Electronic Journal of Statistics
Vol. 16 (2022) 3212–3277
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2019

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

Chiara Amorino

Unité de Recherche en Mathématiques, Université du Luxembourg
e-mail: chiara.amorino@uni.lu

Charlotte Dion-Blanc

LPSM, Sorbonne Université 75005 Paris UMR CNRS 8001
e-mail: charlotte.dion_blanc@sorbonne-universite.fr

Arnaud Gloter

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

and

Sarah Lemler

Université Paris-Saclay, École CentraleSupélec, MICS Laboratory, France
e-mail: sarah.lemler@centralesupelec.fr

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 a long time horizon which remained an open question until now. First, we propose to estimate the volatility coefficient. For that, we introduce a truncation function in our estimation procedure that allows us to take into account the jumps of the process and 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, ensuring its consistency, and then we study an adaptive estimator w.r.t. the regularity. Then, we define an estimator of a sum between the volatility and the jump coefficient modified with the conditional expectation of the intensity of the jumps. We also establish a bound for the empirical risk for the non-adaptive estimators of this sum, the convergence rate up to the regularity of the true function, and an oracle inequality for the final adaptive estimator.

Finally, we give a methodology to recover the jump function in some applications. We conduct a simulation study to measure our estimators' accuracy in practice and discuss the possibility of recovering the jump function from our estimation procedure.

MSC2020 subject classifications: Primary 62G05, 60G55.

*C. Amorino gratefully acknowledges financial support of ERC Consolidator Grant 815703 "STAMFORD: Statistical Methods for High Dimensional Diffusions".

Keywords and phrases: Jump diffusion, Hawkes process, volatility estimation, nonparametric estimation, adaptation.

Received June 2021.

1. Introduction

The present work focuses on the jump-diffusion process introduced in [20]. 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. Some probabilistic results have been established for this model in [20], 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 [19] 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 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 an exterior event, with auto-excitation structure. For example, one can think of the interest rate model (see [26]) in insurance; then, in neurosciences of the evolution of the membrane potential impacted by the signals of the other neurons around it (see [19]). 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 specific 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 generalization of the so-called mutually-exciting-jump diffusion proposed in [5] to study an asset price evolution. This process generalizes 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 stochastic differential equations from the observation of a discrete path is a challenge studied a lot in literature. From a frequentist point of view in the high-frequency context, one can cite [27, 12] and in bayesian, one recently in [1]. Nevertheless, the purpose of this article falls more under the scope of statistics for stochastic processes with jumps. The

literature for the diffusion with jumps from a pure centered Lévy process is large. For example one can refer to [28], [34] and [36].

The first goal of this work is to estimate the volatility coefficient σ^2. As it is well known, in the presence of jumps, the approximate quadratic variation based on the squared increments of X no longer converges to the integrated volatility. As in [33], we base the approach on truncated quadratic variation to estimate the coefficient σ^2. 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 to find a way to approximate the coefficient a. It is important to note that, as presented in [36], in the 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 coefficients 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 [19], we assume that the coupled process (X, λ) is ergodic, stationary, and exponentially $\beta-$mixing. Besides, in this article we obtain that the projection on X of the invariant measure of the process has a density which is lower and upper bounded on compact sets. This property is useful to lead studies of convergence rates for nonparametric estimators since it gives equivalence between empirical and continuous norms. To estimate the volatility in a nonparametric way, as in [4] we consider a truncation of the increments of the quadratic variation of X that allows judging if a jump occurred or 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. We give the convergence rates of these estimators depending on the regularity of the true volatility function. Then, we propose a selection model procedure through a penalized criteria, we obtain non-asymptotic oracle-type inequality for the final 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 a penalized minimum contrast estimation procedure again, and we establish a non-asymptotic oracle inequality for the adaptive estimator. The achieved rates of convergence are similar to the ones obtained in the Lévy jump-diffusion context in [36]. Both adaptive estimators are studied using Talagrand's concentration inequalities.

We then discuss how we can recover a, 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 a correctly.

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 minimizes 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 properties on the process (X_t, λ_t). In Section 3 we present the adaptive estimation procedure for the volatility σ^2 and obtain the consistency and the convergence rate. 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 return to the reason for estimating this function, we detail the estimation procedure and establish bounds for the risks of the non-adaptive estimator and of the adaptive estimator in the regularity. 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 ($i, j \in \{1, \dots, M\}$). 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 [16]). 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 prove with standard arguments the existence of a process $(N_t^{(j)})_{t \geq 0}$ (see for example [17]). 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 [20], 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 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 of 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 remark that the process $N_t^{(j)}$ has jumps of size 1. 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 $[0, T_n]$. For $0 = t_0 \le t_1 \le \dots \le t_n = T_n$, 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, $\forall i \in \{0, \dots, n-1\}$, $c_1 \Delta_{\min} \le \Delta_{n,i} \le c_2 \Delta_n$. We remark we are considering a general case, where the discretization step is not necessarily uniform. However, in case where the discretization step is uniform we clearly have $\Delta_n = \Delta_{n,i}$ for any $i \in \{0, ..., n-1\}$ (which implies that the condition here above is clearly respected with $c_1 = c_2 = 1$) and the time horizon becomes $T_n = n\Delta_n$. Furthermore, we require that

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

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 (3) is more technical. This condition is replaced with a stronger one to obtain Theorem 2 (see Assumption 6 and the discussion below).

Assumption 1 (Assumptions on the coefficients of X)**.** *1. The coefficients b and σ are of class $\mathcal{C}^2$ and there exists a positive constants c such that, for all $x \in \mathbb{R}$, $|b'(x)| + |\sigma'(x)| + |a'(x)| \le c$.*

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. *There exist positive constants* c', q *such that, for all* $x \in \mathbb{R}$, $|b''(x)| + |\sigma''(x)| \leq c'(1 + |x|^q)$.
4. *There exist* $d \geq 0$ *and* $r > 0$ *such that, for all* x *satisfying* $|x| > r$, *we have* $xb(x) \leq -dx^2$.

We remark that, as a consequence of Assumption 1, item 1, the coefficients b, σ and a are globally Lipschitz continuous.

The first three assumptions ensure the existence of a unique solution X (as proven in [20] Proposition 2.3). The last assumption is introduced to study the longtime behavior of X and to ensure its ergodicity (see [20]). Note that the assumption on a can be relaxed (see [20]).

Assumption 2 (Assumptions on the kernels). 1. *Let* H *be a matrix such that* $H_{i,j} := \int_0^\infty h_{i,j}(t)dt = c_{ij}/\alpha$, *for* $1 \leq i, j \leq M$. *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.*

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 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 [19].

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 [20] 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 [20], 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 [20], we take $V : \mathbb{R} \times \mathbb{R}^{M\times M} \to \mathbb{R}_+$ 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 [20] 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 $1 \le i,l \le M$. Then, the following proposition holds true.

Proposition 1. *Suppose that Assumptions 1 and 2 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:*

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

Moreover, both X and λ have bounded moments of any order.

Proposition 1 is proven in the Appendix. Let us now add the third assumption.

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

Then, the process $(X_t,\lambda_t)_{t\ge 0}$ 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 (which exists, see Theorem 2.3 in [19] and proof in [20]). Then, according to Theorem 4.9 in [20], 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

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

Furthermore, from Proposition 3.7 in [20], we know that the measure $\pi^X(dx)$ admits a Lebesgue density $x\mapsto\pi^X(x)$ and it is lower bounded on each compact set of $\mathbb{R}$. In the following lemma we additionally prove that this density is also upper bounded on each compact set.

Lemma 1. *Assume that Assumptions 1 and 2 hold true. Then, for any compact set K of $\mathbb{R}$, there exists a constant $C_K > 0$ such that $\pi^X(x) \le C_K$ for all $x\in K$.*

Finally, we know that for each compact set $A \subset \mathbb{R}$ there exist two positive constants π_0, π_1 such that for $x \in A$ we have

$$0 < \pi_0 \leq \pi_X(x) \leq \pi_1. \tag{7}$$

Let us define the norm with respect to π_X:

$$\|t\|^2_{\pi^X} := \int_A t^2(x)\pi_X(dx).$$

According to Lemma 1 it yields that for a deterministic function t, $\pi_0\|t\|^2 \leq \|t\|_{\pi^X} \leq \pi_1\|t\|^2$.

3. Estimation procedure of the volatility function

With the background introduced in the previous sections, we are now ready to estimate 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 jumps' contribution. To make the latter small as well, we introduce a truncation function (see Lemma 3 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 the best model automatically 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 (15). In Theorem 1 we show that the estimator associated with the selected model realizes the best compromise between automatically the bias term and the penalty term.

3.1. Non-adaptive estimator

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

$$\begin{aligned} X_{t_{i+1}} - X_{t_i} &= \int_{t_i}^{t_{i+1}} b(X_s)ds + \int_{t_i}^{t_{i+1}} \sigma(X_s)dW_s + \int_{t_i}^{t_{i+1}} a(X_{s^-}) \sum_{j=1}^{M} dN_s^{(j)} \\ &= \int_{t_i}^{t_{i+1}} b(X_s)ds + Z_{t_i} + J_{t_i} \end{aligned} \tag{8}$$

where Z, J are given in Equation (9):

$$Z_{t_i} := \int_{t_i}^{t_{i+1}} \sigma(X_s) dW_s, \qquad J_{t_i} := \int_{t_i}^{t_{i+1}} a(X_{s^-}) \sum_{j=1}^{M} dN_s^{(j)}. \tag{9}$$

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 (8) yields

$$T_{t_i} = \frac{1}{\Delta_n}(X_{t_{i+1}} - X_{t_i})^2 = \sigma^2(X_{t_i}) + \bar{A}_{t_i} + B_{t_i} + E_{t_i}, \tag{10}$$

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

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

$$B_{t_i} := \frac{1}{\Delta_n}[Z_{t_i}^2 - \int_{t_i}^{t_{i+1}} \sigma^2(X_s) ds]; \tag{11}$$

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

The term $\bar{A}_{t_i}$ 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 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 $|\zeta| \leq 1$. Also, we define $\varphi_z(.) := \varphi(./z), z > 0$. The idea is to use the size of the increment of the process $\Delta_i X := X_{t_{i+1}} - X_{t_i}$ 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 $\beta < \frac{1}{2}$, 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

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

Now, the just-introduced $\tilde{A}_{t_i}$ is once again a small term, because so $\bar{A}_{t_i}$ was and because the truncation function does not differ a lot from the indicator function, as better justified in Lemma 2 below.

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

Lemma 2. *Suppose that Assumptions 1,2,3 hold. Then, for $\beta \in (0, \frac{1}{2})$ and for any $k \geq 1$,*

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

The proof of Lemma 2 can be found in the Appendix. The same is for the proof of Lemma 3 below, which illustrates the reason why we have introduced a truncation function. Indeed, without the presence of φ, the same Lemma would have held with just a $c\,\Delta_{n,i}$ on 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 3. *Suppose that Assumptions 1,2,3 hold. Then, for $q \geq 1$, for $\beta \in (0, \frac{1}{2})$ and for any $k \geq 1$*

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

From Lemmas 2 and 3 here above, it is possible to prove the following proposition. Also, its proof can be found in the Appendix.

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

In the Proposition here above, it is 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 compact 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 $\mathcal{M}_n$ ~~$\mathcal{M}_n \subset \mathbb{N}$~~ is a set of indexes for the model collection. The contrast function is defined by

$$\gamma_{n,M}(t) := \frac{1}{n}\sum_{i=0}^{n-1}(t(X_{t_i}) - T_{t_i}\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X))^2 \mathbb{1}_A(X_{t_i}) \tag{12}$$

with the T_{t_i} given in Equation (10). The associated mean squares contrast estimator is

$$\widehat{\sigma}_m^2 := \arg\min_{t \in \mathcal{S}_m} \gamma_{n,M}(t). \tag{13}$$

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 indicator function in (12) suppresses the contribution of the data falling outside the compact set A, on which we estimate the unknown function σ^2. However, this indicator function is introduced for

convenience and could be removed without affecting the value of the argmin in (13), as all elements of $\mathcal{S}_m$ are compactly supported on A. 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. *Nesting condition: $(\mathcal{S}_m)_{m\in\mathcal{M}_n}$ is a collection of models such that there exists a space denoted by $\tilde{\mathcal{S}}_n$, belonging to the collection, with $\mathcal{S}_m \subset \tilde{\mathcal{S}}_n$ for all $m \in \mathcal{M}_n$. We denote by N_n the dimension of $\tilde{\mathcal{S}}_n$. It implies that, $\forall m \in \mathcal{M}_n$, $D_m \leq N_n$.*
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.*

Several possible collections of spaces are available: we can consider the collection of dyadic regular piecewise polynomial spaces [DP], the trigonometric spaces [T] or the dyadic wavelet-generated spaces [W] (see Sections 2.2 and 2.3 of [12] for details). As we will see, in the numerical part we will consider the trigonometric spaces while all the results gathered in the theory hold true for no matter which space. However, depending on the collection we consider, we need to require a different condition on the dimension of the space, as stated in the assumption below.

Assumption 5 (Assumptions on the dimension)**.**

1. *There exists a constant $c > 0$ such that $N_n \leq c\frac{\sqrt{n\Delta_n}}{\log n}$ and $\frac{N_n^3}{n} \leq 1$ for the collection [T].*
2. *There exists a constant $c > 0$ such that $N_n \leq c\frac{n\Delta_n}{\log^2 n}$ for collections [DP] and [W].*

We now introduce the empirical norm

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

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 Assumptions 1,2,3,4,5 hold and that $\beta \in (\frac{1}{4}, \frac{1}{2})$. If $\Delta_n \to 0$, $\log n = o(\sqrt{n\Delta_n})$ for $n \to \infty$, then the estimator $\widehat{\sigma}_m^2$ of σ^2 on A given by Equation (13) satisfies*

$$\mathbb{E}\left[\|\widehat{\sigma}_m^2 - \sigma^2\|_n^2\right] \leq 3 \inf_{t\in\mathcal{S}_m} \|t - \sigma^2\|_{\pi^X}^2 + \frac{C_1\sigma_1^4 D_m}{n} + C_2\Delta_n^{4\beta-1} + \frac{C_3\Delta_n^{0\wedge(4\beta-\frac{3}{2})}}{n^2}, \tag{14}$$

with C_1, C_2 and C_3 positive constants.

This result measures the accuracy of our estimator $\widehat{\sigma}_m^2$ for the empirical norm. The right-hand side of the Equation (14) is decomposed into different types of

TABLE 1
Rates of convergence for $\hat{\sigma}_m^2$

	Hawkes-diffusion	Diffusion
$0 < \gamma \leq \frac{2\alpha}{2(2\alpha+1)} \leq 1/2$	${\Delta_n}^{2\beta-1/2}$	Δ_n
$\frac{2\alpha}{2(2\alpha+1)} \leq \gamma \leq \left(\frac{2\alpha}{(2\alpha+1)(4\beta-1)}\right) \wedge 1$	${\Delta_n}^{2\beta-1/2}$	$n^{-\alpha/(2\alpha+1)}$
$\left(\frac{2\alpha}{(2\alpha+1)(4\beta-1)}\right) \wedge 1 \leq \gamma < 1$	$n^{-\alpha/(2\alpha+1)}$	$n^{-\alpha/(2\alpha+1)}$

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 $\|\widehat{\sigma}_m^2 - \sigma^2\|_n^2$ when $\|.\|_n$ and $\|.\|_{\pi^X}$ are not equivalent. This inequality ensures that our estimator $\widehat{\sigma}_m^2$ 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 remaining terms are larger because of the presence of the jumps.

Rates of convergence Let us remind that $\|t - \sigma^2\|_{\pi^X}^2 \leq \pi_1 \|t - \sigma_{|A}^2\|^2$ according to Lemma 1. Thus let us consider $t = \sigma_m^2$ the projection of σ^2 on $\mathcal{S}_m$ for the L^2-norm which realizes the minimum. We assume now that the function of interest $\sigma_{|A}^2$ is in a Besov space $\mathcal{B}_{2,\infty}^{\alpha}$ with regularity $\alpha > 0$ (see *e.g.* [18] for a proper definition). Then it comes that $\|\sigma_m^2 - \sigma_{|A}^2\|^2 \leq C(\alpha) D_m^{-2\alpha}$. Finally, choosing $D_{m_{\text{opt}}} = n^{1/(2\alpha+1)}$, and $\Delta_n = n^{-\gamma}$ with $0 < \gamma < 1$, we obtain the rates of convergences given in Table 1. This table allows to compare the actual rates with the one obtained when $a \equiv 1$ and the process is a simple diffusion process.

