LOCAL POLYNOMIAL ESTIMATION OF THE INTENSITY OF A DOUBLY
STOCHASTIC POISSON PROCESS WITH BANDWIDTH SELECTION
PROCEDURE

THOMAS DESCHATRE

Abstract. We consider a doubly stochastic Poisson process with stochastic intensity \( \lambda_t = nq(X_t) \) where \( X \) is a continuous Itô semimartingale and \( n \) is an integer. Both processes are observed continuously over a fixed period \([0, T]\). An estimation procedure is proposed in a non-parametrical setting for the function \( q \) on an interval \( I \) where \( X \) is sufficiently observed using a local polynomial estimator. A method to select the bandwidth in a non-asymptotic framework is proposed, leading to an oracle inequality. If \( m \) is the degree of the chosen polynomial, the accuracy of our estimator over the H"older class of order \( \beta \) is \( n^{-\beta/2+1} \) if \( m \geq \lfloor \beta \rfloor \) and it is optimal in the minimax sense if \( m \geq \lfloor \beta \rfloor \). A parametrical test is also proposed to test if \( q \) belongs to some parametrical family. Those results are applied to French temperature and electricity spot prices data where we infer the intensity of electricity spot spikes as a function of the temperature.

Mathematics Subject Classification (2010): 60G55, 60J75, 62G05, 62G08, 62M86, 62P05.

Keywords: Doubly stochastic Poisson process, Non parametric estimation, Oracle inequality, Local polynomial estimator, Minimax optimality, Semimartingale, Dependence, Electricity prices, Temperature.

1. Introduction

1.1. Motivation. Jump processes and point processes are used in several domains such as finance, insurance or neuroscience, see [20] for more details. In finance, they allow to model discontinuities in equity prices time series and heavy tails in asset returns [33]. An application for insurance is the model of extreme events such that occurrence times of earthquakes [27]. In neuroscience, these processes model spikes which is a potential difference in the membrane of a neuron [34]. Spikes is a high increase of the potential in the membrane followed by a quick reversion to the initial level of the potential. They are also present in electricity spot prices time series and can be both negative and positive: price level is very high or very low during a short time period before coming back to its original level [5]. One way to model them is to use mean revert jump processes [9, 24].

In all these areas, the frequency of the jumps can be explained by an exogenous variable. Modeling these dependences can not be omitted, because they have an impact on risk management or prediction and help us to understand some behaviors. In [34], the author explains the neural spiking activity with three kind of covariates: the previous spikes, exogenous stimuli and concurrent neural activity. In [3], the authors propose a model for financial contagion. Financial contagion is the fact that a large price move in a market causes large price moves in other markets; jumps are explained by a first jump and in this case we often use multidimensional Hawkes process which are mutually exciting processes, see [3]. For electricity spot prices, spikes are often caused by abnormal temperatures which are not modeled by a jump process. In a general case, when the covariate is not an other jump process, we often use doubly stochastic Poisson processes, which are Poisson processes with stochastic intensity. Two of the most famous models are the Aalen multiplicative
model introduced in [1] where the intensity process is of the form $\alpha_t Y_t$ with $\alpha_t$ a function of time and $Y_t$ a stochastic process and the Cox regression model introduced in [13] where the intensity process is of the form $\alpha_t \exp (\beta^T Z)$ with $Z$ a multi-dimensional random variable. A non parametric version of the Cox model exists where the intensity is of the form $\alpha_t \exp (f(Z))$ [10]. These models are used mainly for life times modeling.

A large literature is dedicated on methods of estimation for intensity estimation of Poisson process, especially in a non parametric setting. In the case of inhomogeneous Poisson processes, [31] and [32] use projection estimators and model selection techniques. Several finite dimensional spaces called models are considered to find projection estimators and they propose a penalty criterion in order to select a model. They work in a non asymptotic framework and a concentration inequality is derived. Furthermore, minimax rates are found over several classes of functions. A different approach for the estimation of the intensity is the use of kernel methods as in [16] and [8]. In [16] and [8], asymptotic properties of the kernel estimator are studied; in [8], methods to select the bandwidth is proposed. In [39], the intensity can be stochastic and is also estimated as a function of time with a kernel estimator in an asymptotic framework. In the context of Cox and Aalen processes, [12] also proposes model selection techniques with projection estimators; local polynomial estimator, which is a generalization of kernel estimators, is proposed by [11] and studied in an asymptotic framework. A method of estimation in asymptotic framework for Cox regression with a time dependent covariable is established in [25]. Lasso penalization is proposed in [22] for Cox regression in the context of high dimensional covariate.

1.2. Objectives and results. In our case, we are interested in a doubly stochastic Poisson process denoted by $N$ where its intensity $\lambda$ is a function of an exogenous covariate which is a stochastic process $X_t$

$$\lambda_t = q(X_t)$$

and our goal is to estimate the function $q$. In this case, conditionally on $(X_s)_{s \geq 0}$, $(N_s)_{s \geq 0}$ is an inhomogeneous Poisson process. We assume that we observe $N$ and $X$ over a time horizon $[0, T]$. We can think of the example of the frequency of electricity spot prices spikes as a function of the temperature depending also on time.

This framework has already been studied. Indeed, [36] proposes a kernel estimator of the function $q$ in the case where $T$ goes to $\infty$ and when $X$ satisfies some asymptotical conditions, which can be for instance stationarity. A kernel estimator is also proposed in [26] where the covariates depends on time and asymptotic properties are studied in the case of n i.i.d. observations and $n$ goes to infinity. In the same context of i.i.d. observations with time dependent covariables, [28] proposes a non parametric estimator based on model selection and asymptotic properties are studied. In [14], $X_t$ corresponds to the fractional part of a Brownian motion and the doubly stochastic Poisson process is used to model the limit order book; an estimation procedure is proposed in an asymptotic framework.

We consider a different framework where $X$ is a continuous Itô semimartingale having a local time $l_T^x$ for $x$ in $\mathbb{R}$. This local time measures the time spend by $X$ around $x$ before time $T$ and verifies properties of Proposition 1. It existence and properties can be insured by low restrictions given in Assumption 1. The function $q$ can be estimated at point $x$ only if $X$ takes this value before time $T$, or if $l_T^x > 0$. We estimate $q$ on an arbitrarily interval $I$ and we work conditionally on the event

$$D(I, \nu) = \{ \omega \in \Omega, \inf_{x \in I} l_T^x (\omega) \geq \frac{\nu \nu}{|I|} \}$$
with \( \nu \in (0, 1] \). We choose to work in a non-parametric framework and in a non-asymptotic framework. To our knowledge, inferring the intensity of a doubly stochastic Poisson process as a function of a continuous Itô semimartingale is not present in the literature.

A local polynomial estimator \( \hat{q}_h \) of \( q \) is proposed in Section 3 with \( h \) a bandwidth parameter. The criteria used to evaluate the performance of our estimator is the \( L_2 \) norm on \( I, \| \cdot \|_I \). We also give a method to select a bandwidth over a finite set \( \mathcal{H} \). We adapt the method of \cite{21} used for density estimation with i.i.d. observations to our context of intensity estimation for doubly stochastic Poisson process. The method consists in approximating the bias of \( \hat{q}_h \) by an estimator of \( \| q_h - q_{h_{\text{min}}} \|_I^2 \) where \( q_h = \mathbb{E} \left( \hat{q}_h | \mathcal{F}_T^X \right) \), \( (\mathcal{F}_t^X)_{0 \leq t \leq T} \) is the natural filtration of \( X \) and \( h_{\text{min}} = \min \mathcal{H} \). Indeed, if \( h_{\text{min}} \) is sufficiently small, the bias of \( \hat{q}_{h_{\text{min}}} \) is negligible and \( \| q_h - q_{h_{\text{min}}} \|_I^2 \approx \| q - q_h \|_I^2 \). A biased estimator of \( \| q_h - q_{h_{\text{min}}} \|_I^2 \) is \( \| \hat{q}_h - \hat{q}_{h_{\text{min}}} \|_I^2 \). Correcting this bias and adding an estimator of the variance term, our method consists in choosing a bandwidth \( \hat{h} \) minimizing a criteria of the form

\[
\| \hat{q}_h - q_{h_{\text{min}}} \|_I^2 + \text{pen}_\alpha(h)
\]

where \( \text{pen}_\alpha(h) \) is a penalty function and \( \alpha > 0 \) a parameter chosen by the statistician weighting the variance. This method is an extension of the one of Goldenshluger and Lepski \cite{17}.

One of our main results is the oracle inequality of Proposition 6. If we write

\[
\lambda_t = n q(X_t)
\]

with \( n \in \mathbb{N}, n \geq 1 \), we obtain

\[
\mathbb{E} \left( \| q - \hat{q}_h \|_I^2 | D(I, \nu) \right) \leq \left( \alpha \vee \frac{1}{\alpha} + O \left( (\log(n)^{-1}) \right) \right) \min_{h \in \mathcal{H}} \mathbb{E} \left( \| q - \hat{q}_h \|_I^2 | D(I, \nu) \right) + O \left( \log(n) \mathbb{E} \left( \| q - q_{h_{\text{min}}} \|_I^2 | D(I, \nu) \right) \right) + O \left( \frac{\log(n \vee |\mathcal{H}|^6)}{n} \right).
\]

This inequality is asymptotically optimal when \( \alpha = 1 \).

Furthermore, if we consider that \( q \) belongs to the Hölder class on \( I \) with parameter \( \beta \), our estimator is optimal in the minimax rate sense and the minimax rate of convergence is \( n^{-\frac{\beta}{\beta + 2}} \) if the degree of the polynomial is larger than \( \lfloor \beta \rfloor \), see Proposition 7. In addition to give a method of estimation for the intensity as a function of \( X \), we have also shown that the method of \cite{21} is adapted to inhomogeneous Poisson process because the case \( X_t = t \) respects the different assumptions.

A second objective is to test if our function \( q \) belongs to some class of parametrical model. Indeed, non parametric estimators are not convenient for operational applications. This objective is achieved in Section 4 where a test is proposed. We test

\[ H_0 : q = q_0 \text{ on } I \text{ for some } \theta \in \Theta \]

against

\[ H_1 : q \neq q_0 \text{ on } I \text{ for all } \theta \in \Theta \]

where \( \Theta \in \mathbb{R}^d \). We consider the contrast \( M_n(\theta) \) defined in (11) that is an unbiased estimator of the distance between \( q \) and \( q_0 \) based on our estimator \( \hat{q}_h \) which is an estimator of \( q \) under both hypothesis. Under \( H_0 \), the minimum of the contrast gives an estimator of \( \theta, \hat{\theta}_n \), converging at the rate \( \sqrt{n} \) towards \( \theta \) when \( n \to \infty \) and a central limit theorem is provided, see Proposition 8 (i). Furthermore, the quantity \( \sqrt{n}M_n(\hat{\theta}_n) \) converges in law towards a normal random variable under
H₀, see Proposition 8 (ii), but to ∞ under H₁, see Proposition 8 (iii). This allows us to propose a critical region for the test.

In Section 5, our estimation procedure is applied on electricity prices and temperature data in order to model the dependence between the spikes frequency of electricity prices and the temperature. In Section 6, results on simulated data are given.

Proof of the oracle inequality is given in Section 7 and other proofs in Section 8.

2. Statistical setting

Let \( (X_t)_{0 \leq t \leq T} \) be a real valued continuous semimartingale of the form

\[
X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s
\]

defined on a filtered probability space \( (Ω, \mathcal{F}, (\mathcal{F}_s)_{0 \leq s \leq T}, \mathbb{P}) \) where \( (W_t)_{0 \leq t \leq T} \) is a standard Brownian motion, \( (b)_{0 \leq t \leq T} \) and \( (σ_t)_{0 \leq t \leq T} \) are càdlàg, progressively measurable and verify

\[
\int_0^T (|μ_s| + σ_s^2) ds < ∞ \text{ almost surely.}
\]

Let also consider the natural filtration of \( X, (\mathcal{F}_s^X)_{0 \leq t \leq T} \)\). Let \( (N_t)_{0 \leq t \leq T} \) be a doubly stochastic Poisson process with intensity \( (λ_t)_{0 \leq t \leq T} \) also defined on \( (Ω, \mathcal{F}, (\mathcal{F}_s)_{0 \leq s \leq T}, \mathbb{P}) \). We observe the two processes on \([0, T]\) with \( T \) finite. We assume that the intensity, which is the function of interest, is of the form

\[
λ_t = nq(X_t)
\]

where \( n \in \mathbb{N}, n \geq 1 \) corresponds to the asymptotic. As we observe the counting process on a finite time horizon, we need to have a sufficient number of jumps during this finite period, which is of order \( \int_0^T λ_u du \) and then in our case of order \( n \). We denote by \( Λ = \int_0^T λ_u du \) the compensator of \( N \) and by \( M = N − Λ \) the compensated Poisson process. Conditionally on \( \mathcal{F}_T^X \), \( N \) is an inhomogeneous Poisson process with deterministic intensity at time \( t \) \( nq(X_t) \). Our aim is to estimate the function \( (q(x), x ∈ I) \) for an arbitrary compact interval \( I \).

Assumption 1. We assume that one of the following assumptions is true:

(i) \( \inf_{0 \leq s \leq T} σ_s ≥ \underline{σ} > 0 \) a deterministic constant and

\[
\mathbb{E} \left( \int_0^T |μ_s| ds + \sup_{0 \leq t \leq T} \left| \int_0^t σ_s dW_s \right| \right) < ∞,
\]

(ii) \( X_t = t \) for all \( t \) in \([0, T]\).

