Non parametric estimation of the diffusion coefficients of a diffusion with jumps

Émeline Schmisser

Version of Record: https://www.sciencedirect.com/science/article/pii/S0304414919301395
Manuscript_b15918faad594324e89de3a7094abe4a

Non parametric estimation of the coefficients of a diffusion with jumps

Émeline Schmisser*

October 8, 2018

Abstract

In this article, we consider a jump diffusion process $(X_t)_{t\geq 0}$, with drift function b, diffusion coefficient σ and jump coefficient ξ^2. This process is observed at discrete times $t = 0, \Delta, \ldots, n\Delta$. The sampling interval Δ tends to 0 and the time interval $n\Delta$ tends to infinity. We assume that $(X_t)_{t\geq 0}$ is ergodic, strictly stationary and exponentially β-mixing. We use a penalized least-square approach to compute adaptive estimators of the functions $\sigma^2 + \xi^2$ and σ^2. We provide bounds for the risks of the two estimators.

Keywords: jump diffusions, model selection, nonparametric estimation
Subject Classification: 62G05, 62M05

1 Introduction

We consider the stochastic differential equation (SDE):

$$dX_t = b(X_{t^-})dt + \sigma(X_{t^-})dW_t + \xi(X_{t^-})dL_t, \qquad X_0 = \eta \tag{1}$$

with η a random variable, $(W_t)_{t\geq 0}$ a Brownian motion independent of η and $(L_t)_{t\geq 0}$ a pure jump centered Lévy process independent of $\left((W_t)_{t\geq 0}, \eta\right)$:

$$L_t = \int_0^t \int_{|z|<1} z\left(\mu(dt,dz) - \nu(dz)dt\right) + \int_0^t \int_{|z|\geq 1} z\mu(dt,dz)$$

where μ is a Poisson measure of intensity $\nu(dz)dt$, with $\int_{\mathbb{R}}(z^2 \wedge 1)\nu(dz) < \infty$. The process $(X_t)_{t\geq 0}$ is assumed to be ergodic, stationary and exponentially β-mixing. It is observed at discrete times $t = 0, \Delta, \ldots, n\Delta$ where the sampling interval Δ tends to 0 and the time of observation $n\Delta$ tends to infinity. Our aim is to construct adaptive non-parametric estimators of $\xi^2 + \sigma^2$ and σ^2 on a compact set A. We do not assume that the jumps are of finite intensity, only that L_Δ is centred (that is $\int_{|z|\geq 1} z\nu(dz) = 0$) and has a moment of order 8.

Diffusions with jumps become powerful tools to model processes in biology, physics, social sciences, medical sciences, economics, and finance. They are used in the study of dynamical systems

*Laboratoire Paul Painlevé, Université Lille 1. email: emeline.schmisser@math.univ-lille1.fr

when the noise is discontinuous or too intensive to be modeled by a Brownian motion, like polymerarization phenomenons (see Berestycki (2004)), telephone noise or infinite capacity dam (Protter and Talay (1997)). In finance, they are used to model a variety of financial applications such as interest rate modelling or capital asset pricing (which includes fair pricing of options). Indeed, the standard model was the Black-Sholes model, but the prices in the market are not continuous and often have jumps. The jump diffusions are therefore more and more used to model asset prices (see Aït-Sahalia and Jacod (2009) and Protter and Talay (1997) for instance).

There already exists numerous articles dealing with adaptive estimation for Lévy processes (see for instance Kappus (2014) and Comte and Genon-Catalot (2009) for pure jump Lévy processes and Neumann and Reiß (2009) and Gugushvili (2009) for more general Lévy processes. The estimators are based on the characteristic function. Non-parametric estimation of the coefficients of a diffusion without jumps is also well known (e.g Hoffmann (1999) or Comte et al. (2007)). However, to our knowledge, there do not exist adaptive estimators for the coefficients of a jump diffusion. Moreover, the non-parametric estimators are pointwise and constructed only in the finite intensity case. Shimizu (2008) construct maximum-likelihood parametric estimators of σ^2 and ξ^2 when the process (X_t) is stationnary. Their estimators converge with rates $\sqrt{n}$ and $\sqrt{n\Delta}$ respectively. Mancini and Renò (2011) and Hanif et al. (2012) both use local times to construct pointwise, non-adaptive estimators in the finite intensity cas. Mancini and Renò (2011) construct kernel estimators of σ^2. Hanif et al. (2012) provide local polynomials estimators of $\sigma^2 + \xi^2$. They both prove the convergence and the asymptotic normality of their estimators.

In this paper, we construct L^2 adaptive estimators of the two functions $\sigma^2+\xi^2$ and σ^2 under the asymptotic framework $n\Delta \to \infty$, $\Delta \to 0$. To estimate the second infinitesimal moment $\sigma^2(x)+\xi^2(x)$ on a compact set A, we consider the following random variables

$$T_{k\Delta} := \frac{(X_{(k+1)\Delta} - X_{k\Delta})^2}{\Delta} = \sigma^2(X_{k\Delta}) + \xi^2(X_{k\Delta}) + \text{noise} + \text{remainder}.$$

We introduce a sequence of increasing subspaces S_m of $L^2(A)$ and we construct a sequence of estimators $\hat{g}_m$ by minimizing over each S_m a contrast function

$$\gamma_{1,n}(t) = \frac{1}{n}\sum_{k=1}^{n}(T_{k\Delta} - t(X_{k\Delta}))^2.$$

We bound the risk, then we introduce a penalty function $\text{pen}_1(m)$ and we minimize on m the function $\gamma_{1,n}(\hat{f}_m) + \text{pen}_1(m)$. Our estimator satisfies an oracle inequality (up to a multiplicative constant). As the penalty $\text{pen}_1(m)$ depends on a unknown constant Σ_1, we construct an estimator $\hat{\Sigma}_1$ and show that the robust estimator obtained by minimizing $\gamma_{1,n}(\hat{f}_m) + \widehat{\text{pen}}_1(m)$ satisfies the same oracle inequality.

To estimate the function σ^2, we consider the randow variables

$$Y_{k\Delta} = \frac{(X_{(k+1)\Delta} - X_{k\Delta})^2}{\Delta}\mathbb{1}_{\left|X_{(k+1)\Delta}-X_{k\Delta}\right| \le \sqrt{\Delta}\ln(n)} = \sigma^2(X_{k\Delta}) + \text{noise} + \text{remainder}.$$

We show that these variables are close to the continuous part of the increments, then we construct an adaptive robust estimator of σ^2 as for $\sigma^2 + \xi^2$. The risk of this estimator depends on the Blumenthal-Getoor index of ν and automatically realizes a bias-variance compromise.

This article is composed as follows: in Section 2, we specify the model and its assumptions. In Sections 3 and 4, we construct the estimators and bound their risks. Section 5 is devoted to the

simulations and proofs are gathered in Section 6. The technical tools (the definition of the Besov spaces, some properties of the Fourier transforms, the Talagrand's and Bennett's inequalities, the Berbee's coupling lemma and the Burkholder-Davis-Gundy inequality) are set in Appendix A.

2 Model

We consider the stochastic differential equation (1). We assume that the following assumptions are fulfilled:

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

This assumption ensures that a unique strong solution of (1) exists.

Assumption A2. *The process (X_t) is ergodic, exponentially β-mixing and there exists a unique invariant probability.*

See Masuda (2007) for sufficient conditions for this assumption. For instance, it is satisfied if:

a. The Lévy measure ν is symetric.

b. The function b is elastic: $\exists c > 0, M \geq 0, \forall x, |x| \geq M,\ xb(x) \leq -c|x|$.

c. The function σ^2 and ξ^2 are bounded.

d. The function σ is bounded by below: $\forall x, \sigma(x) \geq \sigma_0 > 0$.

We can now assume:

Assumption A3.

a. The process (X_t) is stationary.

b. Its stationary measure has a density π which is bounded from below and above on any compact set. In particular, there exists π_0 and π_1 such that for any $x \in A$, there exist two constants $0 < \pi_0 \leq \pi(x) \leq \pi_1 < \infty$.

We need also some technical assumptions on the Lévy measure and on the functions σ^2 and ξ^2:

Assumption A4. *The Lévy measure ν satisfies:*

a. $\nu(\{0\}) = 0$ and $\quad I_8 := \int_{-\infty}^{+\infty} z^8 \nu(dz) < \infty$.

b. $\mathbb{E}\left(X_t^8\right) < \infty$ (it is not always a consequence of the previous item, see also Masuda (2007) for sufficient assumptions).

c. The Blumenthal-Getoor index is strictly less than 2: there exists $\beta \in [0, 2[$ such that $\int_{-1}^{1} |z|^\beta \nu(dz) < \infty$.
This is not a strong assumption, as the Lévy measure already satisfies $\int_{-1}^{1} z^2 \nu(dz) < \infty$.

d. $\int_{-\infty}^{+\infty} z^2 \nu(dz) = 1$. This item ensures the identifiability of the function ξ (if not, we estimate the function $\sigma^2(x) I_2$ where $I_2 := \int_{-\infty}^{+\infty} z^2 \nu(dz)$).

e. The functions σ and ξ are bounded from below and above: $\exists \sigma_1^2, \xi_1^2$ such that

$$\forall x \in \mathbb{R},\ 0 < \sigma^2(x) \leq \sigma_1^2 \quad \textit{and} \quad 0 < \xi^2(x) \leq \xi_1^2.$$

To simplify the notations, let us set

$$B_{k\Delta} = \int_{k\Delta}^{(k+1)\Delta} b(X_s)ds, \quad Z_{k\Delta} = \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s)dW_s \quad \text{and} \quad J_{k\Delta} = \int_{k\Delta}^{(k+1)\Delta} \xi(X_{s^-})dL_s. \quad (2)$$

We introduce the σ-algebra $\mathscr{F}_t = \sigma\left(\eta, (W_s)_{0\leq s\leq t}, (L_s)_{0\leq s\leq t}\right)$.

The following lemma follows directly from the Burkholder-Davis-Gundy inequality.

Lemma 1. *For any $p \geq 1$, if $I_{2p} := \int_{\mathbb{R}} z^{2p}\nu(dz) < \infty$ and $\mathbb{E}\left(X^{2p}\middle|\mathscr{F}_{k\Delta}\right) < \infty$, we have:*

$$\forall u \geq 0, \quad \mathbb{E}\left(\sup_{0\leq s\leq\Delta}(X_{s+u})^{2p}\middle|\mathcal{F}_u\right) \lesssim (1 + |X_u|^{2p})$$

and

$$\forall u \geq 0, \quad \mathbb{E}\left(\sup_{0\leq s\leq\Delta}(X_{s+u} - X_u)^{2p}\right) \lesssim \Delta I_{2p} + \Delta^p \quad (3)$$

where $A \lesssim B$ means $\exists C$, $A \leq CB$ and C does not depend on Δ or on n. Then, as b, σ^2 and ξ^2 are Lipschitz:

$$\begin{aligned}
\mathbb{E}\left(B_{k\Delta}^{2p}\middle|\mathscr{F}_{k\Delta}\right) &= \Delta^{2p}b^{2p}(X_{k\Delta}) + C\Delta^{2p+1/2} \\
\mathbb{E}\left(Z_{k\Delta}^{2p}\middle|\mathscr{F}_{k\Delta}\right) &= \Delta^p\sigma^{2p}(X_{k\Delta})\mathbb{E}\left(N^{2p}\right) + C\Delta^{p+1/2} \quad \textit{where} \quad N \sim \mathscr{N}(0,1) \\
\mathbb{E}\left(J_{k\Delta}^{2p}\middle|\mathscr{F}_{k\Delta}\right) &= \Delta\xi^{2p}(X_{k\Delta})I_{2p} + C\Delta^{3/2}.
\end{aligned}$$

To bound the risk of the adaptive estimator of $g = \sigma^2 + \xi^2$, we need bounded variables (or at least bounded variables with high probability). As exponential moments are needed, the functions σ and ξ have to be bounded. The big jumps must also be under control.

Assumption A5.

a. The Lévy measure ν is sub-exponential:

$$\exists \lambda, C > 0, \quad \forall |z| > 1, \quad \nu(]-z, z[^c) \leq Ce^{-\lambda|z|}.$$

b. The random variables $(X_\Delta, \ldots, X_{n\Delta})$ have exponential moments: $\exists \mu, K > 0$,

$$\mathbb{E}\left(\exp(\mu X_\Delta)\right) \leq K.$$

c. There exists δ, $0 < \delta < 1$, such that $\Delta = O(n^{-\delta})$.

This assumption ensures that there is not too much big jumps in a fixed-time interval. Indeed, the terms $|J_{k\Delta}|$ are bounded by a constant C with high probability. This allow us to bound the terms $B_{k\Delta}$, $Z_{k\Delta}$ and $J_{k\Delta}$:

Lemma 2.

a. *Under Assumptions A1-A4,* $\forall \epsilon \in]0,1[$, $\forall r > 0$, $\mathbb{P}\left(|B_{k\Delta}| \geq \Delta^{1-\epsilon}\right) \lesssim n^{-r}$.

b. *Under A1-A4 and A5e, for any* $r > 0$, $\mathbb{P}\left(|Z_{k\Delta}| \geq r\sigma_1 \Delta^{1/2} \ln(n)\right) \leq 2n^{-r}$.

c. *For the jumps, we have two bounds. Indeed, one term of jump can be quite big, but we can control fairly well a mean of jump terms. Let us set* $q_n = cr\ln(n)/\Delta$. *Under Assumptions A1-A5, for any* $p > 0$, *for any* $r > 0$, *for c large enough,*

$$\mathbb{P}\left(|J_{k\Delta}| \geq \frac{r^2 C_J \ln(n)}{\lambda}\right) \lesssim n^{-r} \quad \text{and} \quad \mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n} J_{k\Delta}^{2p} \geq (r+1)^2 C_p \xi_1^{2p} \Delta \ln^{2p}(n)\right) \lesssim n^{-r}$$

where the constans C_J *and* C_p *will be precised later.*

To estimate σ, we cut off the jumps. The following assumption ensures that the (not too small) jumps can be detected and removed.

Assumption A6.

a. *The function* ξ *is bounded from below:* $\exists \xi_1$, $\forall x \in \mathbb{R}$, $\xi^2(x) \geq \xi_0^2 > 0$.

b. *There exists* $0 < \delta < 1$ *such that* $\Delta = O(n^{-\delta})$.

To estimate the constant in the penalty function, we need an additional assumption on the stationary density. We need the following assumption:

Assumption A7. *The stationary density* π *is continuous on* A.

Our aim is to estimate the functions $g := \sigma^2 + \xi^2$ and σ^2 non parametrically on the compact set A. Estimating directly a function is difficult. To do so, we introduce a sequence of increasing subspaces $(S_m)_{m \in \mathcal{M}_n}$ of $L^2(A)$. Then, as in the linear regression framework, we construct a sequence of estimators by minimizing over each S_m a mean square contrast function $\gamma_n(t)$. Finally we select the "best" estimators $\hat{g}_{\hat{m}}$ and $\hat{\sigma}^2_{\hat{m}}$ thanks to a penalty function. As usual in nonparametric estimation, the risk of our estimators can be decomposed in a variance term and a bias term which depends of the regularity of the estimated function. We choose to use the Besov spaces to caracterize the regularity, which are well adapted to L^2 estimation (particularly for the wavelet decomposition).

First, we need some conditions on the subspaces S_m.

Assumption A8.

a. *The subspaces* S_m *have finite dimension* D_m *and are increasing:* $\forall m$, $S_m \subseteq S_{m+1}$.

b. *The* $\|.\|_{L^2}$ *and* $\|.\|_\infty$ *norms are connected:*

$$\exists \phi_1, \forall m, \forall t \in S_m, \quad \|t\|_\infty^2 \leq \phi_1 D_m \|t\|_{L^2}^2$$

with $\|t\|_{L^2}^2 = \int_A t^2(x)dx$ *and* $\|t\|_\infty = \sup_{x \in A} |t(x)|$. *This implies that for any orthonormal basis* $(\varphi_\lambda)_{\lambda \in \Lambda_m}$ *of* S_m, $\forall x \in \mathbb{R}$,

$$\sum_{\lambda \in \Lambda_m} \varphi_\lambda^2(x) \leq \phi_1 D_m.$$

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

$$\forall \lambda, \operatorname{card}(\lambda', \|\psi_\lambda \psi_{\lambda'}\|_\infty \neq 0) \leq \phi_2.$$

d. For any function t belonging to $\mathscr{B}^\alpha_{2,\infty}$, the Besov space of regularity $\alpha \leq r$ (see Appendix A),

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

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

These conditions are classical in nonparametric estimation. The vectorial subspaces generated by piecewise polynomials of degree r, spline functions of degree r or wavelets of regularity r satisfy these properties (see Meyer (1990), Proposition 4 p50, and DeVore and Lorentz (1993) for the proof of (d)).

3 Estimation of $\sigma^2 + \xi^2$.

To estimate σ^2 for a diffusion process (without jumps), we can consider the random variables

$$T_{k\Delta} = \frac{(X_{(k+1)\Delta} - X_{k\Delta})^2}{\Delta}$$

(see Comte et al. (2007) for instance). For jump diffusions,

$$X_{(k+1)\Delta} - X_{k\Delta} = B_{k\Delta} + Z_{k\Delta} + J_{k\Delta}$$

and therefore $T_{k\Delta}$ is a rough estimator of $\sigma^2(X_{k\Delta}) + \xi^2(X_{k\Delta})$, not of $\sigma^2(X_{k\Delta})$ alone. We can write:

$$T_{k\Delta} = \sigma^2(X_{k\Delta}) + \xi^2(X_{k\Delta}) + E_{k\Delta} + F_{k\Delta} + G_{k\Delta}$$

where

$$\Delta E_{k\Delta} := (B_{k\Delta} + Z_{k\Delta} + J_{k\Delta})^2 - (Z_{k\Delta} + J_{k\Delta})^2 + \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^2 \middle| \mathscr{F}_{k\Delta}\right) - \Delta(\sigma^2(X_{k\Delta}) + \xi^2(X_{k\Delta}))$$

is a remainder term and

$$\Delta F_{k\Delta} := Z^2_{k\Delta} - \mathbb{E}\left(Z^2_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right), \quad \Delta G_{k\Delta} := J^2_{k\Delta} - \mathbb{E}\left(J^2_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right) + 2 J_{k\Delta} Z_{k\Delta}$$

are centred. The random variable $F_{k\Delta}$ comes from the Brownian terms, and $G_{k\Delta}$ from the jump and Brownian terms.

The following lemma derived from Proposition 1 and the Burkholder-Davis-Gundy inequality. It is proved in Section 6.

Lemma 3. *Under Assumptions A1-A4,*

- $\mathbb{E}\left(E^2_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right) \lesssim \Delta$, $\mathbb{E}\left(E^4_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right) \lesssim \Delta$.
- $\mathbb{E}\left(F_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right) = 0$, $\mathbb{E}\left(F^2_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right) = 2\sigma^4(X_{k\Delta}) + C\Delta^{1/2}$, $\mathbb{E}\left(F^4_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right) \lesssim 1$.
- $\mathbb{E}\left(G_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right) = 0$, $\mathbb{E}\left(G^2_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right) = \Delta^{-1}\xi^4(X_{k\Delta})I_4 + C\Delta^{-1/2}$.
- $\mathbb{E}\left((G_{k\Delta} + F_{k\Delta})^2 \middle| \mathscr{F}_{k\Delta}\right) = \Delta^{-1}\xi^4(X_{k\Delta})I_4 + C\Delta^{1/2}$

where C is a constant that does not depend on n neither on Δ and is not the same in two different lines.

3.1 Estimation for fixed m

For any $m \in \mathscr{M}_n = \{m,\ D_m \leq \mathscr{D}_n\}$ where the maximal dimension $\mathscr{D}_n$ satisfies $\mathscr{D}_n \leq \sqrt{n\Delta}/\ln(n)$, we construct an estimator $\hat{g}_m$ of $g = \sigma^2 + \xi^2$ by minimizing on S_m the mean square contrast function

$$\gamma_{1,n}(t) = \frac{1}{n}\sum_{k=1}^{n} \left(t(X_{k\Delta}) - T_{k\Delta}\right)^2.$$

We can always find a function minimizing $\gamma_{1,n}(t)$, but it may not be unique. On the contrary, the random vector $(\hat{g}_m(X_\Delta), \ldots, \hat{g}_m(X_{n\Delta})$ is always uniquely defined. Therefore we consider the empirical risk $\mathscr{R}_n(\hat{g}_m)$, where

$$\mathscr{R}_{1,n}(t) = \mathbb{E}\left(\|t - g_A\|_n^2\right) \quad \text{with} \quad \|t\|_n^2 = \frac{1}{n}\sum_{k=1}^{n} t^2(X_{k\Delta})$$

and $g_A := g\mathbb{1}_A$. We introduce the L^2_π-norm $\|t\|_\pi^2 := \int_A t^2(x)\pi(x)dx$, and $g_{m,\pi}$ the orthogonal projection of g on S_m for the π-norm.

Proposition 4 (Bound of the empirical risk)**.** *Under Assumptions A1-A4 and A8, if $m \in \mathscr{M}_n$, the risk of the estimator $\hat{g}_m$ is bounded by:*

$$\mathscr{R}_{1,n}(\hat{g}_m) \leq \|g_{m,\pi} - g_A\|_\pi^2 + 12\Sigma_1\frac{D_m}{n\Delta} + C\Delta$$

where the constant C does not depend on m, nor on n and Δ and

$$\Sigma_1 := \min\left(\xi_{1,A}^4 I_4, \frac{\Phi_1}{\pi_0} I_4 \int_{\mathbb{R}} \xi^4(z)\pi(z)dz\right)$$

whith $\xi_{1,A} = \sup_{x\in A}\xi(x)$.

