\left[-2\mathfrak{m}_{2.3,3}\frac{(hz+x-x_0)^2}{T}\right]dxdz\\ &= \frac{\mathfrak{c}_6}{n^2t_0^{2q_3}}\end{aligned}$$

where

$$\mathfrak{c}_6=\mathfrak{c}_5\|K\|_1\int_{-\infty}^{\infty}\exp\left[-2\mathfrak{m}_{2.3,3}\frac{(x-x_0)^2}{T}\right]dx$$

and the constant $\mathfrak{c}_5>0$ is not depending on h, N, n and t_0. This concludes the proof.

A.6. **Proof of Proposition 4.3.** The proof of Proposition 4.3 relies on the two following technical lemmas.

Lemma A.1. *Consider a symmetric and continuous function $\varphi_1:\mathbb{R}\rightarrow\mathbb{R}$ such that $\overline{\varphi}_1: z\mapsto z\varphi_1(z)$ belongs to $\mathbb{L}^2(\mathbb{R},dx)$. Consider also $\varphi_2,\psi\in C^0(\mathbb{R})$ having polynomial growth. Under Assumptions 2.1 and 2.2, for every $p>0$, there exists a constant $\mathfrak{c}_{A.1}(p)>0$, not depending on φ_1 and t_0, such that for every $s,t\in[t_0,T]$ satisfying $s<t$,*

$$\begin{aligned}&\int_{-\infty}^{\infty}\mathbb{E}\left[\left(\int_s^t\varphi_1(x-X_u)\varphi_2(X_u)dW_u\right)^2\psi(X_t)^2\right]dx\\ &\qquad\leqslant \mathfrak{c}_{A.1}(p)(t-s)\left[\|\varphi_1\|_2^2+\|\overline{\varphi}_1\|_2^2+\frac{1}{t_0^{1/(2p)}}\left(\int_{-\infty}^{\infty}\varphi_1(z)^{2p}dz\right)^{1/p}\right].\end{aligned}$$

Lemma A.2. *Consider $\varphi\in C^0(\mathbb{R})$. Under Assumptions 2.1 and 2.2, for every $s,t\in[t_0,T]$ such that $s<t$,*

$$\int_{-\infty}^{\infty}\mathbb{E}(K_{h,x}(X_s)\varphi(X_s,X_t))^2dx\leqslant\frac{\mathfrak{c}_{2.2,1}\|K\|_1^2}{t_0^{1/2}}\mathbb{E}[\varphi(X_s,X_t)^2].$$

The proof of Lemma A.1 (resp. Lemma A.2) is postponed to Subsubection A.6.1 (resp. Subsubsection A.6.2).

First of all, note that

$$\begin{aligned}\mathbb{E}(\|\widehat{bf}_{n,N,h}-bf\|_2^2) &\leqslant 2\mathbb{E}(\|\widehat{bf}_{N,h}-bf\|_2^2)+2\mathbb{E}(\|\widehat{bf}_{N,h}-\widehat{bf}_{n,N,h}\|_2^2)\\ &\leqslant 2\left[\|(bf)_h-bf\|_2^2+\frac{\mathfrak{c}_{3.4}(t_0)}{Nh}+\int_{-\infty}^{\infty}\mathbb{E}(\widehat{bf}_{N,h}(x)-\widehat{bf}_{n,N,h}(x))^2dx\right.\\ &\qquad\left.+\int_{-\infty}^{\infty}\mathrm{var}(\widehat{bf}_{N,h}(x)-\widehat{bf}_{n,N,h}(x))dx\right]\\ &=: 2\left[\|(bf)_h-bf\|_2^2+\frac{\mathfrak{c}_{3.4}(t_0)}{Nh}+B_{n,N,h}+V_{n,N,h}\right]\end{aligned}$$

by Proposition 3.4, and note also that

$$\widehat{bf}_{N,h}(x)-\widehat{bf}_{n,N,h}(x)=\frac{1}{N(T-t_0)}\sum_{i=1}^N\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}(K_{h,x}(X_t^i)-K_{h,x}(X_{t_j}^i))dX_t^i.$$

The proof is dissected in two steps. The term $V_{n,N,h}$ is controlled in the first step, and then $B_{n,N,h}$ is controlled in the second one.

Step 1. First of all, by Jensen's inequality,

$$\begin{aligned}
V_{n,N,h} &= \frac{1}{N(T-t_0)^2}\int_{-\infty}^{\infty}\mathrm{var}\left[\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))dX_t\right]dx\\
&\leqslant \frac{2}{N(T-t_0)}\int_{-\infty}^{\infty}\left[\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\mathbb{E}[(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))^2b(X_t)^2]dt\right]dx+V_{n,N,h}^{\sigma}
\end{aligned}$$

with

$$V_{n,N,h}^{\sigma}:=\frac{2}{N(T-t_0)^2}\int_{-\infty}^{\infty}\mathbb{E}\left[\left(\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))\sigma(X_t)dW_t\right)^2\right]dx.$$

In order to control $V_{n,N,h}$ as in the proof of Proposition 4.2, a preliminary bound on $V_{n,N,h}^{\sigma}$ has to be established via the isometry property of Itô's integral:

$$\begin{aligned}
V_{n,N,h}^{\sigma} &= \frac{2}{N(T-t_0)^2}\int_{-\infty}^{\infty}\mathbb{E}\left[\left(\int_{t_0}^{T}\left(\sum_{j=0}^{n-1}(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))\sigma(X_t)\mathbf{1}_{[t_j,t_{j+1}]}(t)\right)dW_t\right)^2\right]dx\\
&= \frac{2}{N(T-t_0)^2}\int_{-\infty}^{\infty}\int_{t_0}^{T}\mathbb{E}\left[\left(\sum_{j=0}^{n-1}(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))\sigma(X_t)\mathbf{1}_{[t_j,t_{j+1}]}(t)\right)^2\right]dtdx\\
&= \frac{2}{N(T-t_0)}\int_{-\infty}^{\infty}\left(\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\mathbb{E}[(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))^2\sigma(X_t)^2]dt\right)dx.
\end{aligned}$$

Then,

$$\begin{aligned}
V_{n,N,h} \leqslant\; & \frac{2}{N(T-t_0)}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\mathbb{E}[(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))^2b(X_t)^2]dxdt\\
&+\frac{2}{N(T-t_0)}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\mathbb{E}[(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))^2\sigma(X_t)^2]dxdt.
\end{aligned}$$

For $\varphi=b$ or $\varphi=\sigma$, by Itô's formula,

$$\begin{aligned}
&\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\mathbb{E}[(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))^2\varphi(X_t)^2]dxdt\\
&\qquad\leqslant \mathfrak{c}_1\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\left[\mathbb{E}\left[\left(\int_{t_j}^{t}K_{h,x}'(X_u)b(X_u)du\right)^2\varphi(X_t)^2\right]\right.\\
&\qquad\qquad+\mathbb{E}\left[\left(\int_{t_j}^{t}K_{h,x}''(X_u)\sigma(X_u)^2du\right)^2\varphi(X_t)^2\right]\\
&\qquad\qquad\left.+\mathbb{E}\left[\left(\int_{t_j}^{t}K_{h,x}'(X_u)\sigma(X_u)dW_u\right)^2\varphi(X_t)^2\right]\right]dxdt
\end{aligned}$$

where the constant $\mathfrak{c}_1 > 0$ is not depending on φ, h, N, n and t_0. Moreover, for every $j \in \{0,\dots,n-1\}$ and $t \in [t_j, t_{j+1}]$, by Lemma A.1 with $p = 1/(1-\varepsilon)$,

$$\begin{aligned}\int_{-\infty}^{\infty}\mathbb{E}\left[\left(\int_{t_j}^{t}K'_{h,x}(X_u)\sigma(X_u)dW_u\right)^2\varphi(X_t)^2\right]dx \leqslant{}& \mathfrak{c}_{A.1}(p)(t-t_j)\int_{-\infty}^{\infty}K'_h(z)^2dz\\ &+\mathfrak{c}_{A.1}(p)(t-t_j)\int_{-\infty}^{\infty}z^2K'_h(z)^2dz\\ &+\frac{\mathfrak{c}_{A.1}(p)}{t_0^{1/(2p)}}(t-t_j)\left(\int_{-\infty}^{\infty}K'_h(z)^{2p}dz\right)^{1/p}\\ \leqslant{}& \mathfrak{c}_2(\varepsilon)(t-t_j)\left[1+\frac{1}{h^3}+\frac{t_0^{-(1-\varepsilon)/2}}{h^{3+\varepsilon}}\right]\end{aligned}$$

where the constant $\mathfrak{c}_2(\varepsilon) > 0$ depends on ε, but not on φ, j, t, h, N, n and t_0. Thus, by Jensen's inequality and Remark 2.4,

$$\begin{aligned}&\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\mathbb{E}[(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))^2\varphi(X_t)^2]dxdt\\ &\quad\leqslant \mathfrak{c}_3(\varepsilon)\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\left[(t-t_j)\int_{t_j}^{t}\mathbb{E}\left(\varphi(X_t)^2b(X_u)^2\int_{-\infty}^{\infty}K'_{h,x}(X_u)^2dx\right)du\right.\\ &\qquad\left.+(t-t_j)\int_{t_j}^{t}\mathbb{E}\left(\varphi(X_t)^2\sigma(X_u)^4\int_{-\infty}^{\infty}K''_{h,x}(X_u)^2dx\right)du+(t-t_j)\left[1+\frac{1}{h^3}+\frac{t_0^{-(1-\varepsilon)/2}}{h^{3+\varepsilon}}\right]\right]dt\\ &\quad\leqslant \frac{\mathfrak{c}_4(\varepsilon)}{\min\{1,t_0^{(1-\varepsilon)/2}\}}(T-t_0)^3\left(\frac{1}{n^2h^3}+\frac{1}{n^2h^5}+\frac{1}{nh^{3+\varepsilon}}\right)\end{aligned}$$

where $\mathfrak{c}_3(\varepsilon)$ and $\mathfrak{c}_4(\varepsilon)$ are two positive constants depending on ε, but not on φ, h, N, n and t_0. Therefore,

$$V_{n,N,h} \leqslant \frac{\mathfrak{c}_5(\varepsilon)}{\min\{1,t_0^{(1-\varepsilon)/2}\}}\cdot\frac{1}{Nnh^{3+\varepsilon}}$$

where the constant $\mathfrak{c}_5(\varepsilon) > 0$ depends on ε, but not on h, N, n and t_0.