Remark 1. Assumption 1 (ii) can be more general. These assumptions are sufficient but not necessary conditions for existence of a local time. In the case \( μ \) deterministic and \( σ = 0 \), we need for all \( 0 ≤ s ≤ T, μ_s ≠ 0 \) but only the case \( μ = 1 \) is interesting for us. Existence of local time with a stochastic drift and a null volatility could also be considered but existing results about local times for absolute continuous processes are not enough in the literature to consider it.

Proposition 1. Let \( X \) the process defined in (1). Under Assumption 1, there exists a function defined on \( \mathbb{R} \times [0, T] \) and denoted \( (x, t) → l^x_t \) verifying

(i) an occupation time formula of the form

\[
\int_0^t f(X_s) ds = \int_\mathbb{R} f(x) l^x_t dx, 0 ≤ t ≤ T
\]
for any measurable function $f$ on $\Omega \times \mathbb{R}$,

(ii) $\mathbb{E} \left( \sup_{x \in \mathbb{R}} l_T^x \right) < \infty$ and

(iii) $x \mapsto l_T^x$ is continuous on $\mathbb{R}$ under Assumption 1 (i) and has one point of discontinuity under Assumption 1 (ii).

As noticed in [18], the estimation of $q(x)$ at point $x \in I$ is meaningful only if the process $X$ hits the point $x$ before time $T$, or if $l_T^x > 0$. Indeed, $l_T^x$ is equal to

$$\lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^T 1_{|X_s - x| \leq \epsilon} ds$$

and measures the time spend by $X$ around the point $x$. For $\nu \in (0, 1]$, let us define the event

$$D(I, \nu) = \{ \omega \in \Omega, \ inf_{x \in I} l_T^x(\omega) \geq \nu T / |I| \}$$

with $|I|$ the Lebesgue measure of $I$. From now, we work conditionally on the event $D(I, \nu)$, and assuming $\mathbb{P}(D(I, \nu)) > 0$. Under (ii), if $X_t = t$, the natural choice of $I$ is $[0, T]$ and $\nu = 1$.

**Remark 2.** We choose $\nu$ being dimensionless, justifying the normalization by $|I|$. Indeed, the local time have dimension equal to time times the inverse of the dimension of $X$. Furthermore, if $I = \bar{[L, I]}$, and if we do the mapping $X_t' = \frac{X_t - L}{|I|}$ for $t \in [0, 1]$, we have $\inf_{x \in [0, 1]} l_T^x(X') = \frac{|I|}{T} \inf_{x \in I} l_T^x$ where $l_T^x(X')$ is the local time of $X'$ at time 1 and point $x$. As $\int_T l_T^x ds \geq \inf_{x \in I} l_T^x |I|$ and $\int_0^T l_T^x ds = \int_0^T 1_{X_s \in I} ds \leq T$, $\inf_{x \in I} l_T^x \leq \frac{T}{|I|}$ and $\nu$ has to be bounded by 1.

3. LOCAL POLYNOMIAL ESTIMATION

Let $m > 0$ be an integer, $K$ a kernel function and $\mathcal{S}^+_{m+1}$ the set of positive definite matrix of $\mathbb{R}^{(m+1) \times (m+1)}$. Let

$$K_h(u) = h^{-1}K \left( \frac{u}{h} \right), \ u \in \mathbb{R}, \ h > 0.$$ Let us consider the local polynomial estimator, for $h > 0$ and $x \in \mathbb{R}$,

$$\hat{\psi}_h(x) = \frac{1}{n} \int_0^T w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} dN_s$$

with

$$w(x, h, z) = U^T(0) B(x, h)^{-1} U(z) 1_{B(x, h) \in \mathcal{S}^+_{m+1}}, \ z \in \mathbb{R},$$

$$U(x) = \begin{pmatrix} 1, x, \frac{x^2}{2!}, ..., \frac{x^m}{m!} \end{pmatrix}^T,$$

and

$$B(x, h) = \int_0^T U \left( \frac{X_s - x}{h} \right) U^T \left( \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} ds.$$ If $B(x, h) \in \mathcal{S}^+_{m+1}$, the estimator $\hat{\psi}_h(x)$ is equal to $U^T(0) \hat{\psi}_h(x)$ with

$$\hat{\psi}_h(x) = \arg\min_{\theta \in \mathbb{R}^{m+1}} -\frac{2}{n} \theta^T \int_0^T U \left( \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} dN_s + \theta^T \int_0^T U \left( \frac{X_s - x}{h} \right) U^T \left( \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} ds \theta.$$
The case \( l = 0 \) corresponds to the classical Nadaraya Watson estimator. The term \( 1_{X_s \in I} \) allows us to avoid issues at the boundaries. We denote by \( q_h \) the conditional expectation of \( \hat{q}_h \) given \( F_h^n \):

\[
q_h (x) = \int_0^T w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} q (X_s) \, ds.
\]

If \( q \) is a polynomial function of degree \( m \) on \( I \), \( q_h (x) \) is equal to \( q (x) \), see Proposition 2.

**Proposition 2.** Let \( x \in I \) and \( h > 0 \) such that \( B (x, h) \in S_{m+1}^+ \) where \( B \) is defined in (4). Let \( Q \) be a polynomial function of degree \( \leq m \). For any realization of the process \( X \), we have

\[
\int_0^T w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} Q (X_s) \, ds = Q (x).
\]

In particular,

\[
\int_0^T w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} ds = 1,
\]

\[
\int_0^T w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} (X_s - x)^k ds = 0, \quad 1 \leq k \leq m.
\]

Proof of Proposition 2 is immediate noticing that \( \int_0^T w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} Q (X_s) \, ds \) is the first component of the quantity minimizing

\[
\int_0^T \left[ \theta^T U \left( \frac{X_s - x}{h} \right) - Q (X_s) \right]^2 K_h (X_s - x) 1_{X_s \in I} \, ds
\]

and using the Taylor’s expansion of a polynomial function; it is similar to the proof of [35, Proposition 1.12].

For \( B (x, h)^{-1} \) to be defined, the positive matrix \( B (x, h) \) must be definite. Assumption 2 is sufficient for \( B (x, h)^{-1} \) to be well defined on the event \( D (I, \nu) \), see Proposition 3.

**Assumption 2.** We assume that:

(i) there exists \( K_{\min} > 0 \) and \( \Delta > 0 \) such that \( K (u) \geq K_{\min} 1_{|u| \leq \Delta} \) for all \( u \) in \( \mathbb{R} \),

(ii) \( K \) has a compact support belonging to \([-1, 1]\) and \( \|K\|_\infty = \sup_{x \in \mathbb{R}} |K (x)| < \infty \).

**Proposition 3.** Let \( 0 < h \leq \frac{4}{3} \frac{|I|}{\Delta} \). Under Assumption 2, on the event \( D (I, \nu) \), the matrix \( B (x, h) \) defined in (4) belongs to \( S_{m+1}^+ \) and for \( x \in I \), \( z \in \mathbb{R} \),

\[
|w (x, h, z) 1_{|z| \leq 1}| \leq \frac{|I|}{A_K \nu T}
\]

with \( w \) defined in Equation (3) and \( A_K \) a constant depending on \( K \).

3.1. Method for bandwidth selection. Our objective is to propose a method in order to choose the bandwidth \( h \). We want for this bandwidth to minimize the \( L_2 \) loss on the interval \( I \), \( \mathbb{E} \left( \|q - \hat{q}_h\|_2^2 | D (I, \nu) \right) \), with \( \|f\|_2^2 = \int_I f (x)^2 \, dx \) for \( f \in L_2 (I) \) and let \( \cdot, \cdot >_I \) the associated scalar product. This loss is equal to the sum of a bias term, \( \mathbb{E} \left( \|q - \hat{q}_h\|_2^2 | D (I, \nu) \right) \), which depends on the regularity of \( q \) and is usually increasing with \( h \), and a variance term \( \mathbb{E} \left( \|\hat{q}_h - q\|_2^2 | D (I, \nu) \right) \), decreasing with \( h \). The theoretical bandwidth minimizing this quantity depends on the function \( q \) itself which is unknown and is called the oracle. One wants to find an estimator of this oracle, \( \hat{h} \), such that \( \mathbb{E} \left( \|q - \hat{q}_{\hat{h}}\|_2^2 | D (I, \nu) \right) \leq (1 + o (1)) \min_{h \in \mathcal{H}} \mathbb{E} \left( \|\hat{q}_h - q\|_2^2 | D (I, \nu) \right) + o (1) \) when \( n \to \infty \), in
order to have a loss with \( \hat{h} \) close to the minimal one; this type of inequality is called an oracle inequality. The usual method to find this estimator of the oracle, which is done in this section, is to find an unbiased estimator of the bias and of the variance then to consider the bandwidth \( h \) minimizing the sum of the two estimators.

In order to select the bandwidth parameter, we use the approach of [21] which is used for density estimation. We consider a finite set \( \mathcal{H} \) of \((0, \infty)\) and \( h_{\min} = \min \mathcal{H} \). The idea is to approximate the bias by \( \mathbb{E} (\|q_h - q_{h_{\min}}\|_I^2 | D(I, \nu)) \) with \( h_{\min} \) sufficiently small. A natural estimator of \( \mathbb{E} (\|q_h - q_{h_{\min}}\|_I^2 | D(I, \nu)) \) is

\[
\|\hat{q}_h - \hat{q}_{h_{\min}}\|_I^2.
\]

The estimator (5) induces a bias equal to the expectation of

\[
\frac{1}{n} \int_0^T \int_I \left( w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) - w \left( x, h_{\min}, \frac{X_s - x}{h_{\min}} \right) K_{h_{\min}} (X_s - x) \right)^2 1_{X_s \in I} dx_q (X_s) \, ds
\]

which can be estimated by the unbiased estimator

\[
\frac{1}{n^2} \int_0^T \int_I \left( w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) - w \left( x, h_{\min}, \frac{X_s - x}{h_{\min}} \right) K_{h_{\min}} (X_s - x) \right)^2 1_{X_s \in I} dx_q (X_s) \, ds
\]

This estimator can be written as

\[
\hat{V}_h + \hat{V}_{h_{\min}} - 2\hat{V}_{h, h_{\min}}
\]

where

\[
\hat{V}_h = \frac{1}{n^2} \int_0^T \int_I \left( w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) \right)^2 1_{X_s \in I} dx_q (X_s) \, ds
\]

and

\[
\hat{V}_{h, h_{\min}} =
\]

\[
\frac{1}{n^2} \int_0^T \int_I \int_I \left( w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) w \left( x, h_{\min}, \frac{X_s - x}{h_{\min}} \right) K_{h_{\min}} (X_s - x) \right)^2 1_{X_s \in I} dx_q (X_s) \, ds 1_{X_s \in I} dx_q (X_s) \, ds.
\]

An unbiased estimator of \( \mathbb{E} (\|q_h - q_{h_{\min}}\|_I^2 | D(I, \nu)) \) is then

\[
\|\hat{q}_h - \hat{q}_{h_{\min}}\|_I^2 - \hat{V}_h - \hat{V}_{h_{\min}} + 2\hat{V}_{h, h_{\min}}.
\]

An unbiased estimator of the variance which is equal to

\[
\mathbb{E} \left( \frac{1}{n} \int_0^T \int_I \left( w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) \right)^2 1_{X_s \in I} dx_q (X_s) \, ds \right) D(I, \nu)
\]

is given by

\[
\hat{V}_h = \frac{1}{n^2} \int_0^T \int_I \left( w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) \right)^2 1_{X_s \in I} dx_q (X_s) \, ds.
\]

In order to choose the bandwidth, we then use the criteria

\[
\|\hat{q}_h - \hat{q}_{h_{\min}}\|_I^2 + \text{pen}_\alpha (h)
\]

where

\[
\text{pen}_\alpha (h) = \alpha \hat{V}_h - \hat{V}_{h_{\min}} + 2\hat{V}_{h, h_{\min}} , \text{ with } \alpha > 0.
\]

The term \( \alpha > 0 \) is used to weight the variance term. The optimal bandwidth \( \hat{h} \) is given by

\[
\hat{h} = \underbrace{\arg\min_h}_{h \in \mathcal{H}} \|\hat{q}_h - \hat{q}_{h_{\min}}\|_I^2 + \text{pen}_\alpha (h).
\]
In the following, we want to derive an oracle inequality for the estimator \( \hat{q}_h \) which has not been done to our knowledge.

3.2. Concentration inequalities. In order to compute this oracle inequality, we first need the two following concentration inequalities, from \([32]\) and \([19]\). The concentration inequality of Proposition 4 is a weak Bernstein inequality, the one of Proposition 5 is an inequality for the Poisson U-statistic. These inequalities will be useful in our case because \( N \) is an inhomogeneous Poisson process conditionally on \( F^X \).

