Non parametric estimation of the jump coefficient of a diffusion with jumps

Émeline Schmisser

Non parametric estimation of the jump coefficient of a diffusion with jumps

Émeline Schmisser
Université de Lille, CNRS, UMR 8524 - Laboratoire Paul Painlevé
Villeneuve d'Ascq, France
email: emeline.schmisser@univ-lille.fr
This work was supported by the Labex CEMPI (ANR-11-LABX-0007-01)

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. This process is supposed to be ergodic, exponentially β-mixing and stationary. It 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 construct a robust, adaptive non-parametric estimator of the function ξ^4 thanks to a penalized least-square approach. We provide bounds of the empirical and L^2-risk of our estimator.

Keywords: jump diffusions, nonparametric estimation, model selection.
Subject classification: 62G05, 62M05.
The code of the simulation study is available on math.univ-lille1.fr/~schmisse/recherche.html

1 Introduction

We consider the jump diffusion given by the one-dimensional stochastic differential equation

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

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 centred Lévy process independent of (W_t, η). It can be written

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

where μ is a Poisson measure of intensity $\nu(dz)dt$. We do not assume that the jumps are of finite intensity, only that L_t is centred and has moments of any order.

The diffusion is observed at discrete times $t = 0, \Delta, \ldots, n\Delta$ under the asymptotic framework: $n\Delta \to \infty$ and $\Delta \to 0$. Our aim is to construct adaptive non-parametric estimator of the jump coefficient ξ^4 on a compact interval $A \subset \mathbb{R}$.

Jump diffusions are used to modelize dynamical systems where the noise is discontinuous or too intensive to be modeled by a Brownian motion, for instance polymerazation phenomenons

(Berestycki (2004)), telephone noise (Protter and Talay (1997)), or finance (see Aït-Sahalia and Jacod (2009) or Protter and Talay (1997) for instance). There are many articles on the estimation for jumps diffusions, although quite few estimate the jump coefficient nonparametrically.

For the parametric estimation, we can cite Shimizu (2006) or Shimizu and Yoshida (2006) who estimate the three coefficients b, σ and ξ for stationary finite intensity processes thanks to a contrast function. Their estimators are consistent and asymptotically normal if the sampling interval Δ is sufficiently small. The estimators of the parameter of b and ξ^2 converge with rate $\sqrt{n\Delta}$, whereas the estimator of the parameter of σ^2 converges with rate $\sqrt{n}$.

For nonparametric estimation, Bandi and Nguyen (2003) use kernel estimators of the infinitesimal moments, $M_1 = b(x)$, $M_2 = \sigma^2(x) + \xi^2(x)\int z^2\nu(dz)$ and $M_4 = \xi^4(x)\int z^4\nu(dz)$. Their estimators are asymptotically normal and converge with rate $(n\Delta)^{-2/5}$. Their estimator do not depends of the regularity of the functions, and are not adaptive. Hanif et al. (2012) and Hanif (2016) construct estimators of the infinitesimal second moment $\sigma^2 + \xi^2\int z^2\nu(dz)$ with reweighted Nadaraya-Watson and Gamma Nadaraya-Watson estimators respectively. Their estimators converge almost surely, and are asymptotically normal.

To estimate σ^2 or ξ^2 non parametrically, a threshold is often used to suppress (or find) the jumps. For instance, Mancini (2009) estimate the integrated volatility. The estimator converge, and in the finite activity cas, the estimator is asymptotically normal. Mancini and Renò (2011) (for the finite intensity case) and Song et al. (2022) (for compound Poisson processes) estimate the diffusion coefficient σ^2, using local time and kernel estimators. Schmisser (2019) construct a robust and adaptive estimator of σ^2 and of the second infinitesimal moment, even for infinite intensity Lévy processes.

Mancini and Renò (2011) for the finite activity case and Park and Wang (2021) for the finite intensity case both construct estimators for the jump coefficient ξ^2, using local times and kernel estimator. The rate of convergence of the kernel estimator of ξ^2 is $(n\Delta)^{-2/5} + \Delta^{1/2}$.

There are two possibilities to construct an estimator of the jump component. Either take the second infinitesimal moment when the increments are big enough (when a jump is suspected): indeed, for finite intensity processes (see for instance Park and Wang (2021, Section 5.2)).

$$\frac{(X_{(k+1)\Delta} - X_{k\Delta})^2}{\Delta}\mathbf{1}_{|X_{(k+1)\Delta}-X_{k\Delta}|\geq\Delta^{1/2-\varepsilon}} = \xi^2(X_{k\Delta})\int z^2\nu(dz) + \text{ centred terms } + \text{ small terms.}$$

The problem of this method is that the small jumps will not be kept. It is not very important if the Lévy process is a compound Poisson process, but if the jumps have infinite activity or infinite intensity, the small terms will not be so small.

In this article, like Bandi and Nguyen (2003), we use the second possibility and consider an estimator based on the fourth moments of the increments:

$$T_{k\Delta} = \frac{(X_{(k+1)\Delta} - X_{k\Delta})^4}{\Delta} = \xi^4(X_{k\Delta})\int z^4\nu(dz) + \text{ centred terms } + \text{ small terms.}$$

The Brownian terms are small because of the power 4, and it only remains the jumps. We introduce a sequence of increasing subspaces S_m of $L^2(A)$ and construct a sequence of estimators $\hat{\xi}^4_m$ by minimizing on each S_m the contrast function $\gamma_n(t) = \frac{1}{n}\sum_{k=1}^n (T_{k\Delta} - t(X_{k\Delta}))^2$. When $\int z^4\nu(dz)$ is known, its L^2-risk is bounded by

$$\mathbb{E}\left(\left\|\hat{\xi}^4 - \xi^4\right\|^2_{L^2(A)}\right) \lesssim \underbrace{\left\|\xi^4 - \xi^4_m\right\|^2_{L^2(A)}}_{\text{bias term}} + \underbrace{\Xi\frac{D_m}{n\Delta}}_{\text{variance term}} + \Delta$$

where D_m is the dimension of the subspace S_m. The bias term decreases when the dimension D_m increases, whereas the variance term increases. To find a good bias-variance compromise, we introduce a penalty, $\widehat{pen}(m)$, proportional to the variance term. To choose the dimension, we minimize the quantity $\gamma_n(\hat{\xi}_m^4 I_4) + \widehat{pen}(m)$ and deduce the "best" dimension $\tilde{m}$. Finally, we prove that the risk of the robust estimator, $\hat{\xi}_{\tilde{m}}^4$, satisfies an oracle inequality.

The paper is divided as follows: the model and its assumptions are stated in Section 2. In Section 3, we construct the estimator and bound its risk. Section 4 is devoted to the simulations and the proofs are gathered in Section 5.

We introduce some notations.

In all the paper, C means a constant that does not depend on n, Δ or the dimension D_m, but may be different from one line to another. We note $A \lesssim B$ if $A \leq CB$ where C is a constant that does not depend on n, Δ or D_m. $\mathbb{R}_+^*$ is the set of positive numbers, and $\mathscr{C}^r$ the set of fonctions t that are r times continuously differentiable.

We set $I_k = \int z^k \nu(dz)$ if this quantity exists.

For a function t in $L^2(A)$, and π a density, $\|t\|_\pi^2 = \int_A t^2(x)\pi(dx)$ and the empirical norm is $\|t\|_n^2 = \frac{1}{n}\sum_{i=1}^n t^2(X_{k\Delta})$. We consider three different projections: t_m is the orthogonal projection of t on S_m (for the L^2-norm), $t_{m,\pi}$ is the orthogonal projection of t on S_m for the $\|.\|_\pi$-norm and $\Pi_m t$ is the orthogonal projection on S_m for the empirical norm $\|.\|_n$.

2 Model and assumptions

We consider the stochastic differential equation given by (1) and assume the following assumptions:

Assumption A1.

a. *The functions b, σ and ξ are Lipschitz.*

b. *The drift function b is elastic:*

$$\exists M > 0, \exists C > 0\ , \forall x,\ |x| \geq M\ , xb(x) \leq -Cx^2.$$

c. *The diffusion and jump functions σ and ξ are bounded from below and above: there exists $\sigma_0, \sigma_1, \xi_0, \xi_1 \in (\mathbb{R}_+^*)^4$ such that*

$$\forall x, 0 < \xi_0 \leq \xi(x) \leq \xi_1 \quad \textit{and} \quad 0 < \sigma_0 \leq \sigma(x) \leq \sigma_1.$$

d. *The Lévy measure has moments of any order: $\forall k \geq 2$, $\int_{\mathbb{R}} |z|^k \nu(dz) < +\infty$.*

e. $I_4 = \int_{\mathbb{R}} z^4 \nu(dz) = 1$.

Item a ensures that SDE (1) has a unique strong solution (see Applebaum (2009, Theorem 6.2.9). According to Masuda (2007, Theorem 2.2 and Lemma 2.4), under Assumptions A1a-d, the process $(X_t)_{t\geq 0}$ has a unique stationary distribution, is ergodic and exponentially β-mixing. Assumption e ensures that the model is identifiable: indeed, if we replace $\nu(.)$ by $c\nu(c.)$ and ξ by ξ/c, we obtain the same process $(X_t)_{t\geq 0}$. If this assumption is not fulfilled, our estimator will estimate $\xi^4(X_{k\Delta})I_4$, wich enables us to see at least the shape of ξ.

We can know assume:

Assumption A2.

a. The process $(X_t)_{t\geq 0}$ is stationary.

b. Its stationary measure has a density π which is bounded on any compact set.

c. the density π is bounded from below and above on the compact of estimation A:

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

By Masuda (2007, Theorem 2.2), it implies that X_t has also moments of any order.
Let us consider 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 proposition helps us to control the increments of the process. It is proved for instance in Applebaum (2009, Theorems 4.4.21 and 4.4.23)).

Proposition 1. *According to the Burkholder-Davis-Gundy inequality, for any $p > 0$, there exists a constant C_p such that:*

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

and

$$\begin{aligned}\mathbb{E}\left(\sup_{s\in[t,t+h]}\left|\int_t^s \xi(X_{u^-})dL_u\right|^p |\mathscr{F}_t\right) \leq\ & C_p\mathbb{E}\left(\left|\int_t^{t+h}\xi^2(X_u)du\right|^{p/2}|\mathscr{F}_t\right) I_2^{p/2} \\ & + C_p\mathbb{E}\left(\int_t^{t+h}|\xi^p(X_u)|\,du\,|\mathscr{F}_t\right) I_p.\end{aligned}$$

To simplify the notations, let us set

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

The following lemma follows almost directly from the Burkhölder-Davis-Gundy inequality (see for instance Schmisser (2014, proof of Proposition 1)).

Lemma 2. *For any integer $p > 0$:*

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

and

$$\forall u \geq 0, \quad \mathbb{E}\left(\sup_{0\leq s\leq t}(X_{s+u}-X_u)^{2p}\middle|\mathscr{F}_u\right) \leq C_{2p}\left(t\xi_1^{2p}I_{2p} + t^p\left(\sigma_1^{2p}+\xi_1^{2p}I_2^p\right)\right).$$

Then, as b, σ^2 and ξ^2 are Lipshitz: $\mathbb{E}\left(|B_{k\Delta}|^{2p}\right) \lesssim \Delta^{2p}$, $\mathbb{E}\left(J_{k\Delta}^{2p}\right) \lesssim \Delta^p$ and $\mathbb{E}\left(J_{k\Delta}\right)^{2p} = \Delta\xi^{2p}(X_{k\Delta})I_{2p} + O(\Delta^{3/2})$.

We need to control the variance term, more precisely the moments of the jumps. The Lévy-Khintchine formula states that $\mathbb{E}\left(e^{iuL_\Delta}\right) = e^{-\Delta\int_{\mathbb{R}}(e^{iuz}-1-iuz)\nu(dz)}$. The following lemma is exactly Lemma 13 of Carpentier et al. (2021).

Lemma 3. *We get the following moments:*

$$\begin{aligned}
\mathbb{E}\left(L_\Delta^2\right) &= I_2\Delta \\
\mathbb{E}\left(L_\Delta^4\right) &= I_4\Delta + 3I_2^2\Delta^2 \\
\mathbb{E}\left(L_\Delta^6\right) &= I_6\Delta + (15I_2I_4 + 10I_3^2)\Delta^2 + 15I_2^3\Delta^3 \\
\mathbb{E}\left(L_\Delta^8\right) &= I_8\Delta + (35I_4^2 + 56I_3I_5 + 28I_2I_6)\Delta^2 + (210I_2^2I_4 + 280I_2I_3^2)\Delta^3 + 105I_2^4\Delta^4.
\end{aligned}$$

To bound the risk the adaptive estimator, we apply a Talagrand's inequality. To this end, we need independent and bounded random variables. As we have an exponentially β-mixing process, we can construct independent random variables thanks to the Berbee's coupling lemma. But to work with nearly bounded variables, we need the following assumption:

Assumption A3.

a. The jumps are sub-exponential: $\exists M, C, \lambda \in (\mathbb{R}_+^*)^3$, $\forall z, |z| \geq M$ $\nu([-z,z]^c) \leq Ce^{-\lambda|z|}$.

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

c. The Blumenthal-Getoor index is strictly less than 2: there exists $\beta < 2$, $\int_{-1}^1 z^\beta \nu(dz) < \infty$. *This is not a very strong assumption, as* $\int_{-1}^1 z^2\nu(dz)$ *is already finite.*

Under Assumption A1 and 3, for any $c \leq \lambda/\xi_1$, any $t > 0$, $\mathbb{E}\left(e^{cX_t}\right) < \infty$.

Proof. Apply Theorem 2.2 of Masuda (2007) with

$$f^*(x) = e^{-cx}\mathbf{1}_{|x|\geq 1} + e^c\left(1 + \frac{c^2}{2} - \frac{5c}{8} + \frac{3c-c^2}{4}x^2 + \frac{c^2-c}{8}x^4\right)\mathbf{1}_{|x|<1}. \qquad \square$$

Under this assumption, the increments can be bounded. The small jumps (smaller than $\sqrt{\Delta}$) have the same comportement than the Brownian motion terms. As we have only sub-exponential-jumps, the size of the bigger jumps is also nearly bounded. The following lemma is proved in Schmisser (2019, Lemma 2):

Lemma 4. *Under Assumptions A1-A3, we have:*

- $\forall \varepsilon > 0$, $\forall r > 0$, $\mathbb{P}\left(|B_{k\Delta}| \geq \Delta^{1-\alpha}\right) \lesssim n^{-r}$.
- $\forall r > 0$, $\mathbb{P}\left(|Z_{k\Delta}| \geq r\sigma_1\Delta^{1/2}\ln(n)\right) \leq 2n^{-r}$.
- $\forall p > 0$, $\forall r > 0$, *for any* C_J *and* $c(p)$ *large enough:*

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

where $q_n \geq 2\ln(n)/\beta$, *and* β *the* β*-mixing coefficient.*

3 Estimation of the jump coefficient

Our aim is to construct an adaptive nonparametric estimator of the jump function ξ on the interval A. To make the diffusion terms small, we consider the fourth moments of the increments:

$$T_{k\Delta} = \frac{(X_{(k+1)\Delta} - X_{k\Delta})^4}{\Delta} = \frac{(B_{k\Delta} + Z_{k\Delta} + J_{k\Delta})^4}{\Delta}.$$

We can write $T_{k\Delta} = \xi^4(X_{k\Delta})I_4 + E_{k\Delta} + F_{k\Delta}$ where

$$\begin{aligned}\Delta E_{k\Delta} &:= (B_{k\Delta} + Z_{k\Delta} + J_{k\Delta})^4 - (Z_{k\Delta} + J_{k\Delta})^4 + \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^4 \middle| \mathscr{F}_{k\Delta}\right) - \xi^4(X_{k\Delta})I_4\Delta \\ \Delta F_{k\Delta} &:= (Z_{k\Delta} + J_{k\Delta})^4 - \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^4 \middle| \mathscr{F}_{k\Delta}\right).\end{aligned}$$

The terms $E_{k\Delta}$ are small, whereas $F_{k\Delta}$ are centred. The following Lemma is proved in Section 5, using Lemma 2.