Step 2. First of all, since Itô's integral restricted to $\mathbb{H}^2$ is a martingale-valued map, since $K_{h,x}$ is a kernel, by Lemma A.2, by Assumptions 2.2 and 2.3, and since b is Lipschitz continuous,

$$\begin{aligned}
&\int_{-\infty}^{\infty}\mathbb{E}(\widehat{bf}_{N,h}(x)-\widehat{bf}_{n,N,h}(x))^2dx\\
&\quad=\int_{-\infty}^{\infty}\left(\frac{1}{T-t_0}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\mathbb{E}((K_{h,x}(X_t)-K_{h,x}(X_{t_j}))b(X_t))dt\right.\\
&\qquad\qquad\left.+\frac{1}{T-t_0}\sum_{j=0}^{n-1}\mathbb{E}\left[\int_{t_j}^{t_{j+1}}(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))\sigma(X_t)dW_t\right]\right)^2dx\\
&\quad\leqslant\frac{2}{T-t_0}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\mathbb{E}(K_{h,x}(X_t)b(X_t)-K_{h,x}(X_{t_j})b(X_{t_j}))^2dxdt\\
&\qquad\qquad+\frac{2}{T-t_0}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\mathbb{E}(|K_{h,x}(X_{t_j})|\cdot|b(X_t)-b(X_{t_j})|)^2dxdt\\
&\quad\leqslant\frac{2\|K\|_1}{T-t_0}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}|K_{h,x}(z)|dx\right)b(z)^2(p_t(x_0,z)-p_{t_j}(x_0,z))^2dzdt\\
&\qquad\qquad+\frac{2\mathfrak{c}_{2.2,1}\|K\|_1^2}{t_0^{1/2}(T-t_0)}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\mathbb{E}[(b(X_t)-b(X_{t_j}))^2]dt\\
&\quad\leqslant\frac{4\|K\|_1^2}{T-t_0}\sum_{j=0}^{n-1}\left[\int_{t_j}^{t_{j+1}}(t-t_j)^2dt\right]\left[\int_{-\infty}^{\infty}b(z)^2\sup_{u\in[t_0,T]}|\partial_up_u(x_0,z)|^2dz\right]\\
&\qquad\qquad+\frac{2\mathfrak{c}_{2.2,1}\|K\|_1^2}{t_0^{1/2}(T-t_0)}\|b'\|_\infty^2\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\mathbb{E}[(X_t-X_{t_j})^2]dt\\
&\quad\leqslant\mathfrak{c}_6\left(\frac{1}{t_0^{2q_3}n^2}+\frac{1}{t_0^{1/2}(T-t_0)}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\mathbb{E}[(X_t-X_{t_j})^2]dt\right)
\end{aligned}$$

where the constant $\mathfrak{c}_6>0$ is not depending on h, N, n and t_0. Moreover, for any $j\in\{0,\dots,n-1\}$ and $t\in[t_j,t_{j+1}]$,

$$X_t-X_{t_j}=\int_{t_j}^t b(X_u)du+\int_{t_j}^t\sigma(X_u)dW_u$$

and then, by Jensen's inequality, the isometry property of Itô's integral and Remark 2.4,

$$\begin{aligned}
\mathbb{E}[(X_t-X_{t_j})^2]&\leqslant(t-t_j)\int_{t_j}^t\mathbb{E}(b(X_u)^2)du+\int_{t_j}^t\mathbb{E}(\sigma(X_u)^2)du\\
&\leqslant(t-t_j)^2\sup_{u\in[t_0,T]}\mathbb{E}(b(X_u)^2)+(t-t_j)\sup_{u\in[t_0,T]}\mathbb{E}(\sigma(X_u)^2)\leqslant\mathfrak{c}_7(t-t_j)
\end{aligned}$$

where the constant $\mathfrak{c}_7>0$ is not depending on j, t, h, N, n and t_0. Therefore,

$$\int_{-\infty}^{\infty}\mathbb{E}(\widehat{bf}_{N,h}(x)-\widehat{bf}_{n,N,h}(x))^2dx\leqslant\frac{\mathfrak{c}_8}{\min\{t_0^{1/2},t_0^{2q_3}\}}\left(\frac{1}{n^2}+\frac{1}{n}\right)$$

where the constant $\mathfrak{c}_8>0$ is not depending on n, N, h and t_0.

Remark A.3. *Assume that b and σ are bounded. Then, in Step 1, for $\varphi = b$ or $\varphi = \sigma$,*

$$\begin{aligned}
&\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\mathbb{E}[(K_{h,x}(X_t)-K_{h,x}(X_{t_j}))^2\varphi(X_t)^2]dxdt\\
&\quad\leqslant \mathfrak{c}_1\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\left[\mathbb{E}\left[\left(\int_{t_j}^{t}K'_{h,x}(X_u)b(X_u)du\right)^2\varphi(X_t)^2\right]\right.\\
&\qquad\qquad+\mathbb{E}\left[\left(\int_{t_j}^{t}K''_{h,x}(X_u)\sigma(X_u)^2du\right)^2\varphi(X_t)^2\right]\\
&\qquad\qquad\qquad\left.+\mathbb{E}\left[\left(\int_{t_j}^{t}K'_{h,x}(X_u)\sigma(X_u)dW_u\right)^2\varphi(X_t)^2\right]\right]dxdt\\
&\quad\leqslant \mathfrak{c}_1\|\varphi\|_{\infty}^2\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{-\infty}^{\infty}\left[\mathbb{E}\left[\left(\int_{t_j}^{t}K'_{h,x}(X_u)b(X_u)du\right)^2\right]\right.\\
&\qquad\qquad\left.+\mathbb{E}\left[\left(\int_{t_j}^{t}K''_{h,x}(X_u)\sigma(X_u)^2du\right)^2\right]+\mathbb{E}\left[\left(\int_{t_j}^{t}K'_{h,x}(X_u)\sigma(X_u)dW_u\right)^2\right]\right]dxdt.
\end{aligned}$$

So, in this special case, the bound on $V_{n,N,h}$ is established by using the exact same arguments than in the proof of Proposition 4.2. In particular, one can take $\varepsilon = 0$, the additional conditions $K \in \mathbb{L}^4(\mathbb{R}, dx)$ and $z \mapsto zK(z)$ belongs to $\mathbb{L}^2(\mathbb{R}, dx)$ are not required, and the bound on $V_{n,N,h}$ is of order $1/(Nnh^3)$ and doesn't depend on t_0. When $\varphi = b$ or $\varphi = \sigma$ is not bounded, since $\varphi(X_t)$ is not $\sigma(W_u)$-measurable for every $u \in [t_j, t)$ ($j \in \{0, \dots, n-1\}$), the Hölder inequality has to be used to get a suitable bound on

$$\int_{-\infty}^{\infty}\mathbb{E}\left[\left(\int_{t_j}^{t}K'_{h,x}(X_u)\sigma(X_u)dW_u\right)^2\varphi(X_t)^2\right]dx$$

(see the proof of Lemma A.1), and for this reason the variance term in the bound of Proposition 4.3 is of order $1/(Nnh^{3+\varepsilon})$ instead of $1/(Nnh^3)$ as when b and σ are bounded.

A.6.1. *Proof of Lemma A.1.* Consider $\varphi(x, z) := \varphi_1(x - z)\varphi_2(z)$ for every $z \in \mathbb{R}$, $q > 0$ such that $1/p + 1/q = 1$, and $s, t \in [0, T]$ such that $s < t$. First of all, by the isometry property of Itô's integral, Burkholder-Davis-Gundy's inequality, Hölder's inequality, Markov's inequality, Remark 2.4, and the generalized Minkowski inequality,

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_s^t\varphi(x,X_u)dW_u\right)^2\psi(X_t)^2\right]\\
&\quad\leqslant x^2\mathbb{E}\left[\left(\int_s^t\varphi(x,X_u)dW_u\right)^2\mathbf{1}_{\psi(X_t)^2\leqslant x^2}\right]\\
&\qquad\quad+\mathbb{E}\left[\left(\int_s^t\varphi(x,X_u)dW_u\right)^{2p}\right]^{1/p}\mathbb{E}(\psi(X_t)^{4q})^{1/(2q)}\mathbb{P}(\psi(X_t)^2>x^2)^{1/(2q)}\\
&\quad\leqslant x^2\int_s^t\mathbb{E}[\varphi(x,X_u)^2]du+\mathfrak{c}_1(p)\mathbb{E}\left[\left(\int_s^t\varphi(x,X_u)^2du\right)^p\right]^{1/p}\mathbb{E}(\psi(X_t)^{4q})^{1/(2q)}\\
&\qquad\qquad\times\left[\frac{\mathbb{E}(\psi(X_t)^2)^{1/(2q)}}{x^{1/q}}\mathbf{1}_{[-1,1]}(x)+\frac{\mathbb{E}(\psi(X_t)^{4q})^{1/(2q)}}{x^2}\mathbf{1}_{\mathbb{R}\setminus[-1,1]}(x)\right]\\
&\quad\leqslant x^2\int_s^t\mathbb{E}[\varphi(x,X_u)^2]du+\mathfrak{c}_2(p)\left(\int_s^t\mathbb{E}[\varphi(x,X_u)^{2p}]^{1/p}du\right)\left(\frac{1}{x^{1/q}}\mathbf{1}_{[-1,1]}(x)+\frac{1}{x^2}\mathbf{1}_{\mathbb{R}\setminus[-1,1]}(x)\right)
\end{aligned}$$

where $\mathfrak{c}_1(p)$ and $\mathfrak{c}_2(p)$ are two positive constants depending on p, but not on x, s, t, φ and t_0. On the one hand, since φ_2 has polynomial growth, by Remark 2.4, for every $u\in[s,t]$,

$$\begin{aligned}\int_{-\infty}^{\infty}x^2\mathbb{E}[\varphi(x,X_u)^2]dx &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x^2\varphi_1(x-z)^2\varphi_2(z)^2p_u(x_0,z)dxdz\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+z)^2\varphi_1(x)^2\varphi_2(z)^2p_u(x_0,z)dxdz\\ &\leqslant 2\left(\int_{-\infty}^{\infty}x^2\varphi_1(x)^2dx\right)\left(\int_{-\infty}^{\infty}\varphi_2(z)^2p_u(x_0,z)dz\right)\\ &\qquad +2\left(\int_{-\infty}^{\infty}\varphi_1(x)^2dx\right)\left(\int_{-\infty}^{\infty}z^2\varphi_2(z)^2p_u(x_0,z)dz\right)\\ &\leqslant \mathfrak{c}_3(\|\overline{\varphi}_1\|_2^2+\|\varphi_1\|_2^2)\end{aligned}$$

where the constant $\mathfrak{c}_3>0$ is not depending on u, φ_1 and t_0. Then,

$$\int_{-\infty}^{\infty}x^2\int_s^t\mathbb{E}[\varphi(x,X_u)^2]dudx\leqslant\mathfrak{c}_3(t-s)(\|\overline{\varphi}_1\|_2^2+\|\varphi_1\|_2^2).$$

