



HAL
open science

Warped Kernel Estimator for I.I.D. Paths of Diffusion Processes

Nicolas Marie, Amélie Rosier

► **To cite this version:**

Nicolas Marie, Amélie Rosier. Warped Kernel Estimator for I.I.D. Paths of Diffusion Processes. 2024. hal-04484908

HAL Id: hal-04484908

<https://hal.science/hal-04484908>

Preprint submitted on 1 Mar 2024

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

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

WARPED KERNEL ESTIMATOR FOR I.I.D. PATHS OF DIFFUSION PROCESSES

NICOLAS MARIE[†] AND AMÉLIE ROSIER^{◊,*}

ABSTRACT. This paper deals with a nonparametric warped kernel estimator \widehat{b} of the drift function computed from independent continuous observations of a diffusion process. A risk bound on \widehat{b} is established. The paper also deals with an extension of the PCO bandwidth selection method for \widehat{b} . Finally, some numerical experiments are provided.

MSC2020 subject classifications. 62G05 ; 62M05.

CONTENTS

1. Introduction	1
2. Risk bound on the warped kernel estimator	3
3. A PCO type bandwidth selection method	4
4. Numerical experiments	5
5. Proofs	7
5.1. Proof of Proposition 2.4	7
5.2. Proof of Theorem 3.2	8
References	17

1. INTRODUCTION

Consider $T > 0$ and the stochastic differential equation

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

where $b, \sigma : \mathbb{R} \rightarrow \mathbb{R}$ 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 Hoffman [22], Kessler [24], Kutoyants [26], Dalalyan [11], Comte et al. [7], etc.), the estimators of the drift function are computed from one path of the stationary solution of Equation (1), which exists and is unique under a restrictive dissipativity condition on b , and converge when T goes to infinity.

Let $\mathcal{I}(\cdot)$ be the Itô map for Equation (1) and, for $N \in \mathbb{N}^*$ copies W^1, \dots, W^N of W , consider

$$X^i := \mathcal{I}(x_0, W^i) ; \forall i \in \{1, \dots, N\}.$$

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 [18], Overgaard et al. [29], Picchini, De Gaetano and Ditlevsen [30], Picchini and Ditlevsen [31], Comte, Genon-Catalot and Samson [8], Delattre and Lavielle [14], Delattre, Genon-Catalot and Samson [13], Dion and Genon-Catalot [17], Delattre, Genon-Catalot and Larédo [12], etc.) and more recently in the nonparametric one (see Comte and Genon-Catalot [5, 6], Comte and Marie [10], Della Maestra and Hoffmann [15], Marie and Rosier [27], Denis et al. [16]).

Key words and phrases. Diffusion processes ; Nonparametric drift estimation ; Warped Kernel estimator ; PCO method.

Under the appropriate conditions on b and σ recalled in Section 2, for every $t \in (0, T]$, the distribution of X_t has a density $p_t(x_0, \cdot)$ such that $s \mapsto p_s(x_0, x)$ belongs to $\mathbb{L}^1([0, T])$ for every $x \in \mathbb{R}$. Then, for a given $t_0 \in [0, T]$, one may define

$$f(x) := \frac{1}{T - t_0} \int_{t_0}^T p_t(x_0, x) dt \quad \text{and} \quad F(x) := \int_{-\infty}^x f(z) dz$$

for every $x \in \mathbb{R}$. Clearly, f is a density function:

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

Let $K : \mathbb{R} \rightarrow \mathbb{R}$ 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 $h \in (0, 1]$. Our paper deals with the warped kernel estimator

$$\widehat{b}_{N,h}(x) := \widehat{\beta}_{N,h}(F; F(x)) ; x \in \mathbb{R}$$

of the drift function b , where

$$\widehat{\beta}_{N,h}(\varphi; z) := \frac{1}{N(T - t_0)} \sum_{i=1}^N \int_{t_0}^T K_h(z - \varphi(X_t^i)) dX_t^i$$

for every $\varphi \in C^0(\mathbb{R})$ and $z \in \mathbb{R}$. From independent copies of X continuously observed on $[0, T]$, $\widehat{b}_{N,h}$ is a natural extension of the warped kernel estimator already well-studied in the nonparametric regression framework (see Chagny [1]). The paper also deals with the adaptive estimator

$$\widehat{b}_N(x) := \widehat{b}_{N,\widehat{h}}(x),$$

where \widehat{h} is selected via a penalized comparison to overfitting (PCO) type criterion. Finally, in practice, the function F is unknown and has to be replaced by the estimator

$$\widehat{F}_N(x) := \frac{1}{N(T - t_0)} \sum_{i=1}^N \int_{t_0}^T \mathbf{1}_{X_t^i \leq x} dt$$

in the definition of $\widehat{b}_{N,h}$. Precisely, $\widehat{b}_{N,h}$ is approximated by

$$\widetilde{b}_{N,h}(x) := \widehat{\beta}_{N,h}(\widehat{F}_N; \widehat{F}_N(x)).$$

Section 2 deals with a risk bound on the warped kernel estimator and Section 3 with a risk bound on the adaptive estimator \widehat{b}_N . Section 4 deals with some numerical experiments on $\widetilde{b}_{N,h}$. The proofs are postponed to Section 5.

Notations and basic definitions:

- The space $C^0(\mathbb{R})$ is equipped with the uniform (semi-)norm $\|\cdot\|_{\infty}$.
- For every $p \in \overline{\mathbb{N}}$, $C_b^p(\mathbb{R}) := \cap_{j=0}^p \{\varphi \in C^p(\mathbb{R}) : \varphi^{(j)} \text{ is bounded}\}$.
- For every $p \geq 1$, $\mathbb{L}^p(\mathbb{R}, dx)$ is equipped with its usual norm $\|\cdot\|_p$:

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

- \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 $\langle \cdot, \cdot \rangle_{\delta}$, and the associated norm by $\|\cdot\|_{\delta}$. For every $A, B \in \overline{\mathbb{R}}$ such that $A < B$, the usual norm on $\mathbb{L}^2([A, B], \delta(x) dx)$ is denoted by $\|\cdot\|_{\delta, A, B}$.

2. RISK BOUND ON THE WARPED KERNEL ESTIMATOR

In the sequel, in order to ensure the existence and the uniqueness of the (strong) solution of Equation (1), and to ensure that the distribution of X_t has a regular enough density $p_t(x_0, \cdot)$ for every $t \in (0, T]$ (see Remark 2.2), b and σ fulfill the following assumption.

Assumption 2.1. *The function b is Lipschitz continuous, $\sigma \in C_b^1(\mathbb{R})$, σ' is Hölder continuous and there exists $\alpha > 0$ such that*

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

Remark 2.2. *By Menozzi et al. [28], Theorem 1.2, for any $t \in (0, T]$, the distribution of X_t has a continuously differentiable and sub-Gaussian density $p_t(x_0, \cdot)$ such that $p_t(x_0, \mathbb{R}) \subset (0, \infty)$, $z \in \mathbb{R} \mapsto \partial_z p_t(x_0, z)$ is also sub-Gaussian and $s \in (0, T] \mapsto p_s(x_0, z)$ belongs to $\mathbb{L}^1([0, T], dt)$ for every $z \in \mathbb{R}$. First, by this last point,*

$$F(\cdot) = \int_{-\infty}^{\cdot} f(z) dz = \frac{1}{T - t_0} \int_{t_0}^T F_s(\cdot) ds \quad \text{with} \quad F_s(\cdot) := \int_{-\infty}^{\cdot} p_s(x_0, z) dz ; \forall s \in (0, T]$$

is well-defined (even when $t_0 = 0$). Moreover, F_t (and then F) is strictly increasing and one-to-one from \mathbb{R} into $(0, 1)$ because $p_t(x_0, \cdot)$ is continuous and $p_t(x_0, \mathbb{R}) \subset (0, \infty)$. Finally, since $p_t(x_0, \cdot)$ is sub-Gaussian, b and σ belong to $\mathbb{L}^2(\mathbb{R}, f(z) dz)$.

The kernel K fulfills the following usual assumption.

Assumption 2.3. *The kernel K is symmetric, continuous and belongs to $\mathbb{L}^2(\mathbb{R}, dx)$.*

The following proposition deals with a risk bound on $\widehat{b}_{N,h}$.