Lemma 5.

- $\mathbb{E}\left(E_{k\Delta}^2\right) \lesssim \Delta$, $\mathbb{E}\left(E_{k\Delta}^4\right) \lesssim \Delta$ *and* $\mathbb{E}\left(E_{k\Delta}^8\right) \lesssim \Delta$.
- $\mathbb{E}\left(F_{k\Delta}|\mathscr{F}_{k\Delta}\right) = 0$, $\mathbb{E}\left(F_{k\Delta}^4\right) \lesssim 1/\Delta^3$ *and*

$$\mathrm{Var}\left(F_{k\Delta}|\mathscr{F}_{k\Delta}\right) = \frac{\xi^8(X_{k\Delta})I_8}{\Delta} + O\left(\frac{1}{\sqrt{\Delta}}\right).$$

3.1 Subspaces of approximation

To construct an adaptive estimator of ξ^4, we consider a sequence of increasing subspaces $S_0 \subseteq \ldots \subseteq S_m, \ldots$ such that $\cup_m S_m$ is dense in $L^2(A)$. We minimize a contrast function $\gamma_n(t)$ on each S_m and then choose the best estimator by introducing a penalty function (see for instance Barron et al. (1999)). The rate of convergence of our estimator will depend on the regularity of the coefficient ξ, i.e. its modulus of smoothness.

Definition (Modulus of smoothness)**.** *The modulus of continuity of a function f is defined by*

$$\omega(f,t) = \sup_{|x-y|\leq t} |f(x) - f(y)|.$$

If f is Lipschitz, the modulus of continuity is proportional to t. If $\omega(f,t) = o(t)$, then f is constant: the modulus of continuity cannot measure higher smoothness.

We define the modulus of smoothness by

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

If $f \in \mathscr{C}^r$ (that is if f is r-times continuously differentiable), then for $1 \leq p \leq \infty$:

$$\omega_r(f,t)_p \leq t^r \omega(f^{(r)},t)_p.$$

Definition (Besov space)**.** *The Besov space* $B^{\alpha}_{2,\infty}$ *is the set of functions:*

$$B^{\alpha}_{2,\infty} = \{f \in L^2, \sup_{t>0} t^{-\alpha}\omega_r(f,t)_2 < \infty\}$$

where $r = \lfloor \alpha + 1 \rfloor$*. The norm on a Besov space is defined by:*

$$\|f\|_{B^{\alpha}_{2,\infty}} := \sup_{t>0} t^{-\alpha}\omega_r(f,t)_2 + \|f\|_{L^2}\,.$$

For more details see DeVore and Lorentz (1993).

Assumption A4. *We consider a series of vectorial subspaces of* $L^2(A)$*,* $(S_m)_{m\in\mathscr{M}_n}$ *satisfying the assumptions*

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:*

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

with $\|t\|^2_{L^2(A)} = \int_A t^2(x)dx$ *and* $\|t\|_{\infty} = \sup_{x\in A}|t(x)|$*. This implies that, for an orthonormal basis* φ_λ *of* S_m*,* $\left\|\sum_{\lambda=1}^{D_m}\varphi_\lambda^2\right\|_{\infty} \leq \phi_1 D_m$*.*

c. *There exists a constant* ϕ_2 *such that for any* $m \in \mathbb{N}$*, there exists an orthonormal basis* $(\psi_\lambda)_{\lambda\in\Lambda}$ *of* S_m *such that*

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

d. *For any function* $t \in \mathscr{B}^{\alpha}_{2,\infty}$*,*

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

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

Assumptions A4 a, b and d are quite standard (see for instance Comte et al. (2007)). Assumption A4 c is an additional assumption, which is not satisfied by the trigonometric basis, for instance. However, it is fulfilled by subspaces generated by piecewise polynomials of degree r, spline functions of degree r, or compactly supported wavelets.

3.2 Estimator with fixed m

We then define the contrast function

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

and for any $m \in \mathscr{M}_n = \{m, D_m^2 \leq (n\Delta)/\ln(n)\}$, we consider the estimator

$$\hat{\xi}^4_m = \arg\min_{t\in S_m}\gamma_n(t). \tag{3}$$

The easiest bound is obtained for the empirical risk.

Proposition 6. *Under Assumptions A1-A4, for any m such as $D_m \in \mathscr{M}_n$:*

$$\mathbb{E}\left(\left\|\hat{\xi}^4_m - \xi^4_A\right\|^2_n\right) \leq \left\|\xi^4 - \xi^4_{m,\pi}\right\|^2_\pi + 12\Xi\frac{D_m}{n\Delta} + C\Delta + \frac{C'}{n}$$

where $\xi_A(x) = \xi(x)\mathbf{1}_{x\in A}$ and

$$\Xi = \min\left(\xi_1^8 I_8, \frac{\phi_1}{\pi_0}\mathbb{E}_\pi\left(\xi^8(X_\Delta)\right) I_8\right).$$

The empirical risk is not very intuitive: indeed, the empirical norm is random, and the stationary measure π is unknown. Nevertheless, the L^2-risk can also be bounded, even if the bound is less sharp.

Corollary 7. *Under Assumptions A1-A4, for any m such as $D_m \leq \mathscr{D}_n = \sqrt{n\Delta/\ln(n)}$:*

$$\mathbb{E}\left(\left\|\hat{\xi}^4_m - \xi^4\right\|^2_\pi\right) \leq 5\left\|\xi^4 - \xi^4_{m,\pi}\right\|^2_\pi + 48\Xi\frac{D_m}{n\Delta} + 4C\Delta.$$

Moreover,

$$\mathbb{E}\left(\left\|\hat{\xi}^4_m - \xi^4\right\|^2_{L^2(A)}\right) \leq 5\frac{\pi_1}{\pi_0}\left\|\xi^4 - \xi^4_m\right\|^2_{L^2(A)} + 48\frac{\Xi}{\pi_0}\frac{D_m}{n\Delta} + \frac{4C}{\pi_0}\Delta.$$

3.3 Adaptive estimator

We obtain a sequence of estimators of ξ^4, and have to choose the best. If ξ^4 belongs to the Besov space $\mathscr{B}^\alpha_{2,\infty}$, then the bias term $\left\|\xi^4 - \xi^4_m\right\|^2_{L^2}$ is smaller than $D_m^{-2\alpha}$ and the risk is optimal for $D_m = (n\Delta)^{-1/(1+2\alpha)}$. The optimal estimator satisfies:

$$\mathbb{E}\left(\left\|\hat{\xi}^4_{m_{opt}} - \xi^4_A\right\|^2_n\right) \lesssim (n\Delta)^{-2\alpha/(2\alpha+1)} + \Delta.$$

The rate $(n\Delta)^{-2\alpha/(2\alpha+1)}$ is standard for nonparamatric estimation of ξ. As the regularity of ξ^4 is unknown, we have to choose the best estimator automatically. To this aim, we introduce the penalty function $pen(m) = \kappa\Xi D_m/(n\Delta)$ and choose the best dimension m:

$$\hat{m} = \arg\min_{m\in\mathscr{M}_n}\{\gamma_n(\hat{\xi}^4_m) + pen(m).\}$$

Our estimator achieves the best rate of convergence (up to a multiplicative constant).

Theorem 8. *Under Assumptions A1-A3, for any $\kappa \geq \kappa_0$ with κ_0 a universal constant,*

$$\mathbb{E}\left(\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_A\right\|^2_n\right) \lesssim \inf_{m\in\mathscr{M}_n}\left(\left\|\xi^4 - \xi^4_{m,\pi}\right\|^2_\pi + \Xi\frac{D_m}{n\Delta}\right) + \frac{1}{n\Delta} + \Delta.$$

The bound of the L^2-risk of the estimator is the same:

Corollary 9.

$$\mathbb{E}\left(\left\|\hat{\xi}^4_{\hat{m}} - \xi^4\right\|^2_{L^2(A)}\right) \lesssim \inf_{m\in\mathscr{M}_n}\left(\left\|\xi^4 - \xi^4_m\right\|^2_{L^2(A)} + \Xi\frac{D_m}{n\Delta}\right) + \frac{1}{n\Delta} + \Delta.$$

3.4 Robust estimator

The constant $\Xi = \min(\xi_1^8 I_8, \frac{\phi_1}{\pi_0}\mathbb{E}_\pi\left(\xi^8(X_\Delta)\right) I_8)$ is unknown, but can be estimated to construct a robust estimator. Let us first estimate $h(x) = \xi^8(x)I_8$. Let us note

$$V_{k\Delta} := \frac{(X_{(k+1)\Delta} - X_{k\Delta})^8}{\Delta} = \xi^8(X_{k\Delta})I_8 + G_{k\Delta} + H_{k\Delta},$$

where $G_{k\Delta}$ is small and $H_{k\Delta}$ is centred:

$$\begin{aligned}\Delta G_{k\Delta} &= (B_{k\Delta} + Z_{k\Delta} + J_{k\Delta})^8 - (Z_{k\Delta} + J_{k\Delta})^8 + \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^8 \middle| \mathscr{F}_{k\Delta}\right) - \xi^8(X_{k\Delta})I_8\\ \Delta H_{k\Delta} &= (Z_{k\Delta} + J_{k\Delta})^8 - \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^8 \middle| \mathscr{F}_{k\Delta}\right).\end{aligned}$$

The following lemma bounds the moments of G and H:

Lemma 10.

$$\mathbb{E}\left(G_{k\Delta}^2 \middle| \mathscr{F}_{k\Delta}\right) \lesssim \Delta, \quad \mathbb{E}\left(G_{k\Delta}^4 \middle| \mathscr{F}_{k\Delta}\right) \lesssim \Delta,$$
$$\mathbb{E}\left(H_{k\Delta} \middle| \mathscr{F}_{k\Delta}\right) = 0, \quad \mathbb{E}\left(H_{k\Delta}^2 \middle| \mathscr{F}_{k\Delta}\right) \lesssim \frac{1}{\Delta}, \quad \mathbb{E}\left(H_{k\Delta}^4 \middle| \mathscr{F}_{k\Delta}\right) \lesssim \frac{1}{\Delta^3}.$$

As previously, we consider the contrast function $\gamma_{2,n}(t) = \frac{1}{n}\sum_{k=1}^n (V_{k\Delta} - t(X_{k\Delta}))^2$ and the estimator $\hat{h}_m = \arg\min_{t\in S_m} \gamma_{2,n}(t)$. This estimator is consistent:

Lemma 11.

$$\mathbb{E}\left(\left\|h_A - \hat{h}_m\right\|_n^2\right) \lesssim \|h - h_m\|_{L^2(A)}^2 + \frac{D_m}{n\Delta} + \Delta$$

where $h_A(x) = h(x)\mathbf{1}_{x\in A}$.

Let us set $\hat{h}_1 = \sup_{x\in A} |\hat{h}_{\ln(n)}(x)|$.

Let us now estimate the second term, $\phi_1/\pi_0\mathbb{E}_\pi\left(\xi^8(X_\Delta)\right) I_8$. We consider the rectangular kernel $K(x) := \mathbf{1}_{|x|\le 1/2}$ and set

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

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

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

Finally, we obtain an estimator of the constant Ξ:

$$\hat{\Xi} := \min\left(\hat{h}_1, \frac{\phi_1}{\hat{\pi}_0}\bar{V}_n\right) + \Delta^{1/4} + 1/\ln(n)$$

with $\bar{V}_n = \frac{1}{n}\sum_{k=1}^n V_{k\Delta}$.The estimator $\hat{\Xi}$ is consistent:

Lemma 12. *Under Assumptions A1-3, we have that*

$$\mathbb{P}\left(|\hat{\Xi} - \Xi| \ge C/\ln(n) + \Delta^{1/4}\right) \lesssim n^{-4}.$$

Let us set

$$\tilde{m} = \arg \inf_{m \in \mathscr{M}_n} \gamma_n(\hat{\xi}_m) + \kappa \hat{\Xi} \frac{D_m}{n\Delta}.$$

The resulting robust estimator $\hat{\xi}_{\tilde{m}}$ is also optimal (up to a multiplicative constant):

Corollary 13. *The robust estimator satisfies the oracle inequality: Under Assumptions A1-A3,*

$$\mathbb{E}\left(\left\|\hat{\xi}^4_{\tilde{m}} - \xi^4_A\right\|_n^2\right) \lesssim \inf_{m \in \mathscr{M}_n} \left(\left\|\xi^4 - \xi^4_{m,\pi}\right\|_\pi^2 + \Xi \frac{D_m}{n\Delta}\right) + \frac{1}{n\Delta} + \Delta.$$

Moreover,

$$\mathbb{E}\left(\left\|\hat{\xi}^4_{\tilde{m}} - \xi^4\right\|^2_{L^2(A)}\right) \lesssim \inf_{m \in \mathscr{M}_n} \left(\left\|\xi^4 - \xi^4_m\right\|^2_{L^2(A)} + \Xi \frac{D_m}{n\Delta}\right) + \frac{1}{n\Delta} + \Delta.$$

3.5 Comparison with the estimator obtained by thresholding

Park and Wang (2021) and Mancini and Renò (2011) construct an estimator of ξ^2 by taking the second infinitesimal moments and keeping only the increments big enough: they consider

$$\frac{(X_{(k+1)\Delta} - X_{k\Delta})^2}{\Delta} \mathbf{1}_{|X_{(k+1)\Delta} - X_{k\Delta}| \geq \Delta^{1/2-\varepsilon}}.$$

Then, they construct a kernel estimator of ξ^2, in the case where the jumps have finite intensity and ξ^2 belongs to $\mathscr{C}^2$. Their estimator is asymptotically normal, and the risk for the optimal bandwith is bounded by

$$\mathbb{E}\left((\hat{\xi}^2(x) - \xi^2(x))^2\right) \lesssim (n\Delta)^{-4/5} + \Delta.$$

Similarly, Schmisser (2019) estimate first the second infinitesimal moment $g = \sigma^2 + \xi^2$, then σ^2 by thresholding. An estimator of ξ^2 can be obtained by substraction. If g and σ^2 belong to a Besov spaces of regularity α_g and α_{σ^2}, the risk of the adaptive estimator $\tilde{\xi}^2$ is bounded by:

$$\mathbb{E}\left(\left\|\tilde{\xi}^2 - \xi^2\right\|^2_{L^2(A)}\right) \lesssim (n\Delta)^{2\alpha_g/(2\alpha_g+1)} + n^{-2\alpha_\sigma/(2\alpha_\sigma+1)} + \Delta^{1-\beta/2}$$

where β is the Blumenthal-Getoor index (that is, $\beta = \inf_{\alpha \geq 0}\{\int (|z|^\alpha \wedge 1)\nu(dz) < \infty\}$). If the Lévy process has finite activity, then $\beta = 0$. If the intensity is finite, $\beta \leq 1$. For both estimators, it is assumed than $I_2 = \int z^2 \nu(dz) = 1$.