On the other hand, since φ_2 has polynomial growth, by Assumption 2.2, for every $x\in\mathbb{R}$,

$$\begin{aligned}\int_s^t\mathbb{E}[\varphi(x,X_u)^{2p}]^{1/p}du &= \int_s^t\left(\int_{-\infty}^{\infty}\varphi_1(z)^{2p}\varphi_2(z+x)^{2p}p_u(x_0,z+x)dz\right)^{1/p}du\\ &\leqslant \frac{\mathfrak{c}_4(p)}{t_0^{1/(2p)}}(t-s)\left(\int_{-\infty}^{\infty}\varphi_1(z)^{2p}dz\right)^{1/p}\end{aligned}$$

where the constant $\mathfrak{c}_4(p)>0$ depends on p, but not on x, s, t, φ_1 and t_0. Then,

$$\begin{aligned}&\int_{-\infty}^{\infty}\left(\int_s^t\mathbb{E}[\varphi(x,X_u)^{2p}]^{1/p}du\right)\left(\frac{1}{x^{1/q}}\mathbf{1}_{[-1,1]}(x)+\frac{1}{x^2}\mathbf{1}_{\mathbb{R}\setminus[-1,1]}(x)\right)dx\\ &\qquad\qquad\leqslant\frac{\mathfrak{c}_5(p)}{t_0^{1/(2p)}}(t-s)\left(\int_{-\infty}^{\infty}\varphi_1(z)^{2p}dz\right)^{1/p}\end{aligned}$$

with

$$\mathfrak{c}_5(p)=\mathfrak{c}_4(p)\left(\int_{-1}^{1}\frac{dx}{x^{1/q}}+\int_{\mathbb{R}\setminus[-1,1]}\frac{dx}{x^2}\right)<\infty.$$

A.6.2. *Proof of Lemma A.2.* Consider $s,t\in[t_0,T]$ such that $s<t$, and let $p_{s,t}$ (resp. $p_{t|s}$) be the density of (X_s,X_t) (resp. the conditional density of X_t with respect to X_s). Moreover, for the sake of readability, $p_s(x_0,.)$ is denoted by $p_s(.)$ in this proof. By Assumption 2.2,

$$\begin{aligned}\mathbb{E}(K_{h,x}(X_s)\varphi(X_s,X_t))^2 &= \|K\|_1^2\left[\int_{-\infty}^{\infty}\frac{K_{h,x}(y)}{\|K_{h,x}\|_1}\left(\int_{-\infty}^{\infty}\varphi(y,z)p_{t|s}(z|y)dz\right)p_s(y)dy\right]^2\\ &\leqslant \|K\|_1\int_{-\infty}^{\infty}|K_{h,x}(y)|\left(\int_{-\infty}^{\infty}\varphi(y,z)p_{t|s}(z|y)dz\right)^2p_s(y)^2dy\\ &\leqslant \|K\|_1\sup_{s\in[t_0,T]}\left\{\sup_{y\in\mathbb{R}}p_s(y)\right\}\int_{-\infty}^{\infty}|K_{h,x}(y)|\left(\int_{-\infty}^{\infty}\varphi(y,z)^2p_{t|s}(z|y)dz\right)p_s(y)dy\\ &\leqslant \frac{\mathfrak{c}_{2.2,1}\|K\|_1}{t_0^{1/2}}\int_{-\infty}^{\infty}|K_{h,x}(y)|\int_{-\infty}^{\infty}\varphi(y,z)^2p_{s,t}(y,z)dzdy.\end{aligned}$$

Therefore,

$$\begin{aligned}\int_{-\infty}^{\infty}\mathbb{E}(K_{h,x}(X_s)\varphi(X_s,X_t))^2dx &\leqslant \frac{\mathfrak{c}_{2.2,1}\|K\|_1}{t_0^{1/2}}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}|K_{h,x}(y)|dx\right)\left(\int_{-\infty}^{\infty}\varphi(y,z)^2p_{s,t}(y,z)dz\right)dy\\ &= \frac{\mathfrak{c}_{2.2,1}\|K\|_1^2}{t_0^{1/2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(y,z)^2p_{s,t}(y,z)dzdy\\ &= \frac{\mathfrak{c}_{2.2,1}\|K\|_1^2}{t_0^{1/2}}\mathbb{E}[\varphi(X_s,X_t)^2].\end{aligned}$$

A.7. **Proof of Theorem 5.2.** Throughout this subsection, $\mathcal{K}$ is a primitive function of the kernel K. The proof of Theorem 5.2.(1) relies on the following technical lemmas proved at the end of this subsection. The proof of Theorem 5.2.(2) is left to the reader because it is similar but simpler than the proof of Theorem 5.2.(1) detailed in this subsection.

Lemma A.4. *Under Assumptions 2.1, 2.2, 3.1 and 5.1,*

$$\frac{1}{T-t_0}\int_{t_0}^{T}K_h(X_t-x)dX_t = \Phi_h(X,x)\ ;\ \forall x\in\mathbb{R},\ \forall h>0,$$

where $(x,h,\varphi)\mapsto\Phi_h(\varphi,x)$ *is the map from* $\mathbb{R}\times(0,\infty)\times C^0([t_0,T];\mathbb{R})$ *into* $\mathbb{R}$ *defined by*

$$\Phi_h(\varphi,x) := \frac{1}{T-t_0}\left[\mathcal{K}\left(\frac{\varphi(T)-x}{h}\right)-\mathcal{K}\left(\frac{\varphi(t_0)-x}{h}\right)-\frac{1}{2h^2}\int_{t_0}^{T}K'\left(\frac{\varphi(t)-x}{h}\right)\sigma(\varphi(t))^2dt\right]$$

for every $x\in\mathbb{R}$, $h>0$ *and* $\varphi\in C^0([t_0,T];\mathbb{R})$. *Moreover,*

1. *For every* $x\in\mathbb{R}$, $h>0$ *and* $\varphi\in C^0([t_0,T];\mathbb{R})$,
$$|\Phi_h(\varphi,x)|\leqslant\frac{2\|\mathcal{K}\|_\infty}{T-t_0}+\frac{\|\sigma\|_\infty^2\|K'\|_\infty}{2h^2}.$$
2. *For every* $h>0$ *and* $\varphi\in C^0([t_0,T];\mathbb{R})$,
$$\|\Phi_h(\varphi,.)\|_{2,\delta}^2\leqslant\frac{6\|\mathcal{K}\|_\infty^2}{(T-t_0)^2}+\frac{\|\delta\|_\infty\|\sigma\|_\infty^4\|K'\|_2^2}{h^3}.$$
3. *There exists a deterministic constant* $\mathfrak{c}_{A.4,1}>0$ *such that, for every* $h,h'>0$,
$$\mathbb{E}(\langle\Phi_h(X^1,.),\Phi_{h'}(X^2,.)\rangle_{2,\delta}^2)\leqslant\mathfrak{c}_{A.4,1}\mathfrak{m}(h')$$
with
$$\mathfrak{m}(h')=\mathbb{E}(\|\Phi_{h'}(X,.)\|_{2,\delta}^2).$$
4. *There exists a deterministic constant* $\mathfrak{c}_{A.4,2}>0$ *such that, for every* $h>0$ *and* $\varphi\in\mathbb{L}^2(\mathbb{R},dx)$,
$$\mathbb{E}(\langle\Phi_h(X,.),\varphi\rangle_{2,\delta}^2)\leqslant\mathfrak{c}_{A.4,2}\|\varphi\|_{2,\delta}^2.$$
5. *There exists a deterministic constant* $\mathfrak{c}_{A.4,3}>0$ *such that, for every* $h,h'\in\mathcal{H}_N$,
$$|\langle\Phi_h(X,.),(bf)_{h'}\rangle_{2,\delta}|\leqslant\mathfrak{c}_{A.4,3}\quad\text{a.s.}$$

Lemma A.5. *Consider*

$$U_{h,h'}(N):=\sum_{i\neq j}\langle\Phi_h(X^i,.)-(bf)_h,\Phi_{h'}(X^j,.)-(bf)_{h'}\rangle_{2,\delta}\ ;\ \forall h,h'\in\mathcal{H}_N. \tag{14}$$

Under Assumptions 2.1, 2.2, 3.1 and 5.1, there exists a deterministic constant $\mathfrak{c}_{A.5}>0$, *not depending on* N, *such that for every* $\theta\in(0,1)$ *and* $\lambda>0$, *with probability larger than* $1-5.4|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h\in\mathcal{H}_N}\left\{\frac{|U_{h,h_0}(N)|}{N^2}-\frac{\theta\mathfrak{m}(h)}{N}\right\}\leqslant\frac{\mathfrak{c}_{A.5}(1+\lambda)^3}{\theta N}$$

and

$$\sup_{h\in\mathcal{H}_N}\left\{\frac{|U_{h,h}(N)|}{N^2}-\frac{\theta\mathfrak{m}(h)}{N}\right\}\leqslant\frac{\mathfrak{c}_{A.5}(1+\lambda)^3}{\theta N}.$$

Lemma A.6. *Consider*

$$V_h(N) := \frac{1}{N}\sum_{i=1}^{N}\|\Phi_h(X^i,.) - (bf)_h\|_{2,\delta}^2 \ ; \ \forall h \in \mathcal{H}_N.$$

Under Assumptions 2.1, 2.2, 3.1 and 5.1, there exists a deterministic constant $\mathfrak{c}_{A.6} > 0$, not depending on N, such that for every $\theta \in (0,1)$ and $\lambda > 0$, with probability larger than $1 - 2|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h\in\mathcal{H}_N}\left\{\frac{1}{N}|V_h(N) - \mathfrak{m}(h)| - \frac{\theta\mathfrak{m}(h)}{N}\right\} \leqslant \frac{\mathfrak{c}_{A.6}(1+\lambda)}{\theta N}.$$

Lemma A.7. *Consider*

$$W_{h,h'}(N) := \langle \widehat{bf}_{N,h} - (bf)_h, (bf)_{h'} - bf\rangle_{2,\delta} \ ; \ \forall h, h' \in \mathcal{H}_N. \tag{15}$$

Under Assumptions 2.1, 2.2, 3.1 and 5.1, there exists a deterministic constant $\mathfrak{c}_{A.7} > 0$, not depending on N, such that for every $\theta \in (0,1)$ and $\lambda > 0$, with probability larger than $1 - 2|\mathcal{H}_N|e^{-\lambda}$,

$$\begin{aligned}
\sup_{h\in\mathcal{H}_N}\{|W_{h,h_0}(N)| - \theta\|(bf)_{h_0} - bf\|_{2,\delta}^2\} &\leqslant \frac{\mathfrak{c}_{A.7}(1+\lambda)^2}{\theta N},\\
\sup_{h\in\mathcal{H}_N}\{|W_{h_0,h}(N)| - \theta\|(bf)_{h} - bf\|_{2,\delta}^2\} &\leqslant \frac{\mathfrak{c}_{A.7}(1+\lambda)^2}{\theta N} \ \textit{and}\\
\sup_{h\in\mathcal{H}_N}\{|W_{h,h}(N)| - \theta\|(bf)_{h} - bf\|_{2,\delta}^2\} &\leqslant \frac{\mathfrak{c}_{A.7}(1+\lambda)^2}{\theta N}.
\end{aligned}$$

A.7.1. *Steps of the proof.* The proof of Theorem 5.2.(1) is dissected in four steps.

