Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 124 (2014) 883–914

stochastic processes and their applications

www.elsevier.com/locate/spa

Non-parametric adaptive estimation of the drift for a jump diffusion process

Émeline Schmisser*

Université Lille 1, Laboratoire Paul Painlevé, France

Received 12 April 2013; received in revised form 24 September 2013; accepted 24 September 2013
Available online 7 October 2013

Abstract

In this article, we consider a jump diffusion process $(X_t)_{t\geq 0}$ observed at discrete times $t = 0, \Delta, \ldots, n\Delta$. The sampling interval Δ tends to 0 and $n\Delta$ tends to infinity. We assume that $(X_t)_{t\geq 0}$ is ergodic, strictly stationary and exponentially β-mixing. We use a penalised least-square approach to compute two adaptive estimators of the drift function b. We provide bounds for the risks of the two estimators.

Keywords: Jump diffusions; Nonparametric estimation; Drift estimation; Model selection

1. Introduction

We consider a general diffusion with jumps:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \xi(X_{t^-})dL_t \quad \text{and} \quad X_0 = \eta \tag{1}$$

where L_t is a centred pure jump Levy process:

$$dL_t = \int_{z\in\mathbb{R}} z\,(\mu(dt, dz) - dt\nu(dz))$$

with μ a random Poisson measure with intensity measure $\nu(dz)dt$ such that $\int_{z\in\mathbb{R}} z^2\nu(dz) < \infty$. The compensated Poisson measure $\tilde{\mu}$ is defined by $\tilde{\mu}(dt, dz) = \mu(dt, dz) - \nu(dz)dt$. The random variable η is independent of $(W_t, L_t)_{t\geq 0}$. Moreover, $(W_t)_{t\geq 0}$ and $(L_t)_{t\geq 0}$ are independent.

* Tel.: +33 320436764.
E-mail address: emeline.schmisser@math.univ-lille1.fr.

0304-4149/$ - see front matter
http://dx.doi.org/10.1016/j.spa.2013.09.012

This process is observed with high frequency (at times $t = 0, \Delta, \dots, n\Delta$ where, as n tends to infinity, the sampling interval $\Delta \to 0$ and the time of observation $n\Delta \to \infty$). It is assumed to be ergodic, stationary and exponentially β-mixing (see [15] for sufficient conditions). Our aim is to construct a non-parametric estimator of b on a compact set A.

The non-parametric estimation of b and σ for a diffusion process observed with high-frequency is well-known (see for instance [10,6]). Diffusion processes with jumps are used in various fields, for instance in finance, for modelling the growth of a population, in hydrology, in medical science, ..., but there exist few results for the non-parametric estimation of b and σ. Mai [13] and Shimizu and Yoshida [21] construct maximum-likelihood estimators of parameters of b. Their estimators reach the standard rate of convergence, $\sqrt{n\Delta}$. Shimizu [20] and Mancini and Renò [14] use a kernel estimator to obtain non parametric threshold estimators of σ. Mancini and Renò [14] also construct a non-parametric truncated estimator of b, but only when L_t is a compound Poisson process. To our knowledge, minimax rates of convergences for non-parametric estimators of b, σ or ξ for jump–diffusion processes are not available in the literature (see [10] or [9] for rates of convergence for diffusion processes).

In this paper, we use model selection to construct two non-parametric estimators of b under the asymptotic framework $\Delta \to 0$ and $n\Delta \to \infty$. This method was introduced by Birgé and Massart [4].

First, we introduce a sequence of linear subspaces $S_m \subseteq L^2(A)$ and, for each m, we construct an estimator $\hat{b}_m$ of b by minimising on S_m the contrast function:

$$\gamma_n(t) = \frac{1}{n}\sum_{k=1}^{n}(Y_{k\Delta} - t(X_{k\Delta}))^2 \quad \text{where } Y_{k\Delta} = \frac{X_{(k+1)\Delta} - X_{k\Delta}}{\Delta}.$$

We obtain a collection of estimators of the drift function b and we bound their risks (Theorem 2). Then, we introduce a penalty function to select the "best" dimension m and we deduce an adaptive estimator $\hat{b}_{\hat{m}}$. Under the assumption that ν is sub-exponential, that is if there exist two positive constants C, λ such that, for z large enough, $\nu([-z, z]^c) \le Ce^{-\lambda z}$, the risk bound of $\hat{b}_{\hat{m}}$ is exactly the same as for a diffusion without jumps (Theorem 4) (see [6] or [10]).

In a second part, we do not assume that ν is sub-exponential and we construct a truncated estimator $\tilde{b}_m$ of b. We minimise the contrast function

$$\tilde{\gamma}_n(t) = \frac{1}{n}\sum_{k=1}^{n}\left(Y_{k\Delta}\mathbb{1}_{|Y_{k\Delta}|\le C_\Delta} - t(X_{k\Delta})\right)^2 \quad \text{where } C_\Delta \propto \sqrt{\Delta}\ln(n)$$

in order to obtain a new estimator $\tilde{b}_m$. As in the first part, we introduce a penalty function to obtain an adaptive estimator $\tilde{b}_{\tilde{m}}$. The risk bound of this adaptive estimator depends on the Blumenthal–Getoor index of ν (Theorems 7 and 10).

In Section 2, we present the model and its assumptions. In Sections 3 and 4, we construct the estimators and bound their risks. Some simulations are presented in Section 5. Proofs are gathered in Section 6.

2. Assumptions

2.1. Assumptions on the model

We consider the following assumptions.

A1. The functions b, σ and ξ are Lipschitz.

A2. 1. The function σ is bounded from below and above:

$$\exists \sigma_0, \sigma_1, \ \forall x \in \mathbb{R}, \quad 0 < \sigma_1 \leq \sigma(x) \leq \sigma_0.$$

2. The function ξ is bounded: $\exists \xi_0, \ \forall x \in \mathbb{R}, \ 0 \leq \xi(x) \leq \xi_0$.
3. The drift function b is elastic: there exists a constant M such that, for any $x \in \mathbb{R}$, $|x| > M$: $xb(x) \lesssim -|x|^2$.
4. The Lévy measure ν satisfies:

$$\nu(\{0\}) = 0, \qquad \int_{-\infty}^{\infty} z^2 \nu(dz) = 1 \quad \text{and} \quad \int_{-\infty}^{\infty} z^4 \nu(dz) < \infty.$$

Under Assumption A1, the stochastic differential equation (1) admits a unique strong solution. According to [15], under Assumptions A1 and A2, the process (X_t) admits a unique invariant probability ϖ and satisfies the ergodic theorem: for any measurable function g such that $\int |g(x)|\varpi(dx) < \infty$, when $T \to \infty$,

$$\frac{1}{T}\int_0^T g(X_s)ds \to \int g(x)\varpi(dx).$$

This distribution has moments of order 4. Moreover, Masuda [15] also ensures that under these assumptions, the process (X_t) is exponentially β-mixing. Furthermore, if there exist two constants c and n_0 such that, for any $x \in \mathbb{R}$, $\xi^2(x) \geq c(1 + |x|)^{-n_0}$, then Ishikawa and Kunita [11] ensure that a smooth transition density exists.

A3. 1. The stationary measure ϖ admits a density π which is bounded from below and above on the compact interval A:

$$\exists \pi_0, \pi_1, \ \forall x \in A, \quad 0 < \pi_1 \leq \pi(x) \leq \pi_0.$$

2. The process $(X_t)_{t\geq 0}$ is stationary ($\eta \sim \varpi(dx) = \pi(x)dx$).

The first part of this assumption is automatically satisfied if $\xi = 0$ (that is if $(X_t)_{t\geq 0}$ is a diffusion process). The following proposition is very useful for the proofs. It is derived from Result 11.

Proposition 1. *Under Assumptions A1– A3, for any $p \geq 1$, there exists a constant $c(p)$ such that, if $\int_{\mathbb{R}} z^{2p}\nu(dz) < \infty$:*

$$\mathbb{E}\left(\sup_{s\in[t,t+h]} (X_s - X_t)^{2p}\right) \leq c(p)h.$$

2.2. *Assumptions on the approximation spaces*

In order to construct an adaptive estimator of b, we use model selection: we compute a collection of estimators $\hat{b}_m$ of b by minimising a contrast function $\gamma_n(t)$ on a vectorial subspace $S_m \subset L^2(A)$; then we choose the best possible estimator using a penalty function $pen(m)$. The collection of vectorial subspaces $(S_m)_{m\in\mathscr{M}_n}$ has to satisfy the following assumption.

A4. 1. The subspaces S_m have finite dimension D_m.
2. The sequence of vectorial subspaces $(S_m)_{m\geq 0}$ is increasing: for any m, $S_m \subseteq S_{m+1}$.

3. Norm connexion: there exists a constant ϕ_1 such that, for any $m \geq 0$, any $t \in S_m$,

$$\|t\|_\infty^2 \leq \phi_1 D_m \|t\|_{L^2}^2$$

where $\|.\|_{L^2}$ is the L^2-norm and $\|.\|_\infty$ is the sup-norm on A.

4. For any $m \in \mathbb{N}$, there exists an orthonormal basis $(\psi_\lambda)_{\lambda \in \Lambda_m}$ of S_m such that

$$\forall \lambda, \quad \mathrm{card}\left(\lambda', \ \|\psi_\lambda \psi_{\lambda'}\|_\infty \neq 0\right) \leq \phi_0$$

where ϕ_0 does not depend on m.

5. For any function t belonging to the unit ball of the Besov space $\mathscr{B}_{2,\infty}^\alpha$,

$$\exists C, \forall m \quad \|t - t_m\|_{L^2}^2 \leq C D_m^{-2\alpha}$$

where t_m is the L^2 orthogonal projection of t on S_m.

The subspaces generated by piecewise polynomials, compactly supported wavelets or spline functions satisfy A4 (see [8,16] for instance).

3. Estimation of the drift

By analogy with [6], we decompose $Y_{k\Delta}$ in the following way:

$$Y_{k\Delta} = \frac{X_{(k+1)\Delta} - X_{k\Delta}}{\Delta} = b(X_{k\Delta}) + I_{k\Delta} + Z_{k\Delta} + T_{k\Delta} \tag{2}$$

where

$$I_{k\Delta} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} (b(X_s) - b(X_{k\Delta}))\, ds, \qquad Z_{k\Delta} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s) dW_s$$
$$T_{k\Delta} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} \xi(X_{s^-}) dL_s.$$

The terms $Z_{k\Delta}$ and $T_{k\Delta}$ are martingale increments. Let us introduce the mean square contrast function

$$\gamma_n(t) = \frac{1}{n} \sum_{k=1}^n (Y_{k\Delta} - t(X_{k\Delta}))^2. \tag{3}$$

We can always minimise $\gamma_n(t)$ on S_m, but the minimiser may be not unique. That is why we introduce the empirical risk

$$\mathscr{R}_n(t) = \mathbb{E}\left(\|t - b_A\|_n^2\right) \quad \text{where } \|t\|_n^2 = \frac{1}{n} \sum_{k=1}^n t^2(X_{k\Delta}) \quad \text{and} \quad t_A = t \mathbb{1}_A. \tag{4}$$

We consider the asymptotic framework:

$$\Delta \to 0, \qquad n\Delta \to \infty.$$

For any $m \in \mathscr{M}_n = \{m, \ D_m \leq \mathscr{D}_n\}$ where $\mathscr{D}_n^2 \leq n\Delta / \ln^2(n)$, we construct the regression-type estimator:

$$\hat{b}_m = \arg\min_{t \in S_m} \gamma_n(t).$$

Theorem 2. *Under Assumptions A1– A4 the risk of the estimator with fixed m satisfies:*

$$\mathscr{R}_n(\hat{b}_m) \leq 3\pi_1 \|b_m - b_A\|_{L^2}^2 + 48(\sigma_0^2 + \xi_0^2)\frac{D_m}{n\Delta} + c\Delta$$

where b_m is the orthogonal (L^2) projection of b_A over the vectorial subspace S_m. The constant c is independent of m, n and Δ.

Except for the constant $(\sigma_0^2 + \xi_0^2)$ in the variance term, this is exactly the bound of the risk that Comte et al. [6] found for a diffusion process without jumps.

The bias term, $\|b_m - b_A\|_{L^2}^2$, decreases when the dimension D_m increases whereas the variance term $(\sigma_0^2 + \xi_0^2)D_m/(n\Delta)$ is proportional to the dimension. Under the classical assumption $n\Delta^2 = O(1)$, the remainder term Δ is negligible. Thus we need to find a good compromise between the bias and the variance term.

Remark 3. If the regularity of the drift function is known, that is, if b belongs to a ball of a Besov space $\mathscr{B}_{2,\infty}^{\alpha}$, then the bias term $\|b_m - b_A\|_{L^2}^2$ is smaller than $D_m^{-2\alpha}$. The best estimator is obtained when the bias term, $\|b_m - b_A\|_{L^2}^2$, and the variance term, $cD_m(n\Delta)^{-1}$, are equal, that is for $D_{m_{opt}} = (n\Delta)^{1/(1+2\alpha)}$. In that case, the estimator risk satisfies:

$$\mathscr{R}_n(\hat{b}_{m_{opt}}) \lesssim (n\Delta)^{-2\alpha/(2\alpha+1)} + \Delta.$$

Let us introduce a penalty function pen such that:

$$pen(m) \geq \kappa(\sigma_0^2 + \xi_0^2)\frac{D_m}{n\Delta}$$

and set:

$$\hat{m} = \arg\min_{m\in\mathscr{M}_n}\left\{\gamma_n(\hat{b}_m) + pen(m)\right\}.$$

We will chose κ later. We denote by $\hat{b}_{\hat{m}}$ the resulting estimator. To bound the risk of the adaptive estimator, an additional assumption is needed.

A5. 1. The Lévy measure ν is symmetric or the function ξ is constant.
2. The Lévy measure ν is sub exponential: there exist $\lambda, C > 0$ such that, for any $|z| > 1$, $\nu(]-z,z[^c) \leq Ce^{-\lambda|z|}$.

Theorem 4. *Under Assumptions A1– A5, there exists a constant κ (depending only on ν) such that, if $\mathscr{D}_n^2 \leq n\Delta/\ln^2(n)$:*

$$\mathbb{E}\left(\left\|\hat{b}_{\hat{m}} - b_A\right\|_n^2\right) \lesssim \inf_{m\in\mathscr{M}_n}\left(\|b_m - b_A\|_{L^2}^2 + pen(m)\right) + \left(\Delta + \frac{1}{n\Delta}\right).$$

Remark 5. We can bound κ theoretically; however, this bound is in practice too large for the simulations. In Section 5, we calibrate κ by simulations (see [6] for instance). If σ and ξ are unknown, it is possible to replace them by rough estimators (in fact, we only need upper bounds of σ_0^2 and ξ_0^2). It is also possible to perform a completely data-driven calibration of the parameters of the penalty (see [2]).

4. Truncated estimator of the drift

Truncated estimators are widely used for the estimation of the diffusion coefficient of a jump diffusion (see for instance [14,20,13]). Our aim is to construct an adaptive estimator of b even if Assumption A5 is not fulfilled. To this end, we cut off the big jumps. Let us introduce the set

$$\Omega_{X,k} = \left\{\omega,\ \left|X_{(k+1)\Delta} - X_{k\Delta}\right| \leq C_\Delta\right\}$$

where $C_\Delta = (b_{\max} + 3)\Delta + (\sigma_0 + 4\xi_0)\sqrt{\Delta}\ln(n)$ (with $b_{\max} = \sup_{x\in A}|b(x)|$). Let us consider the random variables

$$\tilde{Y}_{k\Delta} = \frac{X_{(k+1)\Delta} - X_{k\Delta}}{\Delta}\mathbb{1}_{\Omega_{X,k}}\mathbb{1}_{X_{k\Delta}\in A}.$$

We recall here the definition of the Blumenthal–Getoor index.