The problem is, the first estimator can only be calculated for finite intensity processes, and is not adaptive. Whereas the rate of convergence of the second depends on the Blumenthal-Getoor index, and on the regularity of σ^2.

In order to compare those bounds to the risk of our estimator $\hat{\xi}_{\tilde{m}}$, we can remark that

$$\left\|\sqrt{\hat{\xi}^4_{\tilde{m}}} - \xi^2\right\|^2_{L^2(A)} \leq \frac{\left\|\hat{\xi}^4_{\tilde{m}} - \xi^4_A\right\|^2_{L^2(A)}}{\min_A \xi^2(x)}.$$

Therefore, we get that, if I_4 is known:

$$\mathbb{E}\left(\left\|\sqrt{\hat{\xi}^4_{\tilde{m}}} - \xi^2\right\|^2_{L^2(A)}\right) \lesssim (n\Delta)^{-2\alpha/(2\alpha+1)} + \Delta.$$

This bound is more general than the rate obtained by Park and Wang (2021), and sharper than the one obtained by Schmisser (2019). In particular, if the jump coefficient ξ is more regular than the diffusion coefficient σ^2, the rate is better. Moreover, if the jumps have infinite activity, the remainder term is smaller.

However, we assume in this paper than the fourth moment of the Levy measure, $I_4 = \int z^4 \nu(dz)$, is equal to 1, which is not the standard assumption. If no assumptions are made on the moments of ν, the function ξ is not identifiable. So it can only be estimated up to a multiplicative constant. Even if we want to impose $I_2 = 1$, it is interessant to know the shape of ξ.

4 Simulations

We choose to consider the linear subspaces S_m generated by the spline functions of degree r (that is, piecewise polynomials of degree r and $\mathscr{C}^{r-1}$. It is also possible to consider piecewise polynomials of degree r or subsepaces generated by wavelets. We choose the spline functions because they are explicits, contrary to wavelets functions, and because the resulting estimator is continuous if $r \geq 1$, which is not the case for piecewise polynomials.

We compute the series of estimators $\hat{\xi}^4_{m,r}$ for r between 0 and 4 and m between 0 and $\max(\sqrt{n\Delta/\ln(n)}, 7)$ (for $m = 7$, we already have $D_m = 128$, for m bigger, the computations become too slow). To obtain the adaptive estimator, we minimise first with regard to m, then to r.

We choose $\kappa = 8$ for the simulations. We consider 3 different models. For each model, we let n and Δ vary and realize 200 simulations for each value of (n, Δ). We compute the robust estimator $\hat{\xi}^4_{\tilde{m},\tilde{r}}$, the selected dimension $D_{\tilde{m},\tilde{r}}$, the estimated penalty constant $\hat{\Xi}$ and the L^2-error

$$\text{risk} = \sum_{j=0}^{N-1} \left(\hat{\xi}^4_{\tilde{m},\tilde{r}}(x_j) - \xi^4(x_j)\right)^2 (x_{j+1} - x_j) \simeq \int_A \left(\hat{\xi}^4_{\tilde{m},\tilde{r}}(x) - \xi^4(x)\right)^2 dx$$

for $(x_0, x_1, \ldots, x_N)$ a regular subdivision of A. We also compute the empirical error for each $\hat{\xi}_{m,r}$. Then we deduce the dimension (m_{opt}, r_{opt}) that minimizes the empirical error (denoted by err_{min}). To compare this oracle estimator to the robust one, we compute an oracle *or* which is the mean of the logarithmic ratio between the two errors. To compare our estimator to the one obtained by substraction, we also compute the error obtained by the square root of our estimator, $\text{risk1} = \left\| \sqrt{\hat{\xi}^4_{\tilde{m},\tilde{r}}} - \xi^2_A \right\|^2_{L^2}$ and the risk of the previous estimator obtained by thresholding and substraction in Schmisser (2019): $\text{risk2} := \left\| \tilde{\xi}^2 - \xi^2_A \right\|^2_{L^2}$. We add the empirical risks of the estimators of the diffusion coefficient σ^2 and the second moment $g(x) = \sigma^2(x) + \xi^2(x)$, also computed in Schmisser (2019).

For each model, we provide two tables. The first one show some results for the estimator $\hat{\xi}^4_{\tilde{m},\tilde{r}}$: the mean and standard deviation of the risk of the robust estimator $\hat{\xi}^4_{\tilde{m},\tilde{r}}$ (in fact, the square root of these quantity, to have a norm), a criteria of comparison between the risks of the robust estimator and the 'oracle' estimator; or = mean(ln(risk/err)), the mean and the standard deviation of the estimated penalty constant $\hat{\Xi}$ and the mean and the standard deviation of the selected dimension $D_{\tilde{m},\tilde{r}}$.

On the second table, we write the mean and standard deviations of risk1 and risk2 (more precisely, the square roots of those quantities) in order to compare these two estimators. We also

add the means and standard deviation for the risk of the estimators of g and σ^2.

4.1 Models

In the first two models, L_t is simply a compound Poisson Process:

$$\int_0^t \xi(X_{s^-})dL_s = \sum_{k=1}^{N_t} \xi(X_{t_k^-})\zeta_k$$

where N_t is a Poisson Process of intensity λ, $t_1, \ldots, t_{N_t}$ the times of the jumps, and $(\zeta_k)_{k\geq 0}$ are independent, identically distributed of binomial law: $\mathbb{P}(\zeta_k = \pm 1) = 1/2$.

Model 1: Ornstein-Uhlenbeck Let us consider the process given by the equation:

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

with binomial jumps: $\mathbb{P}(\zeta = 1) = \mathbb{P}(\zeta = -1) = 0.5$ and the intensity of the Poisson process is 1. We have that $I_2 = 1$, $I_4 = 1$, $I_6 = 1$, $I_8 = 1$.

Model 2: non-constant coefficient

$$dX_t = -2X_t dt + \sqrt{2 + 0.5\sin(\pi X_t)}dW_t + \sqrt{\frac{X_{t^-}^2 + 3}{X_{t^-}^2 + 1}}dL_t$$

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

Model 3: nearly stable processes

$$dX_t = -2X_t + \sqrt{\frac{3 + X_t^2}{1 + X_t^2}}dW_t + \sqrt{2 + 0.5\sin(X_{t^-})}dL_t$$

with L_t a nearly stable process: $\nu(dz) = \frac{2-\beta}{2|z|^{1+\beta}}\mathbf{1}_{|z|\leq 1}$. β is the Blumenthal-Getoor index of the process. We take $\beta = 1/2$ and we get that $I_2 = 1$, $I_4 = 3/7$, $I_6 = 3/11$ and $I_8 = 3/15$.

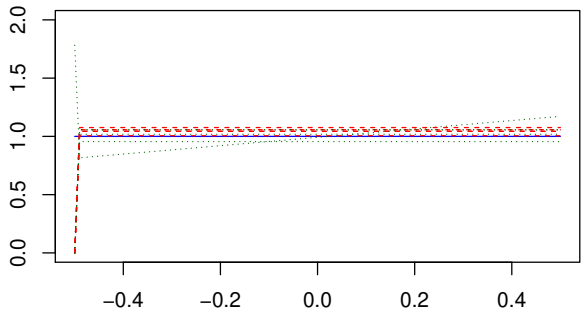
4.2 Results

The risk of the adaptive estimator $\hat{\xi}^4_{\hat{m},\hat{r}}$ decreases when n increases (for fixed Δ). When the period of observation ($n\Delta$) is fixed, the risk seems to be either stable, either decreasing with Δ, especially when Δ is not small enough. When $n\Delta \leq 100$, our estimator nearly always select the smallest dimension, whereas for $n\Delta \geq 10^3$, the algorithm can select more complex models. Our estimator is truly adaptive, and the oracle seems to be quite small.

For the estimated penalty constant $\hat{\Xi}$, it can be noted than when n increases, the variance decreases as well: $\hat{\Xi}$ converges to a constant depending on Δ. When Δ decreases, this constant decreases as well. This is because

$$\bar{V}_n \to \mathbb{E}(V_\Delta) \simeq \Delta^{-1}\mathbb{E}\left((J_\Delta + Z_{k\Delta})^8\right) \geq \mathbb{E}\left(\xi^8(X_\Delta)\right) I_8.$$

Figure 1: Model 1
$dX_t = -2X_t dt + dW_t + dL_t$ with binomial jumps.



−− estimator $\sqrt{\hat{\xi}^4_{\tilde{m},\tilde{r}}}$ ... estimator $\tilde{\xi}^2 = \hat{g}_{\hat{m}} - \hat{\sigma}^2_{\tilde{m}}$.
$n = 10^5, \Delta = 10^{-2}$.
Risk of $\hat{\xi}^4_{\tilde{m},\tilde{r}}$

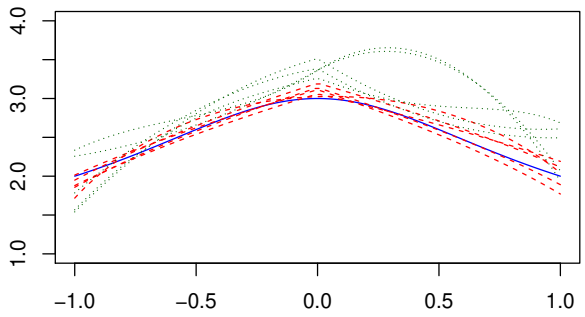
n	Δ	risk			$\hat{\Xi}$		$D_{\tilde{m},\tilde{r}}$	
		mean	sd	or	mean	sd	mean	sd
10^3	10^{-1}	0.67	0.75	0	27.5	43.3	1	0
10^4	10^{-1}	0.61	0.36	0	27.7	24.0	1	0
10^5	10^{-1}	0.60	0.21	0.0014	22.0	11.7	1.06	0.42
10^3	10^{-2}	0.50	0.92	0	4.54	7.99	1	0
10^4	10^{-2}	0.16	0.21	0	4.44	6.22	1	0
10^5	10^{-2}	0.087	0.089	0	3.86	2.49	1	0
10^4	10^{-3}	0.37	0.54	0	2.76	3.50	1	0
10^5	10^{-3}	0.11	0.14	0	2.13	2.42	1	0

Comparison of the estimators $\sqrt{\hat{\xi}^4_{\tilde{m},\tilde{r}}}$ and $\tilde{\xi}^2$.

n	Δ	risk1		risk2		risk g		risk σ^2	
		mean	sd	mean	sd	mean	sd	mean	sd
10^3	10^{-1}	0.28	0.29	0.23	0.22	0.34	0.24	0.42	0.15
10^4	10^{-1}	0.27	0.15	0.20	0.12	0.33	0.14	0.42	0.091
10^5	10^{-1}	0.27	0.12	0.20	0.084	0.25	0.083	0.41	0.049
10^3	10^{-2}	0.22	0.32	0.37	0.43	0.40	0.42	0.12	0.075
10^4	10^{-2}	0.075	0.094	0.15	0.14	0.23	0.14	0.11	0.039
10^5	10^{-2}	0.042	0.043	0.14	0.11	0.21	0.050	0.093	0.067
10^4	10^{-3}	0.18	0.23	0.36	0.43	0.40	0.42	0.10	0.043
10^5	10^{-3}	0.056	0.069	0.15	0.15	0.23	0.15	0.099	0.039

Figure 2: Model 2

$dX_t = -2X_t dt + \sqrt{2 + 0.5\sin(\pi X_t)}dW_t + \sqrt{\frac{X_{t-}^2+3}{X_{t-}^2+1}}dL_t$ with binomial jumps.



−− estimator $\sqrt{\hat{\xi}^4_{\tilde{m},\tilde{r}}}$ ... estimator $\tilde{\xi}^2 = \hat{g}_{\hat{m}} - \hat{\sigma}^2_{\hat{m}}$.

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

Risk of the estimator $\hat{\xi}^4_{\tilde{m},\tilde{r}}$.

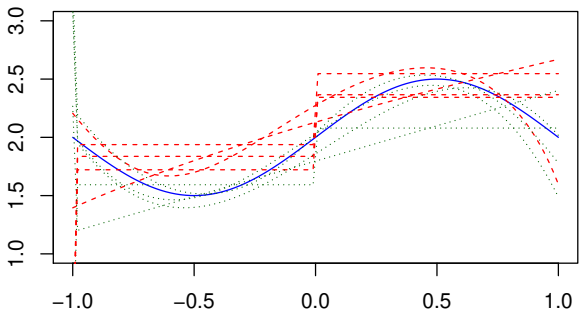
n	Δ	risk			$\hat{\Xi}$		$D_{\tilde{m},\tilde{r}}$	
		mean	sd	or	mean	sd	mean	sd
10^3	10^{-1}	2.87	3.20	0.19	614	870	1	0
10^4	10^{-1}	2.29	1.33	0.33	394	240	1.01	0.07
10^5	10^{-1}	2.10	1.02	0.16	338	141	2.35	1.53
10^3	10^{-2}	3.63	4.92	0.11	128	189	1	0
10^4	10^{-2}	1.86	1.25	0.70	131	174	1	0
10^5	10^{-2}	0.82	0.92	0.44	101	24	3.55	1.07
10^4	10^{-3}	3.25	3.51	0.10	91	50	1	0
10^5	10^{-3}	1.83	1.07	0.94	89	72	1.01	0.07

Comparison of the estimators $\sqrt{\hat{\xi}^4_{\tilde{m},\tilde{r}}}$ and $\tilde{\xi}^2$.

n	Δ	risk1		risk2		risk g		risk σ^2	
		mean	sd	mean	sd	risk	sd	risk	sd
10^3	10^{-1}	0.46	0.54	0.84	0.73	0.99	0.63	1.44	0.29
10^4	10^{-1}	0.37	0.26	0.77	0.38	0.95	0.39	1.43	0.18
10^5	10^{-1}	0.46	0.30	0.78	0.29	0.85	0.32	1.42	0.10
10^3	10^{-2}	0.58	0.75	1.24	1.59	1.25	1.43	0.55	0 0.31
10^4	10^{-2}	0.21	0.25	0.67	0.55	0.71	0.39	0.456	0.19
10^5	10^{-2}	0.48	0.38	0.54	0.35	0.48	0.38	0.43	0.084
10^4	10^{-3}	0.53	0.60	1.21	1.28	1.21	1.28	0.11	0.13
10^5	10^{-3}	0.20	0.22	0.71	0.42	0.71	0.42	0.033	0.026

Figure 3: Model 3:

$$dX_t = -2X_t dt + \sqrt{\frac{X_{t^-}^2+3}{X_{t^-}^2+1}}dW_t + \sqrt{2+0.5\sin(\pi X_{t^-})}dL_t \quad \nu(z) = \frac{1.5}{2|z|^{1.5}}\mathbf{1}_{|z|\le 1}$$



−− estimator $\sqrt{\hat{\xi}^4_{\tilde{m},\tilde{r}}}$ ... estimator $\tilde{\xi}^2 = \hat{g}_{\hat{m}} - \hat{\sigma}^2_{\hat{m}}$.

