ESAIM: PS 22 (2018) 236–260
https://doi.org/10.1051/ps/2018005

ESAIM: Probability and Statistics
www.esaim-ps.org

ADAPTIVE NONPARAMETRIC DRIFT ESTIMATION OF AN INTEGRATED JUMP DIFFUSION PROCESS

Benedikt Funke[1,*] and Émeline Schmisser[2]

Abstract. In the present article, we investigate nonparametric estimation of the unknown drift function b in an integrated Lévy driven jump diffusion model. Our aim will be to estimate the drift on a compact set based on a high-frequency data sample.

Instead of observing the jump diffusion process V itself, we observe a discrete and high-frequent sample of the integrated process

$$X_t := \int_0^t V_s \mathrm{d}s.$$

Based on the available observations of X_t, we will construct an adaptive penalized least-squares estimate in order to compute an adaptive estimator of the corresponding drift function b. Under appropriate assumptions, we will bound the L^2-risk of our proposed estimator. Moreover, we study the behavior of the proposed estimator in various Monte Carlo simulation setups.

Mathematics Subject Classification. 62M09, 62G08

Received May 25, 2017. Accepted January 16, 2018.

1. Introduction

In this paper, we consider a two-dimensional stochastic process $(X_t, V_t)_{t\geq 0}$ such that

$$\begin{aligned} \mathrm{d}X_t &= V_t \mathrm{d}t, \quad X_0 = 0, \\ \mathrm{d}V_t &= b(V_t)\mathrm{d}t + \sigma(V_t)\mathrm{d}W_t + \xi(V_{t_-})\mathrm{d}L_t, \quad V_0 \overset{\mathcal{D}}{=} \eta, \end{aligned} \tag{1.1}$$

where $W = (W_t)_{t\geq 0}$ is a standard Brownian Motion and $L = (L_t)_{t\geq 0}$ is a centered Lévy process with finite variance $\mathbb{E}\left(L_1^2\right) := \int_{\mathbb{R}} y^2 \nu(\mathrm{d}y) < \infty$ such that

$$\mathrm{d}L_t = \int_{\mathbb{R}} z(\mu(\mathrm{d}t, \mathrm{d}z) - \nu(\mathrm{d}z)\mathrm{d}t).$$

Keywords and phrases: Adaptive estimation, integrated jump diffusion, drift estimation, model selection, mean square estimator.

[1] TU Dortmund, Faculty of Mathematics, 44227 Dortmund, Germany.
[2] Université Lille 1, Laboratoire Paul Painlevé, 59655 Villeneuve d'Ascq, France.

[*] Corresponding author: benedikt.funke@math.tu-dortmund.de

Article published by EDP Sciences

W and L are independent and η is independent of both W and L. Moreover, μ denotes the corresponding Poisson random measure of L with intensity measure ν.

Our aim is the nonparametric estimation of the unknown drift function b exclusively based on observations of the first coordinate of (1.1). For our purposes the process X_t will be called an integrated jump diffusion process.

Remark 1.1. We shortly remark, that the system (1.1) is a special case of a two-dimensional stochastic differential equation where no noise is contained in the first coordinate. Moreover, note that the pure jump Lévy process L is a centered L^2-martingale with respect to its augmented canonical filtration under our assumptions.

In many applications in physics, economics or financial mathematics, several occurring stochastic processes can be interpreted as integrated processes, which, for instance, means that at time t they possess cumulatively all information up to this time point. For example, Comte *et al.* [10] refer to a model where V_t denotes the velocity of a particle and X_t represents its coordinate. Further models and applications of such processes in the context of paleoclimate data can be found in Ditlevsen and Sørensen [12] as well as in Lefebvre [17] and Baltazar-Larios and Sørensen [3].

Especially in mathematical finance, if V_t acts as a model for a certain asset price, then X_t denotes the (log-) return of this asset up to time t. If, for instance, only the return series with time lag Δ

$$R_{t,\Delta} := X_{t\Delta} - X_{(t-1)\Delta} = \int_{(t-1)\Delta}^{t\Delta} b(V_t)\mathrm{d}t + \int_{(t-1)\Delta}^{t\Delta} \sigma(V_t)\mathrm{d}W_t + \int_{(t-1)\Delta}^{t\Delta} \xi(V_{t_-})\mathrm{d}L_t$$

is available, our approach allows us to reconstruct the underlying price process and to estimate its drift function; see also Campbell *et al.* [8] for further reasons for investigating the return series rather than the price process itself. In addition, we emphasize that we deal with high-frequency data ($\Delta \to 0$), which is nowadays a common tool for investigating statistical properties of financial processes and which is often readily available to the practitioner. Moreover, we remark that X_t is not assumed to be stationary and a quite simple example for a non-Markovian process with increasing observations as long as V_t stays positive. In addition, many researchers have investigated the estimation of the integrated volatility in stochastic volatility models, which acts a variability measure; see for example Bollerslev and Zhou [7] or Andersen *et al.* [1]. Hence, when assuming that V_t is positive, our model and estimation approach can be applied in this context, too.

Usual estimation schemes for diffusion processes, as for example in Florens-Zmirou [13], Bandi and Phillips [5], Bandi and Nguyen [4] or Comte *et al.* [9], are based on a sample of the original process V. In contrast to this setting, we are now assuming that we cannot observe the process V itself but rather a running integral over this process. In particular, we only observe the first coordinate

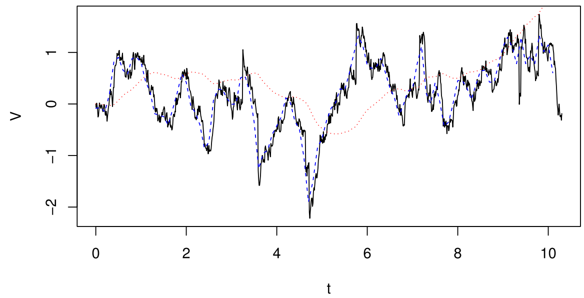
$$X_t = \int_0^t V_s \mathrm{d}s$$

of the original bidimensional process at equidistant time points $k\Delta, k = 1, \ldots, n+2$, over the time interval $[0, T]$, such that

$$T := (n+2)\Delta \to \infty \quad \text{and} \quad \Delta \to 0 \quad \text{as } n \to \infty.$$

Statistical inference for such integrated processes has been, to the best of our knowledge, sporadically investigated. Besides the mentioned articles, further parametric inference has been conducted in some additional works; see for example Gloter [14, 15] as well as Gloter and Gobet [16]. But in general, this topic has not arisen much attention, although it is quite interesting and important for real data applications.

In the nonparametric framework, we are only aware of few works, in which the coefficients of such models have been consistently estimated. For example, Nicolau [20] uses kernel estimators for the pointwise consistent estimation of $b(x)$ and $\sigma^2(x)$. In contrast to the kernel based approach, Comte *et al.* [10] use a model selection approach to construct adaptive nonparametric estimators of b and σ on a fixed compact interval in an integrated



$-\ (V_t)$ $\ldots\ (X_t)$ $\text{- -}\ (\bar{V}_{k\Delta})$

$$\mathrm{d}X_t = V_t \mathrm{d}t, \quad \mathrm{d}V_t = -2V_t + \mathrm{d}W_t + \mathrm{d}L_t, \quad \nu(\mathrm{d}z) = \frac{1}{2}\mathbb{1}_{z=\pm 1}$$

$n = 100$, $\Delta = 10^{-1}$.
V_t is simulated thanks to an Euler scheme.

FIGURE 1. Example of trajectories of (V_t), (X_t) and $(\bar{V}_{k\Delta})$.

diffusion model without jumps. This work extends their approach for estimating ordinary univariate diffusions and was also pursued by Schmisser [21] in the case of univariate jump diffusions.

In view of these two papers, we will conduct an analogous approach for the case of estimating the drift in an integrated jump diffusion model. To the best of our knowledge, adaptive nonparametric inference for the drift function in an integrated jump diffusion model has not been investigated in the literature before. In contrast, empirical likelihood inference for this model has been conducted in Song and Lin [23]. Moreover, a re-weighted kernel estimation procedure has been used by Song *et al.* [24] for estimating the function $\sigma^2 + \xi^2$ and a kernel based approach for estimating b pointwisely has been used in Song [22].

2. Assumptions

Let us at first impose the following assumptions, which guarantee the existence of a unique strong solution (V_t) in equation (1.1); sell also Figure 1 for an example of possible processes under investigation.

A1. i) The functions b, σ and ξ are globally Lipschitz-continuous.

ii) The function σ is bounded away from zero as well as uniformly bounded for all x:

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

iii) The function ξ is non-negative and also bounded:

$$\exists\ \xi_0 \in \mathbb{R}_+ : \forall x \in \mathbb{R} : \quad 0 \leq \xi(x) \leq \xi_0.$$

iv) The function b is elastic (*cf.* Masuda [18]), which means that

$$\exists\ M > 0 : \forall x \in \mathbb{R}, |x| > M : \quad xb(x) \lesssim -x^2.$$

We remark that b cannot be bounded as required in Bandi and Nguyen [4].

v) The Lévy measure ν possesses the properties that

$$\mathrm{Var}\,(L_1) = \int_{\mathbb{R}} y^2 \nu(\mathrm{d}y) = 1, \quad \nu(\{0\}) = 0, \quad \int_{\mathbb{R}} y^4 \nu(\mathrm{d}y) < \infty.$$

Under Assumption A1,i) a unique strong solution (V_t) of (1.1) exists (*cf.* Masuda [18]). Moreover, under A1,i)–iv), this solution is equipped with a unique invariant probability distribution $\varGamma(\mathrm{d}x)$. In addition, V is exponentially β-mixing with mixing coefficient $\beta_V(t)$, which means that

$$\exists\, \gamma > 0: \; \beta_V(t) := \int_{\mathbb{R}} ||P_t(x,\cdot) - \varGamma(\cdot)||_{TV}\varGamma(\mathrm{d}x) = O(e^{-\gamma t}), \quad \text{as } t \to \infty,$$

where $(P_t)_{t\in\mathbb{R}_+}$ denotes the transition probability of the underlying process V and $||\cdot||_{TV}$ defines the total variation norm, see Comte *et al.* [10].

Assumption A1,v) simply ensures that ν has moments up to order 4. Indeed, the condition $\mathrm{Var}\,(L_1) = 1$ is only an identifiability condition.