Corollary 5 (Bound of the L^2-risk)**.** *The bound for the L^2_π-risk is less sharp. Under Assumptions A1-A4 and A8, if $m \in \mathscr{M}_n$*

$$\mathbb{E}\left(\|\hat{g}_m - g_A\|_\pi^2\right) \leq 9\|g - g_{m,\pi}\|_\pi^2 + 24\Sigma_1\frac{D_m}{n\Delta} + 2C\Delta.$$

For any function t on the compact set A, $\|t\|_\pi^2 \leq \pi_1\|t\|_{L^2}^2$. Then, if we denote by g_m the orthogonal projection (L^2) of g on S_m, we get that $\|g_{m,\pi} - g_A\|_\pi^2 \leq \|g_A - g_m\|_\pi^2 \leq \pi_1\|g_A - g_m\|_{L^2}^2$ and, under the same assumptions,

$$\mathbb{E}\left(\|\hat{g}_m - g_A\|_{L^2}^2\right) \leq 9\pi_1\|g - g_m\|_{L^2}^2 + 24\Sigma_1\frac{D_m}{n\Delta} + 2C\Delta.$$

This estimator converges when $n\Delta \to \infty$. This is quite logical: indeed, for a compound Poisson process, we have approximatively $n\Delta$ jumps in the time interval $[0, n\Delta]$. We can remark that, quite naturally, the empirical risk and the L^2_π-risk converges with rate $\sqrt{n\Delta}$.

We have to find a good compromise between the bias term, $\|g_m - g_A\|_{L^2}^2$, which decreases when m increases, and the variance term, proportional to $D_m/(n\Delta)$. If g belongs to the Besov space

$\mathscr{B}^{\alpha}_{2,\infty}(A)$ (with $\alpha \geq 1$), then $\|g_m - g_A\|^2_{L^2}$ is proportional to $D_m^{-2\alpha}$. The risk is then minimal for $m_{opt} = (n\Delta)^{1/(1+2\alpha)}$, and satisfies, when $n\Delta^2 \to 0$,

$$\mathbb{E}\left(\left\|\hat{g}_{m_{opt}} - g_A\right\|^2_{L^2}\right) \lesssim (n\Delta)^{-2\alpha/(2\alpha+1)}.$$

Hanif et al. (2012) assume that σ^2 and ξ^2 belongs to $\mathscr{C}^2$. If $n\Delta^{3/2} \to 0$, they obtain the rate of convergence $(n\Delta)^{-2/5}$. We obtain the same rate of convergence for this regularity.

3.2 Adaptive estimator

We now have a collection of estimators $\hat{g}_0, \ldots, \hat{g}_m, \ldots$. Our aim is to select automatically the dimension $\hat{m}$, without any knowlege of the regularity of g. As the subspaces S_m are increasing, the function $\gamma_{1,n}(\hat{\nu}_m)$ decreases when m increases. To find an adaptive estimator, we need to add a penalty term $\mathrm{pen}_1(m)$. We take a penalty term proportional to the variance, that is $\mathrm{pen}_1(m) = \kappa\Sigma_1 \frac{D_m}{n\Delta}$ and choose the adaptive estimator $\hat{g}_{\hat{m}}$ by minimizing the function

$$\hat{m} = \min_{m \in \mathscr{M}_n} \gamma_{1,n}(\hat{g}_m) + \mathrm{pen}_1(m).$$

The random variables $F_{k\Delta}$ and $G_{k\Delta}$ are bounded with high probability thanks to Lemma 3. As they are also exponentially β-mixing, we can apply the Berbee's coupling lemma and a Talagrand's inequality for β-mixing random variables. We then obtain the following oracle inequality:

Theorem 6. *Under assumptions A1-A5 and A8, there exists a universal constant κ_1 such that for any $\kappa \geq \kappa_1$,*

$$\mathbb{E}\left(\|\hat{g}_{\hat{m}} - g_A\|^2_{L^2}\right) \lesssim \inf_{m \in \mathscr{M}_n} \left\{\|g_m - g_A\|^2_{L^2} + \mathrm{pen}_1(m)\right\} + \Delta + \frac{\ln^3(n)}{n\Delta}.$$

The adaptive estimator $\hat{g}_{\hat{m}}$ automatically realises the best (up to a multiplicative constant) compromise.

3.3 Robust estimator

It remains to find the constant in the penalty. The universal constant κ_1 can be explicitely bounded; the quantity

$$\Sigma_1 = \min\left\{\xi^4_{1,A} I_4 \ , \ \frac{\phi_1}{\pi_0}\int_{\mathbb{R}} \xi^4(x)\pi(x)dx\right\}$$

is unknown but can be bounded. We first estimate the function $h(x) := \xi^4(x)I_4$. Let us set

$$V_{k\Delta} := \frac{(X_{(k+1)\Delta} - X_{k\Delta})^4}{\Delta} = \xi^4(X_{k\Delta})I_4 + \underbrace{L_{k\Delta}}_{\text{small term}} + \underbrace{H_{k\Delta}}_{\text{centred term}}.$$

We consider the estimator

$$\hat{h}_m = \arg\min_{t \in S_m} \gamma_{3,n}(t) \quad \text{where} \quad \gamma_{3,n}(t) = \frac{1}{n}\sum_{k=1}^{n} \left(V_{k\Delta} - t(X_{k\Delta})\right)^2. \tag{4}$$

Let us denote by $\mathscr{R}_{3,n} := \frac{1}{n}\sum_{k=1}^{n}(h(X_{k\Delta} - t(X_{k\Delta}))^2$ the empirical risk.

Proposition 7. *Under Assumptions A1-A5 and A8, for any m such as $D_m \leq \mathscr{D}_n = \sqrt{n\Delta/\ln(n)}$:*

$$\mathbb{E}\left(\mathscr{R}_{3,n}(\hat{h}_m)\right) \leq \|h - h_m\|_\pi^2 I_4 + C\frac{D_m}{n\Delta} + C'\Delta.$$

We set $\hat{h}_1 = \sup_{x\in A} \hat{h}_{\ln(n)}$. Under assumptions A1-A5 and A8,

$$\mathbb{P}\left(\left|\hat{h}_1 - \xi_{1,A}^4 I_4\right| \geq 1/\ln(n)\right) \lesssim n^{-5}.$$

This result is sufficient to bound the penalty (which is all we need). However, the bound will be more precise if we estimate also the second term $\phi_1/\pi_0 I_4 \int_{\mathbb{R}} \xi^4(x)\pi(x)dx$. In the proof of Proposition 7, we show that $\mathbb{E}\left(L_{k\Delta}^2\right) \lesssim \Delta$ and $\mathbb{E}\left(H_{k\Delta}|\,\mathscr{F}_{k\Delta}\right) = 0$. Then, we can remark that

$$\mathbb{E}\left(V_{k\Delta}\right) = \mathbb{E}\left(\xi^4(X_{k\Delta})\right) I_4 + C\Delta^{1/2}.$$

We can estimate $\mathbb{E}\left(V_{k\Delta}\right)$ by the mean $\bar{V}_n$. As we are insterested in the minimum of π, it is better to take the kernel estimator. Indeed, computing the pointzise risk of the kernel estimator is quite natural, and this allows us to control the error on the minimum quite easily. Moreover, the kernel estimator is already implemented in some softwares. We consider the rectangular kernel $K(x) := \mathbb{1}_{|x|\leq 1}/2$ and set

$$\hat{\pi}_h(x) = \frac{1}{n}\sum_{k=1}^{n} K_h\left(X_i - x\right) \quad \text{with} \quad K_h(x) = h^{-1}K(x/h).$$

We take a grid $(x_1, \ldots, x_{\ln^2(n)})$ of equally spaced points of A, and we compute the minimum of $\hat{\pi}_{(n\Delta)^{-1/2})}$ on this grid:

$$\hat{\pi}_0 := \min_{1\leq j\leq \ln^2(n)} \hat{\pi}_{(n\Delta)^{-1/2}}(x_j).$$

Let us set $\hat{\Sigma}_1 = \min\left\{\hat{h}_1\ ,\ \frac{\phi_1}{\hat{\pi}_0}\bar{V}_n\right\}$.

Lemma 8. *Under assumptions A1-A5 and A7-A8, $\hat{\Sigma}_1$ is a consistent estimator of Σ_1 and more precisely, for C large enough*

$$\mathbb{P}\left(|\hat{\Sigma}_1 - \Sigma_1| \geq \frac{1}{\ln(n)}\right) \lesssim n^{-5}.$$

Corollary 9. *Under Assumption A1-A5 and A7-A8, for any $\kappa \geq \kappa_1$, the adaptive estimator $\hat{g}_{\tilde{m}}$ where*

$$\tilde{m} = \arg\min_{m\in\mathscr{M}_n}\left\{\gamma_{1,n}(\hat{g}_m) + \kappa\left(1 + \frac{1}{\ln(n)}\right)\hat{\Sigma}_1\frac{D_m}{n\Delta}\right\}$$

achieves the oracle inequality:

$$\mathbb{E}\left(\|\hat{g}_{\tilde{m}} - g\|_{L^2}^2\right) \lesssim \inf_{m\in\mathscr{M}_n}\left\{\|g_m - g_A\|_{L^2}^2 + \frac{D_m}{n\Delta}\right\} + \Delta + \frac{\ln^3(n)}{n\Delta}.$$

4 Estimation of σ^2.

We have that

$$T_{k\Delta} = \frac{(X_{(k+1)\Delta} - X_{k\Delta})^2}{\Delta} = \sigma^2(X_{k\Delta}) + J_{k\Delta}^2 + \text{small terms} + \text{ centred terms.}$$

The idea is to keep $T_{k\Delta}$ only when there is no jump. The continuous part of the increment is $B_{k\Delta} + Z_{k\Delta}$ and the jump part of the increments is $J_{k\Delta}$. If we could access the continuous part, we would consider the quantities

$$U_{k\Delta} := \frac{(B_{k\Delta} + Z_{k\Delta})^2}{\Delta}.$$

We need to approximate these increments. As the stochastic term $Z_{k\Delta}$ is of order $\Delta^{1/2}$, we can only suppress the jumps of amplitude strictly greater than $\Delta^{1/2}$. Authors such as Gloter et al. (2016) or Mancini and Renò (2011) use a threshold proportional to $\Delta^{1/2-\varepsilon}$. We decide to use a slightly different threshold, proportional to $\ln(n)\Delta^{1/2}$. Then we consider:

$$Y_{k\Delta} = \frac{\left(X_{(k+1)\Delta} - X_{k\Delta}\right)^2}{\Delta} \mathbb{1}_{\Omega_{X,k}}$$

where $\Omega_{X,k} = \left\{\omega,\ \left|X_{(k+1)\Delta} - X_{k\Delta}\right| \leq (\sigma_1 + \xi_1)\ln(n)\Delta^{1/2}\right)\}$. This is a classical method to obtain the continuous increment of a Lévy process or a jump diffusion (see for instance Jacod (2007), , Shimizu and Yoshida (2006) ...)

Lemma 10 (Approximation of the continuous increment)**.** *Under assumptions A1-A4 and A6,*

$$\mathbb{E}\left((U_{k\Delta} - Y_{k\Delta})^{2p}\right) \lesssim \ln^{4p}(n)\Delta^{1-\beta/2} + \frac{1}{n}.$$

We recall that β is the Blumenthal-Getoor index of the process (X_t). For a compound Poisson process, $\beta = 0$ and most of the jumps can be detected and removed. If β is high (close to 2), there is more and more small jumps and it becomes more difficult to detect them. Under Assumptions A1-A4 and A6,

$$\mathbb{P}\left(|U_{k\Delta} - Y_{k\Delta}| \geq \ln^2(n)\right) \lesssim n^{-6}$$

4.1 Estimation for fixed m

We consider the following contrast function and the empirical risk

$$\gamma_{2,n}(t) = \frac{1}{n}\sum_{k=1}^{n}\left(t(X_{k\Delta}) - Y_{k\Delta}\right)^2 \quad \text{and} \quad \mathscr{R}_{2,n}(t) = \mathbb{E}\left(\left\|t - \sigma^2\right\|_n^2\right).$$

Let us set $\hat{\sigma}_m^2 = \arg\inf_{t\in S_m}\gamma_{2,n}(t)$.

Proposition 11. *Under Assumptions A1-A4, A6 and A8, the risk of the estimator $\hat{\sigma}_m$ is bounded by:*

$$\mathscr{R}_{2,n}(\hat{\sigma}_m^2) \leq \left\|\sigma_A^2 - \sigma_m^2\right\|_\pi^2 + \Sigma_2\frac{D_m}{n} + C\Delta^{1-\beta/2}\ln^2(n)$$

where $\sigma_A^2(x) = \sigma^2(x)\mathbb{1}_{x\in A}$ and $\Sigma_2 \leq \min\left\{\sigma_1^4\ ,\ \frac{\phi_1}{\pi_0}\int\sigma^4(x)\pi(x)dx\right\}$.

The bias term $\left\|\sigma_A^2 - \sigma_m^2\right\|_{L^2}^2$ and the variance term $\Sigma_2 D_m n^{-1}$ are the same as for a diffusion without jumps. Nevertheless, the remainder term is Δ^2 for a diffusion process (see for instance Comte et al. (2007)). Even for Poisson processes, the remainder term will be here proportional to $\Delta \ln^4(n)$. This is due to the jumps: to bound the risk, we use a Taylor approximation of $X_{(k+1)\Delta} - X_{k\Delta}$. For a SDE without jumps, we can refine this Taylor approximation to obtain a smaller remainder term. Here, the difference $(X_{(k+1)\Delta} - X_{k\Delta})^{2p}$ is proportional to the probability of having a jump in the time interval $[k\Delta, (k+1)\Delta]$ and does not change with p: we can not refine the approximation.

If σ^2 belongs to the Besov space $\mathscr{B}^{\alpha}_{2,\infty}$, then $\left\|\sigma_A^2 - \sigma_m^2\right\|_{L^2}^2 \lesssim D_m^{-2\alpha}$. The best estimator is obtained for $D_{m_{opt}} = n^{-1/(1+2\alpha)}$ and its risk is bounded by $n^{-2\alpha/(2\alpha+1)} + \Delta^{1-\beta/2}$.

Remark 12. *Let us set $\Delta \sim n^{-\delta}$, with $0 < \delta < 1$. We have the following rates of convergence:*

a	*jump diffusions*	*diffusions*
$0 < \delta \leq \frac{2\alpha}{2(2\alpha+1)} \leq \frac{1}{2}$	$\Delta^{1/2-\beta/4}$	Δ
$\frac{2\alpha}{2(2\alpha+1)} \leq \delta \leq \frac{2\alpha}{(2\alpha+1)(1-\beta/2)} \wedge 1$	$\Delta^{1/2-\beta/4}$	$n^{-\alpha/(2\alpha+1)}$
$\frac{2\alpha}{(2\alpha+1)(1-\beta/2)} \wedge 1 \leq \delta < 1$	$n^{-\alpha/(2\alpha+1)}$	$n^{-\alpha/(2\alpha+1)}$

If $\beta = 0$, the adaptive estimator will reach the rate of convergence $n^{-\alpha/(2\alpha+1)}$ for high frequency data ($n\Delta^{2\alpha+1/(2\alpha)} = O(1)$). This is the minimax rate of convergence for non-parametric estimation of σ^2 for diffusions processes (see for instance Hoffmann (1999)). If β or α is too big (as soon as $\beta(\alpha + 1/2) > 1$), even for high frequency data, the remainder term will be predominant in the risk. Mancini and Renò (2011) set $\delta = 1$ ($\Delta = 1/n$) and assume that the derivative of σ^2 is bounded. They obtain a rate of convergence of $n^{-1/3}$. For δ close to 1, we obtain the same rate if σ^2 belongs to the Besov space $\mathscr{B}^1_{2,\infty}$, that is if $g' \in L^2$.

4.2 Adaptive estimator

Let us introduce a penalty function $\text{pen}_2(m) = \kappa n^{-1}\Sigma_2$ and define the adaptive estimator $\hat{\sigma}^2_{\hat{m}}$:

$$\hat{m} = \arg \min_{m \in \mathscr{M}_n} \gamma_{2,n}\left(\hat{\sigma}_m^2\right) + \text{pen}_2(m)$$

where $\mathscr{M}_n = \{m,\, D_m \leq \mathscr{D}_n\}$. To bound the risk of the estimator $\hat{\sigma}_{\hat{m}}$, we use a Bernstein-type inequality and a decomposition on a lattice.

Theorem 13. *Under Assumptions A1-A4, A6 and A8, there exists κ_2 such that, if $\kappa \geq \kappa_2$, we have the following oracle inequality:*

$$\mathscr{R}_{2,n}(\hat{\sigma}^2_{\hat{m}}) \lesssim \min_{m \in \mathscr{M}_n} \left(\left\|\sigma_A^2 - \sigma_m^2\right\|_{L^2}^2 + \text{pen}_2(m)\right) + \Delta^{1-\beta/2}\ln^4(n) + \frac{1}{n}.$$

Again, the constant $\Sigma_2 = \min\{\sigma_1^4\ ,\ \frac{\phi_1}{\pi_0}\int \sigma^4(x)\pi(x)dx\}$ is unknown. Let us set $\hat{\sigma}_1^2 = \sup_A \hat{\sigma}^2_{\ln(n)}$ and $\hat{M}_4^c = n^{-1}\sum_{k=1}^n \left(Y_{k\Delta}\right)^2$, and

$$\hat{\Sigma}_2 := \min\left(\max \hat{\sigma}_1^4\ ,\ \frac{\phi_1}{3\hat{\pi}_0}\hat{M}_4^c\right).$$

The estimator $\hat{\Sigma}_2$ converges toward Σ_2 with high probability:

Lemma 14. *Under Assumptions A1-A4 and A6-A7,*

$$\mathbb{P}\left(\left|\hat{\Sigma}_2 - \Sigma_2\right| \geq \frac{1}{\ln(n)}\right) \lesssim n^{-5}.$$

Then the estimator with the estimated penalty satisfies also an oracle inequality:

Corollary 15. *Let us set* $\widehat{\text{pen}}_2(m) = \kappa\left(\hat{\Sigma}_2 + \frac{1}{\ln(n)}\right) D_m/(n\Delta)$ *with* $\kappa \geq \kappa_2$ *and* $\tilde{m} = \arg\min_{m\in\mathscr{M}_n} \gamma_{2,n}(\hat{\sigma}^2_m) + \widehat{\text{pen}}_2(m)$. *Then, for any* $\kappa \geq \kappa_1$, *under Asummptions A1-A4 and A6-A8, we have the following oracle inequality:*

$$\mathscr{R}_n(\hat{\sigma}^2_{\tilde{m}}) \lesssim \min_{m\in\mathscr{M}_n}\left(\left\|\sigma^2_A - \sigma^2_m\right\|^2_{L^2} + \text{pen}_2(m)\right) + \Delta^{1-\beta/2}\ln^4(n) + \frac{1}{n}.$$

5 Simulations

5.1 Models

In the first three models, we consider a stochastic process (X_t) such that

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

with L_t a compound Poisson process:

$$L_t = \sum_{k=1}^{N_t} \zeta_k$$

where N_t is a compound Poisson process of intensity 1, and (ζ_k) are centred, independent, and identically distributed random variables. We denote by f the law of ζ_k and we assume that $\mathbb{E}\left(\zeta_k^2\right) = 1$ and that the random variables (ζ_k) are independent of $(\eta, (W_t)_{t\geq 0}, N_t)$. The simulation is done by an Euler scheme: we first simulate the instants and the size of the jumps. We obtain a time vector $t_1, \ldots, t_N$ of the time of the jumps. Thanks to an Euler scheme, we simulate X_t on the sorted vector: $(0, \Delta, \ldots, t_1, \ldots, t_N, \ldots, n\Delta)$. As X_t can considerably vary after a jump, this method of simulation gives us a better accuracy than simulate L_Δ in one step.

The last two models have infinite jump activity: we nearly consider a stable Lévy process (with jumps smaller than 1). We use the same simulation algorithm as for the Poisson process, with an approximate simulaton of the jumps: we only compute the jumps greater than Δ. The jumps smaller than Δ are replaced by a Gaussian.

All these models satisfy all the assumptions that we can check, that is A1, A2, A4, A5 and A6. We estimate the stationary density, and our models also seem to satisfy also Assumptions A3 and A7.

5.2 Method

We change A into the set $[0, 1]$ by an affine transformation. We use the vectorial subspaces generated by the spline functions:

$$S_{m,r} = \text{Vect}\left(\varphi_{r,k,m},\ k \in \mathbb{Z}\right), \quad \text{with} \quad \varphi_{r,k,m} = 2^{m/2} g_r(2^m x - k)\mathbb{1}_{x\in A}$$
$$\text{and} \quad g_r = \mathbb{1}_{x\in A} * \ldots * \mathbb{1}_{x\in A} \quad (r+1) \text{ time}$$

Those subspaces nearly form a multi-resolution analysis of $L^2([0,1])$ of regularity r.

To construct the adaptive estimator, we compute $\hat{f}_{m,r}$ for $D_m \leq \sqrt{n\Delta}$, $0 \leq r \leq 4$ and $0 \leq m \leq 7$. Indeed for $m = 7$, we already have $D_m = 128$. If m was bigger, there will be a memory problem. The constants κ_1 and κ_2 are chosen by numerical calibration (see Comte and Rozenholc (2002), Comte and Rozenholc (2004) for a complete discussion). We take $\kappa_1 = 6$ and $\kappa_2 = 24$. Then we compute the adaptive estimators and the robust estimators. We minimize $\gamma_n(\hat{f}_{m,r}) + \widehat{\text{pen}}(m,r)$ with respect to m, then r.