**Proposition 4.** [32, Equation (2.2)] Let \( T > 0 \). Let \( N \) be an inhomogeneous Poisson process with intensity \( \lambda, \Lambda = \int_0^T \lambda_u du \) and \( M = N - \Lambda \). For all \( u \geq 0 \), with probability larger than \( 1 - e^{-u} \)

\[
\int_0^T f(s) dM_s \leq \sqrt{2u} \int_0^T f^2(s) d\Lambda_s + \sup_{x \in [0,T]} |f(x)| u
\]

**Proposition 5.** [19, Theorem 4.2] Let \( T > 0 \). Let \( N \) be an inhomogeneous Poisson process with intensity \( \lambda, \Lambda = \int_0^T \lambda_u du \) and \( M = N - \Lambda \). For all \( u, s \geq 0 \), with probability larger than \( 1 - 2.77e^{-u} \)

\[
\int_0^T \int_0^s f(u,s) dM_u dM_s \leq 2(1 + \epsilon)^{\frac{2}{3}} C \sqrt{u} + 2\eta(\epsilon) Du + \beta(\epsilon) Bu^{\frac{3}{2}} + \gamma(\epsilon) Au^2
\]

where

\[
\eta(\epsilon) = \sqrt{2\kappa(2 + \epsilon + \epsilon^{-1})}, \quad \beta(\epsilon) = e(1 + \epsilon^{-1})^2 \kappa(\epsilon) + \left(\sqrt{2\kappa(2 + \epsilon + \epsilon^{-1})}\right)^{\frac{1}{2}}(1 + \epsilon)^{\frac{2}{3}}
\]

\[
\gamma(\epsilon) = \left(e(1 + \epsilon^{-1})^2 \kappa(\epsilon)\right)^{\frac{1}{2}} \kappa = 6, \quad \kappa(\epsilon) = 1.25 + \frac{32}{\epsilon}
\]

and

\[
A = \sup_{(u,s) \in [0,T]^2} f(u,s), \quad B^2 = \max\{\sup_{s \leq T} \int_0^s f(u,s)^2 d\Lambda_u, \sup_{u \leq T} \int_0^T f(u,s)^2 d\Lambda_s\},
\]

\[
C^2 = \int_0^T \int_0^s f(u,s)^2 d\Lambda_u d\Lambda_s, \quad D = \sup_{a \in I} a^2 d\Lambda_u = 1, \int_0^T b^2 d\Lambda_s.
\]

3.3. Oracle inequality. For a function \( f \in L_\infty(I) \), we denote by \( \|f\|_{L_\infty,I} \) the norm \( \sup_{x \in I} |f(x)| \). We will also need for the kernel the following norms: \( \|\cdot\|_1, \|\cdot\|_2 = \|\cdot\|_2 \) and \( \|\cdot\|_\infty \) corresponding respectively to the \( L_1, L_2 \) and \( L_\infty \) norms on \( \mathbb{R} \), with the \( L_p \) norm defined by \( \|f\|_p = \left( \int_{\mathbb{R}} |f(x)|^p \right)^{\frac{1}{p}} \), \( p \geq 1 \) and \( \|f\|_{L_\infty} = \sup_{x \in \mathbb{R}} |f(x)| \). Proposition 6 gives an oracle inequality for \( \hat{q}_h \).

**Proposition 6.** Assume 1 and 2. Let \( x \geq 1, \epsilon \in (0,1) \). Let \( H \) a finite subset of \( (0,\infty) \) such that \( \min H = h_{\min} \geq \frac{\|K\|_\infty\|K\|_1|I|}{n} \) and max \( H \leq \frac{2|I|}{3\Delta} \). Let \( \hat{q}_h \) the local polynomial estimator defined in \( (2) \) and \( h \) the bandwidth defined in \( (8) \). With conditional probability given \( F^X \) larger than \( C_1|H|e^{-x} \), on the event \( D(I,\nu) \),

\[
\|\hat{q}_h - q\|^2 \leq C_0(\epsilon,\alpha) \min_{h \in H} \|\hat{q}_h - q\|^2 + C_2(\epsilon,\alpha) \|q_{h_{\min}} - q\|^2
\]

\[
+ C_3(\epsilon,\kappa,\alpha)|I| \left( \frac{(1 + \|q\|_{L_\infty}|I|_T^3 |I|_T}|I|) x^2}{n} + \frac{x^3|I|}{n^2 h_{\min}} \right)
\]
where $C_0 (\epsilon, \alpha) = \alpha + \epsilon$ if $\alpha \geq 1$ and $C_0 (\epsilon, \alpha) = \frac{1}{\epsilon} + \epsilon$ if $0 < \alpha < 1$, $C_1$ is a constant, $C_2 (\epsilon, \alpha)$ is a constant depending only on $\epsilon$ and $\alpha$ and $C_3 (\epsilon, K, \alpha)$ is a constant depending only on $\epsilon$, $K$ and $\alpha$. Furthermore, $C_2 (\epsilon, \alpha) \asymp \frac{1}{\epsilon}$ and $C_3 (\epsilon, K, \alpha) \asymp \frac{1}{\epsilon^2}$ when $\epsilon \to 0$.

We also have,

\[
\mathbb{E} (\| \hat{q}_h - q \|_D^2 | D (I, \nu) \|) \leq \left( \alpha \vee \frac{1}{\alpha} + \frac{\tilde{C}_1}{\log (n)} \right) \min_{\| \hat{h} \|_{\mathcal{H}}} \mathbb{E} (\| \hat{q}_h - q \|_D^2 | D (I, \nu) \|) + \tilde{C}_2 (\alpha) \log (n) \mathbb{E} (\| q_{\min} - q \|_D^2 | D (I, \nu) \|) \\
+ \frac{\tilde{C}_3 (K, \alpha) |I|}{\nu^2 T^2} \left( (1 + \| q_{l, \infty} \|_{l, \infty} \mathbb{E} (\| D (I, \nu) \| |I|) \|) \log (n \vee |\mathcal{H}|)^5 + \log (n \vee |\mathcal{H}|)^6 \right) \\
+ \frac{\tilde{C}_4 \| q \|_T^2}{n^4} + \frac{\tilde{C}_5 (K) |I|}{\nu^2 T^2} \sum_{i=1}^4 \left( \| q \|_{l, \infty, lT} \| \right) \frac{1}{n},
\]

where $\tilde{C}_1$ and $\tilde{C}_4$ are constant, $\tilde{C}_2 (\alpha)$ is a constant that only depends on $\alpha$, $\tilde{C}_3 (K, \alpha)$ is a constant that only depends on $K$ and $\alpha$ and $\tilde{C}_5 (K)$ is a constant that only depends on $K$.

In inequality (10), one can see the presence of an error of order $\log (n) \mathbb{E} (\| q_{\min} - q \|_D^2 | D (I, \nu) \|$.

This error comes from the approximation of the bias $\| q_h - q \|_D^2$ by $\| q_h - q_{\min} \|_D^2$ and is negligible if $h_{\min}$ is small enough and $q$ regular enough. We also remark that the oracle inequality is asymptotically optimal when $\alpha = 1$.

3.4. Adaptive minimax estimation. In this section, we study the performance of the estimator $\hat{q}_h$ in terms of convergence rate. We now work with the asymptotic $n \to \infty$, meaning the number of jumps becomes large when $n \to \infty$. For $\rho, \beta, L > 0$, let $\Lambda_{\rho, \beta} = \{ f : I \to \mathbb{R} : f (x) \geq \rho, \| f \|_{l, \infty} < \infty \} \cap \Sigma (\beta, L, I)$ where $\Sigma (\beta, L, I)$ is the Hölder class on $I$ defined as the set of $l = [\beta]$ differentiable functions $f : I \to \mathbb{R}$ whose derivative $f^{(l)}$ verifies

\[
|f^{(l)} (x) - f^{(l)} (x')| \leq L |x - x'|^{\beta - l}, \forall x, x' \in I,
\]

see [35, Definition 1.2]. We will restrict to the study of $q \in \Lambda_{\rho, \beta}$.

To evaluate the performance of an estimator $\hat{q}_n$ of $q$, we consider the minimax risk

\[
\mathcal{R} (\hat{q}_n, \Lambda_{\rho, \beta}, \varphi_n) = \sup_{q \in \Lambda_{\rho, \beta}} \mathbb{E} \left( \varphi_n^{-2} \int_I \left( \hat{q}_n (N, X, x) - q (x) \right)^2 dx | D (I, \nu) \right).
\]

An estimator $\hat{q}_n^*$ is said to attain an optimal rate of convergence $\varphi_n (\Lambda_{\rho, \beta})$ if

\[
\lim_{n \to \infty} \sup \mathcal{R} (\hat{q}_n^*, \Lambda_{\rho, \beta}, \varphi_n (\Lambda_{\rho, \beta})) < \infty
\]

and no estimator can attain a better rate:

\[
\lim_{n \to \infty} \inf \inf \mathcal{R} (\hat{q}_n, \Lambda_{\rho, \beta}, \varphi_n (\Lambda_{\rho, \beta})) > 0
\]

where the infimum is taken over all estimators.

**Proposition 7.** Assume 1 and 2. Let us consider the set of bandwidth $\mathcal{H} = \{ h > 0 | h \geq \frac{\| I \|}{n}, h \leq \frac{1}{n}, I \in \mathbb{N} \}$. Let $\hat{q}_h$ the local polynomial estimator defined in Proposition 6 and let $m$ be the degree of the corresponding polynomial. In the case where $m \geq [\beta]$, $\hat{q}_h$ is optimal in the minimax sense and the optimal rate of convergence is given by $\varphi (\Lambda_{\rho, \beta}) = \frac{1}{n^{\frac{[\beta]}{2}}}$.

In the case where $m < [\beta]$, the rate of convergence of $\hat{q}_h$ is $\frac{1}{n^{\frac{m}{2m+1}}}$.
4. TEST FOR A PARAMETRIC FAMILY

Let us consider the parametric family $P = \{g_{\theta} (\cdot), \theta \in \Theta\}$ with $\Theta$ a subset of $\mathbb{R}^d$, $d \geq 1$. Our objective is to test if the intensity function $q$ belongs to $P$. Let us consider the two hypothesis:

$$
\begin{align*}
H_0 & : \exists \theta_0 \in \Theta, \ q (\cdot) = g_{\theta_0} (\cdot) \\
H_1 & : q \notin P
\end{align*}
$$

We want to test $H_0$ against $H_1$. Under both hypothesis, one way to estimate $q$ is to use the local polynomial estimator $\hat{q}_h$. As we work in an asymptotic framework, we denote by $h_n$ the optimal bandwidth $\hat{h}$. Under $H_0$, to estimate the parameter $\theta$, let us consider the contrast

$$
M_n (\theta) = \| \hat{q}_h (\cdot) - \int_0^T w (\cdot, h_n, \frac{X_s - \cdot}{h_n}) K_{h_n} (X_s - \cdot) 1_{X_s \in I} g_{\theta} (X_s) ds \|_T^2
$$

(11)

$$
- \frac{1}{n^2} \int_0^T \int_0^T w^2 (x, h_n, \frac{X_s - x}{h_n}) K_{h_n}^2 (X_s - x) 1_{X_s \in I} dx.
$$

The contrast defined in (11) is similar to the one in [2], used in the case of the estimation of the drift and the volatility of an Itô diffusion. However, in [2], the norm is weighted by the density of $X$, that is $L^2$ is our case. As it is important to have a good estimate of $q$ everywhere on $I$, the norm is not weighted in (11). The second term in the right hand side of (11) is a correction of the bias. We can also notice that we use the function $\int_0^T w (\cdot, h_n, \frac{X_s - \cdot}{h_n}) K_{h_n} (X_s - \cdot) 1_{X_s \in I} g_{\theta} (X_s) ds$ and not directly $g_{\theta} (\cdot)$ in order to eliminate the bias term $\int_0^T w (\cdot, h_n, \frac{X_s - \cdot}{h_n}) K_{h_n} (X_s - \cdot) 1_{X_s \in I} (g_{\theta} (X_s) - g_{\theta} (\cdot)) ds$ that would appear and then avoid us to make assumptions on the speed of convergence of this term which depends on the regularity of $g_{\theta}$. An estimator of $\theta$ under $H_0$ is

$$
\hat{\theta}_n = \inf_{\theta \in \Theta} M_n (\theta).
$$

Under classical Assumption 3, this estimator is consistent at a speed rate of $\sqrt{n}$, see Proposition 8 (i). The idea of the test is that under $H_0$, $M_n (\hat{\theta}_n)$ is close to $M (\theta_0)$ which is equal to 0. The rate of convergence is of order $n \sqrt{h_n}$, see Proposition 8 (ii). However, under $H_1$, $M_n (\hat{\theta}_n)$ converges to $\inf_{\theta \in \Theta} \| q - g_{\theta} \|_T^2$ which is different from 0 and then $n \sqrt{h_n} M_n (\hat{\theta}_n)$ goes to $\infty$, see Proposition 8 (iii).

**Assumption 3.** We assume that

(i) The set $\Theta$ is compact in $\mathbb{R}^d$.

(ii) For some $M = M_I > 0$,

$$
\sup_{x \in I} |g_{\theta_1} (x) - g_{\theta_2} (x)| \leq M ||\theta_1 - \theta_2||_d \text{ for } \theta_1, \theta_2 \in \Theta
$$

where $\| \cdot \|_d$ is the $\mathbb{R}^d$ Euclidian norm.