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

Estimation of $\hat{\xi}^4_{\tilde{m},\tilde{r}}$.

n	Δ	risk			$\hat{\Xi}$		$D_{\tilde{m},\tilde{r}}$	
		mean	sd	or	mean	sd	mean	sd
10^3	10^{-1}	3.91	2.83	0.0098	182	222	1	0
10^4	10^{-1}	3.75	1.57	0.015	135	84	1.04	0.18
10^5	10^{-1}	3.73	0.88	0.0081	117	45	1.58	0.55
10^3	10^{-2}	1.41	2.24	0.23	23.0	61.9	1	0
10^4	10^{-2}	0.86	0.66	0.67	17.3	18.4	1.03	0.17
10^5	10^{-2}	0.61	0.35	0.20	15.1	8.1	2.11	0.97
10^4	10^{-3}	1.07	1.33	0.29	12.8	27.9	1	0
10^5	10^{-3}	0.68	0.45	1.16	10.2	17	1.14	0.35

Comparison of the estimators $\sqrt{\hat{\xi}^4_{\tilde{m},\tilde{r}}}$ and $\tilde{\xi}^2$.

n	Δ	risk1		risk2		risk g		risk σ^2	
		mean	sd	mean	sd	mean	sd	mean	sd
10^3	10^{-1}	1.60	1.01	1.33	0.77	0.97	0.58	1.99	0.34
10^4	10^{-1}	1.58	0.61	1.30	0.47	0.94	0.36	1.99	0.21
10^5	10^{-1}	1.66	0.58	1.28	0.26	0.86	0.32	1.99	0.13
10^3	10^{-2}	0.59	0.86	0.93	0.97	0.84	0.72	0.82	0.37
10^4	10^{-2}	0.36	0.42	0.83	0.49	0.65	0.23	0.78	0.23
10^5	10^{-2}	0.59	0.36	0.71	0.31	0.24	0.28	0.76	0.11
10^4	10^{-3}	0.48	0.62	0.84	0.69	0.84	0.70	0.11	0.11
10^5	10^{-3}	0.33	0.50	0.64	0.30	0.65	0.30	0.057	0.038

This is not a drawback, as the penalty should be the variance of the centred terms, more precisely

$$\Xi^{\Delta} = \Delta \min\left(\left\| \mathrm{Var}\left(F_{k\Delta}\,|\mathscr{F}_{k\Delta}\right)\right\|_{\infty}; \frac{\phi_1}{\pi_0}\mathrm{Var}_{\pi}\left(F_{\Delta}\right)\right).$$

Ξ is an estimator of this quantity, that converges when $\Delta \to 0$ and $n\Delta \to \infty$: by Proposition 6, $\Xi^{\Delta} = \Xi + C\Delta^{1/2}$. In the simple case where the jump and diffusion coefficients are constants, then

$$\begin{aligned}\Xi^{\Delta} &= \Delta^{-1}\,\mathrm{Var}\left(F_{k\Delta}\right) = \Delta^{-1}\,\mathrm{Var}\left((\xi L_{k\Delta} + \sigma Z_{k\Delta})^4\right)\\ &= \Delta^{-1}\left(\xi^8\mathbb{E}\left(L_{k\Delta}^8\right) + 28\sigma^2\xi^6\mathbb{E}\left(L_{k\Delta}^6 Z_{k\Delta}^2\right) - \xi^8(\mathbb{E}\left(L_{k\Delta}^4\right))^2\right) + O(\Delta^2)\\ &= \xi^8 I_8 + \Delta\left(\xi^8(34I_4^2 + 28I_2I_6) + \sigma^2\xi^6 28 I_6\right) + O(\Delta^2).\end{aligned}$$

When Δ is not small enough, the term in Δ is not negligeable, neither is the term in $O(\Delta^2)$.

Our estimator $\sqrt{\hat{\xi}^4_{\tilde{m}}}$ is better than $\tilde{\xi}^2$ as soon as the path of discretisation Δ is small enough ($\Delta \leq 10^{-2}$), even for compound Poisson processes.

5 Proof

5.1 Proof of Lemma 5 and 10

According to Proposition 1, we have that

$$\mathbb{E}\left(B_{k\Delta}^{2p}|\mathscr{F}_{k\Delta}\right) \lesssim \Delta^{2p}, \quad \mathbb{E}\left(Z_{k\Delta}^{2p}|\mathscr{F}_{k\Delta}\right) \lesssim \Delta^{p}, \quad \text{and} \quad \mathbb{E}\left(J_{k\Delta}^{2p}|\mathscr{F}_{k\Delta}\right) \lesssim \Delta I_{2p} + O(\Delta^2). \tag{4}$$

Let us set

$$\begin{aligned}\tilde{B}_{k\Delta} &= b(X_{k\Delta})\Delta, & \bar{B}_{k\Delta} &= B_{k\Delta} - \tilde{B}_{k\Delta},\\ \tilde{Z}_{k\Delta} &= \sigma(X_{k\Delta})(W_{(k+1)\Delta} - W_{k\Delta}), & \bar{Z}_{k\Delta} &= Z_{k\Delta} - \tilde{Z}_{k\Delta},\\ \tilde{J}_{k\Delta} &= \xi(X_{k\Delta})(L_{(k+1)\Delta} - L_{k\Delta}), & \bar{J}_{k\Delta} &= J_{k\Delta} - \tilde{J}_{k\Delta}.\end{aligned}$$

Conditionnally to $\mathscr{F}_{k\Delta}$, $\tilde{B}_{k\Delta}$, $\tilde{J}_{k\Delta}$ and $\tilde{Z}_{k\Delta}$ are independent. Moreover, we have that

$$\mathbb{E}\left(\bar{B}_{k\Delta}^{2p}\middle|\,\mathscr{F}_{k\Delta}\right) \lesssim \Delta^{2p+1}, \quad \mathbb{E}\left(\bar{Z}_{k\Delta}^{2p}\middle|\,\mathscr{F}_{k\Delta}\right) \lesssim \Delta^{p+1}, \quad \mathbb{E}\left(\bar{J}_{k\Delta}^{2p}\middle|\,\mathscr{F}_{k\Delta}\right) \lesssim \Delta^{2} I_{2p}. \tag{5}$$

Indeed, by Hölder inequality and Lemma 2, as the drift function b is Lipschitz,

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

By Proposition 1 and the same arguments, we get:

$$\begin{aligned}\mathbb{E}\left(\bar{Z}_{k\Delta}^{2p}\right) &\lesssim \mathbb{E}\left(\left(\int_{k\Delta}^{(k+1)\Delta}(\sigma(X_s)-\sigma(X_{k\Delta}))^2 ds\right)^p\right)\\ &\lesssim \Delta^{p-1}\int_{k\Delta}^{(k+1)\Delta}\mathbb{E}\left((\sigma(X_s)-\sigma(X_{k\Delta})^{2p}\right) ds \lesssim \Delta^{p+1}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left(\bar{J}_{k\Delta}^{2p}\middle|\mathscr{F}_{k\Delta}\right) &\lesssim \mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta}\left(\xi(X_{s^-})-\xi(X_{k\Delta})\right)^{2p} ds\middle|\mathscr{F}_{k\Delta}\right) I_{2p}\\ &\quad+\mathbb{E}\left(\left(\int_{k\Delta}^{(k+1)\Delta}(\xi(X_{s^-})-\xi(X_{k\Delta}))^2 ds\right)^p\middle|\mathscr{F}_{k\Delta}\right) I_2^p\\ &\lesssim \Delta^2 I_{2p}+\Delta^{p+1} I_2^p.\end{aligned}$$

We first bound the moments of the small terms, $E_{k\Delta}$ in Lemma 5 and $G_{k\Delta}$ in Lemma 10.

Bound of $E_{k\Delta}$. Let us decompose the bias term

$$\Delta E_{k\Delta} = (B_{k\Delta}+Z_{k\Delta}+J_{k\Delta})^4-(Z_{k\Delta}+J_{k\Delta})^4+\mathbb{E}\left((Z_{k\Delta}+J_{k\Delta})^4\middle|\mathscr{F}_{k\Delta}\right)-\Delta\xi^4(X_{k\Delta}I_4).$$

We can write $E_{k\Delta}=E_{k\Delta}^{(1)}+E_{k\Delta}^{(2)}+E_{k\Delta}^{(3)}+E_{k\Delta}^{(4)}$ where

$$\begin{aligned}\Delta E_{k\Delta}^{(1)} &= \mathbb{E}\left((J_{k\Delta}+Z_{k\Delta})^4-(\tilde{J}_{k\Delta}+\tilde{Z}_{k\Delta})^4\middle|\mathscr{F}_{k\Delta}\right)\\ \Delta E_{k\Delta}^{(2)} &= \mathbb{E}\left((\tilde{J}_{k\Delta}+\tilde{Z}_{k\Delta})^4-\tilde{J}_{k\Delta}^4\middle|\mathscr{F}_{k\Delta}\right)\\ \Delta E_{k\Delta}^{(3)} &= \mathbb{E}\left(\tilde{J}_{k\Delta}^4\middle|\mathscr{F}_{k\Delta}\right)-\Delta\xi^4(X_{k\Delta})I_4\\ \Delta E_{k\Delta}^{(4)} &= (J_{k\Delta}+Z_{k\Delta}+B_{k\Delta})^4-(J_{k\Delta}+Z_{k\Delta})^4.\end{aligned}$$

The first three terms are conditional expectations, the last is random. For the first term, we factorize the difference of two squares: indeed, for any $x,y\in\mathbb{R}^2$, $x^4-y^4=(x-y)(x+y)(x^2+y^2)$. Then

$$\Delta E_{k\Delta}^{(1)} = \mathbb{E}\left((\bar{J}_{k\Delta}+\bar{Z}_{k\Delta})\left(J_{k\Delta}+\tilde{J}_{k\Delta}+Z_{k\Delta}+\tilde{Z}_{k\Delta}\right)\left((J_{k\Delta}+Z_{k\Delta})^2+(\tilde{J}_{k\Delta}+\tilde{Z}_{k\Delta})^2\right)\middle|\mathscr{F}_{k\Delta}\right).$$

Then, as $\mathbb{E}\left((\bar{J}_{k\Delta}+\bar{Z}_{k\Delta})^2\right)\lesssim\Delta^2$, we obtain by Cauchy-Schwarz inequality and (4):

$$\Delta E_{k\Delta}^{(1)} \lesssim \left(\Delta^2\right)^{1/2}\Delta^{1/4}\Delta^{1/4}\simeq\Delta^{3/2}.$$

Moreover, $E_{k\Delta}^{(2)}$ can be written $\Delta E_{k\Delta}^{(2)}=\sum_{j=1}^4 C_4^j\mathbb{E}\left(\tilde{Z}_{k\Delta}^j\tilde{J}_{k\Delta}^{4-j}\middle|\mathscr{F}_{k\Delta}\right)$. As $\tilde{Z}_{k\Delta}$ and $\tilde{J}_{k\Delta}$ are centred and conditionnally independent, by (4), we obtain:

$$\Delta E_{k\Delta}^{(2)} \lesssim \sum_{j=1}^4\Delta^{j/2}\left(\Delta\mathbf{1}_{j<4}+\mathbf{1}_{j=4}\right)\lesssim\Delta^{3/2}.$$

Furthermore, by Lemma 3, $\mathbb{E}\left(L_{k\Delta}^4\right) = \Delta I_4 + O(\Delta^2)$ and therefore:

$$\Delta E_{k\Delta}^{(3)} = \xi^4(X_{k\Delta})(\mathbb{E}\left(L_\Delta^4\right) - \Delta I_4) \lesssim \Delta^2.$$

As $\Delta E_{k\Delta}^{(4)} = \sum_{j=1}^4 C_4^j B_{k\Delta}^j (Z_{k\Delta} + J_{k\Delta})^{4-j}$, we get that

$$\mathbb{E}\left(\left(\Delta E_{k\Delta}^{(4)}\right)^{2p}\right) \lesssim \sum_{j=1}^4 \mathbb{E}\left(B_{k\Delta}^{2jp}(Z_{k\Delta} + J_{k\Delta})^{8p-2pj}\right). \tag{6}$$

By Cauchy-Schwartz inequality and (4), for $2 \leq j \leq 4$,

$$\begin{aligned}\mathbb{E}\left(B_{k\Delta}^{2jp}(Z_{k\Delta} + J_{k\Delta})^{8p-2pj}\right) &\leq \left(\mathbb{E}\left(B_{k\Delta}^{4jp}\right)\mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^{16p-4pj}\right)\right)^{1/2} \\ &\leq \Delta^{2jp}\left(\mathbf{1}_{j=4} + \Delta^{1/2}\mathbf{1}_{j<4}\right) \leq \Delta^{4p+1/2}. \end{aligned} \tag{7}$$

Moreover, as $B_{k\Delta} = \Delta b(X_{k\Delta}) + \bar{B}_{k\Delta}$, by Equations (4) and (5), for $j = 1$ we obtain:

$$\begin{aligned}\mathbb{E}\left(B_{k\Delta}^{2p}(Z_{k\Delta} + J_{k\Delta})^{6p}\right) &\lesssim \Delta^{2p}\mathbb{E}\left(b(X_{k\Delta})^{2p}\mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^{6p}\middle|\mathscr{F}_{k\Delta}\right)\right) + \mathbb{E}\left(\bar{B}_{k\Delta}^{2p}(Z_{k\Delta} + J_{k\Delta})^{6p}\right) \\ &\lesssim \Delta^{2p+1} + \left(\mathbb{E}\left(\bar{B}_{k\Delta}^{4p}\right)\mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^{12p}\right)\right)^{1/2} \\ &\lesssim \Delta^{2p+1} + \Delta^{2p+1/2}\Delta^{1/2} = \Delta^{2p+1}.\end{aligned}$$

Then $\mathbb{E}\left(\left(\Delta E_{k\Delta}^{(4)}\right)^{2p}\right) \lesssim \Delta^{2p+1}$. As $\mathbb{E}\left(E_{k\Delta}^{2p}\right) \lesssim \mathbb{E}\left(\left(E_{k\Delta}^{(1)} + E_{k\Delta}^{(2)} + E_{k\Delta}^{(3)}\right)^{2p}\right) + \mathbb{E}\left(\left(E_{k\Delta}^{(4)}\right)^{2p}\right)$, we get that, for $p \geq 1$,

$$\mathbb{E}\left(E_{k\Delta}^{2p}\right) \lesssim \Delta.$$

Bound of $G_{k\Delta}$ We recall that

$$\Delta G_{k\Delta} = (B_{k\Delta} + Z_{k\Delta} + J_{k\Delta})^8 - (Z_{k\Delta} + J_{k\Delta})^8 + \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^8\middle|\mathscr{F}_{k\Delta}\right) - \Delta\xi^8(X_{k\Delta})I_8.$$