Definition 6. The Blumenthal–Getoor index of a Lévy measure is

$$\beta = \inf\left\{\alpha \geq 0,\ \int_{|z|\leq 1}|z|^\alpha \nu(dz) < \infty\right\}.$$

A compound Poisson process has $\beta = 0$.

We assume that the following assumption is fulfilled.

A6. 1. For $|x|$ small, $\nu(dx)$ is absolutely continuous with respect to the Lebesgue measure ($\nu(dx) = n(x)dx$) and:

$$\exists\beta \in [0, 2[,\ \exists a_0,\ \forall x \in [-a_0, a_0],\quad n(x) \leq Cx^{-\beta-1}.$$

This implies that the Blumenthal–Getoor index is equal to β.

2. The Lévy measure $\nu(z)$ is symmetric for z small:

$$\exists a_1 < a_0,\ \forall z \in [-a_1, a_1],\quad n(z) = n(-z).$$

3. The function ξ is bounded from below: there exists $\xi_1 > 0$ such that, for any $z \in \mathbb{R}$, $0 < \xi_1 \leq \xi(z)$.
4. The functions σ and ξ are $\mathscr{C}^2$, ξ' and σ' are Lipschitz.

We consider the following asymptotic framework:

$$\frac{n\Delta}{\ln^2(n)} \to \infty, \qquad \Delta^{1-\beta/2}\ln^2(n) \to 0.$$

The truncated estimator $\tilde{b}_m$ is obtained by minimising the contrast function:

$$\tilde{b}_m = \arg\min_{t\in S_m}\tilde{\gamma}_n(t) \quad \text{where } \tilde{\gamma}_n(t) = \frac{1}{n}\sum_{k=1}^{n}\left(\tilde{Y}_{k\Delta} - t(X_{k\Delta})\right)^2.$$

Theorem 7 (*Risk of the Non Adaptive Truncated Estimator*)**.** *Under Assumptions A1– A4 and A6, for any m such that $D_m \leq \mathscr{D}_n$ where $\mathscr{D}_n^2 \leq n\Delta/\ln^2(n)$:*

$$\mathbb{E}\left(\left\|\tilde{b}_m - b_A\right\|_n^2\right) \lesssim \|b_m - b_A\|_{L^2}^2 + (\sigma_0^2 + c\Delta^{1/2-\beta/4})\frac{D_m}{n\Delta} + \Delta^{1-\beta/2}\ln^2(n) + \frac{1}{n\Delta}.$$

The variance term is smaller than for the first estimator, but the remainder term depends on the Blumenthal–Getoor index and is larger than for the first estimator. This remainder term is due to the fact that $\tilde{Y}_{k\Delta} = 0$ every time $|X_{(k+1)\Delta} - X_{k\Delta}| > C_\Delta$: then

$$\left|\mathbb{E}\left(\tilde{Y}_{k\Delta} - b(X_{k\Delta})\right)\right| > |\mathbb{E}(Y_{k\Delta} - b(X_{k\Delta}))|.$$

If L_t is a compound Poisson process, (which implies $\beta = 0$) or if Δ is small enough (see Remark 9), we obtain a better inequality than for the non-truncated estimator.

Remark 8. If ν is not absolutely continuous, we can prove the weaker inequality:

$$\mathbb{E}\left(\left\|\tilde{b}_m - b_A\right\|_n^2\right) \lesssim \|b_m - b_A\|_{L^2}^2 + (\sigma_0^2 + \xi_0^2)\frac{D_m}{n\Delta} + \Delta^{1-\beta}\ln^2(n) + \frac{1}{n\Delta}.$$

In that case, $\tilde{b}_m$ converges towards b_A only if $\beta < 1$, which implies that ν has finite variation ($\int_{\mathbb{R}} |z|\nu(dz) < \infty$). See Remark 18.

Remark 9. Assume that b_A belongs to the Besov space $\mathscr{B}_{2,\infty}^\alpha$ and that $\|b_A\|_{\mathscr{B}_{2,\infty}^\alpha} \leq 1$. The bias–variance compromise $\|b_m - b_A\|_{L^2}^2 + D_m/n\Delta$ is minimal when $m = \log_2(n\Delta)/(1+2\alpha)$, and the risk satisfies:

$$\mathbb{E}\left(\left\|\tilde{b}_m - b_A\right\|_n^2\right) \lesssim (n\Delta)^{-2\alpha/(1+2\alpha)} + \Delta^{1-\beta/2}\ln^2(n).$$

Let us set $\Delta \sim n^{-\gamma}$ with $\gamma > 0$. We have the following convergence rates:

γ	First estimator	Truncated estimator
$0 < \gamma \leq \frac{2\alpha}{4\alpha+1} \leq \frac{1}{2}$	Δ	$\Delta^{1-\beta/2}\ln^2(n)$
$\frac{2\alpha}{4\alpha+1} \leq \gamma \leq \frac{2\alpha}{4\alpha+1-\beta\alpha-\beta/2} \leq \frac{1}{2(1-\beta/4)}$	$(n\Delta)^{-2\alpha/(2\alpha+1)}$	$\Delta^{1-\beta/2}\ln^2(n)$
$\frac{2\alpha+1}{4\alpha+1-\beta\alpha-\beta/2} \leq \gamma < 1$	$(n\Delta)^{-2\alpha/(2\alpha+1)}$	$(n\Delta)^{-2\alpha/(2\alpha+1)}$

If we have sufficiently high frequency data ($n\Delta^{2(1-\beta/4)} = O(1)$), then the rate of convergence is $(n\Delta)^{2\alpha/(2\alpha+1)}$ for the two estimators. The estimator of [13] converges with the corresponding parametric rate, $n\Delta$, if $n\Delta^{3/2-\gamma} = o(1)$ for $\gamma \in]0, 1/2[$.

To construct the adaptive estimator, we use the same penalty function as in the previous section:

$$pen(m) \geq \kappa\left(\sigma_0^2 + \xi_0^2\right)\frac{D_m}{n\Delta}$$

and define the adaptive estimator:

$$\tilde{m} = \arg\min_{m\in\mathscr{M}_n}\left\{\tilde{\gamma}_n(\tilde{b}_m) + pen(m)\right\}.$$

Theorem 10 (*Risk of the Adaptive Truncated Estimator*)**.** *If Assumptions A1– A4 and A6 are satisfied, then there exists κ such that, if $\mathscr{D}_n^2 \leq n\Delta/\ln^2(n)$:*

$$\mathbb{E}\left(\left\|\tilde{b}_{\tilde{m}} - b_A\right\|_n^2\right) \lesssim \min_{m\in\mathscr{M}_n}\left(\|b_m - b_A\|_n^2 + pen(m)\right) + \Delta^{1-\beta/2}\ln^2(n) + \frac{1}{n\Delta}.$$

The adaptive estimator $\tilde{b}_{\tilde{m}}$ automatically realises the bias/variance compromise.

5. Numerical simulations and examples

5.1. Models

We consider the stochastic differential equation:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \xi(X_{t^-})dL_t$$

where L_t is a compound Poisson process of intensity 1: $L_t = \sum_{j=1}^{N_t} \zeta_i$, with N_t a Poisson process of intensity 1 and $(\zeta_1, \dots, \zeta_n)$ are independent and identically distributed random variables independent of (N_t). We denote by f the probability law of ζ_i.

Model 1:

$$b(x) = -2x, \qquad \sigma(x) = \xi(x) = 1 \quad \text{and} \quad f(dz) = \nu(dz) = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}.$$

Model 2:

$$b(x) = -\left(x - 1/4\right)^3 - \left(x + 1/4\right)^3, \qquad \sigma(x) = \xi(x) = 1 \quad \text{and}$$
$$f(dz) = \nu(dz) = \frac{e^{-\lambda|z|}dz}{2}.$$

We can remark that the function b is not Lipschitz and therefore does not satisfy Assumption A1.

Model 3:

We consider the stochastic process of parameters

$$b(x) = -2x + \sin(3x), \qquad \sigma(x) = \xi(x) = \sqrt{\frac{3 + x^2}{1 + x^2}}$$

and

$$f(dz) = \nu(dz) = \frac{1}{4}\sqrt{\frac{\sqrt{24}}{|z|}}e^{-\sqrt{\sqrt{24}|z|}}dz.$$

Let us remark that $\nu = f$ is not sub-exponential and does not satisfy A5. Nevertheless, this model satisfies all the assumptions of Theorem 10.

Model 4:

In this model, the Lévy process is not a compound Poisson process. We set

$$\nu(dz) = \sum_{k=0}^{\infty} 2^{k+2}(\delta_{1/2^k} + \delta_{-1/2^k}), \qquad b(x) = -2x \quad \text{and} \quad \sigma(x) = \xi(x) = 1.$$

The Blumenthal–Getoor index of this process is such that $\beta > 1$.

5.2. Simulation algorithm (compound Poisson case)

We estimate b on the compact interval $A = [-1, 1]$.

1. Simulate random variables $(X_0, X_\Delta, \dots, X_{n\Delta})$ thanks to a Euler scheme with sampling interval $\delta = \Delta/5$. To this end, we use the same simulation scheme as [17]. We simulate the times of the jumps $(\tau_1, \dots, \tau_N, \tau_{N+1})$ with $\tau_N < n\Delta \le \tau_{N+1}$ and we fix $X_0 = 0$.

If $\delta < \tau_1$, we compute

$$X_\delta = \delta b(X_0) + \sqrt{\delta}\sigma(X_0)N \quad \text{with } N \sim \mathcal{N}(0,1).$$

If $\tau_1 < \delta$, we first compute

$$X_{\tau_1} = \tau_1 b(X_0) + \sqrt{\tau_1}\sigma(X_0)N + \xi(X_0)\zeta_1$$

with $N \sim \mathcal{N}(0,1)$ and $\zeta_1 \sim f$ is independent of N. If $\delta < \tau_2$, we compute

$$X_\delta = (\delta - \tau_1)b(X_{\tau_1}) + \sqrt{\delta - \tau_1}\sigma(X_{\tau_1})N'$$

else we compute

$$X_{\tau_2} = (\tau_2 - \tau_1)b(X_{\tau_1}) + \sqrt{\tau_2 - \tau_1}\sigma(X_{\tau_1})N' + \xi(X_{\tau_1})\zeta_2$$

where $N' \sim \mathcal{N}(0,1)$ and ζ_2 has distribution f. N, N', ζ_1 and ζ_2 are independent.

2. Construct the random variables

$$Y_{k\Delta} = \frac{X_{(k+1)\Delta} - X_{k\Delta}}{\Delta} \quad \text{and} \quad \tilde{Y}_{k\Delta} = \frac{X_{(k+1)\Delta} - X_{k\Delta}}{\Delta}\mathbb{1}_{\Omega_{X,k}}\mathbb{1}_{X_{k\Delta}\in A}.$$

3. We consider the vectorial subspaces $S_{m,r}$ generated by the spline functions of degree r (see for instance [19]). In that case $D_{m,r} = \dim(S_{m,r}) = 2^m + r$. For $r \in \{1,2,3\}$ and $m \in \mathcal{M}_n(r) = \{m, D_{m,r} \leq \mathcal{D}_n\}$, we compute the estimators $\hat{b}_{m,r}$ and $\tilde{b}_{m,r}$ by minimising the contrast functions γ_n and $\tilde{\gamma}_n$ on the vectorial subspaces $S_{m,r}$.
4. For the estimation algorithm, we make a selection of m and r as follows. Using the penalty function $pen(m,r) := pen(m) = \kappa(\sigma_0^2 + \xi_0^2)(2^m + r)/n\Delta$, we select the adaptive estimators $\hat{b}_{\hat{m},r}$ and $\tilde{b}_{\tilde{m},r}$, and then choose the best r by minimising $\gamma_n(\hat{b}_{\hat{m},r}) + pen(\hat{m},r)$ and $\tilde{\gamma}_n(\tilde{b}_{\tilde{m},r}) + pen(\tilde{m},r)$.

To calibrate κ, we run a various number of simulations for a model with known parameters and let κ vary. When κ is too small, the value of m selected by the estimation procedure is in general very high (often maximal). When κ is too big, the estimator is always linear even if the true function is not. We used the true value of σ_0^2 and ξ_0^2.

5.3. Results

In Figs. 1–4, we simulate 5 times the process $(X_0, \ldots, X_{n\Delta})$ for $\Delta = 10^{-1}$ and $n = 10^4$ and draw the obtained estimators. The two adaptive estimators are nearly superposed; moreover, they are close to the true function.

In Tables 1–4, for each value of (n, Δ), we simulate 50 trajectories of $(X_0, X_\Delta, \ldots, X_{n\Delta})$. For each path, we construct the two adaptive estimators $\hat{b}_{\hat{m},\hat{r}}$ and $\tilde{b}_{\tilde{m},\tilde{r}}$ and we compute the empirical errors:

$$err_1 = \left\|\hat{b}_{\hat{m},\hat{r}} - b_A\right\|_n^2 \quad \text{and} \quad err_2 = \left\|\tilde{b}_{\tilde{m},\tilde{r}} - b_A\right\|_n^2.$$

In order to check that our algorithm is adaptive, we also compute the minimal errors

$$emin_1 = \min_{m,r}\left\|\hat{b}_{m,r} - b_A\right\|_n^2 \quad \text{and} \quad emin_2 = \min_{m,r}\left\|\tilde{b}_{m,r} - b_A\right\|_n^2$$

and the oracles $oracle_i = err_i/emin_i$. We give the means $\hat{m}_a, \hat{r}_a, \tilde{m}_a$ and $\tilde{r}_a$ of the selected values $\hat{m}, \hat{r}, \tilde{m}$ and $\tilde{r}$. The value $risk_i$ is the mean of err_i over the 50 simulations and or_i is the mean of $oracle_i$. The computation time for one adaptive estimator varies from 0.1 s

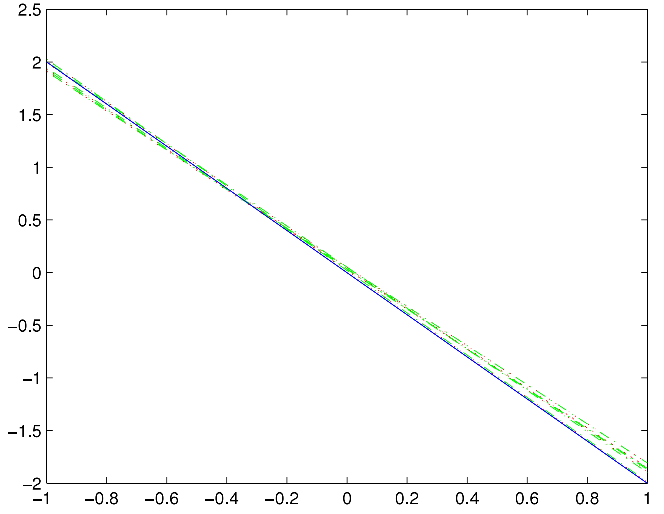

Fig. 1. Model 1: Ornstein–Uhlenbeck and binomial law. $b(x) = -2x$, $\sigma(x) = \xi(x) = 1$ and binomial law.

