Non parametric estimation of the coefficients of a diffusion with jumps

Émeline Schmisser

Abstract

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

1 Introduction

We consider the stochastic differential equation (SDE):

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

(1)

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

\[ L_t = \int_0^t \int_{z \in \mathbb{R}} z (\mu(dt, dz) - \nu(dz)) dt \]

where \(\mu\) 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 \(\beta\)-mixing. It is observed at discrete times \(t = 0, \Delta, \ldots, n\Delta\) where the sampling interval \(\Delta\) tends to 0 and the time of observation \(n\Delta\) tends to infinity. Our aim is to construct adaptive non-parametric estimators of \(\xi^2\) and \(\sigma^2\) on a compact set \(A\).
Diffusions with jumps become powerful tools to model processes in biology, physics, social sciences, medical sciences, economics, and a variety of financial applications such as interest rate modelling or derivative pricing. However, if the non-parametric estimation of the coefficients of a diffusion without jumps is well known (see for instance Hoffmann (1999) or Comte et al. (2007)), to our knowledge, there do not exist adaptive estimators for the coefficients of a jump diffusion, neither minimax rates of convergence. Shimizu (2008) construct maximum-likelihood parametric estimators of $\sigma^2$ and $\xi^2$. Their estimators converge with rates $\sqrt{n}$ and $\sqrt{n\Delta}$ respectively. Mancini and Renò (2011) and Hanif et al. (2012) construct non-parametric estimators of $\sigma^2$ and $\sigma^2 + \xi^2$ thanks to kernel or local polynomials estimators. The estimator of $\sigma^2$ converges with rate $\sqrt{h_n}$, meanwhile the estimator of $\xi^2 + \sigma^2$ converges with rate $\sqrt{n\Delta h}$, where $h$ is the bandwidth of the estimator.

In this paper, we construct non-parametric estimators of $g = \sigma^2 + \xi^2$ and $\sigma^2$ under the asymptotic framework $n\Delta \to \infty$ and $\Delta \to 0$ by model selection. This method was introduced by Birgé and Massart (1998). We consider first 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

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

We bound the risk of $\hat{g}_m$, then we introduce a penalty function $\text{pen}(m)$ and we minimize on $m$ the function $\gamma_n(\hat{g}_m) + \text{pen}(m)$. If the Lévy measure $\nu$ is sub-exponential, the adaptive estimator $\hat{g}_\hat{m}$ satisfies an oracle inequality (up to a multiplicative constant).

To estimate the function $\sigma^2$, we need to cut off the jumps. We minimize over each $S_m$ the contrast function

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

We obtain a sequence of estimators $\hat{\sigma}^2_m$ of $\sigma^2$. The risk of these estimators depends on the Blumenthal-Getoor index of $\nu$. To construct an adaptive estimator, $\hat{\sigma}^2_\hat{m}$, we again introduce a penalty function $\text{pen}(m)$. The estimator $\hat{\sigma}^2_\hat{m}$ automatically realizes a bias-variance compromise. The rates of convergence obtained for $\hat{g}_m$ and $\hat{\sigma}^2_m$ are similar to those obtained by Hanif et al. (2012) and Mancini and Renò (2011).

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.

2 Model

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

2 Model
A 1.  1. The functions $b$, $\sigma$ and $\xi$ are Lipschitz.

2. The functions $\sigma$ and $\xi$ are bounded: $\exists \sigma_0^2, \xi_0^2$ such that
   \[ \forall x \in \mathbb{R}, \ 0 < \sigma^2(x) \leq \sigma_0^2 \quad \text{and} \quad 0 < \xi^2(x) \leq \xi_0^2. \]

Moreover either there exists a positive constant $\sigma_1^2$ such that $\forall x \in \mathbb{R}, \ \sigma^2(x) \geq \sigma_1^2 > 0$, or there exists $\xi_1^2$ such that, $\forall x \in \mathbb{R}, \ \xi^2(x) \geq \xi_1^2 > 0$.

3. The function $b$ is elastic: $\exists M, C, \forall x, |x| > M, \ b(x)x \leq -Cx^2$.

4. The Lévy measure $\nu$ satisfies:
   \[ \nu \{ \{0\} \} = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} z^4 \nu(dz) < \infty \]

and the Blumenthal-Getoor index is strictly less than 2: there exists $\beta \in [0,2[$ such that $\int_{1}^{t} z^\beta \nu(dz) < \infty$. This is a classical assumption (see for instance Mai (2012)). In order to ensure the uniqueness of the function $\xi$, we also assume that $\int_{-\infty}^{+\infty} z^2 \nu(dz) = 1$.

If Assumption A1 is satisfied, SDE (1) as a unique solution. According to Masuda (2007), under assumptions A1.(1-3), the process $(X_t)_{t \geq 0}$ is exponentially $\beta$-mixing and has a unique invariant probability. Moreover, under assumption A1(4), $E(X_t^3) < \infty$. Then we can assume:

A 2. The process $(X_t)_{t \geq 0}$ is stationary, exponentially $\beta$-mixing and its stationary measure has a density $\pi$ which is bounded on any compact set.

The following result is very useful. It comes from Dellacherie and Meyer (1980) or Applebaum (2004).

Burkholder Davis Gundy inequality. Let us consider the filtration
\[ \mathcal{F}_t = \sigma(\eta_t, (W_s)_{0 \leq s \leq t}, (L_s)_{0 \leq s \leq t}). \]

Then, for any $p \geq 2$, there exists a constant $C_p > 0$ such that:
\[
E \left( \sup_{s \leq t} \left| \int_s^t \sigma(X_u) dW_u \right|^p \bigg| \mathcal{F}_t \right) \leq C_p E \left( \left| \int_t^{t+h} \sigma^2(X_u) du \right|^{p/2} \bigg| \mathcal{F}_t \right)
\]

and
\[
E \left( \sup_{s \leq t} \left| \int_s^t \xi(X_u-) dL_u \right|^p \bigg| \mathcal{F}_t \right) \leq C_p E \left( \left| \int_t^{t+h} \xi^2(X_u) du \right|^{p/2} \bigg| \mathcal{F}_t \right) + C_p E \left( \left| \int_t^{t+h} |\xi^p(X_u)| du \right| \bigg| \mathcal{F}_t \right) E \left( \int_\mathbb{R} |z|^p \nu(dz) \right)
\]

The following proposition derives from this result.

**Proposition 1.** For any integer $p$ and any $t \leq 1$:
\[
E \left( \sup_{0 \leq s \leq t} (X_{t+s} - X_u)^{2p} \right) \preceq t.
\]
Now we introduce an increasing sequence of vectorial subspaces \((S_m)_{m\geq 0}\) of \(L^2(A)\) satisfying the following properties:

**A 3.** 1. The subspaces \(S_m\) have finite dimension \(D_m\) and are increasing: \(\forall m, S_m \subseteq S_{m+1}\).

2. The \(\|\cdot\|_L^2\) and \(\|\cdot\|_\infty\) 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 \(\|t\|_L^2 = \int_A t^2(x)dx\) and \(\|t\|_\infty = \sup_{x \in A} |t(x)|\).

3. For any function \(t \in B_{\alpha_2}^{\infty}\), \(\exists c, \forall m, \|t - t_m\|_L^2 \leq c D_m^{-2\alpha}\) where \(t\) is the orthogonal projection \(L^2\) of \(t\) on \(S_m\).

The vectorial subspaces generated by the trigonometric polynomials, the piecewise polynomials, the spline functions and the wavelets satisfy these properties (see Meyer (1990) and DeVore and Lorentz (1993) for the proofs).

### 3 Estimation of \(\sigma^2 + \xi^2\)

Let us set \(Z_{k\Delta} = \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s)dW_s\) and \(J_{k\Delta} = \int_{k\Delta}^{(k+1)\Delta} \xi(X_s)\,dL_s\). To estimate \(\sigma^2\) 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 jump diffusions,

\[
X_{(k+1)\Delta} - X_{k\Delta} = \int_{k\Delta}^{(k+1)\Delta} b(X_s)ds + Z_{k\Delta} + J_{k\Delta}
\]

and therefore

\[
T_{k\Delta} = \sigma^2(X_{k\Delta}) + \xi^2(X_{k\Delta}) + A_{k\Delta} + B_{k\Delta} + E_{k\Delta}
\]

where

\[
A_{k\Delta} = A_{k\Delta}^{(1)} + A_{k\Delta}^{(2)} + A_{k\Delta}^{(3)} + A_{k\Delta}^{(4)}
= \frac{1}{\Delta} \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s)ds \right)^2 + \frac{2}{\Delta} (Z_{k\Delta} + J_{k\Delta}) \int_{k\Delta}^{(k+1)\Delta} (b(X_s) - b(X_{k\Delta})) ds
+ \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} (\sigma^2(X_s) - \sigma^2(X_{k\Delta})) ds + \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} (\xi^2(X_s) - \xi^2(X_{k\Delta})) ds,
\]

\[
B_{k\Delta} = B_{k\Delta}^{(1)} + B_{k\Delta}^{(2)} = 2b(X_{k\Delta})Z_{k\Delta} + \frac{1}{\Delta} \left( \int_{k\Delta}^{(k+1)\Delta} Z_{k\Delta}^2 - \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s)ds \right)
\]
3.1 Estimation for fixed $m$

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

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

Let us bound the empirical risk $\hat{\mathcal{R}}_n(\hat{g}_m)$, where

$$
\hat{\mathcal{R}}_n(t) = \mathbb{E} \left( \|t - g\|_n^2 \right) \text{ with } \|t\|_n^2 = \frac{1}{n} \sum_{k=1}^{n} t^2(X_{k\Delta}).
$$

We set $\|t\|_n^2 = \int_A t^2(x) \pi(x) dx$ and $g_A = g 1_A$.

We have that

$$
\gamma_n(t) = \frac{1}{n} \sum_{k=1}^{n} (t(X_{k\Delta}) - g(X_{k\Delta}) + A_{k\Delta} + B_{k\Delta} + E_{k\Delta})^2
$$

$$
= \|t - g\|_n^2 + \frac{1}{n} \sum_{k=1}^{n} (A_{k\Delta} + B_{k\Delta} + E_{k\Delta})^2
$$

$$
- \frac{2}{n} \sum_{k=1}^{n} (A_{k\Delta} + B_{k\Delta} + E_{k\Delta}) (g(X_{k\Delta}) - t(X_{k\Delta})).
$$

As $\hat{g}_m$ minimizes $\gamma_n(t)$, the inequality $\gamma_n(\hat{g}_m) \leq \gamma_n(g_m)$ holds and then

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

By Cauchy-Schwarz, and as $\hat{g}_m$ and $g_m$ are $A$-supported,

$$
\|\hat{g}_m - g_A\|_n^2 \leq \|g_m - g_A\|_n^2 + \frac{12}{n} \sum_{k=1}^{n} A_{k\Delta}^2 + \frac{1}{12} \|\hat{g}_m - g_m\|_n^2 + \frac{12}{n} \sup_{t \in \mathcal{R}_m} \nu_n^2(t) + \frac{1}{12} \|\hat{g}_m - g_m\|_n^2.
$$
where $\mathcal{B}_m = \{ t \in S_m, \| t \|_2^2 \leq 1 \}$ and $\nu(t) = \frac{1}{n} \sum_{k=1}^n (B_k \Delta + E_k \Delta) t(X_k \Delta)$. Let us set

$$\Omega_n = \left\{ \omega, \forall m \in \mathcal{M}_n, \forall t \in S_m, \left| \frac{\| t \|_n^2}{\| t \|_2^2} - 1 \right| \leq \frac{1}{2} \right\},$$

where the norms $\| \cdot \|_M$ and $\| \cdot \|_n$ are equivalent. The following lemma is proved by Comte et al. (2007) for diffusion processes, but only relies on the $\beta$-mixing and stationary properties.