Proposition 2.4. *Consider $A, B \in \overline{\mathbb{R}}$ such that $A < B$. Under Assumptions 2.1 and 2.3,*

$$\mathbb{E}(\|\widehat{b}_{N,h} - b\|_{f,A,B}^2) \leq \|b_h - b\|_{f,A,B}^2 + \frac{\mathfrak{c}_{2.4}}{Nh}$$

with

$$b_h := [K_h * ((b \circ F^{-1}) \mathbf{1}_{(0,1)})](F(\cdot)) \quad \text{and} \quad \mathfrak{c}_{2.4} = 2\|K\|_2^2 \left(\|b\|_f^2 + \frac{1}{T - t_0} \|\sigma\|_f^2 \right).$$

Let us conclude this section with some remarks about Proposition 2.4:

- Note that the variance term in the risk bound on $\widehat{b}_{N,h}$ stated in Proposition 2.4 is of same order than in the nonparametric regression framework (see Chagny [1]).
- If $b \in \mathbb{L}^2(\mathbb{R}, dx)$, then

$$b_h \xrightarrow[h \rightarrow 0^+]{\mathbb{L}^2} (b \circ F^{-1})(F(\cdot)) = b.$$

- Assume that K is a $[-1, 1]$ -supported kernel, $A, B \in \mathbb{R}$ and that $h \in (0, \kappa(F(A) \wedge (1 - F(B))))$ with $\kappa \in (0, 1)$. Then, for every $x \in [A, B]$,

$$[-1, 1] \subset \left[-\frac{F(A)}{h}, \frac{1 - F(B)}{h} \right] \subset \left[-\frac{F(x)}{h}, \frac{1 - F(x)}{h} \right]$$

and

$$F^{-1}(h \cdot + F(x))([-1, 1]) \subset [F^{-1}((1 - \kappa)F(A)), F^{-1}(\kappa + (1 - \kappa)F(B))] =: I_{A,B,F}.$$

Moreover, since $f(\mathbb{R}) \subset (0, \infty)$, there exists $f_1 > 0$ such that $f(z) \geq f_1$ for every $z \in I_{A,B,F}$. So,

$$\begin{aligned} \|b_h - b\|_{f,A,B}^2 &= \int_A^B \left(\int_0^1 K_h(y - F(x))(b \circ F^{-1})(y) dy - b(x) \right)^2 f(x) dx \\ &= \int_A^B \left[\int_{-F(x)/h}^{(1-F(x))/h} K(y) b(F^{-1}(hy + F(x))) dy - b(x) \right]^2 f(x) dx \\ &= \int_A^B \left(\int_{-1}^1 K(y) [b(F^{-1}(hy + F(x))) - b(F^{-1}(F(x)))] dy \right)^2 f(x) dx \\ &\leq h^2 \|b\|_{\text{Lip}}^2 \left(\sup_{z \in I_{A,B,F}} \frac{1}{(f \circ F^{-1})(z)} \right)^2 \\ &\quad \times \int_{-1}^1 K(y) y^2 \int_A^B f(x) dx dy \leq c_1 h^2 \quad \text{with} \quad c_1 = \frac{\|b\|_{\text{Lip}}^2}{f_1^2} \int_{-1}^1 K(y) y^2 dy. \end{aligned}$$

Therefore, the bias term in the risk bound on $\widehat{b}_{N,h}$ stated in Proposition 2.4 is of order h^2 , as in the nonparametric regression framework (see Chagny [1]).

3. A PCO TYPE BANDWIDTH SELECTION METHOD

This section deals with an extension of the PCO method to the warped kernel estimator studied in this paper. The PCO based adaptive warped kernel estimator offers theoretical guarantees: a risk bound is established in this section.

In this section, consider an additional kernel δ and assume that K , δ and σ fulfill the following technical assumption.

Assumption 3.1. *The kernels K and δ are compactly supported and continuously differentiable on \mathbb{R} , $\delta(I) \subset (0, \infty)$ for every compact subset I of $\text{supp}(\delta)^\circ$, and σ is bounded.*

For instance, the kernel

$$\rho : x \in \mathbb{R} \mapsto \frac{1}{c_\rho} \exp\left(-\frac{1}{1-x^2}\right) \mathbf{1}_{[-1,1]}(x) \quad \text{with} \quad c_\rho = \int_{-1}^1 \exp\left(-\frac{1}{1-y^2}\right) dy$$

satisfies the conditions on both K and δ in Assumption 3.1. In the sequel, for the sake of simplicity, $\text{supp}(K) = [-1, 1]$ and $\text{supp}(\delta) = [-\Delta, \Delta]$ with $\Delta > 0$. Now, let \mathcal{H}_N be a finite subset of $[h_0, \Delta_0]$, where

$$\Delta_0 := \kappa(F(-\Delta) \wedge (1 - F(\Delta))), \quad \kappa \in (0, 1), \quad h_0 \in (0, \Delta_0) \quad \text{and} \quad \frac{\Delta_0^3}{N h_0^3} \leq 1.$$

Finally, consider

$$(2) \quad \widehat{h} \in \arg \min_{h \in \mathcal{H}_N} \{ \|\widehat{b}_{N,h} - \widehat{b}_{N,h_0}\|_\delta^2 + \text{pen}(h) \}$$

with

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

The following theorem deals with a risk bound on the adaptive estimator

$$\widehat{b}_N(x) := \widehat{b}_{N,\widehat{h}}(x); \quad x \in \mathbb{R}.$$

Theorem 3.2. *Under Assumptions 2.1, 2.3 and 3.1, if $t_0 > 0$, then there exist two deterministic constants $\mathfrak{c}_{3.2,1}, \mathfrak{c}_{3.2,2} > 0$, not depending on N , such that for every $\vartheta \in (0, 1)$ and $\lambda > 0$, with probability larger than $1 - \mathfrak{c}_{3.2,1} |\mathcal{H}_N| e^{-\lambda}$,*

$$\|\widehat{b}_N - b\|_\delta^2 \leq (1 + \vartheta) \min_{h \in \mathcal{H}_N} \|\widehat{b}_{N,h} - b\|_\delta^2 + \frac{\mathfrak{c}_{3.2,2}}{\vartheta} \left[\|b_{h_0} - b\|_\delta^2 + \frac{(1 + \lambda)^3}{N} \right].$$

Theorem 3.2 allows to establish a risk bound on our adaptive warped estimator for the usual norm $\|\cdot\|_{2,A,B}$ on $\mathbb{L}^2([A, B], dx)$, where $A, B \in \mathbb{R}$ satisfy $-\Delta < A < B < \Delta$. First, since $\delta(I) \subset (0, \infty)$ for every compact subset I of $\text{supp}(\delta)^\circ = (-\Delta, \Delta)$, there exists $\delta_0 > 0$ such that $\delta(x) > \delta_0$ for every $x \in [A, B]$. Then, for any $\vartheta \in (0, 1)$ and $\lambda > 0$, with probability larger than $1 - \mathfrak{c}_{3.2,1}|\mathcal{H}_N|e^{-\lambda}$,

$$\begin{aligned} \|\widehat{b}_N - b\|_{2,A,B}^2 &\leq \frac{1}{\delta_0} \|\widehat{b}_N - b\|_\delta^2 \\ &\leq \frac{1}{\delta_0} \left[(1 + \vartheta) \min_{h \in \mathcal{H}_N} \|\widehat{b}_{N,h} - b\|_\delta^2 + \frac{\mathfrak{c}_{3.2,2}}{\vartheta} \left[\|b_{h_0} - b\|_\delta^2 + \frac{(1 + \lambda)^3}{N} \right] \right]. \end{aligned}$$

Moreover, since $\text{supp}(\delta) = [-\Delta, \Delta]$ and $f(\mathbb{R}) \subset (0, \infty)$, there exists $f_1 > 0$ such that $f(x) \geq f_1$ for every $x \in \text{supp}(\delta)$, and then

$$\begin{aligned} (1 + \vartheta) \min_{h \in \mathcal{H}_N} \|\widehat{b}_{N,h} - b\|_\delta^2 + \frac{\mathfrak{c}_{3.2,2}}{\vartheta} \left[\|b_{h_0} - b\|_\delta^2 + \frac{(1 + \lambda)^3}{N} \right] \\ \leq \frac{1 \vee \|\delta\|_\infty}{1 \wedge f_1} \left[(1 + \vartheta) \min_{h \in \mathcal{H}_N} \|\widehat{b}_{N,h} - b\|_{f,-\Delta,\Delta}^2 + \frac{\mathfrak{c}_{3.2,2}}{\vartheta} \left[\|b_{h_0} - b\|_{f,-\Delta,\Delta}^2 + \frac{(1 + \lambda)^3}{N} \right] \right]. \end{aligned}$$

Therefore, with probability larger than $1 - \mathfrak{c}_{3.2,1}|\mathcal{H}_N|e^{-\lambda}$,

$$\|\widehat{b}_N - b\|_{2,A,B}^2 \leq \frac{1 \vee \|\delta\|_\infty}{\delta_0(1 \wedge f_1)} \left[(1 + \vartheta) \min_{h \in \mathcal{H}_N} \|\widehat{b}_{N,h} - b\|_{f,-\Delta,\Delta}^2 + \frac{\mathfrak{c}_{3.2,2}}{\vartheta} \left[\|b_{h_0} - b\|_{f,-\Delta,\Delta}^2 + \frac{(1 + \lambda)^3}{N} \right] \right]$$

and, by Proposition 2.4, there exist two constants $\mathfrak{c}_1, \mathfrak{c}_2 > 0$, not depending on N , such that

$$\mathbb{E} \left(\min_{h \in \mathcal{H}_N} \|\widehat{b}_{N,h} - b\|_{f,-\Delta,\Delta}^2 \right) \leq \mathfrak{c}_1 \min_{h \in \mathcal{H}_N} \left\{ h^2 + \frac{1}{Nh} \right\} \quad \text{and} \quad \|b_{h_0} - b\|_{f,-\Delta,\Delta}^2 \leq \mathfrak{c}_2 h_0^2.$$

4. NUMERICAL EXPERIMENTS

First of all, recall that since the function F is unknown in practice, $\widehat{b}_{N,h}(x)$ is approximated by

$$\widetilde{b}_{N,h}(x) = \widehat{\beta}_{N,h}(\widehat{F}_N; \widehat{F}_N(x)) \quad \text{with} \quad \widehat{F}_N(x) = \frac{1}{N(T - t_0)} \sum_{i=1}^N \int_{t_0}^T \mathbf{1}_{X_i \leq x} dt.$$