To obtain Figures 1-5, for each model, we realise 5 simulations and draw the 10 corresponding estimators: 5 with the true penalty constant, and five with the estimated penalty constant (the robust estimator). To construct the tables, for each set (n,Δ) and each model, we make 50 simulations, and for each simulation, we compute the adaptive estimators $\hat{g}_{\hat{m},\hat{r}}$ or $\hat{\sigma}^2_{\hat{m},\hat{r}}$ and the robust estimator $\hat{g}_{\tilde{m},\tilde{r}}$ or $\hat{\sigma}^2_{\tilde{m},\tilde{r}}$, the selected dimension $(\tilde{m},\tilde{r})$, the estimated penalty constant $\hat{\Sigma}_1$ or $\hat{\Sigma}_2$, and the empirical errors

$$\text{risk} = \frac{1}{n}\sum_{k=1}^{n} \left(\hat{g}_{\tilde{m},\tilde{r}}(X_{k\Delta}) - g(X_{k\Delta})\right)^2 \mathbb{1}_{X_{k\Delta}\in A}, \quad \text{err} = \frac{1}{n}\sum_{k=1}^{n} \left(\hat{g}_{\hat{m},\hat{r}}(X_{k\Delta}) - g(X_{k\Delta})\right)^2 \mathbb{1}_{X_{k\Delta}\in A}.$$

Or for estimating σ^2

$$\text{risk} = \frac{1}{n}\sum_{k=1}^{n} \left(\hat{\sigma}^2_{\tilde{m},\tilde{r}}(X_{k\Delta}) - \sigma^2(X_{k\Delta})\right)^2 \mathbb{1}_{X_{k\Delta}\in A}, \quad \text{err} = \frac{1}{n}\sum_{k=1}^{n} \left(\hat{\sigma}^2_{\hat{m},\hat{r}}(X_{k\Delta}) - \sigma^2(X_{k\Delta})\right)^2 \mathbb{1}_{X_{k\Delta}\in A}.$$

We also compute the empirical error for each $\hat{g}_{m,r}$ (or $\hat{\sigma}^2_{m,r}$). Then we deduce the dimension $(m_{\min}, r_{\min})$ that minimizes the empirical error (denoted by $\text{err}_{\min}$). In the tables, we write the following informations:

- mean and standard deviation of the risk of the robust estimators $\hat{g}_{\tilde{m},\tilde{r}}$ and $\hat{\sigma}^2_{\tilde{m},\tilde{r}}$ (in fact, we give the squrare roots of these quantities, that is $\sqrt{\text{mean(risk)}}$ and $\sqrt{\text{sd(risk)}}$ to compare more easily the error and the values of the function).
- A criteria of comparison between the risks of the adaptive estimator and the robust estimator; orpen = mean(ln(risk/err)). The quantity orpen is positive if the robust estimator is worse, and negative if the robust estimator is better.
- An oracle to compare the risks of the robust estimator and the 'oracle' estimator; or = mean(ln(risk/$\text{err}_{\min}$)). This quantity is always positive.
- the mean and the standard deviation of the selected dimension $D_{\tilde{m},\tilde{r}}$.
- the mean and standard deviation of the estimated penalty constant $\hat{\Sigma}_1$ or $\hat{\Sigma}_2$.

The best results for *risk* are written in bold.

5.3 Results

The estimation of σ^2 seems very good for models 1-3, that is when the Blumenthal-Getoor index is equal to 0. For Δ small enough ($\Delta = 10^{-2}$ or 10^{-3} for Model 1, $\Delta = 10^{-3}$ for Models 2 and 3), the risk of the robust adaptive estimator $\hat{\sigma}^2_{\tilde{m},\tilde{r}}$ is inversely proportional to n, that is proportional to

the variance term. When β is not equal to 0 (models 4-5), we overestimate σ^2: this is because the small jumps can not be cut. The risk mostly depends on Δ: the remainder term is predominant. The effect is more important for Model 5 (with Blumenthal-Getoor index equal to 3/2) than for Model 4 (with Blumenthal-Getoor index equal to 1/2): this is consistent with Remark 12. The best results are obtained for $\Delta = 10^{-3}$ and $n = 10^5$.

The results for function $g = \sigma^2 + \xi^2$ are quite different. Indeed, the variance term is bigger (it is proportional to $1/n\Delta$ and not $1/n$). The convergence is good for Ornstein-Uhlenbeck models (Models 1, 4 and 5); when Δ is small enough ($\leq 10^{-2}$), the risk in inversely proportional to $n\Delta$. This is not exactly the case Models 2 and 3, as we have to find a good trade-off between the bias and the variance terms. But, when $\Delta \leq 10^{-2}$, the risk seems to depends only on $n\Delta$. The larger $n\Delta$, the larger the selected dimension: our estimator is really adaptive.

The risk of the robusts estimators seems to be either comparable, either even better (for the Ornstein-Uhlenbeck models) than the adaptive estimator with known penalty constant. For the Ornstein-Uhlenbeck, we tend to overestimate the penalty, which is better than underestimate it.

The constant $\hat{\Sigma}_2$ seems to converge, even very slowly. For $\hat{\Sigma}_1$, it is more difficult to see a pattern. This is quite understandable: the rate of convergence of $\hat{\Sigma}_2$ really depends on Δ and on n, whereas the one of $\hat{\Sigma}_1$ depends on Δ and on $n\Delta$.

6 Proofs

6.1 Proof of Lemma1

The first two inequalities are fairly classical, so we do not give the proof here. The process being stationary, we only prove the result for $k = 1$.

For any function f Lipschitz, by Hölder inequality and equation (3),

$$\mathbb{E}\left(\left(\int_{k\Delta}^{(k+1)\Delta} f(X_s) - f(X_{k\Delta})ds\right)^{2p}\middle|\mathscr{F}_{k\Delta}\right) \leq \Delta^{2p-1}\mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta}(f(X_s) - f(X_{k\Delta}))^{2p}ds\middle|\mathscr{F}_{k\Delta}\right)$$
$$\lesssim \Delta^{2p+1}I_{2p} + \Delta^{3p}. \tag{5}$$

Let us set

$$\tilde{B}_{k\Delta} := \int_{k\Delta}^{(k+1)\Delta}(b(X_s) - b(X_{k\Delta}))ds, \quad \tilde{Z}_{k\Delta} := \int_{k\Delta}^{(k+1)\Delta}(\sigma(X_s) - \sigma(X_{k\Delta}))dW_s,$$
$$\tilde{J}_{k\Delta} := \int_{k\Delta}^{(k+1)\Delta}(\xi(X_s) - \xi(X_{k\Delta}))dL_s. \tag{6}$$

We have that

$$\mathbb{E}\left(B_\Delta^{2p}\middle|\mathscr{F}_0\right) = \mathbb{E}\left(\left(\Delta b(X_{k\Delta}) + \tilde{B}_{k\Delta}\right)^{2p}\middle|\mathscr{F}_0\right) = b(X_0)^{2p}\Delta^{2p} + \sum_{i=1}^{2p} C_{2p}^i b(X_0)^{2p-i}\Delta^{2p-i}\mathbb{E}\left(\tilde{B}_{k\Delta}^i\middle|\mathscr{F}_0\right)$$

Then by, as the function b is Lipschitz, by (5),

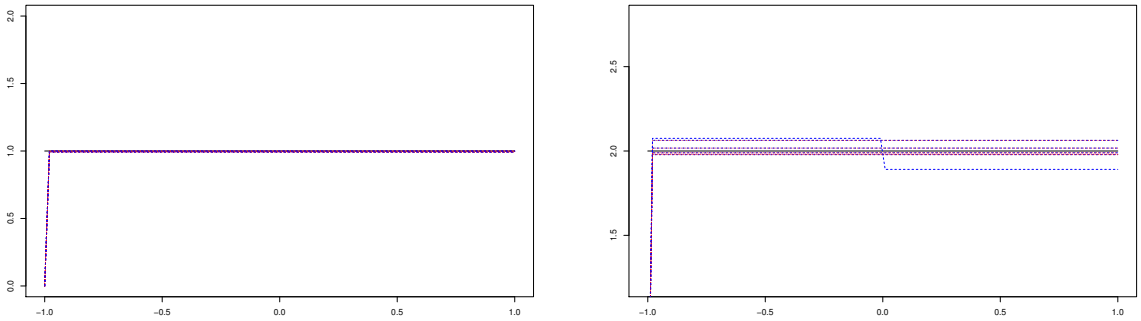
$$\mathbb{E}\left(B_\Delta^{2p} - \Delta^{2p}b(X_0)^{2p}\middle|\mathscr{F}_0\right) \lesssim \sum_{i=1}^{2p}\Delta^{2p-i}\Delta^{i+1/2} \lesssim \Delta^{2p+1/2}.$$

Figure 1: Model 1: Ornstein-Uhlenbeck and binomial jumps ($\beta = 0$)

$$dX_t = -2X_t dt + dW_t + dL_t$$

with binomial jumps: $\mathbb{P}(\zeta_k = 1) = \mathbb{P}(\zeta_k = -1) = 0.5$. Then $\xi_1^4 I_4 = 1$ and $\sigma_1^4 = 1$.

Estimation of σ^2 | Estimation of $\sigma^2 + \xi^2$



$n = 10^5$, $\Delta = 10^{-3}$ | $n = 10^5$, $\Delta = 10^{-2}$

−. : estimator with true penalty constant …: robust estimator

Estimation of $\sigma^2 + \xi^2$

Δ	n	Risk				$D_{\tilde{m},\tilde{r}}$		$\hat{\Sigma}_1$	
		mean	sd	orpen	or	mean	sd	mean	sd
10^{-1}	10^3	0.28	0.26	-0.26	0	1	0	3.8	1.4
10^{-1}	10^4	0.26	0.14	-0.10	0.0046	1.0	0.2	2.9	1.2
10^{-1}	10^5	0.27	0.079	-0.00060	0.12	3.0	0.1	2.1	0.3
10^{-2}	10^3	0.34	0.42	-0.31	0	1	0	4.3	3.7
10^{-2}	10^4	0.11	0.16	-0.19	0	1	0	2.4	1
10^{-2}	10^5	**0.049**	0.054	-0.057	0.087	1.0	0.1	1.5	0.35
10^{-3}	10^4	0.34	0.41	-0.24	0	1	0	4.3	2.8
10^{-3}	10^5	0.11	0.13	-0.16	0	1	0	2.2	0.9

Estimation of σ^2

Δ	n	Risk				$D_{\tilde{m},\tilde{r}}$		$\hat{\Sigma}_2$	
		mean	sd	orpen	or	mean	sd	mean	sd
10^{-1}	10^3	0.74	0.43	-0.15	0.0011	1.0	0.07	7.5	3.7
10^{-1}	10^4	0.74	0.24	-0.034	0.0034	1.125	0.47	4.90	1.05
10^{-1}	10^5	0.75	0.13	-0.0019	0.014	3.0	0.2	4.2	0.3
10^{-2}	10^3	0.050	0.059	-0.035	0	1	0	2.3	2.7
10^{-2}	10^4	0.18	0.14	-0.30	0.061	1.4	1.8	2.1	0.6
10^{-2}	10^5	0.90	0.26	-0.017	0.031	18.6	13.5	4.4	0.7
10^{-3}	10^4	0.015	0.019	-0.055	0	1	0	1.4	1.4
10^{-3}	10^5	**0.0052**	0.0063	0	0	1	0	1.1	0.04

Figure 2: Model 2: square function and Laplace jumps ($\beta = 0$)

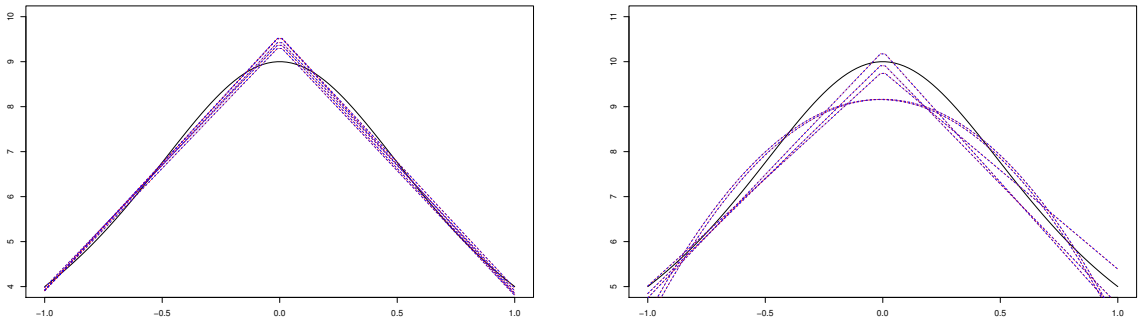
$$dX_t = -2X_t dt + \frac{X_{t^-}^2 + 3}{X_{t^-}^2 + 1} dW_t + dL_t$$

with Laplace jumps:

$$f(dz) = \nu(dz) = 0.5e^{-|z\sqrt{2}|}.$$

Then $\xi_1^4 I_4 = 6$ and $\sigma_1^4 = 81$.

Estimation of σ^2 Estimation of $\sigma^2 + \xi^2$



$n = 10^5$, $\Delta = 10^{-3}$ $n = 10^5$, $\Delta = 10^{-2}$

$-$: true function $-.$: estimator with true penalty constant $\ldots$: robust estimator

Estimation of $\sigma^2 + \xi^2$

Δ	n	Risk				$D_{\tilde{m},\tilde{r}}$		$\hat{\Sigma}_1$	
		mean	sd	orpen	or	mean	sd	mean	sd
10^{-1}	10^3	2.29	1.2	-0.017	0.34	1.0	0.3	30	26
10^{-1}	10^4	2.04	0.9	0.097	0.13	2.1	1.0	26	12
10^{-1}	10^5	1.89	0.4	5.10^{-5}	0.0053	3.0	0.1	19	4
10^{-2}	10^3	2.25	3.5	-0.30	0.85	1.15	0.5	55	198
10^{-2}	10^4	1.29	1.1	0.76	1.20	1.9	1.0	32	41
10^{-2}	10^5	**0.40**	0.3	-0.061	0.21	3	0	20	14
10^{-3}	10^4	1.64	1.7	-0.62	0.71	1.7	1	27	80
10^{-3}	10^5	1.28	1.1	0.71	1.20	1.9	1.0	32	34

Estimation of σ^2

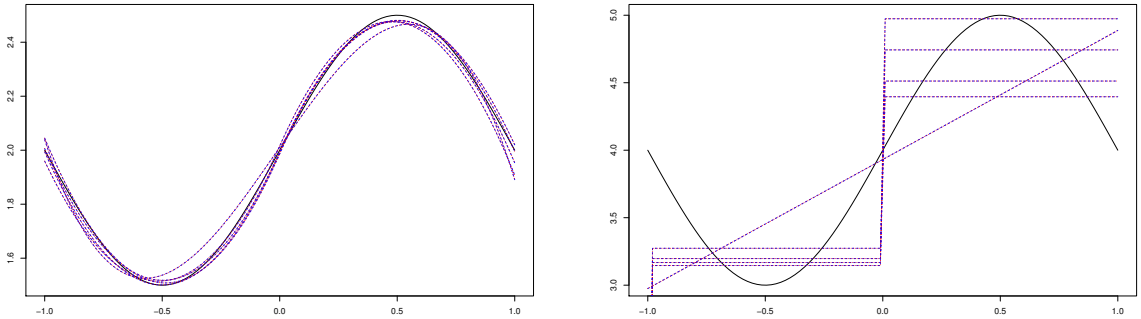
Δ	n	Risk				$D_{\tilde{m},\tilde{r}}$		$\hat{\Sigma}_2$	
		mean	sd	orpen	or	mean	sd	mean	sd
10^{-1}	n^3	1.97	0.8	0	0.52	1	0	48	9.0
10^{-1}	10^4	1.44	0.8	-0.33	0.083	2.6	0.8	43	2.9
10^{-1}	10^5	1.28	0.3	4.10^{-5}	0.013	3.0	0.07	43	1.1
10^{-2}	10^3	1.76	0.7	0.025	1.53	1.0	0.2	67	15.4
10^{-2}	10^4	0.38	0.3	0	0.22	3	0	72	5.6
10^{-2}	10^5	0.26	0.2	-0.016	0.19	3.5	0.8	73	1.7
10^{-3}	10^4	0.32	0.3	0	0.29	3	0	76	6.1
10^{-3}	10^5	**0.14**	0.08	-0.024	0.95	3.3	0.8	78	1.6

Figure 3: Model 3: sine function and Gaussian jumps ($\beta = 0$)

$$dX_t = (-2X_t + \sin(3X_t))dt + \sqrt{2 + 0.5\sin(\pi X_{t^-})}(dW_t + dL_t)$$

with normal jumps: $\zeta_k \sim \mathcal{N}(0,1)$. We have that $\sigma_1^4 = 6.25$ and $\xi_1^4 I_4 = 18.75$.

Estimation of σ^2 Estimation of $\sigma^2 + \xi^2$



$n = 10^5$, $\Delta = 10^{-3}$ $n = 10^5$, $\Delta = 10^{-2}$

$-.$: estimator with true penalty constant ...: robust estimator

Estimation of $\sigma^2 + \xi^2$

Δ	n	Risk				$D_{\tilde{m},\tilde{r}}$		$\hat{\Sigma}_1$	
		mean	sd	orpen	or	mean	sd	mean	sd
10^{-1}	10^3	0.98	0.7	-0.029	0.44	1	0	32	23
10^{-1}	10^4	0.83	0.5	0.10	0.36	1.5	0.5	26	9
10^{-1}	10^5	0.67	0.25	-0.0043	0.024	3.9	0.9	18	3.0
10^{-2}	10^3	1.42	1.5	-0.24	0.17	1	0	53	72
10^{-2}	10^4	0.85	0.6	-0.058	0.78	1.0	0.07	42	26
10^{-2}	10^5	**0.40**	0.3	0.039	0.87	2.0	0.3	27	8
10^{-3}	10^4	1.52	2.0	-0.15	0.20	1	0	63	100
10^{-3}	10^5	0.82	0.5	-0.018	0.82	1.0	0.07	41	21

Estimation of σ^2

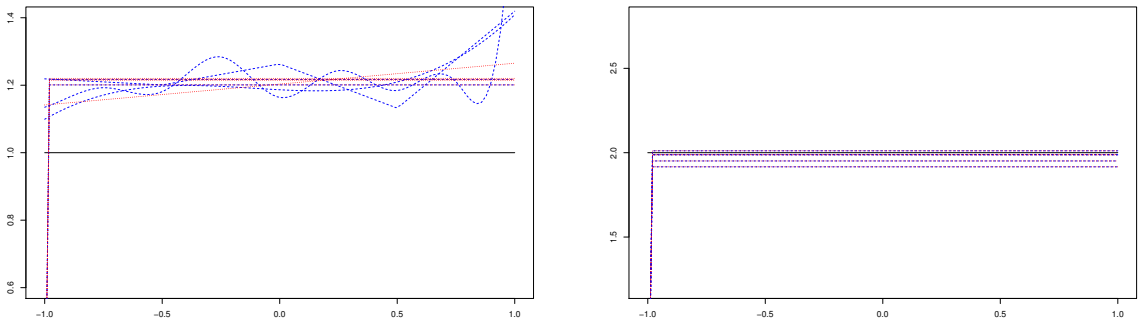
Δ	n	Risk				$D_{\tilde{m},\tilde{r}}$		$\hat{\Sigma}_2$	
		mean	sd	orpen	or	mean	sd	mean	sd
10^{-1}	10^3	0.67	0.4	0.00090	0.21	1.0	0.2	13.0	5.6
10^{-1}	10^4	0.99	0.4	-0.013	0.030	2.3	0.84	12.5	1.8
10^{-2}	10^5	1.22	0.3	-0.0066	0.0048	4.8	0.8	13.8	0.6
10^{-2}	10^3	0.36	0.2	0.10	1.31	1.1	0.32	8.3	4.0
10^{-2}	10^4	0.17	0.1	0.037	0.69	2.6	0.9	6.7	0.6
10^{-2}	10^5	0.20	0.09	-0.00068	0.011	4.1	0.44	6.9	0.22
10^{-3}	10^4	0.14	0.01	0.020	1.40	2.6	0.9	6.4	1.2
10^{-3}	10^5	**0.027**	0.02	0	0.22	4	0	6.2	0.15

Figure 4: Model 4: stable jumps ($\beta = 1/2$)

$$b(x) = -2x \quad \sigma^2(x) = \xi^2(x) = 1, \quad \nu(dz) = \frac{3}{4z^{3/2}} \mathbb{1}_{|z|\leq 1},$$

the jumps have infinite activity. The Blumenthal-Getoor index is equal to $\beta = 1/2$, and $\xi_1^4 I_4 = 3/7$.

Estimation of σ^2 Estimation of $\sigma^2 + \xi^2$



$n = 10^5$, $\Delta = 10^{-3}$ $n = 10^5$, $\Delta = 10^{-2}$

−. : estimator with true penalty constant ...: robust estimator

Estimation of $\sigma^2 + \xi^2$

Δ	n	Risk				$\hat{D}_{\tilde{m},\tilde{r}}$		$\hat{\Sigma}_1$	
		mean	sd	orpen	or	mean	sd	mean	sd
10^{-1}	10^3	0.29	0.2	-0.12	0	1	0	2.6	1.3
10^{-1}	10^4	0.26	0.1	-0.11	0.011	1.1	0.4	2.0	0.6
10^{-1}	10^5	0.27	0.07	-0.00063	0.11	3.0	0.1	1.6	0.2
10^{-2}	10^3	0.24	0.3	-0.090	0	1	0	1.8	1.7
10^{-2}	10^4	0.078	0.09	-0.047	0	1	0	1.2	0.6
10^{-2}	10^5	**0.037**	0.04	0	0.052	1.0	0.1	0.74	0.16
10^{-3}	10^4	0.23	0.3	-0.073	0	1	0	1.8	1.1
10^{-3}	10^5	0.076	0.09	-0.047	0.0084	1.0	0.07	1.0	0.5

Estimation of σ^2

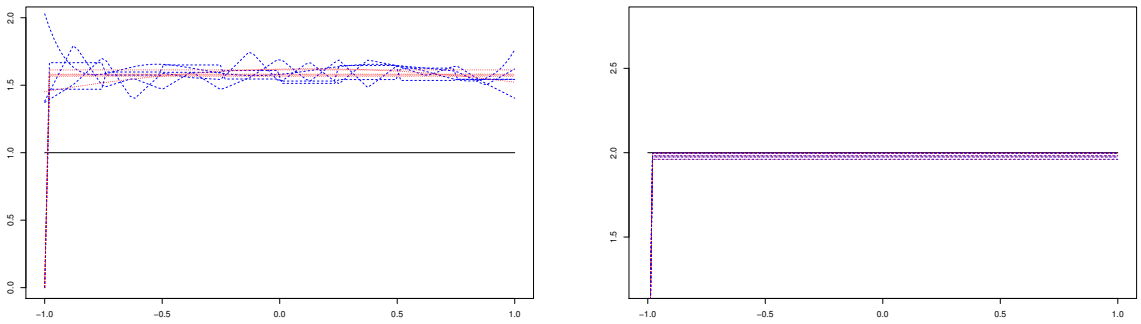
Δ	n	Risk				$D_{\tilde{m},\tilde{r}}$		$\hat{\Sigma}_2$	
		mean	sd	orpen	or	mean	sd	mean	sd
10^{-1}	10^3	0.73	0.4	-0.061	0	1	0	6.7	3.0
10^{-1}	10^4	0.75	0.2	-0.022	0.0015	1.1	0.4	4.7	0.85
10^{-1}	10^5	0.75	0.13	-0.00049	0.014	3	0	4.2	0.25
10^{-2}	10^3	0.53	0.4	-0.14	0.0024	1.0	0.07	7.7	11
10^{-2}	10^4	0.82	0.3	-0.099	0.0038	1.2	0.5	5.0	1.3
10^{-2}	10^5	0.96	0.2	-0.020	0.0018	2.1	3.1	4.5	0.4
10^{-3}	10^4	**0.15**	0.1	-0.041	0.016	1.1	0.4	2.0	0.9
10^{-3}	10^5	**0.21**	0.07	-0.017	0.0078	1.7	1.7	1.7	0.14

Figure 5: Model 5: stable jumps ($\beta = 3/2$)

$$b(x) = -2x \quad \sigma^2(x) = \xi^2(x) = 1, \quad \nu(dz) = \frac{1}{4z^{5/2}} \mathbb{1}_{|z|\leq 1}.$$

Then the Blumethal-Getoor index is equal to 3/2 and $\xi_1^4 I_4 = 1/5$ and $\sigma_1^4 = 1$.

