On the nonparametric inference of coefficients of self-exciting jump-diffusion

Chiara Amorino, Charlotte Dion, Arnaud Gloter, Sarah Lemler

On the nonparametric inference of coefficients of self-exciting jump-diffusion

Chiara Amorino[(1)], Charlotte Dion[(2)], Arnaud Gloter[(3)], Sarah Lemler[(4)]

November 24, 2020

Abstract

In this paper, we consider a one-dimensional diffusion process with jumps driven by a Hawkes process. We are interested in the estimations of the volatility function and of the jump function from discrete high-frequency observations in long time horizon. We first propose to estimate the volatility coefficient. For that, we introduce in our estimation procedure a truncation function that allows to take into account the jumps of the process and we estimate the volatility function on a linear subspace of $L^2(A)$ where A is a compact interval of $\mathbb{R}$. We obtain a bound for the empirical risk of the volatility estimator and establish an oracle inequality for the adaptive estimator to measure the performance of the procedure. Then, we propose an estimator of a sum between the volatility and the jump coefficient modified with the conditional expectation of the intensity of the jumps. The idea behind this is to recover the jump function. We also establish a bound for the empirical risk for the non-adaptive estimator of this sum and an oracle inequality for the final adaptive estimator. We conduct a simulation study to measure the accuracy of our estimators in practice and we discuss the possibility of recovering the jump function from our estimation procedure.

Jump diffusion, Hawkes process, Volatility estimation, Nonparametric, Adaptation
AMS: 62G05, 60G55

1 Introduction

The present work focuses on the jump-diffusion process introduced in [18]. It is defined as the solution of the following equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + a(X_{t^-})\sum_{j=1}^{M} dN_t^{(j)}, \tag{1}$$

where X_{t^-} denotes the process of left limits, $N = (N^{(1)}, \dots, N^{(M)})$ is a M-dimensional Hawkes process with intensity function λ and W is the standard Brownian motion independent of N.

(1) Unité de Recherche en Mathématiques, Université du Luxembourg, chiara.amorino@uni.lu.
Chiara Amorino gratefully acknowledges financial support of ERC Consolidator Grant 815703 STAMFORD: Statistical Methods for High Dimensional Diffusions.

(2) LPSM, Sorbonne Universit 75005 Paris charlotte.dion_blanc@sorbonne-universite.fr.

(3) Laboratoire de Mathématiques et Modélisation d'Evry, CNRS, Univ Evry, Université Paris-Saclay, 91037, Evry, France arnaud.gloter@univ-evry.fr

(4) Université Paris-Saclay, École CentraleSupélec, MICS Laboratory, France, sarah.lemler@centralesupelec.fr

Some probabilistic results have been established for this model in [18], such as the ergodicity and the $\beta-$mixing. A second work has then been conducted to estimate the drift function of the model using a model selection procedure and upper bounds on the risk of this adaptive estimator have been established in [17] in the high frequency observations context.

In this work, we are interested in estimating the volatility function σ^2 and the jump function a. The jumps in this process make estimating these two functions difficult. We assume that discrete observations of a path of X are available, at high frequency and on a large time interval.

1.1 Motivation and state of the art

Let us notice first that this model has practical relevance thinking of continuous phenomenon impacted by exterior event, with auto-excitation structure. For example one can think of interest rate model (see [22]) in insurance; then, in neurosciences of the evolution of the membrane potential impacted by the signals of the other neurons around it (see [17]). Indeed, it is common to describe the *spike train* of a neuron through a Hawkes process which models the auto-excitation of the phenomenon: for a certain type of neurons, when it spikes once, the probability that it will spike again increases. Finally, referring to [7] for a complete review on Hawkes process in finance, the reader can see the considered model as a generalisation of the so called mutually-exciting-jump diffusion proposed in [5] to study an asset price evolution. This process generalises Poisson jumps (or Lévy jumps which have independent increments) with auto-exciting jumps and is more tractable than jumps driven by Lévy process.

Nonparametric estimation of coefficients of a stochastic differential equations from the observation of a discrete path is a challenge that has been studied a lot in literature. From frequentist point of view in the high frequency context one can cite [23, 12] and in bayesian one recently in [1].

Nevertheless, the purpose of this article fall more under the scope of statistic for stochastic processes with jumps. The literature for the diffusion with jumps from a pure centred Lévy process is large. For example one can refer to [29] and [31].

The first goal of this work is to estimate the volatility coefficient σ^2. As is it well known, in presence of jumps the approximate quadratic variation based on the squared increments of X no longer converges to the integrated volatility. As in [28], we base the approach on truncated quadratic variation to estimate the coefficient σ^2. Particularly, instead of truncation, we use a smooth function to filter the jumps as it is done in [4] in the classical jump-diffusion context. The structure of the jumps here is very different from the one induced by the pure-jump Lévy-process. Indeed, the increments are not independent and this implies the necessity to develop a proper methodology as the one presented hereafter.

Secondly, we want to recover coefficient a. It is important to note that, as presented in [31], in classical jump-diffusion framework (where a Lévy process is used instead of the Hawkes process for $M = 1$) it is possible to obtain an estimator for the function $\sigma^2 + a^2$ by considering the quadratic increments (without truncation) of the process. This is no longer the case here, due to the form of the intensity function of the Hawkes process. Indeed, we recover a more complicated function to be estimated as explained in the following.

1.2 Main contribution

The estimations of the volatility function and of the jump function in Model (1) are challenging in the sense that we have to take into account the jumps of the Hawkes process. Statistical inference for the volatility and for the jump function in a jump diffusion model with jumps driven by a Hawkes process has never been studied before. As for the estimation of the drift in [17], we assume that the coupled process (X, λ) is ergodic, stationary and exponentially $\beta-$mixing. In order to estimate the volatility in a non-parametric way, we consider as in [4] a truncation of the increments of the quadratic variation of X that allows to judge if a jump occurred on not

in a time interval. We estimate σ^2 on a collection of subspaces of L^2 by minimizing a least squares contrast over each model and we establish for the obtained estimators a bound on the risk. Then, we propose an adaptive selection of the model and we obtain non-asymptotic oracle inequalities for the adaptive estimator that guarantees its theoretical performance.

In the second part of this work, we are interested in the estimation of the jump function. As it has been said before, it is not possible to recover directly the jump function a from the quadratic increments of X, and what appears naturally is the sum of the volatility and of the product of the square of the jump function and the jump intensity. The jump intensity is hard to control properly and it is unobserved. To overcome such a problem we introduce the conditional expectation of the intensity given the observation of X, which leads us to estimate the sum of the volatility and of the product between a^2 and the conditional expectation of the jump intensity given X. We lead again a penalized minimum contrast estimation procedure and we establish a non-asymptotic oracle inequality for the final adaptive estimator. Both adaptive estimator are studied using Talagrand's concentration inequalities.

We then discuss the estimation of a, obtained as a quotient in which we plug the estimators of σ^2 and $g := \sigma^2 + a^2 \times f$, where f is the conditional expectation of the jump intensity that we do not know in practice. We propose to estimate f using a Nadaraya-Watson estimator. We show that the risk of the estimator of a cumulates the errors coming from the estimation of the three functions σ^2, g and the conditional expectation of the jump intensity, which shows how hard it is to estimate correctly a.

Finally we have conducted a simulation study to observe the behavior of our estimators in practice. We compare the empirical risks of our estimators to the risks of the oracle estimator to which we have access in a simulation study (they correspond to the estimator in the collection of models which minimises the empirical error). We show that we can recover rather well the volatility σ^2 and g from our procedure but it is harder to recover the jump function a.

1.3 Plan of the paper

The model is described in Section 2, some assumptions on the model are discussed and we give some ergodic and $\beta-$mixing properties on the process (X_t, λ_t). In Section 3 we detail the estimation procedure to estimate the volatility σ^2, we establish a bound for the risk of the non-adaptive estimator, then we propose an adaptive procedure to choose the best estimator among the collection of estimators and give the oracle inequality for this final estimator. Section 4 is devoted to the estimation of $\sigma^2 + a^2 \times f$, where f is the expectation of the jump intensity λ given X. In this section, we explain why we have to estimate this function, we detail the estimation procedure and establish bounds for the risks of the non-adaptive estimator and of the adaptive estimator. The estimation of the jump coefficient a is discussed in Section 5. In Section 6 we have conducted a simulation study and give a little conclusion and some perspective to this work in Section 7. Finally, the proofs of the main results are detailed in Section 8 and the technical results are proved in Appendix A.

2 Framework and Assumptions

2.1 The Hawkes process

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We define the Hawkes process for $t \geq 0$ through stochastic intensity representation. We introduce the M-dimensional point process $N_t := (N_t^{(1)}, \ldots, N_t^{(M)})$ and its intensity λ is a vector of non-negative stochastic intensity functions given by a collection of baseline intensities. It consists in positive constants ζ_j, for $j \in \{1, \ldots, M\}$, and in $M \times M$ interaction functions $h_{i,j} : \mathbb{R}^+ \to \mathbb{R}^+$, which are measurable functions $(i, j \in \{1, \ldots, M\})$. For

$i \in \{1, \dots, M\}$ we also introduce $n^{(i)}$, a discrete point measure on $\mathbb{R}^-$ satisfying

$$\int_{\mathbb{R}^-} h_{i,j}(t-s) n^{(i)}(ds) < \infty \qquad \text{for all } t \geq 0.$$

They can be interpreted as initial condition of the process. The linear Hawkes process with initial condition $n^{(i)}$ and with parameters $(\zeta_i, h_{i,j})_{1 \leq i,j \leq M}$ is a multivariate counting process $(N_t)_{t \geq 0}$. It is such that for all $i \neq j$, $\mathbb{P}$ - almost surely, $N^{(i)}$ and $N^{(j)}$ never jump simultaneously. Moreover, for any $i \in \{1, \dots, M\}$, the compensator of $N^{(i)}$ is given by $\Lambda_t^{(i)} := \int_0^t \lambda_s^{(i)} ds$, where λ is the intensity process of the counting process N and satisfies the following equation:

$$\lambda_t^{(i)} = \zeta_i + \sum_{j=1}^{M} \int_0^{t^-} h_{i,j}(t-u) dN_u^{(j)} + \sum_{j=1}^{M} \int_{-\infty}^{0} h_{i,j}(t-u) dn_u^{(j)}.$$

We remark that $N_t^{(j)}$ is the cumulative number of events in the j-th component at time t while $dN_t^{(j)}$ represents the number of points in the time increment $[t, t+dt]$. We define $\tilde{N}_t := N_t - \Lambda_t$ and $\bar{\mathcal{F}}_t := \sigma(N_s, 0 \leq s \leq t)$ the history of the counting process N (see Daley and Vere - Jones [15]). The intensity process $\lambda = (\lambda^{(1)}, \dots, \lambda^{(M)})$ of the counting process N is the $\bar{\mathcal{F}}_t$-predictable process that makes $\tilde{N}_t$ a $\bar{\mathcal{F}}_t$-local martingale.

Requiring that the functions $h_{i,j}$ are locally integrable, it is possible to show with standard arguments the existence of a process $(N_t^{(j)})_{t \geq 0}$ (see for example [16]). We denote as ζ_j the exogenous intensity of the process and as $(T_k^{(j)})_{k \geq 1}$ the non-decreasing jump times of the process $N^{(j)}$.

We interpret the interaction functions $h_{i,j}$ (also called kernel function or transfer function) as the influence of the past activity of subject i on the subject j, while the parameter $\zeta_j > 0$ is the spontaneous rate and is used to take into account all the unobserved signals. In the sequel we focus on the exponential kernel functions defined by

$$h_{i,j} : \mathbb{R}^+ \to \mathbb{R}^+, \quad h_{i,j}(t) = c_{ij} e^{-\alpha t}, \quad \alpha > 0, \quad c_{ij} > 0, \quad 1 \leq i, j \leq M.$$

With this choice of $h_{i,j}$ the conditional intensity process (λ_t) is then Markovian. In this case we can introduce the auxiliary Markov process $Y = Y^{(ij)}$:

$$Y_t^{(ij)} = c_{i,j} \int_0^t e^{-\alpha(t-s)} dN_s^{(j)} + c_{i,j} \int_{-\infty}^{0} e^{-\alpha(t-s)} dn_s^{(j)}, \quad 1 \leq i, j \leq M.$$

The intensity can be expressed in terms of sums of these Markovian processes that is, for all $1 \leq i \leq M$

$$\lambda_t^{(i)} = f_i \left(\sum_{j=1}^{M} Y_{t^-}^{(ij)} \right), \qquad \text{with } f_i(x) = \zeta_i + x.$$

We remark that all the point processes $N^{(j)}$ behave as homogeneous Poisson processes with constant intensity ζ_j, before the first occurrence. Then, as soon as the first occurrence appears for a particular $N^{(i)}$, it affects all the process increasing the conditional intensity through the interaction functions $h_{i,j}$.

Let us emphasized that from the work [18], it is possible to not assume the positiveness of the coefficients $c_{i,j}$, taking then $f_i(x) = (\zeta_i + x)_+$. This is particularly important for the neuronal applications where the neurons can be have *excitatory or inhibitory* behavior.

2.2 Model Assumptions

In this work we consider the following jump-diffusion model. We write the process as $M+1$ stochastic differential equations:

$$\begin{cases} d\lambda_t^{(i)} = -\alpha(\lambda_t^{(i)} - \zeta_i)dt + \sum_{j=1}^{M} c_{i,j} dN_t^{(j)}, \quad i = 1, \dots, M \\ dX_t = b(X_t)dt + \sigma(X_t)dW_t + a(X_{t^-}) \sum_{j=1}^{M} dN_t^{(j)}, \end{cases} \tag{2}$$

with $\lambda_0^{(j)}$ and X_0 random variables independent from the others. In particular, $(\lambda_t^{(1)}, \dots, \lambda_t^{(M)}, X_t)$ is a Markovian process for the general filtration

$$\mathcal{F}_t := \sigma(W_s, N_s^{(j)}, \quad j = 1, \dots, M, \quad 0 \le s \le t).$$

We aim at estimating, in a non-parametric way, the volatility σ and the jump coefficient a starting from a discrete observation of the process X. The process X is indeed observed at high frequency on the time interval $[0, T]$. For $0 = t_0 \le t_1 \le \dots \le t_n = T$, the observations are denoted as X_{t_i}. We define $\Delta_{n,i} := t_{i+1} - t_i$ and $\Delta_n := \sup_{i=0,\dots,n} \Delta_{n,i}$. We are here assuming that $\Delta_n \to 0$ and $n\Delta_n \to \infty$, for $n \to \infty$. We suppose that there exists c_1, c_2 such that, $\forall i \in \{0, \dots, n-1\}$, $c_1 \Delta_{min} \le \Delta_{n,i} \le c_2 \Delta_n$. Furthermore we require that there exists $\varepsilon > 0$ arbitrarily small such that

$$n^{\varepsilon} \log n = o(\sqrt{n\Delta_n}). \tag{3}$$

The size parameter M is fixed and finite all along and asymptotic properties are obtained when $T \to \infty$.

Requiring that the size of the discretization step is always the same, as we do asking that the maximal and minimal discretization steps differ only on a constant, is a pretty classical assumption in our framework. On the other side, the step conditions gathered in (3) is more technical and yet essential to show our main results.

Assumption 1 (Assumptions on the coefficients of X)**.**

1. *The coefficients a, b and σ are globally Lipschitz.*
2. *There exist positive constants a_1 and σ_1 such that $|a(x)| < a_1$ and $0 < \sigma^2(x) < \sigma_1^2$ for all $x \in \mathbb{R}$.*
3. *The coefficients b and σ are of class $\mathcal{C}^2$ and there exist positive constants c, c', q such that, for all $x \in \mathbb{R}$, $|b'(x)| + |\sigma'(x)| + |a'(x)| \le c$ and $|b''(x)| + |\sigma''(x)| \le c'(1 + |x|^q)$.*
4. *There exist $d \ge 0$ and $r > 0$ such that, for all x satisfying $|x| > r$, we have $xb(x) \le -dx^2$.*

The first three assumptions ensure the existence of a strong solution X of the considered stochastic differential equation (the proof can be adapted from [27], under the Lipschitz condition on the jump coefficient a). The last assumption is introduced in order to study the longtime behavior of X and to ensure its ergodicity (see [18]). Note that the assumption on a can be relaxed (see [18]).

Assumption 2 (Assumptions on the kernels)**.**

1. *Let H be a matrix such that $H_{i,j} := \int_0^{\infty} h_{i,j}(t)dt = \frac{c_{ij}}{\alpha}$, for $1 \le i, j \le M$. The matrix H has a spectral radius smaller than 1.*
2. *We suppose that $\sum_{j=1}^{M} \zeta_j > 0$ and that the matrix H is invertible.*

3. For all i, j such that $1 \leq i, j \leq M$, $c_{ij} \leq \alpha$.

The first point of the Assumption 2 here above implies that the process (N_t) admits a version with stationary increments (see [10]). In the sequel we always will consider such an assumption satisfied. The process (N_t) corresponds to the asymptotic limit and (λ_t) is a stationary process. The second point of A2 is needed in order to ensure the positive Harris recurrence of the couple (X_t, λ_t). A discussion about it can be found in Section 2.3 of [17].

2.3 Ergodicity and moments

In the sequel, we repeatedly use the ergodic properties of the process $Z_t := (X_t, \lambda_t)$. From Theorem 3.6 in [18] we know that, under Assumptions 1 and 2, the process $(X_t, \lambda_t)_{t\geq 0}$ is positive Harris recurrent with unique invariant measure $\pi(dx)$. Moreover, in [18], the Foster-Lyapunov condition in the exponential frame implies that, for all $t \geq 0$, $\mathbb{E}[X_t^4] < \infty$ (see Proposition 3.4). In the sequel we need X to have arbitrarily big moments and, therefore, we propose a modified Lyapunov function. In particular, following the ideas in [18], we take $V : \mathbb{R} \times \mathbb{R}^{M\times M}$ such that

$$V(x, y) := |x|^m + e^{\sum_{i,j} m_{ij}|y^{(ij)}|}, \tag{4}$$

where $m \geq 2$ is a constant arbitrarily big and $m_{ij} := \frac{k_i}{\alpha}$, being $k \in \mathbb{R}_+^M$ a left eigenvector of H, which exists and has non-negative components under our Assumption 2 (see [18] below Assumption 3.3).
We now introduce the generator of the process $\tilde{Z}_t := (X_t, Y_t)$, defined for sufficiently smooth test function g by

$$A^{\tilde{Z}} g(x, y) = -\alpha \sum_{i,j=1}^{M} y^{(ij)} \partial_{y^{(ij)}} g(x, y) + \partial_x g(x, y) b(x) + \frac{1}{2}\sigma^2(x)\partial_x^2 g(x, y) \tag{5}$$

$$+ \sum_{j=1}^{M} f_j \left(\sum_{k=1}^{M} y^{(jk)} \right) [g(x + a(x), y + \Delta_j) - g(x, y)],$$

with $(\Delta_j)^{(il)} = c_{i,j}\mathbb{1}_{j=l}$, for all $1 \leq i, l \leq M$. Then, the following proposition holds true.

Proposition 1. *Suppose that A1 and A2 hold true. Let V be as in (4). Then, there exist positive constants d_1 and d_2 such that the following Foster-Lyapunov type drift condition holds:*

$$A^{\tilde{Z}} V \leq d_1 - d_2 V.$$

Proposition 1 is proven in the Appendix. As the process λ is included in Y, and so we can recover it starting from Y, the ergodicity of $\tilde{Z}$ implies the ergodicity of Z as well. As a consequence, both X and λ have bounded moments of any order. Let us now add the third assumption

Assumption 3. (X_0, λ_0) *has probability* π.

Then, the process $(X_t, \lambda_t)_{t\geq 0}$ is in its stationary regime.

We recall that the process Z is called β - mixing if $\beta_Z(t) = o(1)$ for $t \to \infty$ and exponentially β - mixing if there exists a constant $\gamma_1 > 0$ such that $\beta_Z(t) = O(e^{-\gamma_1 t})$ for $t \to \infty$, where β_Z is the β - mixing coefficient of the process Z as defined for a Markov process Z with transition semigroup $(P_t)_{t\in\mathbb{R}^+}$, by

$$\beta_Z(t) := \int_{\mathbb{R}\times\mathbb{R}^M} \|P_t(z, .) - \pi\| \, \pi(dz), \tag{6}$$

where $\|\lambda\|$ stands for the total variation norm of a signed measure λ.

Moreover, it is

$$\beta_X(t) := \int_{\mathbb{R}\times\mathbb{R}^M} \left\| P_t^1(z,.) - \pi^X \right\| \pi(dz),$$

where $P_t^1(z,.)$ is the projection on X of $P_t(z,.)$ such that $P_t^1(z,dx) := P_t(z, dx \times \mathbb{R}^M)$ and $\pi^X(dx) := \pi(dx \times \mathbb{R}^M)$ is the projection of π on the coordinate X. Then, according to Theorem 4.9 in [18], under A1-A3 the process $Z_t := (X_t, \lambda_t)$ is exponentially β-mixing and there exist some constant $K, \gamma > 0$ such that

$$\beta_X(t) \leq \beta_Z(t) \leq Ke^{-\gamma t}.$$

3 Estimation procedure of the volatility function

With the background introduced in the previous sections, we are now ready to deal with the estimation of the volatility function, to whom this section is dedicated. We remind the reader that the procedure is based on the observations $(X_{t_i})_{i=1,\dots,n}$.

First of all, in Subsection 3.1, we propose a non-adaptive estimator based on the squared increments of the process X. To do that, we decompose such increments in several terms, aimed to isolate the volatility function. Regarding the other terms, we can recognize a bias term (which we will show being small), the contribution of the brownian part (which is centered) and the contribution of the jumps. To make the latter small as well we introduce a truncation function (see Lemma 2 below). Thus, we can define a contrast function, based on the truncated squared increments of X, and the associated estimator of the volatility. In Proposition 3, which is the main result of this subsection, we prove a bound for the empirical risk of the volatility estimator we propose.

As the presented estimator depends on the model, in Subsection 3.2 we introduce a fully data driven procedure to select automatically the best model in the sense of the empirical risk. We choose the model such that it minimizes the sum between the contrast and a penalization function, as explained in (13). In Theorem 1 we show that the estimator associated to the selected model realizes automatically the best compromise between the bias term and the penalty term.