A discrete-time approximate version of this estimator is computed on datasets generated by two different SDEs. In each case, the bandwidth of our estimator is selected via the PCO method introduced in Subsection 3. The first model is the so-called Langevin equation, and the second one is a nonlinear SDE with a multiplicative noise:

1. $X_t = x_0 - \int_0^t X_s ds + 0.1 \cdot W_t ; t \in [0, T].$
2. $X_t = x_0 - \int_0^t (X_s + \sin(4X_s)) ds + 0.1 \int_0^t (2 + \cos(X_s)) dW_s ; t \in [0, T].$

The models and the estimator are implemented by taking $N = 100$, $n = 50$, $T = 5$, $x_0 = 2$, $t_0 = 0$ and $K = \delta = \rho$, where ρ is the kernel introduced in Subsection 3. For Model 1, the estimator of the drift function is computed for the bandwidths set

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

and for Model 2, it is computed for the bandwidths set

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

For each of the previous models, on Figures 1 and 2 respectively, the true drift function (in red) and the PCO adaptive 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 Figure 1, one can see that the drift function is well estimated by the PCO adaptive estimator, with a MSE equal to $7.12 \cdot 10^{-4}$ compared to a MSE equal to $5.11 \cdot 10^{-4}$ for the oracle estimator. On Figure 2, one can see that the drift function of Model 2 is still well estimated by our PCO adaptive estimator. However, note that there is a degradation of the MSE, which is equal to $7.47 \cdot 10^{-3}$ for the adaptative estimator and equal to $4.99 \cdot 10^{-3}$ for the oracle estimator. This is

probably related to both the nonlinearity of the drift function and the multiplicative noise. In some cases, especially for Model 2, when the bandwidth is too small, the estimation degrades, but the PCO method selects a higher value of h which performs better. Finally, for each model, Table 1 gathers the

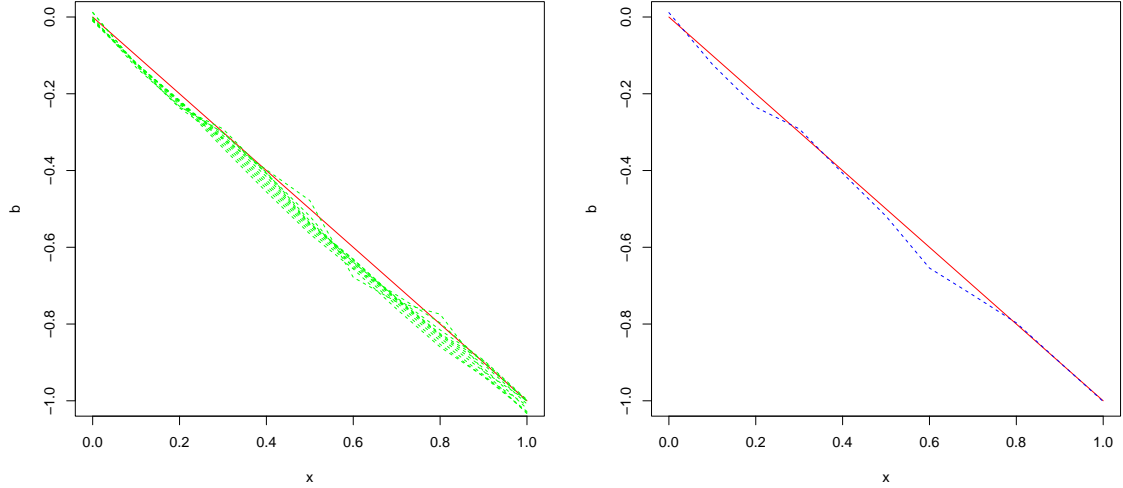


FIGURE 1. PCO adaptive estimator for Model 1 (Langevin equation), $\hat{h} = 0.04$ and $h_{\text{oracle}} = 0.02$.

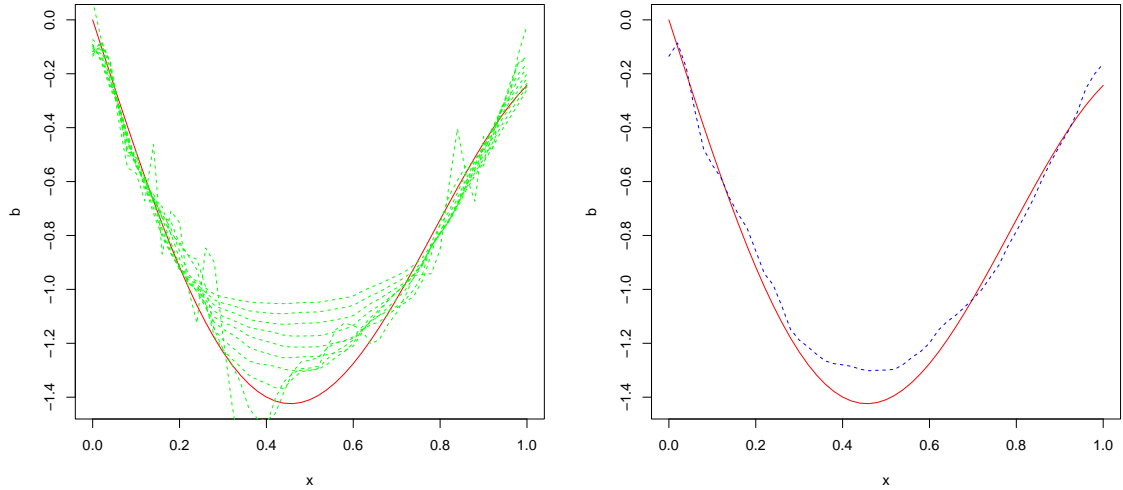


FIGURE 2. PCO adaptive estimator for Model 2, $\hat{h} = 0.05$ and $h_{\text{oracle}} = 0.07$.

mean MSE of 100 PCO estimations of the drift function as well as the mean MSE of the corresponding 100 oracle estimations. The mean MSEs are globally low, but higher for Model 2 with a nonlinear drift function than for Model 1 with a linear one. However, regardless of the complexity of the model, one can notice that for both models, the mean MSE of the PCO estimations remains relatively close to the mean MSE of the corresponding oracle estimations. This means that our PCO method performs well in practice.

	PCO	Oracle
Model 1	$8.30 \cdot 10^{-4}$	$6.28 \cdot 10^{-4}$
Model 2	$5.86 \cdot 10^{-3}$	$4.77 \cdot 10^{-3}$

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

5. PROOFS

5.1. **Proof of Proposition 2.4.** For the sake of readability, without loss of generality, Proposition 2.4 is proved for $t_0 = 0$. First of all,

$$\mathbb{E}(\|\widehat{b}_{N,h} - b\|_{f,A,B}^2) \leq \int_A^B \mathfrak{b}(x)^2 f(x) dx + \int_{-\infty}^{\infty} \mathfrak{v}(x) f(x) dx$$

where $\mathfrak{b}(x)$ (resp. $\mathfrak{v}(x)$) is the bias (resp. the variance) term of $\widehat{b}_{N,h}(x)$ for every $x \in [A, B]$ (resp. $x \in \mathbb{R}$). On the one hand, let us find a suitable bound on the f -weighted integrated variance of our warped kernel estimator. For any $x \in \mathbb{R}$, since X^1, \dots, X^N are independent copies of X ,

$$\begin{aligned} \mathfrak{v}(x) &= \text{var} \left(\frac{1}{NT} \sum_{i=1}^N \int_0^T K_h(F(X_t^i) - F(x)) dX_t^i \right) \leq \frac{1}{NT^2} \mathbb{E} \left[\left(\int_0^T K_h(F(X_t) - F(x)) dX_t \right)^2 \right] \\ &\leq \frac{2}{N} \mathbb{E} \left[\left(\int_0^T K_h(F(X_t) - F(x)) b(X_t) \frac{dt}{T} \right)^2 + \frac{1}{T^2} \left(\int_0^T K_h(F(X_t) - F(x)) \sigma(X_t) dW_t \right)^2 \right]. \end{aligned}$$

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

$$\begin{aligned} \mathfrak{v}(x) &\leq \frac{2}{NT} \int_0^T \mathbb{E}[K_h(F(X_t) - F(x))^2 b(X_t)^2] dt + \frac{2}{NT^2} \int_0^T \mathbb{E}[K_h(F(X_t) - F(x))^2 \sigma(X_t)^2] dt \\ &= \frac{2}{N} \int_{-\infty}^{\infty} K_h(F(z) - F(x))^2 b(z)^2 f(z) dz + \frac{2}{NT} \int_{-\infty}^{\infty} K_h(F(z) - F(x))^2 \sigma(z)^2 f(z) dz. \end{aligned}$$