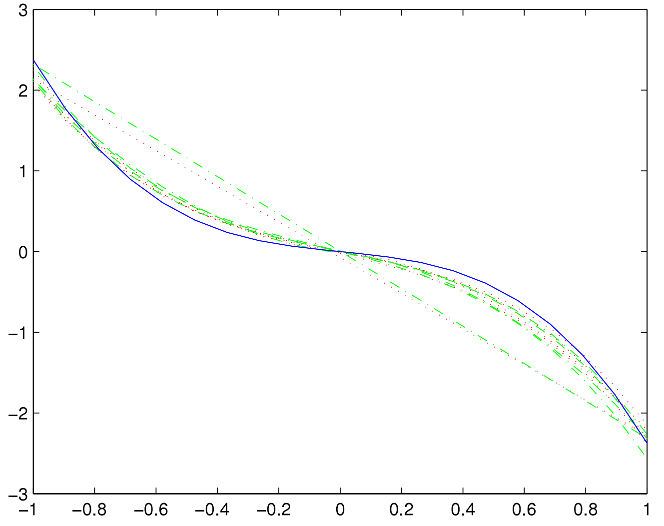

Fig. 2. Model 2: Double well and Laplace law. $b(x) = -(x - 1/4)^3 - (x + 1/4)^3$, $\sigma = \xi = 1$ and Laplace law.

($\Delta = 10^{-1}, n = 10^3$) to 30 s ($\Delta = 10^{-1}, n = 10^4$). The empirical risk is decreasing when the product $n\Delta$ is increasing, which is coherent with the theoretical model. For Model 1, the two estimators are equivalent. When the tails of ν become larger (Models 2 and 3), the truncated estimator is better. The improvement is also more significant when the discretisation distance is smaller. As on the first three models, the processes L_t are compound Poisson processes,

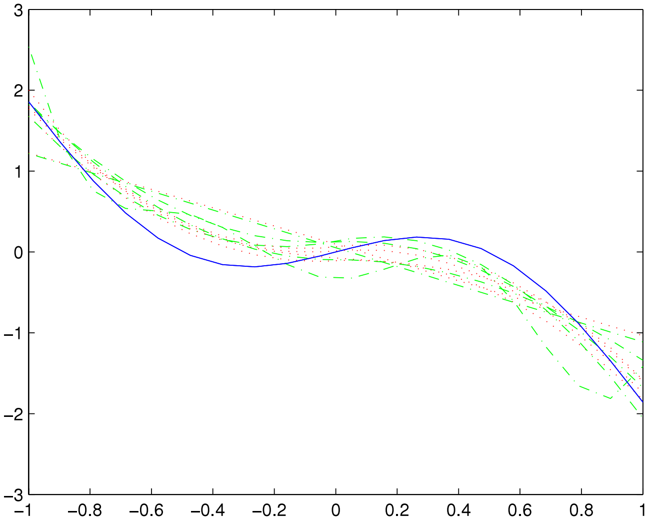

Fig. 3. Model 3: Sine function. $b(x) = -2x + \sin(3x)$, $\sigma(x) = \xi(x) = \sqrt{(3+x^2)/(1+x^2)}$ jumps not sub-exponential.

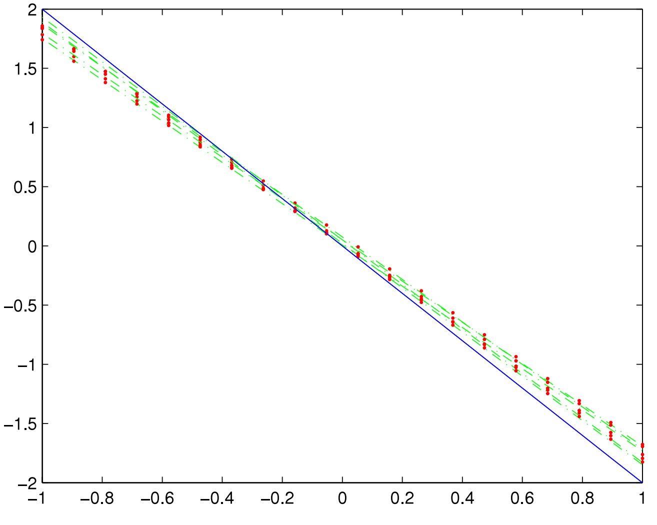

Fig. 4. Model 4: Lévy process. $b(x) = -2x$, $\sigma(x) = \xi(x) = 1$ jumps Lévy.

these results were expected. The truncated estimator seems also more robust: we do not observe aberrant values (like for the first estimator in Table 2). Those aberrant values may be due to the fact that b is not Lipschitz and then $b(X_{k\Delta})$ may be quite large, and to the non-exact simulation by an Euler scheme. For Model 4, the results are slightly better for the first estimator when

Table 1
Model 1: Ornstein–Uhlenbeck and binomial law. $b(x) = -2x$, $\sigma(x) = \xi(x) = 1$ and compound Poisson process (binomial law).

n	Δ	First estimator				Truncated estimator			
		$\hat{m}_a$	$\hat{r}_a$	$risk_1$	or_1	$\tilde{m}_a$	$\tilde{r}_a$	$risk_2$	or_2
10^3	10^{-1}	0	1.02	0.044	1.3	0	1.02	0.044	1.3
10^4	10^{-1}	0	1.02	0.011	1.3	0	1.02	0.011	1.3
10^3	10^{-2}	0	1.02	0.55	1.04	0	1.02	0.55	1.04
10^4	10^{-2}	0	1	0.047	1	0	1	0.047	1
5.10^4	10^{-2}	0.04	1	0.010	1.4	0	1	0.0053	1

$\hat{m}_a, \hat{r}_a$ and $\tilde{m}_a, \tilde{r}_a$: average values of $\hat{m}, \hat{r}$ and $\tilde{m}, \tilde{r}$ on the 50 simulations.
$risk_1$ and $risk_2$: means of the empirical errors of the adaptive estimators.
or_1 and or_2: means of $oracle$ = empirical error of the adaptive estimator/empirical error of the best possible estimator.

Table 2
Model 2: Double well and Laplace law. $b(x) = -(x - 1/4)^3 - (x + 1/4)^3$, $\sigma(x) = \xi(x) = 1$ and Laplace law.

n	Δ	First estimator				Truncated estimator			
		$\hat{m}_a$	$\hat{r}_a$	$risk_1$	or_1	$\tilde{m}_a$	$\tilde{r}_a$	$risk_2$	or_2
10^3	10^{-1}	0.02	1.0	0.12	3.1	0.02	1.0	0.12	3.1
10^4	10^{-1}	1.7	2.1	2e96	51	0.4	2.1	0.04	1.5
10^3	10^{-2}	0.26	1.2	1.8	3.1	0.06	1	0.51	1.4
10^4	10^{-2}	0.12	1.5	0.16	1.8	0.08	1.2	0.13	2.4
5.10^4	10^{-2}	0.30	2.5	0.035	1.6	0.26	2.5	0.019	1.8

$\hat{m}_a, \hat{r}_a$ and $\tilde{m}_a, \tilde{r}_a$: average values of $\hat{m}, \hat{r}$ and $\tilde{m}, \tilde{r}$ on the 50 simulations.
$risk_1$ and $risk_2$: means of the empirical errors of the adaptive estimators.
or_1 and or_2: means of $oracle$ = empirical error of the adaptive estimator/empirical error of the best possible estimator.

Table 3
Model 3: Sine function and jumps not sub-exponential. $b(x) = -2x + \sin(3x), \sigma(x) = \xi(x) = \sqrt{(3 + x^2)/(1 + x^2)}$ and $\nu(dz) \propto e^{-\sqrt{az}}/\sqrt{z}dz$.

n	Δ	First estimator				Truncated estimator			
		$\hat{m}_a$	$\hat{r}_a$	$risk_1$	or_1	$\tilde{m}_a$	$\tilde{r}_a$	$risk_2$	or_2
10^3	10^{-1}	0.34	1.2	0.76	3.6	0.04	1.2	0.28	1.9
10^4	10^{-1}	0.8	2.2	0.082	1.3	0.68	2.2	0.073	1.2
10^3	10^{-2}	0.96	1.2	18	6.3	0.02	1.2	1.3	1.2
10^4	10^{-2}	0.78	1.4	1.5	4.3	0.12	1.4	0.24	3.3
5.10^4	10^{-2}	0.92	2.3	0.24	4.3	0.70	2.3	0.039	1.3

$\hat{m}_a, \hat{r}_a$ and $\tilde{m}_a, \tilde{r}_a$: average values of $\hat{m}, \hat{r}$ and $\tilde{m}, \tilde{r}$ on the 50 simulations.
$risk_1$ and $risk_2$: means of the empirical errors of the adaptive estimators.
or_1 and or_2: means of $oracle$ = empirical error of the adaptive estimator/empirical error of the best possible estimator.

$\Delta = 0.1$, which is due to the fact that the remainder term is greater for the truncated estimator. When $\Delta = 10^{-2}$, the risk of the truncated estimator is lower than for the first estimator.

6. Proofs

Let us introduce the filtration

$$\mathscr{F}_t = \sigma\left(\eta, (W_s)_{0\leq s\leq t}, (L_s)_{0\leq s\leq t}\right).$$

Table 4
Model 4: Lévy process. $b(x) = -2x$, $\sigma(x) = \xi(x) = 1$ and $\nu(dz) = \sum_{k=0}^{\infty} 2^{k+2}(\delta_{2^{-k}} + \delta_{-2^{-k}})$.

n	Δ	First estimator				Truncated estimator			
		$\hat{m}_a$	$\hat{r}_a$	$risk_1$	or_1	$\tilde{m}_a$	$\tilde{r}_a$	$risk_2$	or_2
10^3	10^{-1}	0.04	1.06	0.110	1.86	0.02	1.06	0.111	1.95
10^4	10^{-1}	0.06	1.06	0.0172	1.26	0.06	1.06	0.0176	1.22
10^3	10^{-2}	0.1	1.04	1.17	1.88	0	1.04	0.61	1.12
10^4	10^{-2}	0.04	1.08	0.11	1.25	0.02	1.08	0.068	1.25
5.10^4	10^{-2}	0.08	1.16	0.023	1.71	0	1.16	0.011	1.09

$\hat{m}_a, \hat{r}_a$ and $\tilde{m}_a, \tilde{r}_a$: average values of $\hat{m}, \hat{r}$ and $\tilde{m}, \tilde{r}$ on the 50 simulations.
$risk_1$ and $risk_2$: means of the empirical errors of the adaptive estimators.
or_1 and or_2: means of $oracle$ = empirical error of the adaptive estimator/empirical error of the best possible estimator.

The following result is very useful. It comes from [7, Theorem 92 Chapter VII] and [1, Theorem 4.4.23 p. 265] (Kunita's first inequality).

Result 11 (*Burkholder–Davis–Gundy Inequality*). *We have that, for any $p \geq 2$,*

$$\mathbb{E}\left[\sup_{s\in[t,t+h]}\left|\int_t^s \sigma(X_u)dW_u\right|^p \middle| \mathscr{F}_t\right] \leq C_p\left(\mathbb{E}\left[\left|\int_t^{t+h}\sigma^2(X_u)du\right|^{p/2}\middle|\mathscr{F}_t\right]\right)$$

and, if $\int_{\mathbb{R}}|z|^p\,\nu(dz) < \infty$, as $\int_{\mathbb{R}} z^2\nu(dz) = 1$:

$$\mathbb{E}\left[\sup_{s\in[t,t+h]}\left|\int_t^s \xi(X_{u^-})dL_u\right|^p \middle| \mathscr{F}_t\right] \leq C_p\mathbb{E}\left[\left(\int_t^{t+h}\xi^2(X_u)du\right)^{p/2}\middle|\mathscr{F}_t\right]$$
$$+ C_p\mathbb{E}\left[\left(\int_t^{t+h}|\xi(X_u)|^p\,du\right)\middle|\mathscr{F}_t\right]\int_{\mathbb{R}}|z|^p\,\nu(dz).$$

6.1. *Proof of Theorem 2*

By (3) and (4), we get:

$$\gamma_n(t) = \frac{1}{n}\sum_{k=1}^n (Y_{k\Delta} - t(X_{k\Delta}))^2 = \frac{1}{n}\sum_{k=1}^n (Y_{k\Delta} - b(X_{k\Delta}))^2 + \|b-t\|_n^2$$
$$+ \frac{2}{n}\sum_{k=1}^n (Y_{k\Delta} - b(X_{k\Delta}))\left(b(X_{k\Delta}) - t(X_{k\Delta})\right).$$

As, by definition, $\gamma_n(\hat{b}_m) \leq \gamma_n(b_m)$, we obtain:

$$\left\|\hat{b}_m - b\right\|_n^2 \leq \|b_m - b\|_n^2 + \frac{2}{n}\sum_{k=1}^n (Y_{k\Delta} - b(X_{k\Delta}))\left(\hat{b}_m(X_{k\Delta}) - b_m(X_{k\Delta})\right).$$

By (2), and as $\hat{b}_m$ and b_m are supported by A,

$$\left\|\hat{b}_m - b_A\right\|_n^2 \leq \|b_m - b_A\|_n^2 + \frac{2}{n}\sum_{k=1}^n (I_{k\Delta} + Z_{k\Delta} + T_{k\Delta})\left(\hat{b}_m(X_{k\Delta}) - b_m(X_{k\Delta})\right).$$

Let us introduce the unit ball

$$\mathscr{B}_m = \{t \in S_m,\ \|t\|_\varpi \leq 1\} \quad \text{where } \|t\|_\varpi^2 = \int_A t^2(x)\varpi(dx)$$

and the englobing space $\mathscr{S}_n = \bigcup_{m\in\mathscr{M}_n} S_m$. Let us consider the set

$$\Omega_n = \left\{\omega,\ \forall t \in \mathscr{S}_n,\ \left|\frac{\|t\|_n^2}{\|t\|_\varpi^2} - 1\right| \leq \frac{1}{2}\right\}$$

where the norms $\|.\|_\varpi$ and $\|.\|_n$ are equivalent.

Step 1: *Bound of the risk on* Ω_n. Thanks to the Cauchy–Schwarz inequality, we obtain that, on Ω_n:

$$\left\|\hat{b}_m - b_A\right\|_n^2 \leq \|b_m - b_A\|_n^2 + \frac{1}{12}\left\|\hat{b}_m - b_m\right\|_n^2 + 12\sum_{k=1}^n I_{k\Delta}^2 + \frac{1}{12}\left\|\hat{b}_m - b_m\right\|_\varpi^2 + 12\sup_{t\in\mathscr{B}_m} \nu_n^2(t)$$

where

$$\nu_n(t) = \frac{1}{n}\sum_{k=1}^n (Z_{k\Delta} + T_{k\Delta})t(X_{k\Delta}). \tag{5}$$

On Ω_n, by definition, we have:

$$\left\|\hat{b}_m - b_m\right\|_n^2 \leq 2\left\|\hat{b}_m - b_A\right\|_n^2 + 2\|b_m - b_A\|_n^2 \quad \text{and} \quad \left\|\hat{b}_m - b_m\right\|_\varpi^2 \leq 2\left\|\hat{b}_m - b_m\right\|_n^2.$$

Thus we obtain:

$$\left\|\hat{b}_m - b_A\right\|_n^2 \leq 3\|b_m - b_A\|_n^2 + 24\sum_{k=1}^n I_{k\Delta}^2 + 24\sup_{t\in\mathscr{B}_m}\nu_n^2(t).$$

The following lemma is very useful. It is derived from Proposition 1 and Result 11.