Estimation of σ^2 Estimation of $\sigma^2 + \xi^2$



$n = 10^5$, $\Delta = 10^{-3}$ $n = 10^5$, $\Delta = 10^{-2}$

−. : estimator with true penalty constant ...: robust estimator

Estimation de $\sigma^2 + \xi^2$

Δ	n	Risk				$D_{\tilde{m},\tilde{r}}$		$\hat{\Sigma}_1$	
		mean	sd	orpen	or	mean	sd	mean	sd
10^{-1}	10^3	0.28	0.2	-0.50	0	1	0	2.1	0.9
10^{-1}	10^4	0.27	0.13	-0.16	0.012	1.1	0.39	1.8	0.5
10^{-1}	10^5	0.28	0.070	-0.0032	0.11	3.0	0.1	1.4	0.14
10^{-2}	10^3	0.18	0.20	-0.52	0	1	0	1.1	1.3
10^{-2}	10^4	0.069	0.08	-0.13	0	1	0	0.68	0.3
10^{-2}	10^5	**0.036**	0.035	-0.14	0.020	1.0	0.1	0.44	0.10
10^{-3}	10^4	0.16	0.2	-0.32	0	1	0	0.96	0.7
10^{-3}	10^5	0.054	0.07	-0.2	0.010	1.0	0.07	0.53	0.3

Estimation de σ^2

Δ	n	Risk				$D_{\tilde{m},\tilde{r}}$		$\hat{\Sigma}_2$	
		mean	sd	orpen	or	mean	sd	mean	sd
10^{-1}	10^3	0.74	0.4	-0.042	0	1	0	6.4	2.7
10^{-1}	10^4	0.74	0.21	-0.018	0.0015	1.1	0.3	4.6	0.7
10^{-1}	10^5	0.74	0.12	-0.00020	0.015	3.0	0.07	4.1	0.2
10^{-2}	10^3	0.77	0.42	-0.083	0	1	0	9.3	11
10^{-2}	10^4	0.90	0.30	-0.038	0.00098	1.0	0.25	4.9	0.98
10^{-2}	10^5	0.96	0.18	-0.011	0.00026	1.2	0.52	4.3	0.3
10^{-3}	10^4	**0.53**	0.20	-0.015	0.00076	1.0	0.1	3.3	1.2
10^{-3}	10^5	**0.59**	0.13	-0.0078	0.00051	1.2	0.6	2.8	0.2

We now focus on the Brownian term.

$$\mathbb{E}\left(Z_\Delta^{2p}\middle|\mathscr{F}_0\right) = \mathbb{E}\left(\left(\sigma(X_0)W_\Delta + \tilde{Z}_{k\Delta}\right)^{2p}\middle|\mathscr{F}_0\right)$$
$$= \sigma(X_0)^{2p}\mathbb{E}\left(W_\Delta^{2p}\right) + \sum_{i=1}^{2p} C_{2p}^i \sigma(X_0)^{2p-i}\mathbb{E}\left(W_\Delta^{2p-i}\tilde{Z}_\Delta^i\middle|\mathscr{F}_0\right).$$

By Cauchy-Schwarz,

$$\mathbb{E}\left(Z_\Delta^{2p} - \sigma(X_0)\mathbb{E}\left(W_\Delta^{2p}\right)\middle|\mathscr{F}_0\right) \lesssim \sum_{i=1}^{2p}\left(\mathbb{E}\left(W_\Delta^{4p-2i}\right)\mathbb{E}\left(\tilde{Z}_{k\Delta}^{2i}\middle|\mathscr{F}_0\right)\right)^{1/2} \lesssim \sum_{i=1}^{2p}\left(\Delta^{2p-i}\mathbb{E}\left(\tilde{Z}_\Delta^{2i}\middle|\mathscr{F}_0\right)\right)^{1/2}$$

According to the Burkholder-Davis-Gundy inequality, Hölder inequality and (3),

$$\mathbb{E}\left(\tilde{Z}_\Delta^{2i}\middle|\mathscr{F}_0\right) \lesssim \mathbb{E}\left(\left(\int_0^\Delta (\sigma(X_s)-\sigma(X_{k\Delta}))^2 ds\right)^i\middle|\mathscr{F}_0\right) \lesssim \Delta^{i-1}\mathbb{E}\left(\int_0^\Delta (\sigma(X_s)-\sigma(X_{k\Delta}))^{2i} ds\middle|\mathscr{F}_0\right)$$
$$\lesssim \Delta^{i+1} \tag{7}$$

Therefore, As $\mathbb{E}\left(W_\Delta^{2p}\right) = \Delta^p\mathbb{E}\left(N^{2p}\right)$ where N is a centred reduced gaussian,

$$\mathbb{E}\left(Z_\Delta^{2p} - \sigma(X_0)^{2p}\Delta^p\mathbb{E}\left(N^{2p}\right)\middle|\mathscr{F}_0\right) \lesssim \Delta^{p+1/2}.$$

To bound the jump term, let us first remark that $\mathbb{E}\left(L_\Delta^{2p}\right) = \Delta I_{2p} + C\Delta^2$. Then

$$\mathbb{E}\left(J_{k\Delta}^{2p}\middle|\mathscr{F}_0\right) = \xi^{2p}(X_0)\mathbb{E}\left(L_\Delta^{2p}\right) + \sum_{i=1}^{2p} C_{2p}^i \xi^{2p-i}(X_0)\mathbb{E}\left(L_\Delta^{2p-i}\tilde{J}_\Delta^i\middle|\mathscr{F}_0\right).$$

By the Burkholder-Davis-Gundy inequality and (5), we get:

$$\mathbb{E}\left(\tilde{J}_\Delta^{2p}\right) \lesssim I_{2p}\int_0^\Delta (\xi(X_s)-\xi(X_{k\Delta}))^{2p}\,ds + \left(\int_0^\Delta (\xi(X_s)-\xi(X_0)^2 ds\right)^p$$
$$\lesssim \Delta I_{2p}(\Delta I_{2p} + \Delta^p) + \Delta^p(\Delta I_2)^p \lesssim \Delta^2 I_{2p}^2 + \Delta^{p+1} I_{2p} + \Delta^{2p}. \tag{8}$$

By Cauchy-Schwarz, we obtain:

$$\mathbb{E}\left(J_{k\Delta} - \xi^{2p}(X_0)\mathbb{E}\left(L_\Delta^{2p}\right)\middle|\mathscr{F}_0\right) \lesssim \sum_{i=1}^{2p}\left(\Delta\Delta^2\right)^{1/2} \lesssim \Delta^{3/2}$$

which ends the proof.

6.2 Proof of Lemma 2

We introduce the compensated Poisson measure $\tilde{\mu}(s,z) = \mu(s,z) - \nu(z)s$, and we decompose the Lévy process as follows:

$$L_s^{(1)} = \int_0^s \int_{|z|\leq\Delta^{1/2}} z\tilde{\mu}(ds,dz), \quad L_s^{(2)} = \int_0^s \int_{\Delta^{1/2}<|z|\leq\Delta^{1/4}} z\tilde{\mu}(ds,dz)$$
$$L_s^{(3)} = \int_0^s \int_{\Delta^{1/4}\leq|z|\leq 1} z\tilde{\mu}(ds,dz), \quad L_s^{(4)} = \int_0^s \int_{|z|>1} z\mu(ds,dz) \tag{9}$$

and the random variables $J_{k\Delta}^{(i)} = \int_{k\Delta}^{(k+1)\Delta} \xi(X_{s^-})dL_s^{(i)}$. The Brownian and the very small jumps have a similar behaviour. One big jump term can be quite large, but as many terms are zero, its mean can be bounded.

Bound of the drift terms

As $X_{k\Delta} \in A$, $b(X_{k\Delta})$ is bounded. Let us set $\tilde{B}_{k\Delta} := B_{k\Delta} - \Delta b(X_{k\Delta}) = \int_{k\Delta}^{(k+1)\Delta}(b(X_s) - b(X_{k\Delta}))ds$. By Hölder inequality and Lemma 1, as b is Lipschitz, if $\mathbb{E}\left(X_0^{2p}\right) < \infty$,

$$\begin{aligned}\mathbb{E}\left(\left(\tilde{B}_{k\Delta}\right)^{2p}\middle|\mathscr{F}_{k\Delta}\right) &\leq \mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta}(b(X_s)-b(X_{k\Delta}))^{2p}ds\left(\int_{k\Delta}^{(k+1)\Delta}1^{2p/(2p-1)}\right)^{2p-1}\middle|\mathscr{F}_{k\Delta}\right)\\ &\leq \mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta}(X_s - X_{k\Delta})^{2p}ds\Delta^{2p-1}\middle|\mathscr{F}_{k\Delta}\right) \lesssim \Delta^{2p+1}.\end{aligned} \tag{10}$$

Therefore, by Markov inequality, for any $\epsilon \in]0,1[$, any $p \geq 1$,

$$\mathbb{P}\left(|B_{k\Delta}| \geq \Delta^{1-\epsilon}\right) \lesssim \Delta^{-2p+2p\epsilon}\mathbb{E}\left(\tilde{B}_{k\Delta}^{2p}\right) \lesssim \Delta^{1+2p\epsilon}.$$

By Assumption A5, $\Delta = O(n^{-\delta})$ and then $\mathbb{P}\left(|B_{k\Delta}| \geq \Delta^{1-\epsilon}\right) \lesssim n^{-r}$.

Bound of the Brownian terms

By Markov inequality,

$$\begin{aligned}\mathbb{P}\left(Z_{k\Delta} \geq r\sigma_1\Delta^{1/2}\ln(n)\right) &\leq n^{-r}\mathbb{E}\left[\exp\left(\frac{1}{\sigma_1\Delta^{1/2}}Z_{k\Delta}\right)\right]\\ &\leq n^{-r}\mathbb{E}\left[\exp\left(\frac{1}{\sigma_1^2\Delta}\int_{k\Delta}^{(k+1)\Delta}\sigma^2(X_s)ds\right)\right] \leq n^{-r}.\end{aligned}$$

By symmetry, $\mathbb{P}\left(Z_{k\Delta} \leq -r\sigma_1\Delta^{1/2}\ln(n)\right) \leq n^{-r}$ and $\mathbb{P}\left(|Z_{k\Delta}| \geq r\sigma_1\Delta^{1/2}\ln(n)\right) \leq 2n^{-r}$.

Bound of the small jumps.

The terms $J^{(1)}_{k\Delta}$ are small and can be bounded in the same way as the Brownian terms $Z_{k\Delta}$. As the terms are stationary, we focus on the bound of $J^{(1)}_{\Delta}$. Let us set

$$K_{a,t} = \int_0^t \int_{-\Delta^{1/2}}^{\Delta^{1/2}} \left(e^{a\xi(X_s)z} - 1 - a\xi(X_s)z \right) \nu(dz)ds \quad \text{and} \quad V_{a,t} = a\int_0^t \xi(X_{s^-})dL^{(1)}_s - K_{a,t}.$$

According to Corollary 5.2.2 of Applebaum (2004), for any $a \geq 0$, $t > 0$, $\exp(V_{a,t})$ is a local martingale. We consider a sequence of increasing stopping times (τ_N) such that $\tau_N \to \infty$ and $\exp(V_{a,t\wedge\tau_N})$ is a martingale. Then $\mathbb{E}\left(\exp(V_{a,\Delta\wedge\tau_N}\right) = 1$. We can remark that for any $a \leq 1/(2\xi_1\Delta^{1/2})$,

$$K_{a,t} \leq \int_0^t \int_{-\Delta^{1/2}}^{\Delta^{1/2}} a^2z^2\xi^2(X_s)\nu(dz)ds \leq \xi_1^2a^2t\int_{-\Delta^{1/2}}^{\Delta^{1/2}} z^2\nu(dz).$$

We recall that β is the Blumenthal-Getoor index (see Assumption A2), and therefore

$$\int_{-\Delta^{1/2}}^{\Delta^{1/2}} z^2\nu(dz) = \int_{-\Delta^{1/2}}^{\Delta^{1/2}} z^\beta z^{2-\beta}\nu(dz) \lesssim \Delta^{1-\beta/2}.$$

We get $K_{a,\Delta} \leq \xi_1^2a^2\Delta^{2-\beta/2}$. By Markov's inequality,

$$\begin{aligned} S_1 &:= \mathbb{P}\left(\int_0^{\Delta\wedge\tau_N} \xi(X_{s^-})dL^{(1)}_s \geq 2r\xi_1\Delta^{1/2}\ln(n) \right) \\ &\leq \mathbb{P}\left(V_{a,\Delta\wedge\tau_N} \geq \exp\left(2ra\xi_1\Delta^{1/2}\ln(n) - \xi_1^2a^2\Delta^{2-\beta/2} \right) \right) \\ &\leq \exp\left(-2ra\xi_1\Delta^{1/2}\ln(n) + \xi_1^2a^2\Delta^{2-\beta/2} \right). \end{aligned}$$

Letting $N \to \infty$, and $a = 1/(2\xi_1\Delta^{1/2})$, we obtain $\mathbb{P}\left(J^{(1)}_\Delta \geq 2r\xi_1\Delta^{1/2}\ln(n) \right) \lesssim \exp\left(-r\ln(n)\right) \leq n^{-r}$. We prove in the same way that $\mathbb{P}\left(-J^{(1)}_\Delta \geq 2r\xi_1\Delta^{1/2}\ln(n) \right) \lesssim n^{-r}$. Therefore:

$$\mathbb{P}\left(\left| J^{(1)}_\Delta \right| \geq 2r\xi_1\Delta^{1/2}\ln(n) \right) \lesssim n^{-r}. \tag{11}$$

Bound for the jumps greater than $\Delta^{1/2}$.

We want to bound $\frac{1}{q_n}\sum_{k=jq_n}^{(j+1)q_n-1}(J^{(4)}_{k\Delta})^{2p}$. Let us first prove that the random variables $J^{(4)}_{k\Delta}$ are not too large. We have to bound both the number of jumps in the time interval $[k\Delta,(k+1)\Delta[$ and the size of the jumps.

The probability of having a very large jump is quite small: by Assumption A5,

$$\nu\left(\left[-\frac{(r+2)\ln(n)}{\lambda}, \frac{(r+2)\ln(n)}{\lambda} \right]^c \right) \lesssim n^{-(r+2)}. \tag{12}$$

The number of jumps greater than $(r+2)\ln(n)/\lambda$ on the whole interval $[0,n\Delta]$ follows a Poisson law of parameter smaller than $Cn^{-(r+2)}$. Then

$$\mathbb{P}\left(\mu\left([0,n\Delta], \left[-\frac{(r+2)\ln(n)}{\lambda}, \frac{(r+2)\ln(n)}{\lambda} \right]^c \right) \geq 1 \right) \leq 1 - e^{-cn^{-(r+2)}n\Delta} \simeq \frac{\Delta}{n^{r+1}}.$$

The probability of having more than $(r+2)C$ (with $C = 1/(\delta(2-\beta))$ (see Assumption A5 for the definition of δ)) jumps greater than $\Delta^{1/2}$ in a time interval Δ is very low. Let us set

$$\mathscr{E}_\Delta := \left\{\omega, \forall k, \mu\left([k\Delta,(k+1)\Delta], \left[-\Delta^{1/2},\Delta^{1/2}\right]^c\right) < (r+2)C\right\}.$$

The number of jumps greater than $\Delta^{1/2}$ in the time interval $[k\Delta,(k+1)\Delta]$ is a Poisson variable of parameter $\Delta\int_{|z|\geq\Delta^{1/2}}\nu(dz) \leq \Delta^{2-\beta}\int z^\beta\nu(dz) \lesssim \Delta^{2-\beta}$. Then

$$\mathbb{P}\left(\mathscr{E}_\Delta^c\right) \lesssim n(\Delta^{2-\beta})^{(r+2)C} = n\Delta^{(r+2)/\delta} \lesssim n^{-(r+1)} \tag{13}$$

By (12) and (13),

$$\mathbb{P}\left(|J_{k\Delta}^{(2)+(3)+(4)}| \geq \frac{(r+2)^2C\ln(n)}{\lambda}\right) \lesssim n^{-(r+1)}.$$

Bound of the mean of $J_{k\Delta}^{(4)}$. Most of the terms $J_{k\Delta}^{(4)}$ are equal to 0. To bound $\sum_{k=1}^{q_n}\left(J_{k\Delta}^{(4)}\right)^{2p}$, we have to bound the number of non-zero terms, which is bounded by the number of jumps greater than 1 in the time interval $[jq_n\Delta,(j+1)q_n\Delta]$. We use the following property:

If Y follows a Poisson law of parameter θ_n, then for any $\lambda > 0$,

$$\mathbb{P}\left(Y \geq c_n\right) \leq e^{-\lambda c_n}\mathbb{E}\left(\exp(\lambda Y)\right) \leq e^{-\lambda c_n}\exp\left(\theta_n(e^\lambda-1)\right).$$

This inequality is minimal for $\lambda = \ln(c_n/\theta_n)$ and then: $\mathbb{P}\left(Y \geq c_n\right) \leq \left(\frac{\theta_n}{c_n}\right)^{c_n} e^{-\theta_n+c_n}$. Therefore, for any $\lambda \geq 1$:

$$\mathbb{P}\left(Y \geq \lambda\theta_n\right) \leq e^{-\theta_n(\lambda-1)}. \tag{14}$$

Let us set $v_1 := \nu(]-1,1[^c) \vee 1$ and

$$\mathscr{E}_p^{(4)} := \left\{\sum_{k=1}^{q_n}\left|J_{k\Delta}^{(4)}\right|^{2p} \geq 2\nu_1 q_n\Delta\left(\frac{(r+2)^2C\Delta\ln(n)}{\lambda}\right)^{2p}\right\}.$$

If $\mathscr{E}_p^{(4)}$ is true, then either one $|J_{k\Delta}|$ is greater than $(2r+2)^2C\ln(n)/\lambda$, either there are more than $2v_1\Delta q_n$ jumps. As $\mu\left([0,q_n\Delta[,[-1,1]^c\right)$ is a compound Poisson process of intensity smaller than $v_1\Delta q_n$, we obtain by (14),

$$\begin{aligned}\mathbb{P}\left(\mathscr{E}_p^{(4)}\right) &\lesssim q_n\mathbb{P}\left(|J_{k\Delta}^{(4)}| \geq \frac{(r+2)^2C\ln(n)}{\lambda}\right) + \mathbb{P}\left(\mu\left([0,q_n\Delta[,[-1,1]^c\right) \geq 2v_1\Delta q_n\right)\\ &\lesssim \frac{q_n}{n^{r+1}} + \exp\left(-3q_n\Delta v_1\right) \lesssim \frac{1}{n^{r+1}}.\end{aligned}$$

Bound of the mean of $|J_{k\Delta}^{(2)}+J_{k\Delta}^{(3)}|$. Let us set $L_s^{(i)+(j)} = L_s^{(i)}+L_s^{(j)}$ and $J_{k\Delta}^{(i)+(j)} = J_{k\Delta}^{(i)}+J_{k\Delta}^{(j)}$. By (13), the probability of having more than $(r+2)C$ jumps greater than $\Delta^{1/2}$ in a time interval of length Δ is smaller than $n^{-(r+2)}$. Therefore

$$\mathbb{P}\left(\exists k, \mu\left([k\Delta,(k+1)\Delta],[-\Delta^{1/2},\Delta^{1/2}]^c\right) \geq (r+2)C\right) \lesssim n^{-r+1}.$$

We have:

$$\sum_{k=1}^{q_n} \left| J_{k\Delta}^{(2)+(3)} \right|^{2p} \mathbb{1}_{\mathscr{E}_\Delta} \leq \xi_1^{2p} (r+2)^{2p} C^{p-1} \int_0^{q_n\Delta} |\Delta L_s^{(2+3)}|^{2p}$$

where $\int_0^{q_n\Delta} |\Delta L_s^{(2+3)}|^{2p}$ is the sum of the $2p$-power of all the jumps of size between $\Delta^{1/2}$ and 1 in the time interval $[0, q_n\Delta]$. Then

$$\mathbb{P}\left(\sum_{k=1}^{q_n} |J_{k\Delta}^{(2+3)}|^{2p} \geq c \right) \leq Cn^{-r+1} + \mathbb{P}\left(\int_0^{q_n\Delta} |\Delta L_s^{(2+3)}|^{2p} \geq c/((r+2)^{2p} C^{2p-1} \xi_1^{2p}) \right).$$

As $\int_0^{q_n\Delta} |\Delta L_s^{(2+3)}|^{2p}$ is a compound Poisson process, we have, for any $a \leq 1$:

$$\begin{aligned}
\mathbb{E}\left(\exp\left(a \int_0^{q_n\Delta} |\Delta L_s^{(2+3)}|^{2p} \right) \right) &= \exp\left(q_n\Delta \int_{\Delta^{1/2} \leq |z| \leq 1} (e^{az^{2p}} - 1)\nu(dz) \right) \\
&\leq \exp\left(2aq_n\Delta \int_{\Delta^{1/2} \leq |z| \leq 1} z^{2p} \nu(dz) \right) \leq \exp(2aq_n\Delta I_{2p}).
\end{aligned}$$

Let us set $a = (r+1)/(q_n\Delta)$. By Markov inequality, we obtain:

$$\begin{aligned}
\mathbb{P}\left(\int_0^{q_n\Delta} |\Delta L_s^{(2+3)}|^{2p} \geq q_n \ln(n)\Delta \right) &\leq \mathbb{E}\left(\exp\left(a \int_0^{q_n\Delta} |\Delta L_s^{(2+3)}|^{2p} \right) \right) \exp\left(-a\Delta \ln(n) q_n \right) \\
&= \exp\left(2(r+1) I_{2p} \right) n^{-(r+1)} \lesssim n^{-(r+1)}.
\end{aligned}$$