Since K is symmetric, $K \in \mathbb{L}^2(\mathbb{R}, dx)$, $b, \sigma \in \mathbb{L}^2(\mathbb{R}, f(x) dx)$ and F is one-to-one from \mathbb{R} into $(0, 1)$ (see Assumption 2.3 and Remark 2.2),

$$\begin{aligned} \int_{-\infty}^{\infty} \mathfrak{v}(x) f(x) dx &\leq \frac{2}{N} \int_{\mathbb{R}^2} K_h(F(z) - F(x))^2 \left[b(z)^2 + \frac{1}{T} \sigma(z)^2 \right] f(z) f(x) dx dz \\ &= \frac{2}{N} \int_{-\infty}^{\infty} \left[b(z)^2 + \frac{1}{T} \sigma(z)^2 \right] f(z) \\ &\quad \times \int_{F(\mathbb{R})} K_h(F(z) - y)^2 f(F^{-1}(y)) \frac{dy}{F'(F^{-1}(y))} dz \\ &= \frac{2}{Nh} \int_{-\infty}^{\infty} \left[b(z)^2 + \frac{1}{T} \sigma(z)^2 \right] f(z) \int_{-F(z)/h}^{(1-F(z))/h} K(y)^2 dy dz \\ &\leq \frac{2\|K\|_2^2}{Nh} \left(\|b\|_f^2 + \frac{1}{T} \|\sigma\|_f^2 \right). \end{aligned}$$

On the other hand, let us find a suitable bound on the $[A, B]$ - f -weighted integrated squared-bias of $\widehat{b}_{N,h}$. For every $x \in [A, B]$, since X^1, \dots, X^N are independent copies of X , and since Itô's integral restricted to

\mathbb{H}^2 is a martingale-valued map,

$$\begin{aligned}
\mathfrak{b}(x) + b(x) &= \mathbb{E} \left[\frac{1}{NT} \sum_{i=1}^N \int_0^T K_h(F(X_t^i) - F(x)) dX_t^i \right] = \frac{1}{T} \mathbb{E} \left(\int_0^T K_h(F(X_t) - F(x)) dX_t \right) \\
&= \frac{1}{T} \left[\mathbb{E} \left(\int_0^T K_h(F(X_t) - F(x)) b(X_t) dt \right) + \mathbb{E} \left(\int_0^T K_h(F(X_t) - F(x)) \sigma(X_t) dW_t \right) \right] \\
&= \frac{1}{T} \int_0^T \mathbb{E}(K_h(F(X_t) - F(x)) b(X_t)) dt \\
&= \int_{-\infty}^{\infty} K_h(F(z) - F(x)) b(z) f(z) dz \\
&= \int_0^1 K_h(y - F(x)) b(F^{-1}(y)) dy = [K_h * ((b \circ F^{-1}) \mathbf{1}_{(0,1)})](F(x)).
\end{aligned}$$

Moreover, since $b \in \mathbb{L}^2(\mathbb{R}, f(x)dx)$ by Remark 2.2,

$$\begin{aligned}
\|[K_h * ((b \circ F^{-1}) \mathbf{1}_{(0,1)})](F(\cdot))\|_{f,A,B}^2 &\leq \int_{-\infty}^{\infty} [K_h * ((b \circ F^{-1}) \mathbf{1}_{(0,1)})](F(x))^2 f(x) dx \\
&= \int_0^1 [K_h * ((b \circ F^{-1}) \mathbf{1}_{(0,1)})](y)^2 dy \\
&\leq \|K\|_1^2 \int_0^1 b(F^{-1}(y))^2 dy = \|K\|_1^2 \|b\|_f^2 < \infty.
\end{aligned}$$

Therefore,

$$\int_A^B \mathfrak{b}(x)^2 f(x) dx = \|[K_h * ((b \circ F^{-1}) \mathbf{1}_{(0,1)})](F(\cdot)) - b\|_{f,A,B}^2.$$

This concludes the proof.

5.2. Proof of Theorem 3.2. The proof of Theorem 3.2 relies on the following technical results.

Proposition 5.1. *Under Assumptions 2.1, 2.3 and 3.1,*

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

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

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

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

(1) *There exists a constant $\mathfrak{c}_{5.1,1} > 0$ such that, for every $h \in (0, \Delta_0]$ and $\varphi \in C^0([t_0, T]; \mathbb{R})$,*

$$\|\Phi_h(\varphi, \cdot)\|_{\delta}^2 \leq \frac{\mathfrak{c}_{5.1,1}}{h^3}.$$

(2) *There exists a constant $\mathfrak{c}_{5.1,2} > 0$ such that, for every $h, h' \in (0, \Delta_0]$,*

$$\mathbb{E}(\langle \Phi_h(X^1, \cdot), \Phi_{h'}(X^2, \cdot) \rangle_{\delta}^2) \leq \mathfrak{c}_{5.1,2} \mathfrak{m}(h')$$

with

$$\mathfrak{m}(h') = \mathbb{E}(\|\Phi_{h'}(X, \cdot)\|_{\delta}^2).$$

(3) *There exists a constant $\mathfrak{c}_{5.1,3} > 0$ such that, for every $h \in (0, \Delta_0]$ and $\varphi \in \mathbb{L}^2(\mathbb{R}, dx)$,*

$$\mathbb{E}(\langle \Phi_h(X, \cdot), \varphi \rangle_{\delta}^2) \leq \mathfrak{c}_{5.1,3} \|\varphi\|_{\delta}^2.$$

(4) *If $t_0 > 0$, then there exists a deterministic constant $\mathfrak{c}_{5.1,4} > 0$ such that, for every $h, h' \in \mathcal{H}_N$,*

$$|\langle \Phi_h(X, \cdot), b_{h'} \rangle_{\delta}| \leq \mathfrak{c}_{5.1,4} \quad \text{a.s.}$$

Lemma 5.2. *Consider*

$$(3) \quad U_{h,h'}(N) := \sum_{i \neq j} \langle \Phi_h(X^i, \cdot) - b_h, \Phi_{h'}(X^j, \cdot) - b_{h'} \rangle_\delta ; \forall h, h' \in \mathcal{H}_N.$$

Under Assumptions 2.1, 2.3 and 3.1, there exists a deterministic constant $\mathfrak{c}_{5.2} > 0$, not depending on N , such that for every $\theta \in (0, 1)$ and $\lambda > 0$, with probability larger than $1 - 5.4|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h \in \mathcal{H}_N} \left\{ \frac{|U_{h,h_0}(N)|}{N^2} - \frac{\theta \mathfrak{m}(h)}{N} \right\} \leq \frac{\mathfrak{c}_{5.2}(1+\lambda)^3}{\theta N}$$

$$\text{and } \sup_{h \in \mathcal{H}_N} \left\{ \frac{|U_{h,h}(N)|}{N^2} - \frac{\theta \mathfrak{m}(h)}{N} \right\} \leq \frac{\mathfrak{c}_{5.2}(1+\lambda)^3}{\theta N}.$$

Lemma 5.3. *Consider*

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

Under Assumptions 2.1, 2.3 and 3.1, there exists a deterministic constant $\mathfrak{c}_{5.3} > 0$, not depending on N , such that for every $\theta \in (0, 1)$ and $\lambda > 0$, with probability larger than $1 - 2|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h \in \mathcal{H}_N} \left\{ \frac{1}{N} |V_h(N) - \mathfrak{m}(h)| - \frac{\theta \mathfrak{m}(h)}{N} \right\} \leq \frac{\mathfrak{c}_{5.3}(1+\lambda)}{\theta N}.$$

Lemma 5.4. *Consider*

$$(4) \quad W_{h,h'}(N) := \langle \widehat{b}_{N,h} - b_h, b_{h'} - b \rangle_\delta ; \forall h, h' \in \mathcal{H}_N.$$

Under Assumptions 2.1, 2.3 and 3.1, there exists a deterministic constant $\mathfrak{c}_{5.4} > 0$, not depending on N , such that for every $\theta \in (0, 1)$ and $\lambda > 0$, with probability larger than $1 - 2|\mathcal{H}_N|e^{-\lambda}$,

$$\sup_{h \in \mathcal{H}_N} \{ |W_{h,h_0}(N)| - \theta \|b_{h_0} - b\|_\delta^2 \} \leq \frac{\mathfrak{c}_{5.4}(1+\lambda)^2}{\theta N},$$

$$\sup_{h \in \mathcal{H}_N} \{ |W_{h_0,h}(N)| - \theta \|b_h - b\|_\delta^2 \} \leq \frac{\mathfrak{c}_{5.4}(1+\lambda)^2}{\theta N}$$

$$\text{and } \sup_{h \in \mathcal{H}_N} \{ |W_{h,h}(N)| - \theta \|b_h - b\|_\delta^2 \} \leq \frac{\mathfrak{c}_{5.4}(1+\lambda)^2}{\theta N}.$$

The proof of Proposition 5.1, which is specific to our warped kernel estimator of b , is postponed to Subsubsection 5.2.2. By using our Proposition 5.1 instead of Marie and Rosier [27], Lemma 3, the proofs of Lemmas 5.2, 5.3 and 5.4 remain the same as those of Marie and Rosier [27], Lemmas 4, 5 and 6 respectively.

5.2.1. *Steps of the proof.* The proof of Theorem 3.2.(1) is dissected in four steps.