3.1 Non-adaptive estimator

Let us consider the increments of the process X as follows:

$$\begin{aligned} X_{t_{i+1}} - X_{t_i} &= \int_{t_i}^{t_{i+1}} b(X_s)ds + \int_{t_i}^{t_{i+1}} \sigma(X_s)dW_s + \int_{t_i}^{t_{i+1}} a(X_{s^-}) \sum_{j=1}^{M} dN_s^{(j)} \\ &= \int_{t_i}^{t_{i+1}} b(X_s)ds + Z_{t_i} + J_{t_i} \end{aligned} \tag{7}$$

where Z, J are given in Equation (8):

$$Z_{t_i} := \int_{t_i}^{t_{i+1}} \sigma(X_s)dW_s, \qquad J_{t_i} := \int_{t_i}^{t_{i+1}} a(X_{s^-}) \sum_{j=1}^{M} dN_s^{(j)}. \tag{8}$$

To estimate σ^2 for a diffusion process (without jumps), the idea is to consider the random variables $T_{t_i} := \frac{1}{\Delta_n}(X_{t_{i+1}} - X_{t_i})^2$. Following this idea, we decompose T_{t_i}, in order to isolate the contribution of the volatility computed in X_{t_i}. In particular, Equation (7) yields

$$T_{t_i} = \frac{1}{\Delta_n}(X_{t_{i+1}} - X_{t_i})^2 = \sigma^2(X_{t_i}) + A_{t_i} + B_{t_i} + E_{t_i}, \tag{9}$$

where A, B, E are functions of Z, J:

$$\begin{aligned} A_{t_i} &:= \frac{1}{\Delta_n}\left(\int_{t_i}^{t_{i+1}} b(X_s)ds\right)^2 + \frac{2}{\Delta_n}(Z_{t_i}+J_{t_i})\int_{t_i}^{t_{i+1}}(b(X_s)-b(X_{t_i}))ds \\ &+\frac{1}{\Delta_n}\int_{t_i}^{t_{i+1}}(\sigma^2(X_s)-\sigma^2(X_{t_i}))ds + 2b(X_{t_i})Z_{t_i}, \end{aligned}$$

$$B_{t_i} := \frac{1}{\Delta_n}[Z_{t_i}^2 - \int_{t_i}^{t_{i+1}}\sigma^2(X_s)ds];$$

$$E_{t_i} := 2b(X_{t_i})J_{t_i} + \frac{2}{\Delta_n}Z_{t_i}J_{t_i} + \frac{1}{\Delta_n}J_{t_i}^2.$$

The term A_{t_i} is small, whereas B_{t_i} is centered. In order to make E_{t_i} small as well, we introduce the truncation function $\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})$, for $\beta \in (0, \frac{1}{2})$. It is a smooth version of the indicator function, such that $\varphi(\zeta) = 0$ for each ζ, with $|\zeta| \geq 2$ and $\varphi(\zeta) = 1$ for each ζ, with $|\zeta| \leq 1$. The idea is to use the size of the increment of the process $X_{t_{i+1}} - X_{t_i}$ in order to judge if a jump occurred or not in the interval $[t_i, t_{i+1})$. As it is hard for the increment of X with continuous transition to overcome the threshold $\Delta_{n,i}^\beta$ for $\beta \leq \frac{1}{2}$, we can assert the presence of a jump in $[t_i, t_{i+1})$ if $|X_{t_{i+1}} - X_{t_i}| > \Delta_{n,i}^\beta$. Hence, we consider the random variables

$$T_{t_i}\varphi_{\Delta_{n,i}^\beta}(\Delta_i X) = \sigma^2(X_{t_i}) + \tilde{A}_{t_i} + B_{t_i} + E_{t_i}\varphi_{\Delta_{n,i}^\beta}(\Delta_i X),$$

with

$$\tilde{A}_{t_i} := \sigma^2(X_{t_i})(\varphi_{\Delta_{n,i}^\beta}(\Delta_i X) - 1) + A_{t_i}\varphi_{\Delta_{n,i}^\beta}(\Delta_i X) + B_{t_i}(\varphi_{\Delta_{n,i}^\beta}(\Delta_i X) - 1).$$

Now, the just introduced $\tilde{A}_{t_i}$ is once again a small term, because so A_{t_i} was and because of the fact that the truncation function does not differ a lot from the indicator function, as better justified in Lemma 1 below.

In the sequel, the constant c may change value from line to line.

Lemma 1. *Suppose that A1-A3 hold. Then, for any $k \geq 1$,*

$$\mathbb{E}[|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X) - 1|^k] \leq c\Delta_{n,i}.$$

The proof of Lemma 1 can be found in the Appendix. The same is for the proof of Lemma 2 below, which illustrates the reason why we have introduced a truncation function. Indeed, without the presence of φ, the same Lemma would have held true with just a $c\,\Delta_{n,i}$ in the right hand side. Filtering the contribution of the jumps we can gain an extra $\Delta_{n,i}^{\beta q}$ which, as we will see in Proposition 2, will make the contribution of E_{t_i} small.

Lemma 2. *Suppose that A1-A3 hold. Then, for $q \geq 1$ and for any $k \geq 1$*

$$\mathbb{E}\left[|J_{t_i}|^q \varphi_{\Delta_{n,i}^\beta}^k(\Delta_i X)\right] \leq c\Delta_{n,i}^{1+\beta q}.$$

From Lemmas 1 and 2 here above, it is possible to show the following proposition. Also its proof can be found in the Appendix.

Proposition 2. *Suppose that A1-A3 hold. Then, for $\beta \in (\frac{1}{4}, \frac{1}{2})$,*

1. $\forall \tilde{\varepsilon} > 0$, $\mathbb{E}[\tilde{A}_{t_i}^2] \leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}$, $\quad \mathbb{E}[\tilde{A}_{t_i}^4] \leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}$;

2. $\mathbb{E}[B_{t_i}|\mathcal{F}_{t_i}] = 0$, $\quad \mathbb{E}[B_{t_i}^2|\mathcal{F}_{t_i}] \leq c\sigma_1^4$, $\quad \mathbb{E}[B_{t_i}^4] \leq c$;

3. $\mathbb{E}[|E_{t_i}|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] = c\Delta_{n,i}^{2\beta}$, $\quad \mathbb{E}[E_{t_i}^2\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] \leq c\Delta_{n,i}^{4\beta-1}$, $\quad \mathbb{E}[E_{t_i}^4\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] \leq c\Delta_{n,i}^{8\beta-3}$.

In the Proposition here above it's possible to see in detail in what terms the contribution of $\tilde{A}_{t_i}$ and of the truncation of E_{t_i} are small. Moreover, an analysis of the centered Brownian term B_{t_i} and its powers is proposed.

Based on these variables, we propose a nonparametric estimation procedure for the function $\sigma^2(\cdot)$ on a closed interval A of $\mathbb{R}$. We consider $\mathcal{S}_m$ a linear subspace of $L^2(A)$ such that $\mathcal{S}_m = \text{span}(\varphi_1, \ldots, \varphi_{D_m})$ of dimension D_m, where $(\varphi_i)_i$ is an orthonormal basis of $L^2(A)$. We denote $\tilde{S}_n := \cup_{m\in\mathcal{M}_n}\mathcal{S}_m$, where $\mathcal{M}_n \subset \mathbb{N}$ is a set of indexes for the model collection. The contrast function is defined by

$$\gamma_{n,M}(t) := \frac{1}{n}\sum_{i=0}^{n-1}(t(X_{t_i}) - T_{t_i}\varphi_{\Delta_{n,i}^\beta}(\Delta_i X))^2 \tag{10}$$

with the T_{t_i} given in Equation (9). The associated mean squares contrast estimator is

$$\widehat{\sigma}_m^2 := \arg\min_{t\in\mathcal{S}_m}\gamma_{n,M}(t). \tag{11}$$

We observe that, as $\widehat{\sigma}_m^2$ achieves the minimum, it represents the projection of our estimator on the space $\mathcal{S}_m$. The approximation spaces $\mathcal{S}_m$ have to satisfy the following properties

Assumption 4 (Assumptions on the subspaces)**.**

1. *There exists ϕ_1 such that, for any $t \in \mathcal{S}_m$, $\|t\|_\infty^2 \leq \phi_1 D_m \|t\|^2$.*

2. *The spaces $\mathcal{S}_m$ have finite dimension D_m and are nested: for all $m < m' \in \mathcal{M}_n$, $\mathcal{S}_m \subset \mathcal{S}_{m'}$.*

3. *For any positive d there exists $\tilde{\varepsilon} > 0$ such that, for any $\varepsilon < \tilde{\varepsilon}$, $\sum_{m\in\mathcal{M}_n} e^{-dD_m^{1-\varepsilon}} \leq \Sigma(d)$, where $\Sigma(d)$ denotes a finite constant depending only on d.*

We now introduce the empirical norm

$$\|t\|_n^2 := \frac{1}{n}\sum_{i=0}^{n-1} t^2(X_{t_i}).$$

The main result of this section consists in a bound for $\mathbb{E}[\|\widehat{\sigma}_m^2 - \sigma^2\|_n^2]$, which is gathered in the following proposition. Its proof can be found in Section 8.1.

Proposition 3. *Suppose that A1-A4 hold and that $\beta \in (\frac{1}{4}, \frac{1}{2})$. If $\Delta_n \to 0$ and for some $\varepsilon > 0$ $n^\varepsilon \log n = o(\sqrt{n\Delta_n})$ for $n \to \infty$ and $D_n \leq C\frac{\sqrt{n\Delta_n}}{\log(n)n^\varepsilon}$ for a constant $C > 0$, then the estimator $\widehat{\sigma}_m^2$ of σ^2 on A given by equation (11) satisfies*

$$\mathbb{E}\left[\|\widehat{\sigma}_m^2 - \sigma^2\|_n^2\right] \leq 13\inf_{t\in\mathcal{S}_m}\|t - \sigma^2\|_{\pi^X}^2 + \frac{C_1\sigma_1^4 D_m}{n} + C_2\Delta_n^{4\beta-1} + \frac{C_3\Delta_n^{0\wedge 4\beta-\frac{3}{2}}}{n^2}, \tag{12}$$

with C_1, C_2 and C_3 positive constants.

This inequality measures the performance of our estimator $\hat{\sigma}^2_m$ for the empirical norm. The right hand side of the Equation (12) is decomposed into different types of error. The first term corresponds to the bias term which decreases with the dimension D_m of the space of approximation $\mathcal{S}_m$. The second term corresponds to the variance term, i.e. the estimation error, and contrary to the bias, it increases with D_m. The third term comes from the discretization error and the controls obtained in Proposition 2, taking into account the jumps. Then, the fourth term arise evaluating the norm $\|\widehat{\sigma}^2_m - \sigma^2\|^2_n$ when $\|.\|_n$ are $\|.\|_{\pi^X}$ are not equivalent. This inequality ensures that our estimator $\hat{\sigma}^2_m$ does almost as well as the best approximation of the true function by a function of S_m.

Finally, it should be noted that the variance term is the same as for a diffusion without jumps. Nevertheless, the remainder terms are larger because of the jumps.

3.2 Adaption procedure

We want define a criterion in order to select automatically the best dimension D_m (and so the best model) in the sense of the empirical risk. This procedure should be adaptive, meaning independent of σ^2 and dependent only on the observations. The final chosen model minimizes the following criterion:

$$\widehat{m} := \arg \min_{m \in \mathcal{M}_n} \{\gamma_{n,M}(\widehat{\sigma}^2_m) + \text{pen}_\sigma(m)\}, \tag{13}$$

with $\text{pen}_\sigma(\cdot)$ the increasing function on D_m given by

$$\text{pen}_\sigma(m) := \kappa_1 \frac{D_m}{n} \tag{14}$$

where κ_1 is a constant which has to be calibrated.

Next theorem is proven in Section 8.1.

Theorem 1. *Suppose that A1-A4 hold and that $\beta \in (\frac{1}{4}, \frac{1}{2})$. If $\Delta_n \to 0$ and for some $\varepsilon > 0$ $n^\varepsilon \log n = o(\sqrt{n\Delta_n})$ for $n \to \infty$ and $D_n \leq C \frac{\sqrt{n\Delta_n}}{\log(n)\, n^\varepsilon}$ for $C > 0$, then the estimator $\widehat{\sigma}^2_{\widehat{m}}$ of σ^2 on A given by equations (11) and (13) satisfies*

$$\mathbb{E}\left[\left\|\widehat{\sigma}^2_{\widehat{m}} - \sigma^2\right\|^2_n\right] \leq C_1 \inf_{m \in \mathcal{M}_n} \left\{ \inf_{t \in \mathcal{S}_m} \|t - \sigma^2\|^2_{\pi^X} + \text{pen}_\sigma(m) \right\} + C_2 \Delta_n^{4\beta - 1} + \frac{C_3 \Delta_n^{4\beta - \frac{3}{2}}}{n^2} + \frac{C_4}{n}$$

where $C_1 > 1$ is a numerical constant and C_2, C_3, C_4 are positive constants depending on Δ_n, σ_1 in particular.

This inequality ensures that the final estimator $\hat{\sigma}^2_{\hat{m}}$ realizes automatically the best compromise between the bias term and the penalty term which is of the same order than the variance term.

4 Estimation procedure for both coefficients

In addition to the estimation of the volatility, our goal is to estimate, once again in a nonparametric way, the jump coefficient a. The idea is to study the sum between the volatility and the jump coefficient and to recover consequently a way to estimate a (see Section 5 below). However, what turns out naturally is the volatility plus the product between the jump coefficient and the jump intensity which, as we will see in the sequel, leads to some difficulties. To overcome such difficulties, we must bring ourselves to consider the conditional expectation of the intensity of the jumps with respect to X_{t_i}. In this way we analyze differently the squared increments of the

process X, to highlight the role of the conditional expectation. In particular in the following we use, for the decomposition of the squared increments, ideally the same notation as before: we denote the small bias term as A_{t_i}, the Brownian contribution as B_{t_i} and the jump contribution as E_{t_i}, even if the forms of such terms are no longer the same as in Section 3. In particular, A_{t_i} and E_{t_i} are no longer the same as before and their new definition can be found below, while the Brownian contribution B_{t_i} remains exactly the same. To these, as previously anticipated, a term C_{t_i} deriving from the conditional expectation of the intensity is added.

Besides, as in the previous section, we show that A_{t_i} is small and B_{t_i} is centered. Moreover, in this case we also need the jump part to be centered. Therefore, we consider the compensated measure $d\tilde{N}_t$ instead of dN_t, relocating the difference in the drift.

Let us rewrite the process of interest as:

$$\begin{cases} d\lambda_t^{(j)} = -\alpha(\lambda_t^{(j)} - \zeta_t)dt + \sum_{i=1}^M c_{i,j} dN_t^{(i)} \\ dX_t = (b(X_t) + a(X_{t^-}) \sum_{i=1}^M \lambda_t^{(i)})dt + \sigma(X_t)dW_t + a(X_{t^-}) \sum_{i=1}^M d\tilde{N}_t^{(i)}. \end{cases} \tag{15}$$

We set now

$$J_{t_i} := \int_{t_i}^{t_{i+1}} a(X_{s^-}) \sum_{i=1}^M d\tilde{N}_s^{(i)}. \tag{16}$$

The increments of the process X are such that

$$X_{t_{i+1}} - X_{t_i} = \int_{t_i}^{t_{i+1}} \left(b(X_s) + a(X_{s^-}) \sum_{j=1}^M \lambda_s^{(j)} \right) ds + Z_{t_i} + J_{t_i} \tag{17}$$

where J is given in Equation (16) and Z has not changed and is given in Equation (8). Let us define this time:

$$\begin{aligned} A_{t_i} \;:=\; & \frac{1}{\Delta_n} \left(\int_{t_i}^{t_{i+1}} (b(X_s) + a(X_{s^-}) \sum_{j=1}^M \lambda_s^{(j)}) ds \right)^2 + \frac{1}{\Delta_n} \int_{t_i}^{t_{i+1}} (\sigma^2(X_s) - \sigma^2(X_{t_i})) ds \\ & + \frac{2}{\Delta_n} (Z_{t_i} + J_{t_i}) \left(\int_{t_i}^{t_{i+1}} (b(X_s) - b(X_{t_i})) + (a(X_{s^-}) \sum_{j=1}^M \lambda_s^{(j)} - a(X_{t_i}) \sum_{j=1}^M \lambda_{t_i}^{(j)}) ds \right) \\ & + \frac{1}{\Delta_n} \int_{t_i}^{t_{i+1}} (a^2(X_s) - a^2(X_{t_i})) \sum_{j=1}^M \lambda_s^{(j)} ds + \frac{a^2(X_{t_i})}{\Delta_n} \int_{t_i}^{t_{i+1}} \sum_{j=1}^M (\lambda_s^{(j)} - \lambda_{t_i}^{(j)}) ds \\ & + 2 \left(b(X_{t_i}) + a(X_{t_i}) \sum_{j=1}^M \lambda_{t_i}^{(j)} \right) Z_{t_i} + 2 \left(b(X_{t_i}) + a(X_{t_i}) \sum_{j=1}^M \lambda_{t_i}^{(j)} \right) J_{t_i}, \end{aligned} \tag{18}$$

$$E_{t_i} := \frac{2}{\Delta_n} Z_{t_i} J_{t_i} + \frac{1}{\Delta_n} \left(J_{t_i}^2 - \int_{t_i}^{t_{i+1}} a^2(X_s) \sum_{j=1}^M \lambda_s^{(j)} ds \right). \tag{19}$$

The term A_{t_i} is small, whereas B_{t_i} (which is the same as in the previous section) and E_{t_i} are centered. Moreover, let us define the quantity

$$\sum_{j=1}^M \mathbb{E}[\lambda_{t_i}^{(j)} | X_{t_i}] = \sum_{j=1}^M \frac{\int_{\mathbb{R}^M} z_j \pi(X_{t_i}, z_1, \dots, z_M) dz_1, \dots, dz_M}{\int_{\mathbb{R}^M} \pi(X_{t_i}, z_1, \dots, z_M) dz_1, \dots, dz_M},$$

where π is the invariant density of the process (X, λ), whose existence has been discussed in Section 2.3; and

$$C_{t_i} := a^2(X_{t_i}) \sum_{j=1}^M (\lambda_{t_i}^{(j)} - \mathbb{E}[\lambda_{t_i}^{(j)} | X_{t_i}]).$$

It comes the following decomposition:

$$T_{t_i} = \frac{1}{\Delta_n}(X_{t_{i+1}} - X_{t_i})^2 = \sigma^2(X_{t_i}) + a^2(X_{t_i}) \sum_{j=1}^{M} \mathbb{E}[\lambda_{t_i}^{(j)} | X_{t_i}] + A_{t_i} + B_{t_i} + C_{t_i} + E_{t_i}. \tag{20}$$

In the last decomposition of the squared increments we have isolated the sum of the volatility plus the jump coefficient times the conditional expectation of the intensity with respect to X_{t_i}, which is an object on which we can finally use the same approach as before. Thus, as previously, the other terms need to be evaluate. The term A_{t_i} is small and B_{t_i} and E_{t_i} are centered. Moreover the just added term C_{t_i} is clearly centered, by construction, if conditioned with respect to the random variable X_{t_i} and, as we will see in the sequel, it is enough to get our main results. As explained above Assumption 3, the Foster-Lyapunov condition in the exponential frames implies the existence of bounded moments for λ and so we also get $\mathbb{E}[\lambda_{t_i}^{(j)} | X_{t_i}] < \infty$, for any $j \in \{1, \dots, M\}$.

The properties here above listed are stated in Proposition 4 below, whose proof can be found in the appendix.

Proposition 4. *Suppose that A1 -A3 hold. Then,*

1. $\forall \tilde{\varepsilon} > 0$, $\mathbb{E}[A_{t_i}^2] \leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}$, $\quad \mathbb{E}[A_{t_i}^4] \leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}$;
2. $\mathbb{E}[B_{t_i} | \mathcal{F}_{t_i}] = 0$, $\quad \mathbb{E}[B_{t_i}^2 | \mathcal{F}_{t_i}] \leq c\sigma_1^4$, $\quad \mathbb{E}[B_{t_i}^4] \leq c$;
3. $\mathbb{E}[E_{t_i} | \mathcal{F}_{t_i}] = 0$, $\quad \mathbb{E}[E_{t_i}^2 | \mathcal{F}_{t_i}] \leq \frac{ca_1^4}{\Delta_{n,i}} \sum_{j=1}^{M} \lambda_{t_i}^{(j)}$, $\quad \mathbb{E}[E_{t_i}^4] \leq \frac{c}{\Delta_{n,i}^3}$;
4. $\mathbb{E}[C_{t_i} | X_{t_i}] = 0$, $\quad \mathbb{E}[C_{t_i}^2] \leq c$, $\quad \mathbb{E}[C_{t_i}^4] \leq c$.

From Proposition 4 one can see in detail how small the bias term A_{t_i} is. Moreover, it sheds light to the fact that the Brownian term and the jump term are centered with respect to the filtration $(\mathcal{F}_t)$ while C is centered with respect to the σ-algebra generated by the process X.

4.1 Non-adaptive estimator

Based on variables we have just introduced, we propose a nonparametric estimation procedure for the function

$$g(x) := \sigma^2(x) + a^2(x) f(x) \tag{21}$$

with

$$f(x) = \frac{\sum_{j=1}^{M} \int_{\mathbb{R}^M} z_j \pi(x, z_1, \dots, z_M) dz_1, \dots, dz_M}{\int_{\mathbb{R}^M} \pi(x, z_1, \dots, z_M) dz_1, \dots, dz_M} \tag{22}$$

on a closed interval A. We consider S_m the linear subspace of $L^2(A)$ defined in the previous section for $m \in \mathcal{M}_n$ and satisfying Assumption 4. The contrast function is defined almost as before, since this time we no longer need to truncate the contribution of the jumps. It is, for $t \in \tilde{S}_n$,

$$\gamma_{n,M}(t) := \frac{1}{n} \sum_{i=0}^{n-1} (t(X_{t_i}) - T_{t_i})^2$$

and the T_{t_i} are given in Equation (20) this time. The associated mean squares contrast estimator is

$$\widehat{g}_m := \arg\min_{t \in S_m} \gamma_{n,M}(t). \tag{23}$$

We want to bound the empirical risk $\mathbb{E}[\|\widehat{g}_m - g\|_n^2]$. We state it in next proposition, whose proof can be found in Section 8.2.

Proposition 5. *Suppose that A1-A4 hold. If for some* $\varepsilon > 0$ $\Delta_n \to 0$, $n^\varepsilon \log n = o(\sqrt{n\Delta_n})$ *and* $D_n \leq C\frac{\sqrt{n\Delta_n}}{\log n\, n^\varepsilon}$, *for* $C > 0$, *then the estimator* $\widehat{g}_m$ *of* g *on* A *satisfies, for any* $\tilde{\varepsilon} > 0$,

$$\mathbb{E}\left[\|\widehat{g}_m - g\|_n^2\right] \leq 13 \inf_{t\in\mathcal{S}_m} \|t - g\|_{\pi^X}^2 + \frac{C_1(\sigma_1^4 + a_1^4 + 1)D_m}{n\Delta_n} + C_2\Delta_n^{1-\tilde{\varepsilon}} + \frac{C_3}{n^2\Delta_n^{\frac{3}{2}}}, \tag{24}$$

with C_1, C_2 *and* C_3 *positive constants.*

As in the previous section, this inequality measures the performance of our estimator $\widehat{g}_m^2$ for the empirical norm. The right hand side of Equation (24) is decomposed into four different types of error. The first term corresponds to the bias term which decreases with the dimension D_m of the space of approximation $\mathcal{S}_m$. The second term corresponds to the variance term, i.e. the estimation error, and contrary to the bias, it increases with D_m. The third term comes from the discretization error and the controls obtained in Proposition 4. Then, the fourth term appears when evaluating the norm $\|\widehat{g}_m - g\|_n^2$ when $\|.\|_n$ are $\|.\|_{\pi^X}$ are not equivalent.

Finally, let us compare this result with the bound (12) obtained for the estimator $\widehat{\sigma}_m^2$. The main difference is that the second term is of order $D_m/(n\Delta)$ here, instead of D_m/n as it was previously. As a consequence, in practice the risks will depend mainly on $n\Delta$ for the estimation of g and on n for the estimation of σ^2.

4.2 Adaption procedure

Also for the estimation of g we define a criterion in order to select the best dimension D_m in the sense of the empirical risk. This procedure should be adaptive, meaning independent of g and dependent only on the observations. The final chosen model minimizes the following criterion:

$$\widehat{m} := \arg\min_{m\in\mathcal{M}_n} \{\gamma_{n,M}(\widehat{g}_m) + \mathrm{pen}_g(m)\}, \tag{25}$$

with $\mathrm{pen}_g(\cdot)$ the increasing function on D_m given by

$$\mathrm{pen}_g(m) := \kappa_2 \frac{D_m}{n\Delta_n}, \tag{26}$$

where κ_2 is a constant which has to be calibrated. We remark that $\widehat{m}$ here above introduced is not the same as in Equation (13). The model which minimizes the right hand side of (13) is actually $\widehat{m}_1$ while the one introduced in Equation (25) is $\widehat{m}_2$, but when it does not cause confusion we denote both as $\widehat{m}$ in order to lighten the notation.

We analyse the quantity $\mathbb{E}[\|\widehat{g}_{\widehat{m}} - g\|_n^2]$ in the following theorem, whose proof will be in Section 8.2.