Therefore

$$\mathbb{P}\left(\frac{1}{q_n} \sum_{k=0}^{q_n} |J_{k\Delta}^{(2+3)}|^{2p} \geq (r+2)^{2p} C^{2p-1} \xi_1^{2p} \ln(n)\Delta \right) \lesssim \frac{1}{n^{r+1}}. \tag{15}$$

6.3 Proof of Lemma 3

Bound of the small term. Let us set

$$\begin{aligned}
\Delta E_{k\Delta}^{(1)} &:= (B_{k\Delta} + Z_{k\Delta} + J_{k\Delta})^2 - (Z_{k\Delta} + J_{k\Delta})^2 \\
\Delta E_{k\Delta}^{(2)} &:= \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^2 \middle| \mathscr{F}_{k\Delta} \right) - \Delta\sigma^2(X_{k\Delta}) - \Delta\xi^2(X_{k\Delta})
\end{aligned}$$

We want to prove that $\Delta^2 \mathbb{E}\left(E_{k\Delta}^2\right) \lesssim \Delta^3$ and $\Delta^4 \mathbb{E}\left(E_{k\Delta}^4\right) \lesssim \Delta^5$. We focus first on the term $E_{k\Delta}^{(1)}$. We can write

$$\Delta E_{k\Delta}^{(1)} = B_{k\Delta}^2 + 2B_{k\Delta}(Z_{k\Delta} + J_{k\Delta}) = B_{k\Delta}^2 + 2\Delta b(X_{k\Delta})(Z_{k\Delta} + J_{k\Delta}) + 2\tilde{B}_{k\Delta}(Z_{k\Delta} + J_{k\Delta})$$

where $\tilde{B}_{k\Delta} = \int_{k\Delta}^{(k+1)\Delta} (b(X_s) - b(X_{k\Delta}) ds$. Then, for $p > 0$, by (10) and Lemma 1,

$$\begin{aligned}
\Delta^{2p} \mathbb{E}\left((E_{k\Delta}^{(1)})^{2p} \middle| \mathscr{F}_{k\Delta} \right) &\lesssim \Delta^{4p} + \Delta^{2p} \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^{2p} \middle| \mathscr{F}_{k\Delta} \right) + \mathbb{E}\left(\tilde{B}_{k\Delta}^{2p} (Z_{k\Delta} + J_{k\Delta})^{2p} \middle| \mathscr{F}_{k\Delta} \right) \\
&\lesssim \Delta^{4p} + \Delta^{2p+1} + \left(\mathbb{E}\left(\tilde{B}_{k\Delta}^{4p} \middle| \mathscr{F}_{k\Delta} \right) \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^{4p} \middle| \mathscr{F}_{k\Delta} \right) \right)^{1/2}. \\
&\lesssim \Delta^{4p} + \Delta^{2p+1} + \Delta^{2p+1} \lesssim \Delta^{2p+1}.
\end{aligned} \tag{16}$$

It remains to bound $E^{(2)}_{k\Delta}$. As $\int z^2\nu(dz)=1$, by Lemma 1,

$$\Delta E^{(2)}_{k\Delta} = \mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta}(\sigma^2(X_s)-\sigma^2(X_{k\Delta}))ds\middle|\mathscr{F}_{k\Delta}\right)+\mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta}(\xi^2(X_s)-\xi^2(X_{k\Delta}))ds\middle|\mathscr{F}_{k\Delta}\right)$$
$$+\mathbb{E}\left(2Z_{k\Delta}J_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right).$$

Let us prove that $\mathbb{E}\left(Z_{k\Delta}J_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)=0$. We consider two predictable (with respect to $\mathscr{F}_t$) piecewise stationary processes A_t and B_t. Then

$$\mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta}A_t dW_t\int_{k\Delta}^{(k+1)\Delta}B_s dL_s\right)=\sum_{i,j}\mathbb{E}\left(A_{t_i}B_{t_j}(W_{t_{i+1}}-W_{t_i})(L_{t_{j+1}}-L_{t_j})\middle|\mathscr{F}_{\max(t_i,t_j)}\right)=0.$$

By taking the limit in L^2, we obtain that

$$\mathbb{E}\left(Z_{k\Delta}J_{k\Delta}\right)=0. \tag{17}$$

By Lemma 1, we get that $\Delta E^{(2)}_{k\Delta}\lesssim\Delta^{3/2}$. Therefore $\mathbb{E}\left(E^2_{k\Delta}\right)\lesssim\Delta$ and $\mathbb{E}\left(E^4_{k\Delta}\right)\lesssim\Delta$.

Bound of the centred terms. By (17), we get that $\mathbb{E}\left(G_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)=0$ and by definition, $\mathbb{E}\left(F_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)=0$. Moreover, by Lemma 1, $\mathbb{E}\left(Z^4_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)=3\Delta^2\sigma^4(X_{k\Delta})+C\Delta^{2+1/2}$ and $\mathbb{E}\left(Z^8_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)\lesssim\Delta^4$. Therefore

$$\Delta^2\mathbb{E}\left(F^2_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)=\mathrm{Var}\left(Z^2_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)=2\Delta^2\sigma^4(X_{k\Delta})+C\Delta^{2+1/2}$$

and $\mathbb{E}\left(F^4_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)\lesssim 1$. By Cauchy-Schwarz and Lemma 1, $\mathbb{E}\left(J^2_{k\Delta}Z^2_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)\lesssim\Delta^{3/2}$. By (17), $\mathbb{E}\left(Z_{k\Delta}J_{k\Delta}(J^2_{k\Delta}-\mathbb{E}\left(J^2_{k\Delta}\right)\middle|\mathscr{F}_{k\Delta}\right)=\mathbb{E}\left(Z_{k\Delta}J^3_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)$. We can write

$$Z_\Delta J^3_\Delta=\sigma(X_0)W_\Delta\left(\xi^3(X_0)L^3_\Delta+3\xi^2(X_0)L_\Delta\tilde{J}_\Delta+3\xi(X_0)L_\Delta\tilde{J}^2_\Delta+\tilde{J}^3_\Delta\right)+\tilde{Z}_\Delta J^3_\Delta$$

We have that $\mathbb{E}\left(W_\Delta L^3_\Delta\sigma(X_0)\xi^3(X_0)\middle|\mathscr{F}_0\right)=0$. To bound the other terms, we use Cauchy-Schwarz inequality, Lemma 1, (7) and (8). We get that $\mathbb{E}\left(Z_{k\Delta}J^3_{k\Delta}\right)\lesssim\Delta^{3/2}$. Then

$$\Delta^2\mathbb{E}\left(G^2_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right)=\xi^4(X_{k\Delta})I_4\Delta+C\Delta^{3/2}.$$

By Cauchy-Schwarz, $\mathbb{E}\left(\Delta^2F_{k\Delta}G_{k\Delta}\right)\lesssim\Delta^{3/2}$ and therefore

$$\Delta^2\mathbb{E}\left((F_{k\Delta}+G_{k\Delta})^2\middle|\mathscr{F}_{k\Delta}\right)=\xi^4(X_{k\Delta})I_4\Delta+C\Delta^{3/2}$$

which ends the proof.

6.4 Proof of Proposition 4 (risk of the estimator of $\sigma^2+\xi^2$ for m fixed)

We have that

$$\gamma_{1,n}(t)=\frac{1}{n}\sum_{k=1}^n\left(t(X_{k\Delta})-g(X_{k\Delta})-E_{k\Delta}-F_{k\Delta}-G_{k\Delta}\right)^2$$
$$=\|t-g\|^2_n+\frac{1}{n}\sum_{k=1}^n(E_{k\Delta}+F_{k\Delta}+G_{k\Delta})^2-\frac{2}{n}\sum_{k=1}^n\left(E_{k\Delta}+F_{k\Delta}+G_{k\Delta}\right)\left(g(X_{k\Delta})-t(X_{k\Delta})\right).$$

As $\hat{g}_m$ minimizes $\gamma_{1,n}(t)$, the inequality $\gamma_{1,n}(\hat{g}_m) \leq \gamma_{1,n}(\Pi_m g)$, where $\Pi_m g$ is the orthogonal projection of g on S_m for the $\|.\|_n$-norm, holds and then

$$\|\hat{g}_m - g\|_n^2 \leq \|\Pi_m g - g\|_n^2 + \frac{2}{n}\sum_{k=1}^{n} (E_{k\Delta} + F_{k\Delta} + G_{k\Delta}) (\hat{g}_m(X_{k\Delta}) - \Pi_m g(X_{k\Delta})) .$$

As $\hat{g}_m$ and $\Pi_m g$ are A-supported,

$$\|\hat{g}_m - g_A\|_n^2 \leq \|\Pi_m g - g_A\|_n^2 + \frac{2}{n}\sum_{k=1}^{n} (E_{k\Delta} + F_{k\Delta} + G_{k\Delta}) (\hat{g}_m(X_{k\Delta}) - \Pi_m g(X_{k\Delta})) .$$

where $g_{m,\pi}$ is the orthogonal projection for the $\|.\|_\pi$-norm. As $\|\hat{g}_m - g_A\|_n^2 = \|\hat{g}_m - \Pi_m g\|_n^2 + \|\Pi_m g - g_A\|_n^2$, we get:

$$\|\hat{g}_m - \Pi_m g\|_n^2 \leq \frac{2}{n}\sum_{k=1}^{n} (E_{k\Delta} + F_{k\Delta} + G_{k\Delta}) (\hat{g}_m(X_{k\Delta}) - \Pi_m g(X_{k\Delta})) . \tag{18}$$

By geometric and arithmetic means inequality,

$$\frac{2}{n}\sum_{k=1}^{n} E_{k\Delta}(\hat{g}_m(X_{k\Delta}) - \Pi_m g(X_{k\Delta})) \leq \frac{6}{n}\sum_{k=1}^{n} E_{k\Delta}^2 + \frac{1}{6}\|\hat{g}_m - \Pi_m g\|_n^2 . \tag{19}$$

Let us introduce the unit random ball for the $\|.\|_\pi^2$-norm, $\mathscr{B}_m = \{t \in S_m, \|t\|_\pi \leq 1\}$ and the random function $\nu_{1,n}(t) = \frac{1}{n}\sum_{k=1}^{n}(F_{k\Delta} + G_{k\Delta})t(X_{k\Delta})$. By geometric and arithmetic means inequality,

$$2\nu_{1,n}(\hat{g}_m - \Pi_m g) \leq \|\hat{g}_m - \Pi_m g\|_\pi \sup_{t\in\mathscr{B}_m} \nu_{1,n}(t) \leq \frac{1}{6}\|\hat{g}_m - \Pi_m g\|_\pi^2 + 6 \sup_{t\in\mathscr{B}_m} \nu_{1,n}^2(t). \tag{20}$$

Replacing (19) and (20) in equation (18), we obtain:

$$\|\hat{g}_m - \Pi_m g\|_n^2 \leq \frac{6}{n}\sum_{k=1}^{n} E_{k\Delta}^2 + \frac{1}{6}\|\hat{g}_m - \Pi_m g\|_n^2 + 6 \sup_{t\in\mathscr{B}_m} \nu_{1,n}^2(t) + \frac{1}{6}\|\hat{g}_m - \Pi g_m\|_\pi^2 .$$

Let us set

$$\Omega_n = \left\{ \omega,\ \forall m \in \mathscr{M}_n, \forall t \in S_m, \left| \frac{\|t\|_n^2}{\|t\|_\pi^2} - 1 \right| \leq \frac{1}{2} \right\}$$

where the norms $\|.\|_\pi$ and $\|.\|_n$ are equivalent. The following lemma is proved by Comte et al. (2007) (section 7) for diffusion processes, but only relies on the β-mixing and stationary properties of the process (X_t).

Lemma 16.

$$\mathbb{P}(\Omega_n^c) \leq \frac{c}{n^8}.$$

On Ω_n, any function $t \in S_m$ satisfies $\|t\|_\pi^2 \leq 2\|t\|_n^2$. Then

$$\|\hat{g}_m - \Pi_m g\|_n^2 \mathbb{1}_{\Omega_n} \leq \frac{12}{n}\sum_{k=1}^{n} E_{k\Delta}^2 + 12 \sup_{t\in\mathscr{B}_m} \nu_{1,n}^2(t). \tag{21}$$

By Lemma 3, $\mathbb{E}\left(E_{k\Delta}^2\right) \lesssim \Delta$. It remains to bound $\mathbb{E}\left(\sup_{t\in\mathscr{B}_m} \nu_{1,n}^2(t)\right)$. Let $(\varphi_\lambda)_{1\leq\lambda\leq D_m}$ be an orthonormal (for the $\|.\|_\pi$ norm) basis of S_m. Any function $t \in \mathscr{B}_m$ can be written $t = \sum_{\lambda=1}^{D_m} a_\lambda \varphi_\lambda$ with $\sum_{\lambda=1}^{D_m} a_\lambda^2 \leq 1$. By Cauchy-Schwarz,

$$\sup_{t\in\mathscr{B}_m} \nu_{1,n}^2(t) \leq \sup_{\sum_{\lambda=1}^{D_m} a_\lambda^2 \leq 1} \left(\sum_{\lambda=1}^{D_m} a_\lambda^2\right) \left(\sum_{\lambda=1}^{D_m} \nu_{1,n}^2\left(\varphi_\lambda\right)\right) \leq \sum_{\lambda=1}^{D_m} \nu_{1,n}^2(\varphi_\lambda). \tag{22}$$

As $\mathbb{E}\left(F_{k\Delta} + G_{k\Delta} | \mathscr{F}_{k\Delta}\right) = 0$, we get

$$\mathbb{E}\left(\nu_{1,n}^2(\varphi_\lambda)\right) = \mathbb{E}\left[\left(\frac{1}{n}\sum_{k=1}^n \left(F_{k\Delta} + G_{k\Delta}\right)\varphi_\lambda(X_{k\Delta})\right)^2\right] \leq \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left[\varphi_\lambda^2(X_{k\Delta})\mathbb{E}\left(\left(F_{k\Delta} + G_{k\Delta}\right)^2 \middle| \mathscr{F}_{k\Delta}\right)\right].$$

By Lemma 2 and stationarity,

$$\mathbb{E}\left(\nu_{1,n}^2(\varphi_\lambda)\right) \leq \frac{1}{n}\mathbb{E}\left(\left(\frac{I_4\xi^4(X_0)}{\Delta} + \frac{C}{\sqrt{\Delta}}\right)\varphi_\lambda^2(X_0)\right) = \frac{I_4}{n\Delta}\int_A (\xi^4(y) + C\sqrt{\Delta})\varphi_\lambda^2(y)\pi(y)dy \leq \frac{\xi_{1,A}^4 I_4}{n\Delta}$$

as $(\varphi_\lambda)_{\lambda\in\Lambda}$ is an orthonormal basis of S_m for the L_π^2-norm. Therefore

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m} \nu_{1,n}^2(t)\right) \leq \frac{D_m}{n\Delta}\xi_{1,A}^4 I_4. \tag{23}$$

We can obtain another bound for $\mathbb{E}\left(\sup_{t\in\mathscr{B}_m} \nu_{1,n}^2(t)\right)$. Indeed, if $t \in \mathscr{B}_m$, then $\|t\|_{L^2}^2 \leq 1/\pi_0$. Then if $(\psi_\lambda)_{\lambda\in\Lambda}$ is an orthonormal basis (for the norm L^2) of S_m,

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\mathscr{B}_m} \nu_{1,n}^2(t)\right) &\leq \mathbb{E}\left(\sup_{t\in S_m, \|t\|_{L^2}^2 \leq 1/\pi_0} \nu_{1,n}^2(t)\right) \leq \frac{1}{\pi_0}\sum_{\lambda\in\Lambda} \mathbb{E}\left(\nu_{1,n}^2(\psi_\lambda)\right) \\ &\leq \frac{1}{\pi_0 n^2}\sum_{k=1}^n \mathbb{E}\left(\sum_{\lambda\in\Lambda} \psi_\lambda^2(X_{k\Delta})\left(F_{k\Delta} + G_{k\Delta}\right)^2\right).\end{aligned}$$

By Assumption 8, for any x, $\sum_{\lambda\in\Lambda} \psi^2(x) \leq \phi_1 D_m$ and therefore

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m} \nu_{1,n}^2(t)\right) \leq \frac{\phi_1 D_m}{n\pi_0}\mathbb{E}\left(\frac{\xi^4(X_{k\Delta})I_4}{\Delta} + C\Delta^{-1/2}\right) = \frac{D_m}{n\Delta}\frac{\phi_1 I_4}{\pi_0}\int_{\mathbb{R}} \xi^4(z)\pi(z)dz.$$

It remains to bound the risk on Ω_n^c. The function $\hat{g}_m$ minimises $\gamma_n(t) = n^{-1}\sum_{k=1}^n (T_{k\Delta} - t(X_{k\Delta}))^2$. Therefore $\hat{g}_m$ is the orthogonal projection for the $\|.\|_n$- norm of $(T_\Delta, \ldots, T_{n\Delta})$ on the vectorial subspace: $\{(t(X_\Delta), \ldots, t(X_{n\Delta})), t \in S_m\}$. As $T_{k\Delta} = g(X_{k\Delta}) + E_{k\Delta} + F_{k\Delta} + G_{k\Delta}$, we obtain:

$$\|\hat{g}_m - \Pi_m g\|_n^2 = \|\Pi_m T - g\|_n^2 = \|\Pi_m E + \Pi_m F + \Pi_m G\|_n^2 \leq \|E + F + G\|_n^2.$$

By stationarity and Cauchy-Schwarz:

$$\begin{aligned}\mathbb{E}\left(\|\hat{g}_m - \Pi_m g\|_n^2 \mathbb{1}_{\Omega_n^c}\right) &\lesssim \mathbb{E}\left[\left(\frac{1}{n}\sum_{k=1}^n E_{k\Delta}^2 + F_{k\Delta}^2 + G_{k\Delta}^2\right)\mathbb{1}_{\Omega_n^c}\right] \\ &\lesssim \left[\mathbb{E}\left(E_\Delta^4 + F_\Delta^4 + G_\Delta^4\right)\mathbb{P}\left(\Omega_n^c\right)\right]^{1/2}.\end{aligned}$$

By Lemma 16, $\mathbb{P}(\Omega_n^c) \lesssim 1/n^8$ and therefore

$$\mathbb{E}\left(\|\hat{g}_m - \Pi_m g\|_n^2 \mathbb{1}_{\Omega_n^c}\right) \lesssim \frac{1}{\Delta^3 n^4} \leq \frac{1}{n}.$$

Then

$$\mathbb{E}\left(\|\hat{g}_m - g_A\|_n^2\right) \leq \mathbb{E}\left(\|\Pi_m g - g_A\|_n^2\right) + 12\Sigma_1 \frac{D_m}{n\Delta} + C\Delta.$$

We can remark that, as $\Pi_m g$ is the orthogonal projection of g for the empirical norm on S_m, $\|\Pi_m g - g_A\|_n^2 \leq \|g_{m,\pi} - g_A\|_n^2$. Therefore $\mathbb{E}\left(\|\hat{g}_m - g_A\|_n^2\right) \leq \mathbb{E}\left(\|g_{m,\pi} g - g_A\|_n^2\right) + 12\Sigma_1 \frac{D_m}{n\Delta} + C\Delta$. As, for any function t, deterministic, $\mathbb{E}\left(\|t\|_n^2\right) = \|t\|_\pi^2$, we get:

$$\mathbb{E}\left(\|\hat{g}_m - g_A\|_n^2\right) \leq \|g_{m,\pi} - g\|_\pi^2 + 12\Sigma_1 \frac{D_m}{n\Delta} + C\Delta.$$

6.5 Proof of Corollary 5

We can remark that $\|\hat{g}_m - g\|_\pi^2 = \|\hat{g}_m - g_{m,\pi}\|_\pi^2 + \|g_{m,\pi} - g\|_\pi^2$. As $\hat{g}_m$ and $g_{m,\pi}$ belongs to S_m,

$$\begin{aligned}
\|\hat{g}_m - g_{m,\pi}\|_\pi^2 \mathbb{1}_{\Omega_n} &\leq 2\|\hat{g}_m - g_{m,\pi}\|_n^2 = 2\|\hat{g}_m - \Pi_m g\|_n^2 + 2\|\Pi_m g - g_{m,\pi}\|_n^2 \\
&\leq 2\|\hat{g}_m - \Pi_m g\|_n^2 + 4\|\Pi_m g - g\|_n^2 + 4\|g - g_{m,\pi}\|_n^2 \\
&\leq 2\|\hat{g}_m - \Pi_m g\|_n^2 + 8\|g_{m,\pi} - g\|_n^2. \qquad (24)
\end{aligned}$$

As $\mathbb{E}(\|t\|)_n^2 = \mathbb{E}\left(\|t\|_\pi^2\right)$ for any deterministic function t, we have that:

$$\mathbb{E}\left(\|\hat{g}_m - g_{m,\pi}\|_\pi^2 \mathbb{1}_{\Omega_n}\right) \leq 2\mathbb{E}\left(\|\hat{g}_m - \Pi_m g\|_n^2\right) + 8\|g_{m,\pi} - g\|_\pi^2.$$

And

$$\mathbb{E}\left(\|\hat{g}_m - g_{m,\pi}\|_\pi^2 \mathbb{1}_{\Omega_n^c}\right) \leq \left(\mathbb{E}\left(\|\Pi_m T\|_\pi^4\right) + \|g\|_\pi^4\right)^{1/2} (\mathbb{P}(\Omega_n^c))^{1/2} \lesssim \frac{1}{n}.$$