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

$$\|\widehat{b}_{N,\widehat{h}} - b\|_\delta^2 = \|\widehat{b}_{N,\widehat{h}} - \widehat{b}_{N,h_0}\|_\delta^2 + \|\widehat{b}_{N,h_0} - b\|_\delta^2 - 2\langle \widehat{b}_{N,h_0} - \widehat{b}_{N,\widehat{h}}, \widehat{b}_{N,h_0} - b \rangle_\delta.$$

Then, by (2), for any $h \in \mathcal{H}_N$,

$$\begin{aligned} \|\widehat{b}_{N,\widehat{h}} - b\|_\delta^2 &\leq \|\widehat{b}_{N,h} - \widehat{b}_{N,h_0}\|_\delta^2 + \text{pen}(h) - \text{pen}(\widehat{h}) \\ &\quad + \|\widehat{b}_{N,h_0} - b\|_\delta^2 - 2\langle \widehat{b}_{N,h_0} - \widehat{b}_{N,\widehat{h}}, \widehat{b}_{N,h_0} - b \rangle_\delta \\ &\leq \|\widehat{b}_{N,h} - b\|_\delta^2 + \text{pen}(h) - \text{pen}(\widehat{h}) - 2\langle \widehat{b}_{N,h} - \widehat{b}_{N,\widehat{h}}, \widehat{b}_{N,h_0} - b \rangle_\delta \\ (5) \quad &= \|\widehat{b}_{N,h} - b\|_\delta^2 - \psi_N(h) + \psi_N(\widehat{h}) \end{aligned}$$

where

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

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

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

where

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

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

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

$$|\psi_{1,N}(h)| \leq \frac{\theta \mathbf{m}(h)}{N} + \frac{\mathbf{c}_{5.2}(1+\lambda)^3}{\theta N} \quad \text{and} \quad |\psi_{1,N}(\widehat{h})| \leq \frac{\theta \mathbf{m}(\widehat{h})}{N} + \frac{\mathbf{c}_{5.2}(1+\lambda)^3}{\theta N}.$$

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

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

By Proposition 5.1.(4),

$$|\Psi_{2,N}(h, h')| \leq \frac{1}{N} \sum_{i=1}^N \left| \int_{-\infty}^{\infty} \Phi_h(X^i, x) b_{h'}(x) \delta(x) dx \right| \leq \mathbf{c}_{5.1,4} \quad \text{a.s.}$$

On the other hand, since

$$\begin{aligned} \|b_h \mathbf{1}_{[-\Delta, \Delta]}\|_\infty & \leq \|K\|_1 \|b\|_{\infty, I_{\Delta, F}} \\ \text{and} \quad \|b_{h_0} \delta\|_1 & \leq \|b_{h_0} \mathbf{1}_{[-\Delta, \Delta]}\|_\infty \|\delta\|_1 \leq \|K\|_1 \|b\|_{\infty, I_{\Delta, F}} \end{aligned}$$

by Inequality (11),

$$|\langle b_h, b_{h_0} \rangle_\delta| \leq \|b_h \mathbf{1}_{[-\Delta, \Delta]}\|_\infty \|b_{h_0} \delta\|_1 \leq \|K\|_1^2 \|b\|_{\infty, I_{\Delta, F}}^2.$$

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

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

- By Lemma 5.4 and Cauchy-Schwarz's inequality, with probability larger than $1 - |\mathcal{H}_N|e^{-\lambda}$,

$$\begin{aligned} |\psi_{3,N}(h)| & \leq \frac{\theta}{4} (\|b_h - b\|_\delta^2 + \|b_{h_0} - b\|_\delta^2) + \frac{8\mathbf{c}_{5.4}(1+\lambda)^2}{\theta N} \\ & \quad + 2 \times \frac{1}{2^{1/2}} \left(\frac{\theta}{2}\right)^{1/2} \|b_h - b\|_\delta \times \frac{1}{2^{1/2}} \left(\frac{2}{\theta}\right)^{1/2} \|b_{h_0} - b\|_\delta \\ & \leq \frac{\theta}{2} \|b_h - b\|_\delta^2 + \left(\frac{\theta}{4} + \frac{1}{\theta}\right) \|b_{h_0} - b\|_\delta^2 + \frac{8\mathbf{c}_{5.4}(1+\lambda)^2}{\theta N} \end{aligned}$$

and

$$|\psi_{3,N}(\widehat{h})| \leq \frac{\theta}{2} \|b_{\widehat{h}} - b\|_\delta^2 + \left(\frac{\theta}{4} + \frac{1}{\theta}\right) \|b_{h_0} - b\|_\delta^2 + \frac{8\mathbf{c}_{5.4}(1+\lambda)^2}{\theta N}.$$

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

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

and

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

On the one hand, note that

$$\|\widehat{b}_{N,h} - b\|_\delta^2 - (1 + \theta) \left(\|b_h - b\|_\delta^2 + \frac{\mathbf{m}(h)}{N} \right)$$

can be written

$$\|\widehat{b}_{N,h} - b_h\|_\delta^2 - \frac{(1 + \theta)\mathbf{m}(h)}{N} + 2W_h(N) - \theta \|b_h - b\|_\delta^2,$$

where $W_h(N) := W_{h,h}(N)$ (see (4)). Moreover, for any $h \in \mathcal{H}_N$,

$$(6) \quad \|\widehat{b}_{N,h} - b_h\|_\delta^2 = \frac{U_h(N)}{N^2} + \frac{V_h(N)}{N}$$

with $U_h(N) = U_{h,h}(N)$ (see (3)). So, with probability larger than $1 - \bar{\mathbf{c}}_2 |\mathcal{H}_N| e^{-\lambda}$,

$$\sup_{h \in \mathcal{H}_N} \left\{ \|\widehat{b}_{N,h} - b_h\|_\delta^2 - \frac{(1 + \theta)\mathbf{m}(h)}{N} \right\} \leq \frac{2(\mathbf{c}_{5.2} + \mathbf{c}_{5.3})(1 + \lambda)^3}{\theta N}$$

by Lemmas 5.2 and 5.3, and then

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

by Lemma 5.4. On the other hand, for any $h \in \mathcal{H}_N$,

$$\|b_h - b\|_\delta^2 = \|\widehat{b}_{N,h} - b\|_\delta^2 - \|\widehat{b}_{N,h} - b_h\|_\delta^2 + 2W_h(N).$$

Then,

$$(1 - \theta) \left(\|b_h - b\|_\delta^2 + \frac{\mathbf{m}(h)}{N} \right) - \|\widehat{b}_{N,h} - b\|_\delta^2 \leq 2|W_h(N)| - \theta \|b_h - b\|_\delta^2 + \Lambda_h(N) - \frac{\theta \mathbf{m}(h)}{N}$$

where

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

By Equality (6),

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

By Lemmas 5.3 and 5.2, there exist two deterministic constants $\mathbf{c}_3, \bar{\mathbf{c}}_3 > 0$, not depending N and θ , such that with probability larger than $1 - \bar{\mathbf{c}}_3 |\mathcal{H}_N| e^{-\lambda}$,

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

By Lemma 5.4, with probability larger than $1 - 2|\mathcal{H}_N| e^{-\lambda}$,

$$\sup_{h \in \mathcal{H}_N} \{2|W_h(N)| - \theta \|b_h - b\|_\delta^2\} \leq \frac{4\mathbf{c}_{5.4}(1 + \lambda)^2}{\theta N}.$$

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

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

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

$$|\psi_N(h)| \leq \theta \left(\|b_h - b\|_\delta^2 + \frac{\mathfrak{m}(h)}{N} \right) + \left(\frac{\theta}{2} + \frac{2}{\theta} \right) \|b_{h_0} - b\|_\delta^2 + \frac{\mathfrak{c}_4(1+\lambda)^3}{\theta N}$$

and

$$|\psi_N(\hat{h})| \leq \theta \left(\|b_{\hat{h}} - b\|_\delta^2 + \frac{\mathfrak{m}(\hat{h})}{N} \right) + \left(\frac{\theta}{2} + \frac{2}{\theta} \right) \|b_{h_0} - b\|_\delta^2 + \frac{\mathfrak{c}_4(1+\lambda)^3}{\theta N}.$$

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

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

and

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

By the decomposition (5), there exist two deterministic constants $\mathfrak{c}_6, \bar{\mathfrak{c}}_6 > 0$, not depending on N, θ, h and h_0 , such that with probability larger than $1 - \bar{\mathfrak{c}}_6 |\mathcal{H}_N| e^{-\lambda}$,

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

This concludes the proof.