(iii) For all $x$ in $I$, $\theta \mapsto g_{\theta} (x)$ is three times continuously differentiable. Furthermore, $x \mapsto g_{\theta} (x)$, $x \mapsto \partial_{\theta} g_{\theta} (x)$ are continuous on $I$ and $\frac{\partial_i \partial_{\theta} g_{\theta}}{\partial_{\theta_i} \partial_{\theta_j}}$, $\frac{\partial_i \partial_j \partial_{\theta} g_{\theta}}{\partial_{\theta_i} \partial_{\theta_j} \partial_{\theta_k}}$ are bounded on $I$ for $i, j, k = 1, ..., d$ for all $\theta \in \Theta$.

(iv) For some $\eta = \eta_I > 0$,

$$
\inf_{(x, \theta) \in I \times \Theta} \lambda_{\min} \left( \partial_{\theta} g_{\theta} (x) \partial_{\theta} g_{\theta} (x)^T \right) \geq \eta
$$

where $\lambda_{\min} (A)$ is the smallest eigenvalue of a matrix $A$.

(v) The equality $g_{\theta_1} = g_{\theta_2}$ on $I$ implies $\theta_1 = \theta_2$. 
(vi) $h_n \to 0$ and $n\sqrt{h_n} \to \infty$.
(vii) The intensity function $q$ is continuous on $I$ and the kernel function $K$ is continuous on $\mathbb{R}$.

**Proposition 8.** Let $M_n(\theta)$ and $\hat{\theta}_n$ defined respectively by (11) and (12). We work under Assumption 1, Assumption 2 and Assumption 3. On the event $D(I, \nu)$, conditionally on $F^X_T$, under $H_0$,

(i) 
\[ \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathcal{N} \left( 0, \left( \int_I \partial_{\theta} g_{\hat{\theta}_0}(u) \partial_{\theta} g_{\hat{\theta}_0}(u)^T du \right)^{-1} \int_I \partial_{\theta} g_{\hat{\theta}_0}(u) \partial_{\theta} g_{\hat{\theta}_0}(u)^T g_{\hat{\theta}_0}(u) du \right), \]

(ii) 
\[ n\sqrt{h_n} M_n(\hat{\theta}_n) \xrightarrow{D} \mathcal{N} \left( 0, 2 \int_{\mathbb{R}} \left( \int_{\mathbb{R}} w(u) w(u+p) K(u) K(u+p) du \right)^2 dp \int_I \left( g_{\hat{\theta}_0}(y) \right)^2 dy \right) \]
with
\[ w(u) = U^T(0) \left( \int_{\mathbb{R}} U(z) U^T(z) K(z) dz \right)^{-1} U(u) \]

and under $H_1$,

(iii) 
\[ |n\sqrt{h_n} M_n(\hat{\theta}_n)| \xrightarrow{p} \infty. \]

In order to test the null hypothesis at level $\gamma \in (0, 1)$, we reject $H_0$ when
\[ |M_n(\theta_n)| \geq \hat{c}(\gamma) = n^{-1} h_n^{-\frac{1}{2}} \sqrt{V_n} \Phi^{-1} \left( 1 - \frac{\gamma}{2} \right) \]
where
\[ V_n = A(K) \int_I \left( \int_{\mathbb{R}} \frac{g_{\hat{\theta}_0}(y)}{K_{h_n}(y - X_s) 1_{X_s \in I}} dy \right)^2 dy, \]
\[ A(K) = 2 \left( \int_{\mathbb{R}} K(u) du \right)^2 \int_{\mathbb{R}} \left( \int_{\mathbb{R}} w(u) w(u+p) K(u) K(u+p) du \right)^2 dp \]
only depends on $K$ and $\Phi$ is the cumulative distribution function of a $\mathcal{N}(0, 1)$ random variable.

5. **Dependence between the frequency of electricity spot spikes and temperature**

5.1. **Data.** We dispose of
- the hourly French EPEX spot price between the first January of 2007 and the first of January 2017 not included,
- the hourly French temperature, which is an spatial average of the temperature over 32 cities, between the first January of 2007 and the first of January 2017 not included.

The times series are given in Figure 1 for the year 2010.
5.2. Detection of the jumps. In the spot price time series, we observe spikes that are characteristic of the electricity spot market. A spike can be defined as a jump with a strong mean reversion. We then assume that the spot price $S_t$ has the following dynamic:

$$S_t = Y_t + Z_t$$

with $Y_t$ a continuous Itô semi-martingale and

$$Z_t = \int_0^t \int_\mathbb{R} xe^{-\beta(t-s)} p dt \, dx$$

with $p$ a Poisson measure on $\mathbb{R}^+ \times \mathbb{R}$ with compensator $q = \lambda_t dt \otimes \nu(dx)$. $Z$ has the dynamic

$$dZ_t = -\beta Z_t dt + \int_\mathbb{R} xp dt \, dx$$

corresponding to a mean reverting Poisson process and models the spikes. Let us consider $N$ the Poisson process associated to the jump times of $X$. $N$ has intensity $(\lambda_t)_{0 \leq t \leq T}$ which is the intensity we want to estimate as a function of the temperature. In order to detect the jumps, we use the method of [15] with a threshold equal to $5\bar{\Delta}^{0.49}$ where $\bar{\Delta}$ is the frequency of observations and $\Delta$ is the multipower variation estimator of order 20. We keep only the increments verifying $\Delta_i S_i S_{i+1} S < 0$ as jumps with $\Delta_i S = S_t - S_{t-1}$. When the frequency of observations $\Delta$ goes to 0 and when $\beta$ is large enough, [15] proves that this filtering allows to detect with probability one every spikes under some asymptotic conditions. The data are segmented in periods of one year for the detection of the jumps in order to avoid too much change in the volatility. In [15], the intensity of the Poisson process is constant. Assuming that $\lambda$ is bounded below and above, the results can easily be extended to the case where $\lambda$ is stochastic. Jump times are represented in Figure 1. In the following, we consider that we observe $N$ but we are aware that we only have an estimator of it.

5.3. Dependence with temperature. In this section, we estimate the intensity of the jump process as a function of the temperature, using the method of Section 3. In addition to the statistic interest, quanto options are financial options with temperature and spot price as underlying. They can be used for instance to hedge both volume and price risks. In order to price these options, it is necessary to capture the dependence between the temperature and the spot price. More details
are given in [6] where the dependence between the two is only modeled by a correlation and the spikes are not represented.

The temperature is illustrated in Figure 1 along with the jump times. The spikes seems to happen more often for low temperatures. The observed temperature belongs to the interval $[-8.50, 33.95]$. The temperature is not observed continuously but because of the high frequency of the data and the long range of observation, we pretend that the error due to the discretization is negligible. One wants to estimate the intensity of the spike process as a function of the temperature on the interval $[-5, 33]$ where the temperature is sufficiently observed. To estimate the intensity function, we consider the Epanechnikov kernel $K(u) = \frac{3}{4} (1 - u^2) 1_{|u|\leq 1}$ and the local polynomial estimator with degree 1 considered in Section 3. We choose $h_{\min}$ equal to $\frac{1}{\sqrt{N_I}} \|K\|_1 \|K\|_\infty = 0.13$, where $\|K\|_1 = 1$, $\|K\|_\infty = \frac{3}{4}$ and $N_I = 219$ is the number of jumps in the interval $I$. The tuning parameter of the estimation procedure $\alpha$ is chosen equal to 1. The optimal bandwidth is selected among the set $H = \{h = h_{\min} + 0.1i, \ h \leq 25\}$. The minimum of the criteria is achieved for $\hat{h} = 8.73$ and the estimator for this value of $h$ is given in Figure 2. The estimator takes small negative values for high temperatures, which is caused by the total absence of jumps in this area; one can take the maximum between the estimator and a small positive value for the intensity.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Local polynomial estimator (in the case $\alpha \in \{0.5, 0.75, 1, 1.25, 1.5, 1.75\}$), parametric estimator and maximum likelihood parametric estimator (MLE) of the intensity as a function of the temperature.}
\end{figure}

This result confirms our intuition: spikes happen more often when temperature is low. We now want to test the hypothesis that the intensity is a quadratic function of the temperature:

$q(x) = a_0 \exp(a_1 x)$ for $x \in I$

with $a_0 > 0$, $a_1 < 0$. The constant

$\int_{\mathbb{R}} \left( \int_{\mathbb{R}} w(u) w(u + p) K(u) K(u + p) du \right)^2 dp = \frac{413113}{985600}$
is needed for the test. We find that the null hypothesis is not rejected for a level of confidence at 95% (with a p value equals to 0.083) and that the estimated parameters are \((\hat{a}_0, \hat{a}_1) = (1033.8, -0.2)\). Figure 2 includes the parametric estimator of the intensity as a function of the temperature in the case \(\alpha = 1\). The maximum likelihood estimator (MLE) of the parametric model \(q(x) = a_0 \exp(a_1 x)\) is equal to \((1035.7, -0.17)\) and gives a similar estimator for \(q\), see Figure 2. However, the MLE estimator of \(q\) is closer to the local polynomial estimator than our parametric estimator; this can be caused by the weight introduced in the norm in (11) for \(g_\theta\). In a second time, we test if the
intensity function is independent from the temperature, corresponding to $q$ constant: the test is rejected for a level of confidence at 95% (with a $p$ value equals to 0).

To study the sensibility to the choice of $\alpha$, we perform our estimation procedure for $\alpha \in \{0.25, 0.5, 0.75, 1.25, 1.5, 1.75\}$. Results remains the same for $\alpha \in \{0.5, 0.75, 1.25, 1.5, 1.75\}$: $\hat{h} = 8.73$. However, it differs significantly when $\alpha = 0.25$: $\hat{h} = 0.23$ which is closed to $h_{\text{min}}$ leading to a high bias, which is consistent with a small value of $\alpha$, see Figure 3 for the estimator of the intensity considering this value of $h$. The criteria to minimize in order to find $\hat{h}$ is given in Figure 4 for the different values of $\alpha$. Similar results than [21, Theorem3] can be derived in our context: the minimum penalty we can consider is achieved for $\alpha = 0$, and for $\alpha < 0$ the criteria leads to a value of $h$ closed to $h_{\text{min}}$ with high probability and then induces a high bias. $\alpha = 0.25$ is positive but more likely to produce an under-biased estimator, which has happened in our case. The question of the choice of $\alpha$ remains open but a value of $\alpha$ since the asymptotical one, that is $\alpha = 1$, seems satisfying. This choice is also supported by [37] (corresponding paper in progress) in the case of density estimation where numerical experiments are performed: the tuning parameter can be chosen equals to 1 without impact on the density estimation.

6. Numerical results

In order to evaluate the performance of our estimation procedure, we present some simulation results. To be consistent with data, let us consider a model reproducing the temperature data and the spike times.

As in [7], we model the temperature $\theta_t$ as the sum of a trend seasonality function

$$\Gamma_t = a + bt + c_1 \sin \left( \frac{2\pi t}{365 \times 24} \right) + c_2 \sin \left( \frac{2\pi t}{24} \right)$$

corresponding to yearly and daily seasonality and a diffusion $X_t$ having dynamics

$$dX_t = -\theta X_t dt + \sigma dW_t$$

where $W_t$ is a standard Brownian motion. In [7], the temperature is modeled by a CARMA process with stochastic seasonal volatility but for simplicity we consider the simplest one corresponding to an Ornstein Uhlenbeck process. Using classical estimation procedures, we find $a = 12.06, b = 0.0000072, c_1 = 7.81, c_2 = -3.18, r_1 = -16924.50, r_2 = 10.84, \theta = 0.011, \sigma = 0.46$.

The spike intensity $\lambda_t$ is considered as an exponential function of the temperature $\theta_t$

$$\lambda_t = a_0 \exp (a_1 \theta_t)$$

with $a_0 = 1033.8$ and $a_1 = -0.2$. As the quality of the estimation procedure depends on the interval of estimation, we consider three intervals: $[-1, 29], [-3, 31]$ and $[-5, 33]$. As for the estimation on data, we consider a local polynomial estimator of degree 1 with a Epanechnikov kernel. On each interval, we apply our estimation procedure and our bandwidth selection method with $h_{\text{min}} = \frac{1}{11 \|K\|_2} \frac{1}{200}$ and $H = \{h = h_{\text{min}} + 0.1i, i = 1, \ldots, 11\}$ and we focus the analysis on the case where the tuning parameter $\alpha$ is set to 1. To evaluate the performance on our estimator, we consider the error

$$e = \mathbb{E} \left( \int q(x) - \hat{q}_h(x) \right)^2 dx$$

$$\int q^2(x) dx$$
where $\hat{h}$ is the optimal bandwidth given by our estimation procedure and we compare it to the oracle error

$$e_o = \min_{h \in H} \mathbb{E} \left( \frac{\int_I (q(x) - \hat{q}_h(x))^2 \, dx}{\int_I q^2(x) \, dx} \right).$$

In practice, we consider estimators of these errors, $\hat{e}$ and $\hat{e}_o$. In a second time, we test if the intensity is of the form $a_0 \exp(a_1 x)$ and if it is constant, that is independent of the temperature.