Step 1. This first step provides a suitable decomposition of $\|\widehat{bf}_{N,\widehat{h}} - bf\|_{2,\delta}^2$. First,

$$\begin{aligned}
\|\widehat{bf}_{N,\widehat{h}} - bf\|_{2,\delta}^2 = \ &\|\widehat{bf}_{N,\widehat{h}} - \widehat{bf}_{N,h_0}\|_{2,\delta}^2 + \|\widehat{bf}_{N,h_0} - bf\|_{2,\delta}^2\\
&-2\langle \widehat{bf}_{N,h_0} - \widehat{bf}_{N,\widehat{h}}, \widehat{bf}_{N,h_0} - bf\rangle_{2,\delta}.
\end{aligned}$$

Then, by (11) and the definition of pen(.) (see (12)), for any $h \in \mathcal{H}_N$,

$$\begin{aligned}
\|\widehat{bf}_{N,\widehat{h}} - bf\|_{2,\delta}^2 &\leqslant \|\widehat{bf}_{N,h} - \widehat{bf}_{N,h_0}\|_{2,\delta}^2 + \mathrm{pen}(h) - \mathrm{pen}(\widehat{h})\\
&\quad +\|\widehat{bf}_{N,h_0} - bf\|_{2,\delta}^2 - 2\langle \widehat{bf}_{N,h_0} - \widehat{bf}_{N,\widehat{h}}, \widehat{bf}_{N,h_0} - bf\rangle_{2,\delta}\\
&\leqslant \|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2 + \mathrm{pen}(h) - \mathrm{pen}(\widehat{h})\\
&\quad +\|\widehat{bf}_{N,h_0} - bf\|_{2,\delta}^2 - 2\langle \widehat{bf}_{N,h} - \widehat{bf}_{N,\widehat{h}}, \widehat{bf}_{N,h_0} - bf\rangle_{2,\delta}\\
&= \|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2 - \psi_N(h) + \psi_N(\widehat{h}) \qquad (16)
\end{aligned}$$

where

$$\psi_N(h) := 2\langle \widehat{bf}_{N,h} - bf, \widehat{bf}_{N,h_0} - bf\rangle_{2,\delta} - \mathrm{pen}(h).$$

Let's complete the decomposition of $\|\widehat{bf}_{N,\widehat{h}} - bf\|_{2,\delta}^2$ by writing

$$\psi_N(h) = 2(\psi_{1,N}(h) + \psi_{2,N}(h) + \psi_{3,N}(h)),$$

where

$$\begin{aligned}\psi_{1,N}(h) &:= \frac{1}{(T-t_0)^2N^2}\sum_{i=1}^{N}\left\langle\int_{t_0}^{T}K_h(X_s^i-\cdot)dX_s^i,\int_{t_0}^{T}K_{h_0}(X_s^i-\cdot)dX_s^i\right\rangle_{2,\delta}+\frac{U_{h,h_0}(N)}{N^2}-\frac{1}{2}\text{pen}(h)\\ &= \frac{U_{h,h_0}(N)}{N^2},\\ \psi_{2,N}(h) &:= -\frac{1}{N^2}\left(\sum_{i=1}^{N}\left\langle\frac{1}{T-t_0}\int_{t_0}^{T}K_{h_0}(X_s^i-\cdot)dX_s^i,(bf)_h\right\rangle_{2,\delta}+\right.\\ &\left.\quad+\sum_{i=1}^{N}\left\langle\frac{1}{T-t_0}\int_{t_0}^{T}K_{h}(X_s^i-\cdot)dX_s^i,(bf)_{h_0}\right\rangle_{2,\delta}\right)+\frac{1}{N}\langle(bf)_{h_0},(bf)_h\rangle_{2,\delta}\text{ and}\\ \psi_{3,N}(h) &:= W_{h,h_0}(N)+W_{h_0,h}(N)+\langle(bf)_h-bf,(bf)_{h_0}-bf\rangle_{2,\delta}.\end{aligned}$$

Step 2. This step deals with bounds on $\mathbb{E}(\psi_{j,N}(h))$ and $\mathbb{E}(\psi_{j,N}(\widehat{h}))$ for $j=1,2,3$.

- By Lemma A.5, for any $\lambda>0$ and $\theta\in(0,1)$, with probability larger than $1-5.4|\mathcal{H}_N|e^{-\lambda}$,

$$|\psi_{1,N}(h)|\leqslant\frac{\theta\mathfrak{m}(h)}{N}+\frac{\mathfrak{c}_{A.5}(1+\lambda)^3}{\theta N}\quad\text{and}\quad|\psi_{1,N}(\widehat{h})|\leqslant\frac{\theta\mathfrak{m}(\widehat{h})}{N}+\frac{\mathfrak{c}_{A.5}(1+\lambda)^3}{\theta N}.$$

- On the one hand, for any $h,h'\in\mathcal{H}_N$, consider

$$\Psi_{2,N}(h,h'):=\frac{1}{N}\sum_{i=1}^{N}\langle\Phi_h(X^i,.),(bf)_{h'}\rangle_{2,\delta}.$$

  By Lemma A.4,

$$|\Psi_{2,N}(h,h')|\leqslant\frac{1}{N}\sum_{i=1}^{N}\left|\int_{-\infty}^{\infty}\Phi_h(X^i,x)(bf)_{h'}(x)\delta(x)dx\right|\leqslant\mathfrak{c}_{A.4,3}\quad\text{a.s.}$$

  On the other hand,

$$|\langle(bf)_h,(bf)_{h_0}\rangle_{2,\delta}|\leqslant\|\delta\|_\infty\|K_h*(bf)\|_\infty\|K_{h_0}*(bf)\|_1\leqslant\|\delta\|_\infty\|K\|_1^2\|bf\|_\infty\|bf\|_1.$$

  Then, there exists a deterministic constant $\mathfrak{c}_1>0$, not depending on N and h, such that

$$|\psi_{2,N}(h)|\leqslant\frac{\mathfrak{c}_1}{N}\quad\text{and}\quad|\psi_{2,N}(\widehat{h})|\leqslant\sup_{h'\in\mathcal{H}_N}|\psi_{2,N}(h')|\leqslant\frac{\mathfrak{c}_1}{N}\quad\text{a.s.}$$

- By Lemma A.7 and Cauchy-Schwarz's inequality, with probability larger that $1-|\mathcal{H}_N|e^{-\lambda}$,

$$\begin{aligned}|\psi_{3,N}(h)| &\leqslant \frac{\theta}{4}(\|(bf)_h-bf\|_{2,\delta}^2+\|(bf)_{h_0}-bf\|_{2,\delta}^2)+\frac{8\mathfrak{c}_{A.7}(1+\lambda)^2}{\theta N}\\ &\quad+2\times\frac{1}{2^{1/2}}\left(\frac{\theta}{2}\right)^{1/2}\|(bf)_h-bf\|_{2,\delta}\times\frac{1}{2^{1/2}}\left(\frac{2}{\theta}\right)^{1/2}\|(bf)_{h_0}-bf\|_{2,\delta}\\ &\leqslant\frac{\theta}{2}\|(bf)_h-bf\|_{2,\delta}^2+\left(\frac{\theta}{4}+\frac{1}{\theta}\right)\|(bf)_{h_0}-bf\|_{2,\delta}^2+\frac{8\mathfrak{c}_{A.7}(1+\lambda)^2}{\theta N}\end{aligned}$$

  and

$$|\psi_{3,N}(\widehat{h})|\leqslant\frac{\theta}{2}\|(bf)_{\widehat{h}}-bf\|_{2,\delta}^2+\left(\frac{\theta}{4}+\frac{1}{\theta}\right)\|(bf)_{h_0}-bf\|_{2,\delta}^2+\frac{8\mathfrak{c}_{A.7}(1+\lambda)^2}{\theta N}.$$

Step 3. Let us establish that there exist two deterministic constants $\mathfrak{c}_2,\overline{\mathfrak{c}}_2>0$, not depending on N and θ, such that with probability larger than $1-\overline{\mathfrak{c}}_2|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h\in\mathcal{H}_N}\left\{\|\widehat{bf}_{N,h}-bf\|_{2,\delta}^2-(1+\theta)\left(\|(bf)_h-bf\|_{2,\delta}^2+\frac{\mathfrak{m}(h)}{N}\right)\right\}\leqslant\frac{\mathfrak{c}_2(1+\lambda)^3}{\theta N}$$

and

$$\sup_{h\in\mathcal{H}_N}\left\{\|(bf)_h - bf\|_{2,\delta}^2 + \frac{\mathfrak{m}(h)}{N} - \frac{1}{1-\theta}\|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2\right\} \leqslant \frac{\mathfrak{c}_2(1+\lambda)^3}{\theta(1-\theta)N}.$$

On the one hand, note that

$$\|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2 - (1+\theta)\left(\|(bf)_h - bf\|_{2,\delta}^2 + \frac{\mathfrak{m}(h)}{N}\right)$$

can be written

$$\|\widehat{bf}_{N,h} - (bf)_h\|_{2,\delta}^2 - \frac{(1+\theta)\mathfrak{m}(h)}{N} + 2W_h(N) - \theta\|(bf)_h - bf\|_{2,\delta}^2,$$

where $W_h(N) := W_{h,h}(N)$ (see (15)). Moreover, for any $h \in \mathcal{H}_N$,

$$\|\widehat{bf}_{N,h} - (bf)_h\|_{2,\delta}^2 = \frac{U_h(N)}{N^2} + \frac{V_h(N)}{N} \tag{17}$$

with $U_h(N) = U_{h,h}(N)$ (see (14)). So, with probability larger than $1 - \overline{\mathfrak{c}}_2|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h\in\mathcal{H}_N}\left\{\left|\|\widehat{bf}_{N,h} - (bf)_h\|_{2,\delta}^2 - \frac{\mathfrak{m}(h)}{N}\right| - \frac{\theta\mathfrak{m}(h)}{N}\right\} \leqslant \frac{2(\mathfrak{c}_{A.5} + \mathfrak{c}_{A.6})(1+\lambda)^3}{\theta N}$$

by Lemmas A.5 and A.6, and then

$$\sup_{h\in\mathcal{H}_N}\left\{\|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2 - (1+\theta)\left(\|(bf)_h - bf\|_{2,\delta}^2 + \frac{\mathfrak{m}(h)}{N}\right)\right\} \leqslant \frac{\mathfrak{c}_2(1+\lambda)^3}{\theta N}$$

by Lemma A.7. On the other hand, for any $h \in \mathcal{H}_N$,

$$\|(bf)_h - bf\|_{2,\delta}^2 = \|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2 - \|\widehat{bf}_{N,h} - (bf)_h\|_{2,\delta}^2 - W_h(N).$$

Then,

$$(1-\theta)\left(\|(bf)_h - bf\|_{2,\delta}^2 + \frac{\mathfrak{m}(h)}{N}\right) - \|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2 \leqslant |W_h(N)| - \theta\|(bf)_h - bf\|_{2,\delta}^2 + \Lambda_h(N) - \frac{\theta\mathfrak{m}(h)}{N}$$

where

$$\Lambda_h(N) := \left|\|\widehat{bf}_{N,h} - (bf)_h\|_{2,\delta}^2 - \frac{\mathfrak{m}(h)}{N}\right|.$$