As for the bound of $E_{k\Delta}$, we can write $G_{k\Delta} = G_{k\Delta}^{(1)} + G_{k\Delta}^{(2)} + G_{k\Delta}^{(3)} + G_{k\Delta}^{(4)}$ with

$$\begin{aligned}\Delta G_{k\Delta}^{(1)} &:= \mathbb{E}\left((Z_{k\Delta} + J_{k\Delta})^8\middle|\mathscr{F}_{k\Delta}\right) - \mathbb{E}\left((\tilde{Z}_{k\Delta} + \tilde{J}_{k\Delta})^8\middle|\mathscr{F}_{k\Delta}\right) \\ \Delta G_{k\Delta}^{(2)} &:= \mathbb{E}\left((\tilde{Z}_{k\Delta} + \tilde{J}_{k\Delta})^8\middle|\mathscr{F}_{k\Delta}\right) - \mathbb{E}\left(\tilde{J}_{k\Delta}^8\middle|\mathscr{F}_{k\Delta}\right) \\ \Delta G_{k\Delta}^{(3)} &:= \mathbb{E}\left(\tilde{J}_{k\Delta}^8\middle|\mathscr{F}_{k\Delta}\right) - \Delta\xi^8(X_{k\Delta})I_8 \\ \Delta G_{k\Delta}^{(4)} &:= (B_{k\Delta} + Z_{k\Delta} + J_{k\Delta})^8 - (Z_{k\Delta} + J_{k\Delta})^8.\end{aligned}$$

Let us note $JZ := J_{k\Delta} + Z_{k\Delta}$ and $\widetilde{JZ} := \tilde{J}_{k\Delta} + \tilde{Z}_{k\Delta}$. Then by Cauchy-Schwarz, Equations (4) and (5),

$$\begin{aligned}\Delta G_{k\Delta}^{(1)} &= \mathbb{E}\left((\bar{J}_{k\Delta} + \bar{Z}_{k\Delta})(JZ + \widetilde{JZ})\left(JZ^2 + \widetilde{JZ}^2\right)\left(JZ^4 + \widetilde{JZ}^4\right)\middle|\mathscr{F}_{k\Delta}\right) \\ &\lesssim (\Delta^2)^{1/2}\Delta^{1/4}\Delta^{1/8}\Delta^{1/8} \lesssim \Delta^{3/2}.\end{aligned}$$

Moreover, as $\tilde{Z}_{k\Delta}$ and $\tilde{J}_{k\Delta}$ are conditionnaly independant, we get:

$$\Delta G^{(2)}_{k\Delta} = \sum_{j=1}^{8} C_8^j \mathbb{E}\left(\tilde{Z}^j_{k\Delta}\tilde{J}^{8-j}_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right) \lesssim \sum_{j=1}^{8} \Delta^{j/2}\left(\Delta \mathbf{1}_{j<8} + \mathbf{1}_{j=8}\right) \lesssim \Delta^{3/2}.$$

By Lemma 3:

$$\Delta G^{(3)}_{k\Delta} = \xi^8(X_{k\Delta})(\mathbb{E}\left(L^8_\Delta\right) - \Delta I_8) \lesssim C\Delta^2.$$

Finally, by (5)

$$\begin{aligned}\mathbb{E}\left(\left(\Delta G^{(4)}_{k\Delta}\right)^{2p}\middle|\mathscr{F}_{k\Delta}\right) &\lesssim \sum_{j=2}^{8}\mathbb{E}\left(B^{2pj}_{k\Delta}(Z_{k\Delta}+J_{k\Delta})^{16p-2pj}\middle|\mathscr{F}_{k\Delta}\right)\\ &\quad + \tilde{B}^{2p}_{k\Delta}\mathbb{E}\left((Z_{k\Delta}+J_{k\Delta})^{14p}\middle|\mathscr{F}_{k\Delta}\right) + \mathbb{E}\left(\bar{B}^{2p}_{k\Delta}(Z_{k\Delta}+J_{k\Delta})^{14p}\middle|\mathscr{F}_{k\Delta}\right)\\ &\lesssim \Delta^{2p+1}.\end{aligned}$$

Therefore

$$\mathbb{E}\left(G^2_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right) \lesssim \Delta \quad \text{and} \quad \mathbb{E}\left(G^4_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right) \lesssim \Delta.$$

Let us bound the centred terms, $F_{k\Delta}$ in Lemma 5 and $H_{k\Delta}$ in Lemma 10.

Moments of $F_{k\Delta}$. The term $F_{k\Delta} = \frac{1}{\Delta}\left((Z_{k\Delta}+J_{k\Delta})^4 - \mathbb{E}\left((Z_{k\Delta}+J_{k\Delta})^4\middle|\mathscr{F}_{k\Delta}\right)\right)$ is centred. Moreover, by Equation (4), for any $p>0$,

$$\mathbb{E}\left(F^{2p}_{k\Delta}\right) \lesssim \frac{\mathbb{E}\left(J^{8p}_{k\Delta}\right) + \mathbb{E}\left(Z^{8p}_{k\Delta}\right)}{\Delta^{2p}} \lesssim \frac{1}{\Delta^{2p-1}}.$$

Then $\mathbb{E}\left(F^4_{k\Delta}\right) \lesssim \Delta^{-3}$. It remains to bound the quantity $\mathrm{Var}\left(F_{k\Delta}\,|\mathscr{F}_{k\Delta}\right) - \frac{\xi^8(X_{k\Delta})I_8}{\Delta}$. By Equation (4),

$$\begin{aligned}\Delta^2\,\mathrm{Var}\left(F_{k\Delta}\,|\mathscr{F}_{k\Delta}\right) &= \mathrm{Var}\left((Z_{k\Delta}+J_{k\Delta})^4\,|\mathscr{F}_{k\Delta}\right)\\ &= \mathbb{E}\left((Z_{k\Delta}+J_{k\Delta})^8\middle|\mathscr{F}_{k\Delta}\right) - \left(\mathbb{E}\left((Z_{k\Delta}+J_{k\Delta})^4\middle|\mathscr{F}_{k\Delta}\right)\right)^2\\ &= \mathbb{E}\left((Z_{k\Delta}+J_{k\Delta})^8\middle|\mathscr{F}_{k\Delta}\right) + O(\Delta^2).\end{aligned}$$

As

$$\mathbb{E}\left((Z_{k\Delta}+J_{k\Delta})^8\middle|\mathscr{F}_{k\Delta}\right) - \Delta\xi^8(X_{k\Delta})I_8 = \Delta G^{(1)}_{k\Delta} + \Delta G^{(2)}_{k\Delta} + \Delta G^{(3)}_{k\Delta} \lesssim \Delta^{3/2},$$

we get:

$$\left|\mathrm{Var}\left(F_{k\Delta}\,|\mathscr{F}_{k\Delta}\right) - \frac{\xi^8(X_{k\Delta})I_8}{\Delta}\right| \lesssim \Delta^{-1/2}.$$

Bound of $H_{k\Delta}$. We only need an upper bound for the moments of $H_{k\Delta}$. We have that

$$\mathbb{E}\left(H^2_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right) \le \frac{1}{\Delta^2}\mathbb{E}\left((J_{k\Delta}+Z_{k\Delta})^{16}\middle|\mathscr{F}_{k\Delta}\right) \lesssim \frac{\Delta^8+\Delta}{\Delta^2} = \frac{1}{\Delta}$$

and

$$\mathbb{E}\left(H^4_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right) \le \frac{1}{\Delta^4}\mathbb{E}\left((J_{k\Delta}+Z_{k\Delta})^{32}\middle|\mathscr{F}_{k\Delta}\right) \lesssim \frac{\Delta^{16}+\Delta}{\Delta^4} = \frac{1}{\Delta^3}.$$

5.2 Proof of Proposition 6

We recall that the empirical norm is defined by $\|t\|_n^2 = \frac{1}{n}\sum_{k=1}^n t^2(X_{k\Delta})$. Then for any $t \in S_m$,

$$\begin{aligned}\gamma_n(t) &= \frac{1}{n}\sum_{k=1}^n \left(T_{k\Delta} - \xi^4(X_{k\Delta}) + \xi^4(X_{k\Delta}) - t(X_{k\Delta})\right)^2 \\ &= \frac{1}{n}\sum_{k=1}^n (T_{k\Delta} - \xi^4(X_{k\Delta}))^2 + \left\|t - \xi^4\right\|_n^2 + \frac{2}{n}\sum_{k=1}^n (T_{k\Delta} - \xi^4(X_{k\Delta}))(\xi^4(X_{k\Delta}) - t(X_{k\Delta})). \quad (8)\end{aligned}$$

As $\hat{\xi}_m^4$ minimizes $\gamma_n(t)$, $\gamma_n(\hat{\xi}_m^4) \leq \gamma_n(\Pi_m\xi^4)$ where Π_m is the orthogonal projection on S_m for the $\|.\|_n$-norm. Then by (8),

$$\begin{aligned}\left\|\xi^4 - \hat{\xi}_m^4\right\|_n^2 + \frac{2}{n}\sum_{k=1}^n (T_{k\Delta} - \xi^4(X_{k\Delta}))(\xi^4(X_{k\Delta}) - \hat{\xi}_m^4(X_{k\Delta})) \leq \\ \left\|\Pi_m\xi^4 - \xi^4\right\|_n^2 + \frac{2}{n}\sum_{k=1}^n (T_{k\Delta} - \xi^4(X_{k\Delta}))(\xi^4(X_{k\Delta}) - \Pi_m\xi^4(X_{k\Delta})).\end{aligned}$$

Moreover, as $T_{k\Delta} - \xi^4(X_{k\Delta}) = E_{k\Delta} + F_{k\Delta}$, we get:

$$\left\|\hat{\xi}_m^4 - \xi^4\right\|_n^2 \leq \left\|\Pi_m\xi^4 - \xi^4\right\|_n^2 + \frac{2}{n}\sum_{k=1}^n (E_{k\Delta} + F_{k\Delta})(\hat{\xi}_m^4(X_{k\Delta}) - \Pi_m\xi^4(X_{k\Delta})).$$

We have

$$\left\|\hat{\xi}_m^4 - \xi^4\right\|_n^2 = \left\|\hat{\xi}_m^4 - \Pi_m\xi^4\right\|_n^2 + \left\|\Pi_m\xi^4 - \xi^4\right\|_n^2 \text{ and } \left\|\Pi_m\xi^4 - \xi^4\right\|_n^2 = \left\|\Pi_m\xi^4 - \xi_A^4\right\|_n^2 + \left\|\xi_A^4 - \xi^4\right\|_n^2. \quad (9)$$

By substracting $\left\|\Pi_m\xi^4 - \xi^4\right\|_n^2$ in both sides of the inequality, we obtain:

$$\left\|\hat{\xi}_m^4 - \Pi_m\xi^4\right\|_n^2 \leq \frac{2}{n}\sum_{k=1}^n (E_{k\Delta} + F_{k\Delta})(\hat{\xi}_m^4(X_{k\Delta}) - \Pi_m\xi^4(X_{k\Delta})).$$

By geometric-arithmetic means inequality,

$$\frac{2}{n}\sum_{k=1}^n E_{k\Delta}(\hat{\xi}_m^4(X_{k\Delta}) - \Pi_m\xi^4(X_{k\Delta})) \leq \frac{1}{6}\left\|\hat{\xi}_m^4 - \Pi_m\xi^4\right\|_n^2 + \frac{6}{n}\sum_{k=1}^n E_{k\Delta}^2.$$

We introduce the unit ball for the π-norm $\mathscr{B}_m = \{t \in S_m,\ \|t\|_\pi^2 \leq 1\}$ and the contrast function $\nu_n(t) = \frac{1}{n}\sum_{k=1}^n F_{k\Delta}t(X_{k\Delta})$. Then by geometric-arithmetic means inequality,

$$\begin{aligned}\frac{2}{n}\sum_{k=1}^n F_{k\Delta}(\hat{\xi}_m^4(X_{k\Delta}) - \Pi_m\xi^4(X_{k\Delta})) &\leq \left(\left\|\hat{\xi}_m^4 - \Pi_m\xi^4\right\|_\pi^2 \sup_{t\in\mathscr{B}_m} \nu_n^2(t)\right)^{1/2} \\ &\leq \frac{1}{6}\left\|\hat{\xi}_m^4 - \Pi_m\xi^4\right\|_\pi^2 + 6\sup_{t\in\mathscr{B}_m} \nu_n^2(t).\end{aligned}$$

Collecting terms, we get:

$$\left\|\hat{\xi}_m^4 - \Pi_m \xi^4\right\|_n^2 \leq 6 \sup_{t \in \mathscr{B}_m} \nu_n^2(t) + \frac{6}{n} \sum_{k=1}^{n} E_{k\Delta}^2 + \frac{1}{6} \left\|\hat{\xi}_m^4 - \Pi_m \xi^4\right\|_\pi^2 + \frac{1}{6} \left\|\hat{\xi}_m^4 - \Pi_m \xi^4\right\|_n^2.$$

We have two different norms in the bound of the risk. For any deterministic function t, $\|t\|_\pi^2 = \mathbb{E}\left(\|t\|_n^2\right)$. As $\hat{\xi}^4$ is random, we introduce the space Ω_n on which the norms $\|.\|_n$ and $\|.\|_\pi$ are equivalent:

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

Comte et al. (2007, Lemma 6.1) prove the following lemma for diffusion processes thanks to concentration inequalities. However, its proof only relies on the boundedness of the stationary density π and of the mixing properties of the process $(X_t)_{t\geq 0}$.