**Lemma 3.**

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

$$\mathbb{E} (\| \hat{g}_m - g_A \|_n^2 \mathbb{1}_{\Omega_n}) \leq 3 \| g_m - g_A \|_n^2 + 12 \mathbb{E} (A_k^2 \Delta) + 12 \mathbb{E} \left( \sup_{t \in \mathcal{B}_m} \nu_n(t) \right).$$

On $\Omega_n$, any function $t \in S_m$ satisfies: $\| t \|_2^2 \leq 2 \| t \|_n^2$. Moreover, for any deterministic function $t$, $\mathbb{E} (\| t \|_n^2) = \| t \|_2^2$. Consequently:

$$\mathbb{E} (\| \hat{g}_m - g_A \|_n^2 \mathbb{1}_{\Omega_n}) \leq 3 \| g_m - g_A \|_n^2 + 12 \mathbb{E} (A_k^2 \Delta) + 12 \mathbb{E} \left( \sup_{t \in \mathcal{B}_m} \nu_n(t) \right).$$

By Assumption A2, $\pi$ is bounded on $A$ and then $\| g_m - g_A \|_n^2 \lesssim \| g_m - g_A \|_2^2$. The remainder of the proof is done in Section 6.

**Theorem 4.** Under Assumptions A1-A3, if $m \in \mathcal{M}_n$, the risk of the estimator $\hat{g}_m$ is bounded by:

$$\mathcal{R}_n(\hat{g}_m) \lesssim \| g_m - g_A \|_2^2 + \frac{D_m}{n \Delta} \xi_0^4 + \frac{D_m}{n} (\sigma_4^2 + \sigma_6^2 \xi_0^2) + \frac{1}{n \Delta} + \Delta$$

where $g_m$ is the orthogonal ($L^2$) projection of $g$ on $S_m$.

We have to find a good compromise between the bias term, $\| g_m - g_A \|_2^2$, which decreases when $m$ increases, and the variance term, proportional to $D_m/(n \Delta)$. If $g$ belongs to the Besov space $B_{2, \infty}^s$, then the bias term $\| g_m - g_A \|_2^2 \sim D_m^{-2s}$. The risk is then minimum for $m_{\text{opt}} = (n \Delta)^{1/(1+2s)}$, and satisfies

$$\mathcal{R}_n(\hat{g}_{m_{\text{opt}}}) \lesssim (n \Delta)^{-2s/(2s+1)} + \Delta.$$

### 3.2 Adaptive estimator

To bound the risk of the adaptive estimator, we need the additional assumption:

**A 4.**

1. The Lévy measure $\nu$ is sub-exponential:

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

2. There exist $\eta, \eta > 1$, such that $\Delta^\eta = O(n^{-1})$.  

6
Let us consider the penalty function \( \text{pen}(m) = K_0 \xi_4 D_{\Delta}^4 \) and choose the adaptive estimator \( \hat{g}_m \) by minimizing the function

\[
\hat{m} = \min_{m \in \mathcal{M}} \gamma_n(\hat{g}_m) + \text{pen}(m).
\]

We introduce the function \( p(m, m') = \frac{\text{pen}(m) + \text{pen}(m')}{12} \). For any \( m \in \mathcal{M} \),

\[
\mathbb{E}\left( \| \hat{g}_m - g \|^2_{2, \Omega_n} \right) \lesssim \| g_m - g \|^2_{L_2} + \mathbb{E} \left( A_{\beta, \Delta}^2 \right) + 2 \text{pen}(m)
\]

\[
+ 12 \mathbb{E} \left( \sum_{m' \in \mathcal{M}_n} \left( \sup_{t \in \mathcal{M}_{m, m'}} \nu_n(t) - p(m, m') \right) \right)
\]

where \( \mathcal{M}_{m, m'} = \{ t \in S_m + S_{m'}, \| t \|_\pi \leq 1 \} \). In order to bound the remaining term,

\[
\mathbb{E} \left( \sup_{t \in \mathcal{M}_{m, m'}} \nu_n(t) - p(m, m') \right),
\]

we use the Berbee’s coupling Lemma and a Talagrand’s inequality. Berbee’s coupling Lemma is proved by Viennet (1997). As the random variables \( (X_{k, \Delta}) \) are exponentially \( \beta \)-mixing, it allows us to deal with independent random variables.

**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 \( n = 2p_n q_n \) with \( q_n = 8 \ln(n)/\Delta \). For any \( a \in \{0, 1\} \), \( 1 \leq k \leq p_n \), we consider the random variables

\[
U_{k, a} = (X_{(2(k-1)+a)q_n+1}\Delta, \ldots, X_{(2k-1+a)q_n\Delta}).
\]

There exist random variables \( X_{\Delta}^*, \ldots, X_{n\Delta}^* \) such that

\[
U_{k, a}^* = (X_{(2(k-1)+a)q_n+1}\Delta, \ldots, X_{(2k-1+a)q_n\Delta}^*)
\]

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 \( (a, k) \in \{0, 1\} \times \{1, \ldots, p_n\} \), \( \mathbb{P}\left( U_{k, a} \neq U_{k, a}^* \right) \leq \beta(q_n \Delta) \leq n^{-8} \).

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

The following Talagrand’s inequality is proved by Birgé and Massart (1998) (corollary 2p.354) and Comte and Merlevêde (2002) (p222-223).
Talagrand’s inequality. Let \((X_1, \ldots, X_n)\) be independent identically distributed random variables and \(f_n : \mathcal{B}_{m,m'} \to S_m\) such that
\[ f_n(t) = \frac{1}{n} \sum_{k=1}^{n} F_t(X_k) - E(F_t(X_k)). \]
If
\[ \sup_{t \in \mathcal{B}_{m,m'}} \|F_t\|_{\infty} \leq M, \quad E \left( \sup_{t \in \mathcal{B}_{m,m'}} f_n^2(t) \right) \leq H^2, \quad \sup_{t \in \mathcal{B}_{m,m'}} \text{Var}(F_t(X_k)) \leq V \]
then
\[ E \left( \sup_{t \in \mathcal{B}_{m,m'}} f_n^2(t) - 12H^2 \right) \leq \frac{V}{n} \exp \left( -k_1 \frac{nH^2}{V} \right) + \frac{M^2}{n^2} \exp \left( -k_2 \frac{nH}{M} \right). \]

We then obtain the following oracle inequality:

**Theorem 5.** Under assumptions \(A1-A4\) there exists \(\kappa_0\) such that for any \(\kappa \geq \kappa_0\),
\[ \mathcal{R}_n (\hat{g}_n) \lesssim \inf_{m \in \mathcal{R}_n} \left\{ \|g_m - g_A\|_{L^2} + \text{pen}(m) \right\} + \Delta + \frac{\ln^3(n)}{n\Delta}. \]
The adaptive estimator \(\hat{g}_n\) automatically realises the best (up to a multiplicative constant) compromise.

4 Estimation of \(\sigma^2\).

We have that
\[ T_{k}\Delta = \frac{(X_{(k+1)}\Delta - X_{k}\Delta)^2}{\Delta} = \sigma^2(X_k\Delta) + \frac{1}{\Delta} J_{k}\Delta + \text{small terms} + \text{centred terms}. \]
The idea is to keep \(T_{k}\Delta\) only when there is no jumps. As the stochastic term \(Z_{k}\Delta\) is of order \(\Delta^{1/2}\), we can only suppress the jumps of amplitude greater than \(\Delta^{1/2}\). Then we consider:
\[ Y_{k}\Delta = \frac{(X_{(k+1)}\Delta - X_{k}\Delta)^2}{\Delta} \mathbf{I}_{\Omega_{X,k}} \]
where \(\Omega_{X,k} = \{\omega, |X_{(k+1)}\Delta - X_{k}\Delta| \leq (\sigma_0 + \xi_0) \ln(n) \Delta^{1/2} + \Delta^{1/2}\}\). We have that
\[ Y_{k}\Delta = \sigma^2(X_k\Delta) - \sigma^2(X_k\Delta) \mathbf{I}_{\Omega_{X,k}} + \xi^2(X_k\Delta-) \mathbf{I}_{\Omega_{X,k}} + (A_{k}\Delta + B_{k}\Delta + E_{k}\Delta) \mathbf{I}_{\Omega_{X,k}} \]
\[ = \sigma^2(X_k\Delta) - \sigma^2(X_k\Delta) \mathbf{I}_{\Omega_{X,k}} + (\tilde{A}_{k}\Delta + B_{k}\Delta + \tilde{E}_{k}\Delta) \mathbf{I}_{\Omega_{X,k}} \]
with \(\tilde{A}_{k}\Delta = A^{(1)}_{k}\Delta + A^{(2)}_{k}\Delta + A^{(3)}_{k}\Delta\) and \(\tilde{E}_{k}\Delta = E^{(1)}_{k}\Delta + E^{(2)}_{k}\Delta + \frac{1}{\Delta} \left( f_{k}\Delta \right)^2 \xi(X_{s-})dL_{s}\).
Let us consider
\[ J^{(1)}_{k}\Delta = f^{(k+1)}_{\Delta} \xi(X_{s-})dL^{(1)}_{s}, \]
with
\[ L^{(1)}_{s} = \int_{|z| \leq \Delta^{1/2}} z\mu(dz,ds), \quad L^{(2)}_{s} = \int_{|z| \leq \Delta^{1/2}, \Delta^{1/4}} z\mu(dz,ds), \]
\[ L^{(3)}_{s} = \int_{|z| > \Delta^{1/4}} z\mu(dz,ds) \]
and denote by $N_k = \mu \left(\left[(k\Delta, (k+1)\Delta], [-\Delta^{1/4}, \Delta^{1/4}]\right]\right)$ the number of jumps of amplitude greater than $\Delta^{1/4}$ on the time interval $[k\Delta, (k+1)\Delta]$. We introduce the set

$$\Omega_{N,k} = \left\{ \omega, N_k = 0 \text{ and } \left| \int_{k\Delta}^{(k+1)\Delta} dL_s^{(1)} + dL_s^{(2)} \right| \leq 4 \frac{\sigma_0 + \xi_0}{\xi_1} \Delta^{1/2} \ln(n) \right\}.$$  

The term $B_{k\Delta} \mathbb{1}_{\Omega_{N,k}}$ is no longer centred. Let us set

$$\tilde{B}_{k\Delta} = B_{k\Delta} \mathbb{1}_{\Omega_{N,k}} - \mathbb{E} \left( B_{k\Delta} \mathbb{1}_{\Omega_{N,k}} \mid \mathcal{F}_{k\Delta} \right)$$

and

$$F_{k\Delta} = \left( \tilde{A}_{k\Delta} + \tilde{E}_{k\Delta} \right) \mathbb{1}_{\Omega_{N,k}} - \sigma^2(X_{k\Delta}) \mathbb{1}_{\Omega_{N,k}} + B_{k\Delta} \mathbb{1}_{\Omega_{N,k}} - \mathbb{E} \left( B_{k\Delta} \mathbb{1}_{\Omega_{N,k}} \mid \mathcal{F}_{k\Delta} \right).$$

Then

$$Y_{k\Delta} = \sigma^2(X_{k\Delta}) + F_{k\Delta} + \tilde{B}_{k\Delta}.$$ 

The following assumption is needed.

**A 5.** 

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

2. There exists $\eta, \eta > 1$ such that $\Delta^\eta = O(n^{-1})$.

The following lemmas are proved later.

**Lemma 6.** $\mathbb{P} \left( \Omega_{X,k}^c \right) \lesssim \Delta^{1-\beta/2} + n^{-1}$, $\mathbb{P}(\Omega_{N,k}^c) \lesssim \Delta^{1-\beta/2}$ and $\mathbb{P} \left( \Omega_{X,k} \cap \Omega_{N,k}^c \right) \lesssim \Delta^{2-\beta/2} + n^{-1}$.