Theorem 2. *Suppose that A1-A4 hold. If for some* $\varepsilon > 0$ $\Delta_n \to 0$, $n^\varepsilon \log n = o(\sqrt{n\Delta_n})$ *and* $D_n \leq C\frac{\sqrt{n\Delta_n}}{\log n\, n^\varepsilon}$ *for* $C > 0$, *then the estimator* $\widehat{g}_{\widehat{m}}$ *of* g *on* A *satisfies, for any* $\tilde{\varepsilon} > 0$,

$$\mathbb{E}\left[\|\widehat{g}_{\widehat{m}} - g\|_n^2\right] \leq C_1 \inf_{m\in\mathcal{M}_n} \left\{\inf_{t\in\mathcal{S}_m} \|t - g\|_{\pi^X}^2 + \mathrm{pen}_g(m)\right\} + C_2\Delta_n^{1-\tilde{\varepsilon}} + \frac{C_3}{n^2\Delta_n^{\frac{3}{2}}} + \frac{C_4}{n\Delta_n}$$

where $C_1 > 1$ *is a numerical constants and* C_2, C_3, C_4 *are positive constants depending on* Δ_n, a_1, σ_1 *in particular.*

This oracle inequality guarantees that our final estimator $\hat{g}_{\hat{m}}$ realizes automatically the best compromise between the bias term and the penalty term which is of the same order than the variance term. Since it is more difficult to estimate g because we have to deal with the conditional expectation of the intensity f, the last two error terms are larger than the ones obtained in Theorem 1 for the estimation of σ^2.

5 Estimation of the jump coefficient

The challenge is to get an estimator of the coefficient $a^2(\cdot)$. A natural idea is to replace the conditional expectation in the definition of g given in Equation (21) by an estimator. Let us remind the reader the notation $f(x) := \sum_{j=1}^{M} \mathbb{E}[\lambda_{t_i}^{(j)} | X_{t_i} = x]$ (see Equation (22)) and

$$g(x) = \sigma^2(x) + a^2(x) f(x).$$

The function f can be estimated through a classical estimator for example the Nadaraya-Watson estimator. This is only possible if the intensity of the Hawkes process is known and the jump times observed. We make this assumption in this section and we denote this estimator $\widehat{f}_h$ where $h > 0$ denotes the bandwidth parameter.

Then, to study an estimator of $a(\cdot)$ of the form $\dfrac{\widehat{g}_{m_2}(x) - \widehat{\sigma}^2_{m_1}(x)}{\widehat{f}_h(x)}$ we also assume that $f > f_0$ on A. We then set:

$$\widehat{a}^2_z := \frac{\widehat{g}_{m_2}(x) - \widehat{\sigma}^2_{m_1}(x)}{\widehat{f}_h(x)} \mathbb{1}_{\widehat{f}_h(x) > f_0/2}$$

with $z = (m_1, m_2, h)$. Let us study this estimator, for the empirical norm.

Due to the disjoint support of the two terms and together with Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
\|\widehat{a}^2_z - a^2\|^2_n &= \left\| \left(\frac{(\widehat{g}_{m_2} - g)}{\widehat{f}_h} + \frac{(\sigma^2 - \widehat{\sigma}^2_{m_1})}{\widehat{f}_h} + \frac{(g - \sigma^2)}{f} \frac{f - \widehat{f}_h}{\widehat{f}_h} \right) \mathbb{1}_{\widehat{f}_h > f_0/2} \right\|^2_n + \left\| \frac{g - \sigma^2}{f} \mathbb{1}_{\widehat{f}_h < f_0/2} \right\|^2_n \\
&\leq \frac{12}{f_0^2} \|\widehat{g}_{m_2} - g\|^2_n + \frac{12}{f_0^2} \|\sigma^2 - \widehat{\sigma}^2_{m_1}\|^2_n + 3 \left\| a^2 \left(\frac{f - \widehat{f}_h}{\widehat{f}_h} \right) \mathbb{1}_{\widehat{f}_h > f_0/2} \right\|^2_n \\
&\quad + \frac{1}{n} \sum_{i=0}^{n-1} a^4(X_{t_i}) \mathbb{1}_{\widehat{f}_h(X_{t_i}) < f_0/2}.
\end{aligned}$$

Besides, if $\widehat{f}_h \leq f_0/2$ then $|\widehat{f}_h - f| > f_0/2$ and as $a^2(\cdot) < a_1^2$ finally:

$$\begin{aligned}
\mathbb{E}[\|\widehat{a}^2_z - a^2\|^2_n] &\leq \frac{12}{f_0^2} \mathbb{E}[\|\widehat{g}_{m_2} - g\|^2_n] + \frac{12}{f_0^2} \mathbb{E}[\|\sigma^2 - \widehat{\sigma}^2_{m_1}\|^2_n] + \frac{12 a_1^4}{f_0^2} \mathbb{E}\left[\left\| f - \widehat{f}_h \right\|^2_n \right] \\
&\quad + \frac{a_1^4}{n} \sum_{i=0}^{n-1} \mathbb{P}(|\widehat{f}_h(X_{t_i}) - f(X_{t_i})| > f_0/2).
\end{aligned}$$

And by Markov's inequality, we obtain:

$$\mathbb{E}[\|\widehat{a}^2_z - a^2\|^2_n] \leq \frac{12}{f_0^2} \left(\mathbb{E}[\|\widehat{g}_{m_2} - g\|^2_n] + \mathbb{E}[\|\sigma^2 - \widehat{\sigma}^2_{m_1}\|^2_n] + 2 a_1^4 \mathbb{E}\left[\left\| f - \widehat{f}_h \right\|^2_n \right] \right). \tag{27}$$

This equation teaches us that the empirical risk of the estimator $\widehat{a}_z$ is upper bound by the sum of the three empirical risks of the estimators of the functions g, σ^2, f. The first two are controlled in Theorem 1 and 2. The last one is more classic. The Nadaraya-Watson can be studied with one or two bandwidth parameters. The upper bound of the L^2-risk for two bandwidth is for example done in [13].

Remark 1. *Let us note here that f can be lower bounded by construction. Indeed, its definition jointly with the fact that $\lambda_{t_i}^{(j)} > \zeta_j$ because of the positiveness of $h_{i,j}$, provides us the wanted lower bound. For $\widehat{a}^2_z$ to be an estimator, f_0 must be known or estimated.*

6 Numerical results

In this section we present our numerical study on synthetic data.

6.1 Simulated data

We simulate the Hawkes process N with $M = 1$ for simplicity and here we denote $(T_k)_k$ the sequence of jump times. In fact, the multidimensional structure of the Hawkes process allows to consider a lot of kind of data, but what is impacting the dynamic of X is the cumulative Hawkes process, thus in that sense we do not loose generality taking $M = 1$. In this case, the intensity process is written as

$$\lambda_t = \xi + (\lambda_0 - \xi)e^{-\alpha t} + \sum_{T_k<t} ce^{-\alpha(t-T_k)}.$$

The initial conditions X_0, λ_0 should be simulated according to the invariant distribution (and λ_0 should be larger than $\xi > 0$). This measure of probability is not explicit. Thus we choose: $\lambda_0 = \xi$ and $X_0 = 2$ in the examples. Also, the exogenous intensities ξ is chosen equal to 0.5, the coefficient c is equal to 0.4 and $\alpha = 5$.

Then we simulate $(X_\Delta, \dots X_{(n+1)\Delta})$ from an Euler scheme with a constant time step $\Delta_i = \Delta$. Because of the additional jump term (when $a \neq 0$), to the best of our knowledge it is not possible to use classical more sophisticated scheme. A simulation algorithm is also detailed in [18] Section 2.3.

In order to challenge the proposed methodology, we investigate different kind of models. In this section we present the results for four models which are the following

(a) $b(x) = -4x$, $\sigma(x) = 1$, $a(x) = \sqrt{2 + 0.5\sin(x)}$,

(b) $b(x) = -2x + \sin(x)$, $\sigma(x) = \sqrt{(3 + x^2)/(1 + x^2)}$, $a(x) = 1$,

(c) $b(x) = -2x$, $\sigma(x) = \sqrt{1 + x^2}$, $a(x) = 1$,

(d) $b(x) = -2x$, $\sigma(x) = \sqrt{1 + x^2}$, $a(x) = x\mathbb{1}_{[-5,5]} + 5\mathbb{1}_{(-\infty,-5)} - 5\mathbb{1}_{(5,+\infty)}$.

The drift is chosen linear in order to satisfy the assumptions and as it is not of interest to study the estimation of b here, keeping the same drift coefficient let us focuses on the differences observed due to the coefficients σ, a. For example in models c) and d), σ does not satisfy assumption 1. Let us now detail the numerical estimation strategy.

6.2 Computation of nonparametric estimators

It is important to remind the reader that the estimation procedures are only based on the observations $(X_{k\Delta})_{k=0,\dots,n}$. Indeed, the estimators $\widehat{\sigma}^2_{\widehat{m}}$ and $\widehat{g}_{\widehat{m}}$ of σ^2 and g respectively defined by (11) and (23), are based on the statistics:

$$T_{k\Delta} = \frac{(X_{(k+1)\Delta} - X_{k\Delta})^2}{\Delta},\ k = 0, \dots, n-1.$$

Estimation of σ^2. To compute $\widehat{\sigma}^2_m$ we use a version of the truncated quadratic variation, through a function φ that vanishes when the increments of the data are too large compared to the typical increments of a continuous diffusion process. Precisely, we choose

$$T^{\varphi}_{k\Delta} := T_{k\Delta} \times \varphi\left(\frac{X_{(k+1)\Delta} - X_{k\Delta}}{\Delta^\beta}\right); \quad \varphi(x) = \begin{cases} 1 & |x| < 1 \\ e^{1/3+1/(|x|^2-4)} & \\ 0 & |x| \geq 2 \end{cases}. \tag{28}$$

This choice for the smooth function φ is discussed in [4].

Estimation of g. As far as the estimation of $g := \sigma^2 + a^2 \times f$ is concerned, we do not know the true conditional expectations $f(x_{t_k}) = \mathbb{E}[\lambda_{t_k} | X_{t_k} = x_{t_k}]$ for all k. Thus we compare the estimations of g to the approximate function $\tilde{g}(x) = \sigma^2(x) + a^2(x) \times NW_{\widehat{h}}(x)$ where the function $f(x) = \dfrac{\int z\pi(x,z)dz}{\pi_X(x)}$, which corresponds to $\mathbb{E}[\lambda | X = x]$, is estimated with the classical Nadaraya-Watson estimator NW_h, where h is the bandwidth parameter. To do so we use the R-package `ksmooth`. Then, $\widehat{h}$ is chosen through a cross-validation leave-one-out procedure. We are aware of the fact that the NW estimator can be defined with two bandwidths (one for the numerator and one for the denominator) as it is presented in [13], but we choose the simplest way here.

Choice of the subspaces of $L^2(A)$ The spaces $\mathcal{S}_m$ are generated by the Fourier basis. The maximal dimension D_n is chosen equal to 20 for this study. The theoretical dimension $\lfloor \sqrt{n\Delta}/n^{\varepsilon} \log(n) \rfloor$ is often too small in practice since we have to consider higher dimension to estimate non-regular functions.

In the theoretical part, the estimation is done on a fixed compact interval A. Here it is slightly different. We consider for each model the random data range as the estimation interval. This is more adapted to a real life data set situation.

6.3 Details on the calibration of the constants

Let us remind the reader that the two penalty functions, pen_σ given in Equation (14) and pen_g given in Equation (26), depend on constants named κ_1, κ_2. These constants need to be chosen one for all for each estimator in order to compute the final adaptive estimators $\widehat{\sigma}^2_{\widehat{m}}$ and $\widehat{g}_{\widehat{m}}$. We explain now how these choices are made.

Choice for the universal constants. In order to chose the universal constants κ_1 and κ_2 we investigate models varying b, a, σ^2 (different from those used to validate the procedure later on) for $n \in \{100, 1000, 10000\}$ and $\Delta \in \{0.1, 0.01\}$. We compute Monte-Carlo estimators of the risks $\mathbb{E}[\|\widehat{\sigma}^2_{\widehat{m}} - \sigma^2\|_n^2]$ and $\mathbb{E}[\|\widehat{g}_{\widehat{m}} - \tilde{g}\|_n^2]$. We choose to do $N_{\text{rep}} = 1000$ repetitions to estimate this expectation by the average:

$$\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \|\widehat{\sigma}^{2,(k)}_{\widehat{m}} - \sigma^2\|_n^2 \quad \text{and} \quad \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \|\widehat{g}^{(k)}_{\widehat{m}} - \tilde{g}\|_n^2.$$

Finally comparing the risks as functions of κ_1, κ_2 leads to select values making a good compromise over all experiences. Applying this procedure we finally choose $\kappa_1 = 100$ and $\kappa_2 = 100$.

Choice for the threshold β. The parameter β appears in Equation (28). This parameter helps the algorithm to decide if the process has jumped or not. The theoretical range of values is $(1/4, 1/2)$. We choose to work with $\beta = 1/4 + 0.01$.

Choice for the bandwidth h. The bandwidth h in the Nadaraya-Watson estimator of the conditional expectation is chosen through a leave-one-out cross-validation procedure. Since the true conditional expectation is unknown we focus of the estimation of $\widetilde{g}$ which depends on this estimator anyway. Indeed it is the estimation procedure of g that is evaluated. Other choices for the best bandwidth exist as the Goldenshluger and Lepski method [21] or a Penalized Comparison to Overfitting [26].

6.4 Results: estimation of the empirical risk

As for the calibration phase, we compute Monte-Carlo estimators of the empirical risks. We choose to do $N_{\text{rep}} = 1000$ repetitions to estimate this expectation by the average on the simulations. In the risk tables 1 and 2, we present for the three models and different values of (Δ, n):

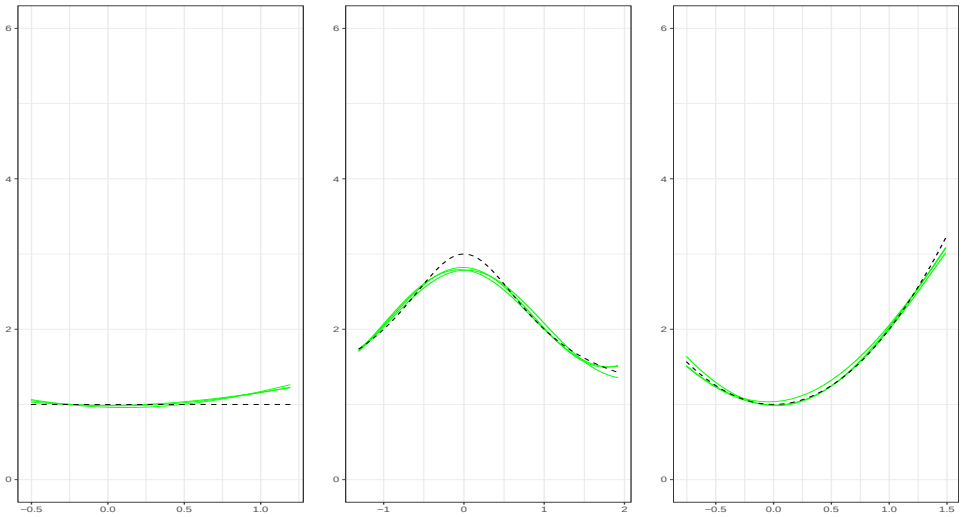

Figure 1: Models (a),(b),(c) with $n = 10000$, $\Delta = 0.01$. Three final estimators are plain green (plain line), true σ^2 plain black (dotted line)

Δ, n	$\Delta = 0.1$ $n = 1000$		$\Delta = 0.1$ $n = 10000$		$\Delta = 0.01$ $n = 10000$	
Model	$\widehat{\sigma}_{\widehat{m}}$	$\widehat{\sigma}_{m^*}$	$\widehat{\sigma}_{\widehat{m}}$	$\widehat{\sigma}_{m^*}$	$\widehat{\sigma}_{\widehat{m}}$	$\widehat{\sigma}_{m^*}$
(a)	0.410 (0.280)	0.361 (0.285)	0.385 (0.122)	0.278 (0.088)	0.015 (0.028)	0.010 (0.023)
(b)	0.187 (1.678)	0.107 (0.989)	0.046 (1.162)	0.027 (1.014)	0.005 (0.015)	0.005 (0.008)
(c)	1.201 (0.216)	0.798 (0.208)	0.452 (0.062)	0.366 (0.042)	0.015 (0.012)	0.008 (0.007)

Table 1: Estimation on a compact interval. Average and standard deviation of the estimated risks $\|\widehat{\sigma}^2_{\widehat{m}} - \sigma^2\|_n^2$ and $\|\widehat{\sigma}^2_{m^*} - \sigma^2\|_n^2$ computed over 1000 repetitions.

the average of the estimated risk over 1000 simulations (MISE) and the standard deviation in the brackets.

Also, we print the result for the *oracle* function in both cases. Indeed, as on simulations we know functions $\sigma^2, \tilde{g}$, we can compute the estimator in the collection $\mathcal{M}_n = \{1, \dots, D_n\}$ which minimises in m the errors $\|\widehat{\sigma}^2_m - \sigma^2\|_n^2$ and $\|\widehat{g}_m - \tilde{g}\|_n^2$. Let us denote the oracle estimators $\widehat{\sigma}^2_{m^*}$ and $\widehat{g}_{m^*}$ respectively. These are not true estimators as they are not available in practice. Nevertheless it is the benchmark. The goal of this numerical study is thus to see how close to the risk results of $\widehat{\sigma}^2_{\widehat{m}}$, $\widehat{g}^2_{\widehat{m}}$ are to the risks of these two *oracle* functions.

Let us detail the result for each estimator.

Estimation of σ^2. Figure 1 shows for models (a),(b),(c), three estimators $\widehat{\sigma}^2_{\widehat{m}}$ in green (light grey) and the true function σ^2 in black (dotted line). We can appreciate here the good reconstruction of the function σ^2 by our estimator.

Table 1 sums up the results of the estimator $\widehat{\sigma}^2_{\widehat{m}}$ for the different models and different parameter choices. We present also the results for the *oracle* estimator $\widehat{\sigma}^2_{m^*}$ as it has been said previously.

The estimations of the MISE and the standard deviation are really close to the oracle ones. As it has been shown in the theoretical part, we can notice that the MISE decreases when n increases. Besides, as the variance term is proportional to $1/n$, when n is fixed and large enough, we can see the clear influence of Δ from 0.1 to 0.01, the MISE are divided at least by 10. The model (c) seems to be the more challenging for the procedure.

Estimation of $\tilde{g}$. Figure 2 shows for each of the three models (a),(b),(c), three estimators $\widehat{g}_{\widehat{m}}$ of $\tilde{g}$ in green (light grey) and function $\tilde{g}$ in black (dotted line). The beams of the three

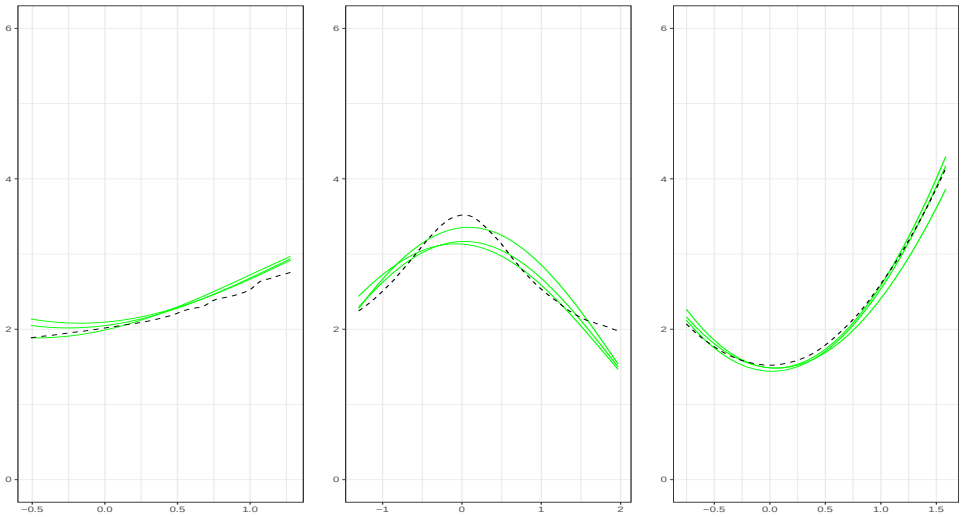

Figure 2: Models (a),(b),(c) with $n = 10000$, $\Delta = 0.01$. Three final estimators of $\tilde{g}$ are plain green (plain line) and $\tilde{g}$ plain black (dotted line).

Δ, n	$\Delta = 0.1$ $n = 1000$		$\Delta = 0.1$ $n = 10000$		$\Delta = 0.01$ $n = 10000$	
Model	$\widehat{g}_{\widehat{m}}$	$\widehat{g}_{m^*}$	$\widehat{g}_{\widehat{m}}$	$\widehat{g}_{m^*}$	$\widehat{g}_{\widehat{m}}$	$\widehat{g}_{m^*}$
(a)	1.363 (0.715)	0.895 (0.606)	0.948 (0.193)	0.735 (0.195)	0.129 (0.141)	0.109 (0.120)
(b)	0.915 (0.520)	0.474 (0.393)	0.313 (0.174)	0.198 (0.079)	0.240 (0.100)	0.098 (0.072)
(c)	0.707 (0.964)	0.311 (0.320)	0.236 (0.202)	0.099 (0.056)	0.073 (0.130)	0.035 (0.035)

Table 2: Estimation on a compact interval. Average and standard deviation of the estimated risks $\|\widehat{g}_{\widehat{m}} - \tilde{g}\|_n^2$ and $\|\widehat{g}_{m^*} - \tilde{g}\|_n^2$ computed over 1000 repetitions.

realisations of the estimator are satisfying.

We observe that the procedure has difficulties in Model (a) and we confirm that impression in Table 2 below with the estimation of the risk. But for the two other models, the estimators seem closer from the true function. The estimation seems to work better in Model (c) than in Model (b) and this is also corroborate by the estimation of the risk given in Table 2.

Table 2 gives the Mean Integrated Squared Errors (MISEs) of the estimator $\widehat{g}_{\widehat{m}}$ obtained from our procedure and of the *oracle* estimator $\widehat{g}_{m^*}$, which is the best one in the collection for the three different models with different values of Δ and n.

As expected, we observe that the MISEs are smaller when n increases and Δ decreases. The different Models (a), (b), (c) gives relatively good results even if as already said, it seems a little bit more difficult to estimate correctly g in Model (a), probably because the volatility σ^2 is constant in this case. For the two other models, the estimators seems to be better. As comparison with the results on the estimation of σ^2, here the variance in proportional to $1/(n\Delta)$ and thus the risks are greater in general.

6.5 Estimation of a^2

As explained is Section 5 the challenge is to get an approximation of the coefficient a from the two previous estimators. A main numerical issue is that, according to the theoretical and numerical results, the best setting for the estimation of σ^2 and g are not the same. Indeed, the smallest Δ is, the best the estimation of σ^2 is, as only large n is important, and on the contrary, $n\Delta$ needs to be large to estimate g properly.