5.2.2. *Proof of Proposition 5.1.* For the sake of readability, Proposition 5.1.(1,2,3) are proved for $t_0 = 0$ without loss of generality. However, the condition $t_0 > 0$ is required in order to prove Proposition 5.1.(4) because f' is involved. First of all, for any $x \in \mathbb{R}$ and $h > 0$, by Itô's formula,

$$\int_{x_0}^{X_T} K_h(F(z) - F(x)) dz = \int_0^T K_h(F(X_t) - F(x)) dX_t + \frac{1}{2h^2} \int_0^T K' \left(\frac{F(X_t) - F(x)}{h} \right) f(X_t) d\langle X \rangle_t.$$

So,

$$\begin{aligned} \int_0^T K_h(F(X_t) - F(x)) dX_t &= \int_{x_0}^{X_T} K_h(F(z) - F(x)) dz \\ &\quad - \frac{1}{2h^2} \int_0^T K' \left(\frac{F(X_t) - F(x)}{h} \right) \sigma(X_t)^2 f(X_t) dt = T\Phi_h(X, x). \end{aligned}$$

(1) Consider $\varphi \in C^0([0, T]; \mathbb{R})$. For any $x \in [-\Delta, \Delta]$ and $h \in (0, \Delta_0]$,

$$(7) \quad [-1, 1] \subset \left[-\frac{F(-\Delta)}{h}, \frac{1-F(\Delta)}{h} \right] \subset \left[-\frac{F(x)}{h}, \frac{1-F(x)}{h} \right]$$

and, since $\text{supp}(K) = [-1, 1]$ and $\text{supp}(\delta) = [-\Delta, \Delta]$,

$$\begin{aligned} \int_{-\infty}^{\infty} \left[\int_{\varphi(0)}^{\varphi(T)} K_h(F(z) - F(x)) dz \right]^2 \delta(x) dx &\leq \int_{-\infty}^{\infty} \left[\int_{-F(x)/h}^{(1-F(x))/h} |K(y)| \frac{dy}{f(F^{-1}(hy + F(x)))} \right]^2 \delta(x) dx \\ &= \int_{-\Delta}^{\Delta} \left(\int_{-1}^1 |K(y)| \frac{dy}{f(F^{-1}(hy + F(x)))} \right)^2 \delta(x) dx. \end{aligned}$$

Moreover, since $h \leq \Delta_0$ and F^{-1} is increasing,

$$(8) \quad F^{-1}(h \cdot + F(x))([-1, 1]) \subset [F^{-1}((1-\kappa)F(-\Delta)), F^{-1}(\kappa + (1-\kappa)F(\Delta))] =: I_{\Delta, F}.$$

So, there exists $f_1 > 0$ such that $f(z) \geq f_1$ for every $z \in I_{\Delta, F}$ because $f(\mathbb{R}) \subset (0, \infty)$, and then

$$\begin{aligned} & \int_{-\infty}^{\infty} \left[\int_{\varphi(0)}^{\varphi(T)} K_h(F(z) - F(x)) dz \right]^2 \delta(x) dx \\ & \leq \frac{1}{f_1^2} \int_{-\Delta}^{\Delta} \left(\int_{-1}^1 |K(z)| dz \right)^2 \delta(x) dx \leq \frac{\|K\|_1^2}{f_1^2}. \end{aligned}$$

The same way, there exists $f_2 > 0$ such that $f(z) \geq f_2$ for every $z \in [-\Delta, \Delta]$, and then

$$\begin{aligned} \int_{-\infty}^{\infty} K' \left(\frac{F(\varphi(t)) - F(x)}{h} \right)^2 \delta(x) dx &= \int_{-\Delta}^{\Delta} K' \left(\frac{F(\varphi(t)) - F(x)}{h} \right)^2 \delta(x) dx \\ &= h \int_{[F(-\Delta) - F(\varphi(t))]/h}^{[F(\Delta) - F(\varphi(t))]/h} K'(y)^2 \frac{\delta(F^{-1}(hy + F(\varphi(t))))}{f(F^{-1}(hy + F(\varphi(t))))} dy \\ &\leq \frac{\|K'\|_2^2 \|\delta\|_{\infty} h}{f_2} \end{aligned}$$

for every $t \in [0, T]$. Thus,

$$\begin{aligned} T^2 \|\Phi_h(\varphi, \cdot)\|_{\delta}^2 &\leq 2 \int_{-\infty}^{\infty} \left[\int_{\varphi(0)}^{\varphi(T)} K_h(F(z) - F(x)) dz \right]^2 \delta(x) dx \\ &\quad + \frac{1}{2h^4} \int_{-\infty}^{\infty} \left[\int_0^T K' \left(\frac{F(\varphi(t)) - F(x)}{h} \right) \sigma(\varphi(t))^2 f(\varphi(t)) dt \right]^2 \delta(x) dx \\ &\leq \frac{2\|K\|_1^2}{f_1^2} + \frac{T}{2h^4} \int_0^T \sigma(\varphi(t))^4 f(\varphi(t))^2 \int_{-\infty}^{\infty} K' \left(\frac{F(\varphi(t)) - F(x)}{h} \right)^2 \delta(x) dx dt \\ &\leq \frac{2\|K\|_1^2}{f_1^2} + \frac{T^2 \|\sigma\|_{\infty}^4 \|f\|_{\infty}^2 \|K'\|_2^2 \|\delta\|_{\infty}}{2f_2 h^3}. \end{aligned}$$

(2) For any $h, h' \in (0, \Delta_0]$,

$$\begin{aligned} & \mathbb{E}(\langle \Phi_h(X^1, \cdot), \Phi_{h'}(X^2, \cdot) \rangle_{\delta}^2) \\ &= \frac{1}{T^4} \mathbb{E} \left[\left(\int_{-\infty}^{\infty} \left(\int_0^T K_h(F(X_t^1) - F(x)) dX_t^1 \right) \left(\int_0^T K_{h'}(F(X_t^2) - F(x)) dX_t^2 \right) \delta(x) dx \right)^2 \right] \\ &\leq \frac{2}{T^4} (\mathbb{E}(A_{h, h'}^2) + \mathbb{E}(B_{h, h'}^2)) \end{aligned}$$

with

$$A_{h, h'} := \int_{-\infty}^{\infty} \left(\int_0^T K_h(F(X_t^1) - F(x)) \sigma(X_t^1) dW_t^1 \right) \left(\int_0^T K_{h'}(F(X_t^2) - F(x)) dX_t^2 \right) \delta(x) dx$$

and

$$B_{h, h'} := \int_{-\infty}^{\infty} \left(\int_0^T K_h(F(X_t^1) - F(x)) b(X_t^1) dt \right) \left(\int_0^T K_{h'}(F(X_t^2) - F(x)) dX_t^2 \right) \delta(x) dx.$$

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

$$\begin{aligned} \mathbb{E}(A_{h, h'}^2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{E} \left[\left(\int_0^T K_h(F(X_t^1) - F(x)) \sigma(X_t^1) dW_t^1 \right) \left(\int_0^T K_h(F(X_t^1) - F(y)) \sigma(X_t^1) dW_t^1 \right) \right] \\ &\quad \times \mathbb{E} \left[\left(\int_0^T K_{h'}(F(X_t^2) - F(x)) dX_t^2 \right) \left(\int_0^T K_{h'}(F(X_t^2) - F(y)) dX_t^2 \right) \right] \delta(x) \delta(y) dx dy. \end{aligned}$$

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

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^T K_h(F(X_t^1) - F(x))\sigma(X_t^1)dW_t^1 \right) \left(\int_0^T K_h(F(X_t^1) - F(y))\sigma(X_t^1)dW_t^1 \right) \right] \\ = \int_0^T \mathbb{E}(K_h(F(X_t^1) - F(x))K_h(F(X_t^1) - F(y))\sigma(X_t^1)^2)dt \\ = T \int_{-\infty}^{\infty} K_h(F(z) - F(x))K_h(F(z) - F(y))\sigma(z)^2 f(z)dz. \end{aligned}$$

Then,

$$\begin{aligned} \mathbb{E}(A_{h,h'}^2) &= T \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_h(F(z) - F(x))K_h(F(z) - F(y))\sigma(z)^2 f(z) \\ &\quad \times \mathbb{E} \left[\left(\int_0^T K_{h'}(F(X_t^2) - F(x))dX_t^2 \right) \left(\int_0^T K_{h'}(F(X_t^2) - F(y))dX_t^2 \right) \right] \delta(x)\delta(y)dxdydz \\ &= T \int_{-\infty}^{\infty} \sigma(z)^2 f(z) \mathbb{E} \left[\left(\int_{-\Delta}^{\Delta} K_h(F(z) - F(x))\delta(x) \int_0^T K_{h'}(F(X_t^2) - F(x))dX_t^2 dx \right)^2 \right] dz. \end{aligned}$$

On the other hand, note that for any $z \in \mathbb{R}$,

$$\begin{aligned} \|K_h(F(z) - F(\cdot))\mathbf{1}_{[-\Delta, \Delta]}(\cdot)\|_1 &= \int_{-\Delta}^{\Delta} |K_h(F(z) - F(y))|dy \\ (9) \qquad \qquad \qquad &= \int_{(F(-\Delta)-F(z))/h}^{(F(\Delta)-F(z))/h} |K(y)| \frac{dy}{f(F^{-1}(hy + F(z)))} \leq \frac{\|K\|_1}{f_2} \end{aligned}$$

and, for every $x \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} f(z)|K_h(F(z) - F(x))|dz = \int_0^1 |K_h(z - F(x))|dz \leq \|K\|_1.$$

Then, since

$$x \in [-\Delta, \Delta] \mapsto \frac{|K_h(F(z) - F(x))|}{\|K_h(F(z) - F(\cdot))\mathbf{1}_{[-\Delta, \Delta]}(\cdot)\|_1}$$

is a density function, by Jensen's inequality,

$$\begin{aligned} \mathbb{E}(A_{h,h'}^2) &\leq T \int_{-\infty}^{\infty} \sigma(z)^2 f(z) \|K_h(F(z) - F(\cdot))\mathbf{1}_{[-\Delta, \Delta]}(\cdot)\|_1 \\ &\quad \times \int_{-\Delta}^{\Delta} |K_h(F(z) - F(x))|\delta(x)^2 \mathbb{E} \left[\left(\int_0^T K_{h'}(F(X_t^2) - F(x))dX_t^2 \right)^2 \right] dx dz \\ &\leq \frac{T\|\sigma\|_{\infty}^2 \|K\|_1^2 \|\delta\|_{\infty}}{f_2} \int_{-\Delta}^{\Delta} \delta(x) \mathbb{E} \left[\left(\int_0^T K_{h'}(F(X_t^2) - F(x))dX_t^2 \right)^2 \right] dx \\ &\leq \frac{T^3 \|\sigma\|_{\infty}^2 \|K\|_1^2 \|\delta\|_{\infty} \mathbf{m}(h')}{f_2}. \end{aligned}$$