**Lemma 7.**

- $\mathbb{E} \left( \tilde{A}_{k\Delta}^2 \mid \mathcal{F}_{k\Delta} \right) \lesssim \Delta$ and $\mathbb{E} \left( \tilde{A}_{k\Delta} \mid \mathcal{F}_{k\Delta} \right) \lesssim \Delta$.
- $\mathbb{E} \left( \tilde{B}_{k\Delta} \mid \mathcal{F}_{k\Delta} \right) = 0$, $\mathbb{E} \left( \tilde{B}_{k\Delta}^2 \mid \mathcal{F}_{k\Delta} \right) \leq \sigma_0^4/n$ and $\mathbb{E} \left( \tilde{B}_{k\Delta}^4 \mid \mathcal{F}_{k\Delta} \right) \lesssim 1$.
- $\mathbb{E} \left( \tilde{E}_{k\Delta}^2 \mathbb{1}_{\Omega_{N,k}} \mid \mathcal{F}_{k\Delta} \right) \lesssim \Delta^{1-\beta/2}$ and $\mathbb{E} \left( \tilde{E}_{k\Delta}^4 \mathbb{1}_{\Omega_{N,k}} \mid \mathcal{F}_{k\Delta} \right) \lesssim 1$.

### 4.1 Estimator for fixed $m$

We consider the following contrast function and the empirical risk

$$\hat{\gamma}_n(t) = \frac{1}{n} \sum_{k=1}^{n} (t(X_{k\Delta}) - Y_{k\Delta})^2 \mathbb{1}_{X_{k\Delta} \in A} \quad \text{and} \quad \mathcal{A}_n(t) = \mathbb{E} \left( \| t - \sigma^2 \|_n^2 \right).$$

Let us set $\hat{\sigma}_m^2 = \arg\inf_{t \in \mathcal{A}_m} \hat{\gamma}_n(t)$.

**Theorem 8.** Under Assumptions **A1**, **A2**, **A3** and **A4** we have that

$$\mathcal{A}_n(\hat{\sigma}_m^2) \leq \| \sigma_2^2 - \sigma_m^2 \|_2^2 + \sigma_0^4 D_m \frac{1}{n} + \Delta^{1-\beta/2} \ln^4(n)$$

where $\sigma_2^2(x) = \sigma^2(x) \mathbb{1}_{x \in A}$. 

9
The bias term \( \|\sigma_A^2 - \sigma_m^2\|_{L^2}^2 \) and the variance term \( \sigma_m^2 D_m n^{-1} \) are the same as for a diffusion without jumps. Nevertheless, the remainder term is \( \Delta^2 \) for a diffusion process (see for instance Comte et al. (2007)). Even for Poisson processes, the remainder term will be here proportional to \( \Delta \ln(n) \).

If \( \sigma^2 \) belongs to \( \mathcal{B}_{2,\infty}^\mathbb{R} \), then \( \|\sigma_A^2 - \sigma_m^2\|_{L^2}^2 \lesssim D_m^{-2\alpha} \). The best estimator is obtained for \( D_{m,\bar{\alpha}} = n^{-1/(1+2\alpha)} \) and its risk is bounded by \( n^{-2\alpha/(2\alpha+1)} + \Delta^{1-\beta/2} \).

**Remark 9.** Let us set \( \Delta \sim n^{-a} \), with \( 0 < a < 1 \). We have the following rates of convergence:

| \( a \) | jumps | diffusions |
|-------|-------|------------|
| \( \frac{2\alpha}{2(2\alpha+1)} \leq a \leq \frac{2\alpha}{2(2\alpha+1) + 1} \) | \( \Delta^{1/2-\beta/4} \) | \( \Delta^{1/2-\beta/4} \) |
| \( \frac{2\alpha}{2(2\alpha+1) + 1} \leq a \leq 1 - \frac{\alpha}{(2\alpha+1)} \) | \( n^{-\alpha/(2\alpha+1)} \) | \( n^{-\alpha/(2\alpha+1)} \) |

If \( \beta \geq 0 \), 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 \( \sigma^2 \) for diffusions processes (see for instance Hoffmann (1999)). If \( \beta \) or \( \alpha \) 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.

### 4.2 Adaptive estimator

Let us introduce a penalty function \( \tilde{\pi}(m) = \kappa n^{-1} \sigma_0^2 \) and define the adaptive estimator \( \tilde{\sigma}_m^2 \):

\[
\tilde{m} = \arg \min_{m \in \mathcal{M}_n} \tilde{\gamma}_n(\tilde{\sigma}_m^2) + \tilde{\pi}(m)
\]

where \( \mathcal{M}_n = \{m, D_m \leq \mathcal{D}_n\} \). As for the adaptive estimator of \( g = \sigma^2 + \xi^2 \), we use the Berbee’s coupling lemma and the Talagrand’s inequality to bound the risk of the estimator \( \tilde{\sigma}_m^2 \).

**Theorem 10.** Under Assumptions \( A\mathbb{I} A\mathbb{Z} \) and \( A\mathbb{X} \) there exists \( \kappa_1 \) such that, if \( \kappa \geq \kappa_1 \), we have the following oracle inequality:

\[
\mathcal{R}_n(\tilde{\sigma}_m^2) \lesssim \min_{m \in \mathcal{M}_n} \left( \|\sigma_A^2 - \sigma_m^2\|_{L^2}^2 + \tilde{\pi}(m) \right) + \Delta^{1-\beta/2} \ln^4(n) + \frac{1}{n}.
\]

**Remark 11.** If Assumptions \( A\mathbb{I} A\mathbb{Z} \) are satisfied, the risk of the estimator \( \xi^2 = g\tilde{m} - \tilde{\sigma}_m^2 \) satisfies the following inequality:

\[
\mathbb{E}(\|\xi^2 - \xi_A^2\|) \lesssim \min_{m \in \mathcal{M}_n} \left\{ \|g\tilde{m} - gA\|_{\mathbb{R}}^2 + \kappa_0 \frac{D_m}{n\Delta} \right\} + \min_{m \in \mathcal{M}_n} \left\{ \|\sigma_m^2 - \sigma_A^2\| + \kappa_1 \frac{D_m}{n} \right\} + \Delta^{1-\beta/2} \ln^2(n).
\]

### 5 Simulations

#### 5.1 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 $(\zeta_k)$ are centred, independent, and identically distributed random variables. We denote by $F$ the law of $\zeta$ and we assume that $\mathbb{E}(\zeta_k^2) = 1$ and that the random variables $(\zeta_k)$ are independent of $(\eta, (W_t)_{t \geq 0}, N_t)$.

### 5.1.1 Model 1: Ornstein Uhlenbeck

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

with binomial jumps: $P(\zeta = 1) = P(\zeta = -1) = 0.5$.

### 5.1.2 Model 2

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

with Laplace jumps: $f(dz) = \nu(dz) = 0.5 e^{-\lambda|x|}$.

### 5.1.3 Model 3

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

with normal jumps: $\zeta_k \sim \mathcal{N}(0, 1)$.

### 5.1.4 Model 4:

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

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

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

### 5.2 Method

We use the vectorial subspaces generated by the spline functions:

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

and $g_r = \mathbbm{1}_{x \in A} * \ldots * \mathbbm{1}_{x \in A}$

Those subspaces form a multi-resolution analysis of $L^2(A)$. We use the same simulation method as in [Rubenthaler (2010)]

To construct the adaptive estimator, we compute $\hat{f}_{m,r}$ for $D_m \leq \sqrt{n \Delta}$, $0 \leq r \leq 4$ and $m \leq 7$ (for $m = 7$, we already have $D_m = 128$. If $m$ was bigger, there will be a memory problem). Then we minimize $\gamma_n(\hat{f}_{m,r}) + \text{pen}(m, r)$ with respect to $m$, then $r$. There is three constants in the penalty function $\text{pen}(m, r)$. The constants $\sigma^2_D$ and $\xi_D$ are unknown, but they can be replaced
by rough estimators, as only an upper bound for $\sigma^4_0$ and $\xi^4_0$ is needed. In our simulations, we took the true value of $\sigma^4_0$ and $\xi^4_0$. The constants $\kappa_0$ and $\kappa_1$ are chosen by numerical calibration (see Comte and Rozenholc (2002, 2004) for a complete discussion). Another way of dealing with the constants of the penalty would be the slope method developed by Arlot and Massart (2009), however, this method is a bit slow.

To obtain Figures 1-4, for each model, we realize 5 simulations and draw the 5 corresponding estimators. To construct Tables (5.3)-(5.3), for each couplet $(n, \Delta)$ and each model, we make 50 simulations, and for each simulation, we compute the adaptive estimator $\hat{g}_{m, \hat{r}}$ or $\hat{\sigma}_{m, \hat{r}}$, the selected dimension $(\hat{m}, \hat{r})$ and the empirical error

$$err = \frac{1}{n} \sum_{k=1}^{n} (\hat{g}_{m, \hat{r}}(X_k\Delta) - g(X_k\Delta))^2 \mathbb{1}_{X_k\Delta \in A}$$

or

$$err = \frac{1}{n} \sum_{k=1}^{n} (\hat{\sigma}^2_{m, \hat{r}}(X_k\Delta) - \sigma^2(X_k\Delta))^2 \mathbb{1}_{X_k\Delta \in A}.$$

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

In the tables, we write the following informations:

- mean of the empirical errors of $\hat{g}_{m, \hat{r}}$ and $\hat{\sigma}^2_{m, \hat{r}}$, risk
- oracle or $= \text{mean}(err/err_{\text{min}})$.
- $m_{\text{est}}$ and $r_{\text{est}}$, means of $\hat{m}$ and $\hat{r}$.
- $t_e$ the mean of the estimation time for one simulation.

5.3 Results

For Models 1-3, for $\Delta$ small enough ($\Delta = 10^{-2}$ or $10^{-3}$ for Model 1, $\Delta = 10^{-3}$ for Models 2 and 3), the risk of the adaptive estimator $\hat{\sigma}^2_{m, \hat{r}}$ is inversely proportional to $n$, that is proportional to the variance term. In Table 5.3 we can see that the risk mostly depends on $\Delta$: the remainder term is predominant. As the Blumethal-Getoor index $\beta > 1$, this is consistent with Remark (9). We can see in Figure 4 that $\sigma^2$ is overestimated: this is because the small jumps can not be cut. This bias decreases with $\Delta$.

The function $g = \sigma^2 + \xi^2$ is more difficult to estimate. Indeed, the variance term is bigger (it is proportional to $1/n\Delta$ and not $1/n$). For $n\Delta$ not big enough ($n\Delta = 1$ or 10), the results can be quite bad. When $\Delta$ is fixed (and small enough so that the remainder term is not preponderant), the risk decreases when $n$ increases.