Lemma 12. 1. $\mathbb{E}\left(I_{k\Delta}^2\right) \leq c\Delta$ *and* $\mathbb{E}\left(I_{k\Delta}^4\right) \leq c\Delta$.
2. $\mathbb{E}\left(Z_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right) = 0$, $\mathbb{E}\left(Z_{k\Delta}^2\middle|\mathscr{F}_{k\Delta}\right) \leq \sigma_0^2/\Delta$ *and* $\mathbb{E}\left(Z_{k\Delta}^4\middle|\mathscr{F}_{k\Delta}\right) \leq c/\Delta^2$.
3. $\mathbb{E}\left(T_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right) = 0$, $\mathbb{E}\left(T_{k\Delta}^2\middle|\mathscr{F}_{k\Delta}\right) \leq \xi_0^2/\Delta$ *and* $\mathbb{E}\left(T_{k\Delta}^4\middle|\mathscr{F}_{k\Delta}\right) \leq c/\Delta^3$.

By Lemma 12, $\mathbb{E}\left[I_{k\Delta}^2\right] \leq \Delta$. It remains to bound $\mathbb{E}\left[\sup_{t\in\mathscr{B}_m}\nu_n^2(t)\right]$. We consider an orthonormal basis $(\varphi_\lambda)_{\lambda\in\Lambda_m}$ of S_m for the L_ϖ^2-norm with $|\Lambda_m| = D_m$. Any function $t \in S_m$ can be written $t = \sum_{\lambda\in\Lambda_m} a_\lambda\varphi_\lambda$ and $\|t\|_\varpi^2 = \sum_{\lambda\in\Lambda_m} a_\lambda^2$. Then:

$$\begin{aligned}\sup_{t\in\mathscr{B}_m}\nu_n^2(t) &= \sup_{\sum_\lambda a_\lambda^2\leq 1}\left(\sum_{\lambda\in\Lambda_m} a_\lambda\nu_n(\varphi_\lambda)\right)^2 \\ &\leq \sup_{\sum_\lambda a_\lambda^2\leq 1}\left(\sum_{\lambda\in\Lambda_m} a_\lambda^2\right)\left(\sum_{\lambda\in\Lambda_m}\nu_n^2(\varphi_\lambda)\right) \\ &= \sum_{\lambda\in\Lambda_m}\nu_n^2(\varphi_\lambda).\end{aligned}$$

It remains to bound $\mathbb{E}\left(\nu_n^2(\varphi_\lambda)\right)$. By (5),

$$\mathbb{E}\left[\nu_n^2(\varphi_\lambda)\right] = \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left[\varphi_\lambda^2(X_{k\Delta})\mathbb{E}\left[(Z_{k\Delta}+T_{k\Delta})^2\middle|\mathscr{F}_{k\Delta}\right]\right] + \frac{2}{n^2}\sum_{k<l}^n \mathbb{E}\left[(Z_{k\Delta}+T_{k\Delta})\varphi_\lambda(X_{k\Delta})\varphi_\lambda(X_{l\Delta})\mathbb{E}\left[Z_{l\Delta}+T_{l\Delta}\middle|\mathscr{F}_{l\Delta}\right]\right].$$

Thanks to Lemma 12, the second term of this inequality is null and we obtain, as $\int_{\mathbb{R}}\varphi_\lambda^2(x)\varpi(dx)=1$:

$$\mathbb{E}\left[\nu_n^2(\varphi_\lambda)\right] \le \frac{2(\sigma_0^2+\xi_0^2)}{n^2\Delta}\sum_{k=1}^n \mathbb{E}\left[\varphi_\lambda^2(X_{k\Delta})\right] = \frac{2(\sigma_0^2+\xi_0^2)}{n\Delta}.$$

Therefore:

$$\mathbb{E}\left[\left\|\hat{b}_m - b_A\right\|_n^2 \mathbb{1}_{\Omega_n}\right] \le 3\left\|b_m - b_A\right\|_n^2 + 48(\sigma_0^2+\xi_0^2)\frac{D_m}{n\Delta} + C\Delta.$$

Step 2: *Bound of the risk on Ω_n^c.* The process $(X_t)_{t\ge 0}$ is exponentially β-mixing, π is bounded from below and above and $n\Delta \to \infty$. The following result is proved for $\xi = 0$ for instance in [6] for diffusion processes, but as it relies only on the β-mixing property, we can apply it.

Result 13.

$$\mathbb{P}\left[\Omega_n^c\right] \le \frac{1}{n^3}.$$

Let us set $e = (e_\Delta, \dots, e_{n\Delta})^*$ where $e_{k\Delta} := Y_{k\Delta} - b(X_{k\Delta}) = I_{k\Delta} + Z_{k\Delta} + T_{k\Delta}$ and $\Pi_m Y = \Pi_m (Y_\Delta, \dots, Y_{n\Delta})^* = \left(\hat{b}_m(X_0), \dots, \hat{b}_m(X_{n\Delta})\right)^*$ where Π_m is the Euclidean orthogonal projection over S_m. Then

$$\left\|\hat{b}_m - b_A\right\|_n^2 = \left\|\Pi_m Y - b_A\right\|_n^2 = \left\|\Pi_m b_A - b_A\right\|_n^2 + \left\|\Pi_m Y - \Pi_m b_A\right\|_n^2 \le \left\|b_A\right\|_n^2 + \left\|e\right\|_n^2.$$

According to Lemma 12, Result 13 and the Cauchy–Schwarz inequality,

$$\mathbb{E}\left[\|e\|_n^2 \mathbb{1}_{\Omega_n^c}\right] \le \left(\mathbb{E}\left[\|e\|_n^4\right]\right)^{1/2}\left(\mathbb{P}\left(\Omega_n^c\right)\right)^{1/2} \le \frac{C}{\left(\Delta^3 n^3\right)^{1/2}} \le \frac{C}{n\Delta}$$

and, as b is bounded on the compact set A,

$$\mathbb{E}\left[\|b_A\|_n^2 \mathbb{1}_{\Omega_n^c}\right] \le \left(\mathbb{E}\left[\|b_A\|_n^4\right]\mathbb{P}\left(\Omega_n^c\right)\right)^{1/2} \lesssim \frac{1}{n^{3/2}}.$$

Collecting the results, we get:

$$\mathbb{E}\left[\left\|\hat{b}_m - b_A\right\|_n^2 \mathbb{1}_{\Omega_n^c}\right] \lesssim \frac{1}{n\Delta}$$

which ends the proof of Theorem 2.

6.2. Proof of Theorem 4

The bound of the risk on Ω_n^c is done exactly in the same way as for the non adaptive estimator. It remains thus to bound the risk on Ω_n. As in the previous proof, we get:

$$\left\|\hat{b}_{\hat{m}} - b_A\right\|_n^2 \mathbb{1}_{\Omega_n} \leq 3\left\|b_m - b_A\right\|_n^2 + \frac{24}{n}\sum_{k=1}^{n} I_{k\Delta}^2 + 2pen(m) - 2pen(\hat{m}) + 24\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_n^2(t)$$

where $\mathscr{B}_{m,m'}$ is the unit ball (for the L^2_{ϖ}-norm) of the subspace $S_m + S_{m'}$: $\mathscr{B}_{m,m'} = \{t \in S_m + S_{m'}, \|t\|_{\varpi} \leq 1\}$. Let us introduce a function $p(m, m')$ such that $12p(m, m') = pen(m) + pen(m')$. We obtain that, on Ω_n, for any $m \in \mathscr{M}_n$:

$$\left\|\hat{b}_{\hat{m}} - b_A\right\|_n^2 \leq 3\left\|b_m - b_A\right\|_n^2 + \frac{24}{n}\sum_{k=1}^{n} I_{k\Delta}^2 + 4pen(m) + 24\sup_{t\in\mathscr{B}_{m,\hat{m}}}\left(\nu_n^2(t) - p(m,\hat{m})\right).$$

It remains to bound

$$\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_n^2(t) - p(m,\hat{m})\right] \leq \sum_{m'}\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,m'}} \nu_n^2(t) - p(m,m')\right]_+ .$$

For this purpose, we use the following proposition proved in [1, Corollary 5.2.2].

Proposition 14 (*Exponential Martingale*). *Let $(Y_t)_{t\geq 0}$ satisfy:*

$$Y_t = \int_0^t F_s dW_s + \int_0^t K_s dL_s - \int_0^t \left[\frac{F_s^2}{2} + \int_{\mathbb{R}}\left(e^{K_s z} - 1 - K_s z\right)\nu(dz)\right] ds$$

where F_s and K_s are locally integrable and predictable processes. If for any $t > 0$,

$$\mathbb{E}\left[\int_0^t \int_{|z|>1} \left|e^{K_s z} - 1\right| \nu(dz) ds\right] < \infty,$$

then e^{Y_t} is a $\mathscr{G}_t$-local martingale where $\mathscr{G}_t = \sigma(W_s, L_s, 0 \leq s \leq t)$.

For any $\varepsilon \leq \varepsilon_1 := (\lambda \wedge 1)/(2\|t\|_\infty \xi_0)$ where λ is defined in Assumption A5, for any $t \geq 0$

$$\int_0^t \int_{|z|\geq 1} (\exp(\varepsilon t(X_{k\Delta})\xi(X_s)z) - 1)\, \nu(dz)\mathbb{1}_{s\in]k\Delta,(k+1)\Delta]} ds < \infty.$$

Let us introduce the two Markov processes

$$A_{\varepsilon,t} := \varepsilon^2 \sum_{k=0}^{n} t^2(X_{k\Delta}) \int_0^t \sigma^2(X_s)\mathbb{1}_{s\in]k\Delta,(k+1)\Delta]} ds$$

and

$$B_{\varepsilon,t} := \sum_{k=0}^{n}\int_0^t \int_{\mathbb{R}} (\exp(\varepsilon t(X_{k\Delta})\xi(X_s)z) - \varepsilon t(X_{k\Delta})\xi(X_s)z - 1) \times \mathbb{1}_{s\in]k\Delta,(k+1)\Delta]}\nu(dz) ds$$

and the following martingale:

$$M_t = \int_0^t \sum_{k=0}^n \mathbb{1}_{s\in]k\Delta,(k+1)\Delta]} t(X_{k\Delta^-}) \left(\sigma(X_s)dW_s + \xi(X_{s^-})dL_s\right).$$

By Proposition 14,

$$Y_{\varepsilon,s} := \varepsilon M_s - A_{\varepsilon,s} - B_{\varepsilon,s}$$

is such that $e^{Y_{\varepsilon,s}}$ is a local martingale.

Bound of $A_{\varepsilon,s}$ and $B_{\varepsilon,s}$. We obtain easily that $A_{\varepsilon,s} \le A_{\varepsilon,(n+1)\Delta} \le \varepsilon^2 n\Delta \|t\|_n^2 \sigma_0^2$. Under Assumption A5, ξ is constant or ν is symmetric, and therefore

$$B_{\varepsilon,s} \le B_{\varepsilon,(n+1)\Delta} \le \Delta \sum_{k=0}^n \int_{\mathbb{R}} \left(\exp\left(\varepsilon t(X_{k\Delta})\xi_0 z\right) - \varepsilon t(X_{k\Delta^-})\xi_0 z - 1\right)\nu(dz).$$

As $\int_{\mathbb{R}} z^2\nu(dz) = 1$, for any $\alpha \le 1$,

$$\int_{-1}^1 \left(\exp(\alpha z) - \alpha z - 1\right)\nu(dz) \le \alpha^2 \int_{-1}^1 z^2\nu(dz) \le \alpha^2.$$

Moreover, by integration by parts, for any $\alpha \le (1\wedge\lambda)/2$,

$$\begin{aligned}\int_{[-1,1]^c} \left(\exp(\alpha z) - \alpha z - 1\right)\nu(dz) \le{}& \left(e^{\alpha} - \alpha - 1\right)\nu([1,+\infty[) \\ &+ \left(e^{-\alpha} + \alpha - 1\right)\nu(]-\infty,-1]) \\ &+ \int_1^{+\infty} \alpha\left(e^{\alpha z} - 1\right)\nu([-z,z]^c)dz.\end{aligned}$$

By Assumption A5, $\nu([-z,z]^c) \le Ce^{-\lambda z}$ and then

$$\begin{aligned}\int_{[-1,1]^c} \left(\exp(\alpha z) - \alpha z - 1\right)\nu(dz) &\le 2\alpha^2\nu\left([-1,1]^c\right) + Ce^{-\lambda}\frac{\alpha}{\lambda}\left(\frac{e^{\alpha}}{1-\alpha/\lambda} - 1\right) \\ &\le C'\alpha^2.\end{aligned}$$

Then $B_{\varepsilon,s} \lesssim n\Delta\varepsilon^2\xi_0^2\|t\|_n^2$. There exists a constant c such that, for any $\varepsilon < \varepsilon_1$,

$$A_{\varepsilon,s} + B_{\varepsilon,s} \le c\frac{n\Delta\varepsilon^2\left(\sigma_0^2 + \xi_0^2\right)\|t\|_n^2}{(1-\varepsilon/\varepsilon_1)}.$$

Bound of $\mathbb{P}\left(\nu_n(t) \ge \eta, \|t\|_n^2 \le \zeta^2\right)$. The process $\exp(Y_{\varepsilon,t})$ is a local martingale; then there exists an increasing sequence (τ_N) of stopping times such that $\lim_{N\to\infty}\tau_N = \infty$ and $\exp(Y_{\varepsilon,t\wedge\tau_N})$ is a $\mathscr{F}_t$-martingale. For any $\varepsilon < \varepsilon_1$, and all N,

$$\begin{aligned}E &:= \mathbb{P}\left(M_{(n+1)\Delta\wedge\tau_N} \ge n\Delta\eta, \|t\|_n^2 \le \zeta^2\right) \\ &\le \mathbb{P}\left(M_{(n+1)\Delta\wedge\tau_N} \ge n\Delta\eta, A_{(n+1)\Delta\wedge\tau_N} + B_{(n+1)\Delta\wedge\tau_N} \le \frac{cn\Delta\varepsilon^2\left(\sigma_0^2+\xi_0^2\right)\zeta^2}{(1-\varepsilon/\varepsilon_1)}\right) \\ &\le \mathbb{E}\left(\exp(Y_{\varepsilon,(n+1)\Delta\wedge\tau_N})\right)\exp\left(-n\Delta\eta\varepsilon + \frac{cn\Delta\varepsilon^2\left(\xi_0^2+\sigma_0^2\right)\zeta^2}{(1-\varepsilon/\varepsilon_1)}\right).\end{aligned}$$

As $\exp(Y_{\varepsilon,t\wedge\tau_N})$ is a martingale, $\mathbb{E}\left(\exp(Y_{\varepsilon,t\wedge\tau_N})\right)=1$ and

$$E\leq\exp\left(-n\Delta\eta\varepsilon+\frac{cn\Delta\varepsilon^2\left(\xi_0^2+\sigma_0^2\right)\zeta^2}{(1-\varepsilon/\varepsilon_1)}\right).$$

Letting N tend to infinity, by dominated convergence, and as $\nu_n(t)=n\Delta M_{(n+1)\Delta}$, we obtain that

$$\mathbb{P}\left(\nu_n(t)\geq\eta,\|t\|_n^2\leq\zeta^2\right)\leq\exp\left(-n\Delta\eta\varepsilon+\frac{cn\Delta\varepsilon^2\left(\xi_0^2+\sigma_0^2\right)\zeta^2}{(1-\varepsilon/\varepsilon_1)}\right).$$

It remains to minimise this inequality in ε. Let us set

$$\varepsilon=\frac{\eta}{2c\left(\sigma_0^2+\xi_0^2\right)\zeta^2/\Delta+\eta/\varepsilon_1}<\varepsilon_1.$$

We get:

$$\mathbb{P}\left(\nu_n(t)\geq\eta,\ \|t\|_n^2\leq\zeta^2\right)\leq\exp\left(-\frac{\eta^2n\Delta}{4c\left(\left(\sigma_0^2+\xi_0^2\right)\zeta^2+c'\eta\xi_0\|t\|_\infty\right)}\right).$$

The following lemma concludes the proof. It is proved thanks to a $L^2_\varpi-L^\infty$ chaining technique. See [5, Proof of Proposition 4] and [18, Appendix D.3].