To overcome this difficulty, we choose a thin discretization of the trajectories of X. We simulate here discrete path of the process X at first with $\Delta = 10^{-3}$, $n = 10^5$. Then, we

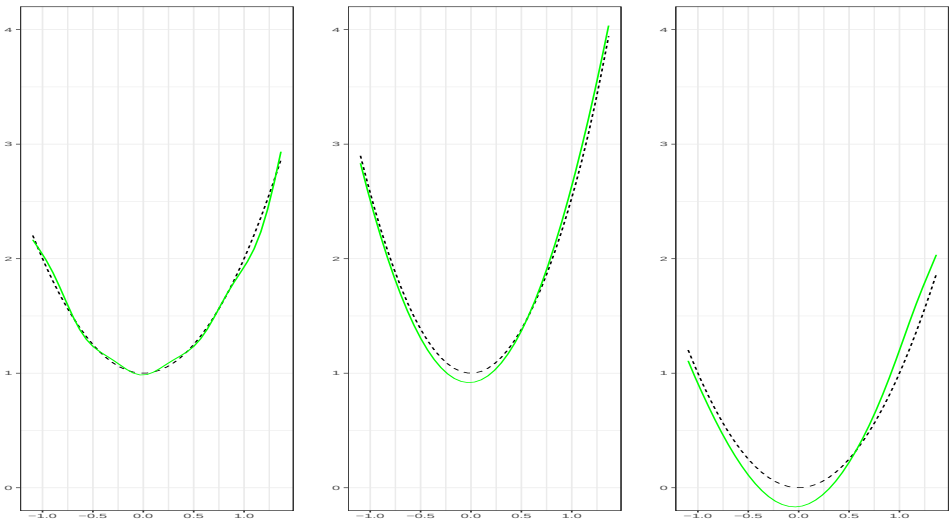

Figure 3: Model (d). Final estimators $\widehat{g}_{\widehat{m}_2}$, $\widehat{\sigma}^2_{\widehat{m}_1}$ and $\widehat{a}$ are plain green (plain line), and true parameters $\tilde{g}$, σ^2 and a^2 in plain black (dotted line) from left to right respectively.

first compute $\widehat{g}_{\widehat{m}_2}$ the estimator of $\tilde{g}$ on all the observations. Secondly, we compute $\widehat{\sigma}^2_{\widehat{m}_1}$ the estimator of σ^2 from a subsample of the discretized observations (one over ten observations thus $\Delta = 0.01,\ n = 10000$).

We finally compute the estimator

$$\widehat{a}^2(x) = \frac{\widehat{g}_{\widehat{m}_2}(x) - \widehat{\sigma}^2_{\widehat{m}_1}(x)}{NW_{\widehat{h}}(x)}.$$

This procedure is slightly different than the one presented in Section 5. Indeed, here we have plugged-in $\widehat{a}^2$ the final estimators of σ^2, g and not computed the all collection $\widehat{a}_z$ with $z = (m_1, m_2, h)$ the parameter to be chosen. Nevertheless, as two procedures to select m_1 and m_2 have been intensively studied before, and the cross-validation method to select h is also well known, this way of doing seems more natural. Besides, the risk bound obtained on $\widehat{a}_z$ in Equation (27) suggests that the better the three functions σ^2, g, f are estimated, the better the estimation of a will be. Nevertheless, one could set up a selection procedure of z in order to minimize the estimation risk on a.

We present on Figure 3 the results obtained on model (d) in which neither σ^2 nor a are constant. Indeed, for the three other models, our procedure has difficulties to estimate properly g, σ^2 and a^2, when one of the parameter of the diffusion jump process is constant. We see that the final estimator $\widehat{a}^2_{\widehat{z}}$ is not so far from the true function a^2 even if there are some fluctuations around the true function. This is understandable because we add the errors coming from the estimations of σ^2 and g as we can see on Inequality (27). Moreover, it should not be forgotten that we do not know exactly g and that we already make an error by estimating $\widetilde{g}$ instead of g, this error is then reflected in the estimate of a^2.

7 Discussion

This paper investigates the jump diffusion model with jumps driven by a Hawkes process. This model is interesting to complete the collection of jump diffusion models and take into account dependency in the jump process. The dynamic of the trajectories obtained from this model is impacted by the Hawkes process which acts independently of the diffusion process.

This work focuses on the estimation of the unknown coefficients σ^2 and a. We propose a classical adaptive estimator of σ^2 based on the truncated increments of the observed discrete

trajectory. this allows to estimate the diffusion coefficient when no jump is detected.

Then, we estimate the sum $g := \sigma^2 + a^2 \times f$. Indeed, it is this function and not $\sigma^2 + a^2$ that can be estimated. The multiplicative term f is the sum of the conditional expectations of the jump process. This function is estimated separately through a Nadaraya-Watson estimator. The proposed estimator of g is built using all increments of the quadratic variation this time.

Furthermore, a main issue is to reach the jump coefficient a from the two first estimators $\widehat{\sigma}^2_{\widehat{m}}$ and $\widehat{g}_{\widehat{m}}$ for which the theoretical and numerical results are convincing. The last section of this article answered this question partially. In fact it is simple to build an estimator of a from the two previous ones, and the estimator of the unknown conditional intensity function f.

Nevertheless, this is possible only if the jumps of the Hawkes process are observed, which is the case of the simulation study. Then, when a real life data arises, the jump times of the counting process must be known to be able to reach a with our methodology. Otherwise, the issue remain an open question.

Then, the proposed estimator $\widehat{a}_z$, with $z = (m_1, m_2, h)$, is a quotient of estimators and the denominator must be lower bounded to insure the proper definition of the estimator. This could be theoretically and numerically carefully studied and be the object for further works.

Moreover, the choice of the $3-$dimensional parameter z could be investigated. Indeed, instead of choosing the triplet $(\widehat{m}_1, \widehat{m}_2, \widehat{h})$ proposed in simulation (where the two first are given in (13) and (25) respectively and the third is the cross-validation bandwidth) one could propose an adaptive estimator of a choosing the triplet, minimizing an estimator of the risk. This is an interesting mathematical question which requires more attention and is beyond the primary purpose of this work.

Finally, our analysis sheds light on the importance to further investigate the conditional intensity function f, dependent on the invariant density π. A future perspective would be to propose a kernel estimator for the invariant density π and to deeply study its behaviour and its asymptotic properties, following the same approach as in [32] and [3]. A projection method is instead considered in [25] in order to estimate the invariant density associated to a piecewise deterministic Markov process (PDMP) process. As a consequence, it will be possible to discuss the properties of the related estimator of f.

To conclude, the innovative procedure that we have presented, could be used to investigate real life data set. For example, some neuronal data as explained in [17] should be interpreted through this model as both X and the jumps times of N are observed. This is a work in progress.

8 Proofs

8.1 Proof of volatility estimation

This section is devoted to the proof of the results stated in Section 3. We start proving Proposition 3.

8.1.1 Proof of Proposition 3

Proof. We want to show an upper bound for the empirical risk $\mathbb{E}[\|\widehat{\sigma}^2_m - \sigma^2\|^2_n]$. First of all we remark that, if t is a deterministic function, then it is $\mathbb{E}[\|t\|^2_n] = \|t\|^2_{\pi^X}$, where $\|t\|^2_{\pi^X} := \int_A t^2(x)\pi_X(dx)$ and $\pi^X(dx) = \pi(dx \times \mathbb{R}^M)$ is the projection on the coordinate X of π, which exists for Theorem 2.3 in [17] (proof in [18]).

By the definition of T_{t_i} we have that

$$\begin{aligned}\gamma_{n,M}(t) &:= \frac{1}{n}\sum_{i=0}^{n-1}\Big(t(X_{t_i}) - T_{t_i}\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X)\Big)^2 \\ &= \frac{1}{n}\sum_{i=0}^{n-1}\Big(t(X_{t_i}) - \sigma^2(X_{t_i}) - (\tilde{A}_{t_i} + B_{t_i} + E_{t_i}\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X))\Big)^2 \\ &= \left\|t-\sigma^2\right\|_n^2 + \frac{1}{n}\sum_{i=0}^{n-1}(\tilde{A}_{t_i} + B_{t_i} + E_{t_i}\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X))^2 \\ &\quad -\frac{2}{n}\sum_{i=0}^{n-1}\Big(\tilde{A}_{t_i} + B_{t_i} + E_{t_i}\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X)\Big)\big(t(X_{t_i}) - \sigma^2(X_{t_i})\big).\end{aligned}$$

As $\widehat{\sigma}_m^2$ minimizes $\gamma_{n,M}(t)$, for any $\sigma_m^2 \in \mathcal{S}_m$ it is $\gamma_{n,M}(\widehat{\sigma}_m^2) \leq \gamma_{n,M}(\sigma_m^2)$ and therefore

$$\left\|\widehat{\sigma}_m^2 - \sigma^2\right\|_n^2 \leq \left\|\sigma_m^2 - \sigma^2\right\|_n^2 + \frac{2}{n}\sum_{i=0}^{n-1}(\tilde{A}_{t_i} + B_{t_i} + E_{t_i}\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X))(\widehat{\sigma}_m^2(X_{t_i}) - \sigma_m^2(X_{t_i})).$$

Let us denote the contrast function

$$\nu_n(t) := \frac{1}{n}\sum_{i=0}^{n-1} B_{t_i} t(X_{t_i}). \tag{29}$$

It follows

$$\begin{aligned}\left\|\widehat{\sigma}_m^2 - \sigma^2\right\|_n^2 &\leq \left\|\sigma_m^2 - \sigma^2\right\|_n^2 + \frac{d}{n}\sum_{i=0}^{n-1}\Big(\tilde{A}_{t_i} + E_{t_i}\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X)\Big)^2 + \frac{1}{d}\left\|\widehat{\sigma}_m^2 - \sigma_m^2\right\|_n^2 \\ &\quad +2\nu_n(\sigma_m^2 - \widehat{\sigma}_m^2).\end{aligned}$$

The linearity of the function ν_n in t implies that

$$2\nu_n(\widehat{\sigma}_m^2 - \sigma_m^2) = 2\|\widehat{\sigma}_m^2 - \sigma_m^2\|_{\pi^X}\nu_n((\widehat{\sigma}_m^2 - \sigma_m^2)/\|\widehat{\sigma}_m^2 - \sigma_m^2\|_{\pi^X}) \leq 2\|\widehat{\sigma}_m^2 - \sigma_m^2\|_{\pi^X}\sup_{t\in\mathcal{B}_m}\nu_n(t),$$

then, using that when $d>0$, we have $2xy \leq \frac{x^2}{d} + dy^2$, we obtain the upper bound

$$2\nu_n(\widehat{\sigma}_m^2 - \sigma_m^2) \leq \frac{1}{d}\|\widehat{\sigma}_m^2 - \sigma_m^2\|_{\pi^X}^2 + d\sup_{t\in\mathcal{B}_m}\nu_n^2(t)$$

where $\mathcal{B}_m = \Big\{t \in S_m : \|t\|_{\pi^X}^2 \leq 1\Big\}$. Finally, using Cauchy-Schwarz's inequality leads to

$$\begin{aligned}\left\|\widehat{\sigma}_m^2 - \sigma^2\right\|_n^2 &\leq \left\|\sigma_m^2 - \sigma^2\right\|_n^2 + \frac{2d}{n}\sum_{i=0}^{n-1}\tilde{A}_{t_i}^2 + \frac{2d}{n}\sum_{i=0}^{n-1}E_{t_i}^2\varphi_{\Delta_{n,i}^{\beta}}^2(\Delta_i X) + \frac{1}{d}\left\|\widehat{\sigma}_m^2 - \sigma_m^2\right\|_n^2 \\ &\quad +d\sup_{\mathcal{B}_m}\nu_n^2(t) + \frac{1}{d}\left\|\widehat{\sigma}_m^2 - \sigma_m^2\right\|_{\pi^X}^2.\end{aligned} \tag{30}$$

Let us set

$$\Omega_n := \left\{\omega, \forall t \in \tilde{S}_n\backslash\{0\}, \left|\frac{\|t\|_n^2}{\|t\|_{\pi^X}^2} - 1\right| \leq \frac{1}{2}\right\}, \tag{31}$$

on which the norms $\|\cdot\|_{\pi^X}$ and $\|\cdot\|_n$ are equivalent. We now act differently to bound the risk on Ω_n and Ω_n^c.

Bound of the risk on Ω_n On Ω_n, it is

$$\left\|\widehat{\sigma}_m^2 - \sigma_m^2\right\|_{\pi^X}^2 \leq 2\left\|\widehat{\sigma}_m^2 - \sigma_m^2\right\|_n^2 \leq 4\left\|\widehat{\sigma}_m^2 - \sigma^2\right\|_n^2 + 4\left\|\sigma^2 - \sigma_m^2\right\|_n^2,$$

where in the last estimation we have used triangular inequality. In the same way we get

$$\left\|\widehat{\sigma}_m^2 - \sigma_m^2\right\|_n^2 \leq 2\left\|\widehat{\sigma}_m^2 - \sigma^2\right\|_n^2 + 2\left\|\sigma^2 - \sigma_m^2\right\|_n^2.$$

Replacing them in (30) we obtain

$$\begin{aligned}\left\|\widehat{\sigma}_m^2 - \sigma^2\right\|_n^2 \quad \leq \quad & \left\|\sigma_m^2 - \sigma^2\right\|_n^2 + \frac{2d}{n}\sum_{i=0}^{n-1}\tilde{A}_{t_i}^2 + \frac{2d}{n}\sum_{i=0}^{n-1}(E_{t_i}\varphi_{\Delta_{n,i}^\beta}(\Delta_i X))^2 + d\sup_{t\in\mathcal{B}_m}\nu_n^2(t) \\ & + \frac{6}{d}\left\|\widehat{\sigma}_m^2 - \sigma^2\right\|_n^2 + \frac{6}{d}\left\|\sigma^2 - \sigma_m^2\right\|_n^2.\end{aligned}$$

We need d to be more than 6, we take $d = 7$ obtaining

$$\left\|\widehat{\sigma}_m^2 - \sigma^2\right\|_n^2 \leq 13\left\|\sigma_m^2 - \sigma^2\right\|_n^2 + \frac{98}{n}\sum_{i=0}^{n-1}\tilde{A}_{t_i}^2 + \frac{98}{n}\sum_{i=0}^{n-1}(E_{t_i}\varphi_{\Delta_{n,i}^\beta}(\Delta_i X))^2 + 49\sup_{t\in\mathcal{B}_m}\nu_n^2(t). \quad (32)$$

We denote as $(\psi_l)_l$ an orthonormal basis of S_m for the $L^2_{\pi^X}$ norm (thus $\int_{\mathbb{R}}\psi_l^2(x)\pi^X(x)dx = 1$). Each $t \in \mathcal{B}_m$ can be written

$$t = \sum_{l=1}^{D_m}\alpha_l\psi_l, \qquad \text{with } \sum_{l=1}^{D_m}\alpha_l^2 \leq 1.$$

Then

$$\sup_{t\in\mathcal{B}_m}\nu_n^2(t) = \sup_{\sum_{l=1}^{D_m}\alpha_l^2\leq 1}\nu_n^2\left(\sum_{l=1}^{D_m}\alpha_l\psi_l\right) \leq \sup_{\sum_{l=1}^{D_m}\alpha_l^2\leq 1}\left(\sum_{l=1}^{D_m}\alpha_l^2\right)\left(\sum_{l=1}^{D_m}\nu_n^2(\psi_l)\right) = \sum_{l=1}^{D_m}\nu_n^2(\psi_l). \quad (33)$$

To study the risk we need to evaluate the expected value. From (32), (33) and using the first and the third points of Proposition 2, we get

$$\mathbb{E}\left[\left\|\widehat{\sigma}_m^2 - \sigma^2\right\|_n^2 \mathbb{1}_{\Omega_n}\right] \leq 13\mathbb{E}\left[\left\|\sigma_m^2 - \sigma^2\right\|_n^2\right] + c\Delta_n^{1-\tilde{\varepsilon}} + c\Delta_n^{4\beta-1} + 49\sum_{l=1}^{D_m}\mathbb{E}[\nu_n^2(\psi_l)]. \quad (34)$$

By the definition (29) of ν_n it is

$$\nu_n(\psi_l) = \frac{1}{n}\sum_{i=0}^{n-1}B_{t_i}\psi_l(X_{t_i}).$$

As B_{t_i} is conditionally centered, using the second point of Proposition 2, it is

$$\sum_{l=1}^{D_m}\mathbb{E}[\nu_n^2(\psi_l)] \leq \frac{c}{n^2}\sum_{i=0}^{n-1}\sum_{l=1}^{D_m}\mathbb{E}[\psi_l^2(X_{t_i})\mathbb{E}[B_{t_i}^2|\mathcal{F}_{t_i}]] \leq \frac{c}{n^2}\sum_{i=0}^{n-1}\sum_{l=1}^{D_m}\sigma_1^4\mathbb{E}[\psi_l^2(X_{t_i})] \leq \frac{c\sigma_1^4 D_m}{n}.$$

Replacing the inequality here above in (34) it yields

$$\mathbb{E}\left[\left\|\widehat{\sigma}_m^2 - \sigma^2\right\|_n^2 \mathbb{1}_{\Omega_n}\right] \leq 13\mathbb{E}\left[\left\|\sigma_m^2 - \sigma^2\right\|_n^2\right] + c\Delta_n^{4\beta-1} + \frac{c\sigma_1^4 D_m}{n}.$$

As on Ω_n the empiric norm and the norm on π^X are equivalent and the reasoning here above applies for no matter what $\sigma^2_m \in \mathcal{S}_m$, it clearly follows

$$\mathbb{E}[\left\|\widehat{\sigma}^2_m - \sigma^2\right\|^2_n \mathbb{1}_{\Omega_n}] \leq 13 \inf_{t\in\mathcal{S}_m} \left\|t - \sigma^2\right\|^2_{\pi^X} + c\Delta_n^{4\beta-1} + \frac{c\sigma_1^4 D_m}{n}. \tag{35}$$

Bound of the risk on Ω_n^c
The complementary space Ω_n^c of Ω_n given in Equation (31) is defined as:

$$\Omega_n^c = \left\{\omega \in \Omega,\ \ \exists t^* \in \tilde{S}_n \backslash \{0\}, \left|\frac{\|t^*\|^2_n}{\|t^*\|^2_{\pi^X}} - 1\right| > 1/2\right\}.$$

Let us set $e = (e_{t_0}, \ldots, e_{t_{n-1}})$, where $e_{t_i} := T_{t_i}\varphi_{\Delta^\beta_{n,i}}(\Delta_i X) - \sigma^2(X_{t_i}) = \tilde{A}_{t_i} + B_{t_i} + E_{t_i}\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)$. Moreover

$$\Pi_m T\varphi = \Pi_m(T_{t_0}\varphi_{\Delta^\beta_{n,0}}(\Delta_0 X), \ldots, T_{t_{n-1}}\varphi_{\Delta^\beta_{n,n-1}}(\Delta_{n-1}X)) = (\widehat{\sigma}^2_m(X_{t_0}), \ldots, \widehat{\sigma}^2_m(X_{t_{n-1}})),$$

where Π_m is the Euclidean orthogonal projection over S_m. Then, according to the projection definition,

$$\begin{aligned}\left\|\widehat{\sigma}^2_m - \sigma^2\right\|^2_n &= \left\|\Pi_m T\varphi - \sigma^2\right\|^2_n = \left\|\Pi_m T\varphi - \Pi_m\sigma^2\right\|^2_n + \left\|\Pi_m\sigma^2 - \sigma^2\right\|^2_n \\ &\leq \left\|T\varphi - \sigma^2\right\|^2_n + \left\|\sigma^2\right\|^2_n = \|e\|^2_n + \left\|\sigma^2\right\|^2_n.\end{aligned}$$

Therefore, from Cauchy -Schwarz inequality and the boundedness of $\sigma^2(x)$,

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{\sigma}^2_m - \sigma^2\right\|^2_n \mathbb{1}_{\Omega_n^c}\right] &\leq \mathbb{E}\left[\|e\|^2_n \mathbb{1}_{\Omega_n^c}\right] + \mathbb{E}\left[\left\|\sigma^2\right\|^2_n \mathbb{1}_{\Omega_n^c}\right] = \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[e^2_{t_i}\mathbb{1}_{\Omega_n^c}] + \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[\sigma^4(X_{t_i})\mathbb{1}_{\Omega_n^c}] \\ &\leq \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[e^4_{t_i}]^{\frac{1}{2}}\mathbb{P}(\Omega_n^c)^{\frac{1}{2}} + \sigma_1^4\mathbb{P}(\Omega_n^c).\end{aligned}$$

From Lemma 6.4 in [17], if $\frac{n\Delta_n}{(\log n)^2} \to \infty$ and $D_n^2 \leq \frac{n\Delta_n}{(\log n)^2}$, then

$$\mathbb{P}(\Omega_n^c) \leq \frac{c_0}{n^4}. \tag{36}$$

In the hypothesis of our proposition we have requested that $n^\varepsilon \log n = o(\sqrt{n\Delta_n})$. As for n going to ∞ we have $\frac{(\log n)^2}{n\Delta_n} < \frac{n^\varepsilon \log n}{\sqrt{n\Delta_n}} \to 0$, the first condition in Lemma 6.4 in [17] hold true. Regarding the bound on D_n, we have assumed $D_n \leq \frac{\sqrt{n\Delta_n}}{\log n\, n^\varepsilon} \leq \frac{\sqrt{n\Delta_n}}{\log n}$ and so we can apply the here above mentioned lemma, which yields (36).
We are left to evaluate $\mathbb{E}[e^4_{t_i}]$. From Proposition 2 it follows

$$\mathbb{E}\left[e^4_{t_i}\right] \leq \mathbb{E}\left[\tilde{A}^4_{t_i} + B^4_{t_i} + E^4_{t_i}\varphi^4_{\Delta^\beta_{n,i}}(\Delta_i X)\right] \leq c\Delta_n^{1-\tilde{\varepsilon}} + c + c\Delta_n^{8\beta-3} \leq c\Delta_n^{0\wedge 8\beta-3}.$$

Putting the pieces together it yields

$$\mathbb{E}\left[\left\|\widehat{\sigma}^2_m - \sigma^2\right\|^2_n \mathbb{1}_{\Omega_n^c}\right] \leq \frac{c\Delta_n^{0\wedge 4\beta-\frac{3}{2}}}{n^2} + \frac{c}{n^4} \leq \frac{c\Delta_n^{0\wedge 4\beta-\frac{3}{2}}}{n^2}. \tag{37}$$

From (35) and (37) it follows

$$\mathbb{E}\left[\left\|\widehat{\sigma}^2_m - \sigma^2\right\|^2_n\right] \leq 13 \inf_{t\in\mathcal{S}_m}\left\|t - \sigma^2\right\|^2_{\pi^X} + \frac{C_1\sigma_1^4 D_m}{n} + C_2\Delta_n^{4\beta-1} + \frac{C_3\Delta_n^{0\wedge 4\beta-\frac{3}{2}}}{n^2}.$$

□

8.1.2 Proof of Theorem 1

Proof. We analyse the quantity $\mathbb{E}[\left\|\widehat{\sigma}^2_{\widehat{m}} - \sigma^2\right\|_n^2]$, acting again in different way depending on whether or not we are on Ω_n. On Ω_n^c the proof can be led as before, getting

$$\mathbb{E}\left[\left\|\widehat{\sigma}^2_{\widehat{m}} - \sigma^2\right\|_n^2 \mathbb{1}_{\Omega_n^c}\right] \leq \frac{c\Delta_n^{0\wedge 4\beta-\frac{3}{2}}}{n^2}. \tag{38}$$

Now we investigate what happens on Ω_n. By the definition of $\widehat{m}$ it is

$$\gamma_{n,M}(\widehat{\sigma}_{\widehat{m}}) + \text{pen}(\widehat{m}) \leq \gamma_{n,M}(\widehat{\sigma}_m) + \text{pen}(m) \leq \gamma_{n,M}(\sigma_m) + \text{pen}(m)$$

and so, acting as before (32), we get

$$\begin{aligned}\mathbb{E}\left[\left\|\sigma^2_{\widehat{m}} - \sigma^2\right\|_n^2 \mathbb{1}_{\Omega_n}\right] \leq 13\mathbb{E}[\left\|\sigma^2_m - \sigma^2\right\|_n^2] + \frac{98}{n}\sum_{i=0}^{n-1}\mathbb{E}[\tilde{A}^2_{t_i}] + \frac{98}{n}\sum_{i=0}^{n-1}\mathbb{E}[(E_{t_i}\varphi_{\Delta_{n,i}^\beta}(\Delta_i X))^2] \\ +49\,\mathbb{E}\left[\sup_{t\in\mathcal{B}_{m,\widehat{m}}}\nu_n^2(t)\right] + 7\text{pen}(m) - 7\mathbb{E}[\text{pen}(\widehat{m})],\end{aligned} \tag{39}$$

where ν_n has been defined in (29) and

$$\mathcal{B}_{m,m'} := \{h \in S_m + S_{m'} : \|h\|_{\pi^X} \leq 1\}.$$

We want to control the term $\mathbb{E}[\sup_{t\in\mathcal{B}_{m,\widehat{m}}}(\nu_n(t))^2]$ and, to do that, we introduce the function $p(m,m')$ which is such that

$$p(m,m') = \frac{1}{49}(\text{pen}(m) + \text{pen}(m')). \tag{40}$$

It is

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

In order to bound the second term in the right hand side here above we want to use Lemma 7 in [30]. We can remark that, for any $p \geq 2$, $\mathbb{E}[|B_{t_i}|^p] \leq \frac{c}{\Delta_n^p}\mathbb{E}[Z^{2p}_{t_i}] + c\sigma_1^{2p}$. According to Proposition 4.2 in Barlow and Yor [8] (B.D.G. inequality with optimal constants) there exists a constant c such that, for any $p > 0$,

$$\mathbb{E}\left[Z^{2p}_{t_i}\right] \leq c^{2p}(2p)^p\Delta_n^p\sigma_1^{2p}.$$

It follows

$$\mathbb{E}[|B_{t_i}|^p] \leq (c^{2p}(2p)^p\sigma_1^{2p} + c\sigma_1^{2p}) \leq 2c^{2p}(2p)^p\sigma_1^{2p}.$$

By Lemma 7 in [30] there exists a constant k such that, for any $m, m' \in \mathcal{M}_n$,

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,m'}}\nu_n^2(t) - kc\sigma_1 p(m,m')\right)_+\right] \leq c\frac{e^{-(D_m+D_{m'})}}{n}. \tag{41}$$

We have said, in the definition of the penalization function pen_σ given in Subsection 3.2, that the constant k_1 has to be calibrated. In particular, we need it to be such that $\frac{k_1}{49} \geq kc\sigma_1$, where σ_1 is the upper bound for the volatility provided in the second point of Assumption 1 and k and c are as in Lemma 7 of [30]. We underline that Lemma 7 in [30] has been proved for a noisy diffusion. However, the same reasoning applies for a jump diffusion (see the proof of Theorem 13 in [31]) and for our framework as well, as it is based on a projection argument and

on algebraic computations which still hold true.
From (41) and the fourth point of Assumption 4 we get

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,m'}}\nu_n^2(t)-p(m,m')\right)_+\right]\leq\frac{c}{n}\sum_{m'\in\mathcal{M}_n}e^{-(D_m+D_{m'})}\leq\frac{c}{n}.$$

It provides us, using also (37) and Proposition 2,

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{\sigma}^2_{\widehat{m}}-\sigma^2\right\|_n^2\right] &\leq 13\mathbb{E}\left[\left\|\sigma^2_m-\sigma^2\right\|_n^2\right]+c\Delta_n^{4\beta-1}+\frac{c}{n^4}+8\mathrm{pen}(m)+\frac{c\Delta_n^{0\wedge(4\beta-\frac{3}{2})}}{n^2}+\frac{c}{n}\\ &\leq C_1\inf_{m\in\mathcal{M}_n}\left\{\inf_{t\in\mathcal{S}_m}\|t-\sigma^2\|_\pi^2+\mathrm{pen}(m)\right\}+C_2\Delta_n^{4\beta-1}+\frac{C_3\Delta_n^{4\beta-\frac{3}{2}}}{n^2}+\frac{C_4}{n}.\end{aligned}$$

□

8.2 Proof of results on estimation of g

In this section we prove the results stated in Section 4.