6 Proofs

6.1 Proof of Theorem 4

By Lemma 2 $\mathbb{E}(A^2_{k\Delta}) \lesssim \Delta$. It remains to bound $\mathbb{E}(\sup_{t \leq T} \nu^2_{\varphi_{\lambda}}(t))$. Let $(\varphi_{\lambda})_{1 \leq \lambda \leq D_m}$ be an orthonormal (for the $\| \cdot \|_\pi$ norm) basis of $S_m$. Any function
Figure 1: Model 1

\[ dX_t = -2X_t dt + dW_t + dL_t, \]  
binomial jumps

Estimation of \( \sigma^2 \)

Estimation of \( \sigma^2 + \xi^2 \)

\( n = 10^5, \Delta = 10^{-3} \)

Figure 2: Model 2

\[ dX_t = -2X_t dt + \frac{X_t^2}{X_t^2 + 1} dW_t + dL_t, \]  
Laplace jumps

Estimation of \( \sigma^2 \)

Estimation of \( \sigma^2 + \xi^2 \)

\( n = 10^5, \Delta = 0.1 \)
Figure 3: Model 3

\[ dX_t = (-2X_t + \sin(3X_t))dt + \sqrt{2 + 0.5 \sin(\pi X_t)}(dW_t + dL_t), \quad \text{Normal jumps} \]

\[
\begin{align*}
\text{Estimation of } \sigma^2 & \\
\text{Estimation of } \sigma^2 + \xi^2 &
\end{align*}
\]

![Graph of Model 3](image)

\[-: \text{true function} \quad -. : \text{estimator} \]
\[n = 10^5, \Delta = 10^{-3} \]

Figure 4: Model 4

\[ n(z) = \sum_{k=1}^{\infty} 2^{k+1}(\delta_{1/2^k} + \delta_{-1/2^k}), \quad b(x) = -2x \quad \text{and} \quad \sigma(x) = \xi(x) = 1. \]

\[
\begin{align*}
\text{Estimation of } \sigma^2 & \\
\text{Estimation of } \sigma^2 + \xi^2 &
\end{align*}
\]

![Graph of Model 4](image)

\[-: \text{true function} \quad -. : \text{estimator} \]
\[n = 10^5, \Delta = 10^{-4} \]

\[-: \text{true function} \quad -. : \text{estimator} \]
\[n = 10^5, \Delta = 10^{-2} \]
Table 1: Model 1

\[ dX_t = -2X_t dt + dW_t + dL_t, \text{ binomial jumps} \]

| \( \Delta \) | \( n \) | risk | oracle | \( m_{est} \) | \( r_{est} \) | \( t_e \) | risk | oracle | \( m_{est} \) | \( r_{est} \) | \( t_e \) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| \( 10^{-1} \) | \( 10^4 \) | 0.075 | 1.00 | 0.00 | 0.92 | 0.78 | 0.93 | 1.56 | 1.92 | 0.92 | 0.78 |
| \( 10^{-1} \) | \( 10^4 \) | 0.061 | 1.03 | 0.02 | 1.30 | 3.61 | 0.63 | 1.12 | 3.44 | 1.30 | 3.61 |
| \( 10^{-2} \) | \( 10^4 \) | 0.066 | 1.14 | 0.16 | 1.46 | 36 | 0.59 | 1.03 | 3.46 | 1.46 | 36 |
| \( 10^{-2} \) | \( 10^4 \) | 0.15 | 1.00 | 0.00 | 0.00 | 0.22 | 0.0026 | 1.71 | 0.02 | 0.00 | 0.22 |
| \( 10^{-2} \) | \( 10^5 \) | 0.015 | 1.00 | 0.00 | 0.00 | 3.60 | 0.0004 | 3.27 | 0.02 | 0.00 | 3.60 |
| \( 10^{-3} \) | \( 10^4 \) | 0.0021 | 1.00 | 0.00 | 0.52 | 36 | 0.00048 | 4.70 | 0.12 | 0.52 | 36 |
| \( 10^{-3} \) | \( 10^4 \) | 4.18 | 1.21 | 0.02 | 0.00 | 0.13 | 0.0020 | 1.00 | 0.00 | 0.00 | 0.13 |
| \( 10^{-3} \) | \( 10^5 \) | 0.12 | 1.00 | 0.00 | 0.00 | 0.58 | 0.0002 | 1.00 | 0.00 | 0.00 | 0.58 |
| \( 10^{-3} \) | \( 10^5 \) | 0.013 | 1.00 | 0.00 | 0.02 | 36 | 0.000022 | 1.57 | 0.00 | 0.02 | 36 |

Table 2: Model 2

\[ dX_t = -2X_t dt + \frac{X_{t-}^2}{X_{t-}^2 + 1} dW_t + dL_t, \text{ Laplace jumps} \]

| \( \Delta \) | \( n \) | risk | oracle | \( m_{est} \) | \( r_{est} \) | \( t_e \) | risk | oracle | \( m_{est} \) | \( r_{est} \) | \( t_e \) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| \( 10^{-1} \) | \( 10^4 \) | 3.53 | 2.46 | 0.00 | 0.02 | 0.78 | 3.02 | 3.04 | 0.00 | 0.02 | 0.78 |
| \( 10^{-1} \) | \( 10^4 \) | 3.07 | 2.05 | 0.00 | 0.42 | 2.54 | 2.23 | 1.68 | 0.14 | 0.42 | 2.54 |
| \( 10^{-1} \) | \( 10^4 \) | 1.51 | 1.01 | 0.52 | 1.58 | 20.4 | 1.26 | 1.01 | 0.46 | 1.58 | 20.4 |
| \( 10^{-2} \) | \( 10^4 \) | 152 | 5.00 | 0.20 | 0.08 | 0.23 | 2.81 | 10.5 | 0.02 | 0.08 | 0.23 |
| \( 10^{-2} \) | \( 10^4 \) | 3.37 | 5.69 | 0.00 | 1.28 | 2.56 | 0.28 | 1.25 | 0.72 | 1.28 | 2.56 |
| \( 10^{-2} \) | \( 10^4 \) | 1.36 | 1.39 | 0.54 | 1.08 | 20.3 | 0.22 | 1.05 | 1.00 | 1.08 | 20.3 |
| \( 10^{-3} \) | \( 10^4 \) | 1600 | 2.87 | 0.02 | 0.20 | 0.14 | 2.34 | 10.2 | 0.16 | 0.20 | 0.14 |
| \( 10^{-3} \) | \( 10^4 \) | 85 | 3.60 | 0.10 | 1.16 | 0.56 | 0.087 | 1.47 | 0.84 | 1.16 | 0.56 |
| \( 10^{-3} \) | \( 10^5 \) | 4.90 | 6.58 | 0.00 | 1.00 | 20.3 | 0.023 | 3.23 | 1.00 | 1.00 | 20.3 |

15
Table 3: Model 3

\[ dX_t = (-2X_t + \sin(3X_t))dt + \sqrt{2 + 0.5 \sin(\pi X_t)}(dW_t + dL_t), \]

Normal jumps

| $\Delta$ | $n$  | risk | oracle | $m_{est}$ | $r_{est}$ | $t_e$ | risk | oracle | $m_{est}$ | $r_{est}$ | $t_e$ |
|---------|------|------|--------|--------|--------|------|------|--------|--------|--------|------|
| $10^{-1}$ | $10^4$ | 1.00 | 1.86 | 0.02 | 1.10 | 0.81 | 15.4 | 7.15 | 4.52 | 1.10 | 0.80 |
| $10^{-1}$ | $10^4$ | 0.56 | 1.27 | 0.48 | 1.20 | 3.43 | 3.43 | 1.71 | 5.72 | 1.20 | 3.39 |
| $10^{-1}$ | $10^4$ | 0.43 | 1.03 | 0.90 | 1.00 | 31.3 | 2.09 | 1.08 | 6.24 | 1.00 | 31.2 |
| $10^{-1}$ | $10^4$ | 24.4 | 28.5 | 0.32 | 0.62 | 0.24 | 2.49 | 8.59 | 1.66 | 0.62 | 0.24 |
| $10^{-1}$ | $10^4$ | 0.78 | 2.57 | 0.12 | 1.30 | 3.41 | 1.46 | 4.34 | 4.88 | 1.30 | 3.38 |
| $10^{-1}$ | $10^4$ | 0.12 | 3.30 | 0.82 | 1.10 | 31.1 | 0.75 | 1.54 | 6.98 | 1.10 | 31.0 |
| $10^{-2}$ | $10^4$ | 82.3 | 3.57 | 0.08 | 0.08 | 0.14 | 0.090 | 5.43 | 0.12 | 0.08 | 0.14 |
| $10^{-2}$ | $10^4$ | 13.1 | 2.51 | 0.18 | 1.14 | 0.60 | 0.019 | 4.61 | 0.80 | 1.14 | 0.61 |
| $10^{-3}$ | $10^4$ | 0.98 | 2.63 | 0.26 | 2.18 | 31.0 | 0.0026 | 1.16 | 0.82 | 2.18 | 30.8 |

Table 4: Model 4

\[ n(z) = \sum_{k=1}^{\infty} 2^{k+1}(\delta_{1/2^k} + \delta_{-1/2^k}), \quad b(x) = -2x \quad \text{and} \quad \sigma(x) = \xi(x) = 1. \]

| $\Delta$ | $n$  | risk | oracle | $m_{est}$ | $r_{est}$ | $t_e$ | risk | oracle | $m_{est}$ | $r_{est}$ | $t_e$ |
|---------|------|------|--------|--------|--------|------|------|--------|--------|--------|------|
| $10^{-1}$ | $10^4$ | 0.074 | 1.00 | 0.00 | 0.14 | 0.86 | 0.56 | 1.02 | 0.06 | 0.14 | 0.89 |
| $10^{-1}$ | $10^4$ | 0.075 | 1.01 | 0.02 | 1.28 | 4.04 | 0.54 | 1.02 | 0.12 | 1.28 | 4.01 |
| $10^{-1}$ | $10^4$ | 0.080 | 1.02 | 0.12 | 1.98 | 37.4 | 0.55 | 1.02 | 0.12 | 1.98 | 37.3 |
| $10^{-1}$ | $10^4$ | 0.039 | 1.00 | 0.00 | 0.42 | 0.25 | 0.96 | 1.19 | 0.70 | 0.42 | 0.25 |
| $10^{-1}$ | $10^4$ | 0.0040 | 1.00 | 0.00 | 0.62 | 4.57 | 0.86 | 1.01 | 0.72 | 0.62 | 4.58 |
| $10^{-2}$ | $10^4$ | 0.0012 | 1.00 | 0.00 | 0.58 | 38.3 | 0.91 | 1.00 | 1.24 | 0.58 | 38.1 |
| $10^{-3}$ | $10^4$ | 1.22 | 1258 | 0.04 | 1.02 | 0.14 | 0.071 | 1.07 | 0.04 | 0.02 | 0.15 |
| $10^{-3}$ | $10^4$ | 0.012 | 1.00 | 0.00 | 0.10 | 0.87 | 0.094 | 1.01 | 0.02 | 0.10 | 0.87 |
| $10^{-3}$ | $10^4$ | 0.0015 | 1.00 | 0.00 | 0.36 | 38.5 | 0.15 | 1.00 | 0.22 | 0.36 | 38.5 |
| $10^{-2}$ | $10^4$ | 0.24 | 1.00 | 0.00 | 0.02 | 0.39 | 0.013 | 1.04 | 0.02 | 0.02 | 0.39 |
| $10^{-2}$ | $10^4$ | 0.021 | 1.00 | 0.00 | 0.04 | 6.31 | 0.014 | 1.00 | 0.00 | 0.04 | 6.26 |
Lemma 2. We can construct independent variables and have same law as

\[ E \{ s \leq 1 \} \]

Then

\[ E \{ B_{kA} + E_{kA} \mid \varphi_{kA} \} = 0, \]

By Lemmas 2 and 3 we obtain:

\[ \sup_{t} \nu_{n}^{2}(t) = \sup_{\sum a_{i} \leq 1} \left( \sum_{\lambda=1}^{D_{m}} a_{\lambda} \varphi_{\lambda} \right) \leq \sum_{\lambda=1}^{D_{m}} \nu_{n}^{2}(\varphi_{\lambda}). \]

According to Lemma 2, we obtain:

\[ \mathbb{E}(\nu_{n}^{2}(\varphi_{\lambda})) = \mathbb{E} \left( \left( \frac{1}{n} \sum_{k=1}^{n} (B_{kA} + E_{kA}) \varphi_{kA}(X_{kA}) \right)^{2} \right) \]

\[ \leq \frac{2}{n^{2}} \sum_{k=1}^{n} \mathbb{E} \left[ \varphi_{kA}^{2}(X_{kA}) \mathbb{E} (B_{kA}^{2} + E_{kA}^{2} \mid \varphi_{kA}) \right] \]

\[ \lesssim \frac{c_{1}^{4}}{n^{4}} + \frac{\sigma_{0}^{4} + \sigma_{0}^{2} \xi_{0}^{2}}{n}. \]