Results are given in Table 1 where 500 simulations of the model during 6 years with a step time of one hour are considered. % converged corresponds to the percentage estimators that have been computed, meaning that their local time was large enough and the matrix $B$ invertible. This percentage diminishes with the length of the interval, and is very low for the last interval. However, this interval corresponds to the one we have estimated the parameters. One explanation is that the model does not capture all the features of the data. For instance, seasonal volatility is not modeled whereas it impacts the number of high and low values taken by the temperature. The two errors $\hat{e}$ and $\hat{e}_o$ increases with the length of the interval, caused by boundary effects: less values of the temperature are observed near the bounds. Furthermore, the ratio between $\hat{e}$ and $\hat{e}_o$ increases: the bandwidth selection procedure is less efficient for larger interval. This corresponds to the term $\frac{1}{\nu^2}$ in the oracle inequality. Columns % exponential and constant corresponds to the percentage of simulation for which the corresponding test has not been rejected at level 95%. Results are satisfying both for exponential and constant test. Estimators of $a_0$ and $a_1$ are consistent with the true parameters but present a small bias, probably due to the form of $M_n(\theta)$ that adds a weight term inside the norm. The mean of $\hat{q}_h$ is represented in Figure 5 for each interval $I$. One can see that there is a bias in the lower boundary for each interval.

| Interval | $\hat{e}$ | $\hat{e}_o$ | % converged | % exponential | % constant | $a_0$ | $a_1$ |
|----------|-----------|-----------|-------------|---------------|------------|-------|-------|
| $[-1, 29]$ | 0.055     | 0.026     | 100         | 97            | 0          | [1006.60, 1059.68] | $[-0.210, -0.222]$ |
| $[-3, 31]$ | 0.082     | 0.03      | 81          | 97            | 0.25       | [1015.67, 1041.91] | $[-0.223, -0.202]$ |
| $[-5, 33]$ | 0.19      | 0.04      | 12.6        | 73            | 0          | [935.77, 935.50]   | $[-0.253, -0.2212]$ |

Table 1. Performance of the local polynomial estimation procedure and parametrical test on different intervals for simulated data.

![Figure 5](image-url)
7. Proof of Proposition 6

In order to prove Proposition 6, we need Proposition 9 and Proposition 10. Proposition 9 gives an approximation of the error by the bias and the variance. This proposition is similar to the one of [23, Proposition 4.1] in the context of density estimation.

During the proof, $\tilde{C}$ denotes a constant that can change from line to line. $\tilde{C}(\cdot)$ denotes a constant depending on $\cdot$ that can also change from line to line.

**Proposition 9.** Assume 1 and 2. Let $x \geq 1, \eta \in (0,1)$. Let $\mathcal{H}$ a finite subset of $(0,\infty)$ such that $\min \mathcal{H} = h_{\min} \geq \frac{\|K\|_{\infty} \|K\|_{1}|I|}{n}$ and $\max \mathcal{H} \leq \frac{2|I|}{\tilde{C}^{\alpha} K}$. Let $\hat{q}_h$ the local polynomial estimator defined in (2) and $h$ the bandwidth defined in (8). With conditional probability given $\mathcal{F}^X_T$ larger than $1 - C|\mathcal{H}|e^{-x}$, on the event $D(I,\nu)$, for any $h \in \mathcal{H}$,

$$\|q - \hat{q}_h\|^2_I \leq (1 + \eta) \left(\|q - q_h\|^2_I + V_h\right) + \tilde{C} \left(1 + \|q\|_{L,\infty}\|I\|_{L,\infty}|I||K||I|\right) x^2 \frac{n A_K^2 \nu^2 T^2 \eta^3}{n A_K^2 \nu^2 T^2 \eta^3}
$$

and

$$\|q - q_h\|^2_I + V_h \leq (1 + \eta) \|q - \hat{q}_h\|^2_I + \tilde{C} \left(1 + \|q\|_{L,\infty}\|I\|_{L,\infty}|I||K||I|\right) x^2 \frac{n A_K^2 \nu^2 T^2 \eta^3}{n A_K^2 \nu^2 T^2 \eta^3}
$$

where

$$V_h = \frac{1}{n} \int_0^T \left( \int_I \left( w (x, h, \frac{X_s - x}{h}) K_h (X_s - x) \right)^2 1_{X_s \in I} dx \right) q (X_s) ds$$

and $A_K$ is a constant depending on $K$ which is introduced in Proposition 4.

**Proposition 10.** Assume 1 and 2. Let $x \geq 1, \theta \in (0,1)$. Let $\mathcal{H}$ a finite subset of $(0,\infty)$ such that $\min \mathcal{H} = h_{\min} \geq \frac{\|K\|_{\infty} \|K\|_{1}|I|}{n}$ and $\max \mathcal{H} \leq \frac{2|I|}{\tilde{C}^{\alpha} K}$. Let $\hat{q}_h$ the local polynomial estimator defined in (2) and $\hat{h}$ the bandwidth defined in (8). Let $\hat{V}_{h,h_{\min}}$ defined in Equation (6) and $\text{pen}(\alpha)$ defined in Equation (7) with $\alpha > 0$. With conditional probability given $\mathcal{F}^X_T$ larger than $1 - C|\mathcal{H}|e^{-x}$, on the event $D(I,\nu)$, for any $h \in \mathcal{H}$,

$$(1 - \theta) \|\hat{q}_h - q\|^2_I \leq (1 + \theta) \|\hat{q}_h - q\|^2_I + \left(\text{pen}_{\alpha} (h) - 2\hat{V}_{h,h_{\min}}\right) - \left(\text{pen}_{\alpha} (\hat{h}) - 2\hat{V}_{h,h_{\min}}\right)
$$

$$+ \frac{C_2}{\theta} \|q_{h_{\min}} - q\|^2_I + \frac{C (K) |I|}{\nu^2 T^2 \theta} \left(1 + \|q\|_{L,\infty}\|I\|_{L,\infty}|I|\right) x^2 \frac{n}{n^2 h_{\min}}$$

where $C_1$ and $C_2$ are constant and $C(K)$ is a constant depending on $K$.

7.1. Proof of Proposition 9. In the following, we work conditionally on $\mathcal{F}^X_T$. We also work on the event $D(I,\nu)$. Conditionally on $\mathcal{F}^X_T$, the process $N$ is a inhomogeneous Poisson process. Proposition 4 and Proposition 5 are then verified taking the conditional probability given $\mathcal{F}^X_T$.

The norm $\|q - \hat{q}_h\|^2_I$ is the sum of $\|q - q_h\|^2_I$ which is a bias term, $\|\hat{q}_h - q_h\|^2_I$ which is a variance term and the cross term $2 < q - q_h, q_h - \hat{q}_h >_I$. In order to control $\|q - \hat{q}_h\|^2_I$ by $\|q - q_h\|^2_I + V_h$, we will control the variance term by $V_h$ and the cross term by $\|q - q_h\|^2_I + V_h$. We consider a real number $x \geq 1$ in the following.
Control of the variance term. First, let us control the term $\|q_h - \hat{q}_h\|^2_I$. This term is equal to
\[
\frac{1}{n^2} \int_I \left( \int_0^T w\left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x) 1_{X_s \in I} \, dM_s \right)^2 \, dx
\]
which can be written as the sum of
\[
\frac{1}{n^2} \int_0^T \int_I w^2\left(x, h, \frac{X_s - x}{h}\right) K_h^2(X_s - x) 1_{X_s \in I} \, dx \, dN_s
\]
and
\[
\frac{2}{n^2} \int_0^T \int_I \int_I w\left(x, h, \frac{X_s - x}{h}\right) \, w\left(x, h, \frac{X_u - x}{h}\right) K_h(X_u - x) K_h(X_s - x) \, dx \, 1_{X_s \in I} \, dM_u \, dM_s.
\]
The term (13) is a simple Poisson integral and can be controlled with Proposition 4. As
\[
\frac{1}{n^2} \int_I w^2\left(x, h, \frac{X_s - x}{h}\right) K_h^2(X_s - x) 1_{X_s \in I} \, dx \leq \frac{\|K\|_1 \|K\|_{\infty} \|I\|^2}{n^2 A_h^2 v^2 T^2 h}
\]
\[
\leq \frac{|I|}{n^2 A_h^2 v^2 T^2}
\]
and
\[
\frac{1}{n^4} \int_0^T \left( \int_I w^2\left(x, h, \frac{X_s - x}{h}\right) K_h^2(X_s - x) 1_{X_s \in I} \, dx \right)^2 \, d\Lambda_s
\]
is bounded by
\[
\frac{1}{n^2} \sup_{z \in I} \int_I w^2\left(x, h, \frac{z - x}{h}\right) K_h^2(z - x) \, dx \, V_h \leq \frac{\|K\|_1 \|K\|_{\infty} V_h |I|^2}{A_h^2 v^2 T^2 h n^2}
\]
\[
\leq \frac{V_h |I|}{n A_h^2 v^2 T^2},
\]
with conditional probability given $\mathcal{F}_T^X$ larger than $1 - 2|I|e^{-\varepsilon}$, on $(D(I, \nu), \mathcal{H})$, for any $h \in \mathcal{H}$,
\[
\left| \frac{1}{n^2} \int_0^T \int_I w^2\left(x, h, \frac{X_s - x}{h}\right) K_h^2(X_s - x) \, dx \, 1_{X_s \in I} \, dN_s - V_h \right| \leq \sqrt{\frac{2|I|}{n A_h^2 v^2 T^2}} V_h + \frac{|I|}{3n A_h^2 v^2 T^2}.
\]
Using Young’s inequality
\[
2ab \leq \epsilon a^2 + \frac{b^2}{\epsilon}, \text{ for all } \epsilon > 0,
\]
we find, for $\theta > 0$, with conditional probability given $\mathcal{F}_T^X$ larger than $1 - 2|I|e^{-\varepsilon}$, on $(D(I, \nu), \mathcal{H})$, for any $h \in \mathcal{H}$,
\[
\left| \frac{1}{n^2} \int_0^T \int_I w^2\left(x, h, \frac{X_s - x}{h}\right) K_h^2(X_s - x) \, dx \, 1_{X_s \in I} \, dN_s - V_h \right| \leq \theta V_h + \tilde{C} \frac{x|I|}{n A_h^2 v^2 T^2 \theta}.
\]

The term (14) is an U-statistics which can be controlled with Proposition 5. With conditional probability given $\mathcal{F}_T^X$ larger than $1 - 6.44e^{-\varepsilon}$, it is dominated in absolute value and on $(D(I, \nu), \mathcal{H})$, by
\[
\tilde{C} \left( C \sqrt{x} + D x + B x^{3} + A x^{2} \right)
\]
with $A, B, C$ and $D$ defined in the following. We have

$$
A = \frac{1}{n^2} \sup_{(u, s)\in[0,T]^2} \int_I \left( w \left( x, h \frac{X_s - x}{h} \right) w \left( x, h \frac{X_u - x}{h} \right) K_h (X_u - x) K_h (X_s - x) \mathbf{1}_{X_u \in I} \mathbf{1}_{X_s \in I} d\lambda_x \right)
$$

\leq \frac{\| K \|_\infty \| K \|_1 \| I \|^2}{n^2 A^2 K \nu^2 T^2}

\leq \frac{|I|}{n^2 A^2 K \nu^2 T^2}.

The term

$$
B^2 = \max \left\{ \sup_{s \in T} \int_0^s \left( \int_I w \left( x, h \frac{X_s - x}{h} \right) w \left( x, h \frac{X_u - x}{h} \right) K_h (X_u - x) K_h (X_s - x) \mathbf{1}_{X_u \in I} \mathbf{1}_{X_s \in I} d\lambda_x \right)^2 d\lambda_u, \sup_{u \leq T} \int_u^T \left( \int_I w \left( x, h \frac{X_s - x}{h} \right) w \left( x, h \frac{X_u - x}{h} \right) K_h (X_u - x) K_h (X_s - x) \mathbf{1}_{X_u \in I} \mathbf{1}_{X_s \in I} d\lambda_x \right)^2 d\lambda_s \right\}
$$

is bounded by, using the occupation time formula,

$$
|I|^4 \sup_{(v, z) \in I^2} \int_I K_h (z - x) K_h (v - x) d\lambda_x \sup_{s \in [0, T]} \int_I \int_I K_h (z - x) K_h (X_u - x) \mathbf{1}_{X_u \in I} d\lambda_x d\lambda_z
$$

\leq \frac{\| K \|_\infty \| K \|_1 \| I \|_\infty \| I \|^4}{n^2 A^4 K \nu^4 T^4}

Using again the occupation time formula and Young’s inequality for convolutions,

$$
C^2 = \int_0^T \int_0^s \left( \int_I w \left( x, h \frac{X_s - x}{h} \right) w \left( x, h \frac{X_u - x}{h} \right) K_h (X_u - x) K_h (X_s - x) \mathbf{1}_{X_u \in I} \mathbf{1}_{X_s \in I} d\lambda_x \right)^2 d\lambda_u d\lambda_s
$$

is bounded by

$$
\frac{\| I_T \|_{L^\infty} \| I_T \|_{L^\infty} \| I_T \|_{L^\infty} \| I_T \|^2}{n^3 A^2 K \nu^2 T^2} \int_0^T \int_I \left( \int_I w \left( x, h \frac{X_s - x}{h} \right) K_h (X_u - x) \mathbf{1}_{X_u \in I} K_h (z - x) d\lambda_x \right)^2 d\lambda_u d\lambda_s
$$

\leq \frac{\| I_T \|_{L^\infty} \| q \|_{L^\infty} \| I_T \|^2}{n^3 A^2 K \nu^2 T^2} \int_0^T \left| w \left( \cdot, h \frac{X_s - \cdot}{h} \right) K_h (X_s - \cdot) \mathbf{1}_{X_s \in I} \mathbf{1} * K_h (\cdot) \right|^2 d\lambda_s
$$

\leq \frac{\| I_T \|_{L^\infty} \| q \|_{L^\infty} \| I_T \|^2}{n^3 A^2 K \nu^2 T^2} \int_0^T \left| w \left( \cdot, h \frac{X_s - \cdot}{h} \right) K_h (X_s - \cdot) \mathbf{1}_{X_s \in I} \mathbf{1} \| K_h \|_2^2 \right|^2 d\lambda_s
$$

\leq \frac{\| I_T \|_{L^\infty} \| q \|_{L^\infty} \| K \|_1^2 \| \mathbf{1} \|^2}{n^2 A^2 K \nu^2 T^2}.