8.2.1 Proof of Proposition 5

Proof. The proof follows the same scheme than the proof of Proposition 3. We want to upper bound the empirical risk $\mathbb{E}[\|\widehat{g}_m-g\|_n^2]$. By the definition of T_{t_i} we have that

$$\gamma_{n,M}(t):=\frac{1}{n}\sum_{i=0}^{n-1}(t(X_{t_i})-T_{t_i})^2=\frac{1}{n}\sum_{i=0}^{n-1}(t(X_{t_i})-g(X_{t_i})-(A_{t_i}+B_{t_i}+C_{t_i}+E_{t_i}))^2$$

$$\begin{aligned}\gamma_{n,M}(t) &= \|t-g\|_n^2+\frac{1}{n}\sum_{i=0}^{n-1}(A_{t_i}+B_{t_i}+C_{t_i}+E_{t_i})^2\\ &\quad-\frac{2}{n}\sum_{i=0}^{n-1}(A_{t_i}+B_{t_i}+C_{t_i}+E_{t_i})(t(X_{t_i})-g(X_{t_i})).\end{aligned}$$

As $\widehat{g}_m$ minimizes $\gamma_{n,M}(t)$, for any $g_m\in\mathcal{S}_m$ it is $\gamma_{n,M}(\widehat{g}_m)\leq\gamma_{n,M}(g_m)$ and therefore

$$\|\widehat{g}_m-g\|_n^2\leq\|g_m-g\|_n^2+\frac{2}{n}\sum_{i=0}^{n-1}(A_{t_i}+B_{t_i}+C_{t_i}+E_{t_i})(\widehat{g}_m(X_{t_i})-g_m(X_{t_i})).$$

Using Cauchy-Schwarz inequality and the fact that, for $d>0$, $2xy\leq\frac{x^2}{d}+dy^2$, we get

$$\begin{aligned}\|\widehat{g}_m-g\|_n^2 &\leq \|g_m-g\|_n^2+\frac{2d}{n}\sum_{i=0}^{n-1}A_{t_i}^2+\frac{1}{d}\|\widehat{g}_m-g_m\|_n^2+2d\sup_{\mathcal{B}_m}\nu_{n,1}^2(t)\\ &\quad+\frac{1}{d}\|\widehat{g}_m-g_m\|_{\pi^X}^2+2d\sup_{\mathcal{B}_m}\nu_{n,2}^2(t),\end{aligned}\tag{42}$$

where $\mathcal{B}_m=\left\{t\in S_m:\|t\|_{\pi^X}^2\leq 1\right\}$ and

$$\nu_{n,1}(t):=\frac{1}{n}\sum_{i=0}^{n-1}(B_{t_i}+E_{t_i})t(X_{t_i}),\quad \nu_{n,2}(t):=\frac{1}{n}\sum_{i=0}^{n-1}C_{t_i}t(X_{t_i}).\tag{43}$$

We still denote Ω_n the space on which the norms $\|\cdot\|_{\pi^X}$ and $\|\cdot\|_n$ are equivalent given by Equation (31). We now act differently to bound the risk on Ω_n and Ω_n^c.

Bound of the risk on Ω_n

On Ω_n, it is

$$\|\widehat{g}_m - g_m\|_{\pi^X}^2 \leq 2\|\widehat{g}_m - g_m\|_n^2 \leq 4\|\widehat{g}_m - g\|_n^2 + 4\|g - g_m\|_n^2,$$

where in the last estimation we have used triangular inequality. Replacing it in (42) we get

$$\begin{aligned}\|\widehat{g}_m - g\|_n^2 \quad \leq \quad & \|g_m - g\|_n^2 + \frac{2d}{n}\sum_{i=0}^{n-1} A_{t_i}^2 + 2d\sup_{\mathcal{B}_m}\nu_{n,1}^2(t) + 2d\sup_{\mathcal{B}_m}\nu_{n,2}^2(t) \\ & + \frac{6}{d}\|\widehat{g}_m - g\|_n^2 + \frac{6}{d}\|g - g_m\|_n^2.\end{aligned}$$

We need d to be more than 6, we take $d = 7$ obtaining

$$\|\widehat{g}_m - g\|_n^2 \leq 13\|g_m - g\|_n^2 + \frac{98}{n}\sum_{i=0}^{n-1} A_{t_i}^2 + 98\sup_{t\in\mathcal{B}_m}\nu_{n,1}^2(t) + 98\sup_{t\in\mathcal{B}_m}\nu_{n,2}^2(t). \tag{44}$$

We now need to introduce a different orthonormal basis of S_m, compared to the one we proposed in Section 8.1, for the estimation of volatility. The reason why it is necessary to change it is that in $\mathbb{E}[E_{t_i}^2|\mathcal{F}_{t_i}]$ we now get a term that depends on λ_{t_i}, which is an extra difficulty compared with the reasoning we applied below (34). Hence, we consider $(\tilde{\psi}_k)_k$ an orthonormal basis of S_m for which

$$\mathbb{E}\left[\tilde{\psi}_k^2(X_{t_i}, l)|\lambda_{t_i} = l\right] = 1, \tag{45}$$

where $\lambda_{t_i} = (\lambda_{t_i}^{(1)}, \ldots, \lambda_{t_i}^{(M)})$. It is possible to build such a basis starting from the one we have introduced in the proof of Proposition 3, through Gram-Schmidt process, for the scalar product in $L^2(\pi(dx|\lambda_{t_i} = l))$, for $l \in \mathbb{R}^m$. Each $t \in \mathcal{B}_m$ can be written

$$t = \sum_{l=1}^{D_m}\alpha_l\tilde{\psi}_l, \qquad \text{with } \sum_{l=1}^{D_m}\alpha_l^2(\lambda_{t_i}) \leq 1.$$

We underline that this time, unlike it was in the estimation of the volatility, the coefficients α_l depend on λ_{t_i}. We omit it in the sequel to lighten the notation. Then, for $j = 1$ and $j = 2$,

$$\sup_{t\in\mathcal{B}_m}\nu_{n,j}^2(t) = \sup_{\sum_{l=1}^{D_m}\alpha_l^2\leq 1}\nu_{n,j}^2\left(\sum_{l=1}^{D_m}\alpha_l\tilde{\psi}_l\right) \leq \sup_{\sum_{l=1}^{D_m}\alpha_l^2\leq 1}\left(\sum_{l=1}^{D_m}\alpha_l^2\right)\left(\sum_{l=1}^{D_m}\nu_{n,j}^2(\tilde{\psi}_l)\right) = \sum_{l=1}^{D_m}\nu_{n,j}^2(\tilde{\psi}_l). \tag{46}$$

To study the risk we need to evaluate the expected value. From (44), (46) and using the first point of Proposition 4, we get

$$\mathbb{E}\left[\|\widehat{g}_m - g\|_n^2\,\mathbb{1}_{\Omega_n}\right] \leq 13\mathbb{E}\left[\|g_m - g\|_n^2\right] + c\Delta_n^{1-\tilde{\varepsilon}} + 98\sum_{l=1}^{D_m}\mathbb{E}[\nu_{n,1}^2(\tilde{\psi}_l)] + 98\sum_{l=1}^{D_m}\mathbb{E}\left[\nu_{n,2}^2(\tilde{\psi}_l)\right]. \tag{47}$$

By the definition (43) of $\nu_{n,1}$ and the points 2 and 3 of Proposition 4, it is

$$\begin{aligned}\sum_{l=1}^{D_m}\mathbb{E}[\nu_{n,1}^2(\tilde{\psi}_l)] \quad &\leq \quad \frac{c}{n^2}\sum_{i=0}^{n-1}\sum_{l=1}^{D_m}\mathbb{E}\left[\tilde{\psi}_l^2(X_{t_i}, \lambda_{t_i})\mathbb{E}[B_{t_i}^2 + E_{t_i}^2|\mathcal{F}_{t_i}]\right] \\ &\leq \quad \frac{c}{n^2}\sum_{i=0}^{n-1}\sum_{l=1}^{D_m}\mathbb{E}\left[\tilde{\psi}_l^2(X_{t_i}, \lambda_{t_i})(c\sigma_1^4 + \frac{ca_1^4}{\Delta_{n,i}}\sum_{j=1}^{M}|\lambda_{t_i}^{(j)}|)\right]\end{aligned}$$

We observe that the first term in the right hand side here above is

$$\frac{c\sigma_1^4}{n^2}\sum_{i=0}^{n-1}\sum_{l=1}^{D_m}\mathbb{E}[\tilde{\psi}_l^2(X_{t_i},\lambda_{t_i})] \leq \frac{cD_m}{n},$$

where we moved to the conditional expectation with respect to $\lambda_{t_i} = (\lambda_{t_i}^{(1)},\ldots,\lambda_{t_i}^{(M)})$. Regarding the second term, we remark it is

$$\begin{aligned}\mathbb{E}\left[\tilde{\psi}_l^2(X_{t_i},\lambda_{t_i})\frac{ca_1^4}{\Delta_{n,i}}\sum_{j=1}^{M}|\lambda_{t_i}^{(j)}|\right] &= \mathbb{E}[\mathbb{E}[\tilde{\psi}_l^2(X_{t_i},\lambda_{t_i})|\lambda_{t_i}]\frac{ca_1^4}{\Delta_{n,i}}\sum_{j=1}^{M}|\lambda_{t_i}^{(j)}|] \\ &= \frac{ca_1^4}{\Delta_{n,i}}\sum_{j=1}^{M}\mathbb{E}[|\lambda_{t_i}^{(j)}|] \leq \frac{ca_1^4}{\Delta_{n,i}},\end{aligned}$$

where in the last inequality we have used the boundedness of the moments of λ and (45). It follows

$$\frac{c}{n^2}\sum_{i=0}^{n-1}\sum_{l=1}^{D_m}\mathbb{E}[\tilde{\psi}_l^2(X_{t_i},\lambda_{t_i})\frac{ca_1^4}{\Delta_{n,i}}\sum_{j=1}^{M}|\lambda_{t_i}^{(j)}|] \leq \frac{cD_m a_1^4}{n\Delta_{n,i}}.$$

Hence,

$$\sum_{l=1}^{D_m}\mathbb{E}\left[\nu_{n,1}^2(\tilde{\psi}_l)\right] \leq \frac{c(\sigma_1^4+a_1^4)D_m}{n\Delta_{n,i}}. \tag{48}$$

In order to evaluate $\mathbb{E}[\nu_{n,2}^2(\tilde{\psi}_l)]$, the following lemma will be useful:

Lemma 3. *Suppose that A1-A3 hold true. Then,*

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=0}^{n-1}C_{t_i}\tilde{\psi}_l(X_{t_i},\lambda_{t_i})\right) \leq \frac{c}{n\Delta_n}.$$

The proof of Lemma 3 is in the appendix. Lemma 3 yields

$$\sum_{l=1}^{D_m}\mathbb{E}\left[\nu_{n,2}^2(\tilde{\psi}_l)\right] \leq \frac{cD_m}{n\Delta_n}.$$

Replacing the inequality here above and (48) in (47) we get, using also that $\Delta_{n,i} \geq c\Delta_{min}$ and the fact that there exist c_1 and c_2 for which $c_1 \leq \frac{\Delta_n}{\Delta_{min}} \leq c_2$,

$$\mathbb{E}\left[\|\widehat{g}_m - g\|_n^2\,\mathbb{1}_{\Omega_n}\right] \leq 13\mathbb{E}\left[\|g_m - g\|_n^2\right] + c\Delta_n^{1-\tilde{\varepsilon}} + \frac{c(\sigma_1^4+a_1^4+1)D_m}{n\Delta_n}.$$

As the choice $g_m \in \mathcal{S}_m$ is arbitrary, we obtain

$$\mathbb{E}\left[\|\widehat{g}_m - g\|_n^2\,\mathbb{1}_{\Omega_n}\right] \leq 13\inf_{t\in\mathcal{S}_m}\|t-g\|_{\pi^X}^2 + c\Delta_n^{1-\tilde{\varepsilon}} + \frac{c(\sigma_1^4+a_1^4+1)D_m}{n\Delta_n}. \tag{49}$$

Bound of the risk on Ω_n^c

Let us set $e = (e_{t_0},\ldots,e_{t_{n-1}})$, where $e_{t_i} := T_{t_i} - g(X_{t_i}) = A_{t_i} + B_{t_i} + C_{t_i} + E_{t_i}$. Moreover

$$\Pi_m T = \Pi_m(T_{t_0},\ldots,T_{t_{n-1}}) = (\widehat{g}_m(X_{t_0}),\ldots,\widehat{g}_m(X_{t_{n-1}})),$$

where Π_m is the Euclidean orthogonal projection over S_m. Then, according to the projection definition,

$$\begin{aligned}\|\widehat{g}_m - g\|_n^2 &= \|\Pi_m T - g\|_n^2 = \|\Pi_m T - \Pi_m g\|_n^2 + \|\Pi_m g - g\|_n^2 \\ &\leq \|T - g\|_n^2 + \|g\|_n^2 = \|e\|_n^2 + \|g\|_n^2 .\end{aligned}$$

Therefore, from Cauchy -Schwarz inequality,

$$\mathbb{E}[\|\widehat{g}_m - g\|_n^2 \mathbb{1}_{\Omega_n^c}] \leq \mathbb{E}[\|e\|_n^2 \mathbb{1}_{\Omega_n^c}] + \mathbb{E}[\|g\|_n^2 \mathbb{1}_{\Omega_n^c}] = \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[e_{t_i}^2 \mathbb{1}_{\Omega_n^c}] + \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[g(X_{t_i})^2 \mathbb{1}_{\Omega_n^c}] \leq$$

$$\frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[e_{t_i}^4]^{\frac{1}{2}} \mathbb{P}(\Omega_n^c)^{\frac{1}{2}} + \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[g(X_{t_i})^4]^{\frac{1}{2}} \mathbb{P}(\Omega_n^c)^{\frac{1}{2}}$$

Moreover, using the boundedness of both a and σ and the fact that $\mathbb{E}[|\lambda_{t_i}|^4] < \infty$, we obtain $\mathbb{E}[g(X_{t_i})^4] < \infty$. We are left to evaluate $\mathbb{E}[e_{t_i}^4]$. From Proposition 4 it follows

$$\mathbb{E}[e_{t_i}^4] \leq \mathbb{E}[A_{t_i}^4 + B_{t_i}^4 + C_{t_i}^4 + E_{t_i}^4] \leq c\Delta_n^{1-\tilde{\varepsilon}} + c + c + \frac{c}{\Delta_{n_i}^3} \leq \frac{c}{\Delta_n^3}.$$

Putting the pieces together it yields

$$\mathbb{E}[\|\widehat{g}_m - g\|_n^2 \mathbb{1}_{\Omega_n^c}] \leq \frac{c}{\Delta_n^{\frac{3}{2}}}\frac{1}{n^2} + \frac{c}{n^2} \leq \frac{c}{n^2\Delta_n^{\frac{3}{2}}}. \tag{50}$$

From (49) and (50) it follows

$$\mathbb{E}[\|\widehat{g}_m - g\|_n^2] \leq 13\mathbb{E}[\|g_m - g\|_n^2] + \frac{C_1(\sigma_1^4 + a_1^4 + 1)D_m}{n\Delta_n} + C_2\Delta_n^{1-\tilde{\varepsilon}} + \frac{C_3}{n^2\Delta_n^{\frac{3}{2}}}.$$

□

8.2.2 Proof of Theorem 2

Proof. We act again in different way depending on whether or not we are on Ω_n. On Ω_n^c the proof can be led as before, getting

$$\mathbb{E}\left[\|\widehat{g}_{\widehat{m}} - g\|_n^2 \mathbb{1}_{\Omega_n^c}\right] \leq \frac{c}{n^2\Delta_n^{\frac{3}{2}}}. \tag{51}$$

Now we investigate what happens on Ω_n. In particular, we analyse what happens on $\mathcal{O} \subset \Omega_n$, a set which will be defined later (see (56)). By the definition of $\widehat{m}$ we have

$$\gamma_{n,M}(\widehat{g}_{\widehat{m}}) + \text{pen}(\widehat{m}) \leq \gamma_{n,M}(\widehat{g}_m) + \text{pen}(m) \leq \gamma_{n,M}(g_m) + \text{pen}(m)$$

and so, following the from of Equation (44), we get

$$\begin{aligned}\mathbb{E}\left[\|\widehat{g}_{\widehat{m}} - g\|_n^2 \mathbb{1}_{\mathcal{O}}\right] &\leq 13\mathbb{E}[\|g_m - g\|_n^2] + \frac{98}{n}\sum_{i=0}^{n-1}\mathbb{E}[A_{t_i}^2] + 98\,\mathbb{E}\left[\sup_{t\in\mathcal{B}_{m,\widehat{m}}} \nu_n^2(t)\mathbb{1}_{\mathcal{O}}\right] \\ &\quad +7\text{pen}(m) - 7\mathbb{E}[\text{pen}(\widehat{m})],\end{aligned}$$

where

$$\nu_n(t) := \frac{1}{n}\sum_{i=0}^{n-1}(B_{t_i} + C_{t_i} + E_{t_i})t(X_{t_i}),$$

and

$$\mathcal{B}_{m,m'} := \{h \in S_m + S_{m'} : \|h\|_{\pi^X} \leq 1\}.$$

In order to control the term $\mathbb{E}[\sup_{t\in\mathcal{B}_{m,\widehat{m}}} \nu_n^2(t)\mathbb{1}_{\mathcal{O}}]$, we introduce the function $p(m,m')$:

$$p(m,m') \leq \frac{1}{98}(\text{pen}(m) + \text{pen}(m')).$$

It is

$$\mathbb{E}[\sup_{t\in\mathcal{B}_{m,\widehat{m}}} \nu_n^2(t)\mathbb{1}_{\mathcal{O}}] \leq \mathbb{E}[p(m,\widehat{m})] + \sum_{m'\in\mathcal{M}_n} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,m'}} \nu_n^2(t) - p(m,m')\right)_+ \mathbb{1}_{\mathcal{O}}\right].$$

Replacing it in (39) and using the first point of Proposition 4 we get

$$\begin{aligned}\mathbb{E}\left[\|\widehat{g}_{\widehat{m}} - g\|_n^2 \mathbb{1}_{\mathcal{O}}\right] &\leq 13\mathbb{E}\left[\|g_m - g\|_n^2\right] + c\Delta_n^{1-\tilde{\varepsilon}} + 98\mathbb{E}[p(m,\widehat{m})] + 7\text{pen}(m) \\ &\quad -7\mathbb{E}[\text{pen}(\widehat{m})] + 98\sum_{m'\in\mathcal{M}_n} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,\widehat{m}}} \nu_n^2(t) - p(m,m')\right)_+ \mathbb{1}_{\mathcal{O}}\right].\end{aligned}$$

We have introduced the function $p(m,m')$ with the purpose to use Talagrand inequality on the last term in the right hand side of the equation here above. We recall the following version of the Talagrand inequality, which has been stated in [31] and proved by Birgé and Massart (1998) [9] (corollary 2p.354) and Comte and Merlevède (2002) [14] (p222-223).

Lemma 4. *Let $T_1,\ldots,T_n$ be independent random variables with values in some Polish space $\mathcal{X}$ and $v_p : \mathcal{B}_{m,m'} \to \mathbb{R}$ such that*

$$v_p(r) := \frac{1}{p}\sum_{j=1}^{p}[r(T_j) - \mathbb{E}[r(T_j)]].$$

Then,

$$\mathbb{E}\left[\left(\sup_{r\in\mathcal{B}_{m,m'}} |v_p(r)|^2 - 2H^2\right)_+\right] \leq c\left(\frac{v}{p}e^{-c\frac{pH^2}{v}} + \frac{M^2}{p^2}e^{-c\frac{pH}{M}}\right), \tag{52}$$

with c a universal constant and where

$$\sup_{r\in\mathcal{B}_{m,m'}} \|r\|_\infty \leq M, \quad \mathbb{E}[\sup_{r\in\mathcal{B}_{m,m'}} |v_p(r)|] \leq H, \quad \sup_{r\in\mathcal{B}_{m,m'}} \frac{1}{p}\sum_{j=1}^{p}\text{Var}(r(T_j)) \leq v.$$

We observe that in Talagrand lemma here above the random variables $T_1,\ldots,T_n$ are supposed to be independent. Starting from our variables we can get independent variables through Berbee's coupling method. We recall it below, it is proved by Viennet in Proposition 5.1 of [33] while an analogous statement in continuous time can be found in [3].