Then

\[ \mathbb{E} \left( \sup_{t \in \mathcal{B}_{m}} \nu_{n}^{2}(t) \right) \lesssim \left( \frac{c_{1}^{4}}{n^{4}} + \frac{\sigma_{0}^{4} + \sigma_{0}^{2} \xi_{0}^{2}}{n} \right) \frac{D_{m}}{n}. \]  

(2)

It remains to bound the risk on \( \Omega_{n}^{c} \). By Lemma 3, \( \mathbb{P}(\Omega_{n}^{c}) \leq 1/n^{8} \). The function \( \tilde{g}_{m} \) is the orthogonal projection (for the \( \| \|_{2} \) norm) of \( (T_{\Delta}, \ldots, T_{n\Delta}) \) on the vectorial subspace \( \{ \{t(X_{kA}), \ldots, t(X_{nA})\}, \ t \in \mathcal{S}_{m} \} \). Let us denote by \( \Pi_{m} \) the orthogonal projection of this subspace. As \( T_{k\Delta} = g(X_{kA}) + A_{kA} + B_{kA} + E_{kA} \), we obtain:

\[ \| \tilde{g}_{m} - gA \|_{2}^{2} = \| \Pi_{m} T - gA \|_{n}^{2} = \| \Pi_{m} gA - gA \|_{2}^{2} + \| \Pi_{m} A + \Pi_{m} B + \Pi_{m} E \|_{2}^{2} \]

\[ \leq \| gA \|_{2}^{2} + \| A + B + E \|_{2}^{2} \]

By stationarity and Cauchy-Schwarz:

\[ \mathbb{E}(\| \tilde{g}_{m} - gA \|_{2}^{2} \mathbb{1}_{\Omega_{n}^{c}}) \lesssim \mathbb{E}(\| gA \|_{2}^{2} \mathbb{1}_{\Omega_{n}^{c}}) + \mathbb{E} \left[ \left( \frac{1}{n} \sum_{k=1}^{n} A_{k\Delta}^{2} + B_{k\Delta}^{2} + E_{k\Delta}^{2} \right) \mathbb{1}_{\Omega_{n}^{c}} \right] \]

\[ \lesssim \left[ \mathbb{E}(\| gA \|_{2}^{4}) + \mathbb{E} \left[ A_{k\Delta}^{4} + B_{k\Delta}^{4} + E_{k\Delta}^{4} \right] \right] \mathbb{P}(\Omega_{n}^{c}) \]

By Lemmas 2 and 3 we obtain:

\[ \mathbb{E}(\| \tilde{g}_{m} - gA \|_{2}^{2} \mathbb{1}_{\Omega_{n}^{c}}) \lesssim \frac{1}{\Delta^{1/2} n^{4}} \lesssim \frac{1}{\Delta^{1/2} n^{4}} \]

6.2 Proof of Theorem 5

First, we apply the Berbee’s coupling lemma to the random vectors \( (B_{kA} + E_{kA}, X_{kA}) \) which are exponentially \( \beta \)-mixing. According to Berbee’s coupling lemma, we can construct independent variables

\[ U_{k,a}^{*} = \frac{1}{q_{n}} \sum_{l=1}^{q_{n}} (B_{k} + E)_{(2(k-1)+a)}^{*} l(t_{X_{kA}^{*}}(2(k-1)+a)q_{n}+l) \Delta \]

such that for \( a \in \{0, 1\} \), the random variables \( (U_{k,a})_{0 \leq k < n} \) are independent and have same law as

\[ U_{k,a} = \frac{1}{q_{n}} \sum_{l=1}^{q_{n}} (B_{k} + E)_{(2(k-1)+a)} l(t_{X_{kA}}(2(k-1)+a)q_{n}+l) \Delta. \]
Let us set 
\[ \Omega^* = \{ \omega, \forall a, \forall k, U_{k,a} = U_{k,a}^* \} , \]
\[ \Omega_{B,a} = \{ \omega, \forall (a,k), |U_{k,a}^*| \leq c \ln^2 (n) D^{1/2} \Delta^{-\alpha} \} , \]
and 
\[ \mathcal{E} = \Omega_{n} \cap \Omega_{B,a} \cap \Omega^* \]
with \( D = D_m + D_{m'} \). By Berbee’s coupling lemma,
\[ \mathbb{P} (\Omega^*) \lesssim n \Delta / n^8 . \]
The following lemma is proved later.

**Lemma 12.** For any \( \alpha > 0 \), there exists a constant \( c \) such that
\[ \mathbb{P} (\Omega_{B,a}^*) \lesssim \frac{1}{n^5} . \]

Then
\[ \mathbb{P} (\mathcal{E}^c) \leq \mathbb{P} (\Omega_{B,a}^c) + \mathbb{P} (\Omega^c) + \mathbb{P} (\Omega_{n}^c) \lesssim \frac{1}{n^5} + \frac{n \Delta}{n^8} + \frac{1}{n^8} \lesssim \frac{1}{n^5} . \]

We can bound \( \mathbb{E} \left( \| \hat{g}_m - g \|_n^2 \mathbb{1}_{\mathcal{E}^c} \right) \) in the same way as we bound the risk of the non-adaptive estimator on \( \Omega_{n}^c \):
\[ \mathbb{E} \left( \| \hat{g}_m - g \|_n^2 \mathbb{1}_{\mathcal{E}^c} \right) \lesssim \frac{1}{\sqrt{n} / 2} \lesssim \frac{1}{n} . \]

It remains to bound the risk on \( \mathcal{E} \). Let us set, for \( a \in \{ 0, 1 \} \),
\[ \nu_{n,a}^*(t) = \frac{1}{p_n} \sum_{k=1}^{p_n} U_{k,a}^* \mathbb{1}_{\mathcal{E}} - \mathbb{E} (U_{k,a}^* \mathbb{1}_{\mathcal{E}}) \]
and \( \nu_{n,a}^*(t) = \nu_{n,0}^*(t) + \nu_{n,1}^*(t) \). We have:
\[ \| \hat{g}_m - g \|_n^2 \mathbb{1}_{\mathcal{E}} \lesssim \frac{1}{n} \sum_{k=1}^{p_n} A_{k,\Delta}^2 + (\mathbb{E} \left[ (B_{k,\Delta}^* + E_{k,\Delta}^* \mathbb{1}_{\mathcal{E}}) \right]^2 + \sup_{t \in \mathcal{F}_{m,m'}} (\nu_{n,a}^*(t))^2 + 2pen (m) . \]

As the random variables \( B_{k,\Delta}^* \) and \( E_{k,\Delta}^* \) are centred,
\[ R_1 := \mathbb{E} \left[ (B_{k,\Delta}^* + E_{k,\Delta}^* \mathbb{1}_{\mathcal{E}}) \right] = - \mathbb{E} \left[ (B_{k,\Delta}^* + E_{k,\Delta}^* \mathbb{1}_{\mathcal{E}}) \right] \]
then by Lemma 2
\[ |R_1| \lesssim \left( \mathbb{E} \left[ (B_{k,\Delta}^* + E_{k,\Delta}^*)^2 \right] \mathbb{P} (\mathcal{E}^c) \right)^{1/2} \lesssim n^{-5/2} \Delta^{-1/2} . \]

Then
\[ \mathbb{E} \left( \| \hat{g}_m - g \|_n^2 \mathbb{1}_{\mathcal{E}} \right) \lesssim \| g_m - g \|_2^2 + \Delta + \frac{1}{n} + 2pen (m) \]
\[ + 12 \mathbb{E} \left[ \sum_{m' \in \mathcal{F}_n} \left( \left( \sup_{t \in \mathcal{F}_{m,m'}} (\nu_{n,a}^*(t))^2 - p(m,m') \right) \right) \mathbb{1}_{\mathcal{E}} \right] \]

18
The functions $\nu_{n,\alpha}(t)$ satisfy the assumptions of Talagrand’s inequality with $M = c \ln^2(n) D^{1/2} \Delta^{-\alpha}$, $V = \frac{1}{q_0} \Delta$, and $H^2 = c' D/(n \Delta)$. Then

$$R_2 := E \left( \left[ \sup_{t \in \mathcal{I}_{m,m'}} (\nu_{n,\alpha}(t))^2 - 12p(m, m') \right] \right) \leq \frac{1}{n \Delta} \exp \left( -c' \frac{p_n \Delta}{n \Delta} D \right) + \frac{c^2 \ln^4(n) D \Delta^{-2\alpha}}{p_n^2} \exp \left( -\frac{c' \sqrt{\Delta \Delta^{-\alpha} D^{1/2} \ln^2(n)}}{\sqrt{n}} \right) \leq \frac{1}{n \Delta} \exp (-cD) + \frac{\ln^6(n)}{n \Delta^2 e^{2\alpha}} \exp \left( -c' \frac{\sqrt{\Delta \Delta^{-\alpha} D^{1/2} \ln^2(n)}}{\ln^2(n)} \right).$$

Consequently, as $\alpha$ is as small as we want:

$$E \left[ \left[ \sup_{t \in \mathcal{I}_{m,m'}} \nu_{n,\alpha}^2(t) - p \ln(m) \right] \right] \leq \frac{1}{n \Delta} \sum_{m'} e^{-cDm,m'} \leq \frac{1}{n \Delta}.$$

**6.3 Proof of Lemma 12**

We have that

$$U_{1,0}^* = \frac{1}{q_0} \sum_{j=1}^{q_0} (B_{j,\Delta}^* + E_{j,\Delta}^*) t(X_{j,\Delta}) \leq \frac{1}{q_0} \sum_{j=1}^{q_0} (J_{j,\Delta}^2 + Z_{j,\Delta}^2) t(X_{j,\Delta}).$$

We know that $|t(X_{j,\Delta})| \leq \|t\|_{\infty} \leq D^{1/2}$. Moreover,

$$P \left( |Z_{k,\Delta}| \geq k \sigma_0 \Delta^{1/2} \ln(n) \right) \leq n^{-k} E \left[ \exp \left( \frac{1}{\sigma_0 \Delta^{1/2} Z_{k,\Delta}} \right) \right] \leq n^{-k} E \left[ \exp \left( \frac{1}{\sigma_0 \Delta} \int_{k \Delta}^{(k+1) \Delta} \sigma^2(X_s) ds \right) \right] \leq n^{-k}. \quad (3)$$

Then

$$P \left( |Z_{k,\Delta}| \geq 6 \sigma_0^2 \Delta \ln(n) \right) \leq n^{-6}. \quad (4)$$

and then

$$\sum_{k=1}^{p_n} \left( \frac{1}{q_0} \sum_{j=1}^{q_0} Z_{j,\Delta}^2 \geq 36 \sigma_0^4 \Delta \ln^2(n) \right) \leq n^{-5}.$$