Since $\beta \in (1/4, 1/2)$, the best choice for β is to choose it close to $1/4$. In this case, the estimator reaches the classical nonparametric convergence rate for high-frequency observations (${n\Delta_n}^{(1+2\alpha)/2\alpha} = O(1)$). Otherwise, most of the time the remainder term will be predominant in the risk. This result is analogous to the jump-diffusion case studied in [36].

We observe that, after having replaced the optimal choice for D_m, which corresponds to $D_{m_{\text{opt}}} = n^{1/(2\alpha+1)}$, the conditions gathered in Assumption 5 become $\gamma \leq \frac{2\alpha-1}{1+2\alpha}$ for collection [T] and $\gamma \leq \frac{2\alpha}{1+2\alpha}$ for collections [DP] and [W].

It is important also to remark that in order to have negligible reminder terms, we want γ to be such that $\gamma \geq \frac{2\alpha}{(2\alpha+1)(4\beta-1)}$. As $\beta \in (\frac{1}{4}, \frac{1}{2})$, the best possible case is to have β in the neighbourhood of $1/2$, so that the right hand side of the inequality here above becomes $\frac{2\alpha}{1+2\alpha}$. It means that, considering the collections [DP] and [W] are respected at the same time up to have a discretization step $\Delta_n = (\frac{1}{n})^{\frac{2\alpha}{1+2\alpha}}$ while considering the collection [T] the two conditions can not be respected at the same time and the only possibility is to take α big. Indeed, if the function of interest is in a Besov space $\mathcal{B}_{2,\infty}^{\infty}$, then the condition in Assumption 5 is respected and the reminder terms are negligible for $\gamma = 1$, which means

$\Delta_n = 1/n$.

This implies that the term $\Delta_n^{4\beta-1}$ always dominates the others. As it is easy to see comparing it with the third point of Proposition 2, it comes from the contribution of the jumps. However, the bound obtained in Proposition 2 on the filtered jumps is optimal. Hence, in order to have negligible reminder terms, the idea is to try to lighten the condition on N_n gathered in Assumption 5, rather than trying to improve the bound on the contribution of the jumps.

3.2. Adaption procedure

We want to 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:

$$\widehat{m}_\sigma := \arg \min_{m \in \mathcal{M}_n} \{\gamma_{n,M}(\widehat{\sigma}_m^2) + \text{pen}_\sigma(m)\}, \tag{15}$$

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

$$\text{pen}_\sigma(m) := \kappa_1 \frac{D_m}{n} \tag{16}$$

where $\kappa_1 > 0$ is a constant which has to be calibrated. Next theorem is proven in Section 8.1.

Theorem 1. *Suppose that Assumptions 1,2,3,4 hold and that* $\beta \in (\frac{1}{4}, \frac{1}{2})$. *If* $\Delta_n \to 0$ *and* $\log n = o(\sqrt{n\Delta_n})$ *for* $n \to \infty$, *then the estimator* $\widehat{\sigma}^2_{\widehat{m}_\sigma}$ *of* σ^2 *on* A *given by equations* (13) *and* (15) *satisfies*

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

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 $\widehat{\sigma}^2_{\widehat{m}_\sigma}$ realizes the best compromise between the bias term and the penalty term, which is of the same order as the variance term. Indeed, it achieves the rates given in Table 1 automatically, without the knowledge of the regularity of the function σ^2.

The convergence of this adaptive estimator is studied in Section 6.

4. Estimation procedure for both coefficients

In addition to the volatility estimation, our goal is to propose a procedure to recover in a non-parametric 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 approach the estimation of 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 leads to some difficulties as we will see in the sequel. 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 the squared increments of the process X differently to highlight the role of the conditional expectation. In the following, we use for the decomposition of the squared increments, 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:

$$\begin{cases} d\lambda_t^{(j)} = -\alpha(\lambda_t^{(j)} - \zeta_t)dt + \sum_{i=1}^M c_{j,i} dN_t^{(i)} \\ dX_t = (b(X_t) + a(X_{t^-}) \sum_{i=1}^M \lambda_t^{(i)})dt + \sigma(X_t)dW_t + a(X_{t^-}) \sum_{i=1}^M d\tilde{N}_t^{(i)}. \end{cases} \tag{17}$$

We set now

$$J_{t_i} := \int_{t_i}^{t_{i+1}} a(X_{s^-}) \sum_{i=1}^M d\tilde{N}_s^{(i)}. \tag{18}$$

The increments of the process X are such that

$$X_{t_{i+1}} - X_{t_i} = \int_{t_i}^{t_{i+1}} \left(b(X_s) + a(X_{s^-}) \sum_{j=1}^M \lambda_s^{(j)} \right) ds + Z_{t_i} + J_{t_i} \tag{19}$$

where J is given in Equation (18) and Z has not changed and is given in Equation (9). Let us define this time:

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

$$+\frac{1}{\Delta_n}\int_{t_i}^{t_{i+1}}(a^2(X_s)-a^2(X_{t_i}))\sum_{j=1}^{M}\lambda_s^{(j)}ds$$
$$+\frac{a^2(X_{t_i})}{\Delta_n}\int_{t_i}^{t_{i+1}}\sum_{j=1}^{M}(\lambda_s^{(j)}-\lambda_{t_i}^{(j)})ds+2\left(b(X_{t_i})+a(X_{t_i})\sum_{j=1}^{M}\lambda_{t_i}^{(j)}\right)Z_{t_i}$$
$$+2\left(b(X_{t_i})+a(X_{t_i})\sum_{j=1}^{M}\lambda_{t_i}^{(j)}\right)J_{t_i}, \tag{20}$$

$$E_{t_i}:=\frac{2}{\Delta_n}Z_{t_i}J_{t_i}+\frac{1}{\Delta_n}\left(J_{t_i}^2-\int_{t_i}^{t_{i+1}}a^2(X_s)\sum_{j=1}^{M}\lambda_s^{(j)}ds\right). \tag{21}$$

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 introduce the quantity

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

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

$$C_{t_i}:=a^2(X_{t_i})\sum_{j=1}^{M}(\lambda_{t_i}^{(j)}-\mathbb{E}[\lambda_{t_i}^{(j)}|X_{t_i}]). \tag{22}$$

It comes the following decomposition:

$$T_{t_i}=\frac{1}{\Delta_n}(X_{t_{i+1}}-X_{t_i})^2=\sigma^2(X_{t_i})+a^2(X_{t_i})\sum_{j=1}^{M}\mathbb{E}[\lambda_{t_i}^{(j)}|X_{t_i}]+A_{t_i}+B_{t_i}+C_{t_i}+E_{t_i}. \tag{23}$$

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 evaluated. 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. Here it is important to remark that the natural choice would have been to estimate directly $\sigma^2(X_{t_i})+a^2(X_{t_i})\sum_{j=1}^{M}\lambda_{t_i}$, rather than $\sigma^2(X_{t_i})+a^2(X_{t_i})\sum_{j=1}^{M}\mathbb{E}[\lambda_{t_i}^{(j)}|X_{t_i}]$. The problem is that λ_{t_i} has its own randomness and is not observed and so the method used before for the estimation of the volatility coefficient can not work anymore. However, replacing $\lambda_{t_i}^{(j)}$ with $\mathbb{E}[\lambda_{t_i}^{(j)}|X_{t_i}]$, our goal turns out being the estimation of g where

$g(X_{t_i}) := \sigma^2(X_{t_i}) + a^2(X_{t_i}) \sum_{j=1}^M \mathbb{E}[\lambda_{t_i}^{(j)} | X_{t_i}]$, is now a function of the observation X_{t_i} and so its non-parametric estimation can be accomplished using the same method as in previous section (see details below).

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 Assumptions 1,2,3 hold. Then,*

1. $\forall \tilde{\varepsilon} > 0$, $\mathbb{E}[A_{t_i}^2] \le c\Delta_{n,i}^{1-\tilde{\varepsilon}}$, $\mathbb{E}[A_{t_i}^4] \le c\Delta_{n,i}^{1-\tilde{\varepsilon}}$;
2. $\mathbb{E}[B_{t_i} | \mathcal{F}_{t_i}] = 0$, $\mathbb{E}[B_{t_i}^2 | \mathcal{F}_{t_i}] \le c\sigma_1^4$, $\mathbb{E}[B_{t_i}^4] \le c$;
3. $\mathbb{E}[E_{t_i} | \mathcal{F}_{t_i}] = 0$, $\mathbb{E}[E_{t_i}^2 | \mathcal{F}_{t_i}] \le \frac{c a_1^4}{\Delta_{n,i}} \sum_{j=1}^M \lambda_{t_i}^{(j)}$, $\mathbb{E}[E_{t_i}^4] \le \frac{c}{\Delta_{n,i}^3}$;
4. $\mathbb{E}[C_{t_i} | X_{t_i}] = 0$, $\mathbb{E}[C_{t_i}^2] \le c$, $\mathbb{E}[C_{t_i}^4] \le c$.

From Proposition 4 one can see in detail how small the bias term A_{t_i} is. Moreover, it sheds light on 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

$$g(x) := \sigma^2(x) + a^2(x) f(x) \tag{24}$$

with

$$f(x) = \frac{\sum_{j=1}^M \int_{\mathbb{R}^M} z_j \pi(x, z_1, \dots, z_M) dz_1, \dots, dz_M}{\int_{\mathbb{R}^M} \pi(x, z_1, \dots, z_M) dz_1, \dots, dz_M} \tag{25}$$

on a closed interval A. One can see that the estimation of g is a natural way to approach the problem of the estimation of the jump coefficient. The same idea can be found for example in [36], where a Lévy-driven stochastic differential equation is considered. The reason why in the above mentioned work the density does not play any role is that it is assumed to be one.

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 $t \in \tilde{S}_n$,

$$\gamma_{n,M}(t) := \frac{1}{n} \sum_{i=0}^{n-1} (t(X_{t_i}) - T_{t_i})^2 \mathbb{1}_A(X_{t_i}) \tag{26}$$

and the T_{t_i} are given in Equation (23) this time. The associated mean squares contrast estimator is

$$\widehat{g}_m := \arg\min_{t \in S_m} \gamma_{n,M}(t). \tag{27}$$

We want to bound the empirical risk $\mathbb{E}[\|\widehat{g}_m - g\|_n^2]$ on the compact A. It is the object of the next result, whose proof can be found in Section 8.2.

Proposition 5. *Suppose that Assumptions 1,2,3,4,5 hold. If* $\Delta_n \to 0$ *and* $\log n = o(\sqrt{n\Delta_n})$*, then the estimator* $\widehat{g}_m$ *of* g *on* A *satisfies, for any* $0 < \tilde{\varepsilon}, \varepsilon < 1$*,*

$$\mathbb{E}\left[\|\widehat{g}_m - g\|_n^2\right] \leq 3 \inf_{t \in \mathcal{S}_m} \|t - g\|_{\pi,x}^2 + \frac{C_1(\sigma_1^4 + a_1^4 + 1)D_m^{1+2\varepsilon}}{n\Delta_n} + C_2\Delta_n^{1-\tilde{\varepsilon}} + \frac{C_3}{n^2\Delta_n^{\frac{3}{2}}}, \tag{28}$$

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

As in the previous section, this inequality measures the performance of our estimator $\widehat{g}_m$ for the empirical norm and the comments given after Proposition 3 hold. The main difference between the proof of Proposition 3 and the proof of Proposition 5 is that, in the first case, we deal with the jumps by introducing the indicator function φ. In this way, the jump part is small and some rough estimations are enough to get rid of them (see point 3 of Proposition 2). From Proposition 4 we can see that for the estimation of both coefficients, instead, the jump contribution (gathered in C_{t_i} and E_{t_i}) is no longer small. However, C_{t_i} and E_{t_i} are both centered (even if with respect to different σ algebras) and we can therefore apply on them the same reasoning as we did for B_{t_i}, which consists in a more detailed analysis. Hence, proving Proposition 5 is more challenging than proving Proposition 3. Evidence of this is for example the fact that, to estimate g, a bound on the variance of C_{t_i} relying on mixing properties is required (see Lemma 4).

Finally, let us compare this result with the bound (14) obtained for the estimator $\widehat{\sigma}_m^2$. The main difference is that, up to a term $D_m^{2\varepsilon}$ for ε arbitrarily small, the second term is of order $D_m/(n\Delta)$ here, instead of D_m/n as it was previously. Consequently, in practice, the risks will depend mainly on $n\Delta$ for the estimation of g and n for the estimation of σ^2.

It is worth remarking that the reason why this extra term $D_m^{2\varepsilon}$ appears relies on the use of the β-mixing property of the process (see Lemma 4). However, in the intensity is of the jump process is constant (Poisson process) or in the case where the process satisfies some stronger mixing properties (such as the ρ-mixing, for example for diffusion processes, see [23]), it is possible to improve the result gathered in Proposition 5 and to lose the $D_m^{2\varepsilon}$.

Rates of convergence Assume now that $g_{|A} \in \mathcal{B}_{\alpha,\infty}^2$ with $\alpha \geq 1$, then, taking $t = g_m^2$ (the projection of g on $\mathcal{S}_m$) produces $\|g_m - g_{|A}\|^2 \leq C(\alpha)D_m^{-2\alpha}$. Choosing $D_{m_{\text{opt}}} = (n\Delta)^{1/(2\alpha+2\varepsilon+1)}$, if $n\Delta^{2-\tilde{\varepsilon}} \to 0$ leads to

$$\mathbb{E}\left[\left\|\widehat{g}_{m_{\text{opt}}} - g\right\|_n^2\right] \leq (n\Delta)^{-2\alpha/(2\alpha+2\varepsilon+1)}.$$

We obtain, up to an ε arbitrarily small, the same rate of convergence for the regularity α as [36] for the estimation of $\sigma^2 + a^2$ in the case of a jump diffusion with a Lévy process instead of the Hawkes process.

In practice the regularity of g is unknown and thus it is necessary to choose the best model in a data driven way. This it the subject of the next paragraph.

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:

$$\widehat{m}_g := \arg \min_{m \in \mathcal{M}_n} \{\gamma_{n,M}(\widehat{g}_m) + \text{pen}_g(m)\}, \tag{29}$$

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

$$\text{pen}_g(m) := \kappa_2 \frac{D_m^{1+2\varepsilon}}{n\Delta_n}, \tag{30}$$

where κ_2 is a constant which has to be calibrated and ε is arbitrarily small.

To establish an oracle-type inequality for the adaptive estimator $\widehat{g}_{\widehat{m}}$, the following further assumption on the discretizion step is essential.

Assumption 6. *There exists $\epsilon > 0$ such that $n^\epsilon \log(n) = o(\sqrt{n\Delta})$ for $n \to \infty$.*

One can be interested in the reason why this condition, stronger than (3) we previously required, is needed. Note that (3) is also the condition required in the discretization scheme proposed in [19] and used in [12] for the nonparametric estimation of coefficients for diffusions. As already said, the proof of the adaptive procedure involves the application of Talagrand inequality. To apply Talagrand, we need to get independent, bounded random variables through Berbee's coupling method and truncation. Intuitively the point is that, in general, such variables are built starting from the Brownian part only (in particular from B_{t_i}, as in (11)). For the adaptive estimation g, instead, also the jumps are involved (Talagrand variables depend on $B_{t_i} + C_{t_i} + E_{t_i}$, with C_{t_i} and E_{t_i} as in (21) and (22), respectively). Then, an extra term n^ϵ appears naturally looking for a bound for the jump part (see Lemma 7 and its proof), which results in the final stronger condition gathered in Assumption 6.

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

Theorem 2. *Suppose that Assumptions 1,2,3,4,5,6 hold. If $\Delta_n \to 0$, then the estimator $\widehat{g}_{\widehat{m}_g}$ of g on A satisfies, for any $1 < \tilde{\varepsilon} < 0$,*

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

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 result guarantees that our final data-driven estimator $\widehat{g}_{\widehat{m}}$ realizes automatically the best compromise between the bias term and the penalty term, and thus reaches the same rate obtained when the regularity of the true function g is known. Note that here 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. A strategy to approach the jump coefficient

The challenge is to get an estimator of the coefficient $a(\cdot)$. Let us first remind the reader of the notation $f(x) := \sum_{j=1}^{M} \mathbb{E}[\lambda_{t_i}^{(j)} | X_{t_i} = x]$ (see Equation (25)) and

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

Thus, a natural idea is to replace f in the previous equation by an estimator, and then, to study an estimator of $a(\cdot)$ of the form $\frac{\widehat{g}(x) - \widehat{\sigma}^2(x)}{\widehat{f}(x)}$. This is not a simple issue and let us discuss the estimation of f later in the section.