Using twice Cauchy Schwarz inequality, $D$ which is equal to

$$
\sup_{\int_0^T a_u d\mu_u = 1} \int_T a_u \int_T b_s \int_I w \left( x, h \frac{X_s - x}{h} \right) w \left( x, h \frac{X_u - x}{h} \right) K_h (X_u - x) K_h (X_s - x) \mathbf{1}_{X_u \in I} \mathbf{1}_{X_s \in I} d\lambda_x d\lambda_u d\lambda_s
$$
can be bounded in the same way than \( C \) and

\[
D \leq \sqrt{\frac{\|I\|_1 \|I\|_\infty \|q\|_1 \|K\|_2^2 V_\theta |I|^2}{nA_K^2 \nu^2 T^2}}.
\]

After applying (15), we obtain for \( \theta \in (0,1] \) that with conditional probability given \( F_T^X \) larger than 

\[
1 - \tilde{C} |\mathcal{H}| e^{-x}, \text{ on } D (I, \nu), \text{ for any } h \in \mathcal{H},
\]

\[
\frac{2}{n^2} \int_0^T \int_0^{x-} \int w \left( x, h, \frac{X_u - x}{h} \right) K_h (X_u - x) 1_{X_u \in I} dM u dM_s
\]

\[
(17)
\]

\[
H \leq \sqrt{\frac{1}{\nu^2 T^2 n^2}} \theta V_\theta + \frac{\tilde{C}}{n A_K^2 \nu^2 T^2} \frac{|I| (1 + \|q\|_1 \|I\|_\infty \|I\|_2^2) \theta^2}{||q_h - q||_1^2 - \|q_h - q\|_I^2 - V_h}.
\]

Combining (16) and (17) with \( \tilde{\theta} = \frac{\theta}{2} \), we find for \( \tilde{\theta} \in (0,1] \), with conditional probability given \( F_T^X \) larger than \( 1 - \tilde{C} |\mathcal{H}| e^{-x} \), on \( D (I, \nu) \), for any \( h \in \mathcal{H}, \)

\[
|\|q_h - q\|_1^2 - V_h| \leq \tilde{\theta} V_\theta + \frac{\tilde{C}}{n A_K^2 \nu^2 T^2} \frac{|I| (1 + \|q\|_1 \|I\|_\infty \|I\|_2^2) \theta^2}{|q_h - q|_I^2 - \|q_h - q\|_1^2 - V_h}.
\]

**Control of the cross term.** The cross term \( <q_h - q, q_h - q>_I \) is equal to

\[
\frac{1}{n} \int_0^T \left( \int_I (q_h (x) - q (x)) w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} dx \right) dM_s
\]

and thus can be controlled using Proposition 4. Using Cauchy Schwarz,

\[
\frac{1}{n} \int_0^T \left( \int_I (q_h (x) - q (x)) w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} dx \right) dM_s \leq \frac{1}{n} |q_h - q|_I \sqrt{\frac{\|K\|_1 \|K\|_\infty |I|^2}{h A_K^2 \nu^2 T^2}}
\]

\[
\leq |q_h - q|_I \sqrt{\frac{|I|^2}{n A_K^2 \nu^2 T^2}}
\]

and for \( \theta > 0 \),

\[
\frac{1}{n} \int_0^T \left( \int_I (q_h (x) - q (x)) w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} dx \right)^2 dM_s \leq \|q_h - q\|_I^2 V_h.
\]

We can also bound this term using the occupation time formula and Young’s inequality for convolution by

\[
\frac{|I|^2 |I|_1 \|q\|_1 \|q\|_\infty |K_h | \|q_h - q\|_I^2}{n A_K^2 \nu^2 T^2} \|K_h \|_2^2 |I|^2 \leq \frac{|I|^2 |I|_1 \|q\|_1 \|K\|_2^2 |I|^2}{n A_K^2 \nu^2 T^2} \|q_h - q\|_1^2
\]

Thus,

\[
\frac{1}{n^2} \int_0^T \left( \int_I (q_h (x) - q (x)) w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} dx \right)^2 dM_s.
\]
is bounded by
\[ \|q - q_h\|^2 \leq \frac{\|l_T\|_I,\infty \|q\|_I,\infty \|K\|_I^2 |I|^2}{nA_k^2 \nu^2 T^2} \]
and for \(\theta, u > 0\),
\[ \sqrt{\frac{2}{n^2}} \int_0^T \left( \int_I (q_h(x) - q(x)) w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} \right)^2 dx \right] \frac{d\Lambda_x}{\theta nA_k^2 \nu^2 T^2} \]
is bounded by
\[ \theta \|q - q_h\|^2 + \frac{1}{2}\sqrt{h} \left( \frac{\|l_T\|_I,\infty \|q\|_I,\infty \|K\|_I^2 |I|^2}{nA_k^2 \nu^2 T^2} \right) \leq \theta \|q - q_h\|^2 + \frac{\theta u h \|q\|_I,\infty \|K\|_I^2 |I|^2}{16\theta u A_k^2 \nu^2 T^2} . \]
Thus, for \(\theta \in (0, 1]\), with \(u = \theta^2\), we obtain, with conditional probability given \(\mathcal{F}_T^X\) larger than 1 \(- \mathcal{C} |\mathcal{H}| e^{-x}\), on \(D(I, \nu)\), for any \(h \in \mathcal{H}\),
\[ \sqrt{\frac{2}{n^2}} \int_0^T \left( \int_I (q_h(x) - q(x)) w \left( x, h, \frac{X_s - x}{h} \right) K_h (X_s - x) 1_{X_s \in I} \right)^2 dx \right] \frac{d\Lambda_x}{\theta nA_k^2 \nu^2 T^2} \]
is bounded by
\[ \theta \left( \|q - q_h\|^2 + V_h \right) + \mathcal{C} \frac{\|l_T\|_I,\infty \|q\|_I,\infty \|K\|_I^2 |I|^2}{\theta nA_k^2 \nu^2 T^2} . \]
Using Proposition 4, bounds (19) and (20), we have for \(\theta \in (0, 1]\), with conditional probability given \(\mathcal{F}_T^X\) larger than 1 \(- \mathcal{C} |\mathcal{H}| e^{-x}\), on \(D(I, \nu)\), for any \(h \in \mathcal{H}\),
\[ 2 |q_h - \hat{q}_h| < q_h - q > I | \leq 3\theta \left( \|q - q_h\|^2 + V_h \right) + \mathcal{C} \frac{|I| (1 + \|q\|_I,\infty \|l_T\|_I,\infty \|I\| \|K\|_I^2) x^2}{\theta nA_k^2 \nu^2 T^2 \theta^3} . \]
Combining (18) and (21), we find for \(\theta \in (0, 1]\), with conditional probability given \(\mathcal{F}_T^X\) larger than 1 \(- \mathcal{C} |\mathcal{H}| e^{-x}\), on \(D(I, \nu)\), for any \(h \in \mathcal{H}\),
\[ \|q - \hat{q}_h\|^2 - \|q - q_h\|^2 \leq 3\theta \|q - q_h\|^2 + (1 + 4\theta) V_h + \mathcal{C} \frac{|I| (1 + \|q\|_I,\infty \|l_T\|_I,\infty \|I\| \|K\|_I^2) x^2}{\theta nA_k^2 \nu^2 T^2 \theta^3} . \]
and
\[ \|q - \hat{q}_h\|^2 - \|q - q_h\|^2 \geq -3\theta \left( \|q - q_h\|^2 + V_h \right) + (1 - \theta) V_h - \mathcal{C} \frac{|I| (1 + \|q\|_I,\infty \|l_T\|_I,\infty \|I\| \|K\|_I^2) x^2}{\theta nA_k^2 \nu^2 T^2 \theta^3} . \]
Taking \(\theta = \frac{2}{4}\) for (22) and \(\theta = \frac{n}{3(1 + \eta)}\) for (27) with \(\eta \in (0, 1]\), the proof is achieved.

7.2. Proof of Proposition 10. We continue to work conditionally on \(\mathcal{F}_T^X\) and on the event \(D(I, \nu)\). The beginning of the proof is similar to the one of [21, Theorem 9]. For any \(h \in \mathcal{H}\),
\[ \|\hat{q}_h - q\|^2 + \text{pen}_\alpha (\hat{h}) \leq \|\hat{q}_h - \hat{q}_h\|^2 + \text{pen}_\alpha (\hat{h}) + \|\hat{q}_h - q\|^2 + 2 < \hat{q}_h - \hat{q}_h, \hat{q}_h - q > I \]
\[ \leq \|\hat{q}_h - \hat{q}_h\|^2 + \text{pen}_\alpha (h) + \|\hat{q}_h - q\|^2 + 2 < \hat{q}_h - \hat{q}_h, \hat{q}_h - q > I \]
\[ \leq \|\hat{q}_h - q\|^2 + 2 \|q - \hat{q}_h\|^2 \geq 2 < \hat{q}_h - q, q - \hat{q}_h > I + \text{pen}_\alpha (h) \]
\[ + 2 < \hat{q}_h - \hat{q}_h, \hat{q}_h - q > I \].
We then have
\[ \| \hat{q}_h - q \|^2 \leq \| \tilde{q}_h - q \|^2 + (\text{pen}_\alpha (h) - 2 \tilde{q}_h - q) q_h - q_h - q > 1 \] (24)

We want to approach \(< \hat{q}_h - q, q_h - q > 1 = S(h, h_{\min}) \) by \( \tilde{V}_h, h_{\min} \) with \( S(h, h') = < \hat{q}_h - q, q_h' - q > 1 \) for \( h, h' \in \mathcal{H} \). We have
\[ < \hat{q}_h - q, q_h - q > 1 = < \hat{q}_h - q, q_h + q_h - q_h - q_h_{\min} + q_h_{\min} - q > 1 \]
\[ = < \hat{q}_h - q, q_h - q_h_{\min} > 1 + S(h, h_{\min}) + S(h_{\min}, h) \]
\[ + < q_h - q, q_h_{\min} - q > 1 . \]

Furthermore, we can easily show that
\[ < \hat{q}_h - q, q_h_{\min} - q > 1 = \tilde{V}_h, h_{\min} + U(h, h_{\min}) \]
with
\[ \tilde{V}_h, h_{\min} = \frac{1}{n^2} \int_0^T \int_I w(x, h, X_s - x) w(x, h_{\min}, X_s - x) K_h(X_s - x) K_{h_{\min}}(X_s - x) 1_{X_s \in I} dxdN_s, \]
\[ U(h, h_{\min}) = \frac{1}{n^2} \int_0^T \int_I (G_h, h_{\min}(X_s, X_s) + G_{h_{\min}, h}(X_s, X_s)) dM_u dM_s \]
and
\[ G_{h, h'}(X_s, X_s) = \int_I w(x, h, X_s - x) w(x, h_{\min} X_s - x) K_h(X_s - x) 1_{X_s \in I} K_{h_{\min}}(X_s - x) 1_{X_s \in I} dx. \]

Thus,
\[ < \hat{q}_h - q, q_h_{\min} - q > 1 = \tilde{V}_h, h_{\min} + U(h, h_{\min}) + S(h, h_{\min}) + S(h_{\min}, h) + < q_h - q, q_h_{\min} - q > 1 . \]

In the following, we consider two real numbers \( \xi \geq 1 \) and \( \theta' \in (0, 1) \).