**Bound of $P \left( |J_{k,\Delta}^{(1)}| \geq 12 \sigma_0 \Delta^{1/2} \ln(n) \right).$**

The terms $J_{k,\Delta}^{(1)}$ are small and can be bounded in the same way as the Brownian terms $Z_{k,\Delta}$. As $\nu$ is symmetric:

$$P \left( |J_{k,\Delta}^{(1)}| \geq 12 \sigma_0 \Delta^{1/2} \ln(n) \right) \leq 2P \left( \exp \left( a J_{k,\Delta}^{(1)} \right) \geq \exp \left( 12 \sigma_0 \Delta^{1/2} \ln(n) \right) \right) \leq 2 \exp \left( -12 \sigma_0 \Delta^{1/2} \ln(n) \right) E \left( \exp \left( a J_{k,\Delta}^{(1)} \right) \right).$$
According to Corollary 5.2.2 of Applebaum [2004],
\[
E \left( \exp \left( a J_{k \Delta}^{(1)} \right) \right) = E \left( \exp \left( \int_{k \Delta}^{(k+1) \Delta} \int_{-\Delta^{1/2}}^{\Delta^{1/2}} e^{a \xi(X_s - z) - 1 - a \xi(X_s - z) + \nu(dz)ds} \right) \right).
\]
Then for any \( a \leq 1/(2 \xi_0 \Delta^{1/2}) \),
\[
E \left( \exp \left( a J_{k \Delta}^{(1)} \right) \right) \leq E \left( \exp \left( \int_{k \Delta}^{(k+1) \Delta} \int_{-\Delta^{1/2}}^{\Delta^{1/2}} a^2 \xi^2(X_s - z) + \nu(dz) \right) \right)
\leq E \left( \exp \left( \xi_0^2 a^2 \Delta^{2-\beta/2} \right) \right).
\]
Let us then set \( a = 1/(2 \xi_0 \Delta^{1/2}) \), we obtain:
\[
\mathbb{P} \left( \left| J_{k \Delta}^{(1)} \right| \geq 12 \xi_0 \Delta^{1/2} \ln(n) \right) \lesssim \exp \left( -6 \ln(n) \right) \leq n^{-6}.
\]

**Bound for the jumps greater than \( \Delta^{1/2} \).**

The probability that \( J_{k \Delta}^{(2)} + J_{k \Delta}^{(3)} \geq \Delta^{1/2} \ln(n) \)
is not small enough. We have to bound both the number of jumps of the time interval \([k \Delta, (k+1) \Delta]\) and the size of the jumps. Let us first consider the jumps greater than 1:
\[
J_{k \Delta}^{(0)} = \int_{k \Delta}^{(k+1) \Delta} \xi(X_s - z) \mu(dz, ds).
\]
The probability of having a very high jump is quite small: by Assumption A4,
\[
\nu \left( \left[ -8 \ln(n), \frac{8 \ln(n)}{\lambda} \right] \right) \lesssim n^{-8}.
\]
The probability of having more than \( C = 8 \eta / (1 - \beta/2) \) (see Assumption A4) jumps greater than 1 on a time interval \( \Delta \) is very low:
\[
Q_1 := \mathbb{P} \left( \mu \left( \left[ [k \Delta, (k+1) \Delta], [-1, 1] \right] \right) \geq C \right)
\leq \mathbb{P} \left( \mu \left( \left[ [k \Delta, (k+1) \Delta], [-\Delta^{1/2}, \Delta^{1/2}] \right] \right) \geq C \right)
\leq \left( \Delta \int_{|z|>\Delta^{1/2}} \nu(dz) \right)^C \lesssim \left( \Delta^{1-\beta/2} \right)^C = \Delta^{8g} \lesssim n^{-8}.
\]
By (6) and (7),
\[
\mathbb{P} \left( \left| J_{k \Delta}^{(0)} \right| \geq \frac{8 C \ln(n)}{\lambda} \right) \lesssim n^{-8}.
\]
Let us set \( v_1 = \nu([-1, 1]^c] \lor 1 \). We have that
\[
Q_2 := \mathbb{P} \left( \frac{1}{q_n} \sum_{k=1}^{q_n} |J_{k}^{0}\alpha| \geq \frac{8G^2}{\lambda^2}v_1\Delta\ln^2(n) \right)
\]
\[
\lesssim q_n\mathbb{P} \left( |J_{k}^{0}\alpha| \geq \frac{8G\ln(n)}{\lambda} \right) + \mathbb{P} \left[ \mu([0, q_n\Delta[-1, 1]^c] \geq v_1\Delta q_n \right]
\]
\[
\lesssim q_n + \sum_{j=v_1\Delta q_n}^{+\infty} \frac{(q_n\Delta v_1)^j}{j!} e^{-q_n\Delta v_1}
\]
\[
\lesssim q_n + \sum_{j=v_1\Delta q_n}^{+\infty} \left( \frac{8\ln(n)v_1e}{j} \right)^j \sqrt{j}e^{-8\ln(n)v_1}
\]
\[
\lesssim q_n + \frac{1}{n}. \quad (8)
\]
Let us now set \( \alpha_0 = 0, \alpha_j = \frac{2\alpha_{j-1}+\alpha}{2} \land \frac{1}{2} \) and
\[
J_{k}^{(\alpha_1)} = \int_{\Delta}^{(k+1)\Delta} \int_{[-\Delta^{\alpha_j-1}, -\Delta^{\alpha_j}] \cup [\Delta^{\alpha_j}, \Delta^{\alpha_j-1}} \xi(X_s)\,dL_s.
\]
By (7),
\[
\mathbb{P} \left( \left| J_{k}^{(\alpha_1)} \right| \geq C\Delta^{\alpha_j-1} \right) \lesssim \frac{1}{n^8}.
\]
We have that \( \nu([-\Delta^{\alpha_j-1}, -\Delta^{\alpha_j}] \cup [\Delta^{\alpha_j}, \Delta^{\alpha_j-1}] \lesssim \Delta^{-\alpha_j}. \) Let us set \( v_2 = \Delta^{\beta \alpha_j} (\nu([-\Delta^{\alpha_j-1}, -\Delta^{\alpha_j}] \cup [\Delta^{\alpha_j}, \Delta^{\alpha_j-1}] \lor 1). \) Then
\[
Q_3 := \mathbb{P} \left( \frac{1}{q_n} \sum_{k=1}^{q_n} \left| J_{k}^{(\alpha_j)} \right| \geq C\Delta^{1-\alpha} \right)
\]
\[
\lesssim \mathbb{P} \left[ \mu([0, q_n\Delta], [-\Delta^{\alpha_j-1}, -\Delta^{\alpha_j}] \cup [\Delta^{\alpha_j}, \Delta^{\alpha_j-1}] \geq q_n\Delta^{1-2\alpha_j-1-\alpha} \right]
\]
\[
+ q_n\mathbb{P} \left( \left| J_{k}^{(\alpha_j)} \right| \geq C\Delta^{\alpha_j-1} \right)
\]
\[
\lesssim q_n + \sum_{i=q_n\Delta^{1-2\alpha_j-1-\alpha}}^{+\infty} \frac{(q_n\Delta v_2 \Delta^{-\beta \alpha_j})^i}{i!} e^{-q_n\Delta v_2 \Delta^{-\beta \alpha_j}}
\]
\[
\lesssim q_n + \frac{1}{n^8}. \quad (9)
\]
Then, by (3), (5), (6) and (8), we obtain:
\[
\mathbb{P} \left( \Omega_{B, \alpha} \right) \lesssim \frac{p_nq_n}{n^6} + \frac{p_nq_n}{n^8} \lesssim n^{-5}.
\]

6.4 Proof of Lemma 6

**Bound of** \( \mathbb{P} \left( \Omega_{X,k}^{\xi} \right). \) We have that \( X_{(k+1)\Delta} = X_{k\Delta} + \int_{k\Delta}^{(k+1)\Delta} b(X_s)\,ds + Z_{k\Delta} + J_{k\Delta}. \) Then
\[
\mathbb{P} \left( \Omega_{X,k}^{\xi} \right) \leq \mathbb{P} \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s)\,ds \geq \Delta^{1/2} \right) + \mathbb{P} \left( |Z_{k\Delta}| \geq \sigma_0\Delta^{1/2} \ln(n) \right)
\]
\[
+ \mathbb{P} \left( |J_{k\Delta}| \geq \xi_0\Delta^{1/2} \ln(n) \right).
\]
By Markov’s inequalities, for any $k \leq 4$:
\[
P \left( \left| \int_{k\Delta}^{(k+1)\Delta} b(X_s) ds \right| \geq \Delta^{1/2} \right) \lesssim \Delta^{-k} \mathbb{E} \left[ \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s) ds \right)^{2k} \right] \lesssim \Delta^k
\]
and by (3), \( \mathbb{P} \left( |Z_{k\Delta}| \geq k\sigma_0 \Delta^{1/2} \ln(n) \right) \lesssim n^{-k} \). Moreover,
\[
P \left( \left| J_{k\Delta}^{(2)} + J_{k\Delta}^{(3)} \right| > 0 \right) \leq \Delta \int_{[-\Delta^{1/2}, \Delta^{1/2}] \nu(dz) \leq \Delta^{1-\beta/2} \int_{[-\Delta^{1/2}, \Delta^{1/2}]^c} z^2 \nu(dz) 
\]
\[
\lesssim \Delta^{1-\beta/2}
\]
and by Markov’s inequality:
\[
P \left( \left| J_{k\Delta}^{(1)} \right| > \xi_0 \Delta^{1/2} \right) \leq \frac{1}{\xi_0 \Delta^{1/2}} \mathbb{E} \left[ \left( J_{k\Delta}^{(1)} \right)^2 \right] 
\]
\[
\leq \frac{1}{\xi_0 \Delta^{1/2}} \Delta \xi_0^2 \int_{-\Delta^{1/2}}^{\Delta^{1/2}} z^2 \nu(dz) \lesssim \Delta^{1-\beta/2} \int_{-\Delta^{1/2}}^{\Delta^{1/2}} z^2 \nu(dz) 
\]
\[
\lesssim \Delta^{1-\beta/2}.
\]

**Bound of \( \mathbb{P} \left( \Omega^c_{N,k} \right) \).**

We have that
\[
P (N_k \geq 1) = \int_{k\Delta}^{(k+1)\Delta} \int_{|z| \geq \Delta^{1/4}} \nu(dz) \leq \Delta^{1-\beta/4} \int_{|z| \geq \Delta^{1/4}} z^2 \nu(dz) \lesssim \Delta^{1-\beta/4}.
\]
Then by (11) and (12), we obtain:
\[
P \left( \Omega^c_{N,k} \right) \leq \mathbb{P} (N_k \geq 1) + \mathbb{P} \left( \left| J_{k\Delta}^{(2)} + J_{k\Delta}^{(3)} \right| > 0 \right) + \mathbb{P} \left( \left| \int_{k\Delta}^{(k+1)\Delta} dL_s^{(1)} \right| \geq \ln(n) \Delta^{1/2} \right) 
\]
\[
\lesssim \Delta^{1-\beta/2}.
\]

**Bound of \( \mathbb{P} \left( \Omega_{X,k} \cap \Omega^c_{N,k} \right) \).**

We have that
\[
P \left( \Omega_{X,k} \cap \{ N_k \geq 1 \} \right) \leq \mathbb{P} (N_k \geq 2) + \mathbb{P} \left( \Omega_{X,k} \cap \{ N_k = 1 \} \right).
\]
Now \( \mathbb{P} (N_k \geq 2) \leq \left( \Delta^{1-\beta/4} \int_{|z| \geq \Delta^{1/4}} z^2 \nu(dz) \right)^2 \lesssim \Delta^{2-\beta/2} \). Moreover, if \( N_k = 1 \), then \( \left| J_{k\Delta}^{(3)} \right| \geq \xi_1 \Delta^{1/4} \) and by conditional independence, we get:
\[
S_1 := \mathbb{P} \left( \Omega_{X,k} \cap \{ N_k = 1 \} \right) 
\]
\[
\leq \mathbb{P} (N_k = 1) \times \mathbb{P} \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s) ds + Z_{k\Delta} + J_{k\Delta}^{(1)} + J_{k\Delta}^{(2)} \leq \xi_1 \Delta^{1/4} \right) 
\]
\[
\leq \mathbb{P} (N_k = 1) \left[ \mathbb{P} \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s) ds \geq \frac{\xi_1 \Delta^{1/4}}{3} \right) + \mathbb{P} \left( |Z_{k\Delta}| \geq \frac{\xi_1 \Delta^{1/4}}{3} \right) \right] 
\]
\[
+ \mathbb{P} (N_k = 1) \mathbb{P} \left( \left| J_{k\Delta}^{(1)} + J_{k\Delta}^{(2)} \right| \geq \frac{\xi_1 \Delta^{1/4}}{3} \right) .
\]
By (11) and (3), $P\left( \left| f_{k\Delta}^{(k+1)\Delta} b(X_s)ds \right| \geq c\Delta^{1/4} \right) \leq \Delta^4$ and, as $\ln(n) \ll \Delta^{-1/4}$, $P\left( |Z_{k\Delta}| \geq c\Delta^{1/4} \right) \leq P\left( |Z_{k\Delta}| \geq \ln(n)\Delta^{1/2} \right) \lesssim n^{-1}$. Moreover, by a Markov inequality, we obtain:

$$P\left( |J_{k\Delta}^{(1)} + J_{k\Delta}^{(2)}| > c\Delta^{1/4} \right) \leq e^{-2\Delta^{1/2}/2} \mathbb{E}\left[ \left( J_{k\Delta}^{(1)} + J_{k\Delta}^{(2)} \right)^2 \right] \leq e^{-2\Delta^{1/2}/2} \xi_0 \Delta^{1/4} \int_{-\Delta^{1/4}}^{\Delta^{1/4}} z^2 \nu(dz) \lesssim \Delta^{1/2}\Delta^{1/2-\beta/4}. $$