Assuming that an estimator of f is known and denoted $\widehat{f}_h$ where $h > 0$ denotes a tunning parameter.

Then, we also assume that $f > f_0$ on A. We then define:

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

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

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

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:

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

And by Markov's inequality, we obtain:

$$\mathbb{E}[\|\widehat{a}_z^2 - a^2\|_n^2] \leq \frac{12}{f_0^2}\left(\mathbb{E}[\|\widehat{g}_{m_2} - g\|_n^2] + \mathbb{E}[\|\widehat{\sigma}_{m_1}^2 - \sigma^2\|_n^2] + 2a_1^4\mathbb{E}\left[\left\|f - \widehat{f}_h\right\|_n^2\right]\right). \quad (31)$$

This equation teaches us that the empirical risk of the estimator $\widehat{a}_z$ is upper bounded 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.

Finally, the triplet of parameters $z = (m_1, m_2, h)$ must be chosen in a collection. A first way to do it is to use the model selection proposed in the paper for $\widehat{\sigma}_{m_1}^2, \widehat{g}_{m_2}$ and then select h through cross-validation for example, obtaining finally $\widehat{z} = (\widehat{m}_\sigma, \widehat{m}_g, \widehat{h})$. Another possible way would consist in defining a new selection procedure for the triplet, for example a Goldenshluger-Lepski type, as it is proposed in [14]. Nevertheless, the authors mention that it seems numerically less performing than to use the one-bandwidth leave-one-out cross validation method in the Nadaraya-watson estimator context.

Then, we choose to study the first methodology in the next Section, because it is directly implementable using the built adaptive estimators of σ^2 and g. Besides, the risk bound obtained on $\widehat{a}_z$ in Equation (31) suggests that the better the three functions σ^2, g, f are estimated, the better the estimation of a will be.

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 $\widehat{a}_z^2$ to be an estimator, f_0 must be known or estimated.*

Estimation of f For sake of simplicity let us assume that $M = 1$. We have that $f(x_{t_k}) = \mathbb{E}[\lambda_{t_k}|X_{t_k} = x_{t_k}]$ for all k. Thus, f depends on the conditional intensity λ; thus, the estimator of f will also depend on the estimator of λ. In addition, to estimate λ, we need more data: we already observe the process X on discrete times, but we also need to observe the jump times (which are not assumed to be known in the above).

Let us assume that we have at our disposal in addition to the $(X_{t_i})_i$'s the sequence T_j's of jump times. Now, as the Hawkes process is assumed to have exponential kernel, this estimation can simply be done using likelihood contrast estimator for example, and we denote $\widehat{\lambda}_t$ the estimator of the intensity process at time t (which do not depend on X). Then, function f can be estimated through a Nadaraya-Watson type estimator defined as

$$\widehat{f}_h^{NW}(x) = \sum_{k=1}^{n} \frac{K_h(x - X_{t_k})}{\sum_{i=1}^{n} K_h(x - X_{t_i})}\widehat{\lambda}_{t_k}.$$

The parameter, h can be chosen using cross-validation for simplicity. Under strong assumptions, the risk of $\widehat{f}_h^{NW}$ is bounded by the risk of the numerator and by the risk of the denominator. Nevertheless here, to get a bound for the

risk of the estimator we need to get a bound for the numerator which does not seem to be a direct computation. This is not solved in the present paper and will be the object of further considerations.

6. Numerical experiments

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 kinds of data, but what is impacting the dynamic of X is the cumulative Hawkes process, thus in that sense we do not lose 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 $a \neq 0$), it is not possible to use classical more sophisticated scheme to the best of our knowledge. A simulation algorithm is also detailed in [20] Section 2.3.

To challenge the proposed methodology, we investigate different kinds 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 to satisfy the assumptions and as it is not of interest to study the estimation of b here, keeping a simple drift coefficient, let us focus on the differences observed due to the coefficients σ and 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 $\widehat{\sigma}^2_{\widehat{m}_\sigma}$ and $\widehat{g}_{\widehat{m}_g}$

of σ^2 and g respectively defined by (13) and (27), 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 standard increments of a continuous diffusion process. Precisely, we choose

$$T^{\varphi}_{k\Delta} := T_{k\Delta} \times \varphi\left(\frac{X_{(k+1)\Delta} - X_{k\Delta}}{\Delta^{\beta}}\right); \quad \varphi(x) = \begin{cases} 1 & |x| < 1 \\ e^{1/3+1/(|x|^2-4)} & \\ 0 & |x| \geq 2 \end{cases}. \tag{32}$$

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 $\tilde{g}(x) = \sigma^2(x) + a^2(x) \times \widehat{f}^{NW}_{\widehat{h}}(x)$ 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 $\widehat{f}^{NW}_h(x)$, 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.

Choice of the subspaces of $L^2(A)$ The spaces $\mathcal{S}_m$ are generated by the trigonometric basis. The maximal dimension N_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} are given in Equation (16) and pen_g given in Equation (30). We consider here the limit scenario where $\varepsilon = 0$ and the penalties are both linear in the dimension. They depend on constants named κ_1, κ_2. These constants need to be chosen once for all for each estimator in order to compute the final adaptive estimators $\widehat{\sigma}^2_{\widehat{m}_{\sigma}}$ and $\widehat{g}_{\widehat{m}_g}$. We explain now how these choices are made.

Choice for the universal constants In order to choose 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 $\mathbb{E}[\|\widehat{\sigma}^2_{\widehat{m}_\sigma} - \sigma^2\|_n^2]$ and $\mathbb{E}[\|\widehat{g}^2_{\widehat{m}_g} - \tilde{g}\|_n^2]$. We choose to do $N_{\text{rep}} = 1000$ repetitions to estimate this expectation by the average:

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

Finally, comparing the risks as functions of κ_1, κ_2 leads to select values making a good compromise overall 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 (32). 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 on the estimation of $\tilde{g}$, 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 [25] or a Penalized Comparison to Overfitting [31].

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 2 and 3, we present for the three models and different values of (Δ, n): 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 $\mathcal{M}_n = \{1, \dots, N_n\}$ 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 the risk results of $\widehat{\sigma}^2_{\widehat{m}_\sigma}$, $\widehat{g}^2_{\widehat{m}_g}$ 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 $\widehat{\sigma}^2_{\widehat{m}_\sigma}$ 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 2 sums up the results of the estimator $\widehat{\sigma}^2_{\widehat{m}_\sigma}$ 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.

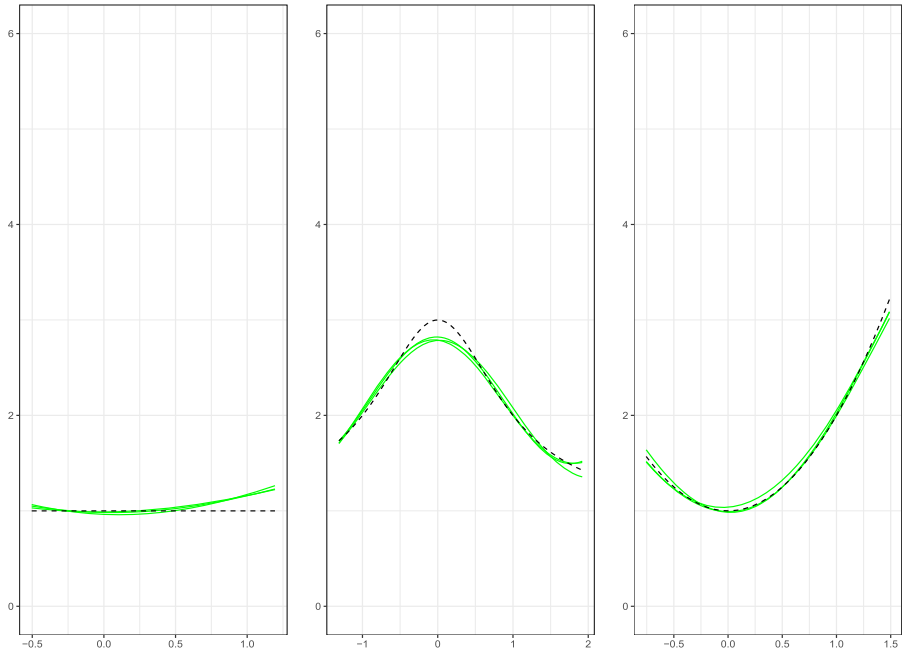

FIG 1. *Models (a),(b),(c) with* $n = 10000$, $\Delta = 0.01$. *Three final estimators are plain green (plain line), true* σ^2 *plain black (dotted line)*

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

Δ, n	Estimator	Model (a)	Model (b)	Model (c)
$\Delta = 0.1$ $n = 1000$	$\widehat{\sigma}_{m^*}$	0.361 (0.285)	0.107 (0.989)	0.798 (0.208)
	$\widehat{\sigma}_{\widehat{m}_\sigma}$	0.410 (0.280)	0.187 (1.678)	1.201 (0.216)
$\Delta = 0.1$ $n = 10000$	$\widehat{\sigma}_{m^*}$	0.278 (0.088)	0.027 (1.014)	0.366 (0.042)
	$\widehat{\sigma}_{\widehat{m}_\sigma}$	0.385 (0.122)	0.046 (1.162)	0.452 (0.062)
$\Delta = 0.01$ $n = 10000$	$\widehat{\sigma}_{m^*}$	0.010 (0.023)	0.005 (0.008)	0.008 (0.07)
	$\widehat{\sigma}_{\widehat{m}_\sigma}$	0.015 (0.028)	0.005 (0.015)	0.015 (0.012)

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 MISEs are divided at least by 10. 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 $\widehat{g}_{\widehat{m}_g}$ of $\tilde{g}$ in green (light gray) and function $\tilde{g}$ in black (dotted line). The beams of the three realizations of the estimator are satisfying.

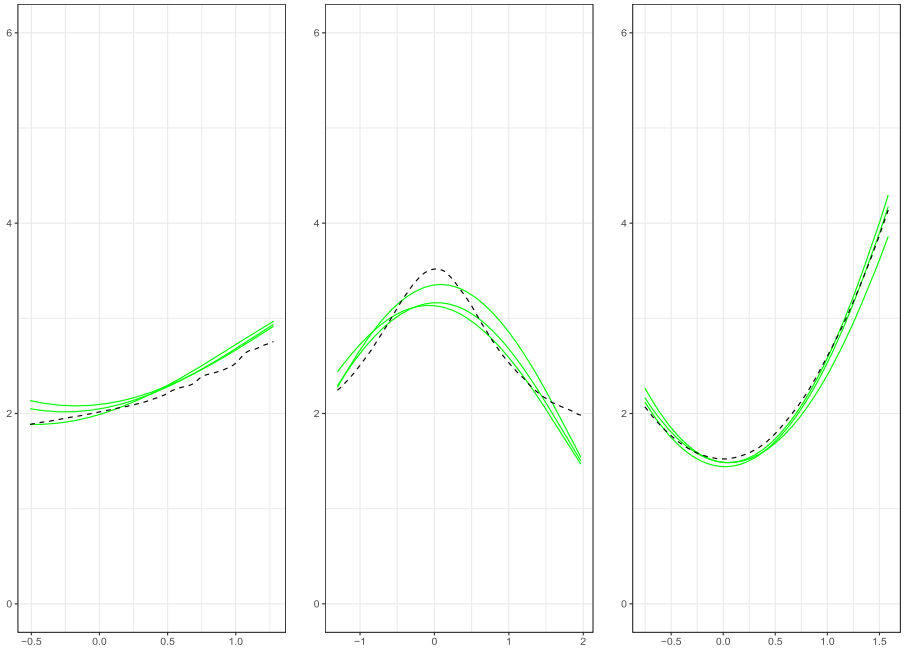

FIG 2. *Models (a),(b),(c) with* $n = 10000$, $\Delta = 0.01$. *Three final estimators of* $\tilde{g}$ *are plain green (plain line) and* $\tilde{g}$ *plain black (dotted line).*

TABLE 3

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

Δ, n	Estimator	Model (a)	Model (b)	Model (c)
$\Delta = 0.1$ $n = 1000$	$\widehat{g}_{m^*}$	0.895 (0.060)	0.474 (0.393)	0.311 (0.320)
	$\widehat{g}_{\widehat{m}_g}$	1.363 (0.715)	0.915 (0.520)	0.707 (0.964)
$\Delta = 0.1$ $n = 10000$	$\widehat{g}_{m^*}$	0.735 (0.195)	0.198 (0.079)	0.099 (0.056)
	$\widehat{g}_{\widehat{m}_g}$	0.948 (0.193)	0.313 (0.174)	0.236 (0.202)
$\Delta = 0.01$ $n = 10000$	$\widehat{g}_{m^*}$	0.109 (0.120)	0.098 (0.072)	0.035 (0.035)
	$\widehat{g}_{\widehat{m}_g}$	0.129 (0.141)	0.240 (0.100)	0.073 (0.130)

We observe that the procedure has difficulties in Model (a), and we confirm that impression in Table 3 below with the estimation of the risk. But for the two other models, the estimators seem closer to the true function. The estimation appears to work better in Model (c) than in Model (b), and this is also corroborated by the estimation of the risk given in Table 3.

Table 3 gives the Mean Integrated Squared Errors (MISEs) of the estimator $\widehat{g}_{\widehat{m}_g}$ 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 seem to be better. Compared with the results on the estimation of σ^2, the variance is proportional to $1/(n\Delta)$, and thus, the risks are greater in general.

6.5. Estimation of a^2

As explained in 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 first compute $\widehat{g}_{\widehat{m}_g}$ the estimator of $\tilde{g}$ on all the observations. Secondly, we compute $\widehat{\sigma}^2_{\widehat{m}_\sigma}$ 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

$$\widehat{a}^2(x) = \frac{\widehat{g}_{\widehat{m}_g}(x) - \widehat{\sigma}^2_{\widehat{m}_\sigma}(x)}{\widehat{f}^{NW}_{\widehat{h}}(x)}.$$

This procedure is presented in Section 5. We have plugged-in $\widehat{a}^2$ the final estimators of σ^2, g.

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 estimating properly g, σ^2 and a^2, when one of the diffusion jump process parameters 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 (31). 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 consider 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.

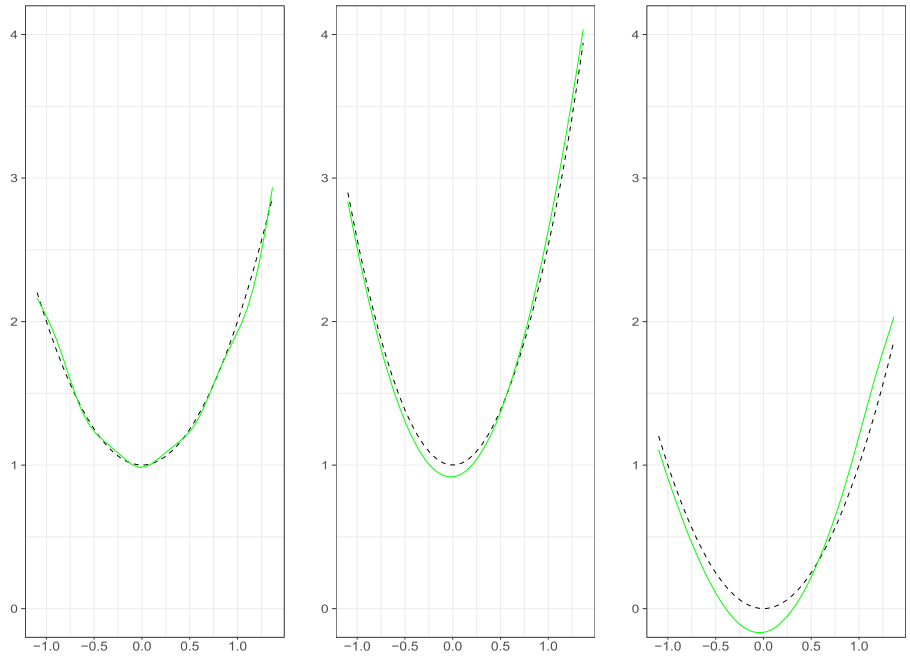

FIG 3. *Model (d). Final estimators* $\widehat{g}_{\widehat{m}_2}$, $\widehat{\sigma}^2_{\widehat{m}_1}$ *and* $\widehat{a}$ *are plain green (plain line), and true parameters* $\tilde{g}$, σ^2 *and* a^2 *in plain black (dotted line) from left to right respectively.*

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 estimating 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 can be 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 $\widehat{\sigma}^2_{\widehat{m}_\sigma}$ and $\widehat{g}_{\widehat{m}_g}$ for which the theoretical and numerical results are convincing. The last section of this article answered this question partially. 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 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 remains 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 ensure the proper

definition of the estimator. This could be theoretically and numerically carefully studied and be the object for further works.