**Control of** \( U(h, h_{\min}) \). The term \( U(h, h_{\min}) \) can be controlled using Proposition 5. With conditional probability given \( \mathcal{F}_T^X \) larger than \( 1 - C|H| e^{-x} \), on \( D(I, \nu) \), for any \( h \in \mathcal{H} \),
\[ |U(h, h_{\min})| \leq \tilde{C} \left( C \sqrt{x} + D x + B x^2 + A x^2 \right) \]
with \( A, B, C \) and \( D \) defined below. We have
\[ A = \frac{1}{n^2} \sup_{(u, v) \in I^2} |G_h, h_{\min}(u, v) + G_{h_{\min}, h}(v, u)| \]
\[ \leq \frac{2}{n^2} \sup_{(u, v) \in I^2} |G_h, h_{\min}(u, v)| \]
\[ \leq \frac{2\|K\|_\infty \|K\|_1 |I|^2}{A^2 K^2 T^2 n^2 h_{\min}}. \]

We have \( B^2 \) equal to
\[ \frac{1}{n^2} \max_{s \leq T} \sup_{u \leq s} \int_0^s (G_h, h_{\min}(X_s, X_s) + G_{h, h_{\min}}(X_s, X_u))^2 dA_u, \sup_{u \leq T} \int_u^T (G_h, h_{\min}(X_u, X_s) + G_{h, h_{\min}}(X_s, X_u))^2 dA_s \]
and bounded by

\[
\frac{1}{n^3} \sup_{u \leq T} \int_0^T (G_{h,h_{\min}}(X_u, X_s) + G_{h,h_{\min}}(X_s, X_u))^2 d\Lambda_s \\
\leq \frac{2}{n^3} \sup_{u \leq T} \int_0^T (G_{h,h_{\min}}(X_u, X_s))^2 d\Lambda_s + \int_0^T (G_{h,h_{\min}}(X_s, X_u))^2 d\Lambda_s.
\]

Using Cauchy-Schwarz inequality,

\[
\frac{1}{n^3} \int_0^T (G_{h,h_{\min}}(X_u, X_s))^2 d\Lambda_s \leq \frac{V_b\|K\|^2}{n^3 A_K^2 \nu^2 T^2 h_{\min}}.
\]

Using the occupation time formula and Young’s inequality for convolutions,

\[
\frac{1}{n^4} \int_0^T (G_{h,h_{\min}}(X_u, X_s))^2 d\Lambda_s \leq \frac{\|q\|_{L,\infty} \|I\|_{L,\infty} \|I\|_4^2 \|K_h(X_u - \cdot) * K_{h_{\min}}(\cdot)\|^2}{n^3 A_K^2 \nu^4 T^4} \\
\leq \frac{\|q\|_{L,\infty} \|I\|_{L,\infty} \|K\|^2 \|K\|_2^2 |I|^4}{h_{\min} n^3 A_K^2 \nu^4 T^4}.
\]

Thus, combining (26) and (27),

\[
B^2 \leq \frac{2V_b\|K\|^2 |I|^2}{A_K^2 n^2 \nu^2 T^2 h_{\min}} \cdot \left( \frac{\|q\|_{L,\infty} \|I\|_{L,\infty} \|I\|_4^2 \|K_h(X_u - \cdot) * K_{h_{\min}}(\cdot)\|^2}{n^3 A_K^2 \nu^4 T^4} \right)
\]

\[
B \leq \sqrt{\frac{2V_b\|K\|^2 |I|^2}{A_K^2 n^2 \nu^2 T^2 h_{\min}}} \cdot \left( \frac{\|q\|_{L,\infty} \|I\|_{L,\infty} \|I\|_4^2 \|K_h(X_u - \cdot) * K_{h_{\min}}(\cdot)\|^2}{n^3 A_K^2 \nu^4 T^4} \right)
\]

and

\[
Bx^2 \leq \frac{\theta'V_h}{3} + \frac{3}{20\theta'} \frac{\|K\|^2 |I|^2 x^3}{A_K^2 n^2 \nu^2 T^2 h_{\min}} + \frac{\|K\|^2 x^3}{2h_{\min} A_K^2 n^2 \nu^2 T^2} + \frac{\|K\|^2 \|q\|_{L,\infty} \|I\|_{L,\infty} |I|^2}{n A_K^2 \nu^2 T^2}
\]

\[
\leq \frac{\theta'V_h}{3} + \frac{1}{\theta'} \frac{\|K\|^2 \|q\|_{L,\infty} \|I\|_{L,\infty} |I|^2}{A_K^2 n^2 \nu^2 T^2 h_{\min}} \cdot \left( \frac{\|q\|_{L,\infty} \|I\|_{L,\infty} \|I\|_4^2 \|K_h(X_u - \cdot) * K_{h_{\min}}(\cdot)\|^2}{n^3 A_K^2 \nu^4 T^4} \right).
\]

To dominate

\[
C^2 = \frac{1}{n^3} \int_0^T \int_0^s (G_{h,h_{\min}}(X_u, X_s) + G_{h,h_{\min}}(X_s, X_u))^2 d\Lambda_u d\Lambda_s
\]

which is bounded by

\[
\frac{4}{n^4} \int_0^T \int_0^T (G_{h,h_{\min}}(X_u, X_s))^2 d\Lambda_u d\Lambda_s,
\]

we use the occupation time formula and Young’s inequality for convolutions:

\[
C^2 \leq \frac{4\|q\|_{L,\infty} \|I\|_{L,\infty} |I|^2}{n^3 A_K^2 \nu^2 T^2} \int_0^T \int_0^T \|w\left(\cdot, h, \frac{X_u - \cdot}{h}\right) K_h(X_u - \cdot) 1_{X_u \in I} 1_{I} * K_{h_{\min}}(\cdot)\|^2 d\Lambda_u
\]

\[
\leq \frac{4\|q\|_{L,\infty} \|I\|_{L,\infty} \|K\|^2 |I|^2}{n^3 A_K^2 \nu^2 T^2} \int_0^T \int_0^T \|w\left(\cdot, h, \frac{X_u - \cdot}{h}\right) K_h(X_u - \cdot) 1_{I} 1_{X_u \in I}\|^2 d\Lambda_u
\]

\[
= \frac{4\|q\|_{L,\infty} \|I\|_{L,\infty} \|K\|^2 |I|^2}{A_K^2 \nu^2 T^2} V_h.
\]

We then have

\[
C \sqrt{x} \leq \frac{\theta'V_h}{3} + C \frac{\|K\|^2 \|q\|_{L,\infty} \|I\|_{L,\infty} \|I\|^2 x}{n A_K^2 \nu^2 T^2 \theta'}.
\]
Finally, with conditional probability given $F^X$ larger than $1 - \hat{C}e^{-\theta x}$, on $D(I, \nu)$,

$$|U(h, h_{\min})| \leq \theta' V_h + \tilde{C} \frac{\|K\|_2^2 \|q\|_{L^\infty} \|l_T\|_{L^\infty} |I|^2 x^2}{nA^2_K \nu^2 T^2 \theta'} + \tilde{C} \frac{\|K\|_2^2 \|K\|_1 |I|^2 x^3}{nA^2_K \nu^2 T^2 \theta^2 n^2 h_{\min}}. \tag{29}$$

**Control of $S$.** We need to control $S(h, h_{\min})$ and $S(h_{\min}, h)$. Let $h, h'$ in $\mathcal{H}$. We can write

$$S(h, h') = \frac{1}{n} \int_0^T \int_I (q_{h'} - q) w \left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x) 1_{X_s \in l} dx dM_s$$

and it is possible to control it with Proposition 4. First, using occupation time formula and noticing that

$$q(x) - q_h(x) = \int_0^T (q(x) - q(X_s)) w \left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x) 1_{X_s \in l} ds$$

using Proposition 2, we have

$$\frac{1}{n} \int_I (q_{h'} - q)(x) w \left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x) 1_{X_s \in l} dx \leq \frac{|I|}{nA_K \nu T} \|K\|_1 \|q - q_{h'}\|_{L^\infty} \leq \frac{2 \|K\|_1^2 \|l_T\|_{L^\infty} \|q\|_{L^\infty} |I|^2}{nA^2_K \nu^2 T^2}.$$

Using the occupation time formula and Young’s inequality for convolutions, the term

$$\frac{1}{n^2} \int_0^T \left( \int_I (q_{h'} - q)(x) w \left(x, h, \frac{X_s - x}{h}\right) K_h(X_s - x) 1_{X_s \in l} dx \right)^2 d\lambda_s$$

is bounded by

$$\frac{\|l_T\|_{L^\infty} \|q\|_{L^\infty} |I|^2}{nA^2_K \nu^2 T^2} \|q_{h'} - q)(\cdot) 1_{\in l} \ast K_h(\cdot)\|_2^2 \leq \frac{\|q_{h'} - q\|_1^2 \|K\|_2^2 \|l_T\|_{L^\infty} \|q\|_{L^\infty} |I|^2}{nA^2_K \nu^2 T^2}.$$
Control of $q_h - q, q_{h_{\text{min}}} - q > 1$. We have:

\[(31) \quad |q_h - q, q_{h_{\text{min}}} - q > 1| \leq \frac{\theta'}{2} \|q_h - q\|_I^2 + \frac{1}{2\theta'} \|q_{h_{\text{min}}} - q\|_I^2.
\]

At the end, combining (25), (29), (30) and (31), we have with conditional probability given $F_T^X$ higher than $1 - C|\mathcal{H}|e^{-x}$, on $D(I, \nu)$, for any $h \in \mathcal{H}$,

\[(32) \quad |q_h - q_{h_{\text{min}}} - q > 1 - \hat{V}_{h,h_{\text{min}}}| \leq \theta' (\|q_h - q\|_I^2 + V_h) + \left(\frac{\theta'}{2} + \frac{1}{2\theta'}\right) \|q_{h_{\text{min}}} - q\|_I^2 + \tilde{C}(K) |I| \left(\frac{\|q_{I,\infty}\|_I}{\nu^2 T^2 \theta'} \left(\frac{1 + \|q_{I,\infty}\|_I}{n} \|l_T\|_{I,\infty} \|I\| \|K\|_I^2 \right) x^2 + \frac{x^3 |I|}{n^2 h_{\text{min}}} \right).
\]

Furthermore, Proposition 9 states that with conditional probability given $F_T^X$ higher than $1 - C|\mathcal{H}|e^{-x}$ for any $h \in \mathcal{H}$, on $D(I, \nu)$,

\[(33) \quad \|q - q_{h_{\text{min}}}\|_I^2 + V_h \leq 2 \|q - \hat{q}_h\|_I^2 + \tilde{C} \frac{|I| (1 + \|q_{I,\infty}\|_I) x^2}{n A_K^2 \nu^2 T^2}.
\]

We combine (24), (32) and (33) and we take $\theta' = \frac{\theta}{4}$ to conclude.

7.3. Proof of Proposition 6.

Proof of (9). Let $x \geq 1$, $\tau = \alpha - 1$, $\epsilon \in (0,1)$ and $\theta \in (0,1)$ depending on $\epsilon$ and specified later. Using Proposition 10, with conditional probability given $F_T^X$ larger than $1 - C|\mathcal{H}|e^{-x}$, on $D(I, \nu)$, for any $h \in \mathcal{H}$,

\[(34) \quad (1 - \theta) \|q_h - q\|_I^2 + \tau \hat{V}_h \leq (1 + \theta) \|q_h - q\|_I^2 + \tau \hat{V}_h + \frac{\tilde{C}}{\nu^2 T^2 \theta'} \left(\frac{1 + \|q_{I,\infty}\|_I}{n} \|l_T\|_{I,\infty} |I| \|K\|_I^2 \right) x^2 + \frac{|I| x^3}{n^2 h_{\text{min}}}.
\]

Equation (16) states that with conditional probability given $F_T^X$ larger than $1 - C|\mathcal{H}|e^{-x}$, on $D(I, \nu)$, for any $h \in \mathcal{H}$,

\[(35) \quad |\hat{V}_h - V_h| \leq \frac{\theta}{2} \hat{V}_h + \tilde{C} \frac{x |I|}{n A_K^2 \nu^2 T^2 \theta'}.
\]

First, let us consider the case $\tau \geq 0$. We then have, with conditional probability given $F_T^X$ larger than $1 - C|\mathcal{H}|e^{-x}$, on $D(I, \nu)$, for any $h \in \mathcal{H}$,

\[(36) \quad (1 - \theta) \|q_h - q\|_I^2 \leq (1 + \theta) \|q_h - q\|_I^2 + \tau \left(1 + \frac{\theta}{2}\right) V_h + \frac{\tilde{C}}{\nu^2 T^2 \theta'} \|q_{h_{\text{min}}} - q\|_I^2 + \frac{\tilde{C} |I| (1 + \|q_{I,\infty}\|_I) x^2}{n \theta^3 \nu^2 T^2}.
\]

Using Proposition 9 with $\eta = \frac{\theta}{2 + \theta}$, with conditional probability given $F_T^X$ larger than $1 - C|\mathcal{H}|e^{-x}$, on $D(I, \nu)$, for any $h \in \mathcal{H}$,

\[\tau V_h \leq \tau \left(1 + \frac{\theta}{2 + \theta}\right) \|q - q_h\|_I^2 + \frac{\tilde{C} |I| (1 + \|q_{I,\infty}\|_I) x^2}{n \theta^3 \nu^2 T^2}.\]
Combining inequality (36) with (34) and (35) and as \( \left( 1 + \frac{\theta}{2} \right) \left( 1 + \frac{\theta}{2 + \theta} \right) = 1 + \theta \), we find that with conditional probability given \( \mathcal{F}_T^X \) larger than 1 – \( \tilde{C} \mathcal{K} e^{-x} \), on \( D (I, \nu) \), for any \( h \in \mathcal{H} \),

\[
(1 - \theta) \| \hat{q}_h - q \|_I^2 \leq \left( 1 + \theta + (1 + \theta) \tau \right) \| \hat{q}_h - q \|_I^2 + \frac{\tilde{C}}{\theta} \| q_{h_{\min}} - q \|_I^2 \\
+ \frac{|I| \tilde{C} (K) \left( \frac{1}{\theta} + \frac{1}{\theta^3} \right) \left( \left( 1 + \| q \|_{L, \infty} \| I_T \|_{L, \infty} |I| \right) x^2 \right)}{n^2 \tau h_{\min}}.
\]