By Equality (17),

$$\Lambda_h(N) = \left|\frac{U_h(N)}{N^2} + \frac{V_h(N)}{N} - \frac{\mathfrak{m}(h)}{N}\right|.$$

By Lemmas A.6 and A.5, there exist two deterministic constants $\mathfrak{c}_3, \overline{\mathfrak{c}}_3 > 0$, not depending N and θ, such that with probability larger than $1 - \overline{\mathfrak{c}}_3|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h\in\mathcal{H}_N}\left\{\Lambda_h(N) - \theta\frac{\mathfrak{m}(h)}{N}\right\} \leqslant \frac{\mathfrak{c}_3(1+\lambda)^3}{\theta N}.$$

By Lemma A.7, with probability larger than $1 - 2|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h\in\mathcal{H}_N}\{|W_h(N)| - \theta\|(bf)_h - bf\|_{2,\delta}^2\} \leqslant \frac{\mathfrak{c}_{A.7}(1+\lambda)^2}{\theta N}.$$

Therefore, with probability larger than $1 - \overline{\mathfrak{c}}_2|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h\in\mathcal{H}_N}\left\{\|(bf)_h - bf\|_{2,\delta}^2 + \frac{\mathfrak{m}(h)}{N} - \frac{1}{1-\theta}\|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2\right\} \leqslant \frac{\mathfrak{c}_2(1+\lambda)^3}{\theta(1-\theta)N}.$$

Step 4. By step 2, there exist two deterministic constants $\mathfrak{c}_4, \overline{\mathfrak{c}}_4 > 0$, not depending on N, θ, h and h_0, such that with probability larger than $1 - \overline{\mathfrak{c}}_4|\mathcal{H}_N|e^{-\lambda}$,

$$|\psi_N(h)| \leqslant \theta\left(\|(bf)_h - bf\|_{2,\delta}^2 + \frac{\mathfrak{m}(h)}{N}\right) + \left(\frac{\theta}{2} + \frac{2}{\theta}\right)\|(bf)_{h_0} - bf\|_{2,\delta}^2 + \frac{\mathfrak{c}_4(1+\lambda)^3}{\theta N}$$

and

$$|\psi_N(\widehat{h})| \leqslant \theta\left(\|(bf)_{\widehat{h}} - bf\|_{2,\delta}^2 + \frac{\mathfrak{m}(\widehat{h})}{N}\right) + \left(\frac{\theta}{2} + \frac{2}{\theta}\right)\|(bf)_{h_0} - bf\|_{2,\delta}^2 + \frac{\mathfrak{c}_4(1+\lambda)^3}{\theta N}.$$

Then, by step 3, there exist two deterministic constants $\mathfrak{c}_5, \overline{\mathfrak{c}}_5 > 0$, not depending on N, θ, h and h_0, such that with probability larger than $1 - \overline{\mathfrak{c}}_5|\mathcal{H}_N|e^{-\lambda}$,

$$|\psi_N(h)| \leqslant \frac{\theta}{1-\theta}\|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2 + \left(\frac{\theta}{2} + \frac{2}{\theta}\right)\|(bf)_{h_0} - bf\|_{2,\delta}^2 + \mathfrak{c}_5\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right)\frac{(1+\lambda)^3}{N}$$

and

$$|\psi_N(\widehat{h})| \leqslant \frac{\theta}{1-\theta}\|\widehat{bf}_{N,\widehat{h}} - bf\|_{2,\delta}^2 + \left(\frac{\theta}{2} + \frac{2}{\theta}\right)\|(bf)_{h_0} - bf\|_{2,\delta}^2 + \mathfrak{c}_5\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right)\frac{(1+\lambda)^3}{N}.$$

By the decomposition (16), there exist two deterministic constants $\mathfrak{c}_6, \overline{\mathfrak{c}}_6 > 0$, not depending on N, θ, h and h_0, such that with probability larger than $1 - \overline{\mathfrak{c}}_6|\mathcal{H}_N|e^{-\lambda}$,

$$\begin{aligned}
\|\widehat{bf}_{N,\widehat{h}} - bf\|_{2,\delta}^2 &\leqslant \|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2 + |\psi_N(h)| + |\psi_N(\widehat{h})| \\
&\leqslant \left(1 + \frac{\theta}{1-\theta}\right)\|\widehat{bf}_{N,h} - bf\|_{2,\delta}^2 + \frac{\theta}{1-\theta}\|\widehat{bf}_{N,\widehat{h}} - bf\|_{2,\delta}^2 \\
&\qquad + \frac{\mathfrak{c}_6}{\theta}\|(bf)_{h_0} - bf\|_{2,\delta}^2 + \frac{\mathfrak{c}_6}{\theta(1-\theta)} \cdot \frac{(1+\lambda)^3}{N}.
\end{aligned}$$

This concludes the proof.

A.7.2. *Proof of Lemma A.4.* First of all, for any $x \in \mathbb{R}$ and $h > 0$, by Itô's formula,

$$\mathcal{K}\left(\frac{X_T - x}{h}\right) = \mathcal{K}\left(\frac{X_{t_0} - x}{h}\right) + \int_{t_0}^T K_h(X_t - x)dX_t + \frac{1}{2h^2}\int_{t_0}^T K'\left(\frac{X_t - x}{h}\right)d\langle X\rangle_t.$$

So,

$$\begin{aligned}
\int_{t_0}^T K_h(X_t - x)dX_t &= \mathcal{K}\left(\frac{X_T - x}{h}\right) - \mathcal{K}\left(\frac{X_{t_0} - x}{h}\right) \\
&\qquad - \frac{1}{2h^2}\int_{t_0}^T K'\left(\frac{X_t - x}{h}\right)\sigma(X_t)^2 dt = (T - t_0)\Phi_h(X, x).
\end{aligned}$$

Lemma A.4.(1) is a straightforward consequence of the previous equality and Lemma A.4.(2) is easy to establish: for every $h > 0$ and $\varphi \in C^0([t_0, T]; \mathbb{R})$,

$$\begin{aligned}
(T - t_0)^2\|\Phi_h(\varphi, .)\|_{2,\delta}^2 &\leqslant 2\int_{-\infty}^{\infty}\mathcal{K}\left(\frac{\varphi(T) - x}{h}\right)^2\delta(x)dx + 4\int_{-\infty}^{\infty}\mathcal{K}\left(\frac{\varphi(t_0) - x}{h}\right)^2\delta(x)dx \\
&\qquad + \frac{1}{h^4}\int_{-\infty}^{\infty}\left[\int_{t_0}^T K'\left(\frac{\varphi(t) - x}{h}\right)\sigma(\varphi(t))^2 dt\right]^2\delta(x)dx \\
&\leqslant 6\|\mathcal{K}\|_\infty^2 + \frac{T - t_0}{h^4}\int_{t_0}^T \sigma(\varphi(t))^4\int_{-\infty}^{\infty}K'\left(\frac{\varphi(t) - x}{h}\right)^2\delta(x)dxdt \\
&\leqslant 6\|\mathcal{K}\|_\infty^2 + \frac{(T - t_0)^2\|\delta\|_\infty\|\sigma\|_\infty^4\|K'\|_2^2}{h^3}.
\end{aligned}$$

Let us prove Lemma A.4.(3,4,5). First, for any $h, h' > 0$,

$$\begin{aligned}
&\mathbb{E}(\langle\Phi_h(X^1, .), \Phi_{h'}(X^2, .)\rangle_{2,\delta}^2) \\
&\quad = \frac{1}{(T - t_0)^4}\mathbb{E}\left[\left(\int_{-\infty}^{\infty}\left(\int_{t_0}^T K_h(X_t^1 - x)dX_t^1\right)\left(\int_{t_0}^T K_{h'}(X_t^2 - x)dX_t^2\right)\delta(x)dx\right)^2\right] \\
&\quad \leqslant \frac{2}{(T - t_0)^4}(\mathbb{E}(A_{h,h'}^2) + \mathbb{E}(B_{h,h'}^2))
\end{aligned}$$

with

$$A_{h,h'} := \int_{-\infty}^{\infty}\left(\int_{t_0}^{T}K_h(X_t^1-x)\sigma(X_t^1)dW_t^1\right)\left(\int_{t_0}^{T}K_{h'}(X_t^2-x)dX_t^2\right)\delta(x)dx \text{ and}$$
$$B_{h,h'} := \int_{-\infty}^{\infty}\left(\int_{t_0}^{T}K_h(X_t^1-x)b(X_t^1)dt\right)\left(\int_{t_0}^{T}K_{h'}(X_t^2-x)dX_t^2\right)\delta(x)dx.$$

Bound on $\mathbb{E}(A_{h,h'}^2)$. Since (X^1,W^1) and X^2 are independent,

$$\begin{aligned}\mathbb{E}(A_{h,h'}^2) = & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbb{E}\left[\left(\int_{t_0}^{T}K_h(X_t^1-x)\sigma(X_t^1)dW_t^1\right)\left(\int_{t_0}^{T}K_h(X_t^1-y)\sigma(X_t^1)dW_t^1\right)\right]\\ & \times\mathbb{E}\left[\left(\int_{t_0}^{T}K_{h'}(X_t^2-x)dX_t^2\right)\left(\int_{t_0}^{T}K_{h'}(X_t^2-y)dX_t^2\right)\right]\delta(x)\delta(y)dxdy.\end{aligned}$$

On the one hand, for every $x,y\in\mathbb{R}$, by the isometry property of Itô's integral and the definition of f,

$$\begin{aligned}&\mathbb{E}\left[\left(\int_{t_0}^{T}K_h(X_t^1-x)\sigma(X_t^1)dW_t^1\right)\left(\int_{t_0}^{T}K_h(X_t^1-y)\sigma(X_t^1)dW_t^1\right)\right]\\ &\quad=\int_{t_0}^{T}\mathbb{E}(K_h(X_t^1-x)K_h(X_t^1-y)\sigma(X_t^1)^2)dt=(T-t_0)\int_{-\infty}^{\infty}K_h(z-x)K_h(z-y)\sigma(z)^2f(z)dz.\end{aligned}$$

Then,

$$\begin{aligned}\mathbb{E}(A_{h,h'}^2) &= (T-t_0)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_h(z-x)K_h(z-y)\sigma(z)^2f(z)\\ &\qquad\times\mathbb{E}\left[\left(\int_{t_0}^{T}K_{h'}(X_t^2-x)dX_t^2\right)\left(\int_{t_0}^{T}K_{h'}(X_t^2-y)dX_t^2\right)\right]\delta(x)\delta(y)dxdydz\\ &= (T-t_0)\int_{-\infty}^{\infty}\sigma(z)^2f(z)\mathbb{E}\left[\left(\int_{-\infty}^{\infty}K_h(z-x)\delta(x)\int_{t_0}^{T}K_{h'}(X_t^2-x)dX_t^2dx\right)^2\right]dz.\end{aligned}$$