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 study its behavior and its asymptotic properties deeply, following the same approach as in [37] and [3]. A projection method is instead considered in [30] to estimate the invariant density associated with a piecewise deterministic Markov process. Consequently, it will be possible to discuss the properties of the related estimator of f.

From the nonparametric estimation point of view, it should be interesting to extend the present estimation work to estimation on the real line instead of on a compact interval. [13] brings a solution to deal with the estimation of the drift function on the all real line from repeated observations. The procedure may be extended to the present framework in future works.

8. Proofs

This section is devoted to the proofs of the results stated in Sections 3 and 4.

One may observe that, concerning non-adaptive estimators, the proof of both Propositions 3 and 5 relies on the same scheme. It consists in introducing the set Ω_n as in (35), on which the norms $\|\cdot\|_{\pi,x}$ and $\|\cdot\|_n$ are equivalent, and to bound the risk on Ω_n and Ω_n^c, respectively. On Ω_n^c, a rough bound on the quantities we are considering is enough, as the probability of Ω_n^c is very small (see (40)). Hence, the idea to bound the risk on Ω_n^c in Proposition 3 and 5 is basically the same. On Ω_n, instead, there are main differences. Indeed, in Proposition 3, it is enough to upper bound roughly both the bias and the jump terms, (to deal more in detail only with the Brownian part), while in Proposition 5 a in-depth study is required for B_{t_i}, C_{t_i} and E_{t_i}. Such difference between the proofs for the estimation of σ and g is more highlighted in the analysis of the adaptive procedure. The proof of both Theorems 1 and 2, indeed, heavily relies on Talagrand inequality and, as for the non-adaptive procedure, for the estimation of σ what really matters is the contribution of B_{t_i}, while for the estimation of g also C_{t_i} and E_{t_i} are involved. It implies that, for the proof of Theorem 2, we are using Berbee's coupling method to get independent variables and truncation to make them bounded, starting from some variables in which also the jumps contribute; which is challenging.

8.1. Proof of volatility estimation

Here we prove all the results stated in Section 3. We start proving Proposition 3.

8.1.1. Proof of Proposition 3

Proof. We want to obtain 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 $\mathbb{E}[\|t\|^2_n] =$

$\|t\|^2_{\pi^X}$.

By the definition of T_{t_i} we have that

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

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

$$\begin{aligned}&\big\|\widehat{\sigma}^2_m - \sigma^2\big\|_n^2 \le \big\|\sigma^2_m - \sigma^2\big\|_n^2 \\ &+ \frac{2}{n}\sum_{i=0}^{n-1}(\tilde{A}_{t_i} + B_{t_i} + E_{t_i}\varphi_{\Delta^\beta_{n,i}}(\Delta_i X))(\widehat{\sigma}^2_m(X_{t_i}) - \sigma^2_m(X_{t_i})),\end{aligned}$$

where in the last sum we can remove the indicator since $\widehat{\sigma}_m$ and σ_m are compactly supported on A. Let us denote the contrast function

$$\nu_n(t) := \frac{1}{n}\sum_{i=0}^{n-1} B_{t_i} t(X_{t_i}). \tag{33}$$

In the sequel, we will repeatedly use that, for $d>0$, it is $2xy \le x^2/d + dy^2$. It follows

$$\begin{aligned}\big\|\widehat{\sigma}^2_m - \sigma^2\big\|_n^2 \;\le\; &\big\|\sigma^2_m - \sigma^2\big\|_n^2 + \frac{d}{n}\sum_{i=0}^{n-1}\Big(\tilde{A}_{t_i} + E_{t_i}\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)\Big)^2 \\ &+\frac{1}{d}\big\|\widehat{\sigma}^2_m - \sigma^2_m\big\|_n^2 + 2\nu_n(\sigma^2_m - \widehat{\sigma}^2_m).\end{aligned}$$

The linearity of the function ν_n in t implies that

$$\begin{aligned}&|2\nu_n(\widehat{\sigma}^2_m - \sigma^2_m)| = 2\|\widehat{\sigma}^2_m - \sigma^2_m\|_{\pi^X}|\nu_n((\widehat{\sigma}^2_m - \sigma^2_m)/\|\widehat{\sigma}^2_m - \sigma^2_m\|_{\pi^X})| \\ &\le 2\|\widehat{\sigma}^2_m - \sigma^2_m\|_{\pi^X}\sup_{t\in\mathcal{B}_m}|\nu_n(t)|,\end{aligned}$$

then, using again that $2xy \le \frac{x^2}{d} + dy^2$, we obtain the upper bound

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

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

$$\begin{aligned} \left\|\widehat{\sigma}^2_m - \sigma^2\right\|^2_n &\leq \left\|\sigma^2_m - \sigma^2\right\|^2_n + \frac{2d}{n}\sum_{i=0}^{n-1} \tilde{A}^2_{t_i} + \frac{2d}{n}\sum_{i=0}^{n-1} E^2_{t_i} \varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X) \\ &\quad + \frac{1}{d}\left\|\widehat{\sigma}^2_m - \sigma^2_m\right\|^2_n + d \sup_{\mathcal{B}_m} \nu^2_n(t) + \frac{1}{d}\left\|\widehat{\sigma}^2_m - \sigma^2_m\right\|^2_{\pi^X}. \quad (34) \end{aligned}$$

Let us set

$$\Omega_n := \left\{ \omega, \forall t \in \tilde{S}_n \backslash \{0\}, \left| \frac{\|t\|^2_n}{\|t\|^2_{\pi^X}} - 1 \right| \leq \frac{1}{2} \right\}, \quad (35)$$

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^c_n.

Bound of the risk on Ω_n On Ω_n, it is

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

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

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

Replacing them in (34) we obtain

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

We need d to be more than 6. We take the optimal choice for d, which corresponds to $d = 12$, obtaining

$$\begin{aligned} \left\|\widehat{\sigma}^2_m - \sigma^2\right\|^2_n &\leq 3\left\|\sigma^2_m - \sigma^2\right\|^2_n + \frac{48}{n}\sum_{i=0}^{n-1} \tilde{A}^2_{t_i} \\ &\quad + \frac{48}{n}\sum_{i=0}^{n-1} (E_{t_i} \varphi_{\Delta^\beta_{n,i}}(\Delta_i X))^2 + 24 \sup_{t \in \mathcal{B}_m} \nu^2_n(t). \quad (36) \end{aligned}$$

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

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

Then,

$$\sup_{t\in\mathcal{B}_m}\nu_n^2(t)=\sup_{\sum_{l=1}^{D_m}\alpha_l^2\le 1}\nu_n^2\left(\sum_{l=1}^{D_m}\alpha_l\psi_l\right)$$
$$\le\sup_{\sum_{l=1}^{D_m}\alpha_l^2\le 1}\left(\sum_{l=1}^{D_m}\alpha_l^2\right)\left(\sum_{l=1}^{D_m}\nu_n^2(\psi_l)\right)=\sum_{l=1}^{D_m}\nu_n^2(\psi_l). \tag{37}$$

To study the risk we need to evaluate the expected value. From (36), (37) and using the first and the third points of Proposition 2, we get

$$\mathbb{E}\left[\left\|\widehat{\sigma}_m^2-\sigma^2\right\|_n^2\mathbb{1}_{\Omega_n}\right]\le 3\mathbb{E}\left[\left\|\sigma_m^2-\sigma^2\right\|_n^2\right]+c\Delta_n^{1-\tilde{\varepsilon}}+c\Delta_n^{4\beta-1}+24\sum_{l=1}^{D_m}\mathbb{E}[\nu_n^2(\psi_l)]. \tag{38}$$

By the definition (33) 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

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

Replacing the inequality here above in (38) it yields

$$\mathbb{E}\left[\left\|\widehat{\sigma}_m^2-\sigma^2\right\|_n^2\mathbb{1}_{\Omega_n}\right]\le 3\mathbb{E}\left[\left\|\sigma_m^2-\sigma^2\right\|_n^2\right]+c\Delta_n^{4\beta-1}+\frac{c\sigma_1^4D_m}{n}.$$

As for any deterministic t it is $\mathbb{E}[\|t\|_n]=\|t\|_{\pi^X}$, it follows

$$\mathbb{E}[\left\|\widehat{\sigma}_m^2-\sigma^2\right\|_n^2\mathbb{1}_{\Omega_n}]\le 3\inf_{t\in\mathcal{S}_m}\left\|t-\sigma^2\right\|_{\pi^X}^2+c\Delta_n^{4\beta-1}+\frac{c\sigma_1^4D_m}{n}. \tag{39}$$

Bound of the risk on Ω_n^c The complementary space Ω_n^c of Ω_n given in Equation (35) is defined as:

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

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

$$\Pi_mT\varphi=\Pi_m(T_{t_0}\varphi_{\Delta_{n,0}^\beta}(\Delta_0X),\dots,T_{t_{n-1}}\varphi_{\Delta_{n,n-1}^\beta}(\Delta_{n-1}X))$$

$$= (\widehat{\sigma}^2_m(X_{t_0}), \dots, \widehat{\sigma}^2_m(X_{t_{n-1}})),$$

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

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

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

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

From Lemma 6.4 in [19], if $\frac{n\Delta_n}{(\log n)^2} \to \infty$ and $N_n \leq \frac{n\Delta_n}{(\log n)^2}$ for [DP] and [W] and $N_n^2 \leq \frac{n\Delta_n}{(\log n)^2}$ for the collection [T], then

$$\mathbb{P}(\Omega_n^c) \leq \frac{c_0}{n^4}. \tag{40}$$

In the hypothesis of our proposition we have requested that $\log n = o(\sqrt{n\Delta_n})$. As for n going to ∞ we have $\frac{(\log n)^2}{n\Delta_n} < \frac{\log n}{\sqrt{n\Delta_n}} \to 0$, the first condition in Lemma 6.4 in [19] hold true. Regarding the bound on N_n, we have required Assumption 5 and so we can apply the here above mentioned lemma, which yields (40).

We are left to evaluate $\mathbb{E}[e_{t_i}^4]$. From Proposition 2 it follows

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

Putting the pieces together it yields

$$\mathbb{E}\left[\left\|\widehat{\sigma}^2_m - \sigma^2\right\|_n^2 \mathbb{1}_{\Omega_n^c}\right] \leq \frac{c\Delta_n^{0\wedge 4\beta-\frac{3}{2}}}{n^2} + \frac{c}{n^4} \leq \frac{c\Delta_n^{0\wedge 4\beta-\frac{3}{2}}}{n^2}. \tag{41}$$

From (39) and (41) it follows

$$\mathbb{E}\left[\left\|\widehat{\sigma}^2_m - \sigma^2\right\|_n^2\right] \leq 6\inf_{t\in\mathcal{S}_m}\left\|t - \sigma^2\right\|_{\pi^X}^2 + \frac{C_1\sigma_1^4 D_m}{n} + C_2\Delta_n^{4\beta-1} + \frac{C_3\Delta_n^{0\wedge 4\beta-\frac{3}{2}}}{n^2}. \quad \square$$

8.1.2. Proof of Theorem 1

Proof. For simplicity in notation we denote $\widehat{m}_\sigma = \widehat{m}$ in the 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

$$\mathbb{E}\left[\left\|\widehat{\sigma}^2_{\widehat{m}} - \sigma^2\right\|_n^2 \mathbb{1}_{\Omega_n^c}\right] \leq \frac{c\Delta_n^{0\wedge 4\beta-\frac{3}{2}}}{n^2}. \tag{42}$$

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

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

and so, acting as before (36), we get

$$\begin{aligned}\mathbb{E}\left[\left\|\sigma^2_{\widehat{m}} - \sigma^2\right\|_n^2 \mathbb{1}_{\Omega_n}\right] &\leq 3\mathbb{E}[\left\|\sigma^2_m - \sigma^2\right\|_n^2] + \frac{48}{n}\sum_{i=0}^{n-1}\mathbb{E}[\tilde{A}^2_{t_i}] \\ &+ \frac{48}{n}\sum_{i=0}^{n-1}\mathbb{E}[(E_{t_i}\varphi_{\Delta^\beta_{n,i}}(\Delta_i X))^2] + 24\,\mathbb{E}\left[\sup_{t\in\mathcal{B}_{m,\widehat{m}}}\nu_n^2(t)\right] \qquad (43)\\ &+ 12\mathrm{pen}(m) - 12\mathbb{E}[\mathrm{pen}(\widehat{m})], \qquad (44)\end{aligned}$$

where ν_n has been defined in (33) 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

$$p(m,m') = \frac{1}{24}(\mathrm{pen}(m) + \mathrm{pen}(m')). \tag{45}$$

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 [35]. 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] there exists a constant c such that, for any $p > 0$,

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

It follows

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

By Lemma 7 in [35] there exists a constant k such that, for any $m, m' \in \mathcal{M}_n$,

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,m'}}\nu_n^2(t) - kc\sigma_1 p(m,m')\right)_+\right] \leq c\frac{e^{-(D_m+D_{m'})}}{n}. \tag{46}$$

We have said, in the definition of the penalization function pen_σ given in Subsection 3.2, that the constant κ_1 has to be calibrated. In particular, we need it to be such that $\frac{\kappa_1}{24} \geq kc\sigma_1$, 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 [35]. We underline that Lemma 7 in [35] has been proved for a noisy diffusion. However, the same reasoning applies for a jump diffusion (see the proof of Theorem 13 in [36]) and for our framework as well, as it is based on a projection argument and on algebraic computations which still hold true.

We remark that Assumption (iii) of Section 2.2 of [35], on the cardinality of the support of the basis, holds true only for the collections [DP] and [W]. However, Lemma 7 of [35] still holds true for the collection [T], up to add the condition $\frac{N_n^3}{n} \leq 1$ as in the first point of Assumption 5. Indeed Lemma 8 of [35] (on which the proof of Lemma 7 relies) does not change considering the collection [T]. Then $\bar{r}_{m,m'}$, as introduced in the proof of Lemma 7 in [35], is now bounded by an extra $D = \max(D_m, D'_m)$, which implies an extra D in the definition of both η_0 and η_k. In particular now we have, using the notation in Lemma 7 of [35], $\eta_k := 2^{-k}(\sqrt{c_3 x_k} + c_4 D x_k)$. It follows, after having replaced x_k,

$$\eta^2 = (\sum_{k=0}^{\infty} \eta_k)^2 \leq c\gamma^2(\frac{D}{n} + \frac{\tau}{n} + \frac{D^4}{n^2} + \frac{D^2\tau^2}{n^2}).$$

Now, $\frac{D^4}{n^2} \leq \frac{D}{n}$ as we have assumed $\frac{N_n^3}{n} \leq 1$ in Assumption 5. Then, following again the proof of Lemma 7 in [35] but substituting the variable τ with y such that $\tau = c\gamma^2(\frac{y}{n} + D^2\frac{y^2}{n^2})$, we get

$$\begin{aligned} E &= C\gamma^2 e^{-D}(\frac{1}{n}\int_0^\infty e^{-y}dy + \frac{2}{n^2}\int_0^\infty D^2 y e^{-y}dy) \\ &\leq c\frac{\gamma^2}{n}e^{-D}(1 + \frac{D^2}{n}). \end{aligned}$$

However, Assumption 5 implies $\frac{D^2}{n} \leq 1$ and so we get that the extra part due to the choice of the collection [T] is negligible. We recover $E \leq c\frac{\gamma^2}{n}e^{-D}$, as in Lemma 7 of [35] and as we wanted.