Using Theorem 2.1 in Masuda [18], we can deduce the ergodicity of (V_t), which means that for all measurable functions $g \in L^1(\varGamma(\mathrm{d}x))$:

$$\frac{1}{T}\int_0^T g(V_s)\mathrm{d}s \longrightarrow \int_{\mathbb{R}} g(x)\varGamma(\mathrm{d}x) \quad \text{a.s., as } T \to \infty.$$

Due to our assumptions on the Lévy measure ν and the Lipschitz-continuity of the coefficients b, σ and ξ, we have that $\mathbb{E}\left(V_t^4\right) < \infty$ for all $t \geq 0$. This can easily be proven by applying the Cauchy-Schwarz inequality successively. We will focus on this property later on.

Moreover, we impose that

A2. vi) $\varGamma$ is absolutely continuous with respect to the Lebesgue measure and, thus, possesses a Lebesgue density π_V such that $\varGamma(\mathrm{d}x) = \pi_V(x)\mathrm{d}x$.

vii) The process (V_t) starts in its invariant law:

$$V_0 \sim \varGamma(\mathrm{d}x)$$

such that (V_t) is stationary.

Remark 2.1. These assumptions are largely congruent to those in Schmisser [21], who investigated the nonparametric estimation of b in the usual non-integrated setting.

We will now concretize our estimation approach. Hence, let us assume that we are aware of a high-frequent data set $\{X_{k\Delta},\; k = 1,\ldots,n+2\}$ of the process (X_t) given by (1.1). As mentioned, the process (V_t) is not observable and has to be approximated. The idea behind our estimation approach relies on the following transformation. We set

$$\bar{V}_{k\Delta} := \bar{V}_k := \frac{1}{\Delta}\left(X_{(k+1)\Delta} - X_{k\Delta}\right) = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} V_s\mathrm{d}s, \quad 1 \leq k \leq n+1.$$

Remark 2.2. We point out that $(\bar{V}_k)_{k\geq 0}$ shares some crucial properties of the underlying process V. According to Comte *et al.* [10], the averaged process $(\bar{V}_k)_{k\geq 0}$ is stationary and exponentially β-mixing, too. The latter fact can be seen due to the fact that

$$\beta_{\bar{V}}(k) \leq \beta_V(k\Delta), \quad k = 1,\ldots,n+1.$$

Let us now start with a very useful proposition acting as a key point for our proofs. The following proposition generalizes Lemmas 7.1–7.3 in Comte *et al.* [10] to the case of integrated jump diffusions.

Proposition 2.3. *Under Assumptions A1 and A2, the following observations hold true:*

a) We have that

$$\bar{V}_k + \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta}(u-k\Delta)dV_u = V_{(k+1)\Delta}, \quad 1 \le k \le n+1.$$

b) For $1 \le k \le n-1$ it holds that

$$Y_{k+1} := \frac{\bar{V}_{k+2}-\bar{V}_{k+1}}{\Delta} = \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)\mathrm{d}V_u,$$

where

$$\psi_k(u) := (u-k\Delta)1_{[k\Delta,(k+1)\Delta[}(u) + ((k+2)\Delta - u)1_{[(k+1)\Delta,(k+2)\Delta[}(u).$$

c) To value the goodness of our used approximation, we state that

$$\mathbb{E}\left((V_{(k+1)\Delta}-\bar{V}_k)^2\right) \lesssim \Delta, \quad 1 \le k \le n+1.$$

d) Additionally, we state that

$$\mathbb{E}\left((V_{(k+1)\Delta}-\bar{V}_k)^4\right) \lesssim \Delta, \quad 1 \le k \le n+1.$$

Based on the sample $\{\bar{V}_k, k = 1,\dots,n+1\}$, we will now propose the drift estimator for the considered model and start with the following decomposition based on Proposition 2.3:

$$\begin{aligned}
Y_{(k+1)\Delta} &:= \frac{\bar{V}_{(k+2)\Delta}-\bar{V}_{(k+1)\Delta}}{\Delta} = \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)\mathrm{d}V_u \\
&= \frac{1}{\Delta^2}\left[\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)b(V_u)\mathrm{d}u + \int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)\sigma(V_u)\mathrm{d}W_u + \int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)\xi(V_{u_-})\mathrm{d}L_u\right] \\
&= b(V_{(k+1)\Delta}) + \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)(b(V_u)-b(V_{(k+1)\Delta}))\mathrm{d}u \\
&\quad + \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)\sigma(V_u)\mathrm{d}W_u + \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)\xi(V_{u_-})\mathrm{d}L_u \\
&= b(\bar{V}_{k\Delta}) + b(V_{(k+1)\Delta}) - b(\bar{V}_{k\Delta}) + \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)(b(V_u)-b(V_{(k+1)\Delta}))\mathrm{d}u \\
&\quad + \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)\sigma(V_u)\mathrm{d}W_u + \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)\xi(V_{u_-})\mathrm{d}L_u \\
&:= b(\bar{V}_{k\Delta}) + R_{k\Delta}^{(1)} + R_{k\Delta}^{(2)} + Z_{k\Delta}^{(1)} + Z_{k\Delta}^{(2)}.
\end{aligned}$$

Hence, $Y_{(k+1)\Delta}$ will act as an approximation of $b(\bar{V}_{k\Delta})$ with

$$R_{k\Delta}^{(1)} = b(V_{(k+1)\Delta}) - b(\bar{V}_{k\Delta}), \quad R_{k\Delta}^{(2)} = \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)(b(V_u)-b(V_{(k+1)\Delta}))\mathrm{d}u$$

and

$$Z_{k\Delta}^{(1)} = \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}(u)\sigma(V_u)\mathrm{d}W_u, \quad Z_{k\Delta}^{(2)} = \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}(u)\xi(V_{u_-})\mathrm{d}L_u.$$

We set

$$R_{k\Delta} := R_{k\Delta}^{(1)} + R_{k\Delta}^{(2)}$$

and

$$Z_{k\Delta} := Z_{k\Delta}^{(1)} + Z_{k\Delta}^{(2)},$$

and denote by

$$\mathscr{F}_t := \sigma(V_0, (W_s)_{0\le s\le t}, (L_s)_{0\le s\le t})$$

the natural filtration of V_t. Let us remark that $\bar{V}_{k\Delta}$ belongs to $\mathscr{F}_{(k+1)\Delta}$ whereas $Z_{k\Delta}$ and $R_{k\Delta}$ belong to $\mathscr{F}_{(k+3)\Delta}$. In order to control the approximation error $R_{k\Delta}$ as well as the noise term $Z_{k\Delta}$ we will need the following lemma.

Lemma 2.4. *Under Assumptions A1 and A2 we have that for* $\Delta \le 1$

a) $\mathbb{E}\left(R_{k\Delta}^2\right) \lesssim \Delta$ and $\mathbb{E}\left(R_{k\Delta}^4\right) \lesssim \Delta$.

b) $\mathbb{E}\left(Z_{k\Delta}^{(1)}|\mathscr{F}_{(k+1)\Delta}\right) = 0$ and $\mathbb{E}\left(Z_{k\Delta}^{(2)}|\mathscr{F}_{(k+1)\Delta}\right) = 0$.

c) $\mathbb{E}\left(\left(Z_{k\Delta}^{(1)}\right)^2|\mathscr{F}_{(k+1)\Delta}\right) \lesssim 1/\Delta$ and $\mathbb{E}\left(\left(Z_{k\Delta}^{(2)}\right)^2|\mathscr{F}_{(k+1)\Delta}\right) \lesssim 1/\Delta$.

d) $\mathbb{E}\left(\left(Z_{k\Delta}^{(1)}\right)^4|\mathscr{F}_{(k+1)\Delta}\right) \lesssim 1/\Delta^2$ and $\mathbb{E}\left(\left(Z_{k\Delta}^{(2)}\right)^4|\mathscr{F}_{(k+1)\Delta}\right) \lesssim 1/\Delta^3$.

3. Spaces of approximation

Let us now turn to our essential aim, namely to estimate nonparametrically the drift function b on a compact set K. To do this, we consider a sequence of nested subspaces $S_0, \dots, S_m, \dots$ such that $\bigcup_{m\in\mathbb{N}_0} S_m$ is dense in $L^2(K)$. 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.* [6]). The rate of convergence of our estimator will depend on the regularity of the drift, *i.e.* its modulus of smoothness.

Definition 3.1 (Modulus of smoothness)**.** The modulus of continuity of a function f at t is defined by

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

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

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

If $f \in \mathscr{C}^r$, then for $1 \leq p \leq \infty$:

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

Definition 3.2 (Besov space)**.** The Besov space $B^\alpha_{2,\infty}$ is the set of functions:

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

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 [11].

We consider a series of nested vectorial subspaces satisfying the following assumptions:

A3. i) The subspaces S_m have finite dimension D_m.

ii) On S_m, the L^2-norm and the L^∞-norm are connected:

$$\exists \phi_1 > 0, \quad \forall m \in \mathbb{N}, \quad \forall s \in S_m, \quad \|s\|^2_\infty \leq \phi_1 D_m \|t\|^2_{L^2}.$$

This implies that, for an orthonormal basis φ_λ of S_m, $\left\| \sum_{\lambda=1}^{D_m} \varphi_\lambda^2 \right\|_\infty \leq \phi_1^2 D_m$.

iii) We can control the bias term: for an integer r called the regularity, there exists a constant $C > 0$ such that for any function $s \in B^\alpha_{2,\infty}$, $\alpha \leq r$, $\forall m \in \mathbb{N}$,

$$\|s - s_m\|_{L^2} \leq C 2^{-m\alpha} \|s\|_{B^\alpha_{2,\infty}},$$

where s_m is the orthogonal projection of s in S_m.

iv) The subspaces are nested: let us set

$$\tilde{\mathscr{M}}_n := \{m \in \mathbb{N},\ D_m \leq N_n\},$$

where N_n is an integer. Then there exists $\mathcal{S}_n$, satisfying properties i), ii) and iii), such that $\forall m \in \tilde{\mathscr{M}}_n$, $S_m \subseteq \mathcal{S}_n$.

Those assumptions are standard for estimation by projection. The subspaces generated by wavelets of regularity r, piecewise polynomials of degree r or trigonometric polynomials satisfy these assumptions (see Meyer [19]); *cf.* Figure 2 for an illustration of an approximation by piecewise linear functions.