Bound on $\mathbb{E}(B_{h,h'}^2)$. First, note that for every $x \in [-\Delta, \Delta]$, by (7) and (8),

$$\begin{aligned} \int_{-\infty}^{\infty} |K_h(F(z) - F(x))|b(z)^2 f(z)dz &= \int_{-F(x)/h}^{(1-F(x))/h} |K(z)|b(F^{-1}(hz + F(x)))^2 dz \\ (10) \qquad \qquad \qquad &= \int_{-1}^1 |K(z)|b(F^{-1}(hz + F(x)))^2 dz \leq \|b\|_{\infty, I_{\Delta, F}}^2 \|K\|_1. \end{aligned}$$

Then, since

$$x \in [-\Delta, \Delta] \mapsto \frac{|K_h(F(X_t(\omega)) - F(x))|}{\|K_h(F(X_t(\omega)) - F(\cdot))\mathbf{1}_{[-\Delta, \Delta]}(\cdot)\|_1}$$

is a density function for every $(t, \omega) \in [0, T] \times \Omega$, by Jensen's inequality and Inequality (9),

$$\begin{aligned} \mathbb{E}(B_{h, h'}^2) &= \mathbb{E} \left[\left(\int_0^T \int_{-\infty}^{\infty} K_h(F(X_t^1) - F(x))b(X_t^1)\delta(x) \int_0^T K_{h'}(F(X_s^2) - F(x))dX_s^2 dx dt \right)^2 \right] \\ &\leq \frac{T\|K\|_1}{f_2} \int_0^T \int_{-\infty}^{\infty} \mathbb{E}(|K_h(F(X_t^1) - F(x))|b(X_t^1)^2) \\ &\quad \times \delta(x)^2 \mathbb{E} \left[\left(\int_0^T K_{h'}(F(X_s^2) - F(x))dX_s^2 \right)^2 \right] dx dt \\ &= \frac{T^2\|K\|_1}{f_2} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |K_h(F(z) - F(x))|b(z)^2 f(z) dz \right) \\ &\quad \times \delta(x)^2 \mathbb{E} \left[\left(\int_0^T K_{h'}(F(X_s^2) - F(x))dX_s^2 \right)^2 \right] dx \\ &\leq \frac{T^2\|b\|_{\infty, I_{\Delta, F}}^2 \|K\|_1^2 \|\delta\|_{\infty}}{f_2} \int_{-\infty}^{\infty} \delta(x) \mathbb{E} \left[\left(\int_0^T K_{h'}(F(X_s^2) - F(x))dX_s^2 \right)^2 \right] dx \\ &\leq \frac{T^4\|b\|_{\infty, I_{\Delta, F}}^2 \|K\|_1^2 \|\delta\|_{\infty} \mathbf{m}(h')}{f_2}. \end{aligned}$$

(3) For any $h \in (0, \Delta_0]$ and $\varphi \in \mathbb{L}^2(\mathbb{R}, dx)$,

$$\begin{aligned} \mathbb{E}(\langle \Phi_h(X, \cdot), \varphi \rangle_{\delta}^2) &= \frac{1}{T^2} \mathbb{E} \left[\left(\int_{-\infty}^{\infty} \varphi(x)\delta(x) \int_0^T K_h(F(X_t) - F(x))dX_t dx \right)^2 \right] \\ &\leq \frac{2}{T^2} (\mathbb{E}(C_h^2) + \mathbb{E}(D_h^2)) \end{aligned}$$

with

$$C_h := \int_{-\infty}^{\infty} \varphi(x)\delta(x) \int_0^T K_h(F(X_t) - F(x))\sigma(X_t)dW_t dx$$

and

$$D_h := \int_{-\infty}^{\infty} \varphi(x)\delta(x) \int_0^T K_h(F(X_t) - F(x))b(X_t)dt dx.$$

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

$$\begin{aligned}
\mathbb{E}(C_h^2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x)\varphi(y)\delta(x)\delta(y) \int_0^T \mathbb{E}(K_h(F(X_t) - F(x))K_h(F(X_t) - F(y))\sigma(X_t)^2) dt dx dy \\
&= T \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x)\varphi(y)\delta(x)\delta(y) \\
&\quad \times \int_{-\infty}^{\infty} K_h(F(z) - F(x))K_h(F(z) - F(y))\sigma(z)^2 f(z) dz dx dy \\
&= T \int_{-\infty}^{\infty} \left(\int_{-\Delta}^{\Delta} K_h(F(z) - F(x))\varphi(x)\delta(x) dx \right)^2 \sigma(z)^2 f(z) dz \\
&\leq T \int_{-\infty}^{\infty} \|K_h(F(z) - F(\cdot))\mathbf{1}_{[-\Delta, \Delta]}(\cdot)\|_1 \\
&\quad \times \left(\int_{-\Delta}^{\Delta} |K_h(F(z) - F(x))|\varphi(x)^2 \delta(x)^2 dx \right) \sigma(z)^2 f(z) dz \\
&\leq \frac{T\|K\|_1\|\sigma\|_{\infty}^2}{f_2} \int_{-\Delta}^{\Delta} \left(\int_{-\infty}^{\infty} f(z)|K_h(F(z) - F(x))| dz \right) \varphi(x)^2 \delta(x)^2 dx \quad \text{by Inequality (9)} \\
&\leq \frac{T\|K\|_1^2\|\sigma\|_{\infty}^2\|\delta\|_{\infty}\|\varphi\|_{\delta}^2}{f_2}.
\end{aligned}$$

Bound on $\mathbb{E}(D_h^2)$. By Jensen's inequality, and Inequalities (9) and (10),

$$\begin{aligned}
\mathbb{E}(D_h^2) &\leq T \int_0^T \mathbb{E} \left[b(X_t)^2 \|K_h(F(X_t) - F(\cdot))\mathbf{1}_{[-\Delta, \Delta]}(\cdot)\|_1 \int_{-\Delta}^{\Delta} |K_h(F(X_t) - F(x))|\varphi(x)^2 \delta(x)^2 dx \right] dt \\
&\leq \frac{T^2\|K\|_1}{f_2} \int_{-\Delta}^{\Delta} \left(\int_{-\infty}^{\infty} |K_h(F(z) - F(x))|b(z)^2 f(z) dz \right) \varphi(x)^2 \delta(x)^2 dx \\
&\leq \frac{T^2\|b\|_{\infty, I_{\Delta, F}}^2\|K\|_1^2\|\delta\|_{\infty}\|\varphi\|_{\delta}^2}{f_2}.
\end{aligned}$$

(4) Since X is a semi-martingale, since the map

$$(t, \omega, x) \mapsto K_h(F(X_t(\omega)) - F(x))b_{h'}(x)\mathbf{1}_{[-\Delta, \Delta]}(x)$$

is measurable and bounded for any $h, h' \in \mathcal{H}_N$, and since

$$A \mapsto \int_A \delta(x) dx \quad \text{is a finite measure,}$$

by the stochastic Fubini theorem, the change of variable formula and Itô's formula,

$$\begin{aligned}
(T - t_0)\langle \Phi_h(X, \cdot), b_{h'} \rangle_{\delta} &= (T - t_0) \int_{-\infty}^{\infty} \Phi_h(X, x)b_{h'}(x)\delta(x) dx \\
&= \int_{t_0}^T \int_{-\infty}^{\infty} K_h(F(X_t) - F(x))b_{h'}(x)\delta(x) dx dX_t \quad \text{a.s.} \\
&= \int_{t_0}^T [K_h * [((b_{h'}\delta/f) \circ F^{-1})\mathbf{1}_{(0,1)}]](F(X_t)) dX_t \\
&= \int_{X_{t_0}}^{X_T} \psi_{h, h'}(z) dz - \frac{1}{2} \int_{t_0}^T \psi'_{h, h'}(X_t)\sigma(X_t)^2 dt
\end{aligned}$$

where

$$\psi_{h, h'}(\cdot) := [K_h * [((b_{h'}\delta/f) \circ F^{-1})\mathbf{1}_{(0,1)}]](F(\cdot)).$$

On the one hand,

$$\begin{aligned} (\psi'_{h,h'}/f) \circ F^{-1} &= K_h * [((b_{h'}\delta/f) \circ F^{-1})\mathbf{1}_{(0,1)}]' \\ &= K_h * [\mathbf{1}_{[F(-\Delta),F(\Delta)]}/(f \circ F^{-1}) \times (b_{h'}\delta'/f + b'_{h'}\delta/f - b_{h'}\delta f'/f^2) \circ F^{-1}]. \end{aligned}$$