Lemma 5. *Let $(M_t)_{t\geq 0}$ be a stationary and exponentially β mixing process observed at discrete times $0 = t_0 \leq t_1 \leq \ldots \leq t_n = T$. Let p_n and q_n be two integers such that $n = 2p_nq_n$. For any $j \in \{0,1\}$ and $1 \leq k \leq p_n$ we consider the random variables*

$$U_{k,j} := (M_{t_{(2(k-1)+j)q_n+1}},\ldots,M_{t_{(2k-1+j)q_n}}).$$

There exist random variables $M_{t_0}^*,\ldots,M_{t_n}^*$ such that

$$U_{k,j}^* := (M^*_{t_{(2(k-1)+j)q_n+1}},\ldots,M^*_{t_{(2k-1+j)q_n}})$$

satisfy the following properties.

- *For any $j \in \{0,1\}$, the random vectors $U^*_{1,j}, \dots, U^*_{p_n,j}$ are independent.*
- *For any $(j,k) \in \{0,1\} \times \{1,\dots,p_n\}$, $U_{k,j}$ and $U^*_{k,j}$ have the same distribution.*
- *For any $(j,k) \in \{0,1\} \times \{1,\dots,p_n\}$, $\mathbb{P}(U_{k,j} \neq U^*_{k,j}) \leq \beta_M(q_n\Delta_{min})$, where β_M is the β-mixing coefficient of the process (M_t).*

We want to apply Berbee's coupling lemma to the random vectors $(B_{t_i} + C_{t_i} + E_{t_i}, X_{t_i})$, that we write as function of $Z_t = (X_t, \lambda_t)$, which is stationary and exponentially β- mixing, as discussed in Section 2.3. We define the σ algebra

$$\tilde{\mathcal{F}}_{t_i} := \sigma(X_s, \lambda_s, s \in (t_i, t_{i+1}]), \tag{53}$$

completed with the null sets. Because of the exponentially β- mixing of (X_t, λ_t) we know it is

$$\beta(\tilde{\mathcal{F}}_{t_i}, \tilde{\mathcal{F}}_{t_j}) \leq ce^{-\gamma|t_j - t_i|}.$$

From (2) and the fact we have assumed $c_{i,j}$ to be inversible, it is possible to write both Z_{t_i} and J_{t_i} in function of X and λ and so they are measurable with respect to $\tilde{\mathcal{F}}_{t_i}$. By the definition of B_{t_i}, C_{t_i} and E_{t_i} it follows that also $B_{t_i} + C_{t_i} + E_{t_i}$ is measurable with respect to $\tilde{\mathcal{F}}_{t_i}$. We can therefore use Berbee's coupling lemma on $(B_{t_i} + C_{t_i} + E_{t_i}, X_{t_i})$.
For $t \in \mathcal{B}_{m,m'}$, according to Berbee's coupling lemma, we can construct

$$U^*_{k,j} := \frac{1}{q_n} \sum_{l=1}^{q_n} (B + C + E)^*_{t_{(2(k-1)+j)q_n+l}} t(X^*_{t_{(2(k-1+j)q_n+l}})$$

such that, for $j \in \{0,1\}$, the random variables $(U^*_{k,j})_{0 \leq k \leq p_n}$ are independent and have the same distribution as

$$U_{k,j} := \frac{1}{q_n} \sum_{l=1}^{q_n} (B + C + E)_{t_{(2(k-1)+j)q_n+l}} t(X_{t_{(2(k-1+j)q_n+l}}).$$

Let us set

$$\Omega^* := \left\{\omega, \forall j, \forall k, U_{k,j} = U^*_{k,j}\right\},$$

by Berbee's coupling lemma it is

$$\mathbb{P}(\Omega^{*,c}) \leq 2p_n\beta_Z(q_n\Delta_{min}) \leq c\frac{n}{q_n}e^{-\gamma q_n \Delta_{min}}.$$

We recall that p_n and q_n are two integers to be chosen such that $2p_nq_n = n$. It is enough to take $q_n := \lfloor \frac{5}{\gamma\Delta_{min}} \log n \rfloor$ in (54) to get

$$\mathbb{P}(\Omega^{*,c}) \leq \frac{c}{n^4 \log n}. \tag{54}$$

We want to apply Talagrand inequality on $v^*_n(t) := v^{0,*}_n(t) + v^{1,*}_n(t)$, where

$$v^{0,*}_n(t) = \frac{1}{p_n} \sum_{k=0}^{p_n} U^*_{k,0}, \qquad v^{1,*}_n(t) = \frac{1}{p_n} \sum_{k=0}^{p_n} U^*_{k,1}.$$

To do that, we first of all observe that, as a consequence of Proposition 4, it is $\mathbb{E}[U^*_{k,j}] = 0$ for any $j \in \{0,1\}$ and any $k \in \{0,\dots,p_n\}$. Now we want to compute the constants M, v and H as defined in Lemma 4. To compute M we introduce the following set

$$\Omega_B := \left\{\omega, \forall j, \forall k, \forall \varepsilon > 0 \, |U^*_{k,j}| \leq \tilde{c}n^{\varepsilon} D^{\frac{1}{2}}\right\}, \tag{55}$$

with $D := D_m + D_{m'}$. The following lemma is proven in the appendix.

Lemma 6. *Suppose that A1-A3 hold. Then there exists $c > 0$ such that*

$$\mathbb{P}(\Omega_B^c) \leq \frac{c}{n^4}.$$

We set

$$\mathcal{O} := \Omega_n \cap \Omega_B \cap \Omega^*. \tag{56}$$

On $\mathcal{O}$ the random variables $|U^*_{k,j}|$ are replaced by $|U^*_{k,j}| \wedge \tilde{c} n^{\varepsilon} D^{\frac{1}{2}}$. As the original variables U and the independent ones U^* are the same on $\mathcal{O}$ even after truncation, the truncated random variables are still independent and we can use Talagrand inequality on them. From (36), (54) and Lemma 6 it follows

$$\mathbb{P}(\mathcal{O}^c) \leq \frac{c}{n^4}.$$

We act on $\mathcal{O}^c$ as we did on Ω_n^c, getting

$$\mathbb{E}[\|\widehat{g}_{\widehat{m}} - g\|_n^2 \mathbb{1}_{\mathcal{O}^c}] \leq \frac{c}{n^2 \Delta_n^{\frac{3}{2}}}. \tag{57}$$

On the other side, on $\mathcal{O}$ we are really going to use Talagrand's inequality to control

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_n(t)^2 - p(m,m') \right)_+ \mathbb{1}_{\mathcal{O}} \right].$$

From the definition of Ω_B, we clearly obtain that $M := c n^{\varepsilon} D^{\frac{1}{2}}$. With the purpose of computing v we observe that for any $t \in \mathcal{B}_{m,m'}$, by stationarity, it is

$$\mathrm{Var}(U^*_{k,j}) = \frac{1}{q_n^2} \sum_{l=1}^{q_n} \mathbb{E}\left[t^2(X^*_{t_l}) C^{*,2}_{t_l} \mathbb{E}_l[B^{*,2}_{t_l} + E^{*,2}_{t_l}] \right].$$

By the second and the third points of Proposition 4 this variance is upper bounded

$$\begin{aligned} \mathrm{Var}(U^*_{k,j}) &\leq \frac{c}{q_n^2} \sum_{l=1}^{q_n} \mathbb{E}\left[t^2(X^*_{t_l}) C^{*,2}_{t_l} (\sigma_1^2 + \frac{a_1^4}{\Delta_n} \sum_{j=1}^{M} \lambda^{(j)}_{t_l}) \right] \\ &\leq \frac{c}{q_n^2} \sum_{l=1}^{q_n} \mathbb{E}\left[t^{2p}(X^*_{t_l}) \right]^{\frac{1}{p}} \mathbb{E}\left[C^{*,2q}_{t_l} \left(\sigma_1^2 + \frac{a_1^4}{\Delta_n} \sum_{j=1}^{M} \lambda^{(j)}_{t_i} \right)^q \right]^{\frac{1}{q}}, \end{aligned} \tag{58}$$

where we have used Holder inequality with q big and p next to 1. We can see $t^{2p}(X^*_{t_l})$ as

$$t^{2+(2p-2)}(X^*_{t_l}) = t^2(X^*_{t_l}) t^{(2p-2)}(X^*_{t_l}) \leq \|t\|_\infty^{2p-2}\, t^2(X^*_{t_l}).$$

As p has been chosen next to 1, $\|t\|_\infty^{\frac{2p-2}{p}} \leq \|t\|_\infty^{\delta} \leq c D^{\frac{\delta}{2}}$ for any δ arbitrarily small. Using also the boundedness of the moments of C and λ it follows that the right hand side of (58) is upper bounded by

$$\frac{c D^{\frac{\delta}{2}}}{q_n \Delta_n} =: v.$$

In order to compute H^2 we observe it is

$$\mathbb{E}\left[\left| \frac{1}{p_n} \sum_{j,k} U^*_{k,j} \right| \right] \leq \sqrt{\mathbb{E}[v_p^{2,*}(t)]}$$

To find an upper bound for the right hand side here above we act in a similar way to how we did before (46): we introduce the orthonormal basis $(\bar\psi_k)_k$ for which $\mathbb{E}[\bar\psi_k^2(X_{t_i},l)|\lambda_{t_i}=l]=1$, such that each $t\in\mathcal{B}_{m,m'}$ can be written as the following

$$t=\sum_{l=1}^{D}\bar\alpha_l\bar\psi_l,\qquad \text{with }\sum_{l=1}^{D}\bar\alpha_l^2(\lambda_{t_i})\le 1.$$

The coefficients $\bar\alpha_l$ depend on λ_{t_i}. We omit it in the sequel to lighten the notation. Similarly to (46), we have

$$\sup_{t\in\mathcal{B}_{m,m'}} v_p^{2,*}(t)=\sup_{\sum_{l=1}^D\bar\alpha_l^2\le 1} v_p^{2,*}\left(\sum_{l=1}^{D}\bar\alpha_l\bar\psi_l\right)\le \sup_{\sum_{l=1}^D\bar\alpha_l^2\le 1}\left(\sum_{l=1}^{D}\bar\alpha_l^2\right)\left(\sum_{l=1}^{D}v_p^{2,*}(\bar\psi_l)\right)=\sum_{l=1}^{D}v_p^{2,*}(\bar\psi_l).$$

Acting exactly as we did in order to get (48) and Lemma 3 on $v_{n,1}^2$ and $v_{n,2}^2$ we obtain

$$\sqrt{\mathbb{E}[v_p^{2,*}(t)]}\le\sqrt{\frac{D}{n\Delta_n}}=:H,$$

We now use Talagrand inequality as in Lemma 4. It follows

$$\begin{aligned}\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,\widehat m}}\nu_n^*(t)^2-2H^2\right)_+\mathbb{1}_{\mathcal{O}}\right] &\le \frac{D^{\frac\delta2}}{p_nq_n\Delta_n}\exp\left(-c\frac{Dp_nq_n\Delta_n}{n\Delta_nD^{\frac\delta2}}\right)+\frac{cn^{2\varepsilon}D}{p_n^2}\exp\left(\frac{-cp_nD^{\frac12}}{\sqrt{n\Delta_n}n^\varepsilon D^{\frac12}}\right)\\ &= \frac{cD^{\frac\delta2}}{n\Delta_n}\exp(-cD^{1-\frac\delta2})+\frac{cn^{2\varepsilon}D}{p_n^2}\exp\left(-\frac{c\sqrt n}{q_n\sqrt{\Delta_n}n^\varepsilon}\right).\end{aligned}$$

We recall that $q_n=c\frac{\log n}{\Delta_{min}}$. We observe that, as Δ_{min} and Δ_n differs only for a constant, $\frac{c\sqrt n}{\sqrt{\Delta_n}q_nn^\varepsilon}=\frac{c\sqrt{n\Delta_n}}{\log n n^\varepsilon}$. Moreover, it goes to ∞ for n going to infinity as we have assumed that $(\log n)n^\varepsilon=o(\sqrt{n\Delta_n})$. Therefore, the second term here above is negligible compared to the first one. It follows, using also the definition of $p(m,\widehat m)$, the fact that for $D>1$ it is $D^{\frac\delta2}e^{-c'D^{1-\frac\delta2}}<ce^{-c'D^{1-\frac\delta2}}$ and the fourth point of Assumption 4,

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,\widehat m}}\nu_n^*(t)^2-p(m,m')\right)_+\right]\le\frac{c}{n\Delta_n}\sum_{m'\in\mathcal{M}_n}D^{\frac\delta2}e^{-c'D^{1-\frac\delta2}}\le\frac{c\,\Sigma(c')}{n\Delta_n}.$$

Replacing it in the equivalent of (52), considering that we are now on $\mathcal{O}$, it follows

$$\begin{aligned}\mathbb{E}\left[\|\widehat g_{\widehat m}-g\|_n^2\,\mathbb{1}_{\mathcal{O}}\right] &\le 13\mathbb{E}[\|g_m-g\|_n^2]+c\Delta_n^{1-\tilde\varepsilon}+98\mathbb{E}[p(m,\widehat m)]\\ &\quad+7\mathrm{pen}(m)-7\mathbb{E}[\mathrm{pen}(\widehat m)]+\frac{c}{n\Delta_n}.\end{aligned}$$

It provides us, using also (57),

$$\begin{aligned}\mathbb{E}[\|\widehat g_{\widehat m}-g\|_n^2] &\le 13\mathbb{E}[\|\widehat g_m-g\|_n^2]+c\Delta_n^{1-\tilde\varepsilon}+\frac{c}{n^2\Delta_n^{\frac32}}+c\mathrm{pen}(m)+\frac{c}{n\Delta_n}\\ &\le c_1\inf_{m\in\mathcal{M}_n}\left\{\inf_{t\in\mathcal{S}_m}\|t-g\|_{\pi^X}^2+\mathrm{pen}(m)\right\}+C_2\Delta_n^{1-\tilde\varepsilon}+\frac{C_3}{n^2\Delta_n^{\frac32}}+\frac{C_4}{n\Delta_n}.\end{aligned}$$

□

A Appendix

For the following proofs, the lemma stated and proved below is a very helpful tool. It provides the size of the increments of both X and λ.

Lemma 7. *Suppose that A1-A3 hold. Then, there exist c_1 and c_2 positive constants such that, for all $t > s$, $|t-s| < 1$ the following hold true*

1. *For all $p \geq 2$, $\mathbb{E}[|X_t - X_s|^p] \leq c_1|t-s|$.*
2. *For all $p \geq 2$ and for any $j \in \{1, \dots, M\}$, $\mathbb{E}[|\lambda_t^{(j)} - \lambda_s^{(j)}|^p] \leq c_2|t-s|$.*
3. *$\mathbb{E}[|\lambda_t - \lambda_s||\mathcal{F}_s] \leq c_3|t-s|(1+|\lambda_s|)$, where $\lambda = (\lambda^{(1)}, \dots, \lambda^{(M)})$ and $|\cdot|$ stands for the euclidean norm.*
4. *For any $j \in \{1, \dots, M\}$, $\sup_{h\in[0,1]} \mathbb{E}[|\lambda_{s+h}^{(j)}||\mathcal{F}_s] \leq |\lambda_s^{(j)}| + c|h|(1+|\lambda_s^{(j)}|)$.*

Proof. We start proving the first point. From the dynamic (2) of the process X we have

$$|X_t - X_s|^p \leq c\left|\int_s^t b(X_u)du\right|^p + c\left|\int_s^t \sigma(X_u)dW_u\right|^p + c\left|\int_s^t a(X_{u^-})\sum_{j=1}^M dN_u^{(j)}\right|^p = I_1 + I_2 + I_3.$$

From Jensen inequality, the polynomial growth of b and the fact that X has bounded moments it follows

$$\mathbb{E}[I_1] \leq c|t-s|^{p-1}\int_s^t \mathbb{E}[|b(X_u)|^p]du \leq c|t-s|^p. \tag{59}$$

Using Burkholder-Davis-Gundy inequality, Jensen inequality and the boundedness of σ it is

$$\mathbb{E}[I_2] \leq c\mathbb{E}\left[\left(\int_s^t \sigma^2(X_u)du\right)^{\frac{p}{2}}\right] \leq c|t-s|^{\frac{p}{2}-1}\int_s^t \mathbb{E}[|\sigma(X_u)|^p]du \leq c\sigma_1^p|t-s|^{\frac{p}{2}}. \tag{60}$$

To evaluate I_3, Kunita inequality will be useful. We refer to the Appendix of [24] for its proof in a general form, while below (A7) on page 52 of [2] can be found an example of its application in a form closer to the one we are going to use. For a compensated Poisson random measure $\tilde{\mu} = \mu - \bar{\mu}$ and a jump coefficient $l(x,z)$, indeed, Kunita inequality provides the following:

$$\begin{aligned}\mathbb{E}\left[\left|\int_0^t\int_{\mathbb{R}} l(X_{s^-},z)\tilde{\mu}(ds,dz)\right|^p\right] \quad &\leq \quad c\mathbb{E}\left[\int_0^t\int_{\mathbb{R}} |l(X_{s^-},z)|^p\bar{\mu}(ds,dz)\right] \\ &\quad + c\mathbb{E}\left[\left|\int_0^t\int_{\mathbb{R}} l^2(X_{s^-},z)\bar{\mu}(ds,dz)\right|^{\frac{p}{2}}\right].\end{aligned}$$

We remark that, up to change the constant c in the right hand side, the equation here above holds with the measure μ instead of the compensated one $\tilde{\mu}$. In the sequel we will apply Kunita inequality on the measure $dN_u^{(j)}$ and the compensated one $d\tilde{N}_u^{(j)}$, for $j \in \{1, ..., M\}$. The compensator is in this case $\lambda^{(j)}(u)du$.
Using on I_3 Kunita inequality together with Jensen inequality and the boundedness of a we get

$$\begin{aligned}\mathbb{E}[I_3] \quad &\leq \quad c\sum_{j=1}^M \mathbb{E}\left[\int_s^t |a(X_{u^-})|^p\lambda_u^{(j)}du + \left(\int_s^t a^2(X_{u^-})\lambda_u^{(j)}du\right)^{\frac{p}{2}}\right] \\ &\leq \sum_{j=1}^M c|a_1|^p\int_s^t \mathbb{E}[\lambda_u^{(j)}]du + c|a_1|^p|t-s|^{\frac{p}{2}-1}\int_s^t \mathbb{E}[|\lambda_u^{(j)}|^{\frac{p}{2}}]du \\ &\leq c|a_1|^p(|t-s| + |t-s|^{\frac{p}{2}}) = c|a_1|^p|t-s|. \end{aligned}\tag{61}$$

From (59), (60) and (61), as $|t-s|<1$, it follows $\mathbb{E}[|X_t - X_s|^p] \leq c_1|t-s|$.

Point 2
Concerning the second point, for any $j \in \{1,\dots,M\}$ it is

$$|\lambda_t^{(j)} - \lambda_s^{(j)}|^p \leq c\left|\alpha\int_s^t(\lambda^{(j)}(u)-\zeta_j)du\right|^p + c\left|\int_s^t\sum_{i=1}^M c_{i,j}dN_u^{(i)}\right|^p.$$

Acting as in the proof of the first point, using as main arguments Jensen inequality, Kunita inequality and the boundedness of the moments of λ, we easily get the wanted estimation.
Point 3
We consider the dynamic of λ gathered in (2) in matrix form and so we have

$$\lambda_t - \lambda_s = \alpha\int_s^t(\lambda_u - \zeta)du + \int_s^t cdN_u =: D_s + G_s,$$

where $\lambda_t = (\lambda_t^{(1)},\dots,\lambda_t^{(M)})$, $c \in \mathbb{R}^M \times \mathbb{R}^M$. We start evaluating D_s. By adding and subtracting λ_s we easily get, denoting as $\mathbb{E}_s[\cdot]$ the quantity $\mathbb{E}[\cdot|\mathcal{F}_s]$,

$$\mathbb{E}_s[|D_s|] \leq c|t-s|(1+|\lambda_s|) + c\int_s^t\mathbb{E}_s[|\lambda_u - \lambda_s|]ds.$$

On G_s we use compensation formula and we apply the same reasoning as before, getting

$$\mathbb{E}_s[|G_s|] \leq \mathbb{E}_s\left[\int_s^t c|\lambda_u|du\right] \leq c|t-s||\lambda_s| + c\int_s^t\mathbb{E}_s[|\lambda_u-\lambda_s|]ds.$$

Putting the pieces together it follows

$$\mathbb{E}_s[|\lambda_t - \lambda_s|] \leq c|t-s|(1+|\lambda_s|) + c\int_s^t\mathbb{E}_s[|\lambda_u-\lambda_s|]ds.$$

We use Gronwall lemma, which yields

$$\mathbb{E}_s[|\lambda_t - \lambda_s|] \leq c|t-s|(1+|\lambda_s|)e^c.$$

Point 4 We observe that, for any $h \in [0,1]$,

$$\mathbb{E}_s[|\lambda_{s+h}^{(j)}|] \leq |\lambda_s^{(j)}| + \mathbb{E}_s[|\lambda_{s+h}^{(j)} - \lambda_s^{(j)}|] \leq |\lambda_s^{(j)}| + c|h|(1+|\lambda_s|),$$

where we have used the just showed third point of this lemma.