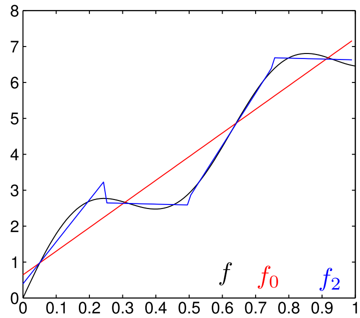
4. Estimation of the drift function

We consider the mean square contrast function

$$\gamma_n(s) := \frac{1}{n}\sum_{k=1}^{n} (s(\bar{V}_{k\Delta}) - Y_{k\Delta})^2 \mathbb{1}_{\bar{V}_{k\Delta} \in K}.$$

For any $m \in \mathscr{M}_n$, where $\mathscr{M}_n := \{m \in \mathbb{N},\ D_m^2 \leq n\Delta/\ln(n)\}$ we consider the contrast estimator

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



1. $S_0 = \{\text{linear functions on } [0,1]\}$.
2. $S_2 = \{\text{linear functions on } [0,1/4[, [1,4,1/2[, [1/2,3/4[, [3/4,1]\}$

FIGURE 2. Example of approximation by piecewise linear functions.

As $V_{k\Delta}$ is not available, we consider the empirical risk

$$\mathscr{R}(s) = \mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n}(s(\bar{V}_{k\Delta}) - b(\bar{V}_{k\Delta}))^2 \mathbb{1}_{\bar{V}_{k\Delta}\in K}\right).$$

The process $\bar{V}_{k\Delta}$ is stationary like $V_{k\Delta}$. We denote by π its stationary density and we assume that this density is bounded from below and above on K:

A4. There exist π_0, π_1 such that for any $x \in K$:

$$0 < \pi_1(x) \leq \pi(x) \leq \pi_0 < \infty.$$

Remark 4.1. Assumption A4 is quite mild, as it only assumes that the stationary density of $\bar{V}_{k\Delta}$ is bounded on a compact set. However, it is not easy to prove. For diffusions, the stationary density of V_t is explicit and Gloter and Gobet [16] give some conditions on the coefficients which ensure the boundedness of π: b and σ have to be bounded, $\mathcal{C}^3$ and their derivatives must also be bounded. However, it is not a necessary condition, as it is also satisfied for Ornstein-Uhlenbeck processes. For jump diffusions, to our knowledge, there do not exist any explicit expression of the stationary density of V_t and it will be quite difficult to express Assumption A4 with respect to certain conditions on the coefficients. However, the simulations show that A4 seems to be satisfied for our Monte Carlo simulation setups.

We obtain the following bound:

Proposition 4.2. *Under Assumptions A1–A4, for any $m \in \mathscr{M}_n$,*

$$\mathscr{R}(\hat{b}_m) \leq \frac{8}{3}\left\|b - b_m\right\|_\pi^2 + 48(\sigma_0^2 + \xi_0^2)\frac{D_m}{n\Delta} + C\Delta + \frac{C'}{n},$$

where b_m is the orthogonal projection of b on S_m and $\|s\|_\pi^2 = \int_K s^2(x)\pi(x)\mathrm{d}x$.

The term $\|b - b_m\|_\pi^2$ is a bias term, which occurs due to the fact that our estimator belongs to S_m. It decreases when m increases. The variance term $D_m/(n\Delta)$ increases with m. Δ and $1/n$ are two remainders terms: Δ appears because the observations are not continuous, it is linked with the difference $b(V_s) - b(\bar{V}_{k\Delta})$, and the term in $1/n$ comes from our approximation method.

We obtain a collection of estimators $(\hat{b}_0, \hat{b}_1, \dots)$ and would like to select the "best" estimator, which is the estimator that minimizes the empirical risk and, in particular, the trade-off between bias and variance terms. If the drift function b belongs to the Besov space $B^{\alpha}_{2,\infty}$, then we have an explicit bound for the bias term:

$$\|b - b_m\|_\pi^2 \leq D_m^{-2\alpha}$$

and the risk of the estimator $\hat{b}_m$ is bounded by

$$\mathscr{R}(\hat{b}_m) \lesssim D_m^{-2\alpha} + \frac{D_m}{n\Delta} + \Delta.$$

This quantity is minimal for $D_{m_{opt}} \propto (n\Delta)^{1/(2\alpha+1)}$. The risk of the optimal estimator $\hat{b}_{m_{opt}}$ satisfies:

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

If $n\Delta^2$ tends to 0, that is if we have high frequency data, $\hat{b}_{m_{opt}}$ converges towards b with the nonparametric rate $(n\Delta)^{-2\alpha/(2\alpha+1)}$.

As we do not usually know the regularity of the drift function b, we now aim at selecting the best estimator without knowing it. Let us introduce the penalty function

$$pen(m) := \kappa(\sigma_0^2 + \xi_0^2) D_m/(n\Delta),$$

which is proportional to the variance term and let us also choose the "best" dimension according to

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

We obtain an adaptive estimator $\hat{b}_{\hat{m}}$. To prove that our estimator selects the "best" dimension m, we make use of Bernstein-type inequalities. We need the following additional assumption in order to control the big jumps of V:

A5. We assume that the Lévy measure ν is sub-exponential:

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

Hence, the tails of the jumps cannot be too heavy.

Remark 4.3. We remark that we only need to control the tail of the jumps. For example, Poisson processes with sub-exponential tails, nearly stable processes or CGMY processes satisfy our assumptions.

We are now ready to state the bound of the L^2-risk of the proposed adaptive drift estimator $\hat{b}_{\hat{m}}$.

Theorem 4.4. *Under Assumptions A1–A5, there exists κ_0 such that for any $\kappa \geq \kappa_0$*

$$\mathbb{E}\left[\left\|\hat{b}_{\hat{m}} - b_m\right\|_n^2\right] \leq \inf_{m \in \mathscr{M}_n} \left\{\frac{8}{3}\|b - b_m\|_\pi^2 + 4pen(m)\right\} + \frac{c}{n} + C\Delta,$$

where $pen(m) = \kappa(\xi_0^2 + \sigma_0^2)D_m/(n\Delta)$ is defined as above.

Remark 4.5. In a comparable model, Song [22] investigated the nonparametric pointwise estimation of the unknown drift b as well as of the function $\sigma^2 + \xi^2$ in an integrated jump diffusion model using a kernel based approach. The resulting estimator is consistent and asymptotically normal distributed possessing a rate of

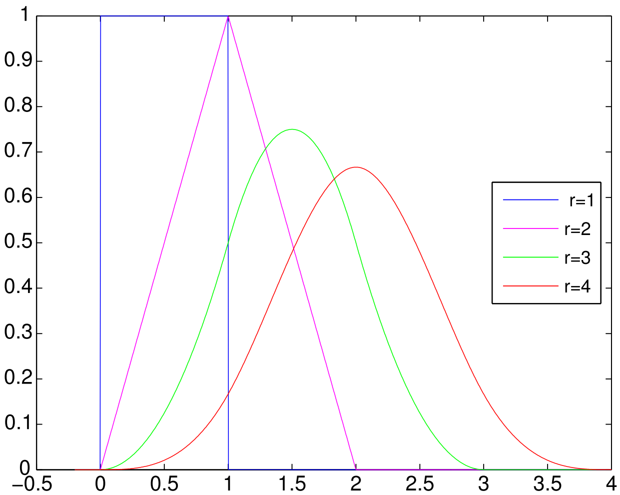


FIGURE 3. Spline functions $\varphi^{(r)}$.

convergence of $\sqrt{n\Delta h}$. In contrast to Song [22], we are interested in estimating adaptively the unknown function b on a compact set K using a model selection approach under quite general assumptions on the driving jump process L. Moreover, we derive the empirical L^2-risk and state assumptions under which the usual nonparametric rate of $(n\Delta)^{-2\alpha/(2\alpha+1)}$ is reached by our estimator as long as $b \in B^{\alpha}_{2,\infty}$.

5. Simulations

In order to practically construct our estimators, we choose the vectorial subspaces generated by spline functions; *cf.* Figure 3.

In that case,

$$S_{m,r} = \left\{ \text{Vect}\left(\left(\varphi^{(r)}_{\lambda,m} \right)_{0 \le m \le 2^m - 1} \right) \right\}, \quad \text{where} \quad \varphi^{(r)}_{\lambda,m}(x) = 2^{m/2} \varphi^{(r)}(2^m(x-\lambda))$$

and

$$\varphi^{(r)} = \mathbb{1}_{[0,1]} * \cdots * \mathbb{1}_{[0,1]}$$

is the r-times convolution product of the indicator function of $[0,1]$. The subspace S_m can also be described as the subspace of all the piecewise polynomials of degree r which belong to $\mathscr{C}^{r-1}$. To obtain the adaptive estimator, we select both (m,r) $(0 \le r \le 7)$ simultaneously. We have the same rate of convergence as if the regularity r was equal to 7.

Let us now focus on the Monte Carlo simulation settings. For each model, we are interested in estimating the drift b on the compact interval $K := [-1,1]$. Thanks to an Euler scheme, we realize for each model five simulations of $(X_0, \dots, X_{n\Delta})$ for the number of observations $n = 10^5$ and the sampling interval $\Delta = 10^{-2}$ and draw the estimators. We also estimate the stationary densities for each model. They look very much alike, therefore we only draw the estimated density of Model 1, *cf.* Figure 4.