Lemma 14. *As*

i) $(X_t)_{t\geq 0}$ *is exponentially* β*-mixing,*

ii) $(X_t)_{t\geq 0}$ *is stationary and its stationary density* π *is bounded from below and above on* A*,*

iii) the vectorial subspaces S_m *satisfy Assumption A4,*

then

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

We first bound the risk on Ω_n, where the $\|.\|_\pi^2$ is controlled. On Ω_n, $\left\|\hat{\xi}_m^4 - \Pi_m \xi^4\right\|_\pi^2 \leq 2\left\|\hat{\xi}_m^4 - \Pi_m \xi^4\right\|_n^2$. Therefore

$$\left\|\hat{\xi}_m^4 - \Pi_m \xi^4\right\|_n^2 \mathbf{1}_{\Omega_n} \leq \frac{12}{n} \sum_{k=1}^{n} E_{k\Delta}^2 + 12 \sup_{t \in \mathscr{B}_m} \nu_n^2(t). \tag{10}$$

The term $\sup_{t\in\mathscr{B}_m} \nu_n^2(t)$ can be bounded in two ways. First, let us consider $(\varphi_\lambda)_{\lambda\in\Lambda_m}$ an orthonormal basis (for the $\|.\|_\pi$-norm) of S_m. As $t \in \mathscr{B}_m$, it can be written $t = \sum_\lambda a_\lambda \varphi_\lambda$ where $\sum_\lambda a_\lambda^2 = 1$. Then, as ν is a linear function

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

By Lemma 5, $\mathbb{E}\left(F_{k\Delta}|\mathscr{F}_{k\Delta}\right)=0$ and therefore:

$$\begin{aligned}\mathbb{E}\left(\nu_n^2(\varphi_\lambda)\right) &= \mathbb{E}\left(\left(\frac{1}{n}\sum_{k=1}^{n}F_{k\Delta}\varphi_\lambda(X_{k\Delta})\right)^2\right)\\ &= \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\left(\varphi_\lambda^2(X_{k\Delta})\mathbb{E}\left(F_{k\Delta}^2|\mathscr{F}_{k\Delta}\right)\right)\\ &+ \frac{2}{n(n-1)}\sum_{j<k}\mathbb{E}\left(\varphi_\lambda(X_{k\Delta})\varphi_\lambda(X_{j\Delta})F_{j\Delta}\mathbb{E}\left(F_{k\Delta}|\mathscr{F}_{k\Delta}\right)\right)\\ &= \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\left(\varphi_\lambda^2(X_{k\Delta})\mathbb{E}\left(F_{k\Delta}^2\big|\mathscr{F}_{k\Delta}\right)\right).\end{aligned}$$

By Lemma 5, we get that

$$\mathbb{E}\left(\nu_n^2(\varphi_\lambda)\right)\leq\frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\left(\varphi_\lambda^2(X_{k\Delta})\left(\frac{\xi^8(X_{k\Delta})I_8}{\Delta}+\frac{C}{\sqrt{\Delta}}\right)\right).$$

Using that $\mathbb{E}\left(\varphi_\lambda^2(X_{k\Delta})\right)=\int_{\mathbb{R}}\varphi_\lambda^2(x)\pi(x)dx=1$, and the boundness of ξ, we obtain

$$\mathbb{E}\left(\nu_n^2(\varphi_\lambda)\right)\leq\frac{\xi_1^8I_8}{n\Delta}+\frac{C}{n\sqrt{\Delta}}.$$

Therefore,

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\nu_n^2(t)\right)\leq\sum_{\lambda}\mathbb{E}\left(\nu_n^2(\varphi_\lambda)\right)\leq\xi_1^8I_8\frac{D_m}{n\Delta}+\frac{CD_m}{n\sqrt{\Delta}}.$$

The second way of bounding $\sup_{t\in\mathscr{B}_n}\nu_n^2(t)$ nearly use the same arguments. Let us remark that if $t\in\mathscr{B}_m$, then $\|t\|_{L^2}^2\leq\frac{1}{\pi_0}\|t\|_\pi^2=\frac{1}{\pi_0}$. Then, if we consider $(\psi_\lambda)_{\lambda\in\Lambda_m}$ an orthonormal basis of S_m for the L^2-norm.

$$\sup_{t\in\mathscr{B}_m}\nu_n^2(t)\leq\sup_{t\in S_m;\|t\|_{L^2}^2\leq\frac{1}{\pi_0}}\nu_n^2(t)\leq\frac{1}{\pi_0}\sum_{\lambda}\nu_n^2(\psi_\lambda).$$

By Assumption 4 b, $\sum_{\lambda\in\Lambda_m}\psi_\lambda^2(X_{k\Delta})\leq\phi_1D_m$, and as the process is stationary:

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\nu_n^2(t)\right)\leq\frac{1}{\pi_0}\frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\left(F_{k\Delta}^2\sum_{\lambda\in\Lambda_m}\psi_\lambda^2(X_{k\Delta})\right)\leq\frac{\phi_1}{\pi_0}\frac{D_m\mathbb{E}\left(F_\Delta^2\right)}{n}.$$

Therefore, as $\mathbb{E}\left(F_\Delta^2\right)=\mathbb{E}\left(\xi^8(X_\Delta)\right)I_8/\Delta+O\left(\frac{1}{\sqrt{\Delta}}\right)$:

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\nu_n^2(t)\right)\leq\Xi\frac{D_m}{n\Delta}\tag{11}$$

where $\Xi=\min\left(\xi_1^8I_8,\frac{\phi_1}{\pi_0}\mathbb{E}_\pi\left(\xi^8(X_\Delta)\right)I_8\right)$. By Lemma 5, $\mathbb{E}\left(E_{k\Delta}^2\right)\lesssim\Delta$. By (10) and (11), we obtain:

$$\mathbb{E}\left(\left\|\hat{\xi}_m^4-\Pi_m\xi^4\right\|_n^2\mathbf{1}_{\Omega_n}\right)\leq12\Xi\frac{D_m}{n\Delta}+c\Delta.\tag{12}$$

As $\Pi_m \xi^4$ is the orthogonal projection for the $\|.\|_n$-norm, $\left\|\Pi_m \xi^4 - \xi_A^4\right\|_n^4 \leq \left\|\xi_{m,\pi}^4 - \xi_A^4\right\|_n^2$. And as $\xi_{m,\pi}$ and ξ are deterministic, $\mathbb{E}\left(\left\|\xi_{m,\pi}^4 - \xi_A^4\right\|_n^2\right) = \left\|\xi_{m,\pi}^4 - \xi_A^4\right\|_\pi^2$. Then by (9)

$$\mathbb{E}\left(\left\|\hat{\xi}_m^4 - \xi_A^4\right\|_n^2 \mathbf{1}_{\Omega_n}\right) \leq \left\|\xi_{m,\pi}^4 - \xi^4\right\|_\pi^2 + \Xi \frac{D_m}{n\Delta} + c\Delta.$$

It remains to bound the risk on Ω_n^c. The function $\hat{\xi}_m^4$ is the orthogonal projection (for the $\|.\|_n$ norm) of $(T_\Delta, \ldots, T_{n\Delta})$ on the vectorial subspace $\{(t(X_\Delta), \ldots, t(X_{n\Delta}))\ , t \in S_m\}$.

We have:

$$\begin{aligned}\left\|\hat{\xi}_m^4 - \xi_A^4\right\|_n^2 &= \left\|\Pi_m T - \xi_A^4\right\|_n^2 = \left\|\Pi_m \xi^4 - \xi_A^4\right\|_n^2 + \|\Pi_m E + \Pi_m F\|_n^2 \\ &\leq \left\|\xi_A^4\right\|_n^2 + \|E + F\|_n^2 .\end{aligned}$$

By stationarity and Cauchy-Schwarz:

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

We obtain by Lemma 5:

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

5.3 Proof of Corollary 7

We again separate the risk on Ω_n (where the empirical norm $\|.\|_n$ is equivalent to the $\|.\|_\pi$-norm) and Ω_n^c. As $\xi_{m,\pi}^4$ is the orthogonal projection of ξ^4 on S_m for the $\|.\|_\pi$- norm:

$$\left\|\hat{\xi}_m^4 - \xi^4\right\|_\pi^2 = \left\|\hat{\xi}_m^4 - \xi_{m,\pi}^4\right\|_\pi^2 + \left\|\xi_{m,\pi}^4 - \xi^4\right\|_\pi^2\}.$$

As $\hat{\xi}_m^4$ and $\xi_{m,\pi}^4$ both belong to the subspace S_m, on Ω_n:

$$\left\|\hat{\xi}_m^4 - \xi_{m,\pi}^4\right\|_\pi^2 \mathbf{1}_{\Omega_n} \leq 2\left\|\hat{\xi}_m^4 - \xi_{m,\pi}^4\right\|_n^2 \leq 4\left\|\hat{\xi}_m^4 - \Pi_m \xi^4\right\|_n^2 + 4\left\|\Pi_m \xi^4 - \xi_{m,\pi}^4\right\|_n^2$$

and as $\Pi_m \xi^4$ is the orthogonal projection of ξ^4 on S_m for the empirical risk, $\left\|\Pi_m \xi^4 - \xi_{m,\pi}^4\right\|_n^2 \leq \left\|\xi_A^4 - \xi_{m,\pi}^4\right\|_n^2$. By (12), we obtain that

$$\mathbb{E}\left(\left\|\hat{\xi}_m^4 - \xi^4\right\|_\pi^2 \mathbf{1}_{\Omega_n}\right) \leq 5\left\|\xi^4 - \xi_{m,\pi}^4\right\|_\pi^2 + 48\Xi \frac{D_m}{n\Delta} + C\Delta.$$

In addition, by Lemma 5 and Lemma 14:

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

which give the bound for the $\|.\|_\pi$-norm.

To bound the L^2-risk, we simply remark that $\|t\|^2_{L^2(A)} \leq \frac{1}{\pi_0}\|t\|^2_\pi$ and that $\left\|\xi^4 - \xi^4_{\pi,m}\right\|^2_\pi \leq \left\|\xi^4 - \xi^4_m\right\|^2_\pi \leq \pi_1 \left\|\xi^4 - \xi^4_m\right\|^2_{L^2(A)}$. Then

$$\mathbb{E}\left(\left\|\hat{\xi}^4_m - \xi^4_A\right\|^2_{L^2(A)}\right) \leq \frac{5\pi_1}{\pi_0}\left\|\xi^4 - \xi^4_m\right\|^2_{L^2(A)} + \frac{48\Xi}{\pi_0}\frac{D_m}{n\Delta} + C\Delta.$$

5.4 Proof of Theorem 8

For any $m \in \mathscr{M}_n$,

$$\gamma_n(\hat{\xi}^4_{\hat{m}}) + pen(\hat{m}) \leq \gamma_n(\Pi_m\xi^4) + pen(m).$$

As previously, we decompose the risk on Ω_n and Ω_n^c. On Ω_n^c, the bound is the same as in the previous proof. On Ω_n, as in the proof of Proposition 6, we have:

$$\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_A\right\|^2_n \leq \left\|\Pi_m\xi^4 - \xi^4_A\right\|^2_n + \frac{1}{12}\left\|\hat{\xi}^4_m - \Pi_m\xi^4\right\|^2_n + \frac{1}{12}\left\|\hat{\xi}^4_m - \Pi_m\xi^4\right\|^2_\pi + 12\sum_{k=1}^n E^2_{k\Delta} + 12\sup_{t\in\mathscr{B}_{m,\hat{m}}}\nu^2_n(t)$$
$$+ pen(m) - pen(\hat{m})$$

where $\mathscr{B}_{m,m'}$ is the unit ball for the $\|.\|_\pi$-norm of the vectorial subspace $S_m + S'_m$. On Ω_n, $\left\|\hat{\xi}^4_m - \Pi_m\xi^4\right\|^2_\pi \leq 2\left\|\hat{\xi}^4_m - \Pi_m\xi^4\right\|^4_n$. By the triangular inequality, $\left\|\hat{\xi}^4_{\hat{m}} - \Pi_m\xi^4\right\|^2_n \leq 2\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_A\right\|^2_n + 2\left\|\Pi_m\xi^4 - \xi^4_A\right\|^2_n$. Then

$$\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_A\right\|^2_n \leq 3\left\|\Pi_m\xi^4 - \xi^4_A\right\|^2_n + 2pen(m) - 2pen(\hat{m}) + 24\sum_{k=1}^n E^2_{k\Delta} + 24\sup_{t\in\mathscr{B}_{m,\hat{m}}}\nu^2_n(t).$$

Let us set $p(m,m') = (pen(m) + pen(m'))/12$. Then

$$\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_A\right\|^2_n \mathbf{1}_{\Omega_n} \leq 3\left\|\Pi_m\xi^4 - \xi^4_A\right\|^2_n + 2pen(m) + 24\sum_{k=1}^n E^2_{k\Delta} + 24(\sup_{t\in\mathscr{B}_{m,\hat{m}}}\nu^2_n(t) - p(m,\hat{m})). \quad (13)$$

It remains to bound

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}}\nu^2_n(t) - p(m,\hat{m})\right) \leq \sum_{m\in\mathscr{M}_n}\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}}\nu^2_n(t) - p(m,m')\right).$$

We can do so using concentration inequalities. The following lemma is deduced from the Berbee's coupling lemma and a Talagrand inequality. It is proved for instance in Schmisser (2019, Appendix A).

Lemma 15 (Talagrand inequality for β-mixing variables)**.** *Let us consider $Z_1, \dots, Z_n$ some random variables exponentially β-mixing. Let us set $b_0 \geq 1/\beta$ where β is the β-mixing coefficient). We define $q_n := 2b_0\ln(n)$, $p_n = n/(2q_n)$. We have that $\beta(q_n) \leq ce^{-2\beta b_0\ln(n)} \lesssim n^{-2}$. Let us consider*

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

If we can find a triplet (M_2, V and H) such that:

$$\forall i, \sup_{t\in\mathscr{F}} \mathrm{Var}\left(\frac{1}{q_n}\sum_{k=iq_n}^{(i+1)q_n} F_t(Z_k)\right) \leq \frac{V}{q_n},$$

$$\sup_{t\in\mathscr{F}}\left\|\frac{1}{q_n}\sum_{k=cq_n}^{(c+1)q_n} F_t(Z_k)\right\|_\infty \leq M_2 \quad \textit{and} \quad \mathbb{E}\left[\sup_{t\in\mathscr{F}}|I_n(t)|\right] \leq \frac{H}{\sqrt{n}},$$

then we have:

$$\mathbb{E}\left[\sup_{t\in\mathscr{F}}\left|I_n^2(t) - \frac{6H^2}{n}\right|\right]_+ \leq C\frac{V}{n}\exp\left(-k_1\frac{H^2}{12V}\right) + C'\frac{M_2^2}{p_n^2}\exp\left(-k_2\frac{\sqrt{p_n}H}{\sqrt{q_n}M_2}\right) + 2\frac{M_2^2}{n^2}$$

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

The random variables $(F_{k\Delta}, X_{k\Delta})$ are exponentially β-mixing, with β-mixing coefficient proportional to Δ. Let us set $F_t(F_{k\Delta}, X_{k\Delta}) := F_{k\Delta}t(X_{k\Delta})$. We want to apply Lemma 15 to $\nu_n(t) = \frac{1}{n}\sum_{k=1}^n F_{k\Delta}t(X_{k\Delta})$. By Equation (11),

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}}\nu_n^2(t)\right) \leq \frac{\Xi(D_m + D_{m'})}{n\Delta},$$

so we can take $H^2 = \frac{\Xi(D_m+D_{m'})}{\Delta}$. Let us set $q_n = c\ln(n)\Delta$ with c such that $\beta(q_n\Delta) \lesssim n^{-2}$. By Lemma 5, as $\mathbb{E}\left(F_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right) = 0$, we have that, for any $t \in \mathscr{B}_{m,m'}$,

$$\mathrm{Var}\left(\frac{1}{q_n}\sum_{k=iq_n}^{(i+1)q_n} F_{k\Delta}t(X_{k\Delta})\right) = \frac{1}{q_n^2}\sum_{k=iq_n}^{(i+1)q_n}\mathbb{E}\left(\mathbb{E}\left(F_{k\Delta}^2\middle|\mathscr{F}_{k\Delta}\right)t^2(X_{k\Delta})\right) \leq \frac{\xi_1^8 I_8}{q_n\Delta}.$$