6.6 Proof of Theorem 6 (oracle inequality for $\hat{g}_{\hat{m}}$)

For any m, $\gamma_{1,n}(\hat{g}_{\hat{m}}) + \mathrm{pen}_1(\hat{m}) \leq \gamma_{1,n}(\Pi_m g) + \mathrm{pen}_1(m)$. Therefore, as

$$\gamma_n(t) = \|t - g\|_n^2 + \frac{1}{n}\sum_{k=1}^n (E_{k\Delta} + F_{k\Delta} + G_{k\Delta})^2 + \frac{2}{n}\sum_{k=1}^n (E_{k\Delta} + F_{k\Delta} + G_{k\Delta})(g(X_{k\Delta}) - t(X_{k\Delta}),$$

we obtain:

$$\|\hat{g}_{\hat{m}} - g_A\|_n^2 \leq \|\Pi_m g - g_A\|_n^2 + \mathrm{pen}_1(m) - \mathrm{pen}_1(\hat{m}) + \frac{2}{n}\sum_{k=1}^n (E_{k\Delta} + F_{k\Delta} + G_{k\Delta})(\hat{g}_{\hat{m}}(X_{k\Delta}) - \Pi_m g(X_{k\Delta})).$$

Let us set $\mathscr{B}_{m,m'} = \{t \in S_m + S_{m'}, \|t\|_\pi \leq 1\}$. Then, by geometric-arithmetic mean inequality,

$$\|\hat{g}_{\hat{m}} - \Pi_m g\|_n^2 \mathbb{1}_{\Omega_n} \leq \frac{12}{n}\sum_{k=1}^n E_{k\Delta}^2 + 12 \sup_{t \in \mathscr{B}_{m,\hat{m}}} \nu_{1,n}^2(t) + 2\,\mathrm{pen}_1(m) - 2\,\mathrm{pen}_1(\hat{m}).$$

The ball $\mathscr{B}_{m,\hat{m}}$ is random, it is not possible to bound $\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}}\nu_n^2(t)\right)$ as previously. We introduce the function $12p_1(m,m') = \mathrm{pen}_1(m) + \mathrm{pen}_1(m')$. Then

$$\mathbb{E}\left(\|\hat{g}_{\hat{m}} - \Pi_m g\|_n^2 \mathbb{1}_{\Omega_n}\right) \leq C\Delta + 12\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_{1,n}^2(t) - p_1(\hat{m},m)\right) + 4\,\mathrm{pen}_1(m).$$

We can write:

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_{1,n}^2(t) - p(\hat{m},m)\right) \leq \sum_{m'} \mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}} \nu_{1,n}^2(t) - p_1(m',m)\right)_+ \tag{25}$$

To bound this term, we apply the Talagrand's inequality for β-mixing random variables (see Appendix A) to $\nu_{1,n}(t)$. The process $(X_t)_{t\geq 0}$ is exponentially β-mixing: there exists c, c' such that the β-mixing coefficient of X_t satisfies $\beta_X(t) \leq ce^{-c't}$. Therefore the random vectors $(F_{k\Delta}+G_{k\Delta}, X_{k\Delta})$ are exponentially β-mixing with β-mixing coefficient smaller than $\beta_X(\Delta) \leq ce^{-c'\Delta}$. Let us set $q_n = \lceil 8\ln(n)/(c'\Delta)\rceil$ and $p_n = \lceil n/(2q_n)\rceil$. We consider the set Ω_B on which the random variables $B_{k\Delta}$, $Z_{k\Delta}$ and $J_{k\Delta}$ are bounded:

$$\Omega_B = \left\{\omega, \bigcap_{k=1}^{n}\left\{|B_{k\Delta}| \leq \Delta^{1-\epsilon}, \quad |Z_{k\Delta}| \leq 7\sigma_1\Delta^{1/2}\ln(n), \quad |J_{k\Delta}| \leq \frac{49C_J\ln(n)}{\lambda}\right\}_{1\leq k\leq n},\right.$$
$$\left.\bigcap_{p\in\{1,2,4\}}\bigcap_{j=0}^{p_n-1}\left\{\sum_{k=jq_n+1}^{(j+1)q_n} J_{k\Delta}^{2p} \leq 64C_p\xi_1^{2p}\Delta\ln^{2p}(n)\right\}_{0\leq j\leq p_n-1}\right\}. \tag{26}$$

By Lemma 2, $\mathbb{P}(\Omega_B^c) \lesssim n^{-6}$. The random variables $F_{k\Delta}$ and $G_{k\Delta}$ are bounded on Ω_B:

$$\forall k, \quad F_{k\Delta} \leq C\ln^2(n) \quad \text{and} \quad \frac{1}{q_n}\sum_{i=1}^{q_n} G_{kq_n+i\Delta} \leq C\ln^2(n)$$

for C a constant large enough. We consider the function $f_t(x,y) = xt(y)$ and set

$$U_j = \frac{1}{q_n}\sum_{k=jq_n}^{(j+1)q_n-1} f_t(F_{k\Delta}+G_{k\Delta}, X_{k\Delta}) = \frac{1}{q_n}\sum_{k=jq_n}^{(j+1)q_n-1}(F_{k\Delta}+G_{k\Delta})\,t(X_{k\Delta}).$$

By Lemma 3, $\mathbb{E}(U_j) = 0$. For any $t \in \mathscr{B}_{m,m'}$, by stationarity,

$$\mathrm{Var}(U_j) = \frac{1}{q_n^2}\sum_{k=1}^{q_n}\mathbb{E}\left(t^2(X_{k\Delta})\mathbb{E}\left((F_{k\Delta}+G_{k\Delta})^2\middle|\mathscr{F}_{k\Delta}\right)\right) \leq \frac{1}{q_n}\mathbb{E}\left(t^2(X_0)\frac{\xi_1^4(X_0)I_4 + C\Delta^{1/2}}{\Delta}\right).$$

Therefore

$$\sup_{t\in\mathscr{B}_{m,m'}}\mathrm{Var}(U_j) \leq \frac{C}{q_n\Delta} := \frac{V}{q_n}.$$

Let us set $D = \max(D_m, D'_m)$. As, if $t \in \mathscr{B}_{m,m'}$, $\|t\|_\infty \lesssim D^{1/2}$, on Ω_B, there exists a constant c such that, for any j, $|U_j| \leq c\ln^4(n)D^{1/2}$. We set $\mathscr{O} = \Omega_n \cap \Omega_B$, $M := c\ln^4(n)D^{1/2}$ and $H^2 := \Sigma_1 D/\Delta$.

By (23), $\mathbb{E}\left(\left|\frac{1}{p_n}\sum U_j\right|\right) \leq \sqrt{\mathbb{E}\left(\nu_{1,n}^2(t)\right)} \leq H/\sqrt{n}$. As the dimension of $S_{m+m'}$ is finite, we can find a countable family $\mathscr{F}$ dense in $\mathscr{B}_{m,m'}$ and we can apply the Talagrand's inequality for β-mixing random variables (see Appendix A). Therefore

$$\begin{aligned}
\mathbb{E}\left(\left[\sup_{t\in\mathscr{B}_{m,m'}} \nu_{n,1}^2(t) - 6H^2\right]_+ \mathbb{1}_{\mathscr{O}}\right) &\lesssim \frac{V}{n}\exp\left(-c'\frac{H^2}{V}\right) + \frac{M^2}{p_n^2}\exp\left(-c'\sqrt{\frac{p_n}{q_n}}\frac{H}{M}\right) + \frac{M}{n^2} \\
&\lesssim \frac{1}{n\Delta}e^{-c'D} + \frac{\ln^8(n)D}{p_n^2}\exp\left(-c'\sqrt{\frac{p_n}{q_n}}\frac{1}{\Delta^{1/2}\ln^4(n)}\right) + \frac{\ln^4(n)D^{1/2}}{n^2} \\
&\lesssim \frac{1}{n\Delta}e^{-c'D} + \frac{\ln^{10}(n)}{n^2\Delta^2}D\exp\left(-c'\frac{\sqrt{n\Delta}}{\ln^5(n)}\right)
\end{aligned}$$

where c' is a constant that varies from one line to another. By (25), as $\sum_{m'\in\mathscr{M}_n} D_{m,m'} \leq \mathscr{D}_n^2 \leq n\Delta$,

$$\mathbb{E}\left(\left[\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_{n,1}^2(t) - p(m,\hat{m})\right]_+\right) \lesssim \frac{1}{n\Delta}\sum_{m'} e^{-cD_{m,m'}} + \frac{\ln^6(n)}{n\Delta} \lesssim \frac{1}{n\Delta}.$$

We can bound $\mathbb{E}\left(\|\hat{g}_{\hat{m}} - \Pi_m g\|_n^2 \mathbb{1}_{\mathscr{O}^c}\right)$ in the same way as we bound the risk of the non-adaptive estimator on Ω_n^c:

$$\mathbb{E}\left(\|\hat{g}_{\hat{m}} - \Pi_m g\|_n^2 \mathbb{1}_{\mathscr{O}^c}\right) \lesssim \frac{1}{\Delta^{3/2}n^{5/2}} \lesssim \frac{1}{n}.$$

Then, for any $m \in \mathscr{M}_n$,

$$\mathbb{E}\left(\|\hat{g}_{\hat{m}} - g_A\|_n^2\right) \leq \|g_{m,\pi} - g_A\|_\pi^2 + \kappa\Sigma_1\frac{D_m}{n\Delta} + C\Delta.$$

It remains to replace the $\|.\|_n$-norm by the $\|.\|_{L^2}$-norm. The proof is the same as in Subsection 6.5.

6.7 Proof of Proposition 7 (risk of the estimator of h for m fixed)

Let us first precise $H_{k\Delta}$ and $L_{k\Delta}$. We can write $\Delta H_{k\Delta} = (Z_{k\Delta} + J_{k\Delta})^4 - \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^4\right)$ and $L_{k\Delta} = L_{k\Delta}^{(1)} + L_{k\Delta}^{(2)}$ with

$$\begin{aligned}
\Delta L_{k\Delta}^{(1)} &:= (B_{k\Delta} + Z_{k\Delta} + J_{k\Delta})^4 - (Z_{k\Delta} + J_{k\Delta})^4 \\
\Delta L_{k\Delta}^{(2)} &:= \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^4\right) - \Delta\xi^4(X_{k\Delta})I_4.
\end{aligned}$$

Let us first prove that

$$\mathbb{E}\left(L_{k\Delta}^2\right) \lesssim \Delta, \quad \mathbb{E}\left(L_{k\Delta}^4\right) \lesssim \Delta \tag{27}$$

$$\mathbb{E}\left(H_{k\Delta}|\mathscr{F}_{k\Delta}\right) = 0, \quad \mathbb{E}\left(H_{k\Delta}^2|\mathscr{F}_{k\Delta}\right) \leq C\Delta^{-1} \quad \text{and} \quad \mathbb{E}\left(H_{k\Delta}^4\right) \lesssim 1/\Delta^3. \tag{28}$$

To bound $L_{k\Delta}^{(1)}$, we use the same techniques as for the bound of $E_{k\Delta}^{(1)}$: we develop $L_{k\Delta}^{(1)}$, use the decomposition $B_{k\Delta} = \Delta b(X_{k\Delta}) + \tilde{B}_{k\Delta}$, and thanks to Cauchy-Schwarz inequality, Lemma 1 and equation (10), we obtain that $\mathbb{E}\left(\left(L_{k\Delta}^{(1)}\right)^p\right) \lesssim \Delta$. By Lemma 1, we get that $\mathbb{E}\left(Z_{k\Delta}^4\right) \leq C\Delta^2$ and

$\mathbb{E}\left(J_{k\Delta}^4\right) = \xi^4(X_{k\Delta})I_4\Delta + C\Delta^{3/2}$. Therefore $|L_{k\Delta}^{(2)}| \lesssim \Delta^{1/2}$. By definition, $\mathbb{E}\left(H_{k\Delta}\right) = 0$ and we derives from Lemma 1 and Cauchy-Schwarz inequality that $\mathbb{E}\left(H_{k\Delta}^2\right) \lesssim \Delta^{-1}$ and $\mathbb{E}\left(H_{k\Delta}^4\right) \lesssim \Delta^{-3}$.

The rest of the proof is done in the same way as in Subsection 6.4. We denote by $\Pi_m h$ the orthogonal projection of h on S_m for the $\|.\|_n$- norm and we get:

$$\left\|\hat{h}_m - \Pi_m h\right\|_n^2 \mathbb{1}_{\Omega_n} \leq \frac{6}{n}\sum_{k=1}^{n} L_{k\Delta}^2 + 6\sum_{\lambda=1}^{D_m} \nu_{3,n}^2(\varphi_\lambda) \quad \text{with} \quad \nu_{3,n}^2(t) := \frac{1}{n}\sum_{k=1}^{n} H_{k\Delta}t(X_{k\Delta}) \tag{29}$$

By (28), we obtain that $\mathbb{E}\left(\nu_{3,n}^2(\varphi_\lambda)\right) \lesssim \frac{1}{n\Delta}$ and therefore

$$\mathbb{E}\left(\left\|\hat{h}_m - \Pi_m h\right\|_n^2 \mathbb{1}_{\Omega_n}\right) \leq \|h - h_{m,\pi}\|_\pi^2 + C\frac{D_m}{n\Delta} + C'\Delta.$$

The bound on Ω_n^c is done in the same way as in Subsection 6.4.

6.8 Proof of Lemma 8 (Estimation of the penalty constant)

Estimation of the first bound of the variance term. Let us set $h_{1,A} = \sup_{x\in A} h(x)$. We have that

$$|\hat{h}_1 - h_{1,A}| \leq \left\|\hat{h}_m - h_A\right\|_{L^\infty(A)} \leq \left\|\hat{h}_m - h_m\right\|_{L^\infty(A)} + \|h_m - h_A\|_{L^\infty(A)}.$$

By DeVore and Lorentz (1993, p182) and Barron et al. (1999, Lemma 12), if h belongs to the Besov space $\mathscr{B}_{2,\infty}^\alpha(A)$, then

$$\|h_m - h_A\|_{L^\infty(A)}^2 \leq CD_m^{1-2\alpha}. \tag{30}$$

As g is Lipschitz on A, $g \in \mathscr{B}_{2,\infty}^1(A)$ and $\|h_m - h_A\|_{L^\infty(A)}^2 \lesssim D_m^{-1}$. As $\hat{h}_m$ and h_m belongs to S_m,

$$\left\|\hat{h}_m - h_m\right\|_{L^\infty(A)}^2 \leq \phi_1 D_m \left\|\hat{h}_m - h_m\right\|_{L^2}^2 \leq \frac{\phi_1 D_m}{\pi_0}\left\|\hat{h}_m - h_m\right\|_\pi^2.$$

On Ω_n,

$$\left\|\hat{h}_m - h_m\right\|_\pi^2 \leq 2\left\|\hat{h}_m - h_m\right\|_n^2 \leq 2\left\|\hat{h}_m - \Pi_m h\right\|_n^2 + 2\|\Pi_m h - h_m\|_n^2$$

and $\|\Pi_m h - h_m\|_n^2 \leq \|h_A - h_m\|_n^2$. Then by (29),

$$|\hat{h}_1 - h_{1,A}|\mathbb{1}_{\Omega_n} \lesssim D_m^{-1} + D_m\|h_m - h_A\|_n^2 + D_m\left(\frac{1}{n}\sum_{k=1}^{n} L_{k\Delta}^2 + \sum_{\lambda} \nu_{3,n}^2(\varphi_\lambda)\right). \tag{31}$$

We can write $\|h_m - h_A\|_n^2 = \|h_m - h_A\|_\pi^2 + \|h_m - h_A\|_n^2 - \|h_m - h_A\|_\pi^2$. As h belongs to the Besov space $\mathscr{B}_{2,\infty}^\alpha$, $\|h_m - h_A\|_\pi^2 \leq D_m^{-2\alpha}$. It remains to bound $\|h_m - h_A\|_n^2 - \|h_m - h_A\|_\pi^2$. As this term is centred, we can apply a Bennet's inequality. We can remark that, when $h \in \mathscr{B}_{2,\infty}^\alpha$, by (30), $\|h_m - h_A\|_\infty^2 \leq CD_m^{1-2\alpha} =: M'$ and

$$\mathrm{Var}\left((h_m - h_A)^2(X_{k\Delta})\right) = \int (h_m - h_A)^4(x)\pi(x)dx \leq \|h_m - h_A\|_\pi^2\|h_m - h_A\|_\infty^2 \leq C'D_m^{1-4\alpha} =: V'.$$

Then

$$\mathbb{P}\left(\left|\|h_m - h_A\|_n^2 - \|h_m - h_A\|_\pi^2\right| \geq cD_m^{-2\alpha}\right) \leq \exp\left(-\frac{c'^2 n D_m^{-4\alpha}}{2(D_m^{1-4\alpha} + D_m^{1-4\alpha}\ln(n)\Delta^{-1}}\right)$$
$$\leq \exp\left(-c'\frac{n\Delta}{\ln(n)}D_m^{-1}\right) \lesssim n^{-8}.$$

On Ω_B (see (26)), by Lemma 2,

$$\frac{\Delta^2}{n}\sum_{k=1}^n \left(L_{k\Delta}^{(1)}\right)^2 \lesssim \frac{1}{n}\sum_{k=1}^n B_{k\Delta}^2 (J_{k\Delta}+Z_{k\Delta})^6 \lesssim \sqrt{\left(\frac{1}{n}\sum_{k=1}^n B_{k\Delta}^4\right) \times \left(\frac{1}{n}\sum_{k=1}^n (Z_{k\Delta}+J_{k\Delta})^{12}\right)} \lesssim \Delta^3 \ln^8(n)$$

and as $|L_{k\Delta}^{(2)}| \leq \Delta^{1/2}$, $\frac{1}{n}\sum_{k=1}^n L_{k\Delta}^2 \lesssim \Delta \ln^8(n)$. To bound $\nu_{3,n}^2(\varphi_\lambda)$, we apply (again) Bennett's inequality. On Ω_B,

$$\frac{1}{q_n}\sum_{k=1}^{q_n} |H_{k\Delta}| \lesssim \frac{\Delta^{-1}}{q_n}\sum_{k=1}^{q_n} Z_{k\Delta}^4 + J_{k\Delta}^4 \lesssim \ln^4(n).$$

We take $V' = C\Delta^{-1}$, $M' = C'\ln^4(n)D_m^{1/2}$. Then, as $\mathbb{E}\left(\nu_{n,3}(\varphi_\lambda)\right) = 0$,

$$\mathbb{P}\left(|\nu_{n,3}(\varphi_\lambda)| \geq \frac{\ln(n)}{\sqrt{n\Delta}}\right) \leq \exp\left(-\frac{-c\ln^2(n)\Delta^{-1}}{\Delta^{-1} + \ln^6(n)D_m^{1/2}\Delta^{-1}/\sqrt{n\Delta}}\right) \leq \exp\left(-\frac{c\ln^2(n)}{1+\frac{\ln^6(n)D_m^{1/2}}{\sqrt{n\Delta}}}\right)$$

and, if $D_m^2 \leq (n\Delta)/\ln^{12}(n)$,

$$\mathbb{P}\left(|\nu_{n,3}(\varphi_\lambda)| \geq \frac{\ln(n)}{\sqrt{n\Delta}}\right) \lesssim n^{-6}.$$

Collecting terms, we get that, as $\alpha \geq 1$,

$$\mathbb{P}\left(|\hat{h}_1 - h_{1,A}| \gtrsim D_m^{-1} + D_m\left(\frac{\ln(n)}{\sqrt{n\Delta}} + \Delta\ln^6(n)\right)\right) \lesssim n^{-5}.$$

By taking $D_m = \ln(n)$, we obtain the expected result.

Estimation of the second bound of the variance term. We first focus on the term $\bar{V}_n$. As $V_{k\Delta} = I_4\xi^4(X_{k\Delta}) + L_{k\Delta} + H_{k\Delta}$, by (28) and (27), $\mathbb{E}\left(V_{k\Delta}\right) = I_4\int \xi^4(x)\pi(x)dx + C\Delta^{1/2}$. Moreover, $\mathrm{Var}\left(V_{k\Delta}\right) \leq \mathbb{E}\left(V_{k\Delta}^2\right) \leq \frac{C}{\Delta} + C'\Delta^{1/2}$ and by Lemma 2, $\frac{1}{q_n}\sum_{k=1}^{q_n} V_{k\Delta} \lesssim \ln^4(n)$. Therefore, by Bennett's inequality,

$$\mathbb{P}\left(|\bar{V}_n - \mathbb{E}\left(V_{k\Delta}\right)| \geq \frac{\ln^7(n)}{\sqrt{n\Delta}}\right) \lesssim \exp\left(-\frac{c\ln^{14}(n)\Delta^{-1}}{\Delta^{-1} + \ln^{12}(n)\Delta^{-1}}\right) \lesssim \exp\left(-c\ln^2(n)\right) \lesssim n^{-5}$$

and

$$\mathbb{P}\left(|\bar{V}_n - \int \xi^4(x)\pi(x)dx| \geq C\Delta^{1/2} + \frac{\ln^7(n)}{\sqrt{n\Delta}}\right) \lesssim n^{-5}.$$

It remains to bound $\hat{\pi}_0 - \pi_0$. We have that $\mathbb{E}\left(\hat{\pi}_h(x)\right) = \pi_h(x)$ where $\pi_h(x) = \int K_h(y-x)\pi(y)dy$. Moreover, as the process (X_t) is exponentially β-mixing, by Berbee's coupling lemma, on Ω^*

$$\begin{aligned}\mathrm{Var}\left(\hat{\pi}_h(x)\right) &= \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n} K_h(X_{k\Delta} - x)\right) \leq \frac{1}{p_n}\mathrm{Var}\left(\frac{1}{q_n}\sum_{k=1}^{q_n} K_h(X_{k\Delta} - x)\right) \\ &\leq \frac{1}{p_n}\mathrm{Var}\left(K_h(X_{k\Delta} - x)\right) \leq \frac{\ln(n)}{n\Delta}\mathbb{E}\left(K_h^2(X_{k\Delta} - x)\right).\end{aligned}$$

By a change of variable, we get that

$$\mathrm{Var}\left(\hat{\pi}_h(x)\right) \leq \frac{\ln(n)}{n\Delta}\int \frac{1}{h^2}K^2\left(\frac{y-x}{h}\right)\pi(y)dy = \frac{\ln(n)}{n\Delta h}\int K^2(u)\pi(x+hu)du \lesssim \frac{\ln(n)}{n\Delta h}.$$

Then

$$\mathbb{E}\left((\hat{\pi}_h(x) - \pi(x))^2\right) \leq \left(\pi_h(x) - \pi(x)\right)^2 + \frac{C\ln(n)}{n\Delta h}.$$

If π is continuous, then the bias term $(\pi_h(x) - \pi(x))^2$ satisfies (see (Tsybakov, 2004, Proposition 1.2 with $\beta = 1$)):

$$\left(\pi_h(x) - \pi(x)\right)^2 \leq C'h. \tag{32}$$

In that case, the best estimator is obtained for $h = 1/\sqrt{n\Delta}$. By Bennett's inequality for β-mixing random variables (see Appendix A), as $|K_h| \leq 1/h = \sqrt{n\Delta}$ and $\mathrm{Var}\left(\sum_{k=1}^{q_n} K_h(X_{k\Delta} - x)\right) \leq Cq_n^2/h = q_n\frac{C'\ln(n)}{\Delta h} = q_n\frac{\ln(n)\sqrt{n}}{\sqrt{\Delta}}$,

$$\begin{aligned}\mathbb{P}\left(|\pi_h(y) - \hat{\pi}_h(y)| \geq \frac{\ln^2(n)}{(n\Delta)^{1/4}}\right) &\leq 2\exp\left(-\frac{c\ln^4(n)n^{1/2}\Delta^{-1/2}}{\ln(n)\Delta^{-1/2}n^{1/2} + \ln^3(n)n^{1/2}\Delta^{-1/2}(n\Delta)^{-1/4}}\right) \\ &\leq 2\exp\left(-\frac{c\ln^4(n)}{\ln(n) + \frac{\ln^3(n)}{(n\Delta)^{1/4}}}\right) \lesssim n^{-6}\end{aligned}$$

As π is continuous, by (32),

$$\begin{aligned}|\hat{\pi}_0 - \pi_0| &\leq \min_{1\leq j\leq \ln^2(n)} |\pi(x_j) - \pi_0| + \max_{1\leq j\leq \ln^2(n)} |\hat{\pi}_{(n\Delta)^{-1/2}}(x_j) - \pi(x_j)| \\ &\lesssim \frac{1}{\ln^2(n)} + \frac{\ln^2(n)}{(n\Delta)^{1/2}} + \max_{1\leq j\leq \ln^2(n)} |\hat{\pi}_{(n\Delta)^{-1/2}}(x_j) - \pi_{(n\Delta)^{-1/2}}(x_j)|\end{aligned}$$

and therefore

$$\mathbb{P}\left(|\hat{\pi}_0 - \pi_0| \geq \frac{1}{\ln(n)} + \frac{\ln^2(n)}{(n\Delta)^{1/4}}\right) \lesssim n^{-5} \tag{33}$$

which concludes the proof.