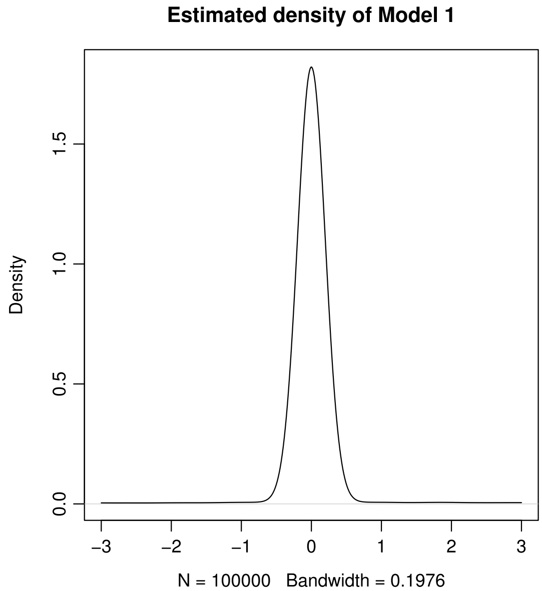
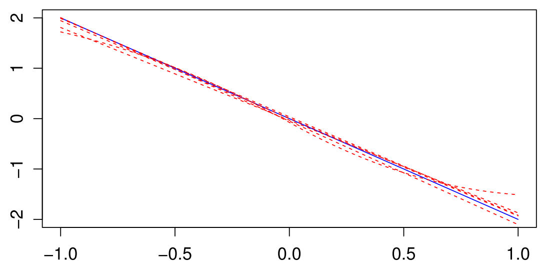


FIGURE 4. Density estimate of $\bar{V}_{k\Delta}$.

$$b(x) = -2x, \quad \sigma(x) = \xi(x) = 1, \quad \nu(\mathrm{d}z) = \frac{1}{2}\mathbb{1}_{z=\pm 1}.$$



Δ	n	risk	or	$\hat{m}$	$\hat{r}$	T_c
10^{-1}	10^3	0.072	1.15	0.04	0.98	0.088
10^{-1}	10^4	0.043	1.00	0	1.02	1.28
10^{-1}	10^5	0.037	1.00	0.02	1.02	34.9
10^{-2}	10^3	0.86	2.83	0.16	0.6	0.031
10^{-2}	10^4	0.076	2.78	0.2	0.98	0.73
10^{-2}	10^5	**0.0055**	1.29	0.04	1.04	11.3
10^{-3}	10^4	0.94	2.87	0.22	0.6	0.22
10^{-3}	10^5	0.068	3.54	0.2	1.02	7.10

FIGURE 5. Model 1.

For each value of (n, Δ), we also realize fifty simulations by an Euler scheme of sampling interval $\delta = \Delta/10$. We compute the estimators $\hat{b}_{\hat{m}}$ and $\hat{b}_{m_{opt}}$ as well as the empirical risks $\hat{R}_n(\hat{b}_{\hat{m}})$ and $\hat{R}_n(\hat{b}_{m_{opt}})$, where

$$\hat{R}_n(t) = \frac{1}{n}\sum_{k=1}^{n}(t(\bar{V}_{k\Delta}) - b(\bar{V}_{k\Delta}))^2.$$

Moreover, we derive the means of $\hat{R}_n(\hat{b}_{\hat{m}})$, denoted by risk, as well as the means of $\hat{m}$ and $\hat{r}$.

In addition, we compute

$$or := \text{mean}\left(\frac{\hat{R}_n(\hat{b}_{\hat{m}})}{\hat{R}_n(\hat{b}_{m_{opt}})}\right)$$

to check that our estimator is really adaptive. Indeed, if the choice of $\hat{m}$ is in some sense good, this quantity should be close to 1. For the sake of completeness, we also give T_c, the mean of the computation times. T_c depends on both $n\Delta$ (and therefore $\mathscr{M}_n$ and the number of estimators $\hat{b}_m$ computed) and n.[1]

From the results it can be seen that for the number of observation $n = 10^5$ and the sampling interval $\Delta = 10^{-2}$, the adaptive estimators are very close to the true drift function (they are nearly superposed). Moreover, the risk of our estimator decreases as the observed time horizon of the underlying process, $T = n\Delta$, increases. This coincides with our theoretical findings in the previous sections. The best results (in bold) are obtained for $n = 10^5, \Delta = 10^{-2}$, that is, Δ small enough, and $n\Delta$ large enough. The oracle is greater for Model 3 (which does not satisfy Assumption A5) than for the other models, especially when $n\Delta$ is big, that is when we can try more models. The choice of the best dimension seems more difficult.

Model 1: Ornstein-Uhlenbeck process with binomial jumps

$$\mathrm{d}V_t = -2V_t\mathrm{d}t + \mathrm{d}W_t + \mathrm{d}L_t$$

with binomial jumps: $\nu(\mathrm{d}z) = \frac{1}{2}\mathbb{1}_{z=\pm 1}$ (Figs. 4 and 5).

Model 2: Cubic function with Laplace jumps

$$\mathrm{d}V_t = \left(-(V_t - 1/4)^3 - (V_t + 1/4)^3\right)\mathrm{d}t + \frac{V_{t^-}^2 + 3}{V_{t^-}^2 + 1}\mathrm{d}W_t + \mathrm{d}L_t$$

with Laplace jumps (Fig. 6):

$$f(\mathrm{d}z) = \nu(\mathrm{d}z) = 0.5e^{-|2^{1/3}z|}.$$

Model 3: Ornstein-Uhlenbeck process with jumps of Student law

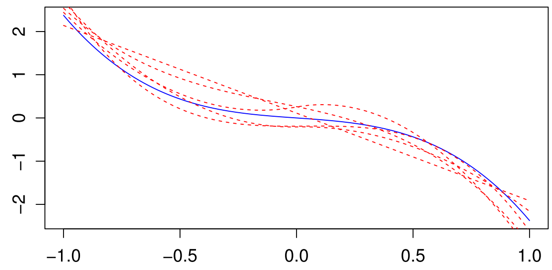
$$\mathrm{d}V_t = -2V_t + \mathrm{d}W_t + \mathrm{d}L_t$$

with L_t a compound Poisson process of intensity $\lambda = 1$ with jumps according to

$$f(\mathrm{d}z) = \nu(\mathrm{d}z) = \frac{1}{\sqrt{6\pi}}\frac{\Gamma(9/2)}{\Gamma(4)}\left(1 + \frac{z^2}{6}\right)^{-\frac{9}{2}}.$$

[1]The programming was done with the software R, the code is available on http://math.univ-lille1.fr/~schmisse/recherche.html.

$$b(x) = -(x + 1/4)^3 - (x - 1/4)^3, \quad \sigma(x) = \frac{x^2 + 3}{x^2 + 1}, \quad \xi^2(x) = 1, \quad \nu(\mathrm{d}z) = 0.5e^{-|2^{1/3}z|}.$$

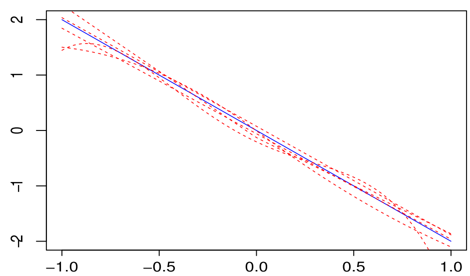


—: true drift - - - : estimators

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

Δ	n	risk	or	$\hat{m}$	$\hat{r}$	T_c
10^{-1}	10^3	0.36	1.23	0.04	0.96	0.077
10^{-1}	10^4	0.27	1.06	0	1	1.00
10^{-1}	10^5	0.25	1.05	0	1	27.2
10^{-2}	10^3	3.58	2.48	0.1	0.2	0.027
10^{-2}	10^4	0.41	1.66	0.1	0.94	0.57
10^{-2}	10^5	**0.12**	1.95	0.2	1.66	8.50
10^{-3}	10^4	3.71	2.34	0.08	0.22	0.17
10^{-3}	10^5	0.45	2.01	0.14	0.98	5.37

FIGURE 6. Model 2.



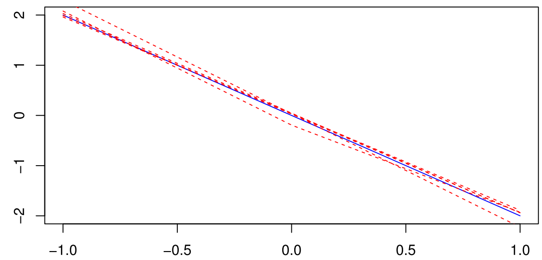
—: true drift - - - : estimators

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

Δ	n	risk	or	$\hat{m}$	$\hat{r}$	T_c
10^{-1}	10^3	0.051	1	0	1	0.12
10^{-1}	10^4	0.038	1	0	1	1.81
10^{-1}	10^5	0.036	1	0	1	37.1
10^{-2}	10^3	0.93	8.37	0.12	0.34	0.034
10^{-2}	10^4	0.060	1.05	0	1.02	0.81
10^{-2}	10^5	**0.0060**	1.97	0.02	1.02	12.6
10^{-3}	10^4	0.94	7.75	0.1	0.4	0.24
10^{-3}	10^5	0.058	1.18	0.02	1.02	7.90

FIGURE 7. Model 3.

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



−: true drift - - : estimators

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

Δ	n	risk	or	$\hat{m}$	$\hat{r}$	T_c
10^{-1}	10^3	0.081	1	0	1	0.092
10^{-1}	10^4	0.043	1	0	1	1.14
10^{-1}	10^5	0.041	1	0	1	34.6
10^{-2}	10^3	0.95	4.45	0.1	0.7	0.031
10^{-2}	10^4	0.080	1.95	0.18	1.04	0.66
10^{-2}	10^5	**0.0051**	1.44	0.04	1.02	10.8
10^{-3}	10^4	0.96	5.12	0.22	0.64	0.20
10^{-3}	10^5	0.068	1.27	0.1	1	6.67

FIGURE 8. Model 4.

This process satisfies Assumptions A1–A2 and A4, but not Assumption A5. Indeed, $\mathbb{E}\left(L_t^8\right) = \infty$ (Fig. 7).

Model 4: Nearly stable Ornstein-Uhlenbeck process

$$\mathrm{d}V_t = -2V_t \mathrm{d}t + \mathrm{d}W_t + \mathrm{d}L_t \quad \text{with} \quad \nu(\mathrm{d}z) = \frac{1}{4z^{5/2}} \mathbb{1}_{|z|\leq 1} \mathrm{d}z.$$

Note that in this model, the jumps have infinite intensity (Fig. 8).

6. Proofs

In this section, we will present the proofs of the stated results. The Burkholder-Davis-Gundy inequality for stochastic integrals driven by L^2-martingales will be one of the keys for the proofs. For the sake of completeness, we will state its formulation at first.