From (46) 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 (41) and Proposition 2,

$$\begin{aligned} \mathbb{E}\left[\left\|\widehat{\sigma}^2_{\widehat{m}} - \sigma^2\right\|_n^2\right] \quad \leq \quad & 3\,\mathbb{E}\left[\left\|\sigma^2_m - \sigma^2\right\|_n^2\right] + c\Delta_n^{4\beta-1} + \frac{c}{n^4} + c\text{pen}(m) \\ & + \frac{c\Delta_n^{0\wedge(4\beta-\frac{3}{2})}}{n^2} + \frac{c}{n} \end{aligned}$$

$$\leq C_1 \inf_{m\in\mathcal{M}_n}\left\{\inf_{t\in\mathcal{S}_m}\|t-\sigma^2\|_{\pi^X}^2+\mathrm{pen}(m)\right\} + C_2\Delta_n^{4\beta-1}+\frac{C_3\Delta_n^{4\beta-\frac{3}{2}}}{n^2}+\frac{C_4}{n}. \quad \square$$

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

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

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

$$\begin{aligned}\|\widehat{g}_m-g\|_n^2 &\leq \|g_m-g\|_n^2+\frac{2d}{n}\sum_{i=0}^{n-1}A_{t_i}^2+\frac{1}{d}\|\widehat{g}_m-g_m\|_n^2+2d\sup_{\mathcal{B}_m}\nu_{n,1}^2(t)\\ &\quad +\frac{1}{d}\|\widehat{g}_m-g_m\|_{\pi^X}^2+2d\sup_{\mathcal{B}_m}\nu_{n,2}^2(t),\end{aligned}\tag{47}$$

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

$$\nu_{n,1}(t):=\frac{1}{n}\sum_{i=0}^{n-1}(B_{t_i}+E_{t_i})t(X_{t_i}),\quad \nu_{n,2}(t):=\frac{1}{n}\sum_{i=0}^{n-1}C_{t_i}t(X_{t_i}).\tag{48}$$

We still denote Ω_n the space on which the norms $\|\cdot\|_{\pi^X}$ and $\|\cdot\|_n$ are equivalent given by Equation (35). 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 (47) we get

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

As before, we take $d = 12$. It yields

$$\|\widehat{g}_m - g\|_n^2 \leq 3\|g_m - g\|_n^2 + \frac{48}{n}\sum_{i=0}^{n-1} A_{t_i}^2 + 48\sup_{t\in\mathcal{B}_m}\nu_{n,1}^2(t) + 48\sup_{t\in\mathcal{B}_m}\nu_{n,2}^2(t). \quad (49)$$

We now need introduce a orthonormal basis of S_m. Hence, we consider $(\tilde{\psi}_k)_k$, an orthonormal basis of S_m for the $L^2_{\pi^X}$ norm, as before. Each $t \in \mathcal{B}_m$ can be written

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

Then, for $j = 1$ and $j = 2$,

$$\sup_{t\in\mathcal{B}_m}\nu_{n,j}^2(t) = \sup_{\sum_{l=1}^{D_m}\alpha_l^2\leq 1}\nu_{n,j}^2\left(\sum_{l=1}^{D_m}\alpha_l\tilde{\psi}_l\right) \quad (50)$$

$$\leq \sup_{\sum_{l=1}^{D_m}\alpha_l^2\leq 1}\left(\sum_{l=1}^{D_m}\alpha_l^2\right)\left(\sum_{l=1}^{D_m}\nu_{n,j}^2(\tilde{\psi}_l)\right) = \sum_{l=1}^{D_m}\nu_{n,j}^2(\tilde{\psi}_l), \quad (51)$$

where we have also used Cauchy-Schwartz inequality. To study the risk we need to evaluate the expected value. From (49), (51) and using the first point of Proposition 4, we get

$$\begin{aligned}&\mathbb{E}\left[\|\widehat{g}_m - g\|_n^2\,\mathbb{1}_{\Omega_n}\right] \\ &\leq 3\mathbb{E}\left[\|g_m - g\|_n^2\right] + c\Delta_n^{1-\tilde{\varepsilon}} + 48\sum_{l=1}^{D_m}\mathbb{E}[\nu_{n,1}^2(\tilde{\psi}_l)] + 48\sum_{l=1}^{D_m}\mathbb{E}\left[\nu_{n,2}^2(\tilde{\psi}_l)\right]. \quad (52)\end{aligned}$$

By the definition (48) of $\nu_{n,1}$ and the points 2 and 3 of Proposition 4, it is

$$\sum_{l=1}^{D_m}\mathbb{E}[\nu_{n,1}^2(\tilde{\psi}_l)] \quad \leq \quad \frac{c}{n^2}\sum_{i=0}^{n-1}\sum_{l=1}^{D_m}\mathbb{E}\left[\tilde{\psi}_l^2(X_{t_i})\mathbb{E}[B_{t_i}^2 + E_{t_i}^2|\mathcal{F}_{t_i}]\right]$$

$$\leq \frac{c}{n^2}\sum_{i=0}^{n-1}\sum_{l=1}^{D_m}\mathbb{E}\left[\tilde{\psi}_l^2(X_{t_i})(c\sigma_1^4+\frac{ca_1^4}{\Delta_{n,i}}\sum_{j=1}^{M}|\lambda_{t_i}^{(j)}|)\right]$$

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

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

Regarding the second term, we remark that, as $\left\|\tilde{\psi}_l\right\|_\infty\leq D_m$ and its norm 2 is bounded by 1, it is

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

where we have used Holder inequality with $p=1+\epsilon$, for $\epsilon>0$ arbitrarily small, and the boundedness of the moments of λ. It follows

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

Hence,

$$\sum_{l=1}^{D_m}\mathbb{E}\left[\nu_{n,1}^2(\tilde{\psi}_l)\right]\leq\frac{c(\sigma_1^4+a_1^4)D_m^{1+2\epsilon}}{n\Delta_{n,i}}. \tag{53}$$

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

Lemma 4. *Suppose that A1-A3 hold true. Then, for any $\varepsilon>0$ arbitrarily small,*

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=0}^{n-1}C_{t_i}\tilde{\psi}_l(X_{t_i})\right)\leq\frac{cD_m^{2\varepsilon}}{n\Delta_n}.$$

The proof of Lemma 4 is in the appendix. Lemma 4 yields

$$\sum_{l=1}^{D_m}\mathbb{E}\left[\nu_{n,2}^2(\tilde{\psi}_l)\right]\leq\frac{cD_m^{1+2\varepsilon}}{n\Delta_n}. \tag{54}$$

Replacing the inequality here above and (53) in (52) 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 $c_1\leq\frac{\Delta_n}{\Delta_{\min}}\leq c_2$,

$$\mathbb{E}\left[\|\widehat{g}_m-g\|_n^2\,\mathbb{1}_{\Omega_n}\right]\leq 3\mathbb{E}\left[\|g_m-g\|_n^2\right]+c\Delta_n^{1-\tilde{\varepsilon}}+\frac{c(\sigma_1^4+a_1^4+1)D_m^{1+2\varepsilon}}{n\Delta_n}.$$

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

$$\mathbb{E}\left[\|\widehat{g}_m - g\|_n^2 \mathbb{1}_{\Omega_n}\right] \leq 3 \inf_{t \in \mathcal{S}_m} \|t - g\|_{\pi^X}^2 + c\Delta_n^{1-\tilde{\varepsilon}} + \frac{c(\sigma_1^4 + a_1^4 + 1)D_m^{1+2\varepsilon}}{n\Delta_n}. \quad (55)$$

Bound of the risk on Ω_n^c Let us set $e = (e_{t_0}, \dots, e_{t_{n-1}})$, where $e_{t_i} := T_{t_i} - g(X_{t_i}) = A_{t_i} + B_{t_i} + C_{t_i} + E_{t_i}$. Moreover

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

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,

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

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

$$\mathbb{E}[\|\widehat{g}_m - g\|_n^2 \mathbb{1}_{\Omega_n^c}] \leq \frac{c}{\Delta_n^{\frac{3}{2}}}\frac{1}{n^2} + \frac{c}{n^2} \leq \frac{c}{n^2\Delta_n^{\frac{3}{2}}}. \quad (56)$$

From (55) and (56) it follows

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

8.2.2. Proof of Theorem 2

Proof. For simplicity in notation we denote $\widehat{m}_g = \widehat{m}$ in the 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

$$\mathbb{E}\left[\|\widehat{g}_{\widehat{m}} - g\|_n^2 \mathbb{1}_{\Omega_n^c}\right] \leq \frac{c}{n^2\Delta_n^{\frac{3}{2}}}. \quad (57)$$

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 (65)). 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, acting as to obtain Equation (49), we get

$$\begin{aligned}\mathbb{E}\left[\|\widehat{g}_{\widehat{m}} - g\|_n^2 \mathbb{1}_{\mathcal{O}}\right] \quad \leq \quad & 3\mathbb{E}[\|g_m - g\|_n^2] + \frac{48}{n}\sum_{i=0}^{n-1}\mathbb{E}[A_{t_i}^2] + 48\,\mathbb{E}\left[\sup_{t\in\mathcal{B}_{m,\widehat{m}}} \nu_n^2(t)\mathbb{1}_{\mathcal{O}}\right] \\ & +12\text{pen}(m) - 12\mathbb{E}[\text{pen}(\widehat{m})],\end{aligned}$$

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 $p(m,m')$:

$$p(m,m') \leq \frac{1}{48}(\text{pen}(m) + \text{pen}(m')).$$

It is

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

Replacing it in (43) and using the first point of Proposition 4 we get

$$\begin{aligned}\mathbb{E}\left[\|\widehat{g}_{\widehat{m}} - g\|_n^2 \mathbb{1}_{\mathcal{O}}\right] \quad \leq \quad & 3\mathbb{E}\left[\|g_m - g\|_n^2\right] + c\Delta_n^{1-\tilde{\varepsilon}} \\ & +48\mathbb{E}[p(m,\widehat{m})] + 12\text{pen}(m) - 12\mathbb{E}[\text{pen}(\widehat{m})] \\ & +48\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,\widehat{m}}} \nu_n^2(t) - p(m,m')\right)_+ \mathbb{1}_{\mathcal{O}}\right] \quad (58)\end{aligned}$$

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 [36] and proved by Birgé and Massart (1998) [9] (corollary 2 p.354) and Comte and Merlevède (2002) [15] (p.222-223).

Lemma 5. *Let $T_1, \ldots, T_p$ 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,

$$\mathbb{E}\left[\left(\sup_{r\in\mathcal{B}_{m,m'}}|v_p(r)|^2-2H^2\right)_+\right]\leq c\left(\frac{v}{p}e^{-c\frac{pH^2}{v}}+\frac{M^2}{p^2}e^{-c\frac{pH}{M}}\right), \tag{59}$$

with c a universal constant and where

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

We observe that in Talagrand lemma here above the random variables $T_1,\dots,T_p$ 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 [38] while an analogous statement in continuous time can be found in [3].

Lemma 6. *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*

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

There exist random variables $M^*_{t_0},\dots,M^*_{t_n}$ such that

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

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 $M_{t_i}=(B_{t_i},C_{t_i},E_{t_i},X_{t_i})$, that we write as a function of (X_t,λ_t), which is stationary and exponentially β- mixing, as discussed in Section 2.3. We define the σ algebra

$$\tilde{\mathcal{F}}_{t_i}:=\sigma(X_s,\lambda_s,s\in(t_i,t_{i+1}]), \tag{60}$$

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

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

Writing the dynamic of $\lambda=(\lambda^{(1)},\dots,\lambda^{(d)})$ in the matrix form, $d\lambda_t=-\alpha(\lambda_t-\zeta)dt+cdN_t$, and since c is invertible, we can get dN_t as a function of $d\lambda_t$ and λ_t. Then, using the invertibility of σ with the second line of (2), we can write dW_t as a function of $dX_t,X_t,d\lambda_t$, and λ_t. Now, by the definition of B_{t_i}, C_{t_i}

and E_{t_i} it follows that $(B_{t_i}, C_{t_i}, E_{t_i})$ is measurable with respect to $\tilde{\mathcal{F}}_{t_i}$. We can therefore use Berbee's coupling on $M_{t_i} = (B_{t_i}, C_{t_i}, E_{t_i}, X_{t_i})$. Let q_n be the size of the blocks, which we will specify later. As it may happen that $2q_n$ does not divide n, we set $p_n = \lfloor n/(2q_n) \rfloor$ and remove from the definition of the contrast function (26) the data corresponding to the indexes $i \in \{2p_nq_n, \dots, n-1\}$. This modification avoids dealing with a last block having a different size, and we can apply Berbee's lemma to $(M_{t_i})_{i=0,\dots,2p_nq_n}$. It yields to the construction of variables $(U^*_{k,j})_{k\in\{1,\dots,p_n\};j=0,1}$ such that $U^*_{k,j}$ has the same law as $U_{k,j} = (B_{t_{(2(k-1)+j)q_n+l}}, C_{t_{(2(k-1)+j)q_n+l}}, E_{t_{(2(k-1)+j)q_n+l}}, X_{t_{(2(k-1)+j)q_n+l}})_{l\in\{1,\dots,q_n\}}$, and for $j \in \{0,1\}$, the random variables $(U^*_{k,j})_{1\le k\le p_n}$ are independent. Let us set

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

by Berbee's coupling lemma it comes that

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

It is enough to take $q_n := \lfloor \frac{5}{\gamma\Delta_{\min}}\log n \rfloor$ in (61) to get

$$\mathbb{P}(\Omega^{*,c}) \le \frac{c}{n^4\log n}. \tag{61}$$

For $t \in \mathcal{B}_{m,m'}$, and $(j,k) \in \{0,1\}\times\{1,\dots,p_n\}$ we define both

$$t(U^*_{k,j}) = \frac{1}{q_n}\sum_{l=1}^{q_n}(B^*_{t_{(2(k-1)+j)q_n+l}} + C^*_{t_{(2(k-1)+j)q_n+l}} + E^*_{t_{(2(k-1)+j)q_n+l}}) \times t(X^*_{t_{(2(k-1)+j)q_n+l}}), \tag{62}$$

and $t(U_{k,j})$ the analogous quantity based on $U_{k,j}$.

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

$$v^{0,*}_p(t) = \frac{1}{p}\sum_{k=1}^{p} t(U^*_{k,0}), \qquad v^{1,*}_p(t) = \frac{1}{p}\sum_{k=1}^{p} t(U^*_{k,1}). \tag{63}$$

With these definitions, we have on the set Ω^*, $v^*_n(t) = \nu_n(t)$ for all $t \in \mathcal{B}_{m,m'}$. Now we want to compute the constants M, v and H as defined in Lemma 5. The random variables $t(U^*_{k,j})$ are not bounded, hence, to compute M, we introduce the following set

$$\Omega_B := \left\{\omega : \forall j, \forall k, \forall t \in \mathcal{B}_{m,m'},\ |t(U^*_{k,j})| \le \tilde{c}n^{\varepsilon_0}D^{\frac{1}{2}}\right\}, \tag{64}$$

with $D := D_m + D_{m'}$ and some $\varepsilon_0 > 0$. The following lemma is proven in the appendix.

Lemma 7. *Suppose that A1-A3 hold. Then there exists $c > 0$ such that*

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

We introduce bounded version of the random variables $t(U^*_{k,j})$ by setting for $M > 0$,

$$t^{(M)}(U^*_{k,j}) = t(U^*_{k,j}) \vee -M \wedge M.$$

With the choice $M := \tilde{c} n^{\varepsilon_0} D^{\frac{1}{2}}$, we have on the event Ω_B that $t^{(M)}(U^*_{k,j}) = t(U^*_{k,j})$, $\forall j, \forall k, , \forall t \in \mathcal{B}_{m,m'}$. We set

$$\mathcal{O} := \Omega_n \cap \Omega_B \cap \Omega^*. \tag{65}$$

From (40), (61) and Lemma 7 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

$$\mathbb{E}[\|\widehat{g}_{\widehat{m}} - g\|_n^2 \mathbb{1}_{\mathcal{O}^c}] \leq \frac{c}{n^2 \Delta_n^{\frac{3}{2}}}. \tag{66}$$

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

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_{m,m'}} \nu_n(t)^2 - p(m,m')\right)_+ \mathbb{1}_{\mathcal{O}}\right]. \tag{67}$$

On $\mathcal{O}$, we have $\nu_n(t) = v^{0,*}_{p_n}(t) + v^{1,*}_{p_n}(t)$ and

$$\begin{aligned} v^{j,*}_p(t) &= \frac{1}{p}\sum_{k=1}^p t^{(M)}(U^*_{k,j}) = \frac{1}{p}\sum_{k=1}^p \left(t^{(M)}(U^*_{k,j}) - \mathbb{E}[t^{(M)}(U^*_{k,j})]\right) + E[t^{(M)}(U^*_{0,0})] \\ &= v^{j,*}_p(t^{(M)}) + \mathbb{E}[t^{(M)}(U^*_{0,0})], \end{aligned}$$

where $v^{j,*}_p$ has been defined in (63).