Lemma 15. *There exists a constant κ such that:*

$$\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,m'}}\nu_n^2(t)-p(m,m')\right]\lesssim\kappa(\xi_0^2+\sigma_0^2)\frac{D^{3/2}}{n\Delta}e^{-\sqrt{D}}$$

where $D=\dim(S_m+S_{m'})$.

As $\sum_D D^{3/2}e^{-\sqrt{D}}\leq\sum_{k=0}^{+\infty}k^{3/2}e^{-\sqrt{k}}<\infty$, we obtain that

$$\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,\hat{m}}}\nu_n^2(t)-p(m,\hat{m})\right]\leq\sum_{m'\in\mathscr{M}_n}\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,m'}}\nu_n^2(t)-p(m,m')\right]\lesssim\kappa\frac{\xi_0^2+\sigma_0^2}{n\Delta}.$$

6.3. Proof of Theorem 7

We recall that

$$\Omega_{X,k}=\left\{\omega,\ \left|X_{(k+1)\Delta}-X_{k\Delta}\right|\leq C_\Delta=(b_{\max}+3)\,\Delta+(\sigma_0+4\xi_0)\sqrt{\Delta}\ln(n)\right\}.$$

Let us introduce the set

$$\Omega_{N,k}=\left\{\omega,N'_{k\Delta}=0\right\}$$

where $N'_{k\Delta}$ is the number of jumps of size larger than $\Delta^{1/4}$ occurring in the time interval $]k\Delta,(k+1)\Delta]$:

$$N'_{k\Delta}=\mu\left(]k\Delta,(k+1)\Delta],\left[-\Delta^{1/4},\Delta^{1/4}\right]^c\right).$$

We have that

$$\begin{aligned}\tilde{Y}_{k\Delta} &= Y_{k\Delta}\mathbb{1}_{\Omega_{X,k}}\mathbb{1}_{X_{k\Delta}\in A}\\ &= b_A(X_{k\Delta}) - b_A(X_{k\Delta})\mathbb{1}_{\Omega_{X,k}^c\cap(X_{k\Delta}\in A)} + I_{k\Delta}\mathbb{1}_{\Omega_{X,k}\cap(X_{k\Delta}\in A)} + \tilde{Z}_{k\Delta} + \tilde{T}_{k\Delta}\\ &\quad + (Z_{k\Delta} + T_{k\Delta})\,\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}^c\cap(X_{k\Delta}\in A)}\\ &\quad + \mathbb{E}\left((Z_{k\Delta} + T_{k\Delta})\,\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)}\middle|\,\mathscr{F}_{k\Delta}\right),\end{aligned}$$

where

$$\tilde{Z}_{k\Delta} = Z_{k\Delta}\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)} - \mathbb{E}\left(Z_{k\Delta}\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)}\middle|\,\mathscr{F}_{k\Delta}\right)$$

and

$$\tilde{T}_{k\Delta} = T_{k\Delta}\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)} - \mathbb{E}\left(T_{k\Delta}\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)}\middle|\,\mathscr{F}_{k\Delta}\right).$$

As previously, we only bound the risk on Ω_n. Let us set

$$\tilde{\nu}_n(t) := \frac{1}{n}\sum_{k=1}^{n} t(X_{k\Delta})\left(\tilde{Z}_{k\Delta} + \tilde{T}_{k\Delta}\right).$$

We have that

$$\begin{aligned}\left\|\tilde{b}_m - b_A\right\|_n^2\mathbb{1}_{\Omega_n} \le 3\left\|b_m - b_A\right\|_n^2 + 24\sup_{t\in\mathscr{B}_m}\tilde{\nu}_n^2(t) + \frac{224}{n}\sum_{k=1}^{n}\left(I_{k\Delta}^2 + b_A^2(X_{k\Delta})\mathbb{1}_{\Omega_{X,k}^c}\right)\\ + \frac{224}{n}\sum_{k=1}^{n}\left(Z_{k\Delta}^2 + T_{k\Delta}^2\right)\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}^c\cap(X_{k\Delta}\in A)}\\ + \frac{224}{n}\sum_{k=1}^{n}\left(\mathbb{E}\left[(Z_{k\Delta} + T_{k\Delta})\,\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)}\middle|\,\mathscr{F}_{k\Delta}\right]\right)^2.\end{aligned}$$

The following lemma is proved later.

Lemma 16. 1. $\mathbb{P}(\Omega_{X,k}^c \cap (X_{k\Delta} \in A)) \lesssim \Delta^{1-\beta/2}$.

2. $\mathbb{P}(\Omega_{X,k} \cap \Omega_{N,k}^c \cap (X_{k\Delta} \in A)) \lesssim \Delta^{2-\beta/2}$.

3. $\left(\mathbb{E}\left[(Z_{k\Delta} + T_{k\Delta})\,\mathbb{1}_{\Omega_{N,k}\cap\Omega_{X,k}\cap(X_{k\Delta}\in A)}\middle|\,\mathscr{F}_{k\Delta}\right]\right)^2 \lesssim \ln^2(n)\Delta^{1-\beta/2}$.

According to Lemma 12, $\mathbb{E}(I_{k\Delta}^2) \le \Delta$. As b is bounded on the compact set A, $\mathbb{E}\Big[b_A^2(X_{k\Delta})\mathbb{1}_{\Omega_{X,k}^c}\Big] \lesssim \mathbb{P}(\Omega_{X,k}^c) \lesssim \Delta^{1-\beta/2}$. Moreover, on $\Omega_{X,k}$,

$$\begin{aligned}(Z_{k\Delta} + T_{k\Delta})^2\,\mathbb{1}_{\Omega_{X,k}\cap(X_{k\Delta}\in A)} &= \left(\frac{X_{(k+1)\Delta} - X_{k\Delta}}{\Delta} - b_A(X_{k\Delta}) - I_{k\Delta}\right)^2\mathbb{1}_{\Omega_{X,k}}\mathbb{1}_{X_{k\Delta}\in A}\\ &\lesssim \frac{\ln^2(n)}{\Delta} + b_A^2(X_{k\Delta}) + I_{k\Delta}^2\end{aligned}$$

and then

$$
\begin{aligned}
E &:= \mathbb{E}\left[(Z_{k\Delta} + T_{k\Delta})^2 \mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}^c\cap(X_{k\Delta}\in A)}\right] \\
&\lesssim \left(\frac{\ln^2(n)}{\Delta} + b_{\max}^2\right)\mathbb{P}\left(\Omega_{X,k}\cap\Omega_{N,k}^c\cap(X_{k\Delta}\in A)\right) + \mathbb{E}\left(I_{k\Delta}^2\right) \\
&\lesssim \ln^2(n)\Delta^{1-\beta/2}.
\end{aligned}
$$

It remains to bound $\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\tilde{\nu}_n^2(t)\right)$. In the same way as in Section 6.1, we get:

$$
\begin{aligned}
\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\tilde{\nu}_n^2(t)\right) &\le \sum_{\lambda\in\Lambda_m}\mathbb{E}\left(\tilde{\nu}_n^2(\varphi_\lambda)\right) \le \frac{2D_m}{n}\mathbb{E}\left(\tilde{Z}_\Delta^2 + \tilde{T}_\Delta^2\right) \\
&\le \frac{2D_m}{n}\mathbb{E}\left(Z_\Delta^2 + T_\Delta^2\right) \le 2\left(\sigma_0^2 + \xi_0^2\right)\frac{D_m}{n\Delta}.
\end{aligned}
$$

We have that $\mathbb{E}\left(\tilde{Z}_\Delta^2\right) \le \mathbb{E}\left(Z_\Delta^2\right) \le \frac{\sigma_0^2}{\Delta}$. Moreover,

$$
\begin{aligned}
\mathbb{E}\left(\tilde{T}_{k\Delta}^2\right) &\lesssim \mathbb{E}\left(T_{k\Delta}^2\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}}\right) - \left(\mathbb{E}\left(T_{k\Delta}\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}}\right)\right)^2 \\
&\lesssim \mathbb{E}\left(T_{k\Delta}^2\mathbb{1}_{\Omega_{N,k}}\right) + \ln^2(n)\Delta^{1-\beta/2} \\
&\lesssim \Delta^{1/2-\beta/4}.
\end{aligned}
$$

Then $\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\tilde{\nu}_n^2(t)\right) \le (n\Delta)^{-1}D_m(\sigma_0^2 + o(1))$.

6.3.1. *Proof of Lemma 16*

Result 17. *Let β be the Blumenthal–Getoor index of L_t. Then:*

$$
\nu([-z,z]^c) \lesssim z^{-\beta}, \qquad \int_{|x|\le z\wedge a_0} x^2\nu(dx) \lesssim z^{2-\beta} \quad \textit{and} \quad \int_{|x|\le z\wedge a_0} x^4\nu(dx) \lesssim z^{4-\beta}.
$$

The constant a_0 is defined in A6.

Bound of $\mathbb{P}(\Omega_{X,k}^c \cap (X_{k\Delta}\in A))$. We have:

$$
\mathbb{P}\left(\Omega_{X,k}^c\cap(X_{k\Delta}\in A)\right) = \mathbb{P}\left(\left\{\left|X_{(k+1)\Delta} - X_{k\Delta}\right| > C_\Delta\right\}\cap(X_{k\Delta}\in A)\right).
$$

We know that $X_{(k+1)\Delta} - X_{k\Delta} = b(X_{k\Delta}) + I_{k\Delta} + Z_{k\Delta} + T_{k\Delta}$. Then

$$
\begin{aligned}
\mathbb{P}\left(\Omega_{X,k}^c\cap(X_{k\Delta}\in A)\right) &\le \mathbb{P}\left(|\Delta I_{k\Delta}| \ge \Delta\right) \\
&\quad + \mathbb{P}\left(|\Delta Z_{k\Delta}| \ge \sigma_0\sqrt{\Delta}\ln(n)\right) + \mathbb{P}\left(|\Delta T_{k\Delta}| \ge \xi_0\sqrt{\Delta}\ln(n)\right).
\end{aligned}
$$

By a Markov inequality and Lemma 12, we obtain:

$$
\mathbb{P}\left(|\Delta I_{k\Delta}| \ge \Delta\right) \le \frac{\mathbb{E}\left(\Delta^2 I_{k\Delta}^2\right)}{\Delta^2} \lesssim \Delta. \tag{6}
$$

By Proposition 14, the process $\exp\left(c\int_0^t \sigma(X_{s-})dW_s - c^2\int_0^t \sigma^2(X_s)ds\right)$ is a local martingale (as σ is bounded, it is in fact a martingale, see [12, pp. 229–232]). Then, by a Markov inequality:

$$\mathbb{P}\left(|\Delta Z_{k\Delta}| \geq \sigma_0\sqrt{\Delta}\ln(n)\right) \leq \frac{2}{n}\mathbb{E}\left[\exp\left(\frac{\sqrt{\Delta}Z_{k\Delta}}{\sigma_0}\right)\right] \lesssim \frac{1}{n}. \tag{7}$$

To bound inequality (5), it remains to bound $\mathbb{P}\left(|\Delta T_{k\Delta}| \geq \xi_0\sqrt{\Delta\ln(n)}\right)$. Let us set

$$T_{k\Delta} = T_{k\Delta}^{(1)} + T_{k\Delta}^{(2)} + T_{k\Delta}^{(3)} \quad \text{where } T_{k\Delta}^{(i)} = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} \xi(X_{s-})dL_s^{(i)}$$

with

$$L_t^{(1)} = \int_0^t \int_{\left[-\sqrt{\Delta},\sqrt{\Delta}\right]} z\tilde{\mu}(ds,dz), \qquad L_t^{(2)} = \int_0^t \int_{\left[-\Delta^{1/4},-\sqrt{\Delta}\right]\cup\left[\sqrt{\Delta},\Delta^{1/4}\right]} z\tilde{\mu}(ds,dz)$$
$$L_t^{(3)} = \int_0^t \int_{[-\Delta^{1/4},\Delta^{1/4}]^c} z\tilde{\mu}(ds,dz).$$

Let us set $N''_{k\Delta} = \mu\left(]k\Delta,(k+1)\Delta],\left[-\sqrt{\Delta},\sqrt{\Delta}\right]^c\right)$. By Result 17, we have:

$$\mathbb{P}\left(\left|T_{k\Delta}^{(2)} + T_{k\Delta}^{(3)}\right| > 0\right) = \mathbb{P}\left(N''_{k\Delta} \geq 1\right) \lesssim \Delta\nu\left(\left[-\sqrt{\Delta},\sqrt{\Delta}\right]^c\right) \lesssim \Delta^{1-\beta/2}.$$

It remains to bound $\mathbb{P}\left[\left|\Delta T_{k\Delta}^{(1)}\right| \geq 2\xi_0\sqrt{\Delta}\ln(n)\right]$. We have that:

$$\mathbb{P}\left[\left|\Delta T_{k\Delta}^{(1)}\right| \geq 2\xi_0\sqrt{\Delta}\ln(n)\right] \leq 2\mathbb{P}\left[\exp\left(\varepsilon\int_{k\Delta}^{(k+1)\Delta}\xi(X_{s-})dL_s^{(1)}\right) \geq n^{2\varepsilon\xi_0\sqrt{\Delta}}\right].$$

By Proposition 14, for any ε,

$$D_t := \exp\left(\varepsilon\int_{k\Delta}^t \xi(X_{s-})dL_s^{(1)} - \int_{k\Delta}^t\int_{|z|\leq\sqrt{\Delta}} (\exp(\varepsilon z\xi(X_{s-}) - 1 - \varepsilon z\xi(X_{s-})))\,\nu(dz)\right)$$

is a local martingale. Let us set $\varepsilon = 1/(2\xi_0\Delta^{1/2})$. There exists an increasing sequence of stopping times τ_N such that, for any N,

$$\begin{aligned}
F &:= \mathbb{P}\left[\exp\left(\frac{1}{2\xi_0\Delta^{1/2}}\int_{k\Delta}^{(k+1)\Delta\wedge\tau_N}\xi(X_{s-})dL_s^{(1)}\right) \geq n\right] \\
&\leq n^{-1}\mathbb{E}\left(\exp\left(\int_{k\Delta}^{(k+1)\Delta\wedge\tau_N}\int_{|z|\leq\sqrt{\Delta}}\left(\exp\left(\frac{z\xi(X_{s-})}{2\xi_0\Delta^{1/2}}\right) - 1 - \frac{z\xi(X_{s-})}{2\xi_0\Delta^{1/2}}\right)\nu(dz)\right)\right) \\
&\leq n^{-1}\exp\left(2\Delta\int_{|z|\leq\sqrt{\Delta}}\frac{\xi_0^2z^2}{4\xi_0^2\Delta}\nu(dz)\right) \leq n^{-1}\exp\left(\int_{\mathbb{R}} z^2\nu(dz)\right) \leq n^{-1}.
\end{aligned}$$