As $P(N_k = 1) \lesssim \Delta^{1-\beta/4}$, we obtain:

$$P(\Omega_{X,k} \cap \{N_k \geq 1\}) \lesssim \Delta^{2-\beta/2}. $$

Let us set $L_s^{(1)+(2)} = L_s^{(1)} + L_s^{(2)}$ and $J_{k\Delta}^{(1)+(2)} = J_{k\Delta}^{(1)} + J_{k\Delta}^{(2)}$. We consider

$$\mathcal{E}_k = \left\{ \left| \int_{k\Delta}^{(k+1)\Delta} dL_s^{(1)+(2)} \right| \leq 4 \frac{\xi_0 + \sigma_0}{\xi_1^{1/2}} \Delta^{1/2} \ln(n) \right\}. $$

We have that

$$\mathcal{E}_k^c \subseteq \left\{ \left| \int_{k\Delta}^{(k+1)\Delta} dL_s^{(1)+(2)} \right| \geq 4(\xi_0 + \sigma_0) \Delta^{1/2} \ln(n) \right\} \cup \left\{ \left| J_{k\Delta}^{(1)+(2)} \right| \geq 2(\xi_0 + \sigma_0) \Delta^{1/2} \ln(n) \right\}. $$

By (10) and (3),

$$S_2 := P\left( \Omega_{X,k} \cap \left\{ \left| J_{k\Delta}^{(1)+(2)} \right| \geq 2(\sigma_0 + \xi_0) \Delta^{1/2} \ln(n) \right\} \cap N_k = 0 \right) \lesssim P\left( \left| \int_{k\Delta}^{(k+1)\Delta} b(X_s)ds + Z_{k\Delta} \right| \geq (\sigma_0 + \xi_0) \Delta^{1/2} \ln(n) \right) \lesssim \Delta^4 + n^{-1}. $$

By the Burkholder Davis Gundy inequality, we obtain that

$$\mathbb{E}\left( \sup_{s \leq (k+1)\Delta} (X_s - X_{k\Delta})^4 : \Omega_{N,k} \right) \lesssim \Delta^{2-\beta/4}. $$

Moreover,

$$S_3 := \mathbb{E}\left( \left( \int_{k\Delta}^{(k+1)\Delta} (\xi(X_s) - \xi(X_{k\Delta}) - \xi(X_{k\Delta} - )) dL_s^{(1)+(2)} \right)^4 \right) \lesssim \left( \int_{k\Delta}^{(k+1)\Delta} \Delta \int_{-\Delta^{1/4}}^{\Delta^{1/4}} z^2 \nu(dz) \right)^2 + \int_{k\Delta}^{(k+1)\Delta} \Delta^{1/4} \int_{-\Delta^{1/4}}^{\Delta^{1/4}} z^4 \nu(dz) \lesssim \Delta^{5-\beta/2} + \Delta^{4-\beta/2} \lesssim \Delta^{4-\beta/2}. $$

(13)
Then by a Markov’s inequality,
\[ Pr \left( \left| \int_{h \Delta}^{(k+1) \Delta} (\xi(X_s^\Delta) - \xi(X_{k \Delta}^\Delta)) dL_s^{(1)+(2)} \right| \geq \Delta^{1/2} \ln(n) \right) \lesssim \Delta^{2-\beta/2} \]
which ends the proof.

### 6.5 Proof of Lemma 7
From the Burkholder Davis Gundy inequality and Proposition 11 we derive easily the bounds for \( \tilde{A}_{h \Delta} \) and \( \tilde{B}_{h \Delta} \). It remains to bound \( E \left( \tilde{E}_{h \Delta}^4 \mid \mathcal{F}_{h \Delta} \right) \) and \( E \left( \tilde{E}_{h \Delta}^4 \mid \mathcal{F}_{h \Delta} \right) \). We first bound \( E \left( \left( J_{h \Delta}^{(1)+(2)} \right)^4 \right) \). We have that
\[ J_{h \Delta}^{(1)+(2)} = \int_{h \Delta}^{(k+1) \Delta} (\xi(X_s^\Delta) - \xi(X_{k \Delta}^\Delta)) dL_s^{(1)+(2)} \]
By [13], \( E \left( \left( \int_{h \Delta}^{(k+1) \Delta} (\xi(X_s^\Delta) - \xi(X_{k \Delta}^\Delta)) dL_s^{(1)+(2)} \right)^4 \right) \lesssim \Delta^{4-\beta/2} \). It remains to bound \( E \left( \left( \int_{h \Delta}^{(k+1) \Delta} dL_s^{(1)+(2)} \mathbb{1}_{\Omega_{h \Delta}} \right)^4 \right) \). This is nearly Proposition 4.5 of Mal [2012]. Let us introduce a nonnegative function \( f \in \mathcal{C}^\infty \) such that
\[
\begin{cases}
  f(x) = x^4 & \text{if } |x| \leq 1 \\
  f(x) = 0 & \text{if } |x| \geq 2.
\end{cases}
\]
Let us set \( f^a(x) = a^4 f(x/a) \). By stationarity, we have
\[
E \left[ \left( \int_{h \Delta}^{(k+1) \Delta} dL_s^{(1)+(2)} \mathbb{1}_{\Omega_{h \Delta}} \right)^4 \right] = E \left[ \left( L_{\Delta}^{(1)+(2)} \mathbb{1}_{\Omega_{h \Delta}} \right)^4 \right] \leq E \left( f^{4/2} \ln(n) \left( L_{\Delta}^{(1)+(2)} \right)^4 \right).
\]
The following result is needed.

**Result 13.** [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
\[ S(\mathbb{R}) = \left\{ h \in \mathcal{C}^\infty, \forall p, q \in \mathbb{N}, \exists C_{p,q}, \forall x \in \mathbb{R}, |x^p h^{(q)}(x)| \leq C_{p,q} \right\}. \]

Then we have the following properties:
1. For any \( h_1, h_2 \in L^2(\mathbb{R}), (a_1, a_2) \in \mathbb{R}^2 \), \( \mathcal{F}(a_1 h_1 + a_2 h_2) = a_1 \mathcal{F}h_1 + a_2 \mathcal{F}h_2 \).
2. For any \( h \in L^2(\mathbb{R}), \mathcal{F}h \in L^2(\mathbb{R}) \) and \( \forall x \in \mathbb{R}, h(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{itx} \mathcal{F}h(t) dt \).
3. For any \( h \in L^2(\mathbb{R}), \mathcal{F}h(.a)(x) = |a| \mathcal{F}h(ax) \).

24
4. 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 \( \mathcal{F}\delta_y(x) = e^{-ixy} \),
\[
h(0) = \int_{\mathbb{R}} h(y)\delta_0(y)dy = \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}h(u)du.
\]

5. For any \( h \in S(\mathbb{R}) \), \( \mathcal{F}h \in S(\mathbb{R}) \) and
\[
\mathcal{F}(h^{(q)})(x) = (ix)^q \mathcal{F}h(x).
\]

By Result [134] we have that
\[
\mathbb{E}\left( f^a(L^{(1)+(2)}_t) \right) = \int_{\mathbb{R}} f^a(x)P_{L^{(1)+(2)}_\Delta}(dx)
= \int_{\mathbb{R}} \mathcal{F}f^a(u)\bar{\phi}_\Delta(u)du
\]
where \( \phi_\Delta \) is the characteristic function of the Lévy process \( L^{(1)+(2)}_\Delta \):
\[
\phi_\Delta(u) = \exp \left( \Delta \int_{-\Delta^{1/4}}^{\Delta^{1/4}} (e^{iux} - 1 - iux)\nu(dx) \right).
\]

By a Taylor development in 0, we obtain that
\[
\phi_\Delta(u) = 1 + \psi_\Delta(u) + R(\Delta, u)
\]
with \( \psi_\Delta(u) = \Delta \int_{-\Delta^{1/4}}^{\Delta^{1/4}} (e^{iux} - 1 - iux)\nu(dx) \). Then
\[
\mathbb{E}\left( f^a(L^{(1)+(2)}_t) \right) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}f^a(u)du + \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}f^a(u)\bar{\psi}_\Delta(u)du
+ \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}f^a(u)R(\Delta, u)du.
\]

By Result [134] \( \int_{\mathbb{R}} \mathcal{F}f^a(u)du = 2\pi f^a(0) = 0 \) and consequently,
\[
\int_{\mathbb{R}} \mathcal{F}f^a(u)\overline{\psi_\Delta(u)}du = \int_{\mathbb{R}} \mathcal{F}f^a(u)\Delta \int_{-\Delta^{1/4}}^{\Delta^{1/4}} (e^{-iux} - 1 + iux)\nu(dx)
= \Delta \int_{-\Delta^{1/4}}^{\Delta^{1/4}} (2\pi f^a(x)\nu(dx) - f^a(0)) + \int_{\mathbb{R}} \mathcal{F}f^a(u)iux\nu(dx).
\]

By Result [135] as \( f^a \in S(\mathbb{R}) \), \( \int_{\mathbb{R}} \mathcal{F}f^a(u)iudu = \int_{\mathbb{R}} \mathcal{F}(f^a)'(u)du = (f^a)'(0) = 0 \). Then
\[
\int_{\mathbb{R}} \mathcal{F}f^{1/2}\ln(n)(u)\overline{\psi_\Delta(u)}du = 2\pi \Delta \int_{-\Delta^{1/4}}^{\Delta^{1/4}} f^{1/2}\ln(n)(x)\nu(dx)
\lesssim \Delta \int_{-2\Delta^{1/2}\ln(n)}^{2\Delta^{1/2}\ln(n)} x^4\nu(dx) \lesssim \Delta^{2-\beta/2}\ln(n)^{4-\beta}.
\]
It remains to bound $E \left( \int_{\mathbb{R}} \mathcal{F} f^\alpha(u) \mathcal{R}(\Delta, u) du \right)$. We have that

$$|\mathcal{R}(\Delta, u)| = \left| e^{\psi_\Delta(u)} - \psi_\Delta(u) - 1 \right| \leq \left| \psi_\Delta^2(u) \right|.$$ 

According to Kappus [2012], $|\psi_\Delta(u)| \lesssim C \Delta |u|^\beta$. By Result 13.3, $\mathcal{F} f^\alpha(u) = a^5 \mathcal{F}(au)$ and therefore

$$\mathbb{E} \left \{ \left( \int_{\mathbb{R}} \mathcal{F} f^\alpha(u) \mathcal{R}(\Delta, u) du \right) \right \} \lesssim \Delta^2 \int_{\mathbb{R}} |\mathcal{F} f^\alpha(u)||u|^{2\beta} du \lesssim \Delta^2 \int_{\mathbb{R}} a^5 |\mathcal{F}(au)||u|^{2\beta} du.$$ 