6.9 Proof of Lemma 10 (Approximation of the continuous terms)

Let us denote by $N_k = \mu\left(]k\Delta,(k+1)\Delta], \left[-\Delta^{1/4},\Delta^{1/4}\right]^c\right)$ the number of jumps of amplitude greater than $\Delta^{1/4}$ in the time interval $]k\Delta,(k+1)\Delta]$. We introduce the set

$$\Omega_{N,k} = \left\{\omega,\ N_k = 0 \quad \text{and} \quad \left|L_{k\Delta}^{(1)+(2)}\right| \leq 4\frac{\sigma_1 + \xi_1}{\xi_0}\Delta^{1/2}\ln(n)\right\}$$

where $L_s^{(i)}$ are defined in (9). The jumps are explicitely bounded on $\Omega_{N,k}$, which is close to $\Omega_{X,k}$. The following lemma is proved later.

Lemma 17.

$$\mathbb{P}\left(\Omega_{X,k}^c\right) \lesssim \Delta^{1-\beta/2}, \quad \mathbb{P}\left(\Omega_{N,k}^c\right) \lesssim \Delta^{1-\beta/2} \quad \textit{and} \quad \mathbb{P}\left(\Omega_{X,k} \cap \Omega_{N,k}^c\right) \lesssim \Delta^{2-\beta/2} + n^{-1}.$$

We can write

$$(U_{k\Delta} - Y_{k\Delta}) = \frac{1}{\Delta}\left(B_{k\Delta} + Z_{k\Delta}\right)^2 \mathbb{1}_{\Omega_{X,k}^c} - \frac{1}{\Delta} J_{k\Delta}^2 \mathbb{1}_{\Omega_{X,k}} - \frac{2}{\Delta} J_{k\Delta}\left(Z_{k\Delta} + B_{k\Delta}\right)\mathbb{1}_{\Omega_{X,k}}.$$

On Ω_B, $|Z_{k\Delta} + B_{k\Delta}| \lesssim \Delta^{1/2}\ln(n)$. Moreover, on $\Omega_{X,k} \cap \Omega_B$, $|J_{k\Delta}| \leq |C\Delta^{1/2}\ln(n) - (B_{k\Delta} + Z_{k\Delta})| \lesssim \Delta^{1/2}\ln(n)$. We obtain the following bound:

$$|U_{k\Delta} - Y_{k\Delta}|\mathbb{1}_{\Omega_B} \lesssim \ln^2(n)\mathbb{1}_{\Omega_{X,k}^c} + \ln^2(n)\mathbb{1}_{\Omega_{X,k}\cap\Omega_{N,k}^c} + \left(\frac{J_{k\Delta}^2}{\Delta} + \ln(n)\frac{|J_{k\Delta}|}{\Delta^{1/2}}\right)\mathbb{1}_{\Omega_{N,k}}. \tag{34}$$

Then by Lemma 17 and Lemma 1,

$$\mathbb{E}\left((U_{k\Delta} - Y_{k\Delta})^{2p}\right) \lesssim \ln^{4p}(n)\Delta^{1-\beta/2} + \frac{1}{n^6} + \mathbb{E}\left(\left(J_{k\Delta}^{4p}\Delta^{-2p} + J_{k\Delta}^{2p}\Delta^{-p}\ln^{2p}(n)\right)\mathbb{1}_{\Omega_{N,k}}\right).$$

It remains to bound $\mathbb{E}\left(J_{k\Delta}^{4p}\Delta^{-2p}\mathbb{1}_{\Omega_{N,k}}\right)$. By equation (8),

$$\mathbb{E}\left(J_{k\Delta}^{2p}\mathbb{1}_{\Omega_{N,k}}\right) \lesssim \mathbb{E}\left(L_{k\Delta}^{2p}\mathbb{1}_{\Omega_{N,k}}\right) + \mathbb{E}\left(\tilde{J}_{k\Delta}^{2p}\mathbb{1}_{\Omega_{N,k}}\right) \lesssim \mathbb{E}\left(L_{k\Delta}^{2p}\right) + \Delta^2 I_{2p}^2 + \Delta^{p+1} I_{2p} + \Delta^{2p}.$$

As we are in $\Omega_{N,k}$, we can replace I_{2p} by $I_{2p,N} := \int_{|z|\leq\Delta^{1/4}} z^{2p}\nu(dz)$. For any $p \geq 1$, $I_{2p,N} \leq \Delta^{2p/4-\beta/4}\int_{|z|\leq\Delta^{1/4}} z^\beta \nu(dz) \lesssim \Delta^{p/2-\beta/4}$. Then

$$\mathbb{E}\left(\tilde{J}_{k\Delta}^{2p}\mathbb{1}_{\Omega_{N,k}}\right) \lesssim \Delta^{2+p-\beta/2}. \tag{35}$$

It remains to bound $\mathbb{E}\left(L_{k\Delta}^{2p}\right)$. We use the following lemma which is nearly Proposition 4.8 of Mai (2014) (the proof is given later):

Lemma 18. *For any integer $p \geq 1$,*

$$\mathbb{E}\left(\left(L_{k\Delta}^{(1)+(2)}\mathbb{1}_{\Omega_{N,k}}\right)^{2p}\right) \lesssim \Delta^{1+p-\beta/2}\ln(n)^{2p-\beta}$$

Then $\mathbb{E}\left((J_{k\Delta}^{(1)+(2)})^{2p}\mathbb{1}_{\Omega_{N,k}}\right) \lesssim \Delta^{1+p-\beta/2}\ln(n)^{2p-\beta}$ and we deduce that $\mathbb{E}\left((U_{k\Delta} - Y_{k\Delta})^{2p}\right) \lesssim \Delta^{1-\beta/2}\ln^{4p}(n)$.

Let us now bound $|U_{k\Delta} - Y_{k\Delta}|$. We have that

$$|U_{k\Delta} - Y_{k\Delta}| \leq \frac{2}{\Delta}\left(B_{k\Delta} + Z_{k\Delta}\right)^2 + \frac{2}{\Delta}J_{k\Delta}^2\mathbb{1}_{\Omega_{X,k}}.$$

And

$$\mathbb{P}\left(\left\{|J_{k\Delta}| \geq (4p+2)\sigma_0\Delta^{1/2}\ln(n)\right\} \cap \Omega_{X,k}\right) \leq \mathbb{P}\left(|B_{k\Delta} + Z_{k\Delta}| \geq (4p+1)\sigma_0\Delta^{1/2}\ln(n)\right).$$

Then, by Lemma 2, $\mathbb{P}\left(|U_{k\Delta} - Y_{k\Delta}| \geq C\sigma_1^2\ln^2(n)\right) \lesssim n^{-5}$ where C is a universal constant.

6.10 Proof of Lemma 17

Bound of $\mathbb{P}\left(\Omega_{X,k}^c\right)$. We have that $X_{(k+1)\Delta} = X_{k\Delta} + B_{k\Delta} + Z_{k\Delta} + J_{k\Delta}$. Then

$$\mathbb{P}\left(\Omega_{X,k}^c\right) \leq \mathbb{P}\left(|B_{k\Delta}| \geq \Delta^{1/2}\right) + \mathbb{P}\left(|Z_{k\Delta}| \geq 6\sigma_1\Delta^{1/2}\ln(n)\right) + \mathbb{P}\left(|J_{k\Delta}| \geq 12\xi_1\Delta^{1/2}\ln(n)\right).$$

By Lemma 2,

$$\mathbb{P}\left(|B_{k\Delta}| \geq \Delta^{1/2}\right) \lesssim n^{-6} \quad \text{and} \quad \mathbb{P}\left(|Z_{k\Delta}| \geq 6\sigma_1\Delta^{1/2}\ln(n)\right) \lesssim n^{-6}.$$

By (11), $\mathbb{P}\left(|J_{k\Delta}^{(1)}| \geq 12\xi_1\Delta^{1/2}\ln(n)\right) \lesssim n^{-6}$. Moreover, by Assumption 4

$$\begin{aligned}\mathbb{P}\left(\left|J_{k\Delta}^{(2+3+4)}\right| > 0\right) &\leq \Delta\int_{[-\Delta^{1/2},\Delta^{1/2}]^c}\nu(dz) \leq \Delta\Delta^{-\beta/2}\int_{[-\Delta^{1/2},\Delta^{1/2}]^c}|z|^\beta\nu(dz) \\ &\lesssim \Delta^{1-\beta/2} \end{aligned} \tag{36}$$

which concludes the proof.

Bound of $\mathbb{P}\left(\Omega_{N,k}^c\right)$. We have that

$$\mathbb{P}\left(N_k \geq 1\right) = \int_{k\Delta}^{(k+1)\Delta}\int_{|z|\geq\Delta^{1/4}}\nu(dz)dt \leq \Delta^{1-\beta/4}\int_{|z|\geq\Delta^{1/4}}|z|^\beta\nu(dz) \lesssim \Delta^{1-\beta/4}.$$

Then by (36) and (11), we obtain:

$$\mathbb{P}\left(\Omega_{N,k}^c\right) \leq \mathbb{P}\left(N_k \geq 1\right) + \mathbb{P}\left(|L_{k\Delta}^{(2)}| > 0\right) + \mathbb{P}\left(\left|L_{k\Delta}^{(1)}\right| \geq c\ln(n)\Delta^{1/2}\right) \lesssim \Delta^{1-\beta/2}.$$

Bound of $\mathbb{P}\left(\Omega_{X,k} \cap \Omega_{N,k}^c\right)$. We have that

$$\mathbb{P}\left(\Omega_{X,k} \cap \{N_k \geq 1\}\right) \leq \mathbb{P}\left(N_k \geq 2\right) + \mathbb{P}\left(\Omega_{X,k} \cap \{N_k = 1\}\right).$$

Now

$$\mathbb{P}\left(N_k \geq 2\right) = \left(\Delta\int_{|z|\geq\Delta^{1/4}}\nu(dz)\right)^2 \leq \left(\Delta^{1-\beta/4}\int_{|z|\geq\Delta^{1/4}}|z|^\beta\nu(dz)\right)^2 \lesssim \Delta^{2-\beta/2}.$$

If $N_k = 1$, there exists a unique jump greater than $\Delta^{1/4}$ and therefore $\left|J_{k\Delta}^{(3+4)}\right| \geq \xi_0\Delta^{1/4}$. By conditional independence, we get:

$$\begin{aligned}\mathbb{P}\left(\Omega_{X,k} \cap \{N_k = 1\}\right) &\leq \mathbb{P}\left(N_k = 1\right) \times \mathbb{P}\left(\left|B_{k\Delta} + Z_{k\Delta} + J_{k\Delta}^{(1+2)}\right| > c\xi_0\Delta^{1/4}\right) \\ &\leq \mathbb{P}\left(N_k = 1\right)\left[\mathbb{P}\left(\Omega_B^c\right) + \mathbb{P}\left(\left|J_{k\Delta}^{(1+2)}\right| \geq \frac{c\xi_0\Delta^{1/4}}{3}\right)\right] \\ &\lesssim n^{-6} + \Delta^{1-\beta/4}\mathbb{P}\left(\left|J_{k\Delta}^{(1+2)}\right| \geq \frac{c\xi_0\Delta^{1/4}}{3}\right)\end{aligned}$$

By a Markov inequality, we obtain:

$$\mathbb{P}\left(\left|J_{k\Delta}^{(1+2)}\right| > c\Delta^{1/4}\right) \lesssim \Delta^{-1/2}\mathbb{E}\left[\left(J_{k\Delta}^{(1+2)}\right)^2\right] \lesssim \Delta^{-1/2}\Delta\int_{-\Delta^{1/4}}^{\Delta^{1/4}} z^2\nu(dz) \lesssim \Delta^{1/2}\Delta^{1/2-\beta/4}.$$

We obtain that $\mathbb{P}\left(\Omega_{X,k}\cap\{N_k\geq 1\}\right)\lesssim\Delta^{2-\beta/2}$. It remains to bound

$$\mathbb{P}\left(\Omega_{X,k}\cap\{N_k=0\}\cap\left\{|L_{k\Delta}^{[1)+(2)}|\geq c\Delta^{1/2}\ln(n)\right\}\right).$$

We can remark that $\xi(X_{k\Delta})L_{k\Delta}^{(1)+(2)} = J_{k\Delta}^{(1)+(2)} - \tilde{J}_{k\Delta}^{(1)+(2)}$, and by Lemma 2,

$$\mathbb{P}\left(\Omega_{X,k}\cap\left\{\left|J_{k\Delta}^{(1)+(2)}\right|\geq(c+1)\Delta^{1/2}\ln(n)\right\}\cap\{N_k=0\}\right)\lesssim\mathbb{P}\left(|B_{k\Delta}+Z_{k\Delta}|\geq c\Delta^{1/2}\ln(n)\right)\lesssim n^{-6}.$$

By (35) and a Markov inequality, we obtain that $\mathbb{P}\left(\left|\tilde{J}_{k\Delta}^{(1)+(2)}\right|\geq\Delta^{1/2}\ln(n)\middle|N_k=0\right)\lesssim\Delta^{2-\beta/2}$ which ends the proof.

6.11 Proof of Lemma 18

Let us introduce a nonnegative function $f\in\mathscr{C}^\infty$ such that

$$\begin{cases} f(x)=x^{2p} & \text{if } |x|\leq 1 \\ f(x)=0 & \text{if } |x|\geq 2. \end{cases} \quad \text{and} \quad \forall x,\ f(x)\leq 1.$$

Let us set $f^a(x)=a^{2p}f(x/a)$ for $a>0$. Then $\mathbb{E}\left[\left(L_\Delta^{(1)+(2)}\right)^{2p}\mathbb{1}_{\Omega_{N,k}}\right]\leq\mathbb{E}\left(f^{c\Delta^{1/2}\ln(n)}(L_\Delta^{(1)+(2)})\right)$. Now $L_\Delta^{(1)+(2)}$ is a pure jump Lévy process, and its characteristic function is known. Following the same proof than Mai (2014, proof of Proposition 4.8), we get that, thanks to the properties of the Fourier transform denoted by $\mathscr{F}$ (see the Appendix),

$$\mathbb{E}\left(f^a(L_t^{(1)+(2)})\right)\leq t\int_{-1}^{1}f^a(x)\nu(dx)+Ct^2\int_{\mathbb{R}}\mathscr{F}f^a(u)|u|^{2\beta}du.$$

As f belongs to the Schwarz space (see Appendix A), so does $\mathscr{F}f$ and therefore , for any $m>0$, there exists a constant C_m such that $|\mathscr{F}f(u))|\leq C_m|u|^{-m}$. Then, as $\mathscr{F}^a(u)=a^{2p+1}\mathscr{F}(au)$, we get that

$$|\mathscr{F}f^a(u)|\leq C_0a^{2p+1}\wedge C_{3+\beta}a^{2p-2-\beta}|u|^{-3-\beta}.$$

Then

$$\mathbb{E}\left(f^a(L_t^{(1)+(2)})\right)\lesssim t\int_{-2a}^{2a}x^{2p}\nu(dx)+t^2a^{2p+1}\int_{[-1,1]}|u|^{2\beta}du+t^2a^{2p-2-\beta}\int_{[-1,1]^c}|u|^{\beta-3}du.$$

As $\beta<2$, $\beta-3<-1$ and all the integrals are finite. Taking $t=\Delta$ and $a=c\Delta^{1/2}\ln(n)$, we get:

$$\mathbb{E}\left(\left(L_{k\Delta}^{(1)+(2)}\mathbb{1}_{\Omega_{N,k}}\right)^{2p}\right)\lesssim\Delta^{1+p-\beta/2}\ln^{2p-\beta}(n).$$

6.12 Proof of Proposition 11 (Risk of the estimator of σ for fixed m)

We have that $Y_{k\Delta} = U_{k\Delta} + Y_{k\Delta} - U_{k\Delta}$. We can write:

$$U_{k\Delta} = \frac{(B_{k\Delta} + Z_{k\Delta})^2}{\Delta} = \sigma^2(X_{k\Delta}) + F_{k\Delta} + \left(\frac{\mathbb{E}\left(Z_{k\Delta}^2\right)}{\Delta} - \sigma^2(X_{k\Delta}) \right) + \frac{2B_{k\Delta}Z_{k\Delta} + B_{k\Delta}^2}{\Delta}$$

Let us set

$$\tilde{E}_{k\Delta} = Y_{k\Delta} - U_{k\Delta} + \left(\frac{\mathbb{E}\left(Z_{k\Delta}^2\right)}{\Delta} - \sigma^2(X_{k\Delta}) \right) + \frac{2B_{k\Delta}Z_{k\Delta} + B_{k\Delta}^2}{\Delta}.$$

By Lemmas 1 and 10, we have that

$$\mathbb{E}\left(\tilde{E}_{k\Delta}^2\right) \lesssim \Delta^{1-\beta/2} \ln^4(n) + \Delta \quad \text{and} \quad \mathbb{E}\left(\tilde{E}_{k\Delta}^4\right) \lesssim \Delta^{1-\beta/2} \ln^8(n).$$

Moreover, by Lemma 3,

$$\mathbb{E}\left(F_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right) = 0, \quad \mathbb{E}\left(F_{k\Delta}^2 \middle| \mathscr{F}_{k\Delta}\right) \leq \frac{\sigma_1^4}{\Delta} \quad \text{and} \quad \mathbb{E}\left(F_{k\Delta}^4 \middle| \mathscr{F}_{k\Delta}\right) \lesssim \frac{1}{\Delta}.$$

The end of the proof is the same as in Subsection 6.4. On Ω_n, we obtain that:

$$\mathbb{E}\left(\left\| \hat{\sigma}_m^2 - \sigma_A^2 \right\|_n^2 \mathbb{1}_{\Omega_n} \right) \leq \left\| \sigma_m^2 - \sigma_A^2 \right\|_\pi^2 + 12\mathbb{E}\left(\tilde{E}_\Delta^2\right) + 12\mathbb{E}\left(\sup_{t \in \mathscr{B}_m} \nu_{n,2}^2(t) \right)$$

where $\nu_{n,2}(t) = n^{-1} \sum_{k=1}^n F_{k\Delta} t(X_{k\Delta})$. By Lemma 3, we get that

$$\mathbb{E}\left(\sup_{t \in S_m} \nu_{n,2}^2(t) \right) \leq \sum_{\lambda \in \Lambda} \mathbb{E}\left(\nu_{n,2}^2(\varphi_\lambda) \right) = \sum_{\lambda} \mathbb{E}\left(F_{k\Delta}^2 \varphi_\lambda^2(X_{k\Delta}) \right) \leq \sigma_1^4 \frac{D_m}{n}$$

where $(\varphi_\lambda)_{\lambda \in \Lambda}$ is an orthonormal basis for the $\|.\|_\pi$- norm. If we take an orthonormal basis for the $\|.\|_{L^2}$-norm, we obtain the second bound

$$\mathbb{E}\left(\sup_{t \in S_m} \nu_{n,2}^2(t) \right) \leq \frac{\phi_1 \int \sigma^4(x)\pi(x)dx}{\pi_0} \frac{D_m}{n}.$$

The bound on the risk on Ω_n^c is done in the same as in Subsection 6.4.