With \( \theta = \frac{\epsilon}{\epsilon + 2 + \epsilon} \), we have

\[
\| \hat{q}_h - q \|_I^2 \leq \left( 1 + \tau + \epsilon \right) \| \hat{q}_h - q \|_I^2 + \tilde{C} (1 + 2 + 2 \tau \epsilon) \| q_{h_{\min}} - q \|_I^2 \\
+ \frac{\tilde{C} (K) |I| (1 + 2 + 2 \tau \epsilon)}{\nu^2 T^2} \left( \left( 1 + \| q \|_{L, \infty} \| I_T \|_{L, \infty} |I| \right) x^2 \right) + \frac{x^3 |I|}{n^2 h_{\min}}.
\]

Inequality (9) is then proved in the case \( \alpha \geq 1 \) with

\[
C_0 (\epsilon, \alpha) = \frac{\alpha + \epsilon}{\alpha} \\
C_2 (\epsilon, \alpha) = \frac{\tilde{C} (2 + 2 \alpha)^2}{\alpha \epsilon}
\]

and

\[
C_3 (\epsilon, K, \alpha) = \frac{\tilde{C} (K) (2 + 2 \alpha)^4}{\alpha \epsilon^3}.
\]

Now, let us consider the case \( -1 < \tau \leq 0 \). According to Proposition 9, with conditional probability on \( \mathcal{F}_T^X \) larger than 1 – \( \tilde{C} \mathcal{K} e^{-x} \), on \( D (I, \nu) \),

\[
\tau V_h \geq \tau (1 + \theta) \| q - \hat{q}_h \|_I^2 + \tau \frac{\tilde{C} (K) |I| (1 + \| q \|_{L, \infty} \| I_T \|_{L, \infty} |I|) x^2}{n \theta^3 \nu^2 T^2}.
\]

We find, combining (37) with (34) and (35), that with conditional probability given \( \mathcal{F}_T^X \) larger than 1 – \( \tilde{C} \mathcal{K} e^{-x} \), on \( D (I, \nu) \), for any \( h \in \mathcal{H} \),

\[
(1 - \theta + \tau (1 + \theta)) \| \hat{q}_h - q \|_I^2 \leq \left( 1 + \theta + (1 + \theta) \tau \right) \| \hat{q}_h - q \|_I^2 + \frac{\tilde{C}}{\theta} \| q_{h_{\min}} - q \|_I^2 \\
+ \frac{|I| \tilde{C} (K) \left( \frac{1}{\theta} + \frac{1}{\theta^3} \right) \left( \left( 1 + \| q \|_{L, \infty} \| I_T \|_{L, \infty} |I| \right) x^2 \right)}{n^2 \tau h_{\min}}.
\]

Taking \( \theta = \frac{\epsilon (\tau + 1)^2}{2 + \tau (1 + \eta)} < 1 \), we obtain, with conditional probability given \( \mathcal{F}_T^X \) larger than 1 – \( \tilde{C} \mathcal{K} e^{-x} \), on \( D (I, \nu) \), for any \( h \in \mathcal{H} \),

\[
\| \hat{q}_h - q \|_I^2 \leq \left( 1 + \frac{1}{\tau + \epsilon} \right) \| \hat{q}_h - q \|_I^2 + \frac{\tilde{C} (2 + \epsilon (1 + \tau))}{2 \epsilon (\tau + 1)^3} \| q_{h_{\min}} - q \|_I^2 \\
+ \frac{\tilde{C} (K) (2 + \epsilon (1 - \tau))^4 |I|}{2 \epsilon^3 (\tau + 1)^7} \left( \left( 1 + \| q \|_{L, \infty} \| I_T \|_{L, \infty} |I| \right) x^2 \right) + \frac{x^3 |I|}{n^2 h_{\min}}.
\]

Inequality (9) is then verified in the case \( \alpha < 1 \) with

\[
C_0 (\epsilon, \alpha) = \frac{\alpha + \epsilon}{\alpha} \\
C_2 (\epsilon, \alpha) = \frac{\tilde{C} (2 + \epsilon (2 - \alpha))^2}{2 \alpha \epsilon^3}
\]
and
\[ C_3 (\epsilon, K, \alpha) = \tilde{C} (K) \frac{(2 + \epsilon (2 - \alpha))^4}{2\epsilon^3 \alpha^7}. \]

**Proof of (10).** Let us use (9) with \( x = 5 \log (n \vee |\mathcal{H}|) \) and \( \epsilon = \tilde{C} (\log (n))^{-1} \). Let \( \mathcal{E} \) be the event on which (9) is true. Integrating with respect to \( \mathcal{F}_T^X \) and dividing by \( \mathbb{P} (D (I, \nu)) \), we find
\[
\mathbb{E} (\| \tilde{q}_h - q \|^2_1 \mathbf{1}_c | D (I, \nu) ) \leq \left( \alpha \vee \frac{1}{\alpha} + \tilde{C} (\log (n))^{-1} \right) \text{min}_{h \in \mathcal{H}} \mathbb{E} (\| \tilde{q}_h - q \|^2_1 | D (I, \nu) )
\]
\[
+ \tilde{C} (K, \alpha) |I| \log (n) \mathbb{E} \left( \| I \|_2 \mathbb{E} (\| I \|_2 | D (I, \nu) ) \log (n \vee |\mathcal{H}|) \right)^2 + \frac{125 |I| \log (n \vee |\mathcal{H}|)^3}{n^2 h_{\text{min}}}.
\]

Let \( N_I = \int_0^T 1_{X_s \in I} dN_s \). On \( \mathcal{E}^c \cap D (I, \nu) \), using Cauchy Schwarz inequality,
\[
\| \tilde{q}_h - q \|^2_1 \leq 2 \| q \|^2_1 + \frac{2 |I|^2}{\nu^2 T^2 A_K^2 n^2} N_I \int_0^T \int_0^T K_h^2 (X_s - x) 1_{X_s \in I} dN_s dx
\]
\[
\leq 2 \| q \|^2_1 + \frac{2 N_I^2 \| K \|_2 \| I \|_2}{\nu^2 T^2 A_K^2 n^2 h}
\]
\[
\leq 2 \| q \|^2_1 + \frac{2 N_I^2 |I|}{\nu^2 T^2 A_K^2 n^2 h}.
\]

Using Cauchy Schwarz inequality again, we have
\[
\mathbb{E} (N_I^2 \mathbf{1}_{c \cap \mathcal{F}_T^X}) \leq \mathbb{E} (N_I^4 \mathbb{F}_T^X)^{1/2} \mathbb{P} (\mathcal{E}^c \cap \mathcal{F}_T^X)^{1/2}.
\]

Using Laplace transform formula, we easily show that \( N_I \) has the law of a Poisson random variable with parameter \( n \int_0^T q (X_s) 1_{X_s \in I} ds \) conditionally on \( \mathcal{F}_T^X \) and
\[
\mathbb{E} (N_I^2 |\mathcal{F}_T^X) \leq \tilde{C} n^2 \left( \| q \|_{L^\infty} + (\| q \|_{L^\infty} T)^2 + (\| q \|_{L^\infty} T)^3 + (\| q \|_{L^\infty} T)^4 \right).
\]

We also have
\[
\mathbb{P} (\mathcal{E}^c |\mathcal{F}_T^X) \leq \frac{|\mathcal{H}|}{n^5 \sqrt{|\mathcal{H}|}}
\]
\[
\leq \frac{\tilde{C}}{n^4}.
\]

Integrating with respect to \( \mathcal{F}_T^X \) and dividing by \( \mathbb{P} (D (I, \nu)) \), we find
\[
\mathbb{E} (\| \tilde{q}_h - q \|^2_1 \mathbf{1}_c | D (I, \nu) ) \leq \tilde{C} (K, \alpha) \frac{|I| |\mathcal{H}|}{\nu^2 T^2 n} + \tilde{C} (K) \frac{|I| \sqrt{\sum_{i=1}^4 (\| q \|_{L^\infty} T)^4}}{\nu^2 T^2 n}.
\]

We obtain (10) combining (38) and (39).
8. Other proofs

8.1. Proof of Proposition 1. Under Assumption 1 (i), we can define a local time in the sense of the continuous semimartingale $L^a_t$ continuous in $t$, cadlag in $x$, see [30, Chapter 6] for more information. [30, Exercise 1.15] states that for every measurable function $h$ on $[0,T] \times \Omega \times \mathbb{R}$.

$$
\int_0^t h(s, X_s) d < X_s, X_s >= \int_0^t da \int_0^t h(s, a) dL^a_s.
$$

Let $f$ be a measurable function on $\Omega \times \mathbb{R}$. As $\sigma_s \geq \sigma > 0$ for every $s \in [0,T]$ almost surely and $d < X_s, X_s >= \sigma^2_a ds$, we have

$$
\int_0^t f(X_s) ds = \int_0^t f(a) da \int_0^t \frac{1}{\sigma^2_s} dL^a_s
$$

and (i) is verified with $l^*_T = \int_0^T \frac{1}{\sigma^2_s} dL^a_s$. According to [4, Equation (III)\gamma], as

$$
\mathbb{E} \left( \sup_{0 \leq t \leq T} \int_0^t \sigma_s dW_s + \int_0^T |\mu_s| ds \right) < \infty,
$$

and because $\sigma_s \geq \sigma$ a.s., we obtain (ii). The continuity of $x \mapsto l^*_T$ follows from [38, Example 2.2.3 (a)] and $\sigma^2_s > 0$ for all $s \in [0,T]$.

Under Assumption 1 (ii), we have $l^*_T = 1_{x \in [0,t]}$ for all $x$ in $\mathbb{R}$ and $t$ in $[0,T]$.

8.2. Proof of Proposition 3. Let $\| \cdot \|_{m+1}$ be the Euclidian norm on $\mathbb{R}^{m+1}$. For a symmetric positive semi-definite matrix $A$, let $\lambda_{\min}(A)$ the smallest eigenvalue of $A$. In the following, we work on $D(I, \nu)$. We have

$$
\lambda_{\min} (B(x, h)) = \inf_{\|v\|_{m+1} = 1} v^T B(x, h) v.
$$

Let $v$ in $\mathbb{R}^{m+1}$ with $\|v\|_{m+1} = 1$. Using the occupation time formula, we have

$$
v^T B(x, h) v = \int_0^T \left( v^T U \left( \frac{X_s - x}{h} \right) \right)^2 K_h(X_s - x) 1_{X_s \in I} ds
$$

$$
= \int_I \left( v^T U \left( \frac{u - x}{h} \right) \right)^2 K_h(u - x) l^0_T du
$$

$$
\geq \frac{\nu T \lambda_{\min}}{|I|h} \int_I \left( v^T U \left( \frac{u - x}{h} \right) \right)^2 1_{|u - x| < |\Delta h|} du.
$$

If $x \in [\min I + \Delta h, \max I - \Delta h]$, if $x \in [\min I, \min I + \Delta h]$, because $h \leq \frac{|I|}{2 \lambda}$,

$$
\frac{1}{h} \int_I \left( v^T U \left( \frac{u - x}{h} \right) \right)^2 1_{|u - x| < |\Delta h|} du = \int_{\mathbb{R}} (v^T U (u))^2 1_{|u| < |\Delta h|} du
$$

$$
\frac{1}{h} \int_I \left( v^T U \left( \frac{u - x}{h} \right) \right)^2 1_{|u - x| < |\Delta h|} du \geq \int_{\mathbb{R}} (v^T U (u))^2 1_{0 < u < \frac{|\Delta h|}{h}} du
$$
and if \( x \in [\max I - \Delta h, \max I] \),
\[
\frac{1}{h} \int_I \left( v^T U \left( \frac{u - x}{h} \right) \right)^2 \mathbf{1}_{|u - x| < \Delta h} du \geq \int_R \left( v^T U (u) \right)^2 \mathbf{1}_{\frac{\Delta}{2} < u < 0} du.
\]

Thus, for all \( x \in I \), \( \lambda_{\min} (B (x, h)) \) is larger than
\[
\frac{\nu T}{|I|} K_{\min} \left( \inf_{\|v\|_{m+1} = 1} \left( \int_R \left( v^T U (u) \right)^2 \mathbf{1}_{0 < u < \frac{\Delta}{2}} du \right), \inf_{\|v\|_{m+1} = 1} \left( \int_R \left( v^T U (u) \right)^2 \mathbf{1}_{\frac{\Delta}{2} < u < 0} du \right) \right)
\]
\[
= \frac{\nu T}{|I|} K_{\min} \left( \int_R U (u) U^T (u) \mathbf{1}_{0 < u < \frac{\Delta}{2}} du, \int_R U (u) U^T (u) \mathbf{1}_{\frac{\Delta}{2} < u < 0} du \right)
\]
and
\[
\min \left( \lambda_{\min} \left( \int_R U (u) U^T (u) \mathbf{1}_{0 < u < \frac{\Delta}{2}} du \right), \lambda_{\min} \left( \int_R U (u) U^T (u) \mathbf{1}_{\frac{\Delta}{2} < u < 0} du \right) \right) > 0
\]

applying [35, Lemma 1.4] with \( K (u) = \mathbf{1}_{0 < u < \frac{\Delta}{2}} \) and \( K (u) = \mathbf{1}_{\frac{\Delta}{2} < u < 0} \).

We have
\[
|w \left( x, h, \frac{X_s - x}{h} \right) \mathbf{1}_{|\frac{X_s - x}{h}| \leq 1} | \leq \| B (x, h) \|^{-1} U \left( \frac{X_s - x}{h} \right) \|_{m+1} \mathbf{1}_{|\frac{X_s - x}{h}| \leq 1}
\]
\[
\leq \frac{1}{\