As $f^\alpha \in S(\mathbb{R})$, $\mathcal{F} f^\alpha \in S(\mathbb{R})$ and then for any $m > 0$, $\exists C_m > 0$, $|\mathcal{F}(u)| \leq C_m |u|^{-m}$. Then, for any $m \in \mathbb{N}$:

$$\mathbb{E} \left( \int_{\mathbb{R}} \mathcal{F} f^{\Delta^{1/2} \ln(n)}(u) \mathcal{R}(\Delta, u) du \right) \lesssim \Delta^2 \int_{\mathbb{R}} a^{5-m} |u|^{2\beta-m} \wedge a^5 |u|^{2\beta} du.$$ 

We choose $m$ such that $2\beta + 1 < m \leq 3 + \beta$. As $\beta < 2$, $m$ always exists. Then $\int_{\mathbb{R}} |u|^{2\beta-m} \wedge |u|^{2\beta} < \infty$ and we get:

$$\mathbb{E} \left( \int_{\mathbb{R}} \mathcal{F} f^{\Delta^{1/2} \ln(n)}(u) \mathcal{R}(\Delta, u) du \right) \lesssim \Delta^2 \left( \Delta^{1/2} \ln(n) \right)^{2-\beta} \lesssim \Delta^{3-\beta/2} \ln(n)^{2-\beta}.$$ 

Then we obtain

$$\mathbb{E} \left[ \left( J_{2\kappa\Delta}^{(1)+(2)} \right)^4 1_{\sigma_k} \right] \lesssim \Delta^{3-\beta/2} (\ln(n))^{4-\beta}. \quad (14)$$

**Bound of $\mathbb{E} \left( \tilde{E}_{2\kappa\Delta}^2 1_{\Omega_{N,k} \cap \Omega_{N,k}} \right)$:**

On $\Omega_{N,k}$,

$$\tilde{E}_{2\kappa\Delta} = \left( 2b(X_{k\Delta}) J_{k\Delta}^{(1)+(2)} \Delta + \frac{J_{k\Delta}^{(1)+(2)}}{\Delta} Z_{k\Delta} + \left( \frac{J_{k\Delta}^{(1)+(2)}}{\Delta} \right)^2 \right) 1_{\sigma_k}.$$ 

Then by (14),

$$\mathbb{E} \left( \tilde{E}_{2\kappa\Delta}^2 1_{\Omega_{N,k} \cap \Omega_{N,k}} \right) \lesssim \frac{\mathbb{E} \left( \left( J_{k\Delta}^{(1)+(2)} \right)^4 1_{\sigma_k} \right)}{\Delta^2} + \mathbb{E} \left( Z_{k\Delta}^4 \right) \lesssim \Delta^{1-\beta/2} \ln(n)^{4-\beta}.$$
6.6 Proof of Theorem 8

As before, we decompose the bound of the risk on $\Omega_n$ and $\Omega_n^c$. We bound the risk on $\Omega_n^c$ in the same way as in the proof of Theorem 4. On $\Omega_n$, we obtain that:

$$\mathbb{E} \left( \tilde{E}_h^2 \mathbbm{1}_{\Omega_{X,k} \cap \Omega_{N,k}^c} \right),$$

where $\tilde{E}_h(t) = n^{-1} \sum_{k=1}^n \tilde{B}_{h,k} t(x_k)$. By Lemma 7, we get that

$$\mathbb{E} \left( F_h^2 \right) \leq \Delta + \frac{\sigma_h^2}{\Delta} \mathbb{E} \left( B_h^2 \mathbbm{1}_{\Omega_{X,k} \cap \Omega_{N,k}^c} \right) + \mathbb{E} \left( B_h^2 \mathbbm{1}_{\Omega_{X,k} \cap \Omega_{N,k}^c} | F_h \right)^2.$$

By Lemma 8, $P \left( |Z_t| \leq \Delta^{1/2} \ln(n) \right) \leq n^{-1}$ and then:

$$\mathbb{E} \left( B_h^2 \mathbbm{1}_{\Omega_{X,k} \cap \Omega_{N,k}^c} \right) \leq \mathbb{E} \left( B_h^2 \mathbbm{1}_{|B_h| \leq \ln^2(n)} \right) + \mathbb{E} \left( \ln^2(n) \mathbbm{1}_{\Omega_{X,k} \cap \Omega_{N,k}^c} \right) \leq n^{-1} + \ln^2(n) P \left( \Omega_{X,k} \cap \Omega_{N,k}^c \right) \leq \Delta^{2-\beta/2}.$$
Then $\mathbb{E}(F^2_{k\Delta}) \leq \Delta^{1-\beta/2}$. We have that

$$
\mathbb{E}\left(\sup_{t \in S_m} \hat{\nu}^2_n(t)\right) \leq \sum_{\lambda \in \Lambda} \mathbb{E}\left(\hat{\nu}^2_\lambda(\varphi_\lambda)\right) \leq \frac{\mathbb{E}(B^2_{k\Delta})}{n} \lesssim \sigma^4_0 \frac{1}{n},
$$

where $(\varphi_\lambda)_{1 \leq \lambda \leq D_m}$ is the orthonormal basis of $S_m$ for the $\|\cdot\|$-norm.

### 6.7 Proof of Theorem 10

We apply the Berbee’s coupling Lemma to the random exponentially $\beta$-mixing vectors $(\tilde{B}_{k\Delta}, X_{k\Delta})$. For any $a \in \{0, 1\}$, we can construct random variables

$$
V^*_a = \frac{1}{q_n} \sum_{l=1}^{q_n} \tilde{B}^*_l(2(k-1)+a)q_n + t| \Delta t (X^*_l(2(k-1)+a)q_n + t)\Delta
$$

independent and of same law as

$$
V_{k,a} = \frac{1}{q_n} \sum_{l=1}^{q_n} (2(k-1)+a)q_n + t| \Delta t (X^*_l(2(k-1)+a)q_n + t)\Delta.
$$

Let us set $\tilde{\Omega}^* = \{\omega, \forall a, \forall k, V^*_{k,a} = V_{k,a}\}$, $\mathbb{P}(\tilde{\Omega}^*) \lesssim n^{-4}$. Let us consider the set $\Omega_Z = \{\omega, \forall k |Z_{k\Delta}| \leq 4\sigma_0 \ln(n)\Delta^{1/2}\}$ on which the random variables $\tilde{B}_{k\Delta}$ are bounded. According to inequality [3], $\mathbb{P}(\Omega^*_Z) \lesssim n^{-4}$.

Let us set $\tilde{\theta} = \Omega_n \cap \Omega_Z \cap \tilde{\Omega}^*$. We bound the risk on $\tilde{\theta}^c$ in the same way as on $\Omega^*_n$. Let us set

$$
\hat{\nu}^*_n(t) = \hat{\nu}^*_n(0,t) + \hat{\nu}^*_n(t) \quad \text{with} \quad \hat{\nu}^*_n(0,t) = \frac{1}{p_n} \sum_{k=1}^{p_n} V^*_{k,a} - \mathbb{E}(V^*_{k,a}).
$$

For any $m \in \mathcal{M}_n$:

$$
\mathbb{E}\left(\|\hat{\mu}_n - \sigma_A \|^2 \right) \leq 3\|\sigma^2_0 - \sigma^2\|^2 + 12\mathbb{E}\left(F^2_{k\Delta}\right) + 12\mathbb{E}\left(\tilde{B}^2_{k\Delta} \mathbb{1}_{\tilde{\theta}}\right)^2 + 2\sigma_0 \ln(n)\sigma_0 - 2\sigma_0 \ln(n)\sigma_0 + \mathbb{E}\left(\sup_{t \in \mathcal{M}_n} (\hat{\nu}^*_n(t))^2\right).
$$

Let us introduce the function $\tilde{p}(m, m') = (\tilde{p}(m) + \tilde{p}(m'))/12$. We have that

$$
\left[\left(\sup_{t \in \mathcal{M}_n} \hat{\nu}^*_n(t) - \tilde{p}(m, m')\right) \mathbb{1}_{\tilde{\theta}}\right]_+ \leq \sum_{m' \in \mathcal{M}_n} \left[\left(\sup_{t \in \mathcal{M}_n} \hat{\nu}^*_n(t) - \tilde{p}(m, m')\right) \mathbb{1}_{\tilde{\theta}}\right]_+.
$$

On $\tilde{\theta}$, for any $a$, the random variables $(V^*_{k,a})$ are independent, centred and bounded. We have that $|V^*_{k,a}| \leq M = \sigma_0^2 \ln^2(n)\Delta^{1/2}$, $\mathbb{E}\left(V^*_{k,a}\right)^2 \leq \tilde{V} = \sigma^2_0/q_n$ and

$$
\mathbb{E}\left(\sup_{t \in \mathcal{M}_n} \hat{\nu}^*_n(t)\right) \leq \tilde{H} = \sigma^4_0 \frac{D}{n}.
$$

28
By the Talagrand’s inequality, we deduce:

\[ R_3 := \mathbb{E} \left[ \sup_{t \in \mathcal{B}_{m,m'}} (\nu_n^*(t))^2 - 12p(m,m') \right]_+ \leq \frac{1}{n} \exp \left( -c \frac{p_n D}{n} \right) + \frac{\ln^4(n)D}{p_n^2} \exp \left( -c' \frac{p_n D}{\ln(n)D^{1/2}} \right) \]

\[ \leq \frac{1}{n} \exp (-cD) + \frac{\ln^6(n)}{n^2\Delta^2} D \exp \left( -c' \sqrt{n\Delta} \frac{\ln^3(n)}{\ln^3(n)} \right). \]

As \( D \leq n\Delta \) and \( \ln^3(n) \ll n\Delta \), we find:

\[ \mathbb{E} \left[ \sup_{t \in \mathcal{B}_{m,m}} \nu_n^2(t) - \text{pen}(m) \right]_+ \leq \frac{1}{n} \sum_{m'} e^{-cD_{m,m'}} + \frac{\ln^4(n)\Delta}{n\Delta} \exp \left( -c' \frac{n\Delta}{\ln^3(n)} \right) \]

\[ \leq \frac{1}{n}. \]

References

Applebaum, D. (2004) *Lévy processes and stochastic calculus*, Cambridge Studies in Advanced Mathematics, volume 93. Cambridge University Press, Cambridge.

Arlot, S. and Massart, P. (2009) Data-driven calibration of penalties for least-squares regression. *Journal of Machine Learning Research*, 10 pp. 245–279.

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

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

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

Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007) Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13 (2) pp. 514–543.

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

Dellacherie, C. and Meyer, P.A. (1980) *Probabilités et potentiel. Chapitres V à VIII*, Actualités Scientifiques et Industrielles [Current Scientific and Industrial Topics], volume 1385. Hermann, Paris, revised edition. Théorie des martingales. [Martingale theory].
DeVore, R.A. and Lorentz, G.G. (1993) *Constructive approximation. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, volume 303. Springer-Verlag, Berlin.

Hanif, M., Wang, H. and Lin, Z. (2012) Reweighted Nadaraya-Watson estimation of jump-diffusion models. *Sci. China Math.*, 55 (5) pp. 1005–1016. URL http://dx.doi.org/10.1007/s11425-011-4340-4

Hoffmann, M. (1999) Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79 (1) pp. 135–163.

Kappus, J. (2012) *Nonparametric adaptive estimation for discretely observed Lévy processes*. Ph.D. thesis, Humboldt-Universität zu Berlin.

Mai, H. (2012) Efficient maximum likelihood estimation for lévy-driven ornstein-uhlenbeck processes.

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

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

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

Rubenthaler, S. (2010) *Probabilités : aspects théoriques et applications en filtrage non linéaire, systèmes de particules et processus stochastiques*. Habilitation à diriger des recherches, Université de Nice-Sophia Antipolis, France.

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

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