On the other hand, for every $z\in\mathbb{R}$, $x\mapsto|K_h(z-x)|/\|K\|_1$ is a density function. Then, by Jensen's inequality,

$$\begin{aligned}\mathbb{E}(A_{h,h'}^2) &\leqslant (T-t_0)\|K\|_1\int_{-\infty}^{\infty}\sigma(z)^2f(z)\int_{-\infty}^{\infty}|K_h(z-x)|\delta(x)^2\mathbb{E}\left[\left(\int_{t_0}^{T}K_{h'}(X_t^2-x)dX_t^2\right)^2\right]dxdz\\ &\leqslant (T-t_0)\|\sigma^2f\|_\infty\|K\|_1^2\|\delta\|_\infty\int_{-\infty}^{\infty}\delta(x)\mathbb{E}\left[\left(\int_{t_0}^{T}K_{h'}(X_t^2-x)dX_t^2\right)^2\right]dx\\ &\qquad\qquad\leqslant (T-t_0)^3\|\sigma^2f\|_\infty\|K\|_1^2\|\delta\|_\infty\mathfrak{m}(h').\end{aligned}$$

Bound on $\mathbb{E}(B_{h,h'}^2)$. Since $x \mapsto |K_h(X_t(\omega)-x)|/\|K\|_1$ is a density function for every $(t,\omega) \in [t_0,T]\times\Omega$, by Jensen's inequality,

$$\begin{aligned}
\mathbb{E}(B_{h,h'}^2) &= \mathbb{E}\left[\left(\int_{t_0}^{T}\int_{-\infty}^{\infty} K_h(X_t^1-x)b(X_t^1)\delta(x)\int_{t_0}^{T} K_{h'}(X_s^2-x)dX_s^2dxdt\right)^2\right]\\
&\leqslant (T-t_0)\|K\|_1\int_{t_0}^{T}\int_{-\infty}^{\infty}\mathbb{E}(|K_h(X_t^1-x)|b(X_t^1)^2)\delta(x)^2\mathbb{E}\left[\left(\int_{t_0}^{T} K_{h'}(X_s^2-x)dX_s^2\right)^2\right]dxdt\\
&= (T-t_0)^2\|K\|_1\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}|K_h(z-x)|b(z)^2f(z)dz\right)\delta(x)^2\mathbb{E}\left[\left(\int_{t_0}^{T} K_{h'}(X_s^2-x)dX_s^2\right)^2\right]dx\\
&\leqslant (T-t_0)^2\|b^2f\|_\infty\|K\|_1^2\|\delta\|_\infty\int_{-\infty}^{\infty}\delta(x)\mathbb{E}\left[\left(\int_{t_0}^{T} K_{h'}(X_s^2-x)dX_s^2\right)^2\right]dx\\
&\qquad\qquad\qquad\qquad \leqslant (T-t_0)^4\|b^2f\|_\infty\|K\|_1^2\|\delta\|_\infty\mathfrak{m}(h').
\end{aligned}$$

Now, for any $h>0$ and $\varphi \in \mathbb{L}^2(\mathbb{R},dx)$,

$$\begin{aligned}
\mathbb{E}(\langle\Phi_h(X,.),\varphi\rangle_{2,\delta}^2) &= \frac{1}{(T-t_0)^2}\mathbb{E}\left[\left(\int_{-\infty}^{\infty}\varphi(x)\delta(x)\int_{t_0}^{T} K_h(X_t-x)dX_tdx\right)^2\right]\\
&\leqslant \frac{2}{(T-t_0)^2}(\mathbb{E}(C_h^2)+\mathbb{E}(D_h^2))
\end{aligned}$$

with

$$\begin{aligned}
C_h &:= \int_{-\infty}^{\infty}\varphi(x)\delta(x)\int_{t_0}^{T} K_h(X_t-x)\sigma(X_t)dW_tdx \text{ and}\\
D_h &:= \int_{-\infty}^{\infty}\varphi(x)\delta(x)\int_{t_0}^{T} K_h(X_t-x)b(X_t)dtdx.
\end{aligned}$$

Bound on $\mathbb{E}(C_h^2)$. By the isometry property of Itô's integral and the definition of f,

$$\begin{aligned}
\mathbb{E}(C_h^2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x)\varphi(y)\delta(x)\delta(y)\int_{t_0}^{T}\mathbb{E}(K_h(X_t-x)K_h(X_t-y)\sigma(X_t)^2)dtdxdy\\
&= (T-t_0)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x)\varphi(y)\delta(x)\delta(y)\int_{-\infty}^{\infty}K_h(z-x)K_h(z-y)\sigma(z)^2f(z)dzdxdy\\
&= (T-t_0)\int_{-\infty}^{\infty}(K_h*(\varphi\delta))(z)^2\sigma(z)^2f(z)dz \leqslant (T-t_0)\|\sigma^2f\|_\infty\|K\|_1^2\|\delta\|_\infty\|\varphi\|_{2,\delta}^2.
\end{aligned}$$

Bound on $\mathbb{E}(D_h^2)$. By the definition of f,

$$\begin{aligned}
\mathbb{E}(D_h^2) &= \mathbb{E}\left[\left(\int_{t_0}^{T} b(X_t)\int_{-\infty}^{\infty}K_h(X_t-x)\varphi(x)\delta(x)dxdt\right)^2\right] \leqslant (T-t_0)\int_{t_0}^{T}\mathbb{E}(b(X_t)^2(K_h*(\varphi\delta))(X_t)^2)dt\\
&\leqslant (T-t_0)^2\int_{-\infty}^{\infty}(K_h*(\varphi\delta))(z)^2b(z)^2f(z)dz \leqslant (T-t_0)^2\|b^2f\|_\infty\|K\|_1^2\|\delta\|_\infty\|\varphi\|_{2,\delta}^2.
\end{aligned}$$

Finally, since X is a semi-martingale and since the map $(t,\omega,x)\mapsto K_h(X_t(\omega)-x)(bf)_{h'}(x)\delta(x)$ is measurable and bounded for any $h,h'\in\mathcal{H}_N$, by the stochastic Fubini theorem and Itô's formula,

$$\begin{aligned}(T-t_0)\langle\Phi_h(X,.),(bf)_{h'}\rangle_{2,\delta} &= (T-t_0)\int_{-\infty}^{\infty}\Phi_h(X,x)(bf)_{h'}(x)\delta(x)dx\\ &= \int_{t_0}^{T}\int_{-\infty}^{\infty}K_h(X_t-x)(bf)_{h'}(x)\delta(x)dxdX_t\quad\text{a.s.}\\ &= \int_{t_0}^{T}[K_h*((bf)_{h'}\delta)](X_t)dX_t\\ &= \Psi_{h,h'}(X_T)-\Psi_{h,h'}(X_{t_0})-\frac{1}{2}\int_{t_0}^{T}\psi_{h,h'}'(X_t)\sigma(X_t)^2dt\end{aligned}$$

where

$$\psi_{h,h'}:=K_h*((bf)_{h'}\delta),\quad\text{and}\quad\Psi_{h,h'}:=\mathcal{K}(./h)*((bf)_{h'}\delta)$$

is a primitive function of $\psi_{h,h'}$. On the one hand,

$$\begin{aligned}\psi_{h,h'}' &= K_h*((bf)_{h'}\delta)'\\ &= K_h*((bf)_{h'}\delta')+K_h*((K_{h'}*(bf)')\delta)\\ &= K_h*((bf)_{h'}\delta')+K_h*((K_{h'}*(bf'))\delta)+K_h*((K_{h'}*(b'f))\delta).\end{aligned}$$

Then, since bf, bf' and $b'f$ are bounded under Assumption 2.2,

$$\begin{aligned}\|\psi_{h,h'}'\|_\infty &\leqslant \|K_h\|_1\|K_{h'}*(bf)\|_\infty\|\delta'\|_\infty+\|K_h\|_1\|K_{h'}*(bf')\|_\infty\|\delta\|_\infty+\|K_h\|_1\|K_{h'}*(b'f))\|_\infty\|\delta\|_\infty\\ &\leqslant \|K\|_1^2\|bf\|_\infty\|\delta'\|_\infty+\|K\|_1^2\|bf'\|_\infty\|\delta\|_\infty+\|K\|_1^2\|b'f\|_\infty\|\delta\|_\infty<\infty.\end{aligned}$$

On the other hand,

$$\|\Psi_{h,h'}\|_\infty\leqslant\|\mathcal{K}(./h)\|_\infty\|(K_{h'}*(bf))\delta\|_1\leqslant\|\mathcal{K}\|_\infty\|\delta\|_\infty\|K\|_1\|bf\|_1<\infty.$$

This concludes the proof because

$$(T-t_0)|\langle\Phi_h(X,.),(bf)_{h'}\rangle_{2,\delta}|\leqslant 2\|\Psi_{h,h'}\|_\infty+\frac{T-t_0}{2}\|\sigma\|_\infty^2\|\psi_{h,h'}'\|_\infty\quad\text{a.s.}$$

A.7.3. *Proof of Lemma A.5.* For any $h,h'\in\mathcal{H}_N$,

$$U_{h,h'}(N)=\sum_{i\neq j}g_{h,h'}(X^i,X^j)$$

with, for every $\varphi_1,\varphi_2\in E=C^0([0,T];\mathbb{R})$,

$$g_{h,h'}(\varphi_1,\varphi_2):=\langle\Phi_h(\varphi_1,.)-(bf)_h,\Phi_{h'}(\varphi_2,.)-(bf)_{h'}\rangle_{2,\delta}.$$

On the one hand, since $\mathbb{E}(g_{h,h'}(\varphi,X))=0$ for every $\varphi\in E$, by Giné and Nickl [17], Theorem 3.4.8, there exists a universal constant $\mathfrak{m}\geqslant 1$ such that for any $\lambda>0$, with probability larger than $1-5.4e^{-\lambda}$,

$$\frac{|U_{h,h'}(N)|}{N^2}\leqslant\frac{\mathfrak{m}}{N^2}(\mathfrak{c}_{h,h'}(N)\lambda^{1/2}+\mathfrak{d}_{h,h'}(N)\lambda+\mathfrak{b}_{h,h'}(N)\lambda^{3/2}+\mathfrak{a}_{h,h'}(N)\lambda^2)$$

where the constants $\mathfrak{a}_{h,h'}(N)$, $\mathfrak{b}_{h,h'}(N)$, $\mathfrak{c}_{h,h'}(N)$ and $\mathfrak{d}_{h,h'}(N)$ are defined and controlled later. First, note that

$$U_{h,h'}(N)=\sum_{i\neq j}(\overline{g}_{h,h'}(X^i,X^j)-\widetilde{g}_{h,h'}(X^i)-\widetilde{g}_{h',h}(X^j)+\mathbb{E}(\overline{g}_{h,h'}(X^i,X^j)))\tag{18}$$

where, for every $\eta,\eta'\in\mathcal{H}_N$ and $\varphi_1,\varphi_2,\psi\in E$,

$$\overline{g}_{\eta,\eta'}(\varphi_1,\varphi_2):=\langle\Phi_\eta(\varphi_1,.),\Phi_{\eta'}(\varphi_2,.)\rangle_{2,\delta}\quad\text{and}\quad\widetilde{g}_{\eta,\eta'}(\psi):=\langle\Phi_\eta(\psi,.),(bf)_{\eta'}\rangle_{2,\delta}=\mathbb{E}(\overline{g}_{\eta,\eta'}(\psi,X)).$$