Proposition 6.1 (Applebaum [2]; denoted as Kunita's first inequality). *Let $V = (V_t)_{t\geq 0}$ be the solution of* (1.1) *and let*

$$\mathscr{F}_t := \sigma(V_0, (W_s)_{s\leq t}, (L_s)_{s\leq t}).$$

Then, under Assumptions A1 and A2 for any $p \geq 2$ such that $\int_{\mathbb{R}} |y|^p \nu(dy) < \infty$ and $\int_{\mathbb{R}} y^2 \nu(dy) = 1$, there exists a deterministic positive constant C_p such that

$$\mathbb{E}\left(\sup_{s\in[t,t+\Delta]} \left| \int_t^s \sigma(V_u) \mathrm{d}W_u \right|^p \middle| \mathscr{F}_t \right) \leq C_p \left(\mathbb{E}\left(\left| \int_t^{t+\Delta} \sigma^2(V_u) \mathrm{d}u \right|^{p/2} \middle| \mathscr{F}_t \right)\right)$$

as well as

$$\mathbb{E}\left(\sup_{s\in[t,t+\Delta]}\left|\int_t^s \xi(V_{u_-})\mathrm{d}L_u\right|^p\middle|\mathscr{F}_t\right) \le C_p\mathbb{E}\left(\left|\int_t^{t+\Delta}\xi^2(V_u)\mathrm{d}u\right|^{p/2}\middle|\mathscr{F}_t\right) +C_p\int_{\mathbb{R}}|y|^p\nu(\mathrm{d}y)\mathbb{E}\left(\left(\int_t^{t+\Delta}|\xi(V_u)|^p\mathrm{d}u\right)\middle|\mathscr{F}_t\right).$$

A consequence of this proposition is the following corollary. Its proof is fairly classical and can be found for instance in Gloter [14, Proposition A] for diffusion processes.

Corollary 6.2. *Let $V=(V_t)_{t\ge 0}$ be defined as in Proposition 6.1. Under Assumptions A1 and A2 it exists a constant $C>0$ such that*

$$\mathbb{E}\left(\sup_{s\in[t,t+\Delta]}(V_s-V_t)^2\right)\le C\Delta,$$

for every $t\ge 0$, provided that $\Delta\le 1$.

Moreover, the fourth moment can also be bounded by

$$\mathbb{E}\left(\sup_{s\in[t,t+\Delta]}(V_s-V_t)^4\right)\le \tilde{C}\Delta,$$

for every $t\ge 0$ provided that $\Delta\le 1$ and whereby $\tilde{C}$ denotes another positive and deterministic constant.

6.1. Proof of Proposition 2.3

We start with the proof of a), which is more or less an interchanging of integrals according to

$$\begin{aligned}\bar{V}_k &= \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta}V_s\mathrm{d}s = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta}(V_{(k+1)\Delta}+V_s-V_{(k+1)\Delta})\mathrm{d}s\\ &= \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta}\left(V_{(k+1)\Delta}-\int_s^{(k+1)\Delta}\mathrm{d}V_u\right)\mathrm{d}s = V_{(k+1)\Delta}-\frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^u\mathrm{d}s\right)\mathrm{d}V_u\\ &= V_{(k+1)\Delta}-\frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta}(u-k\Delta)\mathrm{d}V_u = V_{(k+1)\Delta}+\frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta}(k\Delta-u)\mathrm{d}V_u.\end{aligned}$$

By the use of a), we are able to deduce statement b) as follows:

$$\begin{aligned}Y_{k+1} &= \frac{1}{\Delta}\left(V_{(k+3)\Delta}-\frac{1}{\Delta}\int_{(k+2)\Delta}^{(k+3)\Delta}(u-(k+2)\Delta)\mathrm{d}V_u - V_{(k+2)\Delta}+\frac{1}{\Delta}\int_{(k+1)\Delta}^{(k+2)\Delta}(u-(k+1)\Delta)\mathrm{d}V_u\right)\\ &= \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\left((u-(k+1)\Delta)1_{[(k+1)\Delta,(k+2)\Delta)}(u)+((k+3)\Delta-u)1_{[(k+2)\Delta,(k+3)\Delta)}(u)\right)\mathrm{d}V_u\\ &= \frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)\mathrm{d}V_u.\end{aligned}$$

The proof of c) is based on Corollary 6.2 as well as the Cauchy-Schwarz inequality and is derived as follows:

$$\begin{aligned}\mathbb{E}\left((V_{(k+1)\Delta}-\bar{V}_k)^2\right) &= \frac{1}{\Delta^2}\mathbb{E}\left(\left(\int_{k\Delta}^{(k+1)\Delta}(V_{(k+1)\Delta}-V_s)\mathrm{d}s\right)^2\right)\\ &\le \frac{1}{\Delta^2}\int_{k\Delta}^{(k+1)\Delta}\Delta\mathbb{E}\left((V_{(k+1)\Delta}-V_s)^2\right)\mathrm{d}s \lesssim \Delta.\end{aligned}$$

Statement d) can be deduced by using a) and the Cauchy-Schwarz inequality twice as follows

$$\begin{aligned}\mathbb{E}\left(\left(V_{(k+1)\Delta}-\bar{V}_k\right)^4\right) &= \frac{1}{\Delta^4}\mathbb{E}\left(\left(\int_{k\Delta}^{(k+1)\Delta}(V_{(k+1)\Delta}-V_s)\mathrm{d}s\right)^4\right)\\ &\le \frac{1}{\Delta^4}\cdot\mathbb{E}\left(\left(\Delta\int_{k\Delta}^{(k+1)\Delta}(V_{(k+1)\Delta}-V_s)^2\mathrm{d}s\right)^2\right) \le \frac{1}{\Delta^2}\cdot\mathbb{E}\left(\left(\int_{k\Delta}^{(k+1)\Delta}(V_{(k+1)\Delta}-V_s)^2\mathrm{d}s\right)^2\right)\\ &\le \frac{1}{\Delta^2}\cdot\Delta\int_{k\Delta}^{(k+1)\Delta}\mathbb{E}\left((V_{(k+1)\Delta}-V_s)^4\right)\mathrm{d}s \lesssim \frac{1}{\Delta}\cdot\Delta^2=\Delta.\end{aligned}$$

6.2. Proof of Lemma 2.4

Let us start with a). Obviously, we have that

$$\mathbb{E}\left(R_{k\Delta}^2\right) \le 2\left(\mathbb{E}\left(\left(R_{k\Delta}^{(1)}\right)^2\right)+\mathbb{E}\left(\left(R_{k\Delta}^{(2)}\right)^2\right)\right).$$

By using the Lipschitz-continuity of b as well as Proposition 2.3 we can conclude that

$$\mathbb{E}\left((R_{k\Delta}^{(1)})^2\right)=\mathbb{E}\left(\left(b(V_{(k+1)\Delta})-b(\bar{V}_{k\Delta})\right)^2\right)\le C_b^2\cdot\mathbb{E}\left(\left(V_{(k+1)\Delta}-\bar{V}_{k\Delta}\right)^2\right)\lesssim\Delta,$$

where C_b denotes the Lipschitz constant of the drift function b.

Using the Cauchy-Schwarz inequality as well as the fact that

$$\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}^2(u)\mathrm{d}u=\frac{2\Delta^3}{3},$$

the second term can be handled as follows

$$\begin{aligned}\mathbb{E}\left(\left(R_{k\Delta}^{(2)}\right)^2\right) &= \mathbb{E}\left(\left(\frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}(u)(b(V_u)-b(V_{(k+1)\Delta}))\mathrm{d}u\right)^2\right)\\ &\le \frac{1}{\Delta^4}\int_{(k+1)\Delta}^{(k+3)\Delta}\psi_{k+1}^2(u)\mathrm{d}u\cdot\mathbb{E}\left(\int_{(k+1)\Delta}^{(k+3)\Delta}(b(V_u)-b(V_{(k+1)\Delta}))^2\mathrm{d}u\right)\\ &= \frac{2}{3\Delta}\int_{(k+1)\Delta}^{(k+3)\Delta}\mathbb{E}\left((b(V_u)-b(V_{(k+1)\Delta}))^2\right)\mathrm{d}u \le \frac{2C_b^2}{3\Delta}\int_{(k+1)\Delta}^{(k+3)\Delta}\mathbb{E}\left((V_u-V_{(k+1)\Delta})^2\right)\mathrm{d}u\\ &\lesssim \frac{1}{\Delta}\cdot\Delta^2=\Delta.\end{aligned}$$

The fourth moment of $R_{k\Delta}$ is treated in an analogous manner. At first, it holds that

$$\mathbb{E}\left(R_{k\Delta}^4\right) \leq 8\left(\mathbb{E}\left((R_{k\Delta}^{(1)})^4\right) + \mathbb{E}\left((R_{k\Delta}^{(2)})^4\right)\right).$$

Again by Proposition 2.3, statement d), we have that

$$\mathbb{E}\left(\left(R_{k\Delta}^{(1)}\right)^4\right) = \mathbb{E}\left(\left(b(V_{(k+1)\Delta}) - b(\bar{V}_{k\Delta})\right)^4\right) \leq C_b^4 \cdot \mathbb{E}\left(\left(V_{(k+1)\Delta} - \bar{V}_{k\Delta}\right)^4\right) \lesssim \Delta.$$

In order to derive the second summand, we make use of the Cauchy-Schwarz inequality twice:

$$\begin{aligned}
\mathbb{E}\left(\left(R_{k\Delta}^{(2)}\right)^4\right) &= \mathbb{E}\left(\left(\frac{1}{\Delta^2}\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}(u)(b(V_u) - b(V_{(k+1)\Delta}))\mathrm{d}u\right)^4\right) \\
&\leq \frac{1}{\Delta^8}\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}^2(u)\mathrm{d}u\right)^2 \cdot \mathbb{E}\left(\left(\int_{(k+1)\Delta}^{(k+3)\Delta} (b(V_u) - b(V_{(k+1)\Delta}))^2\mathrm{d}u\right)^2\right) \\
&\leq \frac{4\Delta^6}{9\Delta^8}\mathbb{E}\left(\Delta\int_{(k+1)\Delta}^{(k+3)\Delta} (b(V_u) - b(V_{(k+1)\Delta}))^4\mathrm{d}u\right) \leq \frac{4C_b^4}{9\Delta}\int_{(k+1)\Delta}^{(k+3)\Delta} \mathbb{E}\left((V_u - V_{(k+1)\Delta})^4\right)\mathrm{d}u \\
&\lesssim \frac{1}{\Delta}\cdot\Delta^2 = \Delta,
\end{aligned}$$

which concludes the proof of statement a).

Statement b) is a direct consequence of the fact that both $Z_{k\Delta}^{(1)}$ and $Z_{k\Delta}^{(2)}$ are martingale difference sequences with respect to the canonical filtration $\mathscr{F}_t$. We explicitly remark that $\bar{V}_{k\Delta}$ belongs to $\mathscr{F}_{(k+1)\Delta}$ such that $Z_{k\Delta}$ is centered, conditionally on $\bar{V}_{k\Delta}$, by the use of the martingale property of (W_t) and (L_t).