Then we can choose $V = \xi_1 I_8/\Delta$. It remains to bound the variables $F_{k\Delta}$ with high probability. We have the following decomposition:

$$F_{k\Delta} \lesssim \frac{J_{k\Delta}^4 + Z_{k\Delta}^4}{\Delta}.$$

Let us introduce a set where this quantity is bounded: for any $0 < \alpha < 1$,

$$\Omega_{B,\alpha} = \Big\{\omega, \forall k, \forall p \in \mathbb{N}, p \leq 8, |B_{k\Delta}| \leq \Delta^{1-\alpha}; \quad |Z_{k\Delta}| \leq 6\sigma_1\Delta^{1/2}\ln(n); \quad |J_{k\Delta}| \leq 36\Delta^{-\alpha}\ln(n);$$
$$\frac{1}{q_n}\sum_{k=cq_n}^{(c+1)q_n} J_{k\Delta}^p \leq 49\Delta^{1-\alpha}\ln^p(n)\Big\}.$$

By Lemma 4:

$$\mathbb{P}(\Omega_{B,\alpha}) \lesssim \frac{1}{n^5}. \tag{14}$$

On $\Omega_n \cap \Omega_{B,\alpha}$, if $t \in S_m + S_{m'}$, $t(X_{k\Delta}) \leq (D_m + D_{m'})^{1/2}\sqrt{\phi_1/\pi_0}$, and $F_{k\Delta} \lesssim (J_{k\Delta}^4 + Z_{k\Delta}^4)/\Delta \lesssim C\ln^4(n)\Delta^{-\alpha}$. Then we apply Talagrand's inequality (lemma 15) to $\nu_n(t) = \frac{1}{n}\sum_{k=1}^n t(X_{k\Delta})F_{k\Delta}$ with

$$V = \frac{\xi_1^8 I_8}{\Delta}, \quad H^2 = \frac{\Xi D}{\Delta}, \quad M_2 = C\frac{D^{1/2}\ln^4(n)}{\Delta^\alpha}$$

with $D = D_m + D_{m'}$. By Lemma 15, we obtain:

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}} \nu_n^2(t) - 12H^2\right)_+ &\lesssim \frac{1}{n\Delta}\exp\left(-c\frac{D/\Delta}{1/\Delta}\right) \\ &\quad + \frac{D\ln^8(n)}{p_n^2\Delta^{2\alpha}}\exp\left(-c'\frac{\sqrt{p_n}D^{1/2}/\sqrt{\Delta}}{\sqrt{q_n}D^{1/2}\ln^4(n)/\Delta^\alpha}\right) + \frac{D\ln^8(n)}{n^2\Delta^{2\alpha}} \\ &\lesssim \frac{1}{n\Delta}\exp(-c_1 D) + \frac{D\ln^{10}(n)}{n^2\Delta^{2+2\alpha}}\exp\left(-c_2\frac{\sqrt{n\Delta^{1+2\alpha}}}{\ln^5(n)}\right) + \frac{D\ln^8(n)}{n^2\Delta^{2\alpha}}.\end{aligned}$$

We obtain that

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_n^2(t) - p(m,\hat{m})\right) &\leq \sum_{D_m\in\mathscr{M}_n} \mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}} \nu_n^2(t) - H^2\right) \\ &\leq \sum_{D=1}^{\sqrt{n\Delta}} \mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}} \nu_n^2(t) - H^2\right) \\ &\lesssim \frac{1}{n\Delta} + \frac{1}{n\Delta^{1+2\alpha}} e^{-\sqrt{\frac{n\Delta^{1+2\alpha}}{\ln^5(n)}}} + \frac{\Delta^{1-2\alpha}}{n}.\end{aligned}$$

As we can take α as small as we want, we get that

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_n^2(t) - p(m,\hat{m})\mathbf{1}_{\Omega_n\cap\Omega_{B,\alpha}}\right) \leq \frac{C'}{n\Delta}. \tag{15}$$

The bound of the risk on $(\Omega_n \cap \Omega_{B,\alpha})^c$ is done as in the proof of Proposition 6.

5.5 Proof of Corollary 9

The proof is nearly the same as for the risk of the estimator with m fixed. We have that, for any m:

$$\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_A\right\|_\pi^2 \leq 2\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_{m,\pi}\right\|_\pi^2 + 2\left\|\xi^4_{m,\pi} - \xi^4_A\right\|_\pi^2. \tag{16}$$

And on $\Omega_n \cap \Omega_{B,\alpha}$:

$$\begin{aligned}\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_{m,\pi}\right\|_\pi^2 \mathbf{1}_{\Omega_n} \leq 2\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_{m,\pi}\right\|_n^2 &\leq 4\left\|\hat{\xi}^4_{\hat{m}} - \Pi_m\xi^4\right\|_n^2 + 4\left\|\Pi_m\xi^4 - \xi^4_{m,\pi}\right\|_n^2 \\ &\leq 4\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_A\right\|_n^2 + 4\left\|\xi^4_A - \xi^4_{m,\pi}\right\|_n^2.\end{aligned}$$

As $\left\|\Pi_m\xi^4 - \xi^4_A\right\|_n^2 \leq \left\|\xi^4_{m,\pi} - \xi^4_A\right\|_n^2$, by Equations (13) and (16), we obtain:

$$\begin{aligned}\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_{m,\pi}\right\|_\pi^2 \mathbf{1}_{\Omega_n\cap\Omega_{B,\alpha}} &\leq 8\left\|\hat{\xi}^4_{\hat{m}} - \xi^4_A\right\|_n^2 + 8\left\|\xi^4_A - \xi^4_{m,\pi}\right\|_n^2 + 2\left\|\xi^4_{m,\pi} - \xi^4_A\right\|_\pi^2 \\ &\leq 32\left\|\xi^4_{m,\pi} - \xi^4_A\right\|_n^2 + 2\left\|\xi^4_{m,\pi} - \xi^4_A\right\|_\pi^2 + 16pen(m) \\ &\quad + 192\sum_{k=1}^n E^2_{k\Delta} + 192\sup_{t\in\mathscr{B}_{m,\hat{m}}}\left(\nu_n^2(t) - p(m,\hat{m})\right).\end{aligned}$$

The risk on $(\Omega_n \cap \Omega_{B,\alpha})^c$ is the same as for m fixed, so we get by (15):

$$\mathbb{E}\left(\left\|\hat{\xi}^4_{\hat{m}} - \xi^4\right\|^2_\pi\right) \lesssim \inf_{m\in\mathscr{M}_n}\left(\left\|\xi^4_{m,\pi} - \xi^4\right\|^2_\pi + pen(m)\right) + \frac{1}{n\Delta} + \Delta$$

and

$$\mathbb{E}\left(\left\|\hat{\xi}^4_{\hat{m}} - \xi^4\right\|^2_{L^2(A)}\right) \lesssim \inf_{m\in\mathscr{M}_n}\left(\left\|\xi^4_{m} - \xi^4\right\|^2_{L^2(A)} + pen(m)\right) + \frac{1}{n\Delta} + \Delta.$$

5.6 Proof of Lemma 11

As in the proof of Proposition 6, we decompose the risk on Ω_n and Ω_n^c and get that, on Ω_n,

$$\left\|\hat{h}_m - \Pi_m h\right\|^2_n \mathbf{1}_{\Omega_n} \leq 12 \sup_{t\in\mathscr{B}_m} \nu^2_{n,2}(t) + \frac{12}{n}\sum_{k=1}^n G^2_{k\Delta}$$

where $\nu_{n,2}(t) = \frac{1}{n}\sum_{k=1}^n t(X_{k\Delta})H_{k\Delta}$. By Lemma 10, we have that $\mathbb{E}\left(G^2_{k\Delta}\right) \lesssim \Delta$. Moreover, for $(\varphi_\lambda)_{1\leq\lambda\leq D_m}$ an orthonormal basis for the L^2_π- norm of S_m, we have

$$\mathbb{E}\left(\sup_{t\in S_m}\nu^2_{n,2}(t)\right) \leq \sum_\lambda \mathbb{E}\left(\nu^2_{n,2}(\varphi_\lambda)\right) = \sum_\lambda \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(H^2_{k\Delta}\varphi^2_\lambda(X_{k\Delta})\right) \lesssim \frac{D_m}{n\Delta}.$$

As $\left\|\hat{h}_m - h_A\right\|^2_n = \left\|\hat{h}_m - \Pi_m h\right\|^2_n + \|\Pi_m h - h_A\|^2_n$, and $\mathbb{E}\left(\|\Pi_m h - h_A\|^2_n\right) \leq \mathbb{E}\left(\|h_{m,\pi} - h_A\|^2_n\right) = \|h_{m,\pi} - h_A\|^2_\pi$, we get that

$$\mathbb{E}\left(\left\|\hat{h}_m - h_A\right\|^2_n \mathbf{1}_{\Omega_n}\right) \leq \|h_{m,\pi} - h\|^2_\pi + \frac{CD_m}{n\Delta} + c\Delta.$$

And on Ω_n^c, we get that

$$\mathbb{E}\left(\left\|\hat{h}_m - h_A\right\|^2_n\right) \lesssim \left[\left(\mathbb{E}\left(\|h_A\|^4_n\right) + \mathbb{E}\left(G^4_{k\Delta} + H^4_{k\Delta}\right)\right)\mathbb{P}\left(\Omega_n^c\right)\right]^{1/2} \lesssim \frac{1}{\Delta^{3/2}n^4} \leq \frac{1}{n}$$

which ends the proof.

5.7 Proof of lemma 12

Let us set $h_1 = \sup_{x\in A} h(x)$ and $\hat{h}_1 = \sup_A \hat{h}_{\ln(n)}(x)$.

Bound of $h_1 - \hat{h}_1$. We have that

$$|\hat{h}_1 - h_1| \leq \left\|\hat{h}_{\ln(n)} - h\right\|_{L^\infty(A)} \leq \left\|\hat{h}_{\ln(n)} - h_{\ln(n)}\right\|_{L^\infty(A)} + \left\|h_{\ln(n)} - h\right\|_{L^\infty(A)}.$$

By DeVore and Lorentz (1993, 13, p182) and Barron et al. (1999, 4, Lemma 12), if h belongs to the Besov space $\mathscr{B}^\alpha_{2,\infty}(A)$, then

$$\|h_m - h\|^2_{L^\infty(A)} \leq CD_m^{1-2\alpha}. \tag{17}$$

As ξ is Lipschitz, $h = \xi^8 I_8$ belongs to $\mathscr{B}^1_{2,\infty}(A)$ and $\|h_m - h_A\|^2_{L^\infty(A)} \lesssim D_m^{-1}$. It remains to bound $\left\|\hat{h}_m - h_m\right\|_{L^\infty(A)}$. As $\hat{h}_m$ and h_m belongs to S_m,

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

On Ω_n,

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

Those two terms are bounded thanks to a concentration inequality.

Bound of $\|\Pi_m h - h_m\|^2_n$. As $\Pi_m h$ is the orthogonal projection of h on S_m, $\|\Pi_m h - h_m\|^2_n \leq \|h_m - h_A\|^2_n$. Its expectation is $\|h_m - h_A\|^2_\pi \lesssim D_m^{-2}$. By (17), $\|h_m - h\|^2_{L^\infty(A)} \lesssim D_m^{-1}$. Moreover,

$$\begin{aligned}\operatorname{Var}\left(\sum_{jq_n+1}^{(j+1)q_n} (h_m - h_A)^2(X_{k\Delta})\right) &\leq q_n^2 \mathbb{E}\left((h_m - h_A)^4(X_\Delta)\right) \leq q_n^2 \|h_m - h_A\|^2_{L^\infty} \|h_m - h_A\|^2_\pi \\ &\lesssim q_n^2 D_m^{-3}.\end{aligned}$$

The following lemma is very useful to control the difference between a mean and its expectation. It is proved for instance in Schmisser (2019, Result 24).

Lemma 16 (Bennet inequality for β-mixing variables)**.** *Let us consider $Y_1, \dots, Y_n$ some exponentially β-mixing random variables. Let us set $b_0 \geq 1/\beta$ where β is the β-mixing coefficient. We define $q_n := 2b_0 \ln(n)$, $p_n = n/(2q_n)$. We have that $\beta(q_n) \leq ce^{-2\beta b_0 \ln(n)} \lesssim n^{-2}$. Let us set $\bar{Y}_n = \frac{1}{n}\sum_{k=1}^n Y_k$. If there exists two constants V' and M_2' such that, for any j,*

$$\operatorname{Var}\left(\sum_{k=jq_n+1}^{(j+1)q_n} Y_k\right) \leq q_n V' \quad \text{and} \quad \left|\sum_{k=jq_n+1}^{(j+1)q_n} Y_k - \mathbb{E}\left(Y_k\right)\right| \leq q_n M_2'$$

then, for any $x > 0$,

$$\mathbb{P}\left(\left|\bar{Y}_n - \mathbb{E}\left(\bar{Y}_n\right)\right| \geq x\right) \leq 2\exp\left(-\frac{nx^2}{2(V' + xM_2' q_n)}\right).$$

By Bennett inequality with $V' = q_n D_m^{-3}$ and $M_2' = CD_m^{-1}$, as $D_m = \ln(n)$,

$$\begin{aligned}\mathbb{P}\left(\left|\|h_m - h_A\|^2_n - \|h_m - h_A\|^2_\pi\right| \geq D_m^{-2}\right) &\lesssim \exp\left(-\frac{cnD_m^{-4}}{q_n D_m^{-3} + q_n D_m^{-3}}\right) = \exp\left(-cp_n D_m^{-1}\right) \\ &\lesssim \exp\left(-c\frac{n\Delta}{\ln^2(n)}\right) \lesssim n^{-5}.\end{aligned}$$

Then

$$\mathbb{P}\left(\|h_m - \Pi_m h\|^2_n \geq D_m^{-2}\right) \lesssim n^{-5}. \tag{18}$$

Bound of $\left\|\hat{h}_m - \Pi_m h\right\|_n$. We know that, on Ω_n,

$$\left\|\hat{h}_m - \Pi_m h\right\|_n^2 \leq 12 \sup_{t\in\mathscr{B}_m} \nu_{n,2}^2(t) + \frac{12}{n}\sum_{k=1}^n G_{k\Delta}^2. \tag{19}$$

Moreover, for $(\varphi_\lambda)_{\lambda\in\Lambda_m}$ an orthonormal basis for the L^2_π-norm of S_m,

$$\sup_{t\in\mathscr{B}_m} \nu_{n,2}^2(t) \leq \sum_\lambda \nu_{n,2}^2(\varphi_\lambda) = \sum_\lambda \left(\frac{1}{n}\sum_{k=1}^n H_{k\Delta}\varphi_\lambda(X_{k\Delta})\right)^2$$

and we use again Bennett's inequality. We recall that $|H_{k\Delta}| \lesssim \Delta^{-1}(Z_{k\Delta}^8 + J_{k\Delta}^8)$. For any $t \in \mathscr{B}_m$, this term is centred: $\mathbb{E}\left(\nu_{n,2}(t)\right) = 0$. In addition, as $\mathbb{E}\left(H_{k\Delta}|\mathscr{F}_{k\Delta}\right) = 0$, its variance is bounded by:

$$\mathrm{Var}\left(\sum_{k=jq_n+1}^{(j+1)q_n} H_{k\Delta}\varphi_\lambda(X_{k\Delta})\right) = q_n \mathrm{Var}\left(H_\Delta\varphi_\lambda(X_\Delta)\right) \lesssim \frac{q_n}{\Delta}.$$