When $N \to \infty$, by dominated convergence, we obtain:

$$\mathbb{P}\left(\left|\Delta T_{k\Delta}^{(1)}\right| \geq \xi_0\sqrt{\Delta}\ln(n)\right) \lesssim n^{-1}. \tag{8}$$

Bound of $\mathbb{P}\left(\Omega_{X,k}\cap\Omega_{N,k}^c\cap(X_{k\Delta}\in A)\right)$. We recall that $N'_{k\Delta}=\mu\left(]k\Delta,(k+1)\Delta],[-\Delta^{1/4},\Delta^{1/4}]^c\right)$. We have:

$$\Omega_{N,k}^c=\left\{N'_{k\Delta}=1\right\}\cup\left\{N'_{k\Delta}\geq 2\right\}$$

with

$$\mathbb{P}\left(N'_{k\Delta}=1\right)\lesssim\Delta^{1-\beta/4}\quad\text{and}\quad\mathbb{P}\left(N'_{k\Delta}\geq 2\right)\lesssim\Delta^{2-\beta/2}.$$

Then $\mathbb{P}\left(\Omega_{N,k}^c\cap\left\{N'_{k\Delta}\geq 2\right\}\right)\lesssim\Delta^{2-\beta/2}$. We can write:

$$\begin{aligned}G &:=\mathbb{P}\left(\Omega_{X,k}\cap(X_{k\Delta}\in A)\cap(N'_{k\Delta}=1)\right)\\&\leq\mathbb{P}\left(N'_{k\Delta}=1\right)\mathbb{P}\left(\left|\Delta T_{k\Delta}^{(2)}+\Delta T_{k\Delta}^{(3)}\right|\leq 2C_\Delta\middle|N'_{k\Delta}=1\right)\\&\quad+\mathbb{P}\left(N'_{k\Delta}=1\right)\mathbb{P}\left(\left\{\left|\Delta T_{k\Delta}^{(2)}+\Delta T_{k\Delta}^{(3)}\right|\geq 2C_\Delta\middle|N'_{k\Delta}=1\right\}\cap\Omega_{X,k}\cap(X_{k\Delta}\in A)\right).\end{aligned}$$

By (6)–(8), we obtain:

$$\begin{aligned}H &:=\mathbb{P}\left(\left\{\left|\Delta T_{k\Delta}^{(2)}+\Delta T_{k\Delta}^{(3)}\right|\geq 2C_\Delta\middle|N'_{k\Delta}=1\right\}\cap\Omega_{X,k}\cap(X_{k\Delta}\in A)\right)\\&\leq\mathbb{P}\left(\Delta\left|b_A(X_{k\Delta})+I_{k\Delta}+Z_{k\Delta}+T_{k\Delta}^{(1)}\right|>C_\Delta\right)\\&\lesssim\Delta+n^{-1}.\end{aligned}$$

It remains to bound $J:=\mathbb{P}\left(\left|\Delta T_{k\Delta}^{(2)}+\Delta T_{k\Delta}^{(3)}\right|\leq 2C_\Delta|\,N'_{k\Delta}=1\right)$. If $N'_{k\Delta}=1$, then $\left|\Delta T_{k\Delta}^{(3)}\right|=\left|\int_{k\Delta}^{(k+1)\Delta}\xi(X_{s^-})dL_s^{(3)}\right|\geq\xi_1\Delta^{1/4}$. Then $J\leq\mathbb{P}\left(\Delta\left|T_{k\Delta}^{(2)}\right|\geq\xi_1\Delta^{1/4}-2C_\Delta\right)$. Let us set $n_0=\left\lceil\frac{1}{1-\beta/2}\right\rceil$ and $a=(\xi_0 n_0)^{-1}\left(\xi_1\Delta^{1/4}-2C_\Delta\right)$. We have:

$$\begin{aligned}J &\leq\mathbb{P}\left[\mu(]k\Delta,(k+1)\Delta],[-a,a]^c)\geq 1\right]\\&\quad+\mathbb{P}\left[\mu(]k\Delta,(k+1)\Delta],[-a,-\Delta^{1/2}]\cup[\Delta^{1/2},a])\geq n_0\right]\\&\leq\Delta\nu([-a,a]^c)+\Delta^{n_0}\nu([-\Delta^{1/2},\Delta^{1/2}]^c)^{n_0}\\&\lesssim\Delta^{1-\beta/4}+\Delta.\end{aligned}$$

Then $\mathbb{P}(\Omega_{X,k}\cap\Omega_{N,k}^c)\leq\mathbb{P}(N'_{k\Delta}=1)\Delta^{1-\beta/4}+\mathbb{P}(N'_{k\Delta}=2)\lesssim\Delta^{2-\beta/2}$.

Bound of $\left(\mathbb{E}\left[(Z_{k\Delta}+T_{k\Delta})\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)}\middle|\mathscr{F}_{k\Delta}\right]\right)^2$.

If σ and ξ are constants. Let us set $E:=\left(\mathbb{E}\left[(Z_{k\Delta}+T_{k\Delta})\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)}\middle|\mathscr{F}_{k\Delta}\right]\right)^2$ and

$$\Omega_{I,k}=\left\{\omega,|I_{k\Delta}|\leq 1,\cap|\Delta Z_{k\Delta}|\leq\sigma_0\sqrt{\Delta}\ln(n),\cap\left|\Delta T_{k\Delta}^{(1)}\right|\leq 2\xi_0\sqrt{\Delta}\ln(n)\right\}.$$

By (6)–(8), $\mathbb{P}\left(\Omega_{I,k}^c\right)\leq\Delta+n^{-1}$. Then, by a Markov inequality:

$$E\lesssim\Delta\ln^2(n)+\left(\mathbb{E}\left[(Z_{k\Delta}+T_{k\Delta})\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap\Omega_{I,k}\cap(X_{k\Delta}\in A)}\middle|\mathscr{F}_{k\Delta}\right]\right)^2.$$

Let us introduce the set $\Omega_{ZT,k}:=\left\{\omega,|Z_{k\Delta}+T_{k\Delta}|\leq C_\Delta\Delta^{-1}-b_{\max}-1\right\}$. On $\Omega_{I,k}$, $|I_{k\Delta}|\leq 1$ and therefore:

$$\Omega_{ZT,k}\cap\Omega_{I,k}\subseteq\Omega_{X,k}\cap\Omega_{I,k}\subseteq\left\{\omega,|Z_{k\Delta}+T_{k\Delta}|\leq C_\Delta\Delta^{-1}+b_{\max}+1\right\}\cap\Omega_{I,k}.$$

Then

$$E \lesssim \Delta \ln^2(n) + F^2 + G^2$$

where $F = \mathbb{E}\left[(Z_{k\Delta} + T_{k\Delta}) \mathbb{1}_{\Omega_{ZT,k} \cap \Omega_{N,k} \cap \Omega_{I,k} \cap (X_{k\Delta} \in A)} \middle| \mathscr{F}_{k\Delta} \right]$ and $G = \mathbb{E}\Big[(Z_{k\Delta} + T_{k\Delta}) \mathbb{1}_{\Omega^c_{ZT,k} \cap \Omega_{X,k} \cap \Omega_{N,k} \cap \Omega_{I,k} \cap (X_{k\Delta} \in A)} \Big| \mathscr{F}_{k\Delta} \Big]$. As σ and ξ are constants, the terms

$$Z_{k\Delta} = \frac{\sigma_0}{\Delta} \int_{k\Delta}^{(k+1)\Delta} dW_s \quad \text{and} \quad T_{k\Delta} = \frac{\xi_0}{\Delta} \int_{k\Delta}^{(k+1)\Delta} dL_s$$

are centred and independent. Then $F = 0$. Moreover, on $\Omega_{N,k}$, $T^{(3)}_{k\Delta} = 0$. Then

$$|G| \lesssim \left| \mathbb{E}\left[\left(Z_{k\Delta} + T^{(1)}_{k\Delta} + T^{(2)}_{k\Delta} \right) \mathbb{1}_{\Omega_{X,k} \cap \Omega^c_{ZT,k} \cap \Omega_{N,k} \cap \Omega_{I,k} \cap (X_{k\Delta} \in A)} \middle| \mathscr{F}_{k\Delta} \right] \right|.$$

Let us set $c_b = b_{\max} + 1$. On $\Omega_{I,k} \cap \Omega_{X,k}$, $\left| Z_{k\Delta} + T^{(1)}_{k\Delta} + T^{(2)}_{k\Delta} \right| \lesssim \ln(n) \Delta^{-1/2}$, and

$$\begin{aligned} |G| &\lesssim \frac{\ln(n)}{\sqrt{\Delta}} \left(\mathbb{P}\left(\left| Z_{k\Delta} + T^{(1)}_{k\Delta} + T^{(2)}_{k\Delta} \right| \in \left[C_\Delta \Delta^{-1} - c_b, C_\Delta \Delta^{-1} + c_b \right] \mathbb{1}_{\Omega_{I,k}} \right) \right) \\ &= 2 \frac{\ln(n)}{\sqrt{\Delta}} \int_{\mathbb{R}} \mathbb{P}\left(T^{(2)}_{k\Delta} \in \left[C_\Delta \Delta^{-1} - c_b - x, C_\Delta \Delta^{-1} + c_b - x \right] \mathbb{1}_{\Omega_{I,k}} \right) \\ &\quad \times \mathbb{P}\left(Z_{k\Delta} + T^{(1)}_{k\Delta} \in dx \middle| T^{(2)}_{k\Delta} \in \left[C_\Delta \Delta^{-1} - c_b - x, C_\Delta \Delta^{-1} + c_b - x \right] \mathbb{1}_{\Omega_{I,k}} \right). \end{aligned}$$

On $\Omega_{I,k}$, $\left| Z_{k\Delta} + T^{(1)}_{k\Delta} \right| \leq (\sigma_0 + 2\xi_0) \ln(n) \Delta^{-1/2}$. Then

$$|G| \lesssim \frac{\ln(n)}{\sqrt{\Delta}} \left[\sup_{C \geq \xi_0 \ln(n) \Delta^{-1/2}} \mathbb{P}\left(T^{(2)}_{k\Delta} \in [C, C + 2c_b] \right) \right]. \tag{9}$$

We recall that $L^{(2)}_t$ is a compound Poisson process in which all the jumps are greater than $\sqrt{\Delta}$ and smaller than $\Delta^{1/4}$. Let us denote by τ_i the times of the jumps of size in $\left[\sqrt{\Delta}, \Delta^{1/4} \right]$ and by ζ_i the size of the jumps. We set $a_j = \xi_0^{-1} C\Delta - \sum_{i=1}^{j-1} \zeta_i$ and $c := \xi_0^{-1}(2b_{\max} + 2)$. Then, as ξ is constant equal to ξ_0:

$$\begin{aligned} H &:= \mathbb{P}\left(T^{(2)}_{k\Delta} \in [C, C + 2b_{\max} + 2] \right) \\ &\leq \sum_{j=1}^{\infty} \mathbb{P}\left(j \text{ jumps } \geq \sqrt{\Delta}, \text{last jump } \in [a_j, a_j + c\Delta] \right) \\ &\lesssim 2 \sup_{a \geq \sqrt{\Delta}} \mathbb{P}(1 \text{ jump } \in [a, a + c\Delta]) = 2\Delta \sup_{a \geq \sqrt{\Delta}} \nu([a, a + c\Delta]). \end{aligned}$$

By A6,

$$H \lesssim \Delta \sup_{a \geq \sqrt{\Delta}} \left[\frac{1}{a^\beta} - \frac{1}{(a + c\Delta)^\beta} \right] \lesssim \sqrt{\Delta} \Delta^{1-\beta/2} \tag{10}$$

and, by (9) and (10),

$$E \lesssim \Delta \ln^2(n) + \frac{\ln^2(n)}{\Delta} \Delta \Delta^{2-\beta} \lesssim \Delta \ln^2(n) + \Delta^{2-\beta} \ln^2(n).$$

Remark 18. If ν is not absolutely continuous, inequality (10) is not valid. We obtain:

$$H \lesssim 2\Delta \sup_{a \geq \sqrt{\Delta}} \nu([a, a + c\Delta]) \lesssim \Delta^{1-\beta/2}.$$

Therefore

$$E \leq \Delta \ln^2(n) + G^2 \lesssim \Delta \ln^2(n) + \Delta^{1-\beta} \ln^2(n).$$

If σ or ξ are not constants. The problem is that $Z_{k\Delta}$ and $T_{k\Delta}$ are not symmetric and we cannot apply directly the previous method. We replace them by two centred terms. The following lemma is very useful.

Lemma 19. *Let f be a $\mathscr{C}^2$ function such that f and f' are Lipschitz. Let us set, for any $t \in]k\Delta, (k+1)\Delta]$:*

$$\psi_f(X_{k\Delta}, t) = f'(X_{k\Delta}) \left(\sigma(X_{k\Delta}) \int_{k\Delta}^{t} dW_s + \xi(X_{k\Delta}) \int_{k\Delta}^{t} z\tilde{\mu}(ds, dz) \right).$$

We have:

$$\mathbb{E}\left[\left(f(X_t) - f(X_{k\Delta}) - \psi_f(X_{k\Delta}, t) \right)^2 \mathbb{1}_{\Omega_{N,k}} \mathbb{1}_{X_{k\Delta} \in A} \right] \lesssim \Delta^{2-\beta/4}.$$

Lemma 19 is proved below. Let us set

$$\bar{Z}_{k\Delta} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} \left(\sigma(X_{k\Delta}) + \psi_\sigma(X_{k\Delta,s}) \right) dW_s,$$
$$\bar{T}^{(i)}_{k\Delta} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} \left(\xi(X_{k\Delta}) + \psi_\xi(X_{k\Delta,s}) \right) dL^{(i)}_s \quad \text{and} \quad \bar{T}_{k\Delta} = \bar{T}^{(1)}_{k\Delta} + \bar{T}^{(2)}_{k\Delta} + \bar{T}^{(3)}_{k\Delta}.$$

The terms $\bar{Z}_{k\Delta}$ and $\bar{T}_{k\Delta}$ are symmetric. By Lemma 19,

$$\begin{aligned} &\mathbb{E}\left[\left(\bar{Z}_{k\Delta} - Z_{k\Delta} \right)^2 \mathbb{1}_{\Omega_{N,k}} \mathbb{1}_{X_{k\Delta} \in A} \right] \\ &= \frac{1}{\Delta^2} \mathbb{E}\left[\int_{k\Delta}^{(k+1)\Delta} \left(\sigma(X_s) - \sigma(X_{k\Delta}) - \psi_\sigma(X_{k\Delta,s}) \right)^2 ds \right] \\ &\lesssim \Delta^{1-\beta/4}. \end{aligned} \tag{11}$$

We prove in the same way that

$$\mathbb{E}\left[\left(\bar{T}_{k\Delta} - T_{k\Delta} \right)^2 \mathbb{1}_{\Omega_{N,k}} \mathbb{1}_{X_{k\Delta} \in A} \right] \leq \Delta^{1-\beta/4}. \tag{12}$$

Let us set $U_{k\Delta} = \Delta^{-1} \xi(X_{k\Delta^-}) \int_{k\Delta}^{(k+1)\Delta} dL^{(2)}_s$. By Result 11 and Proposition 1,

$$\mathbb{E}\left[\Delta^2 \left(\bar{T}^{(2)}_{k\Delta} - U_{k\Delta} \right)^2 \right] = \mathbb{E}\left[\int_{k\Delta}^{(k+1)\Delta} \int_{\mathbb{R}} \left(\psi_\xi(X_{k\Delta,s}) \right)^2 z^2 \nu(dz) ds \right] \leq \Delta^{2-\beta/4}. \tag{13}$$