Concerning statement c), we make use of Proposition 6.1 as follows

$$\begin{aligned}
\mathbb{E}\left(\left(Z_{k\Delta}^{(1)}\right)^2 \middle| \mathscr{F}_{(k+1)\Delta}\right) &= \frac{1}{\Delta^4}\mathbb{E}\left(\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}(u)\sigma(V_u)\mathrm{d}W_u\right)^2 \middle| \mathscr{F}_{(k+1)\Delta}\right) \\
&= \frac{1}{\Delta^4}\mathbb{E}\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}^2(u)\sigma^2(V_u)\mathrm{d}u \middle| \mathscr{F}_{(k+1)\Delta}\right) \leq \frac{\sigma_0^2}{\Delta^4}\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}^2(u)\mathrm{d}u \\
&= \frac{\sigma_0^2}{\Delta^4}\cdot\frac{2\Delta^3}{3} \leq \frac{2\sigma_0^2}{3\Delta}.
\end{aligned}$$

In order to handle the Lévy-driven part $Z_{k\Delta}^{(2)}$ we proceed analogously

$$\begin{aligned}
\mathbb{E}\left(\left(Z_{k\Delta}^{(2)}\right)^2 \middle| \mathscr{F}_{(k+1)\Delta}\right) &= \frac{1}{\Delta^4}\mathbb{E}\left(\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}(u)\xi(V_{u_-})\mathrm{d}L_u\right)^2 \middle| \mathscr{F}_{(k+1)\Delta}\right) \\
&= \frac{1}{\Delta^4}\mathbb{E}\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}^2(u)\xi^2(V_u)\mathrm{d}u \middle| \mathscr{F}_{(k+1)\Delta}\right) \leq \frac{\xi_0^2}{\Delta^4}\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}^2(u)\mathrm{d}u \\
&= \frac{\xi_0^2}{\Delta^4}\cdot\frac{2\Delta^3}{3} = \frac{2\xi_0^2}{3\Delta}.
\end{aligned}$$

The fourth conditional moments of $Z^{(1)}_{k\Delta}$ and $Z^{(2)}_{k\Delta}$ can also be treated by Proposition 6.1:

$$\begin{aligned}\mathbb{E}\left(\left(Z^{(1)}_{k\Delta}\right)^4 \middle| \mathscr{F}_{(k+1)\Delta}\right) &= \frac{1}{\Delta^8}\mathbb{E}\left(\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}(u)\sigma(V_u)\mathrm{d}W_u\right)^4 \middle| \mathscr{F}_{(k+1)\Delta}\right)\\ &\lesssim \frac{1}{\Delta^8}\mathbb{E}\left(\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi^2_{k+1}(u)\sigma^2(V_u)\mathrm{d}u\right)^2 \middle| \mathscr{F}_{(k+1)\Delta}\right) \le \frac{\sigma_0^4}{\Delta^8}\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi^2_{k+1}(u)\mathrm{d}u\right)^2\\ &= \frac{\sigma_0^4}{\Delta^8}\cdot\frac{4\Delta^6}{9} \lesssim \frac{1}{\Delta^2}\end{aligned}$$

as well as

$$\begin{aligned}\mathbb{E}\left(\left(Z^{(2)}_{k\Delta}\right)^4 \middle| \mathscr{F}_{(k+1)\Delta}\right) &= \frac{1}{\Delta^8}\mathbb{E}\left(\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi_{k+1}(u)\xi(V_{u_-})\mathrm{d}L_u\right)^4 \middle| \mathscr{F}_{(k+1)\Delta}\right)\\ &\lesssim \frac{1}{\Delta^8}\mathbb{E}\left(\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi^2_{k+1}(u)\xi^2(V_u)\mathrm{d}u\right)^2 \middle| \mathscr{F}_{(k+1)\Delta}\right)\\ &\quad + \frac{1}{\Delta^8}\int_{\mathbb{R}} y^4\nu(\mathrm{d}y)\cdot\mathbb{E}\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi^4_{k+1}(u)\xi^4(V_u)\mathrm{d}u \middle| \mathscr{F}_{(k+1)\Delta}\right)\\ &\le \frac{\xi_0^4}{\Delta^8}\left(\left(\int_{(k+1)\Delta}^{(k+3)\Delta} \psi^2_{k+1}(u)\mathrm{d}u\right)^2 + \int_{\mathbb{R}} y^4\nu(\mathrm{d}y)\int_{(k+1)\Delta}^{(k+3)\Delta} \psi^4_{k+1}(u)\mathrm{d}u\right)\\ &= \frac{\xi_0^4}{\Delta^8}\left(\frac{4\Delta^6}{9} + \int_{\mathbb{R}} y^4\nu(\mathrm{d}y)\frac{2\Delta^5}{5}\right) \lesssim \frac{1}{\Delta^8}\left(\Delta^6+\Delta^5\right) = \frac{1}{\Delta^2}+\frac{1}{\Delta^3} \lesssim \frac{1}{\Delta^3}\end{aligned}$$

with regard on $\Delta \le 1$ and

$$\int_{(k+1)\Delta}^{(k+3)\Delta} \psi^4_{k+1}(u)\mathrm{d}u = \frac{2\Delta^5}{5}.$$

6.3. Proof of Proposition 4.2

We introduce the empirical norm

$$\|s\|_n^2 = \frac{1}{n}\sum_{k=1}^n s^2(\bar{V}_{k\Delta}).$$

We have that

$$\begin{aligned}\gamma_n(s) &= \frac{1}{n}\sum_{k=1}^n \left(s(\bar{V}_{k\Delta}) - Y_{k\Delta}\right)^2 = \frac{1}{n}\sum_{k=1}^n \left(s(\bar{V}_{k\Delta}) - b(\bar{V}_{k\Delta}) + b(\bar{V}_{k\Delta}) - Y_{k\Delta}\right)^2\\ &= \|s-b\|_n^2 + \gamma_n(b) + \frac{2}{n}\sum_{k=1}^n \left(s(\bar{V}_{k\Delta}) - b(\bar{V}_{k\Delta})\right)\left(b(\bar{V}_{k\Delta}) - Y_{k\Delta}\right).\end{aligned}$$

Therefore, as $Y_{k\Delta} = b(\bar{V}_{k\Delta}) + R_{k\Delta} + Z_{k\Delta}$,

$$\gamma_n(s) - \gamma_n(b) = \|s - b\|_n^2 - \frac{2}{n}\sum_{k=1}^{n}(s(\bar{V}_{k\Delta}) - b(\bar{V}_{k\Delta}))(R_{k\Delta} + Z_{k\Delta}).$$

By definition, $\gamma_n(\hat{b}_m) \leq \gamma_n(b_m)$ and thus

$$\left\|b - \hat{b}_m\right\|_n^2 \leq \|b - b_m\|_n^2 + \frac{2}{n}\sum_{k=1}^{n}(\hat{b}_m(\bar{V}_{k\Delta}) - b_m(V_{k\Delta}))(R_{k\Delta} + Z_{k\Delta}). \tag{6.1}$$

By the use of the Cauchy-Schwarz inequality, it holds for any $a > 0$:

$$\frac{2}{n}\sum_{k=1}^{n}(\hat{b}_m(\bar{V}_{k\Delta}) - b_m(V_{k\Delta}))R_{k\Delta} \leq \frac{1}{a}\left\|\hat{b}_m - b_m\right\|_n^2 + \frac{a}{n}\sum_{k=1}^{n}R_{k\Delta}^2. \tag{6.2}$$

Due to Proposition 2.3, it holds that $\mathbb{E}\left(R_{k\Delta}^2\right) \lesssim \Delta$. Let us consider the linear form

$$\nu_n(s) = \frac{1}{n}\sum_{k=1}^{n} s(\bar{V}_{k\Delta})Z_{k\Delta}.$$

Moreover, let us define $\mathscr{B}_m := \left\{s \in S_m, \|s\|_\pi^2 = 1\right\}$, the unit ball (for the $\|.\|_\pi$ norm) of S_m. We have, for any $c > 0$, by the use of the Cauchy-Schwarz inequality:

$$\frac{2}{n}\sum_{k=1}^{n}(\hat{b}_m(\bar{V}_{k\Delta}) - b_m(\bar{V}_{k\Delta}))Z_{k\Delta} \leq 2\left\|\hat{b}_m - b_m\right\|_\pi \cdot \sup_{s\in\mathscr{B}_m} \nu_n(s) \leq \frac{1}{c}\left\|\hat{b}_m - b_m\right\|_\pi^2 + c\sup_{s\in\mathscr{B}_m} \nu_n^2(s). \tag{6.3}$$

Let us introduce the event

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

on which the norms $\|.\|_\pi$ and $\|.\|_n$ are equivalent.

Note that for any deterministic function s, it holds that

$$\mathbb{E}\left[\|s\|_n\right] = \|s\|_\pi .$$

Ω_n happens nearly all the time, as shown by the following Lemma 6.1 from [9].

Lemma 6.3. *As*

i) $\bar{V}_k$ *is exponentially* β*-mixing,*
ii) $\bar{V}_k$ *is stationary and its stationary density* π *is bounded from below and above on* K*,*
iii) the vectorial subspaces S_m *satisfy Assumption A3,*

then

$$\mathbb{P}(\Omega_n^c) \leq c/n^6. \tag{6.4}$$

We first control the risk on Ω_n. Gathering (6.1)–(6.3),

$$\mathbb{E}\left(\left\|b-\hat{b}_m\right\|_n^2 \mathbb{1}_{\Omega_n}\right) \leq \mathbb{E}\left(\|b-b_m\|_n^2\right)+\frac{1}{a}\mathbb{E}\left(\left\|\hat{b}_m-b_m\right\|_n^2\right)+\frac{1}{c}\mathbb{E}\left(\left\|\hat{b}_m-b_m\right\|_\pi^2\right)$$
$$+c\mathbb{E}\left(\sup_{s\in\mathscr{B}_m}\nu_n^2(s)\right)+a\mathbb{E}\left(R_\Delta^2\right).$$

By the triangular inequality, it holds for any norm that

$$\left\|\hat{b}_m-b_m\right\|^2 \leq 2\left\|\hat{b}_m-b\right\|^2+2\left\|b-b_m\right\|^2.$$