Using $\mathbb{E}[t(U^*_{0,0})] = 0$, we deduce

$$|\mathbb{E}[t^{(M)}(U^*_{0,0})]| \leq \mathbb{E}\left[|(t - t^{(M)})(U^*_{0,0})| \mathbb{1}_{\Omega_B^c}\right] \leq \mathbb{E}\left[t(U^*_{0,0})^2\right]^{1/2} \mathbb{P}(\Omega_B^c)^{1/2}.$$

We need the following Lemma whose proof is postponed to the Appendix.

Lemma 8. *We have* $\sup_{t \in \mathcal{B}_{m,m'}} \mathbb{E}\left[t(U^*_{0,0})^2\right] \leq c\frac{D^\delta}{q_n \Delta_n}$, *for* δ *arbitrarily small and some constant* c.

Using Lemma 7 and Lemma 8, we deduce $\sup_{t \in \mathcal{B}_{m,m'}} |E[t^{(M)}(U^*_{0,0})]| \leq c\frac{D^{\delta/2}}{(q_n\Delta_n)^{1/2} n^2} \leq c\,\frac{D^{\delta/2}}{(\ln(n))^{1/2} n^2} \leq c\frac{p(m,m')}{n}$. Hence to control the term (67) it is sufficient to get an upper bound, for n large enough on

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_{m,m'}} v^{j,*}_p(t^{(M)})^2 - \frac{1}{4}p(m,m')\right)_+\right],$$

for $j = 0, 1$. We can apply Lemma 5 to this term. To this purpose, we need to compute the constants M, v and H appearing therein. By construction, we can

use $M = cn^{\varepsilon_0} D^{\frac{1}{2}}$, and by Lemma 8 and stationarity we can take $v = c\frac{D^\delta}{q_n \Delta_n}$. In order to compute H^2 we observe it is

$$\begin{aligned}
\mathbb{E}[\sup_{t\in\mathcal{B}_{m,m'}} |v_p^{j,*}(t^{(M)})|] &\le \mathbb{E}[\sup_{t\in\mathcal{B}_{m,m'}} |v_p^{j,*}(t^{(M)})|\mathbb{1}_{\mathcal{O}}] + \mathbb{E}[\sup_{t\in\mathcal{B}_{m,m'}} |v_p^{j,*}(t^{M})|\mathbb{1}_{\mathcal{O}^c}], \\
&\le \mathbb{E}[\sup_{t\in\mathcal{B}_{m,m'}} |v_p^{j,*}(t)|\mathbb{1}_{\mathcal{O}}] + \sup_{t\in\mathcal{B}_{m,m'}} |\mathbb{E}[t^{(M)}(U^*_{0,0})]| + M\mathbb{P}(\mathcal{O}^c), \\
\le \mathbb{E}[\sup_{t\in\mathcal{B}_{m,m'}} |v_p^{j,*}(t)|] + c\frac{D^{\delta/2}}{q_n\Delta_n n^2} + c\frac{D^{1/2}n^{\varepsilon_0}}{n^4} &\le \sqrt{\mathbb{E}[\sup_{t\in\mathcal{B}_{m,m'}} (v_p^*)^2(t)]} + c\,\frac{D^{1/2}}{n^2},
\end{aligned}$$

where we have used $|t^M(U_{j,k})| \le M$. To find an upper bound for the right hand side here above we act similarly to how we did before (51): we introduce the orthonormal basis $(\bar\psi_k)_k$, such that each $t \in \mathcal{B}_{m,m'}$ can be written as the following

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

Similarly to (51), we have

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

Acting exactly as we did in order to get (53) and Lemma 4 on $v_{n,1}^2$ and $v_{n,2}^2$ (as for Equation (54)) we obtain

$$\sqrt{\mathbb{E}[\sup_{t\in\mathcal{B}_{m,m'}} (v_p^{j,*})^2(t)]} \le c\sqrt{\frac{D^{1+2\varepsilon}}{n\Delta_n}}.$$

In turn we have $\mathbb{E}[\sup_{t\in\mathcal{B}_{m,m'}} |v_p^{j,*}(t^{(M)})|] \le c\sqrt{\frac{D^{1+2\varepsilon}}{n\Delta_n}} =: H$.

We now use Talagrand inequality as in Lemma 5. It follows

$$\begin{aligned}
&\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,\widehat m}} \nu_n^*(t)^2 - 2H^2\right)_+ \mathbb{1}_{\mathcal{O}}\right] \\
&\le \frac{D^\delta}{p_n q_n \Delta_n} \exp\left(-c\frac{D^{1+2\varepsilon} p_n q_n \Delta_n}{n\Delta_n D^\delta}\right) + \frac{cn^{2\varepsilon_0} D}{p_n^2} \exp\left(-c\frac{p_n D^{\frac{1}{2}+\varepsilon}}{\sqrt{n\Delta_n} n^{\varepsilon_0} D^{\frac{1}{2}}}\right) \\
&\le \frac{cD^\delta}{n\Delta_n} \exp(-\frac{c}{2} D^{1+2\varepsilon-\delta}) + \frac{cn^{2\varepsilon_0} D}{p_n^2} \exp\left(-\frac{c\sqrt{n} D^\varepsilon}{2q_n\sqrt{\Delta_n} n_0^\varepsilon}\right),
\end{aligned}$$

where we used $\frac{2p_nq_n}{n} \xrightarrow{n\to\infty} 1$. We recall that $q_n = \lfloor c\frac{\log n}{\Delta_{\min}} \rfloor$. We observe that, as $\Delta_{\min}$ and Δ_n differs only from a constant, $\frac{c\sqrt{n}}{\sqrt{\Delta_n}q_n n^{\varepsilon_0}} \geq \frac{c\sqrt{n\Delta_n}}{\log n n^{\varepsilon_0}}$. Moreover, it goes to ∞ for n going to infinity as we have assumed that $(\log n)n^{\varepsilon} = o(\sqrt{n\Delta_n})$ for some ε and the constant ε_0 can be arbitrarily small. 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 $D^{\delta}e^{-\frac{c}{2}D^{1+2\varepsilon-\delta}} < D^{\delta}e^{-\frac{c}{2}D^{1-\delta}} < c''e^{-c'D^{1-\frac{\delta}{2}}}$ and the fourth point of Assumption 4,

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

Replacing it in the equivalent of (58), considering that we are now on $\mathcal{O}$, it follows

$$\begin{aligned}\mathbb{E}\left[\|\widehat{g}_{\widehat{m}} - g\|_n^2 \mathbb{1}_{\mathcal{O}}\right] &\leq 3\mathbb{E}[\|g_m - g\|_n^2] + c\Delta_n^{1-\tilde{\varepsilon}} + 48\mathbb{E}[p(m,\widehat{m})] \\ &\quad +12\mathrm{pen}(m) - 12\mathbb{E}[\mathrm{pen}(\widehat{m})] + \frac{c}{n\Delta_n}.\end{aligned}$$

It provides us, using also (66),

$$\begin{aligned}\mathbb{E}[\|\widehat{g}_{\widehat{m}} - g\|_n^2] &\leq 3\mathbb{E}[\|\widehat{g}_m - g\|_n^2] + c\Delta_n^{1-\tilde{\varepsilon}} + \frac{c}{n^2\Delta_n^{\frac{3}{2}}} + c\mathrm{pen}(m) + \frac{c}{n\Delta_n} \\ &\leq c_1 \inf_{m\in\mathcal{M}_n}\left\{\inf_{t\in\mathcal{S}_m} \|t-g\|_{\pi^X}^2 + \mathrm{pen}(m)\right\} + C_2\Delta_n^{1-\tilde{\varepsilon}} \\ &\quad + \frac{C_3}{n^2\Delta_n^{\frac{3}{2}}} + \frac{C_4}{n\Delta_n}. \quad \square\end{aligned}$$

Supplementary material

Appendix A

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 9. *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 $j\in\{1,\dots,M\}$, $\sup_{h\in[0,1]}\mathbb{E}[|\lambda_{s+h}^{(j)}||\mathcal{F}_s] \leq |\lambda_s^{(j)}| + c|h|(1+|\lambda_s|)$.*

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

$$\mathbb{E}[I_1] \leq c|t-s|^{p-1}\int_s^t \mathbb{E}[|b(X_u)|^p]du \leq c|t-s|^p. \tag{68}$$

Using Burkholder-Davis-Gundy inequality, Jensen inequality and Assumption 1.2. on σ it is

$$\mathbb{E}[I_2] \leq c\mathbb{E}\left[\left(\int_s^t \sigma^2(X_u)du\right)^{\frac{p}{2}}\right] \leq c|t-s|^{\frac{p}{2}-1}\int_s^t \mathbb{E}[|\sigma(X_u)|^p]du \leq c\sigma_1^p|t-s|^{\frac{p}{2}}. \tag{69}$$

To evaluate I_3, Kunita inequality will be useful. We refer to the Appendix of [29] 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:

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

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

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

$$+c|a_1|^p|t-s|^{p-1}\int_s^t \mathbb{E}[|\lambda_u^{(j)}|^p]du$$
$$\leq c|a_1|^p(|t-s|+|t-s|^{\frac{p}{2}}) \leq c|a_1|^p|t-s|, \tag{70}$$

where we have used that λ has the moment of any order because of Proposition 1. From (68), (69) and (70), 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

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

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

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

Finally, Gronwall lemma 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

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

is the jump-diffusion part and

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

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

$$A_2^{\tilde{Z}}V(x,y) = A_2^{\tilde{Z}}V_2(y) \leq -c_1 V_2(y) + c_2 \mathbb{1}_{K_1}(y), \tag{71}$$

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

$$A_1^{\tilde{Z}}V(x,y) \leq -dm|x|^m + c|x|^{m-2} + \bar{f}(y)(c_1|x|^{m-1} + \ldots + c_m). \tag{72}$$

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 (72), 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,

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

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 [20]). 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 (71) $-c_1 V_2(y)$. The same reasoning applies for $\frac{c}{q}\bar{f}(y)^q$. It follows that

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

which, together with (71), conclude the proof of the first part. Regarding the boundedness of the moments, we use that the Lyapunov function V admits a finite integral with respect to the stationary probability of $\tilde{Z} = (X, Y)$. As the process λ can be recovered as a linear function of Y, the ergodicity $\tilde{Z}$ of implies the ergodicity of $Z = (X, \lambda)$ as well, and we have existence of bounded moments of any order for both X and λ under the stationary law. □

A.2. Proof of Lemma 1

Proof. We first recall the representation of $\pi^X(x)$ given in the proof of Proposition 3.7 in [20]. For $t > 0$, let $L_t = \sup\{s \leq t \mid \exists j,\ \Delta N_s^{(j)} = 1\}$ be the time of the last jump before t, with $L_t = 0$ if there is no such jump. Then, we have for all $x \in \mathbb{R}$

$$\pi^X(x) = \int_{\mathbb{R}\times\mathbb{R}^{M\times M}} \pi(dz)\mathbb{E}_z\left[p_{t-L_t}(X_{L_t}, x)\right],$$

where $(p_s)_{s>0}$ is the family of transition densities associated to the stochastic differential equation $dX_t = b(X_t)dt + \sigma(X_t)dW_t$. From Proposition 1.2 of [24], we know that there exist constants c, C such that for all $s > 0$, $(u, x) \in \mathbb{R}^2$,

$$p_s(u,x) \leq cs^{-1/2} e^{-\frac{(x-u)^2}{Cs}} e^{Csu^2}.$$

We deduce that

$p_s(u,x) \leq cs^{-1/2} e^{-\frac{(x-u)^2}{Cs}} e^{2Cs(x-u)^2 + 2Csx^2} \leq cs^{-1/2} e^{-\frac{(x-u)^2}{2Cs}} e^{2Csx^2}$, if s is smaller than $1/(2C)$. Choosing $t < 1/(2C)$, it yields

$$\pi^X(x) \leq c \int_{\mathbb{R}\times\mathbb{R}^{M\times M}} \pi(dz)\mathbb{E}_z\left[\frac{1}{\sqrt{t-L_t}}\right] e^{2Ctx^2}. \tag{73}$$

We now give an upper bound for $\mathbb{E}_z\left[\frac{1}{\sqrt{t-L_t}}\right]$. Writing

$(t-L_t)^{-1/2} = \frac{1}{2}\int_0^t \frac{1_{\{s\leq L_t\}}ds}{(t-s)^{3/2}} + t^{-1/2}$, we have

$$\mathbb{E}_z\left[\frac{1}{\sqrt{t-L_t}}\right] \leq \frac{1}{2}\int_0^t \frac{P_z(L_t \geq s)}{(t-s)^{3/2}} ds + \frac{1}{t^{1/2}}. \tag{74}$$

From the definition of L_t, it is $P_z(L_t \geq s) = P_z(\exists u \in [s,t], \exists j \ : \ \Delta N_u^{(j)} = 1)$ which implies

$$P_z(L_t \geq s) \leq P_z\left(\sum_{j=1}^{M} \int_{[s,t]} dN_u^{(j)} \geq 1\right) \leq \mathbb{E}_z\left[\sum_{j=1}^{M} \int_{[s,t]} dN_u^{(j)}\right]$$
$$\leq \mathbb{E}_z\left[\sum_{j=1}^{M} \int_s^t \lambda_u^{(j)} du\right] = \sum_{j=1}^{M} \int_s^t \mathbb{E}_z[\lambda_u^{(j)}] du.$$

Using Lemma 9 we get that if $u \leq 1$, $\mathbb{E}_z[\lambda_u^{(j)}] \leq c(1 + |\lambda_0(z)|)$, where $\lambda_0^j(z) = f_j(\sum_{i=1}^M y^i)$ for $z = (x, y)$. Thus, if t is chosen smaller than 1, we have $P_z(L_t \geq s) \leq c(1 + |\lambda_0(z)|)(t - s)$. Using this control with (74) we deduce,

$$\mathbb{E}_z\left[\frac{1}{\sqrt{t - L_t}}\right] \leq \frac{c}{2} \int_0^t \frac{(1 + |\lambda_0(z)|)}{(t-s)^{1/2}} ds + \frac{1}{t^{1/2}} \leq \frac{c}{\sqrt{t}}(1 + |\lambda_0(z)|).$$

From (73), we obtain $\pi^X(x) \leq \frac{ce^{2Ctx^2}}{\sqrt{t}} \int_{\mathbb{R}\times\mathbb{R}^{M\times M}} \pi(dz)(1 + |\lambda_0(z)|)$. By Proposition 1, we know that the intensity λ has finite moments of any order under the stationary measure, and thus $\int_{\mathbb{R}\times\mathbb{R}^{M\times M}} \pi(dz)(1 + |\lambda_0(z)|) \leq c < \infty$. This gives $\pi^X(x) \leq \frac{ce^{2Ctx^2}}{\sqrt{t}}$ for any sufficiently small t, and the lemma follows. □

Remark 2. *The proof of Lemma 1 heavily relies on the integrability near zero of the supremum of the heat kernel and is thus limited to the dimension* 1. *We do not know if it is possible to extend this result to higher dimension for the process* X. *However, it is certainly possible to extend this proof to more general situations, as for instance the case where the jump intensity depends on* X.