Let us now control $\mathfrak{a}_{h,h'}(N)$, $\mathfrak{b}_{h,h'}(N)$, $\mathfrak{c}_{h,h'}(N)$ and $\mathfrak{d}_{h,h'}(N)$:

- **The constant $\mathfrak{a}_{h,h'}(N)$.** Consider

$$\mathfrak{a}_{h,h'}(N) := \sup_{\varphi_1,\varphi_2\in E} |g_{h,h'}(\varphi_1,\varphi_2)|.$$

By (18), Cauchy-Schwarz's inequality and Lemma A.4,

$$\begin{aligned}\mathfrak{a}_{h,h'}(N) &\leqslant 4 \sup_{\varphi_1,\varphi_2\in E} |\langle\Phi_h(\varphi_1,.),\Phi_{h'}(\varphi_2,.)\rangle_{2,\delta}| \leqslant 4\left(\sup_{\varphi_1\in E}\|\Phi_h(\varphi_1,.)\|_{2,\delta}\right)\left(\sup_{\varphi_2\in E}\|\Phi_{h'}(\varphi_2,.)\|_{2,\delta}\right)\\ &\leqslant 4\left[\frac{6\|\mathcal{K}\|_\infty^2}{(T-t_0)^2}+\frac{\|\delta\|_\infty\|\sigma\|_\infty^4\|K'\|_2^2}{h^3}\right]^{1/2}\left[\frac{6\|\mathcal{K}\|_\infty^2}{(T-t_0)^2}+\frac{\|\delta\|_\infty\|\sigma\|_\infty^4\|K'\|_2^2}{(h')^3}\right]^{1/2} \leqslant \frac{\mathfrak{c}_1}{h_0^3}\end{aligned}$$

with

$$\mathfrak{c}_1 = 4\left[\frac{6\|\mathcal{K}\|_\infty^2}{(T-t_0)^2}+\|\delta\|_\infty\|\sigma\|_\infty^4\|K'\|_2^2\right].$$

So, since $(Nh_0^3)^{-1}\leqslant 1$,

$$\frac{\mathfrak{a}_{h,h'}(N)\lambda^2}{N^2} \leqslant \frac{\mathfrak{c}_1\lambda^2}{N^2h_0^3} \leqslant \frac{\mathfrak{c}_1\lambda^2}{N}.$$

- **The constant $\mathfrak{b}_{h,h'}(N)$.** Consider

$$\mathfrak{b}_{h,h'}(N)^2 := N\sup_{\varphi\in E}\mathbb{E}(g_{h,h'}(\varphi,X)^2).$$

By (18), Cauchy-Schwarz's inequality and Lemma A.4,

$$\begin{aligned}\mathfrak{b}_{h,h'}(N)^2 &\leqslant 16N\sup_{\varphi\in E}\mathbb{E}(\langle\Phi_h(\varphi,.),\Phi_{h'}(X,.)\rangle_{2,\delta}^2)\\ &\leqslant 16N\mathbb{E}(\|\Phi_{h'}(X,.)\|_{2,\delta}^2)\sup_{\varphi\in E}\|\Phi_h(\varphi,.)\|_{2,\delta}^2 \leqslant \frac{\mathfrak{c}_2\mathfrak{m}(h')N}{h^3} \quad \text{with} \quad \mathfrak{c}_2 = 4\mathfrak{c}_1.\end{aligned}$$

So, for any $\theta\in(0,1)$, since $(Nh_0^3)^{-1}\leqslant 1$,

$$\begin{aligned}\frac{\mathfrak{b}_{h,h'}(N)\lambda^{3/2}}{N^2} &\leqslant 2\left(\frac{\theta}{3\mathfrak{m}}\right)^{1/2}\frac{\mathfrak{m}(h')^{1/2}}{Nh^{3/2}}\times\left(\frac{3\mathfrak{m}}{\theta}\right)^{1/2}\frac{\mathfrak{c}_2^{1/2}\lambda^{3/2}}{N^{1/2}}\\ &\leqslant \frac{\theta\mathfrak{m}(h')}{3\mathfrak{m}N^2h^3}+\frac{3\mathfrak{c}_2\mathfrak{m}\lambda^3}{\theta N} \leqslant \frac{\theta\mathfrak{m}(h')}{3\mathfrak{m}N}+\frac{3\mathfrak{c}_2\mathfrak{m}\lambda^3}{\theta N}.\end{aligned}$$

- **The constant $\mathfrak{c}_{h,h'}(N)$.** Consider

$$\mathfrak{c}_{h,h'}(N)^2 := N^2\mathbb{E}(g_{h,h'}(X^1,X^2)^2).$$

By (18) and Lemma A.4,

$$\begin{aligned}\mathfrak{c}_{h,h'}(N)^2 &\leqslant 16N^2\mathbb{E}(\langle\Phi_h(X^1,.),\Phi_{h'}(X^2,.)\rangle_{2,\delta}^2)\\ &\leqslant \mathfrak{c}_3\mathfrak{m}(h')N^2 \quad \text{with} \quad \mathfrak{c}_3 = 16\mathfrak{c}_{A.4,1}.\end{aligned}$$

So, as previously,

$$\frac{\mathfrak{c}_{h,h'}(N)\lambda^{1/2}}{N^2} \leqslant \frac{\theta\mathfrak{m}(h')}{3\mathfrak{m}N}+\frac{3\mathfrak{c}_3\mathfrak{m}\lambda}{\theta N}.$$

- **The constant $\mathfrak{d}_{h,h'}(N)$.** Consider

$$\mathfrak{d}_{h,h'}(N) := \sup_{(a,b)\in\mathcal{A}}\mathbb{E}\left[\sum_{i<j}a_i(X^i)b_j(X^j)g_{h,h'}(X^i,X^j)\right],$$

where

$$\mathcal{A} := \left\{(a,b) : \sum_{i=1}^{N-1}\mathbb{E}(a_i(X^i)^2)\leqslant 1 \text{ and } \sum_{j=2}^{N}\mathbb{E}(b_j(X^j)^2)\leqslant 1\right\}.$$

By (18), Cauchy-Schwarz's inequality, Jensen's inequality and Lemma A.4,

$$\begin{aligned}\mathfrak{d}_{h,h'}(N) &\leqslant 4\sup_{(a,b)\in\mathcal{A}}\mathbb{E}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}|a_i(X^i)b_j(X^j)\overline{g}_{h,h'}(X^i,X^j)|\right)\\ &\leqslant 4N\mathbb{E}(\langle\Phi_h(X^1,.),\Phi_{h'}(X^2,.)\rangle_{2,\delta}^2)^{1/2}\leqslant \mathfrak{c}_3^{1/2}\mathfrak{m}(h')^{1/2}N.\end{aligned}$$

So, as previously,

$$\frac{\mathfrak{d}_{h,h'}(N)\lambda}{N^2}\leqslant\frac{\theta\mathfrak{m}(h')}{3\mathfrak{m}N}+\frac{3\mathfrak{c}_3\mathfrak{m}\lambda^2}{\theta N}.$$

Therefore, there exists a deterministic constant $\mathfrak{c}_4>0$, not depending on N, h and h', such that with probability larger than $1-5.4e^{-\lambda}$,

$$\frac{|U_{h,h'}(N)|}{N^2}\leqslant\frac{\theta\mathfrak{m}(h')}{N}+\frac{\mathfrak{c}_4(1+\lambda)^3}{\theta N}.$$

In conclusion, with probability larger than $1-5.4|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h\in\mathcal{H}_N}\left\{\frac{|U_{h,h_0}(N)|}{N^2}-\frac{\theta\mathfrak{m}(h)}{N}\right\}\leqslant\frac{\mathfrak{c}_4(1+\lambda)^3}{\theta N}$$

and

$$\sup_{h\in\mathcal{H}_N}\left\{\frac{|U_{h,h}(N)|}{N^2}-\frac{\theta\mathfrak{m}(h)}{N}\right\}\leqslant\frac{\mathfrak{c}_4(1+\lambda)^3}{\theta N}.$$

A.7.4. *Proof of Lemma A.6.* First, the two following results are used several times in the sequel:

$$\begin{aligned}\|(bf)_h\|_{2,\delta}^2 &\leqslant \|\delta\|_\infty\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K_h(y-x)b(y)f(y)dy\right)^2dx\\ &\leqslant \|\delta\|_\infty\int_{-\infty}^{\infty}b(y)^2f(y)\int_{-\infty}^{\infty}K_h(y-x)^2dxdy\leqslant\frac{\|\delta\|_\infty\|K\|_2^2\|b^2f\|_1}{h}\end{aligned}\tag{19}$$

and

$$\begin{aligned}\mathbb{E}(V_h(N)) &= \mathbb{E}(\|\Phi_h(X,.)-(bf)_h\|_{2,\delta}^2)\\ &= \mathbb{E}(\|\Phi_h(X,.)\|_{2,\delta}^2)+\|(bf)_h\|_{2,\delta}^2-2\int_{-\infty}^{\infty}(bf)_h(x)\mathbb{E}(\Phi_h(X,x))\delta(x)dx\\ &= \mathbb{E}(\|\Phi_h(X,.)\|_{2,\delta}^2)-\|(bf)_h\|_{2,\delta}^2.\end{aligned}\tag{20}$$

Consider

$$v_h(N):=V_h(N)-\mathbb{E}(V_h(N))=\frac{1}{N}\sum_{i=1}^{N}(g_h(X^i)-\mathbb{E}(g_h(X^i)))$$

with

$$g_h(\varphi):=\|\Phi_h(\varphi,.)-(bf)_h\|_{2,\delta}^2\ ;\ \forall\varphi\in E.$$

By Bernstein's inequality, for any $\lambda>0$, with probability larger than $1-2e^{-\lambda}$,

$$|v_h(N)|\leqslant\sqrt{\frac{2\mathfrak{v}_h\lambda}{N}}+\frac{\mathfrak{c}_h\lambda}{N}$$

where

$$\mathfrak{c}_h=\frac{\|g_h\|_\infty}{3}\quad\text{and}\quad\mathfrak{v}_h=\mathbb{E}(g_h(X)^2).$$