As $b-b_m$ is a deterministic function, we have that

$$\mathbb{E}\left[\|b-b_m\|_n^2\right]=\|b-b_m\|_\pi^2.$$

Moreover, on Ω_n, we conclude the relation $\|s\|_\pi^2 \leq 2\|s\|_n^2$. Therefore, it holds that

$$\mathbb{E}\left(\left\|\hat{b}_m-b_m\right\|_\pi^2 \mathbb{1}_{\Omega_n}\right) \leq 4\mathbb{E}\left(\left\|\hat{b}_m-b\right\|_n^2\right)+2\|b-b_m\|_\pi^2$$

and

$$\mathbb{E}\left(\left\|\hat{b}_m-b_m\right\|_n^2\right) \leq 2\mathbb{E}\left(\left\|\hat{b}_m-b\right\|_n^2\right)+2\|b-b_m\|_\pi^2$$

such that consequently

$$\mathbb{E}\left(\left\|b-\hat{b}_m\right\|_n^2 \mathbb{1}_{\Omega_n}\right)\left(1-\frac{2}{a}-\frac{4}{c}\right) \leq \|b-b_m\|_\pi^2\left(1+\frac{2}{a}+\frac{2}{c}\right)+a\mathbb{E}\left(R_\Delta^2\right)+c\mathbb{E}\left(\sup_{s\in\mathscr{B}_m}\nu_n^2(s)\right).$$

Let us set $a=c=12$, then we have

$$\mathbb{E}\left(\left\|b-\hat{b}_m\right\|_n^2 \mathbb{1}_{\Omega_n}\right) \leq \frac{8}{3}\|b-b_m\|_\pi^2+C\Delta+24\mathbb{E}\left(\sup_{s\in\mathscr{B}_m}\nu_n^2(s)\right) \tag{6.5}$$

and, moreover, let us consider (φ_λ), an orthonormal basis (for the norm $\|.\|_\pi$) of S_m. We have that

$$\mathscr{B}_m=\left\{s\in S_m, s=\sum_\lambda a_\lambda\varphi_\lambda, \sum_\lambda a_\lambda^2\leq 1\right\}.$$

Using the Cauchy-Schwarz inequality, we obtain that

$$\mathbb{E}\left(\sup_{s\in\mathscr{B}_m}\nu_n^2(s)\right)=\mathbb{E}\left(\sup_{\sum_\lambda a_\lambda^2\leq 1}\left(\sum_\lambda a_\lambda\nu_n(\varphi_\lambda)\right)^2\right)$$
$$\leq \sup_{\sum_\lambda a_\lambda^2\leq 1}\left(\sum_\lambda\mathbb{E}\left(\nu_n^2(\varphi_\lambda)\right)\right)\left(\sum_\lambda a_\lambda^2\right)\leq\sum_\lambda\mathbb{E}\left(\nu_n^2(\varphi_\lambda)\right).$$

Moreover,

$$\mathbb{E}\left(\nu_n^2(\varphi_\lambda)\right) = \mathbb{E}\left(\left(\frac{1}{n}\sum_{k=1}^n \varphi_\lambda(\bar{V}_{k\Delta})Z_{k\Delta}\right)^2\right)$$
$$= \mathbb{E}\left(\frac{1}{n^2}\sum_{k=1}^n \varphi_\lambda^2(\bar{V}_{k\Delta})Z_{k\Delta}^2\right) + \frac{2}{n^2}\sum_{j<k}\mathbb{E}\left[\varphi_\lambda(\bar{V}_{k\Delta})\varphi_\lambda(\bar{V}_{j\Delta})Z_{k\Delta}Z_{j\Delta}\right].$$

We first bound the square terms:

$$\mathbb{E}\left(\varphi_\lambda^2(\bar{V}_{k\Delta})Z_{k\Delta}^2\right) = \mathbb{E}\left(\varphi_\lambda^2(\bar{V}_{k\Delta})\mathbb{E}\left(Z_{k\Delta}^2|\mathscr{F}_{(k+1)\Delta}\right)\right) \leq \frac{2}{3}\frac{\sigma_0^2+\xi_0^2}{\Delta}\left\|\varphi_\lambda\right\|_\pi^2 = \frac{2}{3}\frac{\sigma_0^2+\xi_0^2}{\Delta}.$$

If $|j-k| \geq 2$, then $Z_{j\Delta} \in \mathscr{F}_{(j+3)\Delta} \subseteq \mathscr{F}_{(k+1)\Delta}$ and the expectation of the product is null:

$$\mathbb{E}\left(\varphi_\lambda(\bar{V}_{k\Delta})\varphi_\lambda(\bar{V}_{j\Delta})Z_{k\Delta}Z_{j\Delta}\right) = \mathbb{E}\left(\varphi_\lambda(\bar{V}_{k\Delta})\varphi_\lambda(\bar{V}_{j\Delta})Z_{j\Delta}\mathbb{E}\left(Z_{k\Delta}|\mathscr{F}_{(k+1)\Delta}\right)\right) = 0$$

and if $j = k-1$, by the Cauchy-Schwarz inequality,

$$\mathbb{E}\left(\varphi_\lambda(\bar{V}_{k\Delta})\varphi_\lambda(\bar{V}_{j\Delta})Z_{j\Delta}Z_{k\Delta}\right) \leq \left(\mathbb{E}\left(\varphi_\lambda^2(\bar{V}_{k\Delta})Z_{k\Delta}^2\right)\mathbb{E}\left(\varphi_\lambda^2(\bar{V}_{j\Delta})Z_{j\Delta}^2\right)\right)^{1/2} \leq \frac{2}{3}\frac{\sigma_0^2+\xi_0^2}{\Delta}.$$

Therefore:

$$\mathbb{E}\left(\sup_{s\in\mathscr{B}_m}\nu_n^2(s)\right) \leq 2\frac{\sigma_0^2+\xi_0^2}{n\Delta}$$

and by (6.5),

$$\mathbb{E}\left(\left\|b-\hat{b}_m\right\|\mathbb{1}_{\Omega_n}\right) \leq \frac{8}{3}\left\|b-b_m\right\|_\pi^2 + C\Delta + 48(\sigma_0^2+\xi_0^2)\frac{D_m}{n\Delta}.$$

It remains to bound the risk on Ω_n^c. We can remark that $(\hat{b}_m(\bar{V}_\Delta), \hat{b}_m(\bar{V}_{2\Delta}), \dots, \hat{b}_m(\bar{V}_{n\Delta}))$ is the orthogonal projection for the $\|.\|_n$-norm of $(Y_\Delta, \dots, Y_{n\Delta})$. We denote this projection by Π_m and define $\mathbf{Y} := (Y_\Delta, \dots, Y_{n\Delta})$, $\mathbf{R} := (R_\Delta, \dots, R_{n\Delta})$ and $\mathbf{Z} := (Z_\Delta, \dots, Z_{n\Delta})$.

We have that $Y_{k\Delta} = b(\bar{V}_{k\Delta}) + R_{k\Delta} + Z_{k\Delta}$ and

$$\left\|b-\hat{b}_m\right\|_n^2 = \left\|b-\Pi_m\mathbf{Y}\right\|_n^2 = \left\|b-\Pi_m b\right\|_n^2 + \left\|\Pi_m\mathbf{R}+\Pi_m\mathbf{Z}\right\|_n^2$$

and, hence, by the Cauchy-Schwarz inequality

$$\mathbb{E}\left[\left\|b-\hat{b}_m\right\|_n^2\mathbb{1}_{\Omega_n^c}\right] \lesssim \left(\frac{1}{n}\sum_{k=1}^n\mathbb{E}\left(b^4(\bar{V}_{k\Delta})\mathbb{1}_{\bar{V}_{k\Delta}\in K}\right)\mathbb{P}(\Omega_n^c)\right)^{1/2}$$
$$+\left(\frac{1}{n}\sum_{k=1}^n(\mathbb{E}\left[R_{k\Delta}^4\right]+\mathbb{E}\left[Z_{k\Delta}^4\right])\right)^{1/2}(\mathbb{P}(\Omega_n^c))^{1/2}.$$

By Lemmas 2.4 and 6.3 we finally conclude that

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

which ends the proof.

6.4. Proof of Theorem 4.4

As previously, we decompose the risk on Ω_n and Ω_n^c. On Ω_n^c, we obtain the same bound as for the non-adaptive estimator. We bound the risk on Ω_n. We have, for any m, like in (6.5):

$$\mathbb{E}\left[\left\|\hat{b}_{\hat{m}}-b\right\|_n^2 \mathbb{1}_{\Omega_n}\right] \leq \frac{8}{3}\left\|b-b_m\right\|_\pi^2+2pen(m)-2pen(\hat{m})+C\Delta+24\mathbb{E}\left[\sup_{s\in\mathscr{B}_{m,\hat{m}}} \nu_n^2(s)\right],$$

where $\mathscr{B}_{m,m'}$ is the unit ball of the set $S_m+S_{m'}$. Let us introduce the function $p(m,m')$ as follows:

$$12p(m,m'):=pen(m)+pen(m').$$

Then

$$\mathbb{E}\left[\left\|\hat{b}_{\hat{m}}-b\right\|_n^2 \mathbb{1}_{\Omega_n}\right] \leq \frac{8}{3}\left\|b-b_m\right\|_\pi^2+4pen(m)+C\Delta+24\mathbb{E}\left[\sup_{s\in\mathscr{B}_{m,\hat{m}}} \nu_n^2(s)-p(m,\hat{m})\right].$$

The problem is to bound $\nu_n^2(s)$ on a random ball. We have:

$$\mathbb{E}\left[\sup_{s\in\mathscr{B}_{m,\hat{m}}} \nu_n^2(s)-p(m,\hat{m})\right] \leq \sum_{m'}\mathbb{E}\left[\sup_{s\in\mathscr{B}_{m,m'}} \nu_n^2(s)-p(m,m')\right].$$

We follow straightly the proof of Theorem 2 in Schmisser [21]. To bound this term, we use a Bernstein inequality and, moreover, we need to apply a Markov inequality on the term $\exp(\nu_n(s))$. The following proposition is exactly Corollary 5.2.2 of Applebaum [2].

Proposition 6.4. *Let F_t and K_t be two locally integrable and previsible processes and let*

$$Y_t:=\int_0^t F_u \mathrm{d}W_u+\int_0^t K_u \mathrm{d}L_u-\int_0^t\left[\frac{F_u^2}{2}+\int_{\mathbb{R}}\left(e^{K_u z}-1-K_u z\right)\nu(\mathrm{d}z)\right]\mathrm{d}u.$$

If

$$\forall t>0,\quad \mathbb{E}\left(\int_0^t\int_{|z|>1}\left|e^{K_u z}-1\right|\nu(\mathrm{d}z)\mathrm{d}u\right)<\infty,$$

then e^{Y_t} is a $\mathscr{F}_t$-local martingale.