6.13 Proof of Theorem 13 (Oracle inequality for the adaptive estimator of σ)

The beginning of the proof is the same as in Subsection 6.6. As, for any m, $\gamma_{n,2}(\hat{\sigma}_{\hat{m}}^2) + \mathrm{pen}_2(\hat{m}) \leq \gamma_{n,2}(\sigma_m^2) + \mathrm{pen}_2(m)$, we get that

$$\left\| \hat{\sigma}_{\hat{m}}^2 - \sigma_A^2 \right\|_n^2 \leq \left\| \sigma_m^2 - \sigma_A^2 \right\|_n^2 + \mathrm{pen}_2(m) - \mathrm{pen}_2(\hat{m}) + \frac{2}{n} \sum_{k=1}^n (\tilde{E}_{k\Delta} + F_{k\Delta})(\hat{\sigma}_{\hat{m}}^2(X_{k\Delta}) - \sigma_m^2(X_{k\Delta})).$$

As previously, we bound the risk on Ω_n and on Ω_n^c. We set $p_2(m,m') = (\mathrm{pen}_2(m) + \mathrm{pen}_2(m'))/12$. On Ω_n, we have:

$$\mathbb{E}\left(\left\|\hat{\sigma}^2_{\hat{m}} - \sigma^2_A\right\|_n^2\right) \leq \left\|\sigma^2_m - \sigma^2_A\right\|_\pi^2 + 4\,\mathrm{pen}_2(m) + C\Delta^{1-\beta/2} + 12\sum_{m'}\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}} \nu^2_{n,2}(t) - p_2(m,m')\right)_+.$$

As in Subsection 6.6, it remains to bound $\mathbb{E}\left(\left[\sup_{t\in\mathscr{B}_{m,m'}} \nu^2_{n,2}(t) - p_2(m,m')\right]_+\right)$. We can remark that $\mathbb{E}\left(F^p_{k\Delta}\right) \leq \frac{1}{\Delta^p}\mathbb{E}\left(Z^{2p}_{k\Delta}\right)$. According to Barlow and Yor (1982, Proposition 4.2 (Burkholder-Davis-Gundy inequality with optimal constants)), there exists a constant c such that, for any $p > 0$:

$$\mathbb{E}\left(Z^{2p}_{k\Delta}\right) \leq c^{2p}(2p)^p\mathbb{E}\left(\left|\int_{k\Delta}^{(k+1)\Delta}\sigma^2(X_s)ds\right|^p\right) \leq c^{2p}(2p)^p\Delta^p\sigma_1^{2p}.$$

Then $\mathbb{E}\left(F^p_{k\Delta}\right) \leq \sigma_1^{2p}c^{2p}(2p)^p$. By Lemma 7 of Schmisser (2012), there exists a constant κ such that, for any $(m,m') \in \mathscr{M}_n^2$,

$$\mathbb{E}\left(\left[\sup_{t\in\mathscr{B}_{m,m'}} \nu^2_{2,n}(t) - p_2(m,m')\right]_+\right) \lesssim \frac{e^{-(D_m+D'_m)}}{n}$$

which ends the proof.

6.14 Proof of Lemma 14 (Estimation of the penalty constant for the estimation of σ^2)

Approximation of the first bound. As in Subsection 6.8, we get an inequality similar to (31):

$$|\hat{\sigma}_1 - \sigma_{1,A}| \leq \left\|\hat{\sigma}^2_m - \sigma^2_A\right\|_\infty^2 \lesssim D_m^{-1} + D_m\left(\frac{1}{n}\sum_{k=1}^n \tilde{E}^2_{k\Delta} + \sum_\lambda \nu^2_{2,n}(\varphi_\lambda)\right).$$

As

$$\tilde{E}_{k\Delta} = (Y_{k\Delta} - U_{k\Delta}) + \left(\frac{\mathbb{E}\left(Z^2_{k\Delta}\right)}{\Delta} - \sigma^2(X_{k\Delta})\right) + \frac{2B_{k\Delta}Z_{k\Delta} + B^2_{k\Delta}}{\Delta},$$

on Ω_B, by Lemma 1,

$$|\tilde{E}_{k\Delta}| \lesssim |Y_{k\Delta} - U_{k\Delta}| + \Delta^{1/2} + \Delta^{1/2}\ln^2(n).$$

We apply Bennett's inequality to $n^{-1}\sum_{k=1}^n (Y_{k\Delta} - U_{k\Delta})^2$. Indeed, by Lemma 10, $\mathbb{E}\left((Y_{k\Delta} - U_{k\Delta})^2\right) \lesssim \Delta^{1-\beta/2}\ln^4(n)$, $|U_{k\Delta} - Y_{k\Delta}| \lesssim \ln^2(n)$ on Ω_B and

$$\mathrm{Var}\left(\sum_{k=1}^{q_n}(Y_{k\Delta} - U_{k\Delta})^2\right) \leq q_n^2\mathbb{E}\left((U_{k\Delta} - Y_{k\Delta})^4\right) \lesssim q_n^2\Delta^{1-\beta/2}\ln^8(n) \lesssim q_n\ln^9(n)\Delta^{-\beta/2}.$$

Then

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n (U_{k\Delta} - Y_{k\Delta})^2 \geq \Delta^{1-\beta/2}\ln^4(n) + \frac{\ln(n)}{\sqrt{n\Delta}}\right) \lesssim \exp\left(-\frac{cn\ln^2(n)/n\Delta}{\Delta^{-\beta/2}\ln^9(n) + \ln^2(n)q_n\ln(n)/\sqrt{n\Delta}}\right)$$
$$\lesssim \exp\left(\frac{-c\ln^2(n)}{\Delta^{1-\beta/2}\ln^9(n) + \ln^4(n)/\sqrt{n\Delta}}\right) \lesssim n^{-6}.$$

Moreover, $\nu_{2,n}(\varphi_\lambda) = n^{-1}\sum_{k=1}^n \varphi_\lambda(X_{k\Delta})F_{k\Delta}$. We recall that $\mathbb{E}\left(\varphi_\lambda(X_{k\Delta})F_{k\Delta}\right) = 0$, $\mathrm{Var}\left(\sum_{k=1}^{q_n}\varphi_\lambda(X_{k\Delta})F_{k\Delta}\right) \leq \sigma_1^4$ and, on Ω_B, $|\varphi_\lambda(X_{k\Delta})F_{k\Delta}| \lesssim D_m \ln(n)$. Therefore, by Bennett's inequality:

$$\mathbb{P}\left(|\nu_{2,n}(\varphi_\lambda)| \geq \frac{\ln^4(n)}{\sqrt{n\Delta}}\right) \lesssim \exp\left(\frac{-cn\ln^2(n)/(n\Delta)}{1+q_n D_m \ln^5(n)/\sqrt{n\Delta}}\right) \lesssim \exp\left(\frac{-c\ln^2(n)}{\ln^{-6}(n)\Delta + \frac{D_m}{\sqrt{n\Delta}}}\right) \lesssim n^{-6}$$

as $D_m \leq \sqrt{n\Delta}$. We get:

$$\mathbb{P}\left(|\hat{\sigma}_1^2 - \sigma_{1,A}^2|^2 \geq \frac{1}{D_m} + D_m\left(\Delta^{1-\beta/2}\ln^4(n) + \frac{\ln^4(n)}{\sqrt{n\Delta}}\right)\right) \lesssim n^{-6}.$$

Taking $D_m = \ln(n)$, we get that

$$\mathbb{P}\left(|\hat{\sigma}_1^2 - \sigma_1^2| \geq 1/\ln(n)\right) \lesssim n^{-6}.$$

Approximation of the second bound. We have that

$$\mathbb{E}\left(Y_{k\Delta}^2\right) = \left(U_{k\Delta} + Y_{k\Delta} - U_{k\Delta}\right)^2 = \mathbb{E}\left(U_{k\Delta}^2\right) + \mathbb{E}\left((Y_{k\Delta} - U_{k\Delta})^2\right) + 2\mathbb{E}\left(U_{k\Delta}(Y_{k\Delta} - U_{k\Delta})\right).$$

By Lemmas 1, 10 and Cauchy-Schwarz,

$$\mathbb{E}\left(Y_{k\Delta}^2 \middle| \mathscr{F}_{k\Delta}\right) = 3\sigma^4(X_{k\Delta}) + C(\Delta^{1/2} + \Delta^{1-\beta/2} + \Delta^{1/2-\beta/4}).$$

and therefore $\mathbb{E}\left(Y_{k\Delta}^2\right) = 3\int \sigma^4(x)\pi(x)dx + C\Delta^{1/2-\beta/4}$ It remains to control the difference with the empirical mean and its expectation by a Bennett inequality. Indeed, on Ω_B, $Y_{k\Delta}^2 \lesssim \ln^4(n)$ and $\mathrm{Var}\left(Y_{k\Delta}^2\right) \lesssim 1$. Then

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^n Y_{k\Delta}^2 - 3\int_{\mathbb{R}}\sigma^4(x)\pi(x)dx\right| \geq C\Delta^{1/2-\beta/4} + \frac{\ln^2(n)}{\sqrt{n\Delta}}\right) \lesssim \exp\left(-\frac{cn\ln^4(n)/(n\Delta)}{1+\ln^4(n)q_n\ln^2(n)/\sqrt{n\Delta}}\right)$$
$$\lesssim \exp\left(-\frac{c\ln^4(n)}{\Delta + \frac{\ln^7(n)}{\sqrt{n\Delta}}}\right) \lesssim n^{-6}.$$

By (33), we get the expected result.

Acknowlegement : The authors wishes to thank Valentine Genon-Catalot and Fabienne Comte for helful discussions. This work was supported in part by the Labex CEMPI (ANR-11-LABX-0007-01)

A Technical tools

Burkholder Davis Gundy inequality. *Let us consider the filtration*

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

Then, for any $p \geq 2$, there exists a constant $C_p > 0$ such that:

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

and

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

Definition of Besov's spaces. *A function g belongs to the Besov space $\mathscr{B}^{\alpha}_{2,\infty}(A)$ if*

- *g belongs to $L^2(A)$*
- *Its modulus of smoothness ω_p satisfies $\sup_{t>0} t^{-\alpha}\omega_r(g,t)_2 < \infty$ with $r = \lfloor \alpha + 1 \rfloor$ and where*

$$\omega_r(g,t)_p := \sup_{0<h\leq t}\|\Delta_h^r(g,.)\|_{L^p(A)} \quad \text{where} \quad \Delta_h^r(g,x) := \sum_{k=0}^{r}(-1)^k C_r^k g(x+kh).$$

This result can be find in Dellacherie and Meyer (1980) or Applebaum (2004).

Berbee's coupling lemma. *Let $(X_t)_{t\geq 0}$ be a stationary and exponentially $\beta-$mixing process observed at discrete times $t = 0, \Delta, \ldots, n\Delta$. Let us set $n = 2p_nq_n$ with $q_n = 8\ln(n)/\Delta$. For any $a \in \{0,1\}$, $1 \leq k \leq p_n$, we consider the random variables*

$$U_{k,a} = \left(X_{((2(k-1)+a)q_n+1)\Delta}, \ldots, X_{(2k-1+a)q_n\Delta}\right).$$

There exist random variables $X^*_{\Delta}, \ldots, X^*_{n\Delta}$ such that

$$U^*_{k,a} = \left(X^*_{((2(k-1)+a)q_n+1)\Delta}, \ldots, X^*_{(2k-1+a)q_n\Delta}\right)$$

satisfy:

- *For any $a \in \{0,1\}$, the random vectors $U^*_{1,a}, U^*_{2,a}, \ldots, U^*_{p_n,a}$ are independent.*
- *For any $(a,k) \in \{0,1\} \times \{1,\ldots,p_n\}$, $U^*_{k,a} \sim U_{k,a}$.*
- *For any $(a,k) \in \{0,1\} \times \{1,\ldots,p_n\}$, $\mathbb{P}\left(U_{k,a} \neq U^*_{k,a}\right) \leq \beta_M(q_n\Delta) \leq n^{-8}$ where β_M is the β-mixing coefficient of (X_t).*

Let us set $\Omega^* = \left\{\omega,\, \forall (k,a) \in \{0,1\} \times \{1,\dots,p_n\},\, U_{k,a} = U^*_{k,a}\right\}$. *Then* $\mathbb{P}\left((\Omega^*)^c\right) \leq n\Delta/n^8$.

Berbee's coupling Lemma is proved by Viennet (1997).

The following Talagrand's inequality is proved by Birgé and Massart (1998) (corollary 2p.354) and Comte and Merlevède (2002) (p222-223).

Talagrand's inequality. *Let* $(X_1,\dots,X_n)$ *be independent identically distributed random variables,* $\mathscr{F}$ *a countable family of functions and define* $I_n : \mathscr{F} \mapsto \mathbb{R}$ *such that*

$$I_n(f) = \frac{1}{n}\sum_{k=1}^{n} f(X_k) - \mathbb{E}\left(f(X_k)\right).$$

If there exists $M, H^2, V \in \mathbb{R}^+$ *such that*

$$\sup_{f\in\mathscr{F}} \|f(X_k)\|_\infty \leq M, \quad \mathbb{E}\left(\sup_{f\in\mathscr{F}} I_n^2(f)\right) \leq H^2, \quad \sup_{f\in\mathscr{F}} \mathrm{Var}\left(f(X_k)\right) \leq V$$

then

$$\mathbb{E}\left(\sup_{f\in\mathscr{F}} I_n^2(f) - 12H^2\right)_+ \lesssim \frac{V}{n}\exp\left(-k_1\frac{nH^2}{V}\right) + \frac{M^2}{n^2}\exp\left(-k_2\frac{nH}{M}\right).$$

where k_1 *and* k_2 *are two explicit constants.*

Bennett inequality. *Let* $(X_1,\dots,X_n)$ *be independent identically distributed random variables such that* $\mathrm{Var}\left(X_i\right) = V$ *and* $|X_i - \mathbb{E}\left(X_i\right)| \leq M$ *p.s (with* $V, M \in \mathbb{R}^+$*. Then, for any* $x > 0$,

$$\mathbb{P}\left(|S_n - \mathbb{E}\left(S_n\right)| \geq x\right) \leq \exp\left(-\frac{x^2}{2(nV + xM/3)}\right)$$

where $S_n = \sum_{i=1}^n X_i$. From Berbee's coupling lemma and Talagrand's and Bennett's inequality, we deduce the following lemma:

Talagrand's and Bennett's inequality for β-mixing variables. *Let us consider* (Z_k) *some random variables, exponentially* β*-mixing, such that the* β*-mixing coefficient of* Z_k *satisfies:*

$$\exists c > 0, \gamma > 0, \forall k,\ \beta_Z(k) \leq ce^{-\gamma k}.$$

We define $q_n := \lceil 2\ln(n)/(\gamma)\rceil$, $p_n = \lceil n/(2q_n)\rceil$. *We have that* $\beta_Z(q_n) \leq ce^{-2\gamma\ln(n)/\gamma} \lesssim n^{-2}$.

Talagrand *Let us consider*

$$I_n(f) = \frac{1}{n}\sum_{k=1}^{n} f(Z_k) - \mathbb{E}\left[f(Z_k)\right]$$

and $\mathscr{F}$ *a countable family of functions. If we can find a triplet (*M*,* V *and* H*) in* $(\mathbb{R}^+)^3$ *such that, for any* $0 \leq j \leq p_n - 1$*:*

$$\forall c, \sup_{f\in\mathscr{F}} \mathrm{Var}\left(\frac{1}{q_n}\sum_{k=jq_n+1}^{(j+1)q_n} f(Z_k)\right) \leq \frac{V}{q_n},$$

$$\forall c, \sup_{f\in\mathscr{F}} \left\|\frac{1}{q_n}\sum_{k=jq_n+1}^{(j+1)q_n} f(Z_k)\right\|_\infty \leq M \quad \textit{and} \quad \mathbb{E}\left[\sup_{f\in\mathscr{F}} |I_n(f)|\right] \leq \frac{H}{\sqrt{n}},$$

then we have, by Talagrand's inequality:

$$\mathbb{E}\left[\sup_{f\in\mathscr{F}}|I_n^2(f)-6H^2|\right]_+ \leq C\frac{V}{n}\exp\left(-k_1\frac{H^2}{12V}\right)+C'\frac{M}{p_n^2}\exp\left(-k_2\frac{\sqrt{p_n}H}{\sqrt{q_n}M}\right)+2\frac{M}{n^2}$$

where C, C', k_1 and k_2 are universal constants.

Benett *Let us set*

$$S_n=\sum_{k=1}^n Z_k \quad \textit{and} \quad \bar{Z}_n=\frac{1}{n}\sum_{k=1}^n Z_k.$$

If we can find V', M' such that, for any $0\leq j\leq p_n-1$

$$\mathrm{Var}\left(\sum_{k=jq_n+1}^{(j+1)q_n} Z_k\right)\leq q_nV' \quad \textit{and} \quad \left|\sum_{k=jq_n+1}^{(j+1)q_n} Z_k-\mathbb{E}\left(Z_k\right)\right|\leq q_nM'$$

then, by Bennett's inequality: $\mathbb{P}\left(|S_n-\mathbb{E}\left(S_n\right)|\geq x\right)\leq 2\exp\left(-\frac{x^2}{2(nV'+xM'q_n)}\right)$ and therefore:

$$\mathbb{P}\left(|\bar{V}_n-\mathbb{E}\left(V_1\right)|\geq x\right)\leq 2\exp\left(-\frac{nx^2}{2(V'+xM'q_n)}\right)$$

Fourier transform.

We denote by $\mathcal{F}h$ the Fourier transform of a function $h\in L^1(\mathbb{R})$:

$$\mathcal{F}h(x)=\int_{\mathbb{R}} f(u)e^{-ixu}du.$$

The Schwarz space is defined as

$$S\left(\mathbb{R}\right)=\left\{h\in\mathscr{C}^\infty,\ \forall p,q\in\mathbb{N},\ \exists C_{p,q},\ \forall x\in\mathbb{R},\ |x^ph^{(q)}(x)|\leq C_{pq}\right\}.$$

Then we have the following properties:

a. *For any $h_1,h_2\in L^2(\mathbb{R})$, $(a_1,a_2)\in\mathbb{R}^2$, $\mathcal{F}(a_1h_1+a_2h_2)=a_1\mathcal{F}h_1+a_2\mathcal{F}h_2$.*

b. *For any $h\in L^2(\mathbb{R})$, $\mathcal{F}h\in L^2(\mathbb{R})$ and $\forall x\in\mathbb{R}$, $h(x)=\frac{1}{2\pi}\int_{\mathbb{R}}e^{itx}\mathcal{F}h(t)dt$.*

c. *For any $h\in L^2(\mathbb{R})$, $\mathcal{F}h(./a)(x)=|a|\mathcal{F}h(ax)$.*

d. *For any functions $h_1,h_2\in L^2(\mathbb{R})$, the Parseval's formula holds:*

$$\int_{\mathbb{R}} h_1(x)\overline{h_2(x)}dx=\frac{1}{2\pi}\int_{\mathbb{R}}\mathcal{F}h_1(u)\overline{\mathcal{F}h_2}(u)du.$$

As $\mathcal{F}\delta_y(x)=e^{-ixy}$,

$$h(0)=\int_{\mathbb{R}} h(y)\delta_0(y)dy=\frac{1}{2\pi}\int_{\mathbb{R}}\mathcal{F}h(u)du$$

e. *For any h in $S(\mathbb{R})$, $\mathcal{F}h\in S(\mathbb{R})$ and*

$$\mathcal{F}(h^{(q)})(x)=(ix)^q\mathcal{F}h(x).$$

References

Yacine Aït-Sahalia and Jean Jacod. Estimating the degree of activity of jumps in high frequency data. *Ann. Statist.*, 37(5A):2202–2244, 2009. ISSN 0090-5364.

David Applebaum. *Lévy processes and stochastic calculus*, volume 93 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2004.

M. T. Barlow and M. Yor. Semimartingale inequalities via the Garsia-Rodemich-Rumsey lemma, and applications to local times. *J. Funct. Anal.*, 49(2):198–229, 1982.

Andrew Barron, Lucien Birgé, and Pascal Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999.

Julien Berestycki. Exchangeable fragmentation-coalescence processes and their equilibrium measures. *Electron. J. Probab.*, 9:no. 25, 770–824 (electronic), 2004.

Lucien Birgé and Pascal Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998.

F. Comte and V. Genon-Catalot. Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Process. Appl.*, 119(12):4088–4123, 2009.

F. Comte and Y. Rozenholc. Adaptive estimation of mean and volatility functions in (auto-)regressive models. *Stochastic Process. Appl.*, 97(1):111–145, 2002.

F. Comte and Y. Rozenholc. A new algorithm for fixed design regression and denoising. *Ann. Inst. Statist. Math.*, 56(3):449–473, 2004.

Fabienne Comte and Florence Merlevède. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.*, 6:211–238 (electronic), 2002. New directions in time series analysis (Luminy, 2001).

Fabienne Comte, Valentine Genon-Catalot, and Yves Rozenholc. Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13(2):514–543, 2007.

Claude Dellacherie and Paul-André Meyer. *Probabilités et potentiel. Chapitres V à VIII*, volume 1385 of *Actualités Scientifiques et Industrielles [Current Scientific and Industrial Topics]*. Hermann, Paris, revised edition, 1980. Théorie des martingales. [Martingale theory].

Ronald A. DeVore and George G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

Arnaud Gloter, Dasha Loukianova, and Hilmar Mai. Jump filtering and efficient drift estimation for lévy-driven sde's. 2016.

Shota Gugushvili. Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat.*, 21(3):321–343, 2009.

Muhammad Hanif, HanChao Wang, and ZhengYan Lin. Reweighted Nadaraya-Watson estimation of jump-diffusion models. *Sci. China Math.*, 55(5):1005–1016, 2012.

Marc Hoffmann. L_p estimation of the diffusion coefficient. *Bernoulli*, 5(3):447–481, 1999.

Jean Jacod. Asymptotic properties of power variations of Lévy processes. *ESAIM Probab. Stat.*, 11:173–196, 2007.

Johanna Kappus. Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Process. Appl.*, 124(1):730–758, 2014.

Hilmar Mai. Efficient maximum likelihood estimation for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 20(2):919–957, 2014.

Cecilia Mancini and Roberto Renò. Threshold estimation of Markov models with jumps and interest rate modeling. *J. Econometrics*, 160(1):77–92, 2011.

Hiroki Masuda. Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.*, 117(1):35–56, 2007.

Yves Meyer. *Ondelettes et opérateurs. I.* Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris, 1990. Ondelettes. [Wavelets].

Michael H. Neumann and Markus Reiß. Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248, 2009.

Philip Protter and Denis Talay. The Euler scheme for Lévy driven stochastic differential equations. *Ann. Probab.*, 25(1):393–423, 1997.

Emeline Schmisser. Non-parametric estimation of the diffusion coefficient from noisy data. *Stat. Inference Stoch. Process.*, 15(3):193–223, 2012.

Yasutaka Shimizu. Some remarks on estimation of diffusion coefficients for jump-diffusions from finite samples. *Bull. Inform. Cybernet.*, 40:51–60, 2008.

Yasutaka Shimizu and Nakahiro Yoshida. Estimation of parameters for diffusion processes with jumps from discrete observations. *Stat. Inference Stoch. Process.*, 9(3):227–277, 2006.

Alexandre B. Tsybakov. *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2004.

G. Viennet. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 1997.