Let us introduce the set

$$\bar{\Omega}_{I,k} = \left\{\omega, |I_{k\Delta}| + \left|Z_{k\Delta} - \bar{Z}_{k\Delta}\right| + \left|T_{k\Delta} - \bar{T}_{k\Delta}\right| \le 3\right\}$$
$$\bigcap \left\{\left|\Delta \bar{Z}_{k\Delta}\right| \le \sigma_0\sqrt{\Delta}\ln(n) + \Delta, \left|\Delta \bar{T}^{(1)}_{k\Delta}\right| \le 2\xi_0\sqrt{\Delta}\ln(n) + \Delta\right\}$$
$$\bigcap \left\{\left|\Delta(\bar{T}^{(2)}_{k\Delta} - U_{k\Delta})\right| \le \xi_0\sqrt{\Delta}\right\}.$$

By (6)–(8), (11)–(13) and Markov inequalities, we obtain:

$$\mathbb{P}\left(\bar{\Omega}^c_{I,k}\right) \lesssim \Delta^{1-\beta/4} + \frac{1}{n}. \tag{14}$$

Then

$$E := \left(\mathbb{E}\left[(Z_{k\Delta} + T_{k\Delta})\, \mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)}\middle| \mathscr{F}_{k\Delta}\right]\right)^2$$
$$\lesssim \Delta^{1-\beta/2}\ln^2(n) + \left(\mathbb{E}\left[\left(\bar{Z}_{k\Delta} + \bar{T}_{k\Delta}\right) \mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)\cap\bar{\Omega}_{I,k}}\middle| \mathscr{F}_{k\Delta}\right]\right)^2. \tag{15}$$

Let us introduce the set:

$$\bar{\Omega}_{ZT,k} := \left\{\omega, \left|\bar{Z}_{k\Delta} + \bar{T}_{k\Delta}\right| \le C_\Delta \Delta^{-1} - b_{\max} - 3\right\}.$$

We have that

$$\bar{\Omega}_{ZT,k} \cap \bar{\Omega}_{I,k} \subseteq \Omega_{X,k} \cap \bar{\Omega}_{I,k} \subseteq \left\{\omega, \left|\bar{Z}_{k\Delta} + \bar{T}_{k\Delta}\right| \le C_\Delta\Delta^{-1} + b_{\max} + 3\right\} \cap \bar{\Omega}_{I,k}.$$

Given the filtration $\mathscr{F}_{k\Delta}$, the sum $\bar{Z}_{k\Delta} + \bar{T}_{k\Delta}$ is symmetric. Then

$$\mathbb{E}\left[\left(\bar{Z}_{k\Delta} + \bar{T}_{k\Delta}\right) \mathbb{1}_{\bar{\Omega}_{ZT,k}\cap\Omega_{N,k}\cap(X_{k\Delta}\in A)}\middle| \mathscr{F}_{k\Delta}\right] = 0.$$

Moreover, on $\Omega_{N,k}$, $\bar{T}^{(3)}_{k\Delta} = 0$. Then, by (15),

$$E \lesssim \Delta^{1-\beta/2}\ln^2(n) + G^2 + H^2$$

where $G := \mathbb{E}\left[\left(\bar{Z}_{k\Delta} + \bar{T}^{(1)}_{k\Delta} + \bar{T}^{(2)}_{k\Delta}\right) \mathbb{1}_{\Omega_{X,k}\cap\Omega^c_{ZT,k}\cap\Omega_{N,k}\cap\Omega_{I,k}\cap(X_{k\Delta}\in A)}\middle| \mathscr{F}_{k\Delta}\right]$ and $H := \mathbb{E}\left[\left(\bar{Z}_{k\Delta} + \bar{T}^{(1)}_{k\Delta} + \bar{T}^{(2)}_{k\Delta}\right) \mathbb{1}_{\Omega_{X,k}\cap\Omega_{ZT,k}\cap\Omega_{N,k}\cap\Omega^c_{I,k}\cap(X_{k\Delta}\in A)}\middle| \mathscr{F}_{k\Delta}\right]$. We have that $H^2 \lesssim \Delta^{-1}\ln^2(n)$ $\mathbb{P}^2(\Omega^c_{I,k}) \lesssim \Delta^{1-\beta/2}\ln^2(n)$. The end of the proof is the same as in the case of σ and ξ constants. We obtain that

$$|G| \lesssim \frac{\ln(n)}{\sqrt{\Delta}} \sup_{C\ge\kappa_0\ln(n)\Delta^{-1/2}} \mathbb{P}\left(U_{k\Delta} \in [C, C + 2b_{\max} + 6]\right) \lesssim \sqrt{\Delta}\Delta^{1-\beta/2}.$$

6.3.2. Proof of Lemma 19

According to the Itô formula (see for instance [1, Theorem 4.4.7 p. 251]), we have that

$$f(X_t) - f(X_{k\Delta}) = I_1 + I_2 + I_3 + I_4$$

where

$$I_1 = \int_{k\Delta}^t f'(X_s)\sigma(X_s)dW_s,$$

$$I_2 = \int_{k\Delta}^{t}\int_{\mathbb{R}} (f(X_{s^-} + z\xi(X_{s^-})) - f(X_{s^-}))\,\tilde{\mu}(ds, dz)$$
$$I_3 = \int_{k\Delta}^{t}\int_{z\in\mathbb{R}} \left[f(X_s + z\xi(X_s)) - f(X_s) - z\xi(X_s)f'(X_s)\right]\nu(dz)ds$$
$$I_4 = \int_{k\Delta}^{t}\left[f'(X_s)b(X_s) + f''(X_s)\sigma^2(X_s)/2\right]ds.$$

By Proposition 1, for any $t \le (k+1)\Delta$, we have:

$$\begin{aligned} Q &:= \mathbb{E}\left[\left(I_1 - f'(X_{k\Delta})\sigma(X_{k\Delta})\int_{k\Delta}^{t} dW_s\right)^2\right] \\ &= \mathbb{E}\left[\left(\int_{k\Delta}^{t}\left(\sigma(X_s)f'(X_s) - \sigma(X_{k\Delta})f'(X_{k\Delta})\right)dW_s\right)^2\right] \\ &= \int_{k\Delta}^{t}\left(\sigma(X_s)f'(X_s) - \sigma(X_{k\Delta})f'(X_{k\Delta})\right)^2 ds \lesssim \Delta^2. \end{aligned}$$

We can write:

$$\begin{aligned} E &:= \mathbb{E}\left[\left(I_2 - f'(X_{k\Delta})\xi(X_{k\Delta^-})\int_{k\Delta}^{t} dL_s^{(1)} + dL_s^{(2)}\right)^2 \mathbb{1}_{\Omega_{N,k}}\right] \\ &\le 2\int_{k\Delta}^{t}\int_{|z|\le\Delta^{1/4}} \mathbb{E}\left[\left(f(X_s + z\xi(X_s)) - f(X_s) - z\xi(X_s)f'(X_s)\right)^2\right]\nu(dz)ds \\ &\quad + 2\int_{k\Delta}^{t}\int_{|z|\le\Delta^{1/4}} \mathbb{E}\left[z^2\left(\xi(X_s)f'(X_s) - \xi(X_{k\Delta})f'(X_{k\Delta})\right)^2\right]\nu(dz)ds. \end{aligned}$$

The function f is $\mathscr{C}^2$; then, by the Taylor formula, for any $s \in [k\Delta, t]$, $z \in \mathbb{R}$, there exists $\zeta_{s,z}$ in $[X_s, X_s + z\xi(X_s)]$ such that:

$$f(X_s + z\xi(X_s)) - f(X_s) - z\xi(X_s)f'(X_s) = \frac{z^2\xi^2(X_s)}{2} f''(\zeta_{s,z}).$$

Then, as ξ and f'' are bounded:

$$\mathbb{E}\left[\left(f(X_s) + z\xi(X_s) - f(X_s) - z\xi(X_s)f'(X_s)\right)^2\right] = \frac{z^4}{4}\mathbb{E}\left[\left(\xi(X_s)f''(\zeta_{s,z})\right)^2\right] \lesssim z^4$$

and, by Result 17, for any $t \le (k+1)\Delta$,

$$\begin{aligned} F &:= \int_{k\Delta}^{t}\int_{|z|\le\Delta^{1/4}} \mathbb{E}\left[\left(f(X_s) + z\xi(X_s) - f(X_s) - z\xi(X_s)f'(X_s)\right)^2\right]\nu(dz)ds \\ &\lesssim \Delta\int_{|z|\le\Delta^{1/4}} z^4\nu(dz) \lesssim \Delta^{2-\beta/4}. \end{aligned}$$

The functions ξ and f' are Lipschitz; then by Proposition 1,

$$\mathbb{E}\left[z^2\left(\xi(X_s)f'(X_s)-\xi(X_{k\Delta})f'(X_{k\Delta})\right)^2\right]\lesssim z^2\mathbb{E}\left[(X_s-X_{k\Delta})^2\right]\lesssim \Delta z^2$$

and consequently, for any $t\le(k+1)\Delta$:

$$\int_{k\Delta}^t\int_{|z|\le\Delta^{1/4}}\mathbb{E}\left[z^2\left(\xi(X_s)f'(X_s)-\xi(X_{k\Delta})f'(X_{k\Delta})\right)^2\right]\nu(dz)ds\lesssim\Delta^{2-\beta/4};$$

then $E\lesssim\Delta^{2-\beta/4}$. By the same way, we obtain that

$$\mathbb{E}\left[I_3^2\right]\le\mathbb{E}\left[\int_{k\Delta}^t\int_{|z|\le\Delta^{1/4}}\left(\frac{z^2\xi^2(X_s)}{2}f''(\zeta_{s,z})\right)^2\nu(dz)ds\right]\lesssim\Delta^{2-\beta/4}.$$

The functions b and f' are Lipschitz and f'' and σ are bounded; then, for any $t\le(k+1)\Delta$:

$$\mathbb{E}\left[I_4^2\right]\lesssim\Delta\int_{k\Delta}^t\left(1+\mathbb{E}\left[X_s^4\right]\right)ds\lesssim\Delta^2.$$

Then, for any $t\le(k+1)\Delta$:

$$\mathbb{E}\left[\left(f(X_t)-f(X_{k\Delta})-\psi_f(X_{k\Delta},t)\right)\right]\le\Delta^{2-\beta/4}.$$

6.4. Proof of Theorem 10

As previously, we only bound the risk on Ω_n. As in Section 6.2, we introduce the function $p(m,m')$ such that $p(m,m')=12(pen(m)+pen(m'))$. On Ω_n, for any $m\in\mathscr{M}_n$, we have:

$$\begin{aligned}\left\|\tilde b_{\tilde m}-b_A\right\|_n^2\le\ &3\left\|b_m-b_A\right\|_n^2+\frac{224}{n}\sum_{k=1}^n b_A^2(X_{k\Delta})\mathbb{1}_{\Omega_{X,k}^c}+I_{k\Delta}^2\\&+2\left(Z_{k\Delta}^2+T_{k\Delta}^2\right)\mathbb{1}_{\Omega_{X,k}\cap\Omega_{Z,k}^c}\\&+\frac{224}{n}\sum_{k=1}^n\left(\mathbb{E}\left[(Z_{k\Delta}+T_{k\Delta})\mathbb{1}_{\Omega_{X,k}\cap\Omega_{Z,k}}\middle|\mathscr{F}_{k\Delta}\right]\right)^2\\&+24\sup_{t\in\mathscr{B}_{m,\hat m}}\left(\tilde\nu_n^2(t)-p(m,\tilde m)\right)+4pen(m).\end{aligned}$$

It remains only to bound

$$\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,\hat m}}\left(\tilde\nu_n^2(t)-p(m,\tilde m)\right)\right]\le\sum_{m'}\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,m'}}\left(\tilde\nu_n^2(t)-p(m,\tilde m)\right)\right].$$

As in the proof of Theorem 4, we bound the quantity

$$\mathbb{E}\left[\exp\left(\varepsilon t(X_{k\Delta})\left(\tilde Z_{k\Delta}+\tilde T_{k\Delta}\right)\right)\middle|\mathscr{F}_{k\Delta}\right].$$

We have that

$$\mathbb{E}\left[\exp\left(\varepsilon t(X_{k\Delta})Z_{k\Delta}\right)\mathbb{1}_{\Omega_{N,k}}\middle|\mathscr{F}_{k\Delta}\right]\le\exp\left(\frac{\varepsilon^2\sigma_0^2t^2(X_{k\Delta})}{2\Delta}\right).$$

The truncated Lévy process $\tilde{L}_t = \int_0^t \int_{|z|\leq \Delta^{1/4}} z\tilde{\mu}(ds,dz)$ satisfies Assumption A5 and then there exists a constant c such that:

$$\mathbb{E}\left[\exp\left(\varepsilon t(X_{k\Delta})T_{k\Delta}\right)\mathbb{1}_{\Omega_{N,k}}\middle|\mathscr{F}_{k\Delta}\right] \leq \exp\left(\frac{c\varepsilon^2\xi_0^2 t^2(X_{k\Delta})}{\Delta\left(1-\varepsilon/\varepsilon_1\right)}\right).$$

As $Z_{k\Delta}\mathbb{1}_{\Omega_{N,k}}$ and $T_{k\Delta}\mathbb{1}_{\Omega_{N,k}}$ are centred, we obtain:

$$\mathbb{E}\left[\exp\left(\varepsilon\left|t(X_{k\Delta})\left(Z_{k\Delta}+T_{k\Delta}\right)\right|\right)\mathbb{1}_{\Omega_{N,k}}\middle|\mathscr{F}_{k\Delta}\right] \leq 2\exp\left(\frac{c\varepsilon^2\left(\sigma_0^2+\xi_0^2\right)t^2(X_{k\Delta})}{\Delta\left(1-\varepsilon/\varepsilon_1\right)}\right)$$

and then

$$\mathbb{E}\left[\exp\left(\varepsilon\left|t(X_{k\Delta})\left(\tilde{Z}_{k\Delta}+\tilde{T}_{k\Delta}\right)\right|\right)\mathbb{1}_{\Omega_{N,k}\cap\Omega_{X,k}}\middle|\mathscr{F}_{k\Delta}\right]$$
$$\leq 2\exp\left(\frac{c\varepsilon^2\left(\sigma_0^2+\xi_0^2\right)t^2(X_{k\Delta})}{\Delta\left(1-\varepsilon/\varepsilon_1\right)}\right).$$

We conclude as in the proof of Theorem 4.