A.3. Proof of Lemma 2

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,

$$\mathbb{E}[|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X) - 1|^k] \leq c\mathbb{E}[\mathbb{1}_{\{|\Delta_i X| > \Delta_{n,i}^\beta\}}]$$
$$= c\mathbb{E}\left[\mathbb{1}_{\left\{|\Delta_i X| > \Delta_{n,i}^\beta, |J_{t_i}| \leq \frac{\Delta_{n,i}^\beta}{2}\right\}}\right] + c\mathbb{E}\left[\mathbb{1}_{\left\{|\Delta_i X| > \Delta_{n,i}^\beta, |J_{t_i}| > \frac{\Delta_{n,i}^\beta}{2}\right\}}\right]. \quad (75)$$

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

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

The first term in the right hand side of (75) is

$$c\mathbb{E}\left[\mathbb{1}_{\left\{|\Delta_i X^c|>\frac{\Delta_{n,i}^\beta}{2}\right\}}\right] = c\mathbb{P}(|\Delta_i X^c| > \frac{\Delta_{n,i}^\beta}{2}) \leq \frac{c\,\mathbb{E}\left[|\Delta_i X^c|^r\right]}{\Delta_{n,i}^{\beta r}} \leq c\Delta_{n,i}^{r(\frac{1}{2}-\beta)}, \tag{76}$$

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 (75), instead, we have to introduce the set

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

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

$$\mathbb{P}(N_{i,n}^c) \leq \sum_{j=1}^M \mathbb{P}(|\Delta_i N^{(j)}| \geq 1) \leq \sum_{j=1}^M \mathbb{E}[|\Delta_i N^{(j)}|] \leq cM\Delta_{n,i}. \tag{77}$$

On $N_{i,n}$, instead, $\forall j\ |\Delta_i N^{(j)}| = 0$ and so $(N_{i,n}) \cap \left\{|J_{t_i}| > \frac{\Delta_{n,i}^\beta}{2}\right\} = \emptyset$. It follows that the second term in the right hand side of (75) is

$$\begin{aligned} &c\mathbb{E}\left[\mathbb{1}_{\left\{|\Delta_i X|>\Delta_{n,i}^\beta, |J_{t_i}|>\frac{\Delta_{n,i}^\beta}{2}, N_{i,n}\right\}}\right] + c\mathbb{E}[\mathbb{1}_{\left\{|\Delta_i X|>\Delta_{n,i}^\beta, |J_{t_i}|>\frac{\Delta_{n,i}^\beta}{2}, N_{i,n}^c\right\}}] \\ &\leq c\mathbb{P}(N_{i,n}^c) \leq c\Delta_{n,i}. \end{aligned}$$

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

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

A.4. Proof of Lemma 3

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

$$\begin{aligned} &\mathbb{E}[|J_{t_i}|^q \varphi_{\Delta_{n,i}^\beta}^k(\Delta_i X)] \\ &= \mathbb{E}\left[|J_{t_i}|^q \varphi_{\Delta_{n,i}^\beta}^k(\Delta_i X)\mathbb{1}_{\{|J_{t_i}|>3\Delta_{n,i}^\beta\}}\right] + \mathbb{E}\left[|J_{t_i}|^q \varphi_{\Delta_{n,i}^\beta}^k(\Delta_i X)\mathbb{1}_{\{|J_{t_i}|\leq 3\Delta_{n,i}^\beta\}}\right]. \end{aligned} \tag{78}$$

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

$$\mathbb{E}\left[|J_{t_i}|^q \varphi^k_{\Delta_{n,i}^{\beta}}(\Delta_i X)\mathbb{1}_{\{|J_{t_i}|>3\Delta_{n,i}^{\beta}\}}\right] \leq \mathbb{E}\left[|J_{t_i}|^q \mathbb{1}_{\{|\Delta_i X^c|>\Delta_{n,i}^{\beta}\}}\right]$$

$$\leq \mathbb{E}[|J_{t_i}|^{qp_1}]^{\frac{1}{p_1}} \mathbb{E}\left[\mathbb{1}_{\{|\Delta_i X^c|>\Delta_{n,i}^{\beta}\}}\right]^{\frac{1}{p_2}} \leq c\Delta_{n,i}^{\frac{1}{p_1}} \Delta_{n,i}^{\frac{r}{p_2}(\frac{1}{2}-\beta)} \leq c\Delta_{n,i}^{r(\frac{1}{2}-\beta)-\varepsilon},$$

where we have used first of all Holder inequality and then Kunita inequality and (76). We remark it is possible to use Kunita inequality only for $qp_1 \geq 2$. However, as in the estimation here above the power of Δ is arbitrarily large, we can always choose p_1 such that, for any $q \geq 1$, qp_1 is bigger than 2. In order to evaluate the second term of (78), we introduce again the set $N_{i,n}$ defined in the proof of Lemma 2. 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 (77), we have

$$\mathbb{E}\left[|J_{t_i}|^q \varphi^k_{\Delta_{n,i}^{\beta}}(\Delta_i X)\mathbb{1}_{\{|J_{t_i}|\leq 3\Delta_{n,i}^{\beta}, N_{i,n}^c\}}\right] \leq c\Delta_{n,i}^{\beta q}\mathbb{P}(N_{i,n}^c) \leq c\Delta_{n,i}^{1+\beta q}.$$

By the arbitrariness of r it follows

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

as we wanted. □

A.5. Proof of Proposition 2

Proof. As the second point is useful in order to prove 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

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

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

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

Point 1 We analyse the behaviour of

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

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

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

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

$$\mathbb{E}[\tilde{A}_{t_i}^2] \leq c\sigma_1^4\Delta_{n,i} + \mathbb{E}[A_{t_i}^2\varphi_{\Delta_{n,i}^\beta}^2(\Delta_i X)] + c\sigma_1^2\Delta_{n,i}^{1-\tilde{\varepsilon}}, \tag{79}$$

for $\tilde{\varepsilon} > 0$ arbitrarily small. We are left to study $A_{t_i}^2\varphi_{\Delta_{n,i}^\beta}^2$. From its definition, recalling that φ is a bounded function, we obtain

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

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

$$I_1 \leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}^3\int_{t_i}^{t_{i+1}}\mathbb{E}[b^4(X_s)]ds \leq c\Delta_{n,i}^2. \tag{80}$$

On I_2 we use first of all Holder inequality. Then, on the first we use B.D.G. and Kunita inequalities, as in (69) and (70), while on the second the finite increments theorem, the boundedness of b' and the first point of Lemma 9:

$$\begin{aligned}
I_2 \quad & \leq \quad \frac{c}{\Delta_{n,i}^2}\mathbb{E}[(Z_{t_i} + J_{t_i})^8]^{\frac{1}{2}}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} b(X_s) - b(X_{t_i})ds\right)^4\right]^{\frac{1}{2}} \\
& \leq \quad \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}^{\frac{1}{2}}\Delta_{n,i}^{\frac{3}{2}}\mathbb{E}\left[\int_{t_i}^{t_{i+1}} c|X_s - X_{t_i}|^4 ds\right]^{\frac{1}{2}} \leq c\Delta_{n,i}.
\end{aligned} \tag{81}$$

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 9 will be once again useful.

$$I_3 \leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}\mathbb{E}\left[\int_{t_i}^{t_{i+1}} c|X_s - X_{t_i}|^2 ds\right] \leq c\Delta_{n,i}. \tag{82}$$

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

$$I_4 \leq c\mathbb{E}[b(X_{t_i})^4]^{\frac{1}{2}}\mathbb{E}[Z_{t_i}^4]^{\frac{1}{2}} \leq c\Delta_{n,i}. \tag{83}$$

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 (79) 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 To prove the estimations on the jumps gathered in the third point of Proposition 3 we repeatedly use Lemma 3. 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

$$\mathbb{E}[|E_{t_i}|\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X)] \leq c\mathbb{E}[|b(X_{t_i})||J_{t_i}|\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X)]$$

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

thus, because, as $\beta\in(0,\frac{1}{2})$, we can always find an $\varepsilon>0$ such that $\frac{1}{2}+\beta-\varepsilon>2\beta$. Hence, we set $1/q=1-\varepsilon$. It comes

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

In analogous way we obtain

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

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.6. Proof of Proposition 4

Proof. Point 1

Regarding the first point, we first 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 (20) we can denote

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

Replacing $\tilde{b}$ with b, we already know from (80), (81), (82) and (83) that

$$\mathbb{E}[\bar{I}_1^2 + \bar{I}_2^2 + \bar{I}_3^2 + \bar{I}_6^2] \leq c\Delta_{n,i}. \tag{84}$$

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

$$\begin{aligned}
\mathbb{E}[\bar{I}_4^2] &\leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}\int_{t_i}^{t_{i+1}} \mathbb{E}\left[(a^2(X_s) - a^2(X_{t_i}))^2 \left(\sum_{j=1}^M \lambda_s^{(j)}\right)^2\right] ds \\
&\leq \frac{c}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}} \mathbb{E}[(a^2(X_s) - a^2(X_{t_i}))^{2p}]^{\frac{1}{p}} \mathbb{E}\left[\left(\sum_{j=1}^M \lambda_s^{(j)}\right)^{2q}\right]^{\frac{1}{q}} ds \\
&\leq \frac{c}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}} \Delta_{n,i}^{\frac{1}{p}} ds \leq c\Delta_{n,i}^{1-\tilde{\varepsilon}},
\end{aligned} \tag{85}$$

where we have also used the boundedness of the moments of λ and set $1/p = 1 - \tilde{\varepsilon}$. On $\bar{I}_5$ we use that $a(x) \leq a_1$ and the second point of Lemma 9, getting

$$\mathbb{E}[\bar{I}_5^2] \leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}\int_{t_i}^{t_{i+1}} \sum_{j=1}^M \mathbb{E}[(\lambda_s^{(j)} - \lambda_{t_i}^{(j)})^2] ds \leq c\Delta_{n,i}. \tag{86}$$

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

$$\mathbb{E}[\bar{I}_7^2] \leq c\mathbb{E}[\tilde{b}(X_{t_i})^2 J_{t_i}^2] \leq \mathbb{E}[\tilde{b}(X_{t_i})^{2p}]^{\frac{1}{p}} \mathbb{E}[J_{t_i}^{2q}]^{\frac{1}{q}} \leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}, \tag{87}$$

where in the last inequality we have chosen p big and q next to 1. In particular, we have taken $1/q = 1 - \tilde{\varepsilon}$. From (84), (85), (86) and (87) 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

$$\mathbb{E}[\bar{I}_1^4 + \bar{I}_2^4 + \bar{I}_3^4 + \bar{I}_6^4] \leq c\Delta_{n,i}. \tag{88}$$

Acting as in (85) we get

$$\mathbb{E}[\bar{I}_4^4] \leq \frac{c}{\Delta_{n,i}^4}\Delta_{n,i}^3\int_{t_i}^{t_{i+1}} \mathbb{E}\left[(a^2(X_s) - a^2(X_{t_i}))^4 \left(\sum_{j=1}^M \lambda_s^{(j)}\right)^4\right] ds \tag{89}$$

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

where we have chosen $1/p = 1 - \tilde{\varepsilon}$. In the same way, acting as in (86) we obtain

$$\mathbb{E}[\bar{I}_5^4] \leq \frac{c}{\Delta_{n,i}^4}\Delta_{n,i}^3 \int_{t_i}^{t_{i+1}} \sum_{j=1}^{M} \mathbb{E}[(\lambda_s^{(j)} - \lambda_{t_i}^{(j)})^4] ds \leq c\Delta_{n,i}. \tag{90}$$

We conclude the proof of the point 2 by observing that

$$\mathbb{E}[\bar{I}_7^4] \leq c\mathbb{E}[\tilde{b}(X_{t_i})^{4p}]^{\frac{1}{p}} \mathbb{E}[J_{t_i}^{4q}]^{\frac{1}{q}} \leq c\Delta_{n,i}^{1-\tilde{\varepsilon}}, \tag{91}$$

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

$$\mathbb{E}_i[E_{t_i}^2] \leq \frac{c}{\Delta_{n,i}^2}\mathbb{E}_i[Z_{t_i}^2 J_{t_i}^2] + \frac{c}{\Delta_{n,i}^2}\mathbb{E}_i[J_{t_i}^4 + \left(\int_{t_i}^{t_{i+1}} a^2(X_{s^-}) \sum_{j=1}^{M} \lambda_s^{(j)} ds\right)^2]. \tag{92}$$

We show that the first term in the right hand side of the equation (92) is negligible if compared to the second one. By a conditional version of Holder, BDG and Kunita inequalities we get

$$\frac{c}{\Delta_{n,i}^2}\mathbb{E}_i[Z_{t_i}^2 J_{t_i}^2] \leq \frac{c}{\Delta_{n,i}^2}\mathbb{E}_i[Z_{t_i}^{2p}]^{\frac{1}{p}} \mathbb{E}_i[J_{t_i}^{2q}]^{\frac{1}{q}} \leq \frac{c}{\Delta_{n,i}^2}\Delta_{n,i}\Delta_{n,i}^{\frac{1}{q}} \leq c\Delta_{n,i}^{-\varepsilon}, \tag{93}$$

for any $\varepsilon > 0$, setting $1/q = 1 - \varepsilon$. To study the last term in the right hand side of (92) 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

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

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 9 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 (93) in (92) 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

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

$$\leq \frac{c}{\Delta_{n,i}^4}(\Delta_{n,i}^2\Delta_{n,i}^{1-\varepsilon} + \Delta_{n,i} + \Delta_{n,i}^3\Delta_{n,i}) \leq \frac{c}{\Delta_{n,i}^3}.$$

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.7. Proof of Lemma 4

Proof. It is

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

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 [21])

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

with α the coefficient of α-mixing and p, q and r such that $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1$. Using that

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

in our case the inequality here above becomes

$$|\mathrm{Cov}(f(X_{t_i}, \lambda_{t_i}), f(X_{t_j}, \lambda_{t_j}))| \leq c \|f(X_{t_i}, \lambda_{t_i})\|_p \left\|f(X_{t_j}, \lambda_{t_j})\right\|_q e^{-\frac{1}{r}\gamma|t_i - t_j|}.$$

From the definition of f and the boundedness of a and the existence of moments of λ we have

$$\begin{aligned}\|f(X_{t_i}, \lambda_{t_i})\|_p &\leq c \sum_{j=1}^{M} \left\|(\lambda_{t_i}^{(j)} - \mathbb{E}[\lambda_{t_i}^{(j)}|X_{t_i}])\right\|_{pp_1} \left\|\tilde{\psi}_l(X_{t_i})\right\|_{pp_2} \\ &\leq c \sum_{j=1}^{M} \left\|\tilde{\psi}_l(X_{t_i})\right\|_{pp_2},\end{aligned}$$

with p_1 and p_2 such that $\frac{1}{p_1}+\frac{1}{p_2}=1$. We remark that, as we are going to bound both the L^p and the L^q norm of $f(X_{t_i}, \lambda_{t_i})$, it seems natural to chose $p=q$, in order to repeat twice the same estimation. Then, as $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1$ and we need $r>0$, we obtain $p=q>2$. We can then chose, for $\varepsilon>0$ arbitrarily small, $pp_2 = 2+\varepsilon$. It implies $p_1 = \frac{2+\varepsilon}{2+\varepsilon-p}$, which leads us to chose $2<p=2+\tilde{\varepsilon}<2+\varepsilon$, for some $\tilde{\varepsilon}<\varepsilon$. Then, using that the L^2 norm of $\tilde{\psi}_l$ is smaller than 1 and that we can bound $\tilde{\psi}_l(x)$ by D_m, we obtain

$$\left\|\tilde{\psi}_l(X_{t_i})\right\|_{2+\varepsilon} \leq c \left\|\tilde{\psi}_l(X_{t_i})\right\|_{\infty}^{\varepsilon} \left\|\tilde{\psi}_l(X_{t_i})\right\|_2 \leq c D_m^{\varepsilon},$$

which provides

$$\|f(X_{t_i}, \lambda_{t_i})\|_p \leq c M D_m^{\varepsilon}.$$

In a similar way, it is easy to see that

$$\|f(X_{t_i}, \lambda_{t_i})\|_q \leq c M D_m^{\varepsilon}.$$

We now introduce a partition of $(0, T_n]$ (where T_n is the time horizon) 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

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

where the last inequality is a consequence of the following estimation: for each $k_1, k_2 \in \{0, \dots, n-1\}$ it is $|t_{k_1,h_1} - t_{k_2,h_2}| \geq |k_1 - k_2|\frac{T_n}{n} - \frac{2T_n}{n}$.

We remark here that, as we are considering the general case where the discretization step is not necessarily uniform, we can not replace $\frac{T_n}{n}$ with simply Δ_n: we have to keep it like this and compare it with $\Delta_{\min}$ and $\Delta_{\max}$, which is equal to Δ_n by definition.

Now we observe that the exponent does not depend on h anymore, hence the last term here above can be upper bounded by
$\frac{ce^{\frac{2\gamma}{r}\frac{T_n}{n}}}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} M_{k_1}M_{k_2}e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}}$.

Moreover, remarking that the length of each interval A_k is $\frac{T_n}{n}$, it is easy to see 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 $e^{\frac{2\gamma}{r}\frac{T_n}{n}} \leq e^{\frac{2\gamma}{r}\Delta_n} \leq c$.