Moreover, by Inequality (19) and Lemma A.4,

$$\begin{aligned}\mathfrak{c}_h &= \frac{1}{3}\sup_{\varphi\in E}\|\Phi_h(\varphi,.)-(bf)_h\|_{2,\delta}^2\leqslant\frac{2}{3}\left(\sup_{\varphi\in E}\|\Phi_h(\varphi,.)\|_{2,\delta}^2+\|(bf)_h\|_{2,\delta}^2\right)\\ &\leqslant\frac{\mathfrak{c}_1}{h^3}\quad\text{with}\quad\mathfrak{c}_1=\frac{2}{3}\left[\frac{6\|\mathcal{K}\|_\infty^2}{(T-t_0)^2}+\|\delta\|_\infty\|\sigma\|_\infty^4\|K'\|_2^2+\|\delta\|_\infty\|K\|_2^2\|b^2f\|_1\right]\end{aligned}$$

and, by Inequality (19), Equality (20) and Lemma A.4,

$$\begin{aligned}\mathfrak{v}_h &\leqslant \|g_h\|_\infty \mathbb{E}(V_h(N)) \leqslant \frac{3\mathfrak{c}_1}{h^3}(\mathbb{E}(\|\Phi_h(X,.)\|_{2,\delta}^2) - \|(bf)_h\|_{2,\delta}^2)\\ &\leqslant \frac{\mathfrak{c}_2\mathfrak{m}(h)}{h^3} \quad \text{with} \quad \mathfrak{c}_2 = 3\mathfrak{c}_1.\end{aligned}$$

Then, for any $\theta \in (0,1)$, since $(Nh_0^3)^{-1} \leqslant 1$, with probability larger than $1 - 2e^{-\lambda}$,

$$\begin{aligned}|v_h(N)| &\leqslant 2\sqrt{\frac{\mathfrak{c}_2\mathfrak{m}(h)\lambda}{Nh^3}} + \frac{\mathfrak{c}_1\lambda}{Nh^3}\\ &\leqslant \theta\mathfrak{m}(h) + \frac{(\mathfrak{c}_1+\mathfrak{c}_2)\lambda}{\theta Nh^3} \leqslant \theta\mathfrak{m}(h) + \frac{(\mathfrak{c}_1+\mathfrak{c}_2)\lambda}{\theta}.\end{aligned}$$

So, with probability larger than $1 - 2|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h\in\mathcal{H}_N}\left\{\frac{|v_h(N)|}{N} - \frac{\theta\mathfrak{m}(h)}{N}\right\} \leqslant \frac{(\mathfrak{c}_1+\mathfrak{c}_2)\lambda}{\theta N}.$$

Therefore, by Equality (20), with probability larger that $1 - 2|\mathcal{H}_N|e^{-\lambda}$,

$$\begin{aligned}&\sup_{h\in\mathcal{H}_N}\left\{\frac{1}{N}|V_h(N) - \mathbb{E}(\|\Phi_h(X,.)\|_{2,\delta}^2)| - \frac{\theta\mathfrak{m}(h)}{N}\right\}\\ &\qquad\leqslant \sup_{h\in\mathcal{H}_N}\left\{\frac{|v_h(N)|}{N} - \frac{\theta\mathfrak{m}(h)}{N}\right\} + \frac{1}{N}\|K_h*(bf)\|_{2,\delta}^2 \leqslant \frac{(\mathfrak{c}_1+\mathfrak{c}_2+\|\delta\|_\infty\|K\|_1^2\|bf\|_2^2)(1+\lambda)}{\theta N}.\end{aligned}$$

A.7.5. *Proof of Lemma A.7.* For any $h, h' \in \mathcal{H}_N$,

$$W_{h,h'}(N) = \frac{1}{N}\sum_{i=1}^N (g_{h,h'}(X^i) - \mathbb{E}(g_{h,h'}(X^i)))$$

with, for every $\varphi \in E$,

$$g_{h,h'}(\varphi) := \langle\Phi_h(\varphi,.), (bf)_{h'} - bf\rangle_{2,\delta}.$$

By Bernstein's inequality, for any $\lambda > 0$, with probability larger than $1 - 2e^{-\lambda}$,

$$|W_{h,h'}(N)| \leqslant \sqrt{\frac{2\mathfrak{v}_{h,h'}\lambda}{N}} + \frac{\mathfrak{c}_{h,h'}\lambda}{N}$$

where

$$\mathfrak{c}_{h,h'} = \frac{\|g_{h,h'}\|_\infty}{3} \quad \text{and} \quad \mathfrak{v}_{h,h'} = \mathbb{E}(g_{h,h'}(X)^2).$$

Moreover, by Lemma A.4,

$$\begin{aligned}\mathfrak{c}_{h,h'} &= \frac{1}{3}\sup_{\varphi\in E}|\langle\Phi_h(\varphi,.), (bf)_{h'} - bf\rangle_{2,\delta}| \leqslant \frac{1}{3}\|(bf)_{h'} - bf\|_{2,\delta}\sup_{\varphi\in E}\|\Phi_h(\varphi,.)\|_{2,\delta}\\ &\leqslant \frac{\mathfrak{c}_1}{h^{3/2}}\|(bf)_{h'} - bf\|_{2,\delta} \quad \text{with} \quad \mathfrak{c}_1 = \frac{1}{3}\left[\frac{6\|\mathcal{K}\|_\infty^2}{(T-t_0)^2} + \|\delta\|_\infty\|\sigma\|_\infty^4\|K'\|_2^2\right]^{1/2}\end{aligned}$$

and

$$\mathfrak{v}_{h,h'} \leqslant \mathbb{E}(\langle\Phi_h(X,.), (bf)_{h'} - bf\rangle_{2,\delta}^2) \leqslant \mathfrak{c}_{A.4,2}\|(bf)_{h'} - bf\|_{2,\delta}^2.$$

Then, for any $\theta \in (0,1)$, with probability larger than $1 - 2e^{-\lambda}$,

$$\begin{aligned}|W_{h,h'}(N)| &\leqslant 2\sqrt{\frac{\mathfrak{c}_{A.4,2}\lambda}{N}\|(bf)_{h'} - bf\|_{2,\delta}^2} + \frac{\mathfrak{c}_1\lambda}{Nh^{3/2}}\|(bf)_{h'} - bf\|_{2,\delta}\\ &\leqslant \theta\|(bf)_{h'} - bf\|_{2,\delta}^2 + \frac{2\mathfrak{c}_{A.4,2}\lambda}{\theta N} + \frac{2\mathfrak{c}_1^2\lambda^2}{\theta N^2h^3} \leqslant \theta\|(bf)_{h'} - bf\|_{2,\delta}^2 + \frac{2(\mathfrak{c}_{A.4,2}+\mathfrak{c}_1^2)(1+\lambda)^2}{\theta N}.\end{aligned}$$

So, with probability larger than $1-2|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h\in\mathcal{H}_N}\{|W_{h,h_0}(N)|-\theta\|(bf)_{h_0}-bf\|_{2,\delta}^2\}\leqslant\frac{\mathfrak{c}_{A.7}(1+\lambda)^2}{\theta N},$$

$$\sup_{h\in\mathcal{H}_N}\{|W_{h_0,h}(N)|-\theta\|(bf)_{h}-bf\|_{2,\delta}^2\}\leqslant\frac{\mathfrak{c}_{A.7}(1+\lambda)^2}{\theta N}\text{ and}$$

$$\sup_{h\in\mathcal{H}_N}\{|W_{h,h}(N)|-\theta\|(bf)_{h}-bf\|_{2,\delta}^2\}\leqslant\frac{\mathfrak{c}_{A.7}(1+\lambda)^2}{\theta N}.$$

A.8. **Proof of Corollary 5.3.** On the one hand, as in the proof of Proposition 3.5 and since $\delta(x)>m$ for every $x\in[A,B]$,

$$\begin{aligned}\mathbb{E}(\|\widehat{b}_{N,\widehat{h},\widehat{h}'}-b\|_{f,A,B}^2)&\leqslant\frac{\mathfrak{c}_{3.5}}{m^2}[\mathbb{E}(\|\widehat{bf}_{N,\widehat{h}}-bf\|_{2,A,B}^2)+2\mathbb{E}(\|\widehat{f}_{N,\widehat{h}'}-f\|_2^2)]\\&\leqslant\frac{2\mathfrak{c}_{3.5}}{m^3}[\mathbb{E}(\|\widehat{bf}_{N,\widehat{h}}-bf\|_{2,\delta}^2)+\mathbb{E}(\|\widehat{f}_{N,\widehat{h}'}-f\|_2^2)].\end{aligned}$$

On the other hand, by Theorem 5.2 and *union bounds*,

$$\begin{aligned}\mathbb{E}(\|\widehat{bf}_{N,\widehat{h}}-bf\|_{2,\delta}^2)&\leqslant(1+\vartheta)\min_{h\in\mathcal{H}_N}\mathbb{E}(\|\widehat{bf}_{N,h}-bf\|_{2,\delta}^2)+\frac{\mathfrak{c}_{5.2,2}}{\vartheta}\left(\|(bf)_{h_0}-bf\|_{2,\delta}^2+\frac{1}{N}\right)\\&\leqslant(1\vee\|\delta\|_\infty)\left[(1+\vartheta)\min_{h\in\mathcal{H}_N}\mathbb{E}(\|\widehat{bf}_{N,h}-bf\|_2^2)+\frac{\mathfrak{c}_{5.2,2}}{\vartheta}\left(\|(bf)_{h_0}-bf\|_2^2+\frac{1}{N}\right)\right]\end{aligned}$$

and

$$\mathbb{E}(\|\widehat{f}_{N,\widehat{h}'}-f\|_2^2)\leqslant(1+\vartheta)\min_{h'\in\mathcal{H}_N}\mathbb{E}(\|\widehat{f}_{N,h'}-f\|_2^2)+\frac{\overline{\mathfrak{c}}_{5.2,2}}{\vartheta}\left(\|f_{h_0'}-f\|_2^2+\frac{1}{N}\right).$$

Therefore,

$$\begin{aligned}\mathbb{E}(\|\widehat{b}_{N,\widehat{h},\widehat{h}'}-b\|_{f,A,B}^2)\leqslant\frac{2\mathfrak{c}_{3.5}(1\vee\|\delta\|_\infty)}{m^3}\Bigg[&(1+\vartheta)\min_{(h,h')\in\mathcal{H}_N\times\mathcal{H}_N'}\{\mathbb{E}(\|\widehat{bf}_{N,h}-bf\|_2^2)+\mathbb{E}(\|\widehat{f}_{N,h'}-f\|_2^2)\}\\&+\frac{\mathfrak{c}_{5.3}}{\vartheta}\left(\|(bf)_{h_0}-bf\|_2^2+\|f_{h_0'}-f\|_2^2+\frac{1}{N}\right)\Bigg].\end{aligned}$$

Acknowledgments. Many thanks to Fabienne Comte and Valentine Genon-Catalot for their constructive comments about this work.

†,◇Laboratoire Modal'X, Université Paris Nanterre, Nanterre, France
Email address: nmarie@parisnanterre.fr

◇ESME Sudria, Paris, France
Email address: amelie.rosier@esme.fr