□

A.1 Proof of Proposition 1

Proof. We write $V(x,y) = V_1(x) + V_2(y)$, where $V_1(x) = |x|^m$ for m arbitrarily big and $V_2(y) = e^{\sum_{i,j} m_{ij}|y^{(ij)}|}$. From the definition (5) of $A^{\tilde{z}}$ we have

$$A^{\tilde{z}}V = A_1^{\tilde{z}}V + A_2^{\tilde{z}}V,$$

where

$$\begin{aligned} A_1^{\tilde{z}}V(x,y) &:= \partial_x V(x,y)b(x) + \frac{1}{2}\sigma^2(x)\partial_x^2 V(x,y) + \sum_{j=1}^M f_j(\sum_{k=1}^M y^{(jk)})[V_1(x+a(x)) - V_1(x)] \\ &= m|x|^{m-1}b(x) + \frac{1}{2}\sigma^2(x)m(m-1)|x|^{m-2} + \sum_{j=1}^M f_j\left(\sum_{k=1}^M y^{(jk)}\right)[|x+a(x)|^m - |x|^m] \end{aligned}$$

is the diffusion part and

$$\begin{aligned} A_2^{\tilde{z}}V(x,y) &:= A^{\tilde{z}}V(x,y) - A_1^{\tilde{z}}V(x,y) \\ &= -\alpha\sum_{i,j=1}^{M} y^{(ij)}\partial_{y^{(ij)}}V(x,y) + \sum_{j=1}^{M} f_j(\sum_{k=1}^{M} y^{(jk)})[V_2(y+\Delta_j) - V_2(y)], \end{aligned}$$

is the jump part of the generator. The arguments of the proof of Proposition 4.5 in [11] imply that

$$A_2^{\tilde{z}}V(x,y) = A_2^{\tilde{z}}V_2(y) \leq -c_1V_2(y) + c_2\mathbb{1}_{K_1}(y), \tag{62}$$

with c_1 and c_2 some positive constants and K_1 some compact of $\mathbb{R}^{M\times M}$. Moreover, denoting $\bar{f}(y) := \sum_{j=1}^{M} f_j(\sum_{k=1}^{M} y^{(jk)})$ the total jump rate, it is

$$A_1^{\tilde{z}}V(x,y) = m|x|^{m-1}b(x) + \frac{1}{2}\sigma^2(x)m(m-1)|x|^{m-2} + \bar{f}(y)[|x+a(x)|^m - |x|^m].$$

From the drift condition on b gathered in the fourth point of Assumption 1 and the boundedness of both σ^2 and a it follows

$$A_1^{\tilde{z}}V(x,y) \leq -dm|x|^m + c|x|^{m-2} + \bar{f}(y)(c_1|x|^{m-1} + \ldots + c_m). \tag{63}$$

We observe that, for any x such that $|x| > r$, $|x|^{m-2}$ is negligible compared to $|x|^m = V_1(x)$. To study the last term in the right hand side of (63), we choose $1 < p < 2$ and $q > 2$ such that $p(m-1) < m$ (i e $p < 1 + \frac{1}{m-1}$) and $\frac{1}{p} + \frac{1}{q} = 1$. Then,

$$\bar{f}(y)(c_1|x|^{m-1} + \ldots + c_m) \leq \frac{c}{p}(c_1|x|^m + \ldots + c_m)^p + \frac{c}{q}\bar{f}(y)^q.$$

The first term is again negligible compared to $|x|^m = V_1(x)$, being $p(m-1) < m$. To estimate the second one we observe that, for each $y \in \mathbb{R}^M$ the total jump rate $\bar{f}(y)$ can be seen as $\sum_{i=1}^{M}(\zeta_i + \sum_j y^{(ij)})$ (see page 12 in [18]). Therefore, it is

$$\bar{f}(y) \leq \bar{c} + \tilde{c}\sum_{i,j=1}^{M}|y^{(ij)}| \leq \bar{c} + \tilde{c}_2\log(V_2(y)),$$

which is negligible with respect to the negative term of (62) $-c_1V_2(y)$. The same reasoning applies for $\frac{c}{q}\bar{f}(y)^q$. It follows that

$$A_1^{\tilde{z}}V(x,y) \leq -dm|x|^m + o(V_1(x)) + o(V_2(y))$$

which, together with (62), conclude the proof. □

A.2 Proof of Lemma 1

Proof. By the definition of φ, for any $k \geq 1$ $|\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X) - 1|^k$ is different from zero only if $|\Delta_i X| > \Delta_{n,i}^{\beta}$. Therefore,

$$\begin{aligned} \mathbb{E}[|\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X) - 1|^k] &\leq c\mathbb{E}[\mathbb{1}_{\left\{|\Delta_i X| > \Delta_{n,i}^{\beta}\right\}}] \\ &= c\mathbb{E}\left[\mathbb{1}_{\left\{|\Delta_i X| > \Delta_{n,i}^{\beta}, |J_{t_i}| \leq \frac{\Delta_{n,i}^{\beta}}{2}\right\}}\right] + c\mathbb{E}\left[\mathbb{1}_{\left\{|\Delta_i X| > \Delta_{n,i}^{\beta}, |J_{t_i}| > \frac{\Delta_{n,i}^{\beta}}{2}\right\}}\right] \end{aligned} \tag{64}$$

We denote as $\Delta_i X^c$ the increment of the continuous part of X, which is

$$\Delta_i X^c := X^c_{t_{i+1}} - X^c_{t_i} = \int_{t_i}^{t_{i+1}} b(X_s)ds + Z_{t_i}.$$

The first term in the right hand side of (64) is

$$c\mathbb{E}\left[\mathbb{1}_{\left\{|\Delta_i X^c|>\frac{\Delta^\beta_{n,i}}{2}\right\}}\right] = c\mathbb{P}(|\Delta_i X^c| > \frac{\Delta^\beta_{n,i}}{2}) \leq \frac{c\,\mathbb{E}\left[|\Delta_i X^c|^r\right]}{\Delta^{\beta r}_{n,i}} \leq c\Delta^{r(\frac{1}{2}-\beta)}_{n,i}, \tag{65}$$

where we have used Markov inequality and a classical estimation for the continuous increments of X (see for example point 6 of Lemma 1 in [4]). In order to evaluate the second term in the right hand side of (64), instead, we have to introduce the set

$$N_{i,n} := \left\{\sum_{j=1}^M |\Delta_i N^{(j)}| := \sum_{j=1}^M |N^{(j)}_{t_{i+1}} - N^{(j)}_{t_i}| \leq \frac{4\Delta^\beta_{n,i}}{a_1}\right\}.$$

We observe that, on $N^c_{i,n}$, there exists $j \in \{1,\dots,M\}$ such that $|\Delta_i N^{(j)}| \neq 0$. Therefore,

$$\mathbb{P}(N^c_{i,n}) \leq \mathbb{P}(|\Delta_i N^{(j)}| \geq 1) \leq \mathbb{E}[|\Delta_i N^{(j)}|] \leq c\Delta_{n,i}. \tag{66}$$

On $N_{i,n}$, instead, $\forall j\ |\Delta_i N^{(j)}| = 0$ and so $(N_{i,n}) \cap \left\{|J_{t_i}| > \frac{\Delta^\beta_{n,i}}{2}\right\} = \emptyset$. It follows that the second term in the right hand side of (64) is

$$c\mathbb{E}\left[\mathbb{1}_{\left\{|\Delta_i X|>\Delta^\beta_{n,i},|J_{t_i}|>\frac{\Delta^\beta_{n,i}}{2},N_{i,n}\right\}}\right] + c\mathbb{E}[\mathbb{1}_{\left\{|\Delta_i X|>\Delta^\beta_{n,i},|J_{t_i}|>\frac{\Delta^\beta_{n,i}}{2},N^c_{i,n}\right\}}] \leq c\mathbb{P}(N^c_{i,n}) \leq c\Delta_{n,i}.$$

Putting the pieces together, as r is arbitrary, it follows

$$\mathbb{E}\left[\left|\varphi_{\Delta^\beta_{n,i}}(\Delta_i X) - 1\right|^k\right] \leq c\Delta_{n,i}.$$

□

A.3 Proof of Lemma 2

Proof. Again, we act differently depending on whether the jumps are big or not:

$$\mathbb{E}[|J_{t_i}|^q \varphi^k_{\Delta^\beta_{n,i}}(\Delta_i X)] = \mathbb{E}\left[|J_{t_i}|^q \varphi^k_{\Delta^\beta_{n,i}}(\Delta_i X)\mathbb{1}_{\left\{|J_{t_i}|>\frac{\Delta^\beta_{n,i}}{2}\right\}}\right] + \mathbb{E}\left[|J_{t_i}|^q \varphi^k_{\Delta^\beta_{n,i}}(\Delta_i X)\mathbb{1}_{\left\{|J_{t_i}|\leq\frac{\Delta^\beta_{n,i}}{2}\right\}}\right]. \tag{67}$$

By the definition of φ it is different from 0 only if $|\Delta_i X| \leq 2\Delta^\beta_{n,i}$. As $\Delta_i X = \Delta_i X^c + J_{t_i}$, it is

$$\mathbb{E}\left[|J_{t_i}|^q \varphi^k_{\Delta^\beta_{n,i}}(\Delta_i X)\mathbb{1}_{\left\{|J_{t_i}|>\frac{\Delta^\beta_{n,i}}{2}\right\}}\right] \leq \mathbb{E}\left[|J_{t_i}|^q \mathbb{1}_{\left\{|\Delta_i X^c|>\frac{3\Delta^\beta_{n,i}}{2}\right\}}\right] \leq \mathbb{E}[|J_{t_i}|^{qp_1}]^{\frac{1}{p_1}}\mathbb{E}\left[\mathbb{1}_{\left\{|\Delta_i X^c|>\frac{3\Delta^\beta_{n,i}}{2}\right\}}\right]^{\frac{1}{p_2}}$$

$$\leq c\Delta^{\frac{1}{p_1}}_{n,i}\Delta^{\frac{r}{p_2}(\frac{1}{2}-\beta)}_{n,i} c\Delta^{r(\frac{1}{2}-\beta)-\varepsilon}_{n,i},$$

where we have used first of all Holder inequality and then Kunita inequality and (65). In order to evaluate the second term of (67), we introduce again the set $N_{i,n}$. On $N_{i,n}$ the increments $\Delta_i N^{(j)}$ are null and so $|J_{t_i}| = 0$. On $N_{i,n}^c$ instead, using also (66), we have

$$\mathbb{E}\left[|J_{t_i}|^q \varphi^k_{\Delta^\beta_{n,i}}(\Delta_i X)\mathbb{1}_{\left\{|J_{t_i}|\leq \frac{\Delta^\beta_{n,i}}{2}, N^c_{i,n}\right\}}\right] \leq c\Delta^{\beta q}_{n,i}\mathbb{P}(N^c_{i,n}) \leq c\Delta^{1+\beta q}_{n,i}.$$

By the arbitrariness of r it follows

$$\mathbb{E}[|J_{t_i}|^q \varphi^k_{\Delta^\beta_{n,i}}(\Delta_i X)] \leq c\Delta^{1+\beta q}_{n,i},$$

as we wanted. □

A.4 Proof of Proposition 2

Proof. As the second point is useful in order to show the first one, we start proving point 2.
Point 2
By definition we know that B_{t_i} is centered. In the sequel we denote as $\mathbb{E}_i[\cdot]$ the conditional expected value $\mathbb{E}[\cdot|\mathcal{F}_{t_i}]$. Regarding the second moment, it is

$$\mathbb{E}_i[B^2_{t_i}] \leq \frac{1}{\Delta^2_{n,i}}\mathbb{E}_i\left[Z^4_{t_i} + (\int_{t_i}^{t_{i+1}} \sigma^2(X_s)ds)^2\right] \leq \frac{c}{\Delta^2_{n,i}}\mathbb{E}_i\left[\left(\int_{t_i}^{t_{i+1}} \sigma^2(X_s)ds\right)^2\right] \leq c\sigma_1^4$$

where we have used, sequentially, BDG inequality, Jensen inequality and the boundedness of σ. Using the same arguments we show the following:

$$\mathbb{E}_i[B^4_{t_i}] \leq \frac{1}{\Delta^4_{n,i}}\mathbb{E}_i\left[Z^8_{t_i} + \left(\int_{t_i}^{t_{i+1}} \sigma^2(X_s)ds\right)^4\right] \leq \frac{c}{\Delta^4_{n,i}}\mathbb{E}_i\left[\left(\int_{t_i}^{t_{i+1}} \sigma^2(X_s)ds\right)^4\right] \leq c\sigma_1^8.$$

Point 1
We analyse the behaviour of

$$\tilde{A}_{t_i} = \sigma^2(X_{t_i})(\varphi_{\Delta^\beta_{n,i}}(\Delta_i X) - 1) + A_{t_i}\varphi_{\Delta^\beta_{n,i}}(\Delta_i X) + B_{t_i}(\varphi_{\Delta^\beta_{n,i}}(\Delta_i X) - 1).$$

From Holder inequality, the boundedness of σ and a repeated use of Lemma 2 we get

$$\mathbb{E}[\tilde{A}^2_{t_i}] \leq c\sigma_1^4\Delta_{n,i} + \mathbb{E}[A^2_{t_i}\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)] + \mathbb{E}[B^{2p}_{t_i}]^{\frac{1}{p}} c\Delta^{\frac{1}{q}}_{n,i}.$$

We evaluate the moments of B_{t_i} acting as in the proof of the first point and we choose p big and q next to 1, getting

$$\mathbb{E}[\tilde{A}^2_{t_i}] \leq c\sigma_1^4\Delta_{n,i} + \mathbb{E}[A^2_{t_i}\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)] + c\sigma_1^2\Delta^{1-\tilde{\varepsilon}}_{n,i}, \tag{68}$$

for $\tilde{\varepsilon} > 0$ arbitrarily small. We are left to study $A^2_{t_i}\varphi^2_{\Delta^\beta_{n,i}}$. From its definition, recalling that φ is a bounded function, we obtain

$$\begin{aligned}\mathbb{E}[A^2_{t_i}\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)] \leq\ & \frac{c}{\Delta^2_{n,i}}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} b(X_s)ds\right)^4\right] + \frac{c}{\Delta^2_{n,i}}\mathbb{E}\left[(Z_{t_i} + J_{t_i})^2\left(\int_{t_i}^{t_{i+1}} b(X_s) - b(X_{t_i})ds\right)^2\right] \\ & + \frac{c}{\Delta^2_{n,i}}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} \sigma^2(X_s) - \sigma^2(X_{t_i})ds\right)^2\right] + 4\mathbb{E}[b^2(X_{t_i})Z^2_{t_i}] =: \sum_{j=1}^4 I_j.\end{aligned}$$

Using Jensen inequality, the polynomial growth of b and the existence of bounded moments of X we get

$$I_1 \leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}^3 \int_{t_i}^{t_{i+1}} \mathbb{E}[b^4(X_s)]ds \leq c\Delta_{n,i}^2. \tag{69}$$

On I_2 we use first of all Holder inequality. Then, on the first we use B.D.G. and Kunita inequalities, as in (60) and (61), while on the second the finite increments theorem, the boundedness of b' and the first point of Lemma 7:

$$\begin{aligned} I_2 &\leq \frac{c}{\Delta_{n,i}^2}\mathbb{E}[(Z_{t_i}+J_{t_i})^4]^{\frac{1}{2}}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} b(X_s)-b(X_{t_i})ds\right)^4\right]^{\frac{1}{2}} \\ &\leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}^{\frac{1}{2}}\Delta_{n,i}^{\frac{3}{2}}\mathbb{E}\left[\int_{t_i}^{t_{i+1}} c|X_s-X_{t_i}|^4 ds\right]^{\frac{1}{2}} \leq c\Delta_{n,i}. \end{aligned} \tag{70}$$

In order to study the behaviour of I_3, Jensen inequality, the finite increment theorem, the boundedness of the derivative of σ^2 and the first point of Lemma 7 will be once again useful.

$$I_3 \leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}\mathbb{E}\left[\int_{t_i}^{t_{i+1}} c|X_s-X_{t_i}|^2 ds\right] \leq c\Delta_{n,i}. \tag{71}$$

From Holder inequality, the polynomial growth of b, the boundedness of the moments of X and BDG inequality we obtain

$$I_4 \leq c\mathbb{E}[b(X_{t_i})^4]^{\frac{1}{2}}\mathbb{E}[Z_{t_i}^4]^{\frac{1}{2}} \leq c\Delta_{n,i}. \tag{72}$$

Putting the pieces together it follows that, for any $\tilde{\varepsilon}>0$,

$$\mathbb{E}[\tilde{A}_{t_i}^2] \leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}.$$

We now evaluate $\mathbb{E}[\tilde{A}_{t_i}^4]$. Acting as above (68) it easily follows

$$\mathbb{E}[\tilde{A}_{t_i}^4] \leq c\Delta_{n,i}^{1-\tilde{\varepsilon}} + \mathbb{E}[A_{t_i}^4\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X)].$$

Replacing the definition of A_{t_i} we get that $\mathbb{E}[A_{t_i}^4\varphi_{\Delta_{n,i}^{\beta}}^4(\Delta_i X)]$ is again the sum of 4 terms, that we now denote as $\tilde{I}_1,\ldots,\tilde{I}_4$. Using exactly the same arguments as in the study of $\mathbb{E}[A_{t_i}^4\varphi_{\Delta_{n,i}^{\beta}}^4(\Delta_i X)]$ we easily get

$$\tilde{I}_1 \leq \frac{c}{\Delta_{n,i}^4}\Delta_{n,i}^7 \int_{t_i}^{t_{i+1}} \mathbb{E}[b^8(X_s)]ds \leq c\Delta_{n,i}^4,$$

$$\begin{aligned} \tilde{I}_2 &\leq \frac{c}{\Delta_{n,i}^4}\mathbb{E}[(Z_{t_i}+J_{t_i})^4]^{\frac{1}{2}}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} b(X_s)-b(X_{t_i})ds\right)^8\right]^{\frac{1}{2}} \\ &\leq \frac{c}{\Delta_{n,i}^4}(\Delta_{n,i}+\Delta_{n,i}^{\frac{1}{2}})\Delta_{n,i}^{\frac{7}{2}}\mathbb{E}\left[\int_{t_i}^{t_{i+1}} c|X_s-X_{t_i}|^8 ds\right]^{\frac{1}{2}} \leq c\Delta_{n,i}, \end{aligned}$$

$$\tilde{I}_3 \leq \frac{c}{\Delta_{n,i}^4}\Delta_{n,i}^3\mathbb{E}\left[\int_{t_i}^{t_{i+1}} c|X_s-X_{t_i}|^4 ds\right] \leq c\Delta_{n,i},$$

$$\tilde{I}_4 \leq \mathbb{E}[b(X_{t_i})^8]^{\frac{1}{2}}\mathbb{E}[Z_{t_i}^8]^{\frac{1}{2}} \leq c\Delta_{n,i}^2.$$

The four equations here above provide the wanted result.

Point 3

In order to show the estimations on the jumps gathered in the third point of Proposition 3 we repeatedly use Lemma 2. Using also Holder inequality with p big and q next to 1, BDG inequality, the polynomial growth of b and the boundedness of the moments of X it is

$$\begin{aligned}\mathbb{E}[|E_{t_i}|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] &\leq c\mathbb{E}[|b(X_{t_i})||J_{t_i}|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] + \frac{c}{\Delta_{n,i}}\mathbb{E}[|Z_{t_i}||J_{t_i}|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] \\ &\quad + \frac{c}{\Delta_{n,i}}\mathbb{E}\left[|J_{t_i}|^2\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)\right] \\ &\leq c\mathbb{E}[|b(X_{t_i})|^p]^{\frac{1}{p}}\mathbb{E}[|J_{t_i}|^q\varphi^q_{\Delta_{n,i}^\beta}(\Delta_i X)]^{\frac{1}{q}} + \frac{c}{\Delta_{n,i}}\mathbb{E}[|Z_{t_i}|^p]^{\frac{1}{p}}\mathbb{E}[|J_{t_i}|^q\varphi^q_{\Delta_{n,i}^\beta}(\Delta_i X)]^{\frac{1}{q}} \\ &\quad + \frac{c}{\Delta_{n,i}}\Delta_{n,i}^{1+2\beta}\end{aligned}$$

thus, because, as $\beta \in (0, \frac{1}{2})$, we can always find an $\varepsilon > 0$ such that $\frac{1}{2} + \beta - \varepsilon > 2\beta$, it comes

$$\mathbb{E}[|E_{t_i}|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] \leq\leq c\Delta_{n,i}^{\frac{1}{q}+\beta} + \frac{c}{\Delta_{n,i}}\Delta_{n,i}^{\frac{1}{2}}\Delta_{n,i}^{\frac{1}{q}+\beta} + c\Delta_{n,i}^{2\beta} = c\Delta_{n,i}^{1+\beta-\varepsilon} + c\Delta_{n,i}^{\frac{1}{2}+\beta-\varepsilon} + c\Delta_{n,i}^{2\beta} = c\Delta_{n,i}^{2\beta}.$$

In analogous way we obtain

$$\begin{aligned}\mathbb{E}[|E_{t_i}|^2\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] &\leq c\mathbb{E}[|b(X_{t_i})|^{2p}]^{\frac{1}{p}}\mathbb{E}[|J_{t_i}|^{2q}\varphi^q_{\Delta_{n,i}^\beta}(\Delta_i X)]^{\frac{1}{q}} + \frac{c}{\Delta_{n,i}^2}\mathbb{E}[|Z_{t_i}|^{2p}]^{\frac{1}{p}}\mathbb{E}[|J_{t_i}|^{2q}\varphi^q_{\Delta_{n,i}^\beta}(\Delta_i X)]^{\frac{1}{q}} \\ &\quad + \frac{c}{\Delta_{n,i}^2}\mathbb{E}[|J_{t_i}|^4\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] \\ &\leq c\Delta_{n,i}^{\frac{1}{q}+2\beta} + \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}\Delta_{n,i}^{\frac{1}{q}+2\beta} + \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}^{1+4\beta} \\ &= c\Delta_{n,i}^{1+2\beta-\varepsilon} + c\Delta_{n,i}^{2\beta-\varepsilon} + c\Delta_{n,i}^{4\beta-1} = c\Delta_{n,i}^{4\beta-1},\end{aligned}$$

where the last inequality is, again, consequence of the fact that we can always find $\varepsilon > 0$ for which $2\beta - \varepsilon > 4\beta - 1$. Finally, acting as before,

$$\mathbb{E}[|E_{t_i}|^4\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] \leq c\Delta_{n,i}^{1+4\beta-\varepsilon} + \frac{c}{\Delta_{n,i}^4}\Delta_{n,i}^2\Delta_{n,i}^{1+4\beta-\varepsilon} + \frac{c}{\Delta_{n,i}^4}\Delta_{n,i}^{1+8\beta} = c\Delta_{n,i}^{8\beta-3}.$$

□

A.5 Proof of Proposition 4

Proof. Point 1

Regarding the first point, we first of all introduce $\tilde{b}(X_s) := b(X_s) + a(X_{s^-})\sum_{j=1}^M \lambda_s^{(j)} ds$. We observe that, as b has polynomial growth, a is bounded and both λ and X have bounded moments of any order, then $\tilde{b}$ has bounded moments of any order as well. Recalling that A_{t_i} is given as in (18) we can denote

$$A_{t_i} =: \sum_{j=1}^7 \bar{I}_j.$$

Replacing $\tilde{b}$ with b, we already know from (69), (70), (71) and (72) that

$$\mathbb{E}[\bar{I}_1^2 + \bar{I}_2^2 + \bar{I}_3^2 + \bar{I}_6^2] \leq c\Delta_{n,i}. \tag{73}$$

We now consider $\bar{I}_4$. From Assumption 1 we know the function a is Lipschitz and with bounded derivative. Therefore, we use the finite increments theorem followed by the first point of Lemma

7. It provides us, using also Jensen inequality and Holder inequality with q big and p next to 1,

$$\begin{aligned}\mathbb{E}[\bar{I}_4^2] &\leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}\int_{t_i}^{t_{i+1}}\mathbb{E}\left[(a^2(X_s)-a^2(X_{t_i}))^2\left(\sum_{j=1}^{M}\lambda_s^{(j)}\right)^2\right]ds \\ &\leq \frac{c}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}}\mathbb{E}[(a^2(X_s)-a^2(X_{t_i}))^{2p}]^{\frac{1}{p}}\mathbb{E}\left[\left(\sum_{j=1}^{M}\lambda_s^{(j)}\right)^{2q}\right]^{\frac{1}{q}}ds \\ &\leq \frac{c}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}}\Delta_{n,i}^{\frac{1}{p}}ds \leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}, \end{aligned} \tag{74}$$

where we have also used the boundedness of the moments of λ. On $\bar{I}_5$ we use that $a(x)\leq a_1$ and the second point of Lemma 7, getting

$$\mathbb{E}[\bar{I}_5^2]\leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}\int_{t_i}^{t_{i+1}}\sum_{j=1}^{M}\mathbb{E}[(\lambda_s^{(j)}-\lambda_{t_i}^{(j)})^2]ds\leq c\Delta_{n,i}. \tag{75}$$

To conclude the proof of the bound on $\mathbb{E}[A_{t_i}^2]$ we are left to evaluate $\bar{I}_7$. We do that through Holder and Kunita inequalities. It yields

$$\mathbb{E}[\bar{I}_7^2]\leq c\mathbb{E}[\bar{b}(X_{t_i})^2J_{t_i}^2]\leq \mathbb{E}[\bar{b}(X_{t_i})^{2p}]^{\frac{1}{p}}\mathbb{E}[J_{t_i}^{2q}]^{\frac{1}{q}}\leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}, \tag{76}$$

where in the last inequality we have chosen p big and q next to 1. From (73), (74), (75) and (76) it follows

$$\mathbb{E}[A_{t_i}^2]\leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}.$$

Concerning the fourth moment of A_{t_i}, as before we know from Proposition 2 that

$$\mathbb{E}[\bar{I}_1^4+\bar{I}_2^4+\bar{I}_3^4+\bar{I}_6^4]\leq c\Delta_{n,i}. \tag{77}$$

Acting as in (74) we get

$$\mathbb{E}[\bar{I}_4^4]\leq \frac{c}{\Delta_{n,i}^4}\Delta_{n,i}^3\int_{t_i}^{t_{i+1}}\mathbb{E}\left[(a^2(X_s)-a^2(X_{t_i}))^4\left(\sum_{j=1}^{M}\lambda_s^{(j)}\right)^4\right]ds \tag{78}$$

$$\leq \frac{c}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}}\mathbb{E}[(a^2(X_s)-a^2(X_{t_i}))^{4p}]^{\frac{1}{p}}\mathbb{E}\left[\left(\sum_{j=1}^{M}\lambda_s^{(j)}\right)^{4q}\right]^{\frac{1}{q}}ds\leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}.$$

In the same way, acting as in (75) we obtain

$$\mathbb{E}[\bar{I}_5^4]\leq \frac{c}{\Delta_{n,i}^4}\Delta_{n,i}^3\int_{t_i}^{t_{i+1}}\sum_{j=1}^{M}\mathbb{E}[(\lambda_s^{(j)}-\lambda_{t_i}^{(j)})^4]ds\leq c\Delta_{n,i}. \tag{79}$$

We conclude the proof of the point 2 by observing that

$$\mathbb{E}[\bar{I}_7^4]\leq c\mathbb{E}[\bar{b}(X_{t_i})^{4p}]^{\frac{1}{p}}\mathbb{E}[J_{t_i}^{4q}]^{\frac{1}{q}}\leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}, \tag{80}$$

by the boundedness of the moments of $\tilde{b}$ and Kunita inequality.