Moreover, for every $x \in \mathbb{R}$, by (8),

$$(11) \quad |b_{h'}(x)\mathbf{1}_{[-\Delta,\Delta]}(x)| = \left| \mathbf{1}_{[-\Delta,\Delta]}(x) \int_{-\infty}^{\infty} K_{h'}(y - F(x))[(b \circ F^{-1})\mathbf{1}_{(0,1)}](y)dy \right| \\ \leq \|K\|_1 \sup_{(x,y) \in [-\Delta,\Delta] \times [-1,1]} |[(b \circ F^{-1})\mathbf{1}_{(0,1)}](h'y + F(x))| \leq \|K\|_1 \|b\|_{\infty, I_{\Delta, F}}$$

and

$$|b'_{h'}(x)\mathbf{1}_{[-\Delta,\Delta]}(x)| = \left| \mathbf{1}_{[-\Delta,\Delta]}(x) f(x) \int_{-\infty}^{\infty} K_{h'}(y - F(x))[(b \circ F^{-1})'\mathbf{1}_{(0,1)}](y)dy \right| \\ \leq \|K\|_1 \sup_{(x,y) \in [-\Delta,\Delta] \times [-1,1]} \left| \left[\frac{b' \circ F^{-1}}{f \circ F^{-1}} \mathbf{1}_{(0,1)} \right] (h'y + F(x)) \right| \leq \frac{\|K\|_1 \|b'\|_{\infty, I_{\Delta, F}}}{f_1}.$$

Then, since f and f' are bounded by Remark 2.2,

$$\begin{aligned} \|\psi'_{h,h'}\|_{\infty} &\leq \frac{\|f\|_{\infty} \|K_h\|_1}{f_2} \left(\frac{\|\delta'\|_{\infty} \|b_{h'}\mathbf{1}_{[-\Delta,\Delta]}\|_{\infty}}{f_2} + \frac{\|\delta\|_{\infty} \|b'_{h'}\mathbf{1}_{[-\Delta,\Delta]}\|_{\infty}}{f_2} + \frac{\|\delta\|_{\infty} \|f'\|_{\infty} \|b_{h'}\mathbf{1}_{[-\Delta,\Delta]}\|_{\infty}}{f_2^2} \right) \\ &\leq \frac{\|f\|_{\infty} \|K\|_1^2}{f_2} \left(\frac{\|\delta'\|_{\infty} \|b\|_{\infty, I_{\Delta, F}}}{f_2} + \frac{\|\delta\|_{\infty} \|b'\|_{\infty, I_{\Delta, F}}}{f_1 f_2} + \frac{\|\delta\|_{\infty} \|f'\|_{\infty} \|b\|_{\infty, I_{\Delta, F}}}{f_2^2} \right) < \infty. \end{aligned}$$

On the other hand, by Inequality (11), by (7) and by (8),

$$\begin{aligned} \int_{-\infty}^{\infty} |\psi_{h,h'}(x)|dx &= \int_{-\infty}^{\infty} \left| \int_0^1 K_h(F(x) - y) \left(\frac{b_{h'}\delta}{f} \right) (F^{-1}(y))dy \right| dx \\ &= \int_0^1 \left| \int_{-\infty}^{\infty} K_h(x - F(y)) b_{h'}(y) \delta(y) dy \right| \frac{dx}{f(F^{-1}(x))} \\ &\leq \|K\|_1 \|b\|_{\infty, I_{\Delta, F}} \int_0^1 \int_{-\Delta}^{\Delta} |K_h(x - F(y))| \delta(y) \frac{dy dx}{f(F^{-1}(x))} \\ &= \|K\|_1 \|b\|_{\infty, I_{\Delta, F}} \int_{-\Delta}^{\Delta} \delta(y) \int_{-F(y)/h}^{(1-F(y))/h} |K(x)| \frac{dx}{f(F^{-1}(hx + F(y)))} dy \\ &\leq \frac{\|K\|_1 \|b\|_{\infty, I_{\Delta, F}}}{f_1} \int_{-\Delta}^{\Delta} \delta(y) \int_{-1}^1 |K(x)| dx dy = \frac{\|K\|_1^2 \|b\|_{\infty, I_{\Delta, F}}}{f_1}. \end{aligned}$$

This concludes the proof because

$$(T - t_0) |\langle \Phi_h(X, \cdot), b_{h'} \rangle_{\delta}| \leq \int_{-\infty}^{\infty} |\psi_{h,h'}(x)| dx + \frac{T - t_0}{2} \|\sigma\|_{\infty}^2 \|\psi'_{h,h'}\|_{\infty} \quad \text{a.s.}$$

REFERENCES

- [1] G. Chagny. *Adaptive Warped Kernel Estimators*. Scandinavian Journal of Statistics 42, 2, 336-360, 2015.
- [2] A. Cohen, M.A. Leviatan and D. Leviatan. *On the Stability and Accuracy of Least Squares Approximations*. Foundations of Computational Mathematics 13, 819-834, 2013.
- [3] F. Comte. *Estimation non-paramétrique. 2e édition*. Spartacus IDH, 2017.
- [4] 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.
- [5] 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.
- [6] F. Comte and V. Genon-Catalot. *Drift Estimation on Non Compact Support for Diffusion Models*. Stochastic Processes and Their Applications 134, 174-207, 2021.
- [7] 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.

- [8] F. Comte, V. Genon-Catalot and A. Samson. *Nonparametric Estimation for Stochastic Differential Equations with Random Effects*. Stochastic Processes and Their Applications 123, 2522-2551, 2013.
- [9] F. Comte and N. Marie. *On a Nadaraya-Watson Estimator with Two Bandwidths*. Electronic Journal of Statistics 15, 1, 2566-2607, 2021.
- [10] F. Comte and N. Marie. *Nonparametric Drift Estimation from Diffusions with Correlated Brownian Motions*. Journal of Multivariate Analysis 198, 23 pages, 2023.
- [11] A. Dalalyan. *Sharp Adaptive Estimation of the Trend Coefficient for Ergodic Diffusion*. The Annals of Statistics 33, 6, 2507-2528, 2005.
- [12] M. Delattre, V. Genon-Catalot and C. Larédo. *Parametric Inference for Discrete Observations of Diffusion Processes with Mixed Effects*. Stochastic Processes and Their Applications 128, 1929-1957, 2018.
- [13] M. Delattre, V. Genon-Catalot and A. Samson. *Maximum Likelihood Estimation for Stochastic Differential Equations with Random Effects*. Scandinavian Journal of Statistics 40, 322-343, 2013.
- [14] M. Delattre and M. Lavielle. *Coupling the SAEM Algorithm and the Extended Kalman Filter for Maximum Likelihood Estimation in Mixed-Effects Diffusion Models*. Statistics and Its Interface 6, 519-532, 2013.
- [15] L. Della Maestra and M. Hoffmann. *Nonparametric Estimation for Interacting Particle Systems: McKean-Vlasov Models*. Probability Theory and Related Fields 182, 551-613, 2022.
- [16] 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*. Bernoulli 27, 4, 2675-2713, 2021.
- [17] C. Dion and V. Genon-Catalot. *Bidimensional Random Effect Estimation in Mixed Stochastic Differential Model*. Statistical Inference Stochastic Processes 19, 131-158, 2016.
- [18] S. Ditlevsen and A. De Gaetano. *Mixed Effects in Stochastic Differential Equation Models*. REVSTAT 3, 137-153, 2005.
- [19] E. Giné and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, 2015.
- [20] 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.
- [21] H. Halconruy and N. Marie. *Kernel Selection in Nonparametric Regression*. Mathematical Methods of Statistics 29, 1, 32-56, 2020.
- [22] M. Hoffmann. *Adaptive Estimation in Diffusion Processes*. Stochastic Processes and Their Applications 79, 135-163, 1999.
- [23] C. Houdré and P. Reynaud-Bouret. *Exponential Inequalities, with Constants, for U-statistics of Order Two*. Stochastic Inequalities and Applications, vol. 56 of Progress in Probability, 55-69, Birkhauser, 2003.
- [24] M. Kessler. *Simple and Explicit Estimating Functions for a Discretely Observed Diffusion Process*. Scandinavian Journal of Statistics 27, 1, 65-82, 2000.
- [25] S. Kusuoka and D. Stroock. *Applications of the Malliavin Calculus, Part II*. J. Fac. Sci. Univ. Tokyo 32, 1-76, 1985.
- [26] Y. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer, 2004.
- [27] N. Marie and A. Rosier. *Nadaraya-Watson Estimator for IID Paths of Diffusion Processes*. Scandinavian Journal of Statistics 50, 2, 589-637, 2023.
- [28] 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.
- [29] R. Overgaard, N. Jonsson, C. Tornøe, H. Madsen. *Non-Linear Mixed Effects Models with Stochastic Differential Equations: Implementation of an Estimation Algorithm*. Journal of Pharmacokinetics and Pharmacodynamics 32, 85-107, 2005.
- [30] U. Picchini, A. De Gaetano and S. Ditlevsen. *Stochastic Differential Mixed-Effects Models*. Scandinavian Journal of Statistics 37, 67-90, 2010.
- [31] U. Picchini and S. Ditlevsen. *Practical Estimation of High Dimensional Stochastic Differential Mixed-Effects Models*. Computational Statistics and Data Analysis 55, 1426-1444, 2011.

Email address: nmarie@parisnanterre.fr

Email address: amelie.rosier@esme.fr

[†]LABORATOIRE MODAL'X, UNIVERSITÉ PARIS NANTERRE, NANTERRE, FRANCE

[◇]ESME, IVRY-SUR-SEINE, FRANCE

*UMR MIA-PARIS-SACLAY, INRAE, FRANCE