On $\Omega_{B,\alpha}$, as $\|\varphi_\lambda\|_{L^\infty}^2 \leq \frac{\phi_1}{\pi_0} D_m \|\varphi\|_{L^2}^2 = \frac{\phi_1}{\pi_0} D_m$, we have:

$$\left|\sum_{k=jq_n+1}^{(j+1)q_n} H_{k\Delta}\varphi_\lambda(X_{k\Delta})\right| \leq D_m^{1/2}\sqrt{\frac{\phi_1}{\pi_0}}\sum_{k=jq_n+1}^{(j+1)q_n}|H_{k\Delta}| \lesssim q_n\sqrt{D_m}\ln^8(n)\Delta^{-\alpha}.$$

And by Bennett's inequality (Lemma 16) with $V' = C/\Delta$ and $M_2' = C'\sqrt{D_m}\ln^8(n)\Delta^{-\alpha}$, we obtain:

$$\mathbb{P}\left(|\nu_{n,2}(\varphi_\lambda)| \geq \frac{1}{\sqrt{D_m}\ln^9(n)}\right) \lesssim \mathbb{P}\left(\Omega_{B,\alpha}^c\right) + \exp\left(-c\frac{n\Delta^{1+\alpha}}{D_m\ln^{18}(n)}\right) \lesssim n^{-5} + \exp\left(-\frac{n\Delta^{1+\alpha}}{\ln^{19}(n)}\right) \lesssim n^{-5}.$$

Then

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in\mathscr{B}_m}\nu_{n,2}^2(t) \geq \frac{1}{\ln^{18}(n)}\right) &\leq \mathbb{P}\left(\sum_\lambda \nu_{n,2}^2(\varphi_\lambda) \geq \frac{1}{\ln^{18}(n)}\right)\\ &\leq \sum_\lambda \mathbb{P}\left(|\nu_{n,2}^2(\varphi_\lambda)| \geq \frac{1}{D_m\ln^{18}(n)}\right) \lesssim D_m n^{-5}.\end{aligned} \tag{20}$$

Let us now bound $G_{k\Delta}$. We have that $G_{k\Delta} = G_{k\Delta}^{(1)} + G_{k\Delta}^{(2)} + G_{k\Delta}^{(3)} + G_{k\Delta}^{(4)}$, with $|G_{k\Delta}^{(1)} + G_{k\Delta}^{(2)} + G_{k\Delta}^{(3)}| \lesssim \Delta^{1/2}$. Moreover, on $\Omega_{B,\alpha}$, by Lemma 4,

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=1}^n\left(G_{k\Delta}^{(4)}\right)^2\right| &\lesssim \frac{\Delta^{-2}}{n}\sum_{k=1}^n\sum_{j=1}^8 B_{k\Delta}^{2j}(Z_{k\Delta}^{16-2j} + J_{k\Delta}^{16-2j})\\ &\lesssim \Delta^{-2}\sum_{j=1}^8 \Delta^{2j-2j\alpha}\left(\Delta^{8-j}\ln^{16-2j}(n) + \frac{1}{n}\sum_{k=1}^n J_{k\Delta}^{16-2j}\right)\\ &\lesssim \sum_{j=1}^8 \Delta^{2j-2j\alpha-2}\left(\Delta^{8-j}\ln^{16-2j}(n) + \Delta^{1-\alpha}\mathbf{1}_{j<8}\ln^{16-2j}(n) + \mathbf{1}_{j=8}\right)\\ &\lesssim \Delta^{1-3\alpha}\ln^{16}(n).\end{aligned}$$

Then, for n large enough:

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} G_{k\Delta}^2 \geq \Delta \ln(n) + \Delta^{1-3\alpha}\ln^{17}(n)\right) \leq \mathbb{P}\left(\Omega_{B,\alpha}^c\right) \lesssim n^{-5}. \tag{21}$$

Therefore, by Equations (19), (21) and (20), we obtain:

$$\mathbb{P}\left(\left\|\hat{h}_m - \Pi_m h\right\|_n^2 \geq \Delta^{1-3\alpha}\ln^{17}(n) + \frac{1}{\ln^9(n)}\right) \leq \mathbb{P}\left(\Omega_n^c\right) + D_m n^{-5} + n^{-5} \lesssim D_m n^{-5}. \tag{22}$$

We recall that, on Ω_n:

$$|\hat{h}_1 - h_1| \lesssim \left\|h_{\ln(n)} + h\right\|_{L^\infty(A)} + \ln(n)\left\|h_{\ln(n)} - \Pi_{\ln(n)}h\right\|_n^2 + \ln(n)\left\|\hat{h}_{\ln(n)} - \Pi_{\ln(n)}h\right\|_n^2.$$

By Equations (17), (18) and (22), we obtain:

$$\mathbb{P}\left(|\hat{h}_1 - h_1| \gtrsim \frac{1}{\ln(n)} + \frac{\ln(n)}{\ln^9(n)} + \ln^{18}(n)\Delta^{1-3\alpha}\right) = \mathbb{P}\left(|\hat{h}_1 - h_1| \gtrsim \frac{1}{\ln(n)}\right) \lesssim \mathbb{P}\left(\Omega_n^c\right) + n^{-4} \lesssim n^{-4}.$$

Bound of $\bar{V}_n - \mathbb{E}\left(\xi^8(X_\Delta)\right) I_8$. To bound this quantity, we use Bennett's inequality.

By Lemma 10,

$$\mathbb{E}\left(\bar{V}_n\right) = \mathbb{E}\left(\xi^8(X_{k\Delta})I_8 + G_{k\Delta} + H_{k\Delta}\right) = \mathbb{E}\left(\xi^8(X_{k\Delta})\right) I_8 + C'\Delta^{1/2}.$$

Moreover,

$$\begin{aligned}\mathrm{Var}\left(\sum_{j=1}^{q_n} V_{j\Delta}\right) &\lesssim \mathrm{Var}\left(\sum_{j=1}^{q_n}\xi^8(X_{j\Delta})\right) + \mathrm{Var}\left(\sum_{j=1}^{q_n} G_{j\Delta}\right) + \mathrm{Var}\left(\sum_{j=1}^{q_n} H_{j\Delta}\right)\\ &\lesssim q_n^2 + q_n^2\Delta + \frac{q_n}{\Delta} = q_n^2 + \frac{q_n}{\Delta}.\end{aligned}$$

And on $\Omega_{B,\alpha}$:

$$\left|\sum_{j=1}^{q_n} V_{j\Delta}\right| \lesssim \sum_{j=1}^{q_n}\frac{B_{j\Delta}^8 + Z_{j\Delta}^8 + J_{j\Delta}^8}{\Delta} \lesssim q_n \ln^8(n)\Delta^{-\alpha}.$$

Then, with $V' = q_n + 1/\Delta$ and $M_2' = \ln^8(n)\Delta^{-\alpha}$, by Bennett's inequality, we get on $\Omega_{B,\alpha}$:

$$\begin{aligned}\mathbb{P}\left(\left|\bar{V}_n - \mathbb{E}\left(\xi^8(X_\Delta)\right) I_8\right| \geq \Delta^{1/4} + x\right) &\leq \exp\left(-\frac{nx^2}{2(V' + M_2' x q_n)}\right)\\ &= \exp\left(-\frac{Cnx^2}{\frac{1}{\Delta} + \frac{\ln(n)}{\Delta} + \frac{\ln^9(n)x\Delta^{-\alpha}}{\Delta}}\right)\end{aligned}$$

Taking $x = 1/\ln(n)$, as $\Delta = O(n^{-\delta})$ and $\ln(n) = o(1/\Delta)$, we get:

$$\mathbb{P}\left(\left|\bar{V}_n - \mathbb{E}\left(\xi^8(X_\Delta)I_8\right)\right| \geq \frac{1}{\ln(n)} + \Delta^{1/4}\right) \lesssim \exp\left(-\frac{cn\Delta}{\ln^2(n)\left(1 + \ln(n) + \ln^8(n)\Delta^{-\alpha}\right)}\right) \lesssim n^{-5}.$$

Bound of $\pi_0 - \hat{\pi}_0$. We have that $\hat{\pi}_h(x) = \frac{1}{n}\sum_{k=1}^{n} K_h(X_{k\Delta}) - x$. The expectation of this term is $\mathbb{E}\left(\hat{\pi}_h(x)\right) = \pi_h(x) := \int K_h(y-x)\pi(y)dy$. We have the following bounds:

$$\mathrm{Var}\left(\sum_{j=kq_n+1}^{(k+1)q_n} K_h(X_{j\Delta})\right) \lesssim \frac{q_n^2}{h} \quad \text{and} \quad \left|\sum_{j=kq_n+1}^{(k+1)q_n} K_h(X_{j\Delta})\right| \lesssim \frac{q_n}{h}.$$

Then by Bennett's inequality (lemma 16) with $V' = Cq_n/h$ and $M_2' = C'/h$, we get, for any x:

$$\mathbb{P}\left(|\hat{\pi}_h(x) - \pi_h(x)| \geq c\frac{\ln^2(n)}{(n\Delta)^{1/4}}\right) \lesssim \exp\left(-\frac{nc^2\ln^4(n)/(n\Delta)^{1/2}}{\frac{Cq_n}{h} + \frac{C'cq_n\ln^2(n)}{(n\Delta)^{1/4}h}}\right) \lesssim \exp\left(-\frac{C''nh\ln^4(n)}{(n\Delta)^{1/2}q_n}\right).$$

Let us take $h = (n\Delta)^{-1/2}$, as q_n is proportional to $\frac{\ln(n)}{\Delta}$, we obtain:

$$\mathbb{P}\left(|\hat{\pi}_h(x) - \pi_h(x)| \geq c\frac{\ln^2(n)}{(n\Delta)^{1/4}}\right) \lesssim \exp\left(-C\ln^3(n)\right).$$

Moreover, as π is continuous, by Tsybakov (2004, Proposition 1.2 with $\beta = 1$): $|\pi_h(x) - \pi(x)| \lesssim C'h$. Let $(x_1, \ldots, x_{\ln^2(n)})$ be equally spaced points on the interval A. Then

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

and therefore

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

which ends the proof.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Y. Aït-Sahalia and J. Jacod. Estimating the degree of activity of jumps in high frequency data. *Ann. Statist.*, 37(5A):2202–2244, 2009. ISSN 0090-5364,2168-8966. doi: 10.1214/08-AOS640. URL https://doi.org/10.1214/08-AOS640.

D. Applebaum. *Lévy processes and stochastic calculus*, volume 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2009. ISBN 978-0-521-73865-1. doi: 10.1017/CBO9780511809781. URL https://doi.org/10.1017/CBO9780511809781.

F. M. Bandi and T. H. Nguyen. On the functional estimation of jump-diffusion models. volume 116, pages 293–328. 2003. doi: 10.1016/S0304-4076(03)00110-6. URL https://doi.org/10.1016/S0304-4076(03)00110-6. Frontiers of financial econometrics and financial engineering.

A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999. ISSN 0178-8051,1432-2064. doi: 10.1007/s004400050210. URL https://doi.org/10.1007/s004400050210.

J. Berestycki. Exchangeable fragmentation-coalescence processes and their equilibrium measures. *Electron. J. Probab.*, 9:no. 25, 770–824, 2004. ISSN 1083-6489. doi: 10.1214/EJP.v9-227. URL https://doi.org/10.1214/EJP.v9-227.

A. Carpentier, C. Duval, and E. Mariucci. Total variation distance for discretely observed Lévy processes: a Gaussian approximation of the small jumps. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(2):901–939, 2021. ISSN 0246-0203,1778-7017. doi: 10.1214/20-aihp1102. URL https://doi.org/10.1214/20-aihp1102.

F. Comte, V. Genon-Catalot, and Y. Rozenholc. Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13(2):514–543, 2007. ISSN 1350-7265,1573-9759. doi: 10.3150/07-BEJ5173. URL https://doi.org/10.3150/07-BEJ5173.

R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. ISBN 3-540-50627-6.

M. Hanif. A nonparametric approach to the estimation of jump-diffusion models with asymmetric kernels. *Cogent Math.*, 3:Art. ID 1179247, 15, 2016. ISSN 2331-1835. doi: 10.1080/23311835.2016.1179247. URL https://doi.org/10.1080/23311835.2016.1179247.

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

C. Mancini. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand. J. Stat.*, 36(2):270–296, 2009. ISSN 0303-6898,1467-9469. doi: 10.1111/j.1467-9469.2008.00622.x. URL https://doi.org/10.1111/j.1467-9469.2008.00622.x.

C. Mancini and R. Renò. Threshold estimation of Markov models with jumps and interest rate modeling. *J. Econometrics*, 160(1):77–92, 2011. ISSN 0304-4076,1872-6895. doi: 10.1016/j.jeconom.2010.03.019. URL https://doi.org/10.1016/j.jeconom.2010.03.019.

H. Masuda. Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.*, 117(1):35–56, 2007. ISSN 0304-4149,1879-209X. doi: 10.1016/j.spa.2006.04.010. URL https://doi.org/10.1016/j.spa.2006.04.010.

J. Y. Park and B. Wang. Nonparametric estimation of jump diffusion models. *J. Econometrics*, 222(1):688–715, 2021. ISSN 0304-4076,1872-6895. doi: 10.1016/j.jeconom.2020.07.020. URL https://doi.org/10.1016/j.jeconom.2020.07.020.

P. Protter and D. Talay. The Euler scheme for Lévy driven stochastic differential equations. *Ann. Probab.*, 25(1):393–423, 1997. ISSN 0091-1798,2168-894X. doi: 10.1214/aop/1024404293. URL https://doi.org/10.1214/aop/1024404293.

E. Schmisser. Non-parametric adaptive estimation of the drift for a jump diffusion process. *Stochastic Process. Appl.*, 124(1):883–914, 2014. ISSN 0304-4149,1879-209X. doi: 10.1016/j.spa.2013.09.012. URL https://doi.org/10.1016/j.spa.2013.09.012.

E. Schmisser. Non parametric estimation of the diffusion coefficients of a diffusion with jumps. *Stochastic Process. Appl.*, 129(12):5364–5405, 2019. ISSN 0304-4149,1879-209X. doi: 10.1016/j.spa.2019.03.003. URL https://doi.org/10.1016/j.spa.2019.03.003.

Y. Shimizu. M-estimation for discretely observed ergodic diffusion processes with infinitely many jumps. *Stat. Inference Stoch. Process.*, 9(2):179–225, 2006. ISSN 1387-0874,1572-9311. doi: 10.1007/s11203-005-8113-y. URL https://doi.org/10.1007/s11203-005-8113-y.

Y. Shimizu and N. Yoshida. Estimation of parameters for diffusion processes with jumps from discrete observations. *Stat. Inference Stoch. Process.*, 9(3):227–277, 2006. ISSN 1387-0874,1572-9311. doi: 10.1007/s11203-005-8114-x. URL https://doi.org/10.1007/s11203-005-8114-x.

K. Song, Y. Song, and H. Wang. Threshold reweighted Nadaraya-Watson estimation of jump-diffusion models. *Probab. Uncertain. Quant. Risk*, 7(1):31–44, 2022. ISSN 2095-9672,2367-0126. doi: 10.3934/puqr.2022003. URL https://doi.org/10.3934/puqr.2022003.

A. B. Tsybakov. *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2004. ISBN 3-540-40592-5.