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.8. Proof of Lemma 7

Proof. In order to estimate the probability of the complementary of the set Ω_B, as defined in (64), we first of all observe that

$\Omega_B^c \subset \cup_{j,k}\left\{\sup_{t\in\mathcal{B}_{m,m'}}|t(U^*_{k,j})| \geq \tilde{c}n^{\varepsilon_0}D^{\frac{1}{2}}\right\}$. Now we find an upper bound for the probability of Ω_B^c focusing on what happens for $j = 1$ and $k = 0$. Recalling the definition of $U^*_{j,k}$ in (62) and using that, as $t \in \mathcal{B}_{m,m'}$ whose dimension is D, $\|t\|_\infty \leq cD^{\frac{1}{2}}$, we can write, for any $\epsilon > 0$ arbitrarily small,

$$\begin{aligned}
&\mathbb{P}(\sup_{t\in\mathcal{B}_{m,m'}}|t(U^*_{0,1})| \geq \tilde{c}n^{\varepsilon_0}D^{\frac{1}{2}}) \leq \mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|B^*_{t_k} + C^*_{t_k} + E^*_{t_k}| \geq \tilde{c}n^{\varepsilon_0}\right) \\
&\leq \mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|B^*_{t_k}| \geq \frac{\tilde{c}}{3}n^{\varepsilon_0}\right) + \mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|C^*_{t_k}| \geq \frac{\tilde{c}}{3}n^{\varepsilon_0}\right) \qquad (94) \\
&+ \mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|E^*_{t_k}| \geq \frac{\tilde{c}}{3}n^{\varepsilon_0}\right).
\end{aligned}$$

From the definition of B it is

$$\frac{1}{q_n}\sum_{k=1}^{q_n}|B^*_{t_k}| \leq \frac{c}{q_n\Delta_n}\sum_{k=1}^{q_n} Z^2_{t_k} + c. \tag{95}$$

Moreover, using Markov inequality and the boundedness of σ,

$$\begin{aligned}\mathbb{P}\left(|Z_{t_k}| \geq c\sigma_1\Delta_n^{\frac{1}{2}}\log n\right) &= \mathbb{P}\left(e^{\frac{|Z_{t_k}|}{\sigma_1\sqrt{\Delta_n}}} \geq n^c\right) \leq \frac{1}{n^c}\mathbb{E}\left[e^{\frac{|Z_{t_k}|}{\sigma_1\sqrt{\Delta_n}}}\right] \\ &\leq \frac{1}{n^c}\mathbb{E}\left[e^{\frac{c'}{\Delta_n\sigma_1^2}\int_{t_k}^{t_{k+1}}\sigma^2(X_s)ds}\right] \leq \frac{c'}{n^c}. \end{aligned} \tag{96}$$

Therefore, as the constant c in (95) can be moved in the other side of the inequality in the first probability of (94) and so it turns out not being influential, the first probability of (94) is upper bounded by $\frac{q_n}{n^c}$, which is arbitrarily small. Concerning the second term of (94), we use Markov inequality and the fact that C has bounded moments. We get, $\forall r \geq 1$,

$$\mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|C^*_{t_k}| \geq \frac{\tilde{c}}{3}n^{\varepsilon_0}\right) \leq \sum_{k=1}^{q_n}\mathbb{P}\left(|C^*_{t_k}| \geq \frac{\tilde{c}}{3}n^{\varepsilon_0}\right) \leq c\sum_{k=1}^{q_n}\frac{\mathbb{E}[|C^*_{t_k}|^r]}{n^{r\varepsilon_0}} \leq \frac{cq_n}{n^{r\varepsilon_0}}.$$

Regarding the third term of (94) we observe that, replacing the value of q_n we get

$$\mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|E^*_{t_k}| \geq \frac{\tilde{c}}{3}n^{\varepsilon_0}\right) = \mathbb{P}\left(\sum_{k=1}^{q_n}|E^*_{t_k}| \geq \frac{\tilde{c}}{3}n^{\varepsilon_0}\frac{\log n}{\Delta_n}\right). \tag{97}$$

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

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

The right hand side of (97) is upper bounded by

$$\mathbb{P}\left(I_1 \geq \frac{\tilde{c}}{9}n^{\varepsilon_0}\frac{\log n}{\Delta_n}\right) + \mathbb{P}\left(I_2 \geq \frac{\tilde{c}}{9}n^{\varepsilon_0}\frac{\log n}{\Delta_n}\right) + \mathbb{P}\left(I_3 \geq \frac{\tilde{c}}{9}n^{\varepsilon_0}\frac{\log n}{\Delta_n}\right).$$

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 $\frac{\tilde{c}}{9}n^{\varepsilon_0}\frac{\log n}{\Delta_n}$ is arbitrarily small as a consequence of (96). $I_{1,2}$ is instead equal to I_2 and so it is enough to study such a term. From Markov, Holder, BDG and Kunita inequalities we have

$$\mathbb{P}\left(I_3 \geq \frac{\tilde{c}}{9}n^{\varepsilon_0}\frac{\log n}{\Delta_n}\right) \leq \frac{\mathbb{E}[(I_3)^r]}{(n^{\varepsilon_0}\log n\Delta_n^{-1})^r} \leq \frac{c\Delta_n^{-r}t^r_{q_n}}{(n^{\varepsilon_0}\log n\Delta_n^{-1})^r} \leq \frac{c}{n^{\varepsilon_0 r}},$$

where we underline that the order of t_{q_n} is $cq_n\Delta_n = c\frac{\log n}{\Delta_{\min}}\Delta_n \leq c\log n$. It is arbitrarily small. Concerning I_2, we want to estimate $\mathbb{P}(\sum_{k=0}^{q_n-1} J_{t_k}^2 \geq \frac{c}{9}n^{\varepsilon_0}\log n)$. We now consider two different possibilities, starting from the definition of the following set

$$A := \left\{\exists \tilde{k} \in \{0,\dots,q_n-1\} \text{ such that } J_{t_{\tilde{k}}}^2 \geq n^{\frac{\varepsilon_0}{2}}\right\}.$$

Then

$$\begin{aligned}&\mathbb{P}\left(\sum_{k=0}^{q_n-1} J_{t_k}^2 \geq \frac{c}{9}n^{\varepsilon_0}\log n\right)\\ &= \mathbb{P}\left(\sum_{k=0}^{q_n-1} J_{t_k}^2 \geq \frac{c}{9}n^{\varepsilon_0}\log n, A\right) + \mathbb{P}\left(\sum_{k=0}^{q_n-1} J_{t_k}^2 \geq \frac{c}{9}n^{\varepsilon_0}\log n, A^c\right).\end{aligned}$$

We observe that Markov inequality and Kunita inequality yield

$$\mathbb{P}\left(\sum_{k=0}^{q_n-1} J_{t_k}^2 \geq \frac{c}{9}n^{\varepsilon_0}\log n, A\right) \leq \mathbb{P}(A) \leq \sum_{k=0}^{q_n-1}\frac{\mathbb{E}[(J_{t_{\tilde{k}}})^{2r}]}{n^{\frac{\varepsilon_0 r}{2}}} \leq \frac{\Delta_n q_n}{n^{\frac{\varepsilon_0 r}{2}}} = \frac{c\log n}{n^{\frac{\varepsilon_0 r}{2}}},$$

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 $J_{t_k}^2 < n^{\frac{\varepsilon}{2}}$. 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

$$\begin{aligned}\mathbb{P}\left(\sum_{k=0}^{q_n-1} J_{t_k}^2 \geq \frac{c}{9}n^{\varepsilon_0}\log n, A^c\right) &\leq \mathbb{P}\left(\Delta N_q > \frac{c}{9}n^{\frac{\varepsilon_0}{2}}\log n\right) \leq c\frac{\mathbb{E}[(\Delta N_q)^r]}{(n^{\frac{\varepsilon_0}{2}}\log n)^r}\\ &\leq \frac{c\left(t_{q_n}^r + t_{q_n}\right)}{(n^{\frac{\varepsilon_0}{2}}\log n)^r} \leq \frac{c((\log n)^r + \log n)}{(n^{\frac{\varepsilon_0}{2}}\log n)^r} \leq \frac{c}{n^{\frac{\varepsilon_0 r}{2}}},\end{aligned}$$

where again we have used Markov inequality and we got a quantity arbitrarily small choosing $r \geq 1$ large enough. We put all the pieces together and we observe we can choose in particular r for which

$$\mathbb{P}\left(\frac{1}{q_n}\sum_{k=1}^{q_n}|E_{t_k}^*| \geq \frac{\tilde{c}}{3}n^{\varepsilon_0}\right) \leq \frac{c}{n^5}.$$

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

$$\mathbb{P}(\Omega_B^c) \leq \sum_{j=0,1;k\in\{1,\dots,p_n\}} \mathbb{P}\left(\sup_{t\in\mathcal{B}_{m,m'}}|t(U_{k,j}^*)| \geq \tilde{c}n^{\varepsilon_0}D^{\frac{1}{2}}\right) \leq \frac{cp_n}{n^5} \leq \frac{c}{n^4}. \quad \square$$

A.9. Proof of Lemma 8

Proof. We observe that for any $t \in \mathcal{B}_{m,m'}$, by (62) and Proposition 4, it is

$$\begin{aligned}\mathbb{E}[t(U^*_{0,0})^2] &= \mathrm{Var}(U^*_{0,0}) \\ &\leq \frac{4}{q_n^2}\sum_{l=1}^{q_n}\mathbb{E}\left[t^2(X^*_{t_l})\mathbb{E}_l[B^{*,2}_{t_l}+E^{*,2}_{t_l}]\right] \\ &\quad + 4\mathrm{Var}\left(\frac{1}{q_n}\sum_{l=1}^{q_n} t(X^*_{t_l})C^*_{t_l}\right) =: V_1 + V_2.\end{aligned}$$

By the second and the third points of Proposition 4, we can upper bound V_1 as

$$\begin{aligned}V_1 &\leq \frac{c}{q_n^2}\sum_{l=1}^{q_n}\mathbb{E}\left[t^2(X^*_{t_l})(\sigma_1^2 + \frac{a_1^4}{\Delta_n}\sum_{j=1}^{M}\lambda^{(j)}_{t_l})\right] \\ &\leq \frac{c}{q_n^2}\sum_{l=1}^{q_n}\mathbb{E}\left[t^{2p}(X^*_{t_l})\right]^{\frac{1}{p}}\mathbb{E}\left[\left(\sigma_1^2 + \frac{a_1^4}{\Delta_n}\sum_{j=1}^{M}\lambda^{(j)}_{t_i}\right)^q\right]^{\frac{1}{q}},\end{aligned}$$

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

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

From Assumption 4, $\|t\|_\infty^{\frac{2p-2}{p}} \leq cD_m^{\frac{2p-2}{2p}} \leq cD_m^\delta$, for any δ arbitrarily small, as p has been chosen next to 1. Using also the boundedness of the moments of λ it follows that

$$V_1 \leq \frac{D_m^\delta}{q_n^2}\frac{c}{\Delta_n}q_n = \frac{cD_m^\delta}{q_n\Delta_n}.$$

Using the same arguments as in the proof of Lemma 4, remarking that the sum over n is now replaced by the sum over q_n and that t now plays the same role as $\tilde{\psi}$, being such that its L^2 norm is smaller than 1 and it is bounded by D_m, we can show

$$V_2 \leq \frac{c}{q_n\Delta_n}\|t\|_\infty^{2\delta} \leq \frac{c}{q_n\Delta_n}D_m^{2\delta},$$

for any $\delta > 0$. It concludes the proof of the lemma. □

Acknowledgments

The authors would like to thank the anonymous referees for their helpful remarks that helped to improve the first version of the paper.

References

[1] Abraham, K. Nonparametric Bayesian posterior contraction rates for scalar diffusions with high-frequency data. Bernoulli 25 (2019), no. 4A, 2696–2728. MR4003562

[2] Amorino, C. and Gloter, A., 2020. Contrast function estimation for the drift parameter of ergodic jump diffusion process. Scandinavian Journal of Statistics, 47(2), pp.279–346. MR4157142

[3] Amorino, C. and Gloter, A., 2021. Invariant density adaptive estimation for ergodic jump–diffusion processes over anisotropic classes. Journal of Statistical Planning and Inference, 213, pp.106–129. MR4184676

[4] Amorino, C. and Gloter, A., 2020. Unbiased truncated quadratic variation for volatility estimation in jump diffusion processes. Stochastic Processes and their Applications, 130(10), pp.5888–5939. MR4140022

[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, 2475–2499. MR3054533

[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):198–229, 1982. MR0680660

[9] Birgé, L. and Massart, P. (1998) Minimum contrast estimators on sieves: exponential bounds and rates of convergence. Bernoulli, 4 (3) pp. 329–375. MR1653272

[10] Brémaud, P. Massoulié, L. (1996). Stability of nonlinear hawkes processes. The Annals of Probability pp. 1563–1588. MR1411506

[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. MR3646432

[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. MR2331262

[13] Comte, F. and Genon-Catalot, V. 2020 Nonparametric drift estimation for iid paths of stochastic differential equations. The Annals of Statistics, 48 (6), pp. 3336–3365 MR4185811

[14] Comte, F. and Marie, N. (2020) On a Nadaraya-Watson Estimator with Two Bandwidths. Preprint HAL hal-02457079. MR4255331

[15] Comte, F. and Merlevède, F. (2002) Adaptive estimation of the stationary density of discrete and continuous time mixing processes. ESAIM Probab. Statist., 6 pp. 211–238 (electronic). New directions in time series analysis (Luminy, 2001). MR1943148

[16] 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. MR2371524
[17] Delattre, S., Fournier, N., Hoffmann, M. (2016). Hawkes processes on large networks. The Annals of Applied Probability 26, 216–261. MR3449317
[18] DeVore, R. and Lorentz G. (1993). Constructive approximation. Springer Science & Business Media. MR1261635
[19] Dion, C., Lemler, S. (2019). Nonparametric drift estimation for diffusions with jumps driven by a Hawkes process. Statistical Inference for Stochastic Processes, 1–27. MR4136702
[20] Dion, C., Lemler, S., Löcherbach, E. (2019). Exponential ergodicity for diffusions with jumps driven by a Hawkes process. arXiv preprint arXiv:1904.06051. MR4136702
[21] Doukhan, P. (2012). Mixing: properties and examples (Vol. 85). Springer Science and Business Media. MR1312160
[22] Embrechts, P., Liniger, T. and Lu Lin. (2011) Multivariate hawkes processes: an application to financial data. Journal of Applied Probability. 48(A) 367–378. MR2865638
[23] Genon-Catalot, V., Jeantheau, T., and Larédo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. Bernoulli, 1051–1079. MR1809735
[24] E. Gobet. (2002). Lan property for ergodic diffusions with discrete observations. Ann. Inst. H. Poincaré Probab. Statist., 38(5), 711–737. MR1931584
[25] Goldenshluger, A. & Lepski., O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. The Annals of Statistics, 39(3) 1608–1632. MR2850214
[26] 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 MR3383817
[27] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. Stochastic processes and their Applications, 79(1), 135–163. MR1670522
[28] Jacod, J., Todorov, V. (2009). Testing for common arrivals of jumps for discretely observed multidimensional processes. The Annals of Statistics, 37(4), 1792–1838. MR2533472
[29] Jacod, J., Protter, P. (2011). Discretization of processes (Vol. 67). Springer Science and Business Media. MR2859096
[30] Krell, N., Schmisser, E. (2019). Nonparametric estimation of jump rates for a specific class of Piecewise Deterministic Markov Processes. arXiv preprint arXiv:1901.10166. MR4303887
[31] Lacour, C., Massart, P. and Rivoirard, V. (2017). Estimator Selection: a New Method with Applications to Kernel Density Estimation. Sankhya A 79, 298–335. MR3707423
[32] Le Gall, J. (2010). Calcul stochastique et processus de markov. Notes de cours.
[33] Mancini, C., and Reno, R. (2011). Threshold estimation of Markov models with jumps and interest rate modeling. Journal of Econometrics, 160(1), 77–92. MR2745869

[34] Neumann, M. H. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. Bernoulli, 15(1), 223–248. MR2546805

[35] Schmisser, E. (2012). Nonparametric estimation of the diffusion coefficient from noisy data. Statistical inference for stochastic processes, 15(3), 193–223. MR2979650

[36] Schmisser, E. (2019). Nonparametric estimation of the diffusion coefficients of a diffusion with jumps. Stochastic Processes and their Applications, 129(12), 5364–5405. MR4025707

[37] Strauch, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. The Annals of Statistics, 46(6B), 3451–3480. MR3852658

[38] Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. Probability theory and related fields, 107(4), 467–492. MR1440142