We set

$$F_u:=\frac{1}{\Delta}\sum_{k=1}^n s(\bar{V}_{k\Delta})\sigma(V_u)\psi_{k+1}(u),\qquad K_u:=\frac{1}{\Delta}\sum_{k=0}^n s(\bar{V}_{k\Delta})\xi(V_{u_-})\psi_{k+1}(u),$$

$$A_{\varepsilon,t} := \frac{\varepsilon^2}{2}\int_0^t F_u^2 \mathrm{d}u, \qquad B_{\varepsilon,t} := \int_0^t \int_{\mathbb{R}} \left(e^{\varepsilon K_u z} - \varepsilon K_u z - 1\right)\nu(\mathrm{d}z)\mathrm{d}u$$

and consider

$$M_t := \int_0^t F_u \mathrm{d}W_u + \int_0^t K_u \mathrm{d}L_u, \quad \text{and} \quad Y_{\varepsilon,t} = \varepsilon M_t - A_{\varepsilon,t} - B_{\varepsilon,t}.$$

As $\psi_{k+1}(u) \le \Delta$, $|K_u| \le n\,\|s\|_\infty\,\xi_0$ and for $\varepsilon \le \varepsilon_1 := (\lambda \wedge 1)/(2n\,\|s\|_\infty\,\xi_0)$, by Assumption A5,

$$\mathbb{E}\left(\int_0^t \int_{|z|>1} \left|e^{\varepsilon K_u z} - 1\right| \nu(\mathrm{d}z)\right) < \infty.$$

Then, by Proposition 6.4, $e^{Y_{\varepsilon,t}}$ is a local martingale for $\varepsilon \le \varepsilon_1$. It remains to compute its expectation.

We can remark that

$$\int_0^t \psi_k^2(u)\mathrm{d}u = \frac{2\Delta^3}{3}, \quad \int_0^t \psi_k(u)\psi_{k+1}(u)\mathrm{d}u = \frac{\Delta^3}{6}$$

and the function $\psi_k\psi_j$ is identically null if $|k-j| \ge 2$. Then

$$\begin{aligned} A_{\varepsilon,t} &\le \frac{\varepsilon^2}{2\Delta^2}\sigma_0^2 \sum_{0\le k,j\le n} s(\bar{V}_{k\Delta})s(\bar{V}_{j\Delta}) \int_0^t \psi_{k+1}(u)\psi_{j+1}(u)\mathrm{d}u \\ &\le \frac{\varepsilon^2}{2\Delta^2}\sigma_0^2\,\|s\|_n^2 \sum_{k=0}^n \left(\frac{2\Delta^3}{3} + 2\frac{\Delta^3}{6}\right) = \frac{\varepsilon^2}{2}\Delta n\,\|s\|_n^2\,\sigma_0^2. \end{aligned}$$

Moreover, if $\alpha \le 1 \wedge \lambda$,

$$\int_{\mathbb{R}} (e^{\alpha z} - \alpha z - 1)\nu(\mathrm{d}z) \le C\alpha^2.$$

Then, if $\varepsilon \le \varepsilon_1$,

$$B_{\varepsilon,t} \lesssim \int_0^t \varepsilon^2 K_u^2 \mathrm{d}u \lesssim \frac{\varepsilon^2}{\Delta^2}\xi_0^2 \sum_{k,j} s(\bar{V}_{j\Delta})s(\bar{V}_{k\Delta}) \int_0^t \psi_{k+1}(u)\psi_{j+1}(u)\mathrm{d}u \lesssim \varepsilon^2\Delta n\,\|s\|_n^2\,\xi_0^2.$$

Hence, there exists a $c > 0$, such that for any $\varepsilon \le \varepsilon_1$,

$$A_{\varepsilon,t} + B_{\varepsilon,t} \le cn\Delta\varepsilon^2(\sigma_0^2 + \xi_0^2)\,\|s\|_n^2\,.$$

Using the fact that $\nu_n(s) = \frac{1}{n\Delta}M_{n\Delta}$, we conclude for $\varepsilon \le \varepsilon_1$:

$$\begin{aligned} \mathbb{P}\left(\nu_n(s) \ge \eta, \|s\|_n^2 \le \zeta^2\right) &\le \mathbb{P}\left(e^{\varepsilon M_{n\Delta}} \ge e^{\varepsilon n\Delta\eta}, A_{\varepsilon,n\Delta} + B_{\varepsilon,n\Delta} \le cn\Delta\varepsilon^2(\sigma_0^2 + \xi_0^2)\zeta^2\right) \\ &\le \mathbb{P}\left(e^{Y_{\varepsilon,n\Delta}} \ge \exp\left(n\Delta\eta\varepsilon - cn\Delta\varepsilon^2(\sigma_0^2 + \xi_0^2)\zeta^2\right)\right). \end{aligned}$$

We choose

$$\varepsilon = \frac{\eta}{2c(\sigma_0^2+\xi_0^2)\zeta^2+\eta/\varepsilon_1}$$

such that $\varepsilon \leq \varepsilon_1$ as well as

$$n\Delta\eta\varepsilon - cn\Delta\varepsilon^2(\sigma_0^2+\xi_0^2)\zeta^2 \leq -\frac{\eta^2 n\Delta}{4c((\sigma_0^2+\xi_0^2)\zeta^2+c'\eta\xi_0\left\|s\right\|_\infty}.$$

Let us consider a sequence $\{\tau_N\}$ of increasing stopping times such that $\lim_{N\to\infty}\tau_N=\infty$. Then, as $e^{Y_{\varepsilon,t}}$ is a local martingale, the following equality for the corresponding expectation holds:

$$\mathbb{E}\left(e^{Y_{\varepsilon,t\wedge\tau_N}}\right) = \mathbb{E}\left(e^{Y_{\varepsilon,0}}\right) = 1.$$

Moreover, we have that:

$$\mathbb{P}\left(e^{Y_{\varepsilon,t\wedge\tau_N}} \geq a\right) \leq e^{-a}.$$

Letting $N\to\infty$, we obtain $\mathbb{P}\left(e^{Y_{\varepsilon,t}} \geq a\right) \leq e^{-a}$ such that

$$\mathbb{P}\left(\nu_n(s)\geq\eta, \left\|s\right\|_n^2 \leq \zeta^2\right) \leq \exp\left(-\frac{\eta^2 n\Delta}{4c\left((\sigma_0^2+\xi_0^2)\zeta^2+c'\eta\xi_0\left\|s\right\|_\infty\right)}\right).$$

To conclude the proof, we use a $L^2_\pi - L_\infty$ chaining technique (see [21], Prop. 20).

Let $D_{m,m'} := dim(S_m+S_{m'})$. We finally obtain that

$$\mathbb{E}\left(\sup_{s\in\mathscr{B}_{m,m'}}\nu_n^2(s)-p(m,m')\right) \lesssim (\xi_0^2+\sigma_0^2)\frac{D^{3/2}_{m,m'}}{n\Delta}e^{-\sqrt{D_{m,m'}}}$$

and, therefore, as $\sum_{m'} D^{3/2}_{m,m'}e^{-\sqrt{D_{m,m'}}} = O(1)$,

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

which concludes the proof.

References

[1] T. Andersen, T. Bollerslev, F. Diebold and P. Labys, The distribution of exchange rate volatility. *J. Am. Stat. Assoc.* **91** (2001) 42–55.

[2] D. Applebaum, Lévy Processes and Stochastic Calculus, 2nd edn. Part of *Cambridge Studies in Advanced Mathematics*. Cambridge, UK (2009).

[3] F. Baltazar-Larios and M. Sørensen, Maximum likelihood estimation for integrated diffusion processes, in Contemporary Quantitative Finance: Essays in Honor of Eckhard Platen, edited by C. Chiarella and A. Novikov. Springer Heidelberg, Germany (2010).

[4] F. Bandi and T. Nguyen, On the functional estimation of jump-diffusion models. *J. Econom.* **116** (2003) 293–328.

[5] F. Bandi and P. Phillips, Fully nonparametric estimation of scalar diffusion models. *Econometrica* **71** (2003) 241–243.

[6] A. Barron, L. Birgé and P. Massart, Risk bounds for model selection *v*ia penalization. *Probab. Theory Relat. Fields* **113** (1999) 301–413.

[7] T. Bollerslev and H. Zhou, Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *J. Econom.* **109** (2002) 33–65.

[8] J. Campbell, A. Lo and A. MacKinlay, The Econometrics of Financial Markets. Princeton University Press, Princeton (1997).

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

[10] F. Comte, V. Genon-Catalot and Y. Rozenholc, Nonparametric adaptive estimation for integrated diffusions. *Stoch. Process. Appl.* **119** (2009) 811–834.

[11] R.A. DeVore and G.G. Lorentz, Constructive Approximation. Vol. 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg (1993).

[12] S. Ditlevsen and M. Sørensen, Inference for observations of integrated diffusion processes. *Scand. J. Stat.* **31** (2004) 417–429.

[13] D. Florens-Zmirou, On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.* **30** (1993) 790–804.

[14] A. Gloter, Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM: PS* **4** (2000) 205–227.

[15] A. Gloter, Parameter estimation for a discretely observed integrated process. *Scand. J. Stat.* **33** (2006) 83–104.

[16] A. Gloter and E. Gobet, LAMN property for hidden processes: the case of integrated diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** (2008) 104–128.

[17] M. Lefebvre, On the inverse of the first hitting time problem for bidimensional processes. *J. Appl. Probab.* **34** (1997) 610–622.

[18] H. Masuda, Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stoch. Process. Appl.* **117** (2007) 35–56.

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

[20] J. Nicolau, Nonparametric estimation of second-order stochastic differential equations. *Econom. Theory* **23** (2007) 880–898.

[21] E. Schmisser, Nonparametric adaptive estimation of the drift for a jump diffusion process. *Stoch. Process. Appl.* **124** (2014) 883–914.

[22] Y. Song, Nonparametric estimation for second-order diffusion models in high frequency data. *Singap. Econ. Rev.* **63** (2017) 1–31.

[23] Y. Song and Z. Lin, Empirical likelihood inference for the second-order jump-diffusion model. *Stat. Probab. Lett.* **83** (2013) 184–195.

[24] Y. Song, Z. Lin and H. Wang, Re-weighted nadaraya-watson estimation of second-order jump-diffusion model. *J. Stat. Plan. Inference* **143** (2013) 730–744.