7. Auxiliary proofs

7.1. Decomposition on a lattice

Proposition 20. *If there exist some constants c_1, c_2 and K independent of D, n, Δ, b and σ and two constants α and β independent of n and D such that, for any function $t \in S_m + S'_m$:*

$$\forall \eta,\zeta>0,\ \forall t\in S_m+S_{m'}\ \|t\|_\infty \leq C\zeta,$$
$$\mathbb{P}\left(f_n(t)\geq\eta,\|t\|_n^2\leq\zeta^2\right)\leq K\exp\left(-\frac{\eta^2 n\beta}{\left(c_1\alpha^2\zeta^2+2Cc_2\alpha\eta\zeta\right)}\right),$$

then there exist some constants C and κ depending only on ν such that, if $D \leq n\beta$:

$$\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,m'}} f_n^2(t)-\frac{\kappa\alpha^2 D}{n\beta}\right]_+ \leq CK\frac{\kappa\alpha^2 D^{3/2}e^{-\sqrt{D}}}{n\beta}.$$

Let us consider an orthonormal (for the L^2_ϖ-norm) basis $(\psi_\lambda)_{\lambda\in\Lambda_{m,m'}}$ of $S_{m,m'} = S_m + S_{m'}$ such that

$$\forall\lambda,\quad \operatorname{card}\left(\left\{\lambda',\ \|\psi_\lambda\psi_{\lambda'}\|\neq 0\right\}\right)\leq\phi_2.$$

Let us set

$$\bar{r}_{m,m'} = \frac{1}{\sqrt{D}}\sup_{\beta\neq 0}\frac{\left\|\sum_\lambda \beta_\lambda\psi_\lambda\right\|_\infty}{|\beta|_\infty}.$$

We obtain that

$$\left\|\sum_\lambda \beta_\lambda\psi_\lambda\right\|_\infty \leq \phi_2\,|\beta|_\infty\sup_\lambda\|\psi_\lambda\|_\infty \quad \text{and}\ \|\psi_\lambda\|_\infty\leq\sqrt{D}\,\|\psi_\lambda\|_{L^2}\leq\pi_1\sqrt{D}\,\|\psi_\lambda\|_\varpi\,;$$

then

$$\bar{r}_{m,m'} \leq \bar{r} := \phi_2 \pi_1.$$

We need a lattice of which the infinite norm is bounded. We use Lemma 9 of [3].

Result 21. *There exists a δ_k-lattice T_k of $L^2_\varpi \cap (S_m + S_{m'})$ such that*

$$\left|T_k \cap \mathscr{B}_{m,m'}\right| \leq \left(5/\delta^k\right)^D$$

where $\delta_k = 2^{-k}/5$. Let us denote by $p_k(u)$ the orthogonal projection of u on T_k. For any $u \in S_{m,m'}$, $\|u - p_k(u)\|_\pi \leq \delta_k$ and

$$\sup_{u \in p_k^{-1}(t)} \|u - t\|_\infty \leq \bar{r}_{m,m'} \delta_k \leq \bar{r} \delta_k.$$

Let us set $H_k = \ln(\left|T_k \cap \mathscr{B}_{m,m'}\right|)$. We have that:

$$H_k \leq D \ln(5/\delta_k) = D\left(k \ln(2) + \ln(5/\delta_0)\right) \leq C(k+1)D.$$

The decomposition of u_k on the δ_k-lattice must be done very carefully: the norms $\|u_k - u_{k-1}\|_\varpi$ and $\|u_k - u_{k-1}\|_\infty$ must be controlled. Let us set

$$\mathscr{E}_k = \left\{u_k \in T_k \cap \mathscr{B}_{m,m'},\ \|u - u_k\|_\varpi \leq \delta_k \text{ et } \|u - u_k\|_\infty \leq \bar{r}\delta_k\right\}.$$

We have that $\ln(|\mathscr{E}_k|) \leq H_k$. For any function $u \in \mathscr{B}_{m,m'}$, there exist a series $(u_k)_{k\geq 0} \in \prod_k \mathscr{E}_k$ such that

$$u = u_0 + \sum_{k=1}^{\infty} (u_k - u_{k-1}).$$

Let us consider $(\eta_k)_{k\geq 0}$ and $\eta \in \mathbb{R}$ such that $\eta_0 + \sum_{k=1}^{\infty} \eta_k \leq \eta$. We obtain:

$$\begin{aligned} &\mathbb{P}\left(\sup_{u \in \mathscr{B}_{m,m'}} |f_n(u)| > \eta\right) \\ &\quad \leq \mathbb{P}\left(\exists (u_k) \in \prod \mathscr{E}_k,\ \left|f_n(u_0) + \sum_{k=1}^{\infty} f_n(u_k - u_{k-1})\right| > \eta_0 + \sum_{k=1}^{\infty} \eta_k\right) \\ &\quad \leq P_1 + \sum_{k=1}^{\infty} P_{2,k} \end{aligned} \tag{16}$$

where

$$P_1 = \sum_{u_0 \in \mathscr{E}_0} \mathbb{P}\left(|f_n(u_0)| > \eta_0\right) \quad \text{and} \quad P_{2,k} = \sum_{u_k \in \mathscr{E}_k} \mathbb{P}\left(|f_n(u_k - u_{k-1})| > \eta_k\right).$$

As $u_0 \in T_0$, $\|u_0\|_\varpi \leq 1$ and $\|u_0\|_\infty \leq \bar{r}\sqrt{D}$. Moreover, $\|u_0\|_n^2 \leq 3/2\ \|u_0\|_\varpi^2 \leq 3\delta_0/2$. Then

$$\mathbb{P}\left(|f_n(u_0)| > \eta_0\right) = \mathbb{P}\left(|f_n(u_0)| > \eta_0,\ \|u_0\|_n^2 \leq 3\delta_0/2\right).$$

There exist two constants c'_1 and c'_2 depending only on δ_0 and $\bar{r}$ such that

$$\mathbb{P}\left(|f_n(u_0)| > \eta_0\right) \leq K \exp\left(-\frac{n\beta\eta_0^2}{c'_1\alpha^2 + 2c'_2\sqrt{D}\alpha\eta_0}\right).$$

Let us set x_0 such that $\eta_0 = \alpha\left(\sqrt{c'_1\left(x_0/\beta\right)} + c'_2\sqrt{D}\left(x_0/\beta\right)\right)$. Then:

$$x_0 \leq \frac{\beta\eta_0^2}{c'_1\alpha^2 + 2c'_2\sqrt{D}\alpha\eta_0}$$

and

$$\mathbb{P}\left(f_n(u_0) > \eta_0\right) \leq K \exp\left(-nx_0\right).$$

Then

$$P_1 \leq K \sum_{u_0 \in \mathscr{E}_0} \exp\left(-nx_0\right) \leq K \exp\left(H_0 - nx_0\right). \tag{17}$$

We have that

$$\|u_k - u_{k-1}\|_\pi^2 \leq 2\left(\|u - u_{k-1}\|_\pi^2 + \|u - u_k\|_\pi^2\right) \leq 5\delta_{k-1}^2/2;$$

then $\|u_k - u_{k-1}\|_n^2 \leq 15\delta_{k-1}^2/4$. As $u_{k-1}, u_k \in \mathscr{E}_{k-1} \times \mathscr{E}_k$, it follows that $\|u_k - u_{k-1}\|_\infty^2 \leq 5\delta_{k-1}^2\bar{r}^2/2$. There exists two constants c_3 and c_4 such that:

$$\begin{aligned}\mathbb{P}_n\left(|f_n(u_k - u_{k-1})| > \eta_k\right) &= \mathbb{P}_n\left(|f_n(u_k - u_{k-1})| > \eta_k,\ \|u_k - u_{k-1}\|_n^2 \leq 15\delta_{k-1}^2/4\right)\\ &\leq K \exp\left(-\frac{n\beta\eta_k^2}{c_3\alpha^2\delta_{k-1}^2 + 2c_4\alpha\delta_{k-1}}\right).\end{aligned}$$

Let us fix x_k such that $\eta_k = \delta_{k-1}a\left(\sqrt{c_3\left(x_k/\beta\right)} + c_4\left(x_k/\beta\right)\right)$. We obtain:

$$x_k \leq \frac{\beta\eta_k^2}{c_3\alpha^2\delta_{k-1}^2 + 2c_4\alpha\delta_{k-1}}$$

and

$$\mathbb{P}\left(|f_n(u_k - u_{k-1})| > \eta_k\right) \leq K \exp\left(-nx_k\right).$$

Then, $P_{2,k} \leq K \exp\left(H_{k-1} + H_k - nx_k\right)$ and

$$P_2 = \sum_{k=1}^{\infty} P_{2,k} \leq K \sum_{k=1}^{\infty} \exp\left(H_{k-1} + H_k - nx_k\right). \tag{18}$$

Let us set $\tau > 0$ and choose (x_k) (and then (η_k)) such that

$$\begin{cases}\sqrt{D}nx_0 = H_0 + D + \tau\\ nx_k = H_{k-1} + H_k + (k+1)D + \tau.\end{cases}$$

Collecting the results, we obtain, by (16)–(18):

$$\mathbb{P}\left(\sup_{u\in\mathscr{B}_{m,m'}} |f_n(u)| > \eta\right) \le C\left(e^{-D}e^{-\tau} + e^{-\sqrt{D}}e^{-\tau/\sqrt{D}}\right). \tag{19}$$

It remains to compute η^2. We denote by C a constant depending only on δ_0 and $\bar{r}$. This constant may vary from one line to another. We have that:

$$\eta = \sum_{k=0}^{\infty} \eta_k \le C\alpha\left(\sum_{k=1}^{\infty} \delta_{k-1}\left(\sqrt{\frac{x_k}{\beta}} + \frac{x_k}{\beta}\right)\right) + \alpha\left(\sqrt{\frac{x_0}{\beta}} + \sqrt{D}\frac{x_0}{\beta}\right).$$

Let us recall that $H_k = C(k+1)D$. Then, $nx_k = C(3k+2)D + \tau$, $\sqrt{D}nx_0 = CD + \tau$ and

$$\sum_{k=0}^{\infty} \frac{\delta_{k-1}x_k}{\beta} \le \frac{1}{n\beta}\sum_{k=0}^{\infty} 2^{-(k-1)}(C(3k+2)D + \tau) \le C\frac{D+\tau}{n\beta}.$$

Moreover,

$$\sum_{k=0}^{\infty} \delta_{k-1}\sqrt{\frac{x_k}{\beta}} \le C\frac{\sqrt{D} + \sqrt{\tau}}{\sqrt{n\beta}}.$$

As $D/n\beta \le 1$, there exists a constant κ such that

$$\eta^2 \le \kappa\alpha^2\left(\frac{D}{n\beta} + 2\frac{\tau}{n\beta} + \frac{\tau^2}{n^2\beta^2}\right).$$

Then, according to (19):

$$\mathbb{P}\left(\sup_{u\in\mathscr{B}_{m,m'}} f_n^2(u) > \kappa\alpha^2\left(\frac{D}{n\beta} + 2\frac{\tau}{n\beta} + \frac{\tau^2}{n^2\beta^2}\right)\right) \le C\left(e^{-D-\tau} + e^{-\sqrt{D}-\tau/\sqrt{D}}\right). \tag{20}$$

Furthermore

$$\begin{aligned} E &:= \mathbb{E}\left(\left[\sup_{u\in\mathscr{B}_{m,m'}} f_n^2(u) - \kappa a^2\frac{D}{n\beta}\right]_+\right) \\ &= \int_0^{\infty} \mathbb{P}\left(\sup_{u\in\mathscr{B}_{m,m'}} f_n^2(u) > \kappa a^2\frac{D}{n\beta} + \tau\right) d\tau. \end{aligned}$$

Setting $\tau = \kappa\alpha^2\left(2y/n\beta + y^2/n^2\beta^2\right)$, it follows:

$$E = C\gamma^2\int_0^{\infty} \mathbb{P}\left(\sup_{u\in\mathscr{B}_{m,m'}} f_n^2(u) > \kappa\alpha^2\left(\frac{D}{n\beta} + 2\frac{y}{n\beta} + \frac{y^2}{n^2\beta^2}\right)\right)\left(\frac{2}{n\beta} + \frac{2y}{n^2\beta^2}\right) dy.$$

By (20),

$$\begin{aligned} E &= C\kappa\alpha^2\left(e^{-D} + e^{-\sqrt{D}}\right)\left(\frac{1}{n\beta}\int_0^{\infty} ye^{-y/\sqrt{D}}dy\right) \\ &\le C\frac{\kappa\alpha^2}{n\beta}D^{3/2}e^{-\sqrt{D}}. \end{aligned}$$

Acknowledgements

The author wishes to thank M. Reiss and V. Genon-Catalot for helpful discussions.

References

[1] D. Applebaum, Lévy Processes and Stochastic Calculus, in: Cambridge Studies in Advanced Mathematics, vol. 93, Cambridge University Press, Cambridge, 2004.
[2] S. Arlot, P. Massart, Data-driven calibration of penalties for least-squares regression, Journal of Machine Learning Research 10 (2009) 245–279.
[3] A. Barron, L. Birgé, P. Massart, Risk bounds for model selection via penalization, Probability Theory and Related Fields 113 (3) (1999) 301–413.
[4] L. Birgé, P. Massart, Minimum contrast estimators on sieves: exponential bounds and rates of convergence, Bernoulli 4 (3) (1998) 329–375.
[5] F. Comte, Adaptive estimation of the spectrum of a stationary gaussian sequence, Bernoulli 7 (2) (2001) 267–298.
[6] F. Comte, V. Genon-Catalot, Y. Rozenholc, Penalized nonparametric mean square estimation of the coefficients of diffusion processes, Bernoulli 13 (2) (2007) 514–543.
[7] C. Dellacherie, P.A. Meyer, Probabilités et potentiel. Chapitres V à VIII, revised ed., in: Actualités Scientifiques et Industrielles (Current Scientific and Industrial Topics), vol. 1385, Hermann, Paris, 1980. Théorie des martingales. [Martingale theory].
[8] R.A. DeVore, G.G. Lorentz, Constructive Approximation, in: Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 303, Springer-Verlag, Berlin, 1993.
[9] E. Gobet, M. Hoffmann, M. Reiß, Nonparametric estimation of scalar diffusions based on low frequency data, The Annals of Statistics 32 (5) (2004) 2223–2253.
[10] M. Hoffmann, Adaptive estimation in diffusion processes, Stochastic Processes and their Applications 79 (1) (1999) 135–163.
[11] Y. Ishikawa, H. Kunita, Malliavin calculus on the Wiener–Poisson space and its application to canonical SDE with jumps, Stochastic Processes and their Applications 116 (12) (2006) 1743–1769.
[12] R.S. Liptser, A.N. Shiryaev, Statistics of Random Processes. I, expanded ed., in: Applications of Mathematics (New York), vol. 5, Springer-Verlag, Berlin, 2001. General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.
[13] H. Mai, Efficient maximum likelihood estimation for Lévy-driven Ornstein–Uhlenbeck processes, 2012.
[14] C. Mancini, R. Renò, Threshold estimation of Markov models with jumps and interest rate modeling, Journal of Econometrics 160 (1) (2011) 77–92.
[15] H. Masuda, Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps, Stochastic Processes and their Applications 117 (1) (2007) 35–56.
[16] Y. Meyer, Ondelettes et opérateurs. I, in: Actualités Mathématiques (Current Mathematical Topics), Hermann, Paris, 1990, Ondelettes. [Wavelets].
[17] S. Rubenthaler, Probabilités: aspects théoriques et applications en filtrage non linéaire, systèmes de particules et processus stochastiques, Habilitation à diriger des recherches, Université de Nice–Sophia Antipolis, France, 2010.
[18] E. Schmisser, Estimation non paramétrique pour des processus de diffusion, Ph.D. Thesis, Université Paris Descartes, 2010.
[19] E. Schmisser, Penalized nonparametric drift estimation for a multidimensional diffusion process, Statistics 47 (1) (2013) 61–84. URL http://dx.doi.org/10.1080/02331888.2011.591931.
[20] Y. Shimizu, Some remarks on estimation of diffusion coefficients for jump-diffusions from finite samples, Bulletin of Informatics and Cybernetics 40 (2008) 51–60.
[21] Y. Shimizu, N. Yoshida, Estimation of parameters for diffusion processes with jumps from discrete observations, Statistical Inference for Stochastic Processes 9 (3) (2006) 227–277.