Point 2

We observe that B_{t_i} is defined in the same way in Section 3 and Section 4. Therefore, the second

point has already been showed in point 2 of Proposition 2.

Point 3
By the definition of E_{t_i} it clearly follows $\mathbb{E}_i[E_{t_i}] = 0$. We now analyse

$$\mathbb{E}_i[E_{t_i}^2] \leq \frac{c}{\Delta_{n,i}^2}\mathbb{E}_i[Z_{t_i}^2 J_{t_i}^2] + \frac{c}{\Delta_{n,i}^2}\mathbb{E}_i[J_{t_i}^4 + \left(\int_{t_i}^{t_{i+1}} a^2(X_{s^-})\sum_{j=1}^{M}\lambda_s^{(j)}ds\right)^2]. \tag{81}$$

We are going to show that the first term in the right hand side of the equation (81) is negligible if compared to the second one. By a conditional version of Holder, BDG and Kunita inequalities we get

$$\frac{c}{\Delta_{n,i}^2}\mathbb{E}_i[Z_{t_i}^2 J_{t_i}^2] \leq \frac{c}{\Delta_{n,i}^2}\mathbb{E}_i[Z_{t_i}^{2p}]^{\frac{1}{p}}\mathbb{E}_i[J_{t_i}^{2q}]^{\frac{1}{q}} \leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}\Delta_{n,i}^{\frac{1}{q}} \leq c\Delta_{n,i}^{-\varepsilon}, \tag{82}$$

for any $\varepsilon > 0$. To study the last term in the right hand side of (81) we recall it is $J_{t_i} = \int_{t_i}^{t_{i+1}} a(X_{s^-})\sum_{j=1}^{M} d\tilde{N}_s^{(j)}$. Therefore, from conditional Kunita inequality, we have

$$\begin{aligned}
\frac{c}{\Delta_{n,i}^2}\mathbb{E}_i\left[J_{t_i}^4 + \left(\int_{t_i}^{t_{i+1}} a^2(X_{s^-})\sum_{j=1}^{M}\lambda_s^{(j)}ds\right)^2\right] &\leq \frac{c}{\Delta_{n,i}^2}\mathbb{E}_i\left[\int_{t_i}^{t_{i+1}} a^4(X_{s^-})\sum_{j=1}^{M}\lambda_s^{(j)}ds\right. \\
&\quad \left. +2\left(\int_{t_i}^{t_{i+1}} a^2(X_{s^-})\sum_{j=1}^{M}\lambda_s^{(j)}ds\right)^2\right] \\
&\leq \frac{ca_1^4}{\Delta_{n,i}^2}(1+\Delta_{n,i})\int_{t_i}^{t_{i+1}}\mathbb{E}_i\left[\sum_{j=1}^{M}\lambda_s^{(j)}\right]ds,
\end{aligned}$$

where we have also used Jensen inequality on the last term here above, which is the reason why we get an extra $\Delta_{n,i}$. From the fourth point of Lemma 7 it follows that the equation here above is upper bounded by $\frac{ca_1^4}{\Delta_{n,i}}\sum_{j=1}^{M}\lambda_{t_i}^{(j)}$, plus a negligible term. Replacing it and (82) in (81) it follows

$$\mathbb{E}_i[E_{t_i}^2] \leq \frac{ca_1^4}{\Delta_{n,i}}\sum_{j=1}^{M}\lambda_{t_i}^{(j)} + c\Delta_{n,i}^{-\varepsilon} \leq \frac{ca_1^4}{\Delta_{n,i}}\sum_{j=1}^{M}\lambda_{t_i}^{(j)},$$

where the last inequality is a consequence of the fact that λ is always strictly more than zero. Regarding the fourth moment of E_{t_i}, from Kunita, Holder and Jensen inequality we have

$$\begin{aligned}
\mathbb{E}[E_{t_i}^4] &\leq \frac{c}{\Delta_{n,i}^4}\mathbb{E}[Z_{t_i}^{4p}]^{\frac{1}{p}}\mathbb{E}[J_{t_i}^{4q}]^{\frac{1}{q}} + \frac{c}{\Delta_{n,i}^4}\mathbb{E}_i\left[J_{t_i}^8 + \left(\int_{t_i}^{t_{i+1}} a^2(X_{s^-})\sum_{j=1}^{M}\lambda_s^{(j)}ds\right)^4\right] \\
&\leq \frac{c}{\Delta_{n,i}^4}(\Delta_{n,i}^2\Delta_{n,i}^{1-\varepsilon} + \Delta_{n,i} + \Delta_{n,i}^3\Delta_{n,i}) \leq \frac{c}{\Delta_{n,i}^3}.
\end{aligned}$$

Point 4
The result follows directly from the definition of C_{t_i} and the boundedness of a and of the moments of λ.

□

A.6 Proof of Lemma 3

Proof. It is

$$C_{t_i}\tilde{\psi}_l(X_{t_i},\lambda_{t_i}) = a^2(X_{t_i})\sum_{j=1}^{M}(\lambda_{t_i}^{(j)} - \mathbb{E}[\lambda_{t_i}^{(j)}|X_{t_i}])\tilde{\psi}_l(X_{t_i},\lambda_{t_i}) =: f(X_{t_i},\lambda_{t_i}).$$

Since

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\lambda_{t_i})\right) \leq \frac{1}{n^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\mathrm{Cov}(f(X_{t_i},\lambda_{t_i}), f(X_{t_j},\lambda_{t_j})),$$

we need to estimate the covariance.
As explained in Section 2.3 we know that, under our assumptions, the process $Z := (X,\lambda)$ is β-mixing with exponential decay. It means that there exists $\gamma > 0$ such that

$$\beta_X(t) \leq \beta_Z(t) \leq Ce^{-\gamma t}.$$

If the process Y is β- mixing, then it is also α-mixing and so the following estimation holds (see Theorem 3 in Section 1.2.2 of [19])

$$|\mathrm{Cov}(Y_{t_i}, Y_{t_j})| \leq c\,\|Y_{t_i}\|_p \left\|Y_{t_j}\right\|_q \alpha^{\frac{1}{r}}(Y_{t_i}, Y_{t_j})$$

with p, q and r such that $\frac{1}{p}+\frac{1}{q}+\frac{1}{r} = 1$. Using that

$$\alpha(Z_{t_i}, Z_{t_j}) \leq \beta_Z(|t_i - t_j|) \leq Ce^{-\gamma|t_i - t_j|},$$

in our case the inequality here above becomes

$$|\mathrm{Cov}(f(X_{t_i},\lambda_{t_i}), f(X_{t_j},\lambda_{t_j}))| \leq ce^{-\frac{1}{r}\gamma|t_i - t_j|},$$

where we have also used the definition of f and the boundedness of a and the existence of moments of λ to include the two norms in the constant c.
We introduce a partition of $(0,T_n]$ based on the sets $A_k := (k\frac{T_n}{n}, (k+1)\frac{T_n}{n}]$, for which $(0,T_n] = \cup_{k=0}^{n-1} A_k$. Now each point t_i in $(0,T_n]$ can be seen as $t_{k,h}$, where k identifies the particular set A_k to which the point belongs while, defining M_k as $|A_k|$, h is a number in $\{1,\dots,M_k\}$ which enumerates the points in each set. It follows

$$\begin{aligned}
\frac{c}{n^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} e^{-\frac{1}{r}\gamma|t_i - t_j|} &\leq \frac{c}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}\sum_{h_1=1}^{M_{k_1}}\sum_{h_2=1}^{M_{k_2}} e^{-\frac{1}{r}\gamma|t_{k_1,h_1} - t_{k_2,h_2}|} \\
&\leq \frac{ce^{\frac{1}{r}\frac{T_n}{n}}}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}\sum_{h_1=1}^{M_{k_1}}\sum_{h_2=1}^{M_{k_2}} e^{-\frac{1}{r}\gamma|k_1 - k_2|\frac{T_n}{n}},
\end{aligned}$$

where the last inequality is a consequence of the following estimation: for each $k_1, k_2 \in \{0,\dots,n-1\}$ it is $|t_{k_1,h_1} - t_{k_2,h_2}| \geq |k_1 - k_2|\frac{T_n}{n} - \frac{T_n}{n}$.
Now we observe that the exponent does not depend on h anymore, hence the last term here above can be upper bounded by $\frac{ce^{\frac{1}{r}\frac{T_n}{n}}}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} M_{k_1}M_{k_2}e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}}$.
Moreover, remarking that the length of each interval A_k is $\frac{T_n}{n}$, it is easy to show that we can always upper bound M_k with $\frac{T_n}{n}\frac{1}{\Delta_{min}}$, with $T_n = \sum_{i=0}^{n-1}\Delta_{n,i} \leq n\Delta_n$ and so $M_k \leq \frac{\Delta_n}{\Delta_{min}}$, that we have assumed bounded by a constant c_1.
Furthermore, still using that $T_n \leq n\Delta_n$, we have $e^{\frac{1}{r}\frac{T_n}{n}} \leq e^{\frac{1}{r}\Delta_n} \leq c$.

To conclude, we have to evaluate $\frac{c}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}}$. We define $j := k_1 - k_2$ and we apply a change of variable, getting

$$\begin{aligned}\frac{c}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}} &\leq \frac{c}{n^2}\sum_{j=-(n-1)}^{n-1} e^{-\frac{1}{r}\gamma|j|\frac{T_n}{n}}|n-j| \leq \frac{c}{n}\sum_{j=-(n-1)}^{n-1} e^{-\frac{1}{r}\gamma|j|\Delta_{min}} \\ &\leq \frac{c}{n(1-e^{-\frac{1}{r}\gamma\Delta_{min}})} \leq \frac{c}{T_n},\end{aligned}$$

as we wanted.

□

A.7 Proof of Lemma 6

Proof. In order to estimate the probability of the complementary of the set Ω_B, as defined in (55), we first of all observe that, as $t \in \mathcal{B}_{m,m'}$ whose dimension is D, it is

$$|t(X_{t_k})| \leq \|t\|_\infty \leq cD^{\frac{1}{2}}.$$

Now we find an upper bound for the probability of Ω_B^c focusing on what happens for $j = 1$ and $k = 0$. It is

$$\mathbb{P}(|U_{0,1}^*| \geq \tilde{c}n^\varepsilon) \leq \mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|B_{t_k}^* + C_{t_k}^* + E_{t_k}^*| \geq \tilde{c}n^\varepsilon\right) \leq$$

$$\leq \mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|B_{t_k}^*| \geq \frac{\tilde{c}}{3}n^\varepsilon\right) + \mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|C_{t_k}^*| \geq \frac{\tilde{c}}{3}n^\varepsilon\right) + \mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|E_{t_k}^*| \geq \frac{\tilde{c}}{3}n^\varepsilon\right). \tag{83}$$

From the definition of B it is

$$\frac{1}{q_n}\sum_{k=1}^{q_n}|B_{t_k}^*| \leq \frac{c}{q_n\Delta_n}\sum_{k=1}^{q_n} Z_{t_k}^2 + c. \tag{84}$$

Moreover, using Markov inequality and the boundedness of σ,

$$\begin{aligned}\mathbb{P}\left(|Z_{t_k}| \geq c\sigma_1\Delta_n^{\frac{1}{2}}\log n\right) &= \mathbb{P}\left(e^{\frac{|Z_{t_k}|}{\sigma_1\sqrt{\Delta_n}}} \geq n^c\right) \leq \frac{1}{n^c}\mathbb{E}\left[e^{\frac{|Z_{t_k}|}{\sigma_1\sqrt{\Delta_n}}}\right] \\ &\leq \frac{1}{n^c}\mathbb{E}\left[e^{\frac{c'}{\Delta_n\sigma_1^2}\int_{t_k}^{t_{k+1}}\sigma^2(X_s)ds}\right] \leq \frac{c'}{n^c}.\end{aligned} \tag{85}$$

Therefore, as the constant c in (84) can be moved in the other side of the inequality in the first probability of (83) and so it turns out not being influential, the first probability of (83) is upper bounded by $\frac{q_n}{n^c}$, which is arbitrarily small. Concerning the second term of (83), we use Markov inequality and the fact that C has bounded moments. We get, $\forall r \geq 1$,

$$\mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|C_{t_k}^*| \geq \frac{\tilde{c}}{3}n^\varepsilon\right) \leq q_n\mathbb{P}\left(|C_{t_k}^*| \geq \frac{\tilde{c}}{3}n^\varepsilon\right) \leq cq_n\frac{\mathbb{E}[|C_{t_k}^*|^r]}{n^{r\varepsilon}} \leq \frac{cq_n}{n^{r\varepsilon}}.$$

Regarding the third term of (83) we observe that, replacing the value of q_n we get

$$\mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|E_{t_k}^*| \geq \frac{\tilde{c}}{3}n^\varepsilon\right) = \mathbb{P}\left(\sum_{k=1}^{q_n}|E_{t_k}^*| \geq \frac{\tilde{c}}{3}n^\varepsilon\frac{\log n}{\Delta_n}\right). \tag{86}$$

We now recall that, from the definition of E_{t_k} it is

$$\sum_{k=1}^{q_n} |E^*_{t_k}| \leq \left| \frac{2}{\Delta_n} \sum_{k=1}^{q_n} Z_{t_k} J_{t_k} \right| + \left| \frac{1}{\Delta_n} \sum_{k=1}^{q_n} J^2_{t_k} \right| + \left| \frac{1}{\Delta_n} \int_0^{t_{q_n}} a(X_{s^-}) \sum_{j=1}^{M} \lambda^{(j)}(s) ds \right|$$
$$=: I_1 + I_2 + I_3.$$

The right hand side of (86) is upper bounded by

$$\mathbb{P}\left(I_1 \geq \frac{\tilde{c}}{9} n^{\varepsilon} \frac{\log n}{\Delta_n} \right) + \mathbb{P}\left(I_2 \geq \frac{\tilde{c}}{9} n^{\varepsilon} \frac{\log n}{\Delta_n} \right) + \mathbb{P}\left(I_3 \geq \frac{\tilde{c}}{9} n^{\varepsilon} \frac{\log n}{\Delta_n} \right).$$

Concerning the first one, we observe it is

$$I_1 \leq \frac{1}{\Delta_n} \sum_{k=1}^{q_n} (Z^2_{t_k} + J^2_{t_k}) = I_{1,1} + I_{1,2}.$$

The probability that $I_{1,1}$ is bigger than $\frac{\tilde{c}}{9} n^{\varepsilon} \frac{\log n}{\Delta_n}$ is arbitrarily small as a consequence of (85). $I_{1,2}$ is instead equal to I_3 and so it is enough to study such a term. From Markov, Holder, BDG and Kunita inequalities we have

$$\mathbb{P}\left(I_3 \geq \frac{\tilde{c}}{9} n^{\varepsilon} \frac{\log n}{\Delta_n} \right) \leq \frac{\mathbb{E}[(I_3)^r]}{(n^{\varepsilon} \log n \Delta_n^{-1})^r} \leq \frac{c \Delta_n^{-r} t^r_{q_n}}{(n^{\varepsilon} \log n \Delta_n^{-1})^r} \leq \frac{c}{n^{\varepsilon r}},$$

where we underline that the order of t_{q_n} is $c q_n \Delta_n = c \frac{\log n}{\Delta_{min}} \Delta_n = c \log n$. It is arbitrarily small. Concerning I_2, we want to estimate $\mathbb{P}(\sum_{k=0}^{q_n - 1} J^2_{t_k} \geq \frac{c}{9} n^{\varepsilon} \log n)$. We now consider two different possibilities, starting from the definition of the following set

$$A := \left\{ \exists \tilde{k} \in \{0, \dots, q_n - 1\} \text{ such that } J^2_{t_{\tilde{k}}} \geq n^{\frac{\varepsilon}{2}} \right\}.$$

Then

$$\mathbb{P}\left(\sum_{k=0}^{q_n - 1} J^2_{t_k} \geq \frac{c}{9} n^{\varepsilon} \log n \right) = \mathbb{P}\left(\sum_{k=0}^{q_n - 1} J^2_{t_k} \geq \frac{c}{9} n^{\varepsilon} \log n, A \right) + \mathbb{P}\left(\sum_{k=0}^{q_n - 1} J^2_{t_k} \geq \frac{c}{9} n^{\varepsilon} \log n, A^c \right).$$

We observe that Markov inequality and Kunita inequality yield

$$\mathbb{P}\left(\sum_{k=0}^{q_n - 1} J^2_{t_k} \geq \frac{c}{9} n^{\varepsilon} \log n, A \right) \leq \mathbb{P}(A) \leq q_n \frac{\mathbb{E}[(J_{t_{\tilde{k}}})^{2r}]}{n^{\frac{\varepsilon r}{2}}} \leq \frac{\Delta_n q_n}{n^{\frac{\varepsilon r}{2}}} = \frac{c \log n}{n^{\frac{\varepsilon r}{2}}},$$

which is arbitrarily small by the arbitrariness of r. We remark that on A^c, for every $k \in \{0, \dots, q_n - 1\}$, it is $J^2_{t_k} < n^{\frac{\varepsilon}{2}}$. Therefore, to have the sum of them bigger than $\frac{c}{9} n^{\varepsilon} \log n$ we should have at least $\frac{c}{9} \log n n^{\frac{\varepsilon}{2}}$ jumps. Hence, denoting as ΔN_q the number of jumps in $[0, t_{q_n}]$, we have

$$\begin{aligned} \mathbb{P}\left(\sum_{k=0}^{q_n - 1} J^2_{t_k} \geq \frac{c}{9} n^{\varepsilon} \log n, A^c \right) &\leq \mathbb{P}\left(\Delta N_q > \frac{c}{9} n^{\frac{\varepsilon}{2}} \log n \right) \leq c \frac{\mathbb{E}[(\Delta N_q)^r]}{(n^{\frac{\varepsilon}{2}} \log n)^r} \\ &\leq \frac{c\, t_{q_n}}{(n^{\frac{\varepsilon}{2}} \log n)^r} \leq \frac{c}{(\log n)^{r-1} n^{\frac{\varepsilon r}{2}}}, \end{aligned}$$

where again we have used Markov inequality and we got a quantity arbitrarily small. We put all the pieces together and we observe we can choose in particular r for which

$$\mathbb{P}\left(\frac{1}{q_n} \sum_{k=1}^{q_n} |E^*_{t_k}| \geq \frac{\tilde{c}}{3} n^{\varepsilon} \right) \leq \frac{c}{n^4}.$$

In the same way it is possible to choose r and $\tilde{c}$ such that

$$\mathbb{P}(\Omega_B^c) \leq \frac{c}{n^4}.$$

□

References

[1] Abraham, K. Nonparametric Bayesian posterior contraction rates for scalar diffusions with high-frequency data. Bernoulli 25 (2019), no. 4A, 2696–2728.

[2] Amorino, C., Gloter, A. (2020). Contrast function estimation for the drift parameter of ergodic jump diffusion process. Scandinavian Journal of Statistics, 47(2), 279-346.

[3] Amorino, C., Gloter, A. (2020). Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes. arXiv preprint arXiv:2001.07422.

[4] Amorino, C., Gloter, A. (2020). Unbiased truncated quadratic variation for volatility estimation in jump diffusion processes. Stochastic Processes and their Applications.

[5] Ait-Sahalia, Y., Cacho-Diaz J. and Roger J.A. (2015) Modeling financial contagion using mutually exciting jump processes. Journal of Financial Economics. 117(3) 585 - 606.

[6] Bacry, E., Delattre, S., Hoffmann, M., Muzy, J.-F. (2013). Some limit theorems for hawkes processes and application to financial statistics. Stochastic Processes and their Applications 123, 24752499.

[7] Bacry, E., Mastromatteo I. and Muzy J.F. (2015) Hawkes processes in finance. Market Microstructure and Liquidity. (1) 1550005.

[8] M. T. Barlow and M. Yor. Semimartingale inequalities via the Garsia-Rodemich-Rumsey lemma, and applications to local times. J. Funct. Anal., 49(2):198229, 1982.

[9] Birgé, L. and Massart, P. (1998) Minimum contrast estimators on sieves: exponential bounds and rates of convergence. Bernoulli , 4 (3) pp. 329375.

[10] Brémaud, P. Massoulié, L. (1996). Stability of nonlinear hawkes processes. The Annals of Probability pp. 15631588.

[11] Clinet S. and Yoshida N. (2017) Statistical inference for ergodic point processes and application to limit order book. Stochastic Processes and their Applications, 127(6):1800 1839.

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

[13] Comte, F. and Marie, N. (2020) On a Nadaraya-Watson Estimator with Two Bandwidths. Preprint HAL hal-02457079.

[14] Comte, F. and Merlevède, F. (2002) Adaptive estimation of the stationary density of discrete and continuous time mixing processes. ESAIM Probab. Statist., 6 pp. 211238 (electronic). New directions in time series analysis (Luminy, 2001).

[15] Daley, D. J. Vere-Jones, D. (2007). An introduction to the theory of point processes: volume II: general theory and structure. Springer Science and Business Media.

[16] Delattre, S., Fournier, N., Hoffmann, M. (2016). Hawkes processes on large networks. The Annals of Applied Probability 26, 216261.

[17] Dion, C., Lemler, S. (2019). Nonparametric drift estimation for diffusions with jumps driven by a Hawkes process. Statistical Inference for Stochastic Processes, 1-27.

[18] Dion, C., Lemler, S., Lcherbach, E. (2019). Exponential ergodicity for diffusions with jumps driven by a Hawkes process. arXiv preprint arXiv:1904.06051.

[19] Doukhan, P. (2012). Mixing: properties and examples (Vol. 85). Springer Science and Business Media.

[20] Embrechts, P., Liniger, T. and Lu Lin. (2011) Multivariate hawkes processes: an application to financial data. Journal of Applied Probability. 48(A) 367–378.

[21] Goldenshluger, A. & Lepski., O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. The Annals of Statistics, 39(3) 16081632.

[22] Gomez-Valle, L. and Martinez-Rodriguez, J. (2016). Estimation of risk-neutral processes in single-factor jump-diffusion interest rate models. Journal of Computational and Applied Mathematics. 291, 48–57

[23] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. Stochastic processes and their Applications, 79(1), 135-163.

[24] Jacod, J., Protter, P. (2011). Discretization of processes (Vol. 67). Springer Science and Business Media.

[25] Krell, N., Schmisser, E. (2019). Nonparametric estimation of jump rates for a specific class of Piecewise Deterministic Markov Processes. arXiv preprint arXiv:1901.10166.

[26] Lacour, C., Massart, P. and Rivoirard, V. (2017). Estimator Selection: a New Method with Applications to Kernel Density Estimation. Sankhya A 79, 298335.

[27] Le Gall, J. (2010). Calcul stochastique et processus de markov. Notes de cours .

[28] Mancini, C., and Reno, R. (2011). Threshold estimation of Markov models with jumps and interest rate modeling. Journal of Econometrics, 160(1), 77-92.

[29] Neumann, M. H. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. Bernoulli, 15(1), 223-248.

[30] Schmisser, E. (2012). Nonparametric estimation of the diffusion coefficient from noisy data. Statistical inference for stochastic processes, 15(3), 193-223.

[31] Schmisser, E. (2019). Nonparametric estimation of the diffusion coefficients of a diffusion with jumps. Stochastic Processes and their Applications, 129(12), 5364-5405.

[32] Strauch, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. The Annals of Statistics, 46(6B), 3451-3480.

[33] Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. Probability theory and related fields, 107(4), 467-492.