Annals of the Institute of Statistical Mathematics (2024) 76:209–234
https://doi.org/10.1007/s10463-023-00881-7

On a projection least squares estimator for jump diffusion processes

Hélène Halconruy[1] · Nicolas Marie[2]

Received: 31 October 2022 / Revised: 17 April 2023 / Accepted: 20 July 2023 /
Published online: 11 September 2023

Abstract

This paper deals with a projection least squares estimator of the drift function of a jump diffusion process X computed from multiple independent copies of X observed on $[0, T]$. Risk bounds are established on this estimator and on an associated adaptive estimator. Finally, some numerical experiments are provided.

Keywords Projection least squares estimator · Model selection · Jump diffusion processes

1 Introduction

Let $Z = (Z_t)_{t\in[0,T]}$ be the compound Poisson process defined by

$$Z_t := \sum_{n=1}^{\nu_t} \zeta_n = \int_0^t \int_{-\infty}^{\infty} z\mu(ds, dz)$$

for every $t \in [0, T]$, where $\nu = (\nu_t)_{t\in[0,T]}$ is a (usual) Poisson process of intensity $\lambda > 0$, independent of the ζ_n's which are i.i.d. random variables of probability distribution π, and μ is the Poisson random measure of intensity $m(ds, dz) := \lambda\pi(dz)ds$ defined by

$$\mu([0, t] \times dz) := |\{s \in [0, t] : Z_s - Z_{s^-} \in dz\}|; \forall t \in [0, T].$$

✉ Nicolas Marie
nmarie@parisnanterre.fr

Hélène Halconruy
helene.halconruy@devinci.fr

1 Léonard de Vinci Pôle universitaire, Research Center, 12 avenue Léonard de Vinci, 92400 Courbevoie, France

2 Laboratoire Modal'X, Université Paris Nanterre, 200 avenue de la République, 92001 Nanterre, France

In the sequel, Z is *replaced* by the centered martingale $\mathfrak{Z} = (\mathfrak{Z}_t)_{t\in[0,T]}$ defined by

$$\mathfrak{Z}_t := Z_t - \int_0^t \int_{-\infty}^{\infty} zm(ds, dz) = Z_t - \mathfrak{c}_\zeta \lambda t$$

for every $t \in [0, T]$, where $\mathfrak{c}_\zeta$ is the (common) expectation of the ζ_n's. Now, let us consider the stochastic differential equation

$$X_t = x_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s + \int_0^t \gamma(X_s)d\mathfrak{Z}_s;\ t \in [0, T], \tag{1}$$

where $x_0 \in \mathbb{R}$, $B = (B_t)_{t\in[0,T]}$ is a Brownian motion independent of Z, $b \in C^1(\mathbb{R})$ and its derivative is bounded, and $\sigma, \gamma : \mathbb{R} \to \mathbb{R}$ are bounded Lipschitz continuous functions such that $\inf_{x\in\mathbb{R}} \sigma(x)^2 \wedge \gamma(x)^2 > 0$. Under these conditions on b, σ and γ, Equation (1) has a unique (strong) solution $X = (X_t)_{t\in[0,T]}$.

As for continuous diffusion processes, the major part of the estimators of the drift function in stochastic differential equations driven by jump processes is computed from one path of the solution to Equation (1) and converges when $T \to \infty$ (see Schmisser, 2014; Gloter, et al. 2018; Amorino et al. 2022, etc.). The existence and the uniqueness of the ergodic stationary solution to Equation (1) is then required and obtained thanks to a restrictive dissipativity condition on b. For stochastic differential equations driven by a pure-jump Lévy process, some authors have also studied estimation methods based on high frequency observations, on a fixed time interval, of one path of the solution (see Clément and Gloter, 2019, 2020).

Now, consider $X^i := \mathcal{I}(x_0, B^i, \mathfrak{Z}^i)$ for every $i \in \{1, \dots, N\}$, where $\mathcal{I}(.)$ is the Itô map associated with Equation (1) and $(B^1, \mathfrak{Z}^1), \dots, (B^N, \mathfrak{Z}^N)$ are $N \in \mathbb{N}^*$ independent copies of $(B, \mathfrak{Z})$. The estimation of the drift function b from a continuous-time or a discrete-time observation of $(X^1, \dots, X^N)$ is a functional data analysis problem already investigated in the parametric and in the nonparametric frameworks for continuous diffusion processes (see Ditlevsen and De Gaetano, 2005; Picchini and Ditlevsen, 2011; Delattre, Genon-Catalot and Samson, 2013; Comte and Genon-Catalot, 2020b; Denis, et al. 2021; Marie and Rosier, 2023, etc.). Up to our knowledge, no such estimator of the drift function has been already proposed for jump diffusion processes. So, our paper deals with a projection least squares estimator $\widehat{b_m}$ of b computed from $X^1, \dots, X^N$, which means that $\widehat{b}_m$ is minimizing the objective function

$$\tau \longmapsto \gamma_N(\tau) := \frac{1}{NT} \sum_{i=1}^{N} \left(\int_0^T \tau(X_s^i)^2 \mathrm{d}s - 2 \int_0^T \tau(X_s^i) \mathrm{d}X_s^i \right)$$

on a m-dimensional function space $\mathcal{S}_m$. Precisely, risk bounds are established on $\widehat{b}_m$ and on the adaptive estimator $\widehat{b}_{\widehat{m}}$, where

$$\widehat{m} = \arg \min_{m\in\widehat{\mathcal{M}}_N} \{-\|\widehat{b}_m\|_N^2 + \text{pen}(m)\}$$

with $\widehat{\mathcal{M}}_N \subset \{1, \dots, N\}$,

$$\mathrm{pen}(m) := \mathfrak{c}_{\mathrm{cal}}\frac{m}{N};\ \forall m \in \mathbb{N}^*$$

and $\mathfrak{c}_{\mathrm{cal}} > 0$ is a constant to calibrate in practice.

In Sect. 2, a detailed definition of the projection least squares estimator of b is provided. Section 3 deals with a risk bound on $\widehat{b}_m$ and Sect. 4 with a risk bound on the adaptive estimator $\widehat{b}_{\widehat{m}}$. Finally, some numerical experiments are provided in Sect. 5. The proofs (resp. tables and figures) are postponed to Appendix A(resp. Appendix B).

Notations and basic definitions

- $\mathfrak{c}_{\zeta^n} := \mathbb{E}(\zeta_1^n)$ for every $n \in \mathbb{N}^*$.
- Consider $d \in \mathbb{N}^*$. The j-th component of any $\mathbf{x} \in \mathbb{R}^d$ is denoted by $\mathbf{x}_j$ or $[\mathbf{x}]_j$.
- For every $k \in \mathbb{N}^*$, $\|.\|_{k,d}$ is the norm on $\mathbb{R}^d$ defined by

$$\|\mathbf{x}\|_{k,d} := \left(\sum_{j=1}^{d} |\mathbf{x}_j|^k\right)^{1/k};\ \forall \mathbf{x} \in \mathbb{R}^d.$$

- The spectral norm on the space $\mathcal{M}_d(\mathbb{R})$ of the $d \times d$ real matrices is denoted by $\|.\|_{\mathrm{op}}$:

$$\|\mathbf{A}\|_{\mathrm{op}} := \sup_{x\in\mathbb{R}^d\,:\,\|x\|_{2,d}=1} \|\mathbf{A}\mathbf{x}\|_{2,d};\ \forall \mathbf{A} \in \mathcal{M}_d(\mathbb{R}).$$

2 A projection least squares estimator of the drift function

2.1 The objective function

Assume that the probability distribution of X_s has a density $p_s(x_0, .)$ with respect to Lebesgue's measure for every $s \in (0, T]$ that $s \mapsto p_s(x_0, x)$ belongs to $\mathbb{L}^1([0, T], dt)$ for every $x \in \mathbb{R}$ which legitimates to consider the density function f_T defined by

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

and that

$$\int_{-\infty}^{\infty} b(x)^4 f_T(x)dx < \infty.$$

Remark 1 Assume that b and π satisfy the following additional conditions:

(1) The function b belongs to the Kato class

$$\mathbb{K}_2 := \left\{\varphi : \mathbb{R} \to \mathbb{R} : \lim_{\delta\to 0}\sup_{x\in\mathbb{R}} \int_0^{\delta}\int_{-\infty}^{\infty} |\varphi(x+y)+\varphi(x-y)|s^{1/2}(|y|+s^{1/2})^{-3}\mathrm{d}y ds = 0\right\}.$$

(2) The Lévy measure $\pi_\lambda(.) := \lambda\pi(.)$ has a density θ with respect to Lebesgue's measure. Moreover, there exists $\alpha \in (0,2)$ such that $z \in \mathbb{R} \mapsto \theta(z)|z|^{1+\alpha}$ is bounded, and if $\alpha = 1$, then

$$\int_{r<|z|<R} z\theta(z)dz = 0;\ \forall R > r > 0.$$

By Chen et al. (2017), Theorem 1.1 and the remark p. 126, l. 5-7, in Amorino and Gloter (2021), for every $s \in (0,T]$, the probability distribution of X_s has a density $p_s(x_0,.)$ with respect to Lebesgue's measure, and there exist two constants $\mathfrak{c}_p, \mathfrak{m}_p > 0$, not depending on s, such that

$$p_s(x_0,x) \leqslant \mathfrak{c}_p\left[s^{-1/2}\exp\left(-\mathfrak{m}_p\frac{(x-x_0)^2}{s}\right) + \frac{s}{(s^{1/2}+|x-x_0|)^{1+\alpha}}\right];\ \forall x \in \mathbb{R}. \quad (2)$$

So,

- f_T is well-defined and even bounded, which is crucial in Sect. 4. Indeed, since $-1/2 < 1-(1+\alpha)/2 < 1/2$, for every $x \in \mathbb{R}$,

$$0 \leqslant f_T(x) \leqslant \frac{\mathfrak{c}_p}{T}\left(\int_0^T s^{-1/2}\mathrm{d}s + \int_0^T s^{1-(1+\alpha)/2}\mathrm{d}s\right) = \frac{\mathfrak{c}_p}{T}\left[2T^{1/2} + \frac{T^{2-(1+\alpha)/2}}{2-(1+\alpha)/2}\right] < \infty.$$

- If there exists a constant $\mathfrak{c}_b > 0$ and $\varepsilon \in (0,\alpha)$ as close as possible to 0 such that $|b(x)| \leqslant \mathfrak{c}_b(1+|x|)^{(\alpha-\varepsilon)/4}$ for every $x \in \mathbb{R}$, then

$$b(x)^4 p_s(x_0,x) \underset{x\to\pm\infty, s\to 0^+}{=} O\left(\frac{s^{-1/2}}{|x-x_0|^{1+\varepsilon}}\right),$$

  which leads to

$$\int_{-\infty}^{\infty} b(x)^4 f_T(x)\mathrm{d}x < \infty.$$

Now, let us consider the objective function $\gamma_N(.)$ defined by

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

for every $\tau \in \mathcal{S}_m$, where $m \in \{1, \dots, N_T\}$, $N_T := [NT] + 1$, $\mathcal{S}_m := \text{span}\{\varphi_1, \dots, \varphi_m\}$, $\varphi_1, \dots, \varphi_{N_T}$ are continuous functions from I into $\mathbb{R}$ such that $(\varphi_1, \dots, \varphi_{N_T})$ is an orthonormal family in $\mathbb{L}^2(I, dx)$, and $I \subset \mathbb{R}$ is a non-empty interval. For any $\tau \in \mathcal{S}_m$,

$$\begin{aligned}\mathbb{E}(\gamma_N(\tau)) =& \frac{1}{T}\int_0^T \mathbb{E}(|\tau(X_s) - b(X_s)|^2)\mathrm{d}s - \frac{1}{T}\int_0^T \mathbb{E}(b(X_s)^2)\mathrm{d}s \\ =& \int_{-\infty}^{\infty} (\tau(x) - b(x))^2 f_T(x)\mathrm{d}x - \int_{-\infty}^{\infty} b(x)^2 f_T(x)\mathrm{d}x.\end{aligned}$$

Then, the closer τ is to b, the smaller $\mathbb{E}(\gamma_N(\tau))$. For this reason, the estimator of b minimizing $\gamma_N(.)$ is studied in this paper.

2.2 The projection least squares estimator and some related matrices

Consider

$$J := \sum_{j=1}^{m} \theta_j \varphi_j \quad \text{with} \quad \theta_1, \dots, \theta_m \in \mathbb{R}.$$

Then,

$$\begin{aligned}\nabla\gamma_N(J) =& \left(\frac{1}{NT}\sum_{i=1}^{N}\left(2\sum_{\ell=1}^{m}\theta_\ell \int_0^T \varphi_j(X_s^i)\varphi_\ell(X_s^i)ds - 2\int_0^T \varphi_j(X_s^i)dX_s^i\right)\right)_{j\in\{1,\dots,m\}} \\ =& 2(\widehat{\mathbf{\Psi}}_m(\theta_1, \dots, \theta_m)^* - \widehat{\mathbf{X}}_m)\end{aligned}$$

where

$$\widehat{\mathbf{\Psi}}_m := \left(\frac{1}{NT}\sum_{i=1}^{N}\int_0^T \varphi_j(X_s^i)\varphi_\ell(X_s^i)\mathrm{d}s\right)_{j,\ell\in\{1,\dots,m\}}$$

and

$$\widehat{\mathbf{X}}_m := \left(\frac{1}{NT}\sum_{i=1}^{N}\int_0^T \varphi_j(X_s^i)\mathrm{d}X_s^i\right)_{j\in\{1,\dots,m\}}.$$

The symmetric matrix $\widehat{\mathbf{\Psi}}_m$ is positive semidefinite because

$$\mathbf{u}^*\widehat{\mathbf{\Psi}}_m\mathbf{u} = \frac{1}{NT}\sum_{i=1}^{N}\int_0^T \left(\sum_{j=1}^{m}\mathbf{u}_j\varphi_j(X_s^i)\right)^2 ds \geqslant 0$$

for every $\mathbf{u} \in \mathbb{R}^m$. If in addition $\widehat{\mathbf{\Psi}}_m$ is invertible, it is positive definite, and then

$$\widehat{b}_m = \sum_{j=1}^{m} \widehat{\theta}_j \varphi_j \quad \text{with} \quad \widehat{\boldsymbol{\theta}}_m := (\widehat{\theta}_1, \dots, \widehat{\theta}_m)^* = \widehat{\boldsymbol{\Psi}}_m^{-1} \widehat{\mathbf{X}}_m \tag{3}$$

is the only minimizer of $\gamma_N(.)$ on $\mathcal{S}_m$, called the projection least squares estimator of b.

Remarks

(1) $\widehat{\boldsymbol{\Psi}}_m = (\langle \varphi_j, \varphi_\ell \rangle_N)_{j,\ell}$, where

$$\langle \varphi, \psi \rangle_N := \frac{1}{NT} \sum_{i=1}^{N} \int_0^T \varphi(X_s^i)\psi(X_s^i)ds$$

for every continuous functions $\varphi, \psi : \mathbb{R} \to \mathbb{R}$.

(2) $\widehat{\mathbf{X}}_m = (\langle b, \varphi_j \rangle_N)_j^* + \widehat{\mathbf{E}}_m$, where

$$\widehat{\mathbf{E}}_m := \left(\frac{1}{NT} \sum_{i=1}^{N} \int_0^T \varphi_j(X_s^i)(\sigma(X_s^i)dB_s^i + \gamma(X_s^i)d\mathfrak{Z}_s^i) \right)_{j \in \{1,\dots,m\}}^* .$$

Let us introduce the two following deterministic matrices related to the previous random ones:

$$\boldsymbol{\Psi}_{m,\sigma} := NT\mathbb{E}(\widehat{\mathbf{E}}_m \widehat{\mathbf{E}}_m^*) \quad \text{and} \quad \boldsymbol{\Psi}_m := \mathbb{E}(\widehat{\boldsymbol{\Psi}}_m) = (\langle \varphi_j, \varphi_\ell \rangle_{f_T})_{j,\ell},$$

where $\langle .,. \rangle_{f_T}$ is the usual scalar product in $\mathbb{L}^2(I, f_T(x)dx)$. The following lemma will be crucial in Sect. 3 to evaluate the order of the variance term in the risk bound on our projection least squares estimator of b.

Lemma 1 $\mathrm{trace}(\boldsymbol{\Psi}_m^{-1/2} \boldsymbol{\Psi}_{m,\sigma} \boldsymbol{\Psi}_m^{-1/2}) \leqslant (\|\sigma\|_\infty^2 + \lambda \mathfrak{c}_{\zeta^2} \|\gamma\|_\infty^2)m.$

Let us conclude this section with few words about the extension of the projection least squares estimation for multidimensional diffusion processes, and on the reason why the probability distribution of X_s, $s \in (0, T]$, needs to have a density with respect to Lebesgue's measure with a sharp bound as (2).

Remarks

(1) Assume that X is a d-dimensional diffusion process with $d \in \mathbb{N}^*$. A natural extension of the objective function γ_N would be defined by

$$\gamma_{d,N}(\tau) := \frac{1}{NT} \sum_{i=1}^{N} \left(\int_0^T \|\tau(X_s^i)\|_{2,d}^2 ds - 2 \int_0^T \langle \tau(X_s^i), dX_s^i \rangle_{2,d} \right) ; \tau \in \mathcal{S}_{d,m},$$

where $\mathcal{S}_{d,m} := \mathrm{span}(\{\varphi_{j_1} \otimes \cdots \otimes \varphi_{j_d}; j_1, \dots, j_d \in \{1, \dots, m\}\})^d$. This is out of the scope of our paper, but note that one could get an expression of the solution of the minimization problem $\min_{\mathcal{S}_{d,m}} \gamma_{d,N}$ similar to (3) but involving hypermatrices in the spirit of Dussap (2021). So, except in the very special case where the components of X are independent, to extend our estimation method to the multidimensional framework is not straightforward.

(2) By the Fubini–Tonelli theorem, for every continuous functions $\varphi, \psi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}(\langle \varphi, \psi \rangle_N) = \int_{-\infty}^{\infty} \varphi(x)\psi(x)\mathbb{P}_T(\mathrm{d}x) \quad \text{with} \quad \mathbb{P}_T(\mathrm{d}x) := \frac{1}{T}\int_0^T \mathbb{P}_{X_s}(dx)ds,$$

and then, one could think that there is no need to assume that $\mathbb{P}_{X_s}(\mathrm{d}x) = p_s(x_0, x)dx$ for any $s \in (0, T]$. However, since the drift function b may be unbounded, a sharp bound on $\mathbb{P}_{X_s}(dx)$ as (2) is required to show that

$$\int_{-\infty}^{\infty} b(x)^4 \mathbb{P}_T(\mathrm{d}x) < \infty$$

as assumed in the beginning of Sect. 2.1. Note also that if the Malliavin covariance (matrix) of X_s, $s \in (0, T]$ is almost surely invertible (Bouleau–Hirsch's condition), then, the probability distribution of X_s has a density with respect to Lebesgue's measure, but not necessarily with a sharp bound as (2). For instance, X_s satisfies the Bouleau–Hirsch condition under Assumption 2.4 on b, σ and γ in Bichteler and Jacod (1983).

3 Risk bound on the projection least squares estimator

This section deals with a risk bound on the truncated estimator

$$\widetilde{b}_m := \widehat{b}_m \mathbf{1}_{\Lambda_m},$$

where

$$\Lambda_m := \left\{ L(m)(\|\widehat{\mathbf{\Psi}}_m^{-1}\|_{\mathrm{op}} \vee 1) \leqslant \mathfrak{c}_T \frac{NT}{\log(NT)} \right\}$$

with

$$L(m) = 1 \vee \left(\sup_{x \in I} \sum_{j=1}^{m} \varphi_j(x)^2 \right) \quad \text{and} \quad \mathfrak{c}_T = \frac{3\log(3/2) - 1}{8T}.$$

On the event Λ_m, $\widehat{\mathbf{\Psi}}_m$ is invertible because

$$\inf\{\mathrm{sp}(\widehat{\mathbf{\Psi}}_m)\} \geqslant \frac{L(m)}{\mathfrak{c}_T} \cdot \frac{\log(NT)}{NT},$$

and then, $\widetilde{b}_m$ is well-defined. In the sequel, m fulfills the following assumption.

Assumption 1 $L(m)(\|\mathbf{\Psi}_m^{-1}\|_{\mathrm{op}} \vee 1) \leqslant \frac{\mathfrak{c}_T}{2} \cdot \frac{NT}{\log(NT)}$.

The above condition is a generalization of the so-called *stability condition* introduced in the nonparametric regression framework in Cohen et al. (2013), and already extended to the independent copies of continuous diffusion processes framework in Comte and Genon-Catalot (2020b).

Theorem 1 *Under Assumption 1, there exists a constant* $\mathfrak{c}_1 > 0$, *not depending on* m *and* N, *such that*

$$\mathbb{E}(\|\widetilde{b}_m - b_I\|_N^2) \leqslant \min_{\tau \in \mathcal{S}_m} \|\tau - b_I\|_{f_T}^2 + \mathfrak{c}_1\left(\frac{m}{NT} + \frac{1}{N}\right). \tag{4}$$

Remarks

- Note that Theorem 1 says first that the bound on the variance of $\widetilde{b}_m$ is of order m/N as in the nonparametric regression framework, in which it is optimal (see Comte and Genon-Catalot, 2020a, Theorem 1). For this reason, (4) should be near to optimality, but it's a difficult challenge, out of the scope of this paper, to establish a lower bound on $\widehat{b}_m$. Note that even for continuous diffusion processes, no lower bound has been established (see Comte and Genon-Catalot, 2020b).
- The order of the bias term

$$\min_{\tau \in \mathcal{S}_m} \|\tau - b_I\|_{f_T}^2,$$

as well as $L(m)$ and $\|\mathbf{\Psi}_m^{-1}\|_{\mathrm{op}}$, depend on the φ_j's. Let us evaluate them for the trigonometric basis, which is compactly supported, and for the Hermite basis, which is $\mathbb{R}$-supported:

(1) Assume that $I = [\ell, \mathtt{r}]$ with $\ell, \mathtt{r} \in \mathbb{R}$ satisfying $\ell < \mathtt{r}$ and that

$$\varphi_1(x) := \sqrt{\frac{1}{\mathtt{r} - \ell}}\mathbf{1}_{[\ell,\mathtt{r}]}(x),$$
$$\varphi_{2j+1}(x) := \sqrt{\frac{2}{\mathtt{r} - \ell}} \sin\left(2\pi j \frac{x - \ell}{\mathtt{r} - \ell}\right)\mathbf{1}_{[\ell,\mathtt{r}]}(x) \quad \text{and}$$
$$\varphi_{2j}(x) := \sqrt{\frac{2}{\mathtt{r} - \ell}} \cos\left(2\pi j \frac{x - \ell}{\mathtt{r} - \ell}\right)\mathbf{1}_{[\ell,\mathtt{r}]}(x)$$

for every $x \in [\ell, \mathtt{r}]$ and $j \in \mathbb{N}^*$. On the one hand, since $\cos(.)^2 + \sin(.)^2 = 1$, there exists a constant $\mathfrak{c}_\varphi > 0$, not depending on m, N and T, such that

$$L(m) = 1 \vee \left[\sup_{x\in I}\sum_{j=1}^{m}\varphi_j(x)^2\right] \leqslant \mathfrak{c}_{\varphi}m.$$

Moreover, since under the same conditions than in Remark 1, there exists $\underline{\mathfrak{m}} > 0$ such that $f_T(.) \geqslant \underline{\mathfrak{m}}$ on I by Chen et al. (2017), Theorem 1.3,

$$\begin{aligned}\|\mathbf{\Psi}_m^{-1}\|_{\mathrm{op}} = \frac{1}{\lambda_{\min}(\mathbf{\Psi}_m)} =& \left(\inf_{\theta\,:\,\|\theta\|_{2,m}=1}\sum_{j,\ell=1}^{m}\theta_j\theta_\ell[\mathbf{\Psi}_m]_{j,\ell}\right)^{-1}\\ =& \left[\inf_{\theta\,:\,\|\theta\|_{2,m}=1}\int_{\ell}^{\mathtt{r}}\left(\sum_{j=1}^{m}\theta_j\varphi_j(x)\right)^2 f_T(x)\mathrm{d}x\right]^{-1} \leqslant \frac{1}{\underline{\mathfrak{m}}}.\end{aligned}$$

Then,

$$m \leqslant \frac{1}{\mathfrak{c}_{\varphi}(\underline{\mathfrak{m}}^{-1}\vee 1)}\cdot\frac{\mathfrak{c}_T}{2}\cdot\frac{NT}{\log(NT)}$$

satisfies Assumption 1. On the other hand, consider the Fourier-Sobolev space

$$\mathbb{W}_2^{\beta}([\ell,\mathtt{r}]) := \left\{\varphi\in C^{\beta}([\ell,\mathtt{r}];\mathbb{R})\ :\ \int_{\ell}^{\mathtt{r}}\varphi^{(\beta)}(x)^2\mathrm{d}x < \infty\right\}$$

with $\beta\in\mathbb{N}^*$, and assume that $b_{|I}\in\mathbb{W}_0^{\beta}([\ell,\mathtt{r}])$. By DeVore and Lorentz (1993), Corollary 2.4 p. 205, there exists a constant $\mathfrak{c}_{\beta,\ell,\mathtt{r}} > 0$, not depending on m and N, such that

$$\|\Pi_m(b_I) - b_I\|^2 \leqslant \mathfrak{c}_{\beta,\ell,\mathtt{r}}m^{-2\beta},$$

where Π_m is the orthogonal projection from $\mathbb{L}^2(I,dx)$ onto $\mathcal{S}_m$. Since under appropriate conditions on b and π, as explained in Remark 1, f_T is upper bounded on I by a constant $\overline{\mathfrak{m}} > 0$,

$$\begin{aligned}\min_{\tau\in\mathcal{S}_m}\|\tau - b_I\|_{f_T}^2 \leqslant& \overline{\mathfrak{m}}\|\Pi_m(b_I) - b_I\|^2\\ \leqslant& \overline{\mathfrak{c}}_{\beta,\ell,\mathtt{r}}m^{-2\beta}\quad\text{with}\quad\overline{\mathfrak{c}}_{\beta,\ell,\mathtt{r}} = \mathfrak{c}_{\beta,\ell,\mathtt{r}}\overline{\mathfrak{m}}.\end{aligned}$$

In conclusion, by Theorem 1, there exists a constant $\overline{\mathfrak{c}}_{1,1} > 0$, not depending m and N, such that

$$\mathbb{E}(\|\widetilde{b}_m - b_I\|_N^2) \leqslant \overline{\mathfrak{c}}_{1,1}\left(m^{-2\beta} + \frac{m}{N}\right),$$

and then, the bias-variance tradeoff is reached by (the risk bound on) $\widetilde{b}_m$ for m of order $N^{1/(1+2\beta)}$.

(2) Assume that $\varphi_j = h_{j-1}$ for any $j\in\{1,\dots,m\}$, where $(h_n)_{n\in\mathbb{N}}$ is the Hermite basis: for every $x\in\mathbb{R}$ and $n\in\mathbb{N}$,

$$h_n(x) := (2^n n!\sqrt{\pi})^{-1/2} H_n(x) e^{-x^2/2} \quad \text{with} \quad H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}.$$

On the one hand, by Abramowitz and Stegun (1964), $\|\varphi_j\|_\infty \leqslant \pi^{-1/4}$ and then $L(m) \leqslant m$. Moreover, assume that the conditions (1) and (2) in Remark 1 are satisfied by b and π with $\alpha \in [1,2)$. By Chen et al. (2017), Theorem 1.3, for every $s \in (0,T]$, there exist two constants $\bar{\mathfrak{c}}_p, \overline{\mathfrak{m}}_p > 0$, not depending on s, such that for any $x \in \mathbb{R}$,

$$\begin{aligned} p_s(x_0,x) \geqslant & \bar{\mathfrak{c}}_p \left[s^{-1/2} \exp\left(-\overline{\mathfrak{m}}_p \frac{(x-x_0)^2}{s} \right) + \frac{s}{(s^{1/2} + |x-x_0|)^{1+\alpha}} \right] \\ \geqslant & \frac{\bar{\mathfrak{c}}_p s}{(2T + 4x_0^2 + 4x^2)^{(1+\alpha)/2}} \geqslant \frac{\bar{\mathfrak{c}}_p s}{[4 \vee (2T + 4x_0^2)]^{(1+\alpha)/2}} \cdot \frac{1}{(1+x^2)^{(1+\alpha)/2}}. \end{aligned}$$

So,

$$f_T(x) \geqslant \frac{\mathfrak{c}_{f_T}}{(1+x^2)^{(1+\alpha)/2}} \quad \text{with} \quad \mathfrak{c}_{f_T} = \frac{\bar{\mathfrak{c}}_p T}{2[4 \vee (2T + 4x_0^2)]^{(1+\alpha)/2}}.$$

Since $\alpha \in [1,2)$, by Comte and Genon-Catalot (2020a), Proposition 9, there exists a constant $\mathfrak{c}_{\boldsymbol{\Psi}} > 0$, not depending on m, such that

$$\|\boldsymbol{\Psi}_m^{-1}\|_{\mathrm{op}} \leqslant \mathfrak{c}_{\boldsymbol{\Psi}} m^{(1+\alpha)/2}.$$

Then,

$$m \leqslant \left[\frac{1}{\mathfrak{c}_{\boldsymbol{\Psi}} \vee 1} \cdot \frac{\mathfrak{c}_T}{2} \cdot \frac{NT}{\log(NT)} \right]^{2/(3+\alpha)}$$

satisfies Assumption 1. On the other hand, assume that $b \in \mathbb{W}_{\mathrm{H}}^{\beta}(D)$, where $\mathbb{W}_{\mathrm{H}}^{\beta}(D)$ is the Hermite-Sobolev ball defined by

$$\mathbb{W}_{\mathrm{H}}^{\beta}(D) := \left\{ \varphi \in \mathbb{L}^2(\mathbb{R}) : \sum_{n=0}^{\infty} n^{\beta} \langle \varphi, h_n \rangle^2 \leqslant D \right\}$$

with $\beta > (3+\alpha)/2 - 1$ and $D > 0$. Then, $\|\Pi_m(b) - b\|^2 \leqslant Dm^{-\beta}$ (see Belomestny et al. (2019), Section 4.2), and since f_T is upper bounded on $\mathbb{R}$ by a constant $\overline{\mathfrak{m}} > 0$ (see Remark 1),

$$\begin{aligned} \min_{\tau \in \mathcal{S}_m} \|\tau - b\|_{f_T}^2 \leqslant & \overline{\mathfrak{m}} \|\Pi_m(b) - b\|^2 \\ \leqslant & \mathfrak{c}_{\mathrm{H}} m^{-\beta} \quad \text{with} \quad \mathfrak{c}_{\mathrm{H}} = D\overline{\mathfrak{m}}. \end{aligned}$$

In conclusion, by Theorem 1, there exists a constant $\bar{\mathfrak{c}}_{1,2} > 0$, not depending m and N, such that

$$\mathbb{E}(\|\widetilde{b}_m - b\|_N^2) \leqslant \bar{\mathfrak{c}}_{1,2} \left(m^{-\beta} + \frac{m}{N} \right),$$

and then the bias-variance tradeoff is reached by (the risk bound on) $\widetilde{b}_m$ for m of order $N^{1/(1+\beta)}$.

4 Model selection

First of all, let us state some additional assumptions on the φ_j's, on the Lévy measure π_λ and on the density function f_T.

Assumption 2 The spaces $\mathcal{S}_m$, $m \in \{1, \dots, N_T\}$, satisfy the following conditions:

(1) There exists a constant $\mathfrak{c}_\varphi \geqslant 1$, not depending on N, such that for every $m \in \{1, \dots, N_T\}$, the basis $(\varphi_1, \dots, \varphi_m)$ of $\mathcal{S}_m$ satisfies

$$L(m) = 1 \vee \left[\sup_{x \in I} \sum_{j=1}^{m} \varphi_j(x)^2\right] \leqslant \mathfrak{c}_\varphi^2 m.$$

(2) (Nested spaces) For every $m, m' \in \{1, \dots, N_T\}$, if $m > m'$, then $\mathcal{S}_{m'} \subset \mathcal{S}_m$.

The proof of the oracle inequality satisfied by the adaptive estimator relies on a Bernstein type inequality (see Lemma 5). In order to control the *big jumps* of X in the proof of Lemma 5, the Lévy measure π_λ needs to fulfill the following assumption.

Assumption 3 The Lévy measure π_λ is sub-exponential: there exist positive constants $\mathfrak{a}, \mathfrak{b} > 0$ such that, for every $x > 1$,

$$\pi_\lambda((-x, x)^c) \leqslant \mathfrak{a} e^{-\mathfrak{b}|x|}.$$

Assumption 4 The density function f_T is bounded.

Note that Remark 1 provides a sufficient condition for Assumption 4 to be satisfied by f_T. Now, let us introduce the set

$$\widehat{\mathcal{M}}_N := \left\{ m \in \{1, \dots, N_T\} \, : \, \mathfrak{c}_\varphi^2 m (\|\widehat{\boldsymbol{\Psi}}_m^{-1}\|_{\mathrm{op}}^2 \vee 1) \leqslant \mathfrak{d}_T \frac{NT}{\log(NT)} \right\}$$

with

$$\mathfrak{d}_T = \min\left\{ \frac{\mathfrak{c}_T}{2}, \frac{1}{64 \mathfrak{c}_\varphi^2 T(\|f_T\|_\infty + \sqrt{\mathfrak{c}_T/2}/(3\mathfrak{c}_\varphi))} \right\},$$

as well as its theoretical counterpart

$$\mathcal{M}_N := \left\{ m \in \{1, \dots, N_T\} : \mathfrak{c}_\varphi^2 m(\|\mathbf{\Psi}_m^{-1}\|_{\text{op}}^2 \vee 1) \leqslant \frac{\mathfrak{d}_T}{4} \cdot \frac{NT}{\log(NT)} \right\}.$$

Let us also consider

$$\widehat{m} := \arg \min_{m \in \widehat{\mathcal{M}}_N} \{-\|\widehat{b}_m\|_N^2 + \text{pen}(m)\},$$

where

$$\text{pen}(m) := \mathfrak{c}_{\text{cal}} \frac{m}{NT}; \ \forall m \in \{1, \dots, N_T\}$$

and $\mathfrak{c}_{\text{cal}} > 0$ is a deterministic constant to calibrate.

Theorem 2 *Under Assumptions 1, 2, 3 and 4, there exists a constant $\mathfrak{c}_2 > 0$, not depending on N, such that*

$$\mathbb{E}(\|\widehat{b}_{\widehat{m}} - b_I\|_N^2) \leqslant \mathfrak{c}_2 \left[\min_{m \in \mathcal{M}_N} \left\{ \min_{\tau \in \mathcal{S}_m} \|\tau - b_I\|_{f_T}^2 + \frac{m}{NT} \right\} + \frac{1}{N} \right].$$

Note that Theorem 2 provides a risk bound on the adaptive estimator $\widehat{b}_{\widehat{m}}$ of same order than the minimal risk bound on $\widetilde{b}_m$ (see Theorem 1) for m taken in $\mathcal{M}_N$.

5 Numerical experiments

First of all, let us recall that all figures and tables related to this section are postponed to Appendix B.

Some numerical experiments on our estimation method of b in Equation (1) are presented in this subsection when the common distribution π of the ζ_n's is standard normal, which implies that $\mathfrak{Z} = Z$, and the intensity of the (usual) Poisson process ν is $\lambda = 0.5$. The estimation method investigated on the theoretical side in Sects. 3 and 4 is implemented here for the three following models:

(1) $X_t = 0.5 - \int_0^t X_s ds + 0.5 B_t + Z_t$ (linear/additive noise).

(2) $X_t = 0.5 + 0.5 \int_0^t \sqrt{1 + X_s^2} ds + 0.5 B_t + Z_t$ (nonlinear/additive noise).

(3) $X_t = 0.5 + 0.5 \int_0^t \sqrt{1 + X_s^2} ds + 0.5 \int_0^t (1 + \cos(X_s)^2) dB_s + Z_t$ (nonlinear/multiplicative noise).

For each of the three previous models, our adaptive estimator of b is computed on $I = [-3, 3]$ from $N = 400$ paths of the process X observed along the dissection $\{lT/n; l = 1, \dots, n\}$ of $[0, T]$ with $n = 200$ and $T = 5$, when $(\varphi_1, \dots, \varphi_m)$ is the m-dimensional trigonometric basis for every $m \in \{1, \dots, 6\}$. This experiment is repeated 100 times, and the means and the standard deviations of the MISE of $\widehat{b}_{\widehat{m}}$ are stored in Table 1. Moreover, for each model, 10 estimations (dashed black curves) of b (red curve) are plotted on Figs. 1, 2 and 3.

On the one hand, on average, the MISE of our adaptive estimator is slightly increasing with the *complexity* of the model: 0.1251 (Model 1) < 0.1469 (Model 2)< 0.1825 (Model 3). The same remark holds for its standard deviation: 0.0950 (Model 1) < 0.1688 (Model 2) < 0.1928 (Model 3). This means that the more the model is *complex*, the more the quality of the estimation degrades, and it is visible on Figs. 1, 2 and 3. However, for all the three previous models, the MISE of $\widehat{b}_{\widehat{m}}$ remains good and of same order than for models without jump component (see Comte and Genon-Catalot, 2020b, Sect. 4). On the other hand, the model selection procedure works well because it does not systematically select the lowest or the highest possible value of m (see $\widehat{m}$ on Figs. 1, 2 and 3). For Model 1, the procedure is stable because $\widehat{m}$ has a low estimated standard deviation (0.4830). The procedure remains satisfactory for Models 2 and 3, with $\mathrm{StD}(\widehat{m}) = 1.1353$ and $\mathrm{StD}(\widehat{m}) = 1.2867$, respectively, but not as much as for Model 1.

Appendix A. Proofs

A.1 Proof of Lemma 1

First, let us show that the symmetric matrix $\mathbf{\Psi}_{m,\sigma}$ is positive semidefinite. Indeed, for any $y \in \mathbb{R}^m$,

$$\begin{aligned} y^*\mathbf{\Psi}_{m,\sigma}y =& \frac{1}{NT}\sum_{j,\ell=1}^{m} y_j y_\ell \sum_{i,k=1}^{N}\mathbb{E}\Bigg(\left(\int_0^T \varphi_j(X_s^i)(\sigma(X_s^i)dB_s^i + \gamma(X_s^i)d\mathfrak{Z}_s^i)\right) \\ & \times\left(\int_0^T \varphi_\ell(X_s^k)(\sigma(X_s^k)dB_s^k + \gamma(X_s^k)d\mathfrak{Z}_s^k)\right)\Bigg) \\ =& \frac{1}{NT}\mathbb{E}\left[\left(\sum_{i=1}^{N}\int_0^T \tau_y(X_s^i)(\sigma(X_s^i)dB_s^i + \gamma(X_s^i)d\mathfrak{Z}_s^i)\right)^2\right] \geqslant 0 \quad \text{with} \quad \tau_y(.) := \sum_{j=1}^{m} y_j\varphi_j(.). \end{aligned}$$

Moreover, by the isometry property of Itô's integral (with respect to B), by the isometry type property of the stochastic integral with respect to $\mathfrak{Z}$, and since σ and γ are bounded, for every $j \in \{1, \dots, m\}$,

$$\begin{aligned} y^*\mathbf{\Psi}_{m,\sigma}y =& \frac{1}{T}\mathbb{E}\left[\left(\int_0^T \tau_y(X_s)\sigma(X_s)dB_s\right)^2\right] + \frac{1}{T}\mathbb{E}\left[\left(\int_0^T \tau_y(X_s)\gamma(X_s)d\mathfrak{Z}_s\right)^2\right] \\ =& \frac{1}{T}\int_0^T \mathbb{E}(\tau_y(X_s)^2\sigma(X_s)^2)ds + \frac{\lambda\mathfrak{c}_{\zeta^2}}{T}\int_0^T \mathbb{E}(\tau_y(X_s)^2\gamma(X_s)^2)ds \qquad (5) \\ \leqslant& (\|\sigma\|_\infty^2 + \lambda\mathfrak{c}_{\zeta^2}\|\gamma\|_\infty^2)\int_{-\infty}^{\infty}\left(\sum_{j=1}^{m} y_j\varphi_j(x)\right)^2 f_T(x)dx = (\|\sigma\|_\infty^2 + \lambda\mathfrak{c}_{\zeta^2}\|\gamma\|_\infty^2)\|\mathbf{\Psi}_m^{1/2}y\|_{2,m}^2. \end{aligned}$$

Therefore, since $\mathbf{\Psi}_{m,\sigma}$ is positive semidefinite, and by Inequality (),

$$\begin{aligned}\mathrm{trace}(\mathbf{\Psi}_m^{-1/2}\mathbf{\Psi}_{m,\sigma}\mathbf{\Psi}_m^{-1/2}) \leqslant & m\|\mathbf{\Psi}_m^{-1/2}\mathbf{\Psi}_{m,\sigma}\mathbf{\Psi}_m^{-1/2}\|_{\mathrm{op}}\\ =& m\cdot\sup\{y^*\mathbf{\Psi}_{m,\sigma}y; y\in\mathbb{R}^m \text{ and } \|\mathbf{\Psi}_m^{1/2}y\|_{2,m}=1\}\\ \leqslant&(\|\sigma\|_\infty^2+\lambda\mathfrak{c}_{\zeta^2}\|\gamma\|_\infty^2)m.\end{aligned}$$

□

A.2 Proof of Theorem 1

The proof of Theorem 1 relies on the two following lemmas.

Lemma 2 *There exists a constant* $\mathfrak{c}_2 > 0$, *not depending on m and N, such that*

$$\mathbb{E}(|\widehat{\mathbf{E}}_m^*\widehat{\mathbf{E}}_m|^2) \leqslant \mathfrak{c}_2\frac{mL(m)^2}{N^2}.$$

Lemma 3 *Consider the event*

$$\Omega_m := \left\{\sup_{\tau\in\mathcal{S}_m}\left|\frac{\|\tau\|_N^2}{\|\tau\|_{f_T}^2}-1\right|\leqslant\frac{1}{2}\right\}.$$

Under Assumptions 1, there exists a constant $\mathfrak{c}_3 > 0$, *not depending on m and N, such that*

$$\mathbb{P}(\Omega_m^c)\leqslant\frac{\mathfrak{c}_3}{N^7} \quad\text{and}\quad \mathbb{P}(\Lambda_m^c)\leqslant\frac{\mathfrak{c}_3}{N^7}.$$

The proof of Lemma 2 is postponed to Subsubsection a.2.2, and the proof of Lemma 3 remains the same as the proof of Comte and Genon-Catalot (2020b), Lemma 6.1, because $(B^1,Z^1),\dots,(B^N,Z^N)$ are independent.

A.2.1 Steps of the proof

First of all,

$$\begin{aligned}\|\widetilde{b}_m-b_I\|_N^2 =& \|b_I\|_N^2\mathbf{1}_{\Lambda_m^c}+\|\widehat{b}_m-b_I\|_N^2\mathbf{1}_{\Lambda_m}\\ =&\|b_I\|_N^2\mathbf{1}_{\Lambda_m^c}+\|\widehat{b}_m-b_I\|_N^2\mathbf{1}_{\Lambda_m\cap\Omega_m}+\|\widehat{b}_m-b_I\|_N^2\mathbf{1}_{\Lambda_m\cap\Omega_m^c} =: U_1+U_2+U_3.\end{aligned}$$

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

- **Bound on** $\mathbb{E}(U_1)$. By Cauchy-Schwarz's inequality,

$$\mathbb{E}(U_1) \leqslant \mathbb{E}(\|b_I\|_N^4)^{1/2}\mathbb{P}(\Lambda_m^c)^{1/2} \leqslant \mathbb{E}\left(\frac{1}{T}\int_0^T b_I(X_t)^4 dt\right)^{1/2}\mathbb{P}(\Lambda_m^c)^{1/2}$$
$$\leqslant \mathfrak{c}_1\mathbb{P}(\Lambda_m^c)^{1/2} < \infty \quad \text{with} \quad \mathfrak{c}_1 = \left(\int_{-\infty}^{\infty} b_I(x)^4 f_T(x)dx\right)^{1/2} < \infty.$$

- **Bound on** $\mathbb{E}(U_2)$. Let $\Pi_{N,m}(.)$ be the orthogonal projection from $\mathbb{L}^2(I, f_T(x)dx)$ onto $\mathcal{S}_m$ with respect to the empirical scalar product $\langle .,. \rangle_N$. Then,

$$\|\widehat{b}_m - b_I\|_N^2 = \|\widehat{b}_m - \Pi_{N,m}(b_I)\|_N^2 + \min_{\tau\in\mathcal{S}_m}\|b_I - \tau\|_N^2. \tag{6}$$

As in the proof of Comte and Genon-Catalot (2020b), Proposition 2.1, on Ω_m,

$$\|\widehat{b}_m - \Pi_{N,m}(b_I)\|_N^2 = \widehat{\mathbf{E}}_m^*\widehat{\mathbf{\Psi}}_m^{-1}\widehat{\mathbf{E}}_m \leqslant 2\widehat{\mathbf{E}}_m^*\mathbf{\Psi}_m^{-1}\widehat{\mathbf{E}}_m.$$

So,

$$\mathbb{E}(\|\widehat{b}_m - \Pi_{N,m}(b_I)\|_N^2\mathbf{1}_{\Lambda_m\cap\Omega_m}) \leqslant 2\mathbb{E}\left(\sum_{j,\ell=1}^m [\widehat{\mathbf{E}}_m]_j[\widehat{\mathbf{E}}_m]_\ell[\mathbf{\Psi}_m^{-1}]_{j,\ell}\right)$$
$$= \frac{2}{NT}\sum_{j,\ell=1}^m [\mathbf{\Psi}_{m,\sigma}]_{j,\ell}[\mathbf{\Psi}_m^{-1}]_{j,\ell} = \frac{2}{NT}\mathrm{trace}(\mathbf{\Psi}_m^{-1/2}\mathbf{\Psi}_{m,\sigma}\mathbf{\Psi}_m^{-1/2}).$$

Then, by Equality (6) and Lemma 1,

$$\mathbb{E}(U_2) \leqslant \mathbb{E}\left(\min_{\tau\in\mathcal{S}_m}\|b_I - \tau\|_N^2\right) + \frac{2}{NT}\mathrm{trace}(\mathbf{\Psi}_m^{-1/2}\mathbf{\Psi}_{m,\sigma}\mathbf{\Psi}_m^{-1/2})$$
$$\leqslant \min_{\tau\in\mathcal{S}_m}\|b_I - \tau\|_{f_T}^2 + \frac{2m}{NT}(\|\sigma\|_\infty^2 + \lambda\mathfrak{c}_{\zeta^2}\|\gamma\|_\infty^2).$$

- **Bound on** $\mathbb{E}(U_3)$. Since

$$\|\widehat{b}_m - \Pi_{N,m}(b_I)\|_N^2 = \widehat{\mathbf{E}}_m^*\widehat{\mathbf{\Psi}}_m^{-1}\widehat{\mathbf{E}}_m,$$

by the definition of the event Λ_m, and by Lemma 2,

$$\mathbb{E}(\|\widehat{b}_m - \Pi_{N,m}(b_I)\|_N^2\mathbf{1}_{\Lambda_m\cap\Omega_m^c}) \leqslant \mathbb{E}(\|\widehat{\mathbf{\Psi}}_m^{-1}\|_{\mathrm{op}}|\widehat{\mathbf{E}}_m^*\widehat{\mathbf{E}}_m|\mathbf{1}_{\Lambda_m\cap\Omega_m^c})$$
$$\leqslant \frac{\mathfrak{c}_T}{L(m)}\cdot\frac{NT}{\log(NT)}\mathbb{E}(|\widehat{\mathbf{E}}_m^*\widehat{\mathbf{E}}_m|^2)^{1/2}\mathbb{P}(\Omega_m^c)^{1/2}$$
$$\leqslant \frac{\mathfrak{c}_2 m^{1/2}}{\log(NT)}\mathbb{P}(\Omega_m^c)^{1/2},$$

where the constant $\mathfrak{c}_2 > 0$ doesn't depend on m and N. Moreover,

$$\min_{\tau\in\mathcal{S}_m}\|\tau - b_I\|_N^2 \leqslant \|b_I\|_N^2 \quad \text{because} \quad 0 \in \mathcal{S}_m,$$

and then,

$$\begin{aligned}\|\widehat{b}_m - b_I\|_N^2 =& \|\widehat{b}_m - \Pi_{N,m}(b_I)\|_N^2 + \min_{\tau\in\mathcal{S}_m} \|\tau - b_I\|_N^2 \\ \leqslant& \|\widehat{b}_m - \Pi_{N,m}(b_I)\|_N^2 + \|b_I\|_N^2.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}(U_3) \leqslant& \mathbb{E}(\|\widehat{b}_m - \Pi_{N,m}(b_I)\|_N^2 \mathbf{1}_{\Lambda_m\cap\Omega_m^c}) + \mathbb{E}(\|b_I\|_N^2 \mathbf{1}_{\Lambda_m\cap\Omega_m^c}) \\ \leqslant& \frac{\mathfrak{c}_2 m^{1/2}}{\log(NT)}\mathbb{P}(\Omega_m^c)^{1/2} + \mathfrak{c}_1 \mathbb{P}(\Omega_m^c)^{1/2}.\end{aligned}$$

So,

$$\begin{aligned}\mathbb{E}(\|\widetilde{b}_m - b_I\|_N^2) \leqslant& \min_{\tau\in\mathcal{S}_m} \|b_I - \tau\|_{f_T}^2 + \frac{2m}{NT}(\|\sigma\|_\infty^2 + \lambda\mathfrak{c}_{\zeta^2}\|\gamma\|_\infty^2) \\ &+ \mathfrak{c}_2\frac{\sqrt{m\mathbb{P}(\Omega_m^c)}}{\log(NT)} + \mathfrak{c}_1(\mathbb{P}(\Lambda_m^c)^{1/2} + \mathbb{P}(\Omega_m^c)^{1/2}).\end{aligned}$$

Therefore, by Lemma 3, there exists a constant $\mathfrak{c}_3 > 0$, not depending on m and N, such that

$$\mathbb{E}(\|\widetilde{b}_m - b_I\|_N^2) \leqslant \min_{\tau\in\mathcal{S}_m} \|b_I - \tau\|_{f_T}^2 + \mathfrak{c}_3\left(\frac{m}{NT} + \frac{1}{N}\right). \qquad \square$$

A.2.2 Proof of Lemma 2

In the sequel, the quadratic variation of any piecewise continuous stochastic process $(\Gamma_t)_{t\in[0,T]}$ is denoted by $([\![\Gamma]\!]_t)_{t\in[0,T]}$. First of all, note that since B and Z are independent, for every $j \in \{1, \dots, m\}$,

$$\begin{aligned}\left[\!\!\left[\int_0^\cdot \varphi_j(X_s)(\sigma(X_s)dB_s + \gamma(X_s)d\mathfrak{Z}_s)\right]\!\!\right]_T =& \int_0^T \varphi_j(X_s)^2\sigma(X_s)^2 ds + \int_0^T \varphi_j(X_s)^2\gamma(X_s)^2 dZ_s^{(2)} \\ = \int_0^T \varphi_j(X_s)^2(\sigma(X_s)^2 + \mathfrak{c}_{\zeta^2}\lambda\gamma(X_s)^2)ds + \int_0^T \varphi_j(X_s)^2\gamma(X_s)^2 d\mathfrak{Z}_s^{(2)}&\end{aligned}$$

where, for every $t \in [0, T]$,

$$Z_t^{(2)} := [\![\mathfrak{Z}]\!]_t = \sum_{n=1}^{\nu_t} \zeta_n^2 \quad \text{and} \quad \mathfrak{Z}_t^{(2)} := Z_t^{(2)} - \mathfrak{c}_{\zeta^2}\lambda t.$$

By Jensen's inequality and Burkholder-Davis-Gundy's inequality (see Dellacherie and Meyer, 1980, p. 303), there exists a constant $\mathfrak{c}_1 > 0$, not depending on m and N, such that

$$\begin{aligned}\mathbb{E}(|\widehat{\mathbf{E}}_m^*\widehat{\mathbf{E}}_m|^2) &\leqslant m\sum_{j=1}^{m}\mathbb{E}([\widehat{\mathbf{E}}_m]_j^4) \leqslant \frac{\mathfrak{c}_1 m}{N^4T^4}\sum_{j=1}^{m}\mathbb{E}\left(\left[\left[\sum_{i=1}^{N}\int_0^{\cdot}\varphi_j(X_s^i)(\sigma(X_s^i)dB_s^i+\gamma(X_s^i)d\mathfrak{Z}_s^i)\right]\right]_T^2\right)\\
&\leqslant \frac{2\mathfrak{c}_1 m}{N^2T^4}\sum_{j=1}^{m}\mathbb{E}\left(\left[\left[\int_0^{\cdot}\varphi_j(X_s)(\sigma(X_s)dB_s+\gamma(X_s)d\mathfrak{Z}_s)\right]\right]_T^2\right)\\
&\leqslant \frac{4\mathfrak{c}_1 m}{N^2T^4}\sum_{j=1}^{m}\left[\mathbb{E}\left[\left(\int_0^T\varphi_j(X_s)^2(\sigma(X_s)^2+\mathfrak{c}_{\zeta^2}\lambda\gamma(X_s)^2)ds\right)^2\right]\right.\\
&\qquad\left.+\mathbb{E}\left[\left(\int_0^T\varphi_j(X_s)^2\gamma(X_s)^2d\mathfrak{Z}_s^{(2)}\right)^2\right]\right].\end{aligned}$$

By Jensen's inequality,

$$\begin{aligned}\left(\int_0^T\varphi_j(X_s)^2(\sigma(X_s)^2+\mathfrak{c}_{\zeta^2}\lambda\gamma(X_s)^2)ds\right)^2 &= T^2\left(\int_0^T\varphi_j(X_s)^2(\sigma(X_s)^2+\mathfrak{c}_{\zeta^2}\lambda\gamma(X_s)^2)\frac{ds}{T}\right)^2\\
&\leqslant T\int_0^T\varphi_j(X_s)^4(\sigma(X_s)^2+\mathfrak{c}_{\zeta^2}\lambda\gamma(X_s)^2)^2ds.\end{aligned}$$

Moreover, since $\|\mathbf{x}\|_{4,m}\leqslant\|\mathbf{x}\|_{2,m}$ for every $\mathbf{x}\in\mathbb{R}^d$,

$$\sup_{x\in I}\sum_{j=1}^{m}\varphi_j(x)^4=\sup_{x\in I}\|(\varphi_j(x))_j\|_{4,m}^4\leqslant\sup_{x\in I}\|(\varphi_j(x))_j\|_{2,m}^4\leqslant L(m)^2.$$

So, by applying twice the Fubini–Tonelli theorem,

$$\begin{aligned}&\sum_{j=1}^{m}\mathbb{E}\left[\left(\int_0^T\varphi_j(X_s)^2(\sigma(X_s)^2+\mathfrak{c}_{\zeta^2}\lambda\gamma(X_s)^2)ds\right)^2\right]\\
&\leqslant 2T\sum_{j=1}^{m}\int_0^T\mathbb{E}(\varphi_j(X_s)^4(\sigma(X_s)^4+\mathfrak{c}_{\zeta^2}^2\lambda^2\gamma(X_s)^4))ds\\
&=2T^2\int_{-\infty}^{\infty}\underbrace{\left(\sum_{j=1}^{m}\varphi_j(x)^4\right)}_{\leqslant L(m)^2}(\sigma(x)^4+\mathfrak{c}_{\zeta^2}^2\lambda^2\gamma(x)^4)f_T(x)dx,\end{aligned}$$

and by the isometry type property of the stochastic integral with respect to $\mathfrak{Z}^{(2)}$,

$$\begin{aligned}\sum_{j=1}^{m}\mathbb{E}\left[\left(\int_0^T\varphi_j(X_s)^2\gamma(X_s)^2d\mathfrak{Z}_s^{(2)}\right)^2\right] &= \lambda\mathfrak{c}_{\zeta^4}\sum_{j=1}^{m}\int_0^T\mathbb{E}(\varphi_j(X_s)^4\gamma(X_s)^4)ds\\
&\leqslant\lambda\mathfrak{c}_{\zeta^4}TL(m)^2\int_{-\infty}^{\infty}\gamma(x)^4f_T(x)dx.\end{aligned}$$

Therefore,

$$\mathbb{E}(|\widehat{\mathbf{E}}_m^*\widehat{\mathbf{E}}_m|^2) \leqslant \frac{\mathfrak{c}_2}{N^2T^2} mL(m)^2$$

with

$$\mathfrak{c}_2 = 8\mathfrak{c}_1\left(\int_{-\infty}^{\infty}(\sigma(x)^4 + \mathfrak{c}_{\zeta^2}^2\lambda^2\gamma(x)^4)f_T(x)dx + \lambda\mathfrak{c}_{\zeta^4}\int_{-\infty}^{\infty}\gamma(x)^4 f_T(x)dx\right). \qquad \square$$

A.3 Proof of Theorem 2

Let us consider the events

$$\Omega_N := \bigcap_{m\in\mathcal{M}_N^+}\Omega_m \quad \text{and} \quad \Xi_N := \{\mathcal{M}_N \subset \widehat{\mathcal{M}}_N \subset \mathcal{M}_N^+\},$$

where

$$\mathcal{M}_N^+ := \left\{m \in \{1,\dots,N_T\} : \mathfrak{c}_\varphi^2 m(\|\mathbf{\Psi}_m^{-1}\|_{\mathrm{op}}^2 \vee 1) \leqslant 4\mathfrak{d}_T\frac{NT}{\log(NT)}\right\},$$

and let us recall that

$$\mathfrak{d}_T = \min\left\{\frac{\mathfrak{c}_T}{2}, \frac{1}{64\mathfrak{c}_\varphi^2 T(\|f_T\|_\infty + \sqrt{\mathfrak{c}_T/2}/(3\mathfrak{c}_\varphi))}\right\}.$$

As a reminder, the sets $\widehat{\mathcal{M}}_N$ and $\mathcal{M}_N$ introduced in Sect. 4 are, respectively, defined by

$$\widehat{\mathcal{M}}_N = \left\{m \in \{1,\dots,N_T\} : \mathfrak{c}_\varphi^2 m(\|\widehat{\mathbf{\Psi}}_m^{-1}\|_{\mathrm{op}}^2 \vee 1) \leqslant \mathfrak{d}_T\frac{NT}{\log(NT)}\right\}$$

and

$$\mathcal{M}_N = \left\{m \in \{1,\dots,N_T\} : \mathfrak{c}_\varphi^2 m(\|\mathbf{\Psi}_m^{-1}\|_{\mathrm{op}}^2 \vee 1) \leqslant \frac{\mathfrak{d}_T}{4}\cdot\frac{NT}{\log(NT)}\right\}.$$

The proof of Theorem 2 relies on the three following lemmas.

Lemma 4 *Under Assumptions* 1, 2 *and* 3, *there exists a constant* $\mathfrak{c}_4 > 0$, *not depending on* N, *such that*

$$\mathbb{P}(\Xi_N^c) \leqslant \frac{\mathfrak{c}_4}{N^6}.$$

Lemma 5 (*Bernstein type inequality*) *Consider the empirical process*

$$\nu_N(\tau) := \frac{1}{NT}\sum_{i=1}^{N}\int_0^T \tau(X_s^i)(\sigma(X_s^i)dB_s^i + \gamma(X_s^i)d\mathfrak{Z}_s^i); \tau \in \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_{N_T}.$$

Under Assumption 3, for every $\xi, v > 0$,

$$\mathbb{P}(\nu_N(\tau) \geqslant \xi, \|\tau\|_N^2 \leqslant v^2) \leqslant \exp\left[-\frac{NT\xi^2}{4[\mathfrak{c}_5(\|\sigma\|_\infty^2 + \|\gamma\|_\infty^2)v^2 + \xi\|\tau\|_{\infty,I}\|\gamma\|_\infty]}\right]$$

with

$$\mathfrak{c}_6 = \frac{1}{2}\left[1 \vee \int_{-\infty}^{\infty} e^{\mathfrak{b}|z|/2}\pi_\lambda(dz)\right].$$

Lemma 6 *Under Assumptions 1 and 3, there exists a constant $\mathfrak{c}_7 > 0$, not depending on N, such that for every $m \in \mathcal{M}_N$,*

$$\mathbb{E}\left[\left(\left[\sup_{\tau\in\mathcal{B}_{m,m'}} \nu_n(\tau)\right]^2 - p(m,\widehat{m})\right)_+ \mathbf{1}_{\Xi_N\cap\Omega_N}\right] \leqslant \frac{\mathfrak{c}_8}{NT}$$

where, for every $m' \in \mathcal{M}_N$,

$$\mathcal{B}_{m,m'} := \left\{\tau \in \mathcal{S}_{m\wedge m'} : \|\tau\|_{f_T} = 1\right\} \quad \text{and} \quad p(m,m') := \frac{\mathfrak{c}_{\text{cal}}}{8}\cdot\frac{m \vee m'}{NT}.$$

The proof of Lemma 5 is postponed to Subsubsection A.3.2. Lemma 6 is a consequence of Lemma 5 thanks to the $\mathbb{L}_{f_T}^2$-$\mathbb{L}^\infty$ chaining technique (see Comte, 2001, Proposition 4). Finally, the proof of Lemma 4 remains the same as the proof of Comte and Genon-Catalot (2020b), Eq. (6.17), because $(B^1, \mathfrak{Z}^1), \ldots, (B^N, \mathfrak{Z}^N)$ are independent.

A.3.1 Steps of the proof

First of all,

$$\begin{aligned}\|\widehat{b}_{\widehat{m}} - b_I\|_N^2 &= \|\widehat{b}_{\widehat{m}} - b_I\|_N^2\mathbf{1}_{\Xi_N^c} + \|\widehat{b}_{\widehat{m}} - b_I\|_N^2\mathbf{1}_{\Xi_N}\\ &=: U_1 + U_2.\end{aligned} \tag{7}$$

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

- **Bound on** $\mathbb{E}(U_1)$. Since

  $$\|\widehat{b}_{\widehat{m}} - \Pi_{N,\widehat{m}}(b_I)\|_N^2 = \widehat{\mathbf{E}}_{\widehat{m}}^*\widehat{\mathbf{\Psi}}_{\widehat{m}}^{-1}\widehat{\mathbf{E}}_{\widehat{m}},$$

  by the definition of $\widehat{\mathcal{M}}_N$, and by Lemma 2,

$$\begin{aligned}\mathbb{E}(\|\widehat{b}_{\widehat{m}} - \Pi_{N,\widehat{m}}(b_I)\|_N^2 \mathbf{1}_{\Xi_N^c}) \leqslant & \mathbb{E}(\|\widehat{\mathbf{\Psi}}_{\widehat{m}}^{-1}\|_{\mathrm{op}} |\widehat{\mathbf{E}}_{NT}^* \widehat{\mathbf{E}}_{NT}| \mathbf{1}_{\Xi_N^c})\\ & \leqslant \left[\mathfrak{d}_T \frac{NT}{\log(NT)}\right]^{1/2} \mathbb{E}(|\widehat{\mathbf{E}}_{NT}^* \widehat{\mathbf{E}}_{NT}|^2)^{1/2} \mathbb{P}(\Xi_N^c)^{1/2} \leqslant \frac{\mathfrak{c}_1 N}{\log(NT)} \mathbb{P}(\Xi_N^c)^{1/2},\end{aligned}$$

where the constant $\mathfrak{c}_1 > 0$ does not depend on N. Then,

$$\begin{aligned}\mathbb{E}(U_1) \leqslant & \mathbb{E}(\|\widehat{b}_{\widehat{m}} - \Pi_{N,\widehat{m}}(b_I)\|_N^2 \mathbf{1}_{\Xi_N^c}) + \mathbb{E}(\|b_I\|_N^2 \mathbf{1}_{\Xi_N^c})\\ \leqslant & \frac{\mathfrak{c}_1 N}{\log(NT)} \mathbb{P}(\Xi_N^c)^{1/2} + \mathfrak{c}_2 \mathbb{P}(\Xi_N^c)^{1/2}\end{aligned}$$

with

$$\mathfrak{c}_2 = \left(\int_{-\infty}^{\infty} b_I(x)^4 f_T(x) dx\right)^{1/2}.$$

So, by Lemma 4, there exists a constant $\mathfrak{c}_3 > 0$, not depending on N, such that

$$\mathbb{E}(U_1) \leqslant \frac{\mathfrak{c}_3}{N}.$$

- **Bound on** $\mathbb{E}(U_2)$. Note that

$$\begin{aligned}U_2 = & \|\widehat{b}_{\widehat{m}} - b_I\|_N^2 \mathbf{1}_{\Xi_N \cap \Omega_N^c} + \|\widehat{b}_{\widehat{m}} - b_I\|_N^2 \mathbf{1}_{\Xi_N \cap \Omega_N}\\ =: & U_{2,1} + U_{2,2}.\end{aligned}$$

On the one hand, by Lemma 3, there exists a constant $\mathfrak{c}_4 > 0$, not depending on N, such that

$$\mathbb{P}(\Xi_N \cap \Omega_N^c) \leqslant \sum_{m \in \mathcal{M}_N^+} \mathbb{P}(\Omega_m^c) \leqslant \frac{\mathfrak{c}_4}{N^6}.$$

Then, as for $\mathbb{E}(U_1)$, there exists a constant $\mathfrak{c}_5 > 0$, not depending on N, such that

$$\mathbb{E}(U_{2,1}) \leqslant \frac{\mathfrak{c}_5}{N}.$$

On the other hand,

$$\gamma_N(\tau') - \gamma_N(\tau) = \|\tau' - b\|_N^2 - \|\tau - b\|_N^2 - 2\nu_N(\tau' - \tau)$$

for every $\tau, \tau' \in \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_{N_T}$. Moreover, since

$$\widehat{m} = \arg\min_{m \in \widehat{\mathcal{M}}_N} \{-\|\widehat{b}_m\|_N^2 + \mathrm{pen}(m)\} = \arg\min_{m \in \widehat{\mathcal{M}}_N} \{\gamma_N(\widehat{b}_m) + \mathrm{pen}(m)\},$$

for every $m \in \widehat{\mathcal{M}}_N$,

$$\gamma_N(\widehat{b}_{\widehat{m}}) + \text{pen}(\widehat{m}) \leqslant \gamma_N(\widehat{b}_m) + \text{pen}(m). \tag{8}$$

On the event $\Xi_N = \{\mathcal{M}_N \subset \widehat{\mathcal{M}}_N \subset \mathcal{M}_N^+\}$, Inequality (8) remains true for every $m \in \mathcal{M}_N$. Then, on Ξ_N, for any $m \in \mathcal{M}_N$, since $\mathcal{S}_m + \mathcal{S}_{\widehat{m}} \subset \mathcal{S}_{m\vee\widehat{m}}$ under Assumption 2,

$$\begin{aligned}\|\widehat{b}_{\widehat{m}} - b_I\|_N^2 \leqslant & \|\widehat{b}_m - b_I\|_N^2 + 2\|\widehat{b}_{\widehat{m}} - \widehat{b}_m\|_{f_T}\nu_N\left(\frac{\widehat{b}_{\widehat{m}} - \widehat{b}_m}{\|\widehat{b}_{\widehat{m}} - \widehat{b}_m\|_{f_T}}\right) + \text{pen}(m) - \text{pen}(\widehat{m})\\ \leqslant & \|\widehat{b}_m - b_I\|_N^2 + \frac{1}{8}\|\widehat{b}_{\widehat{m}} - \widehat{b}_m\|_{f_T}^2\\ & + 8\left(\left[\sup_{\tau\in\mathcal{B}_{m,\widehat{m}}} |\nu_N(\tau)|\right]^2 - p(m,\widehat{m})\right)_+ + \text{pen}(m) + 8p(m,\widehat{m}) - \text{pen}(\widehat{m}).\end{aligned}$$

Since $\|.\|_{f_T}^2 \mathbf{1}_{\Omega_N} \leqslant 2\|.\|_N^2 \mathbf{1}_{\Omega_N}$ on $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_{\max(\mathcal{M}_N^+)}$, and since $8p(m,\widehat{m}) \leqslant \text{pen}(m) + \text{pen}(\widehat{m})$, on $\Xi_N \cap \Omega_N$,

$$\|\widehat{b}_{\widehat{m}} - b_I\|_N^2 \leqslant 3\|\widehat{b}_m - b_I\|_N^2 + 4\text{pen}(m) + 16\left(\left[\sup_\tau \in \mathcal{B}_m, \widehat{m}|\nu_N(\tau)|\right]^2 - p(m,\widehat{m})\right)_+.$$

So, by Lemma 6,

$$\begin{aligned}\mathbb{E}(U_{2,2}) \leqslant & \min_{m\in\mathcal{M}_N}\{\mathbb{E}(3\|\widehat{b}_m - b_I\|_N^2 \mathbf{1}_{\Xi_N}) + 4\text{pen}(m)\} + \frac{16\mathfrak{c}_6}{NT}\\ \leqslant & \mathfrak{c}_6 \min_{m\in\mathcal{M}_N}\left\{\inf_{\tau\in\mathcal{S}_m}\|\tau - b_I\|_{f_T}^2 + \frac{m}{NT}\right\} + \frac{\mathfrak{c}_6}{N}\end{aligned}$$

where $\mathfrak{c}_6 > 0$ is a deterministic constant not depending on N.

□

A.2.3.2 Proof of Lemma 5

Consider $\tau \in \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_{N_T}$ and, for any $i \in \{1,\dots,N\}$, let $M^i(\tau) = (M^i(\tau)_t)_{t\in[0,T]}$ be the martingale defined by

$$M^i(\tau)_t := \int_0^t \tau(X_s^i)(\sigma(X_s^i)dB_s^i + \gamma(X_s^i)d\mathfrak{Z}_s^i);\ \forall t \in [0,T].$$

Moreover, for every $\varepsilon > 0$, consider

$$Y_\varepsilon^i(\tau) := \varepsilon M^i(\tau) - A_\varepsilon^i(\tau) - B_\varepsilon^i(\tau),$$

where $A_\varepsilon^i(\tau) = (A_\varepsilon^i(\tau)_t)_{t\in[0,T]}$ and $B_\varepsilon^i(\tau) = (B_\varepsilon^i(\tau)_t)_{t\in[0,T]}$ are the stochastic processes defined by

$$A_\varepsilon^i(\tau)_t := \frac{\varepsilon^2}{2}\int_0^t \tau(X_s^i)^2\sigma(X_s^i)^2 ds$$
$$\text{and} \quad B_\varepsilon^i(\tau)_t := \int_0^t \left[\int_{-\infty}^{\infty}(e^{\varepsilon z\tau(X_s^i)\gamma(X_s^i)} - \varepsilon z\tau(X_s^i)\gamma(X_s^i) - 1)\pi_\lambda(dz)\right]ds$$

for every $t \in [0,T]$. The proof is dissected in three steps.

Step 1. Note that for any $i \in \{1,\dots,N\}$ and $t \in [0,T]$,

$$|\tau(X_t^i)\gamma(X_t^i)| \leqslant \|\tau\|_{\infty,I}\|\gamma\|_\infty$$

and then, by Assumption 3,

$$\mathbb{E}\left(\int_0^t \int_{|z|>1} |e^{\varepsilon z\tau(X_s^i)\gamma(X_s^i)} - 1|\pi_\lambda(dz)ds\right) < \infty$$

for any $\varepsilon \in (0,\varepsilon^*)$ with $\varepsilon^* = (\mathfrak{b} \wedge 1)/(2\|\tau\|_{\infty,I}\|\gamma\|_\infty)$. So, $(\exp(Y_\varepsilon^i(\tau)_t))_{t\in[0,T]}$ is a local martingale by Applebaum (2009), Corollary 5.2.2. In other words, there exists an increasing sequence of stopping times $(T_n^i)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} T_n^i = \infty$ a.s. and $(\exp(Y_\varepsilon^i(\tau)_{t\wedge T_n^i})_{t\in[0,T]}$ is a martingale. Therefore, by Lebesgue's theorem and Markov's inequality, for every $\rho > 0$, the stochastic process $Y_{N,\varepsilon}(\tau) := Y_\varepsilon^1(\tau) + \cdots + Y_\varepsilon^N(\tau)$ satisfies

$$\begin{aligned}\mathbb{P}(e^{Y_{N,\varepsilon}(\tau)_T} > \rho) &= \lim_{n\to\infty}\mathbb{P}\left(\exp\left[\sum_{i=1}^N Y_\varepsilon^i(\tau)_{T\wedge T_n^i}\right] > \rho\right)\\ &\leqslant \frac{1}{\rho}\lim_{n\to\infty}\mathbb{E}(\exp(Y_\varepsilon^1(\tau)_{T\wedge T_n^1}))^N = \frac{1}{\rho}\mathbb{E}(\exp(Y_\varepsilon^1(\tau)_0))^N = \frac{1}{\rho}.\end{aligned}$$

Step 2. For any $\varepsilon \in (0,\varepsilon^*)$ and $t \in [0,T]$, let us find suitable bounds on

$$A_{N,\varepsilon}(\tau)_t := \sum_{i=1}^N A_\varepsilon^i(\tau)_t \quad \text{and} \quad B_{N,\varepsilon}(\tau)_t := \sum_{i=1}^N B_\varepsilon^i(\tau)_t.$$

On the one hand,

$$A_{N,\varepsilon}(\tau)_t \leqslant \frac{\varepsilon^2\|\sigma\|_\infty^2}{2}\sum_{i=1}^N\int_0^t \tau(X_s^i)^2 ds \leqslant \frac{\varepsilon^2\|\sigma\|_\infty^2\|\tau\|_N^2 NT}{2}. \tag{9}$$

On the other hand, for every $\beta \in (-\mathfrak{b}/2,\mathfrak{b}/2)$, by Taylor's formula and Assumption 3,

$$\begin{aligned}\int_{-\infty}^{\infty}(e^{\beta z} - \beta z - 1)\pi_\lambda(dz) &= \beta^2\int_{-\infty}^{\infty}\left(\int_0^1 (1-\theta)e^{\theta\beta z}d\theta\right)\pi_\lambda(dz)\\ &\leqslant \frac{\mathfrak{c}_1}{2}\beta^2 \quad \text{with} \quad \mathfrak{c}_1 = \int_{-\infty}^{\infty} e^{\mathfrak{b}|z|/2}\pi_\lambda(dz) < \infty.\end{aligned}$$

Since $\varepsilon \in (0, \varepsilon^*)$, one can take $\beta = \varepsilon\tau(X_s^i)\gamma(X_s^i)$ for any $s \in [0, t]$ and $i \in \{1, \dots, N\}$, and then

$$B_{N,\varepsilon}(\tau)_t \leqslant \frac{\mathfrak{c}_1\varepsilon^2}{2}\sum_{i=1}^{N}\int_0^t \tau(X_s^i)^2\gamma(X_s^i)^2 ds \leqslant \frac{\mathfrak{c}_1\varepsilon^2\|\gamma\|_\infty^2\|\tau\|_N^2 NT}{2}. \tag{10}$$

Therefore, Inequalities (9) and (10) lead to

$$A_{N,\varepsilon}(\tau)_t + B_{N,\varepsilon}(\tau)_t \leqslant \mathfrak{c}_2\varepsilon^2(\|\sigma\|_\infty^2 + \|\gamma\|_\infty^2)\|\tau\|_N^2 NT \quad \text{with} \quad \mathfrak{c}_2 = \frac{1}{2}(1 \vee \mathfrak{c}_1).$$

Step 3 (conclusion). Consider $M_N(\tau) := M^1(\tau) + \cdots + M^N(\tau)$. For any $\varepsilon \in (0, \varepsilon^*)$ and $\xi, v > 0$, thanks to Step 2,

$$\begin{aligned}
\mathbb{P}(&\nu_N(\tau) \geqslant \xi, \|\tau\|_N^2 \leqslant v^2) \\
&\leqslant \mathbb{P}(e^{\varepsilon M_N(\tau)_T} \geqslant e^{NT\varepsilon\xi}, A_{N,\varepsilon}(\tau)_T + B_{N,\varepsilon}(\tau)_T \\
&\leqslant \mathfrak{c}_2\varepsilon^2(\|\sigma\|_\infty^2 + \|\gamma\|_\infty^2)NTv^2) \leqslant \mathbb{P}(e^{Y_{N,\varepsilon}(\tau)_T} \\
&\geqslant \exp(NT\varepsilon\xi - \mathfrak{c}_2\varepsilon^2(\|\sigma\|_\infty^2 + \|\gamma\|_\infty^2)NTv^2)).
\end{aligned}$$

Moreover, taking

$$\varepsilon = \frac{\xi}{2\mathfrak{c}_2(\|\sigma\|_\infty^2 + \|\gamma\|_\infty^2)v^2 + \xi/\varepsilon^*} < \varepsilon^*$$

leads to

$$\begin{aligned}
NT\varepsilon\xi - \mathfrak{c}_2\varepsilon^2(\|\sigma\|_\infty^2 + \|\gamma\|_\infty^2)NTv^2 &= \frac{NT\xi^2[\mathfrak{c}_2(\|\sigma\|_\infty^2 + \|\gamma\|_\infty^2)v^2 + \xi/\varepsilon^*]}{[2\mathfrak{c}_2(\|\sigma\|_\infty^2 + \|\gamma\|_\infty^2)v^2 + \xi/\varepsilon^*]^2} \\
&\geqslant \frac{NT\xi^2}{4[\mathfrak{c}_2(\|\sigma\|_\infty^2 + \|\gamma\|_\infty^2)v^2 + \xi/\varepsilon^*]}.
\end{aligned}$$

Therefore, by Step 1,

$$\mathbb{P}(\nu_N(\tau) \geqslant \xi, \|\tau\|_N^2 \leqslant v^2) \leqslant \exp\left(-\frac{NT\xi^2}{4[\mathfrak{c}_2(\|\sigma\|_\infty^2 + \|\gamma\|_\infty^2)v^2 + \xi\|\tau\|_{\infty,I}\|\gamma\|_\infty]}\right).$$

□

Appendix B

See Table 1 and Figs. (1, 2 and 3)

Table 1 Means and StD of the MISE of $\widehat{b}_{\widehat{m}}$ (100 repetitions)

	Model 1	Model 2	Model 3
Mean MISE	0.1251	0.1469	0.1825
StD MISE	0.0950	0.1688	0.1928

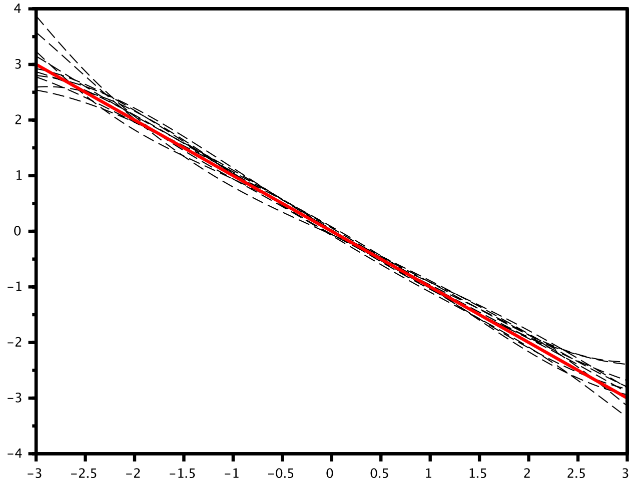

Fig. 1 Plots of b and of 10 adaptive estimations for Model 1 ($\overline{\widehat{m}} = 5.3$)

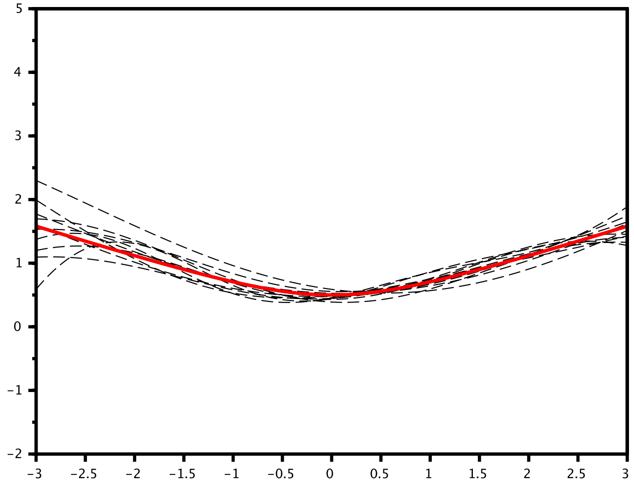

Fig. 2 Plots of b and of 10 adaptive estimations for Model 2 ($\overline{\widehat{m}} = 4.2$)

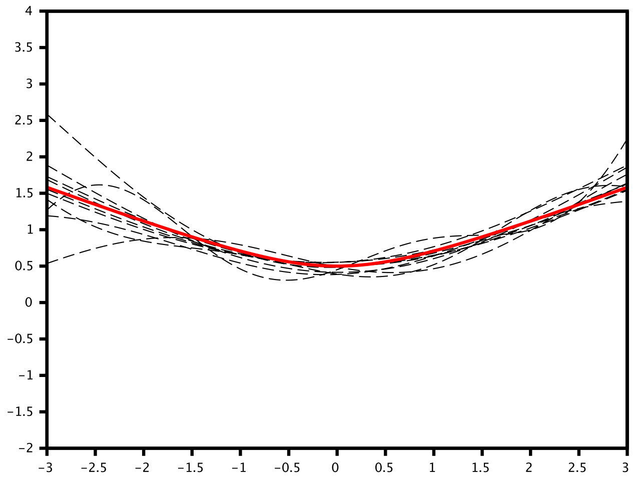

Fig. 3 Plots of b and of 10 adaptive estimations for Model 3 ($\overline{\widehat{m}} = 4.1$)

References

Abramowitz, M., Stegun I.A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. USA: Dover Publications.

Amorino, C., Gloter, A. (2021). Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes. *Journal of Statistical Planning and Inference*, *213*, 106–129.

Amorino, C., Dion-Blanc, C., Gloter, A., Lemler, S. (2022). On the nonparametric inference of coefficients of self-exciting jump-diffusion. *Electronic Journal of Statistics*, *16*, 3212–3277.

Applebaum, C. (2009). *Lévy Processes and Stochastic Calculus*. Cambridge: Cambridge University Press.

Belomestny, D., Comte, F., Genon-Catalot, V. (2019). Sobolev-Hermite versus Sobolev nonparametric density estimation on $\mathbb{R}$. *Annals of the Institute of Statistical Mathematics*, *71*(1), 29–62.

Bichteler, K., Jacod, J. (1983). Calcul de Malliavin pour les diffusions avec sauts : existence d'une densité dans le cas unidimensionnel. *Séminaire de Probabilités*, *17*, 132–157.

Chen, Z.Q., Hu, E., Xie, L., Zhang, X. (2017). Heat Kernels for nonsymmetric diffusion operators with jumps. *Journal of Differential Equations*, *263*(10), 6576–6634.

Clément, E., Gloter, A. (2019). Estimating Functions for SDE Driven by Stable Lévy Processes. *Annales de l'Institut Henri Poincaré (B)*, *55*(3), 1316–1348.

Clément, E., Gloter, A. (2020). Joint estimation for sde driven by locally stable Lévy processes. *Electronic Journal of Statistics*, *14*(2), 2922–2956.

Cohen, A., Davenport, M.A., Leviatan, D. (2013) On the stability and accuracy of least squares approximations. *Foundations of Computational Mathematics*, *13*, 819–834.

Comte, F. (2001). Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Bernoulli*, *7*(2), 267–298.

Comte, F., Genon-Catalot, V. (2020). Regression function estimation as a partly inverse problem. *Annals of the Institute of Statistical Mathematics*, *72*(4), 1023–1054.

Comte, F., Genon-Catalot, V. (2020). Nonparametric drift estimation for i.i.d. paths of stochastic differential equations. *The Annals of Statistics*, *48*(6), 3336–3365.

Delattre, M., Genon-Catalot, V., Samson, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics*, *40*, 322–343.

Dellacherie, C.A., Meyer, P.A. (1980). *Probabilités et potentiel : théorie des martingales*. Paris: Hermann.

Denis, C., Dion-Blanc, C., Martinez M. (2021). A ridge estimator of the drift from discrete repeated observations of the solution of a stochastic differential equation. *Bernoulli*, *27*, 2675–2713.
DeVore, R.A., Lorentz, G.G. (1993). *Constructive Approximation*. Berlin, Heidelberg: Springer.
Ditlevsen, S., De Gaetano, A. (2005). Mixed effects in stochastic differential equation models. *REVSTAT*, *3*, 137–153.
Dussap, F. (2021). Anisotropic multivariate deconvolution using projection on the laguerre basis. *Journal of Statistical Planning and Inference*, *215*, 23–46.
Gloter, A., Loukianova, D., Mai, H. (2018). Jump filtering and efficient drift estimation for Lévy-Driven SDEs. *The Annals of Statistics*, *46*(4), 1445–1480.
Marie, N., Rosier, A. (2023). Nadaraya-Watson Estimator for I.I.D paths of diffusion processes. *Scandinavian Journal of Statistics*, *50*(2), 589–637.
Picchini, U., Ditlevsen, S. (2011) Practical estimation of high dimensional stochastic differential mixed-effects models. *Computational Statistics and Data Analysis*, *55*, 1426–1444.
Schmisser, E. (2014). Non-parametric adaptive estimation of the drift for a jump diffusion process. *Stochastic Processes and their Applications*, *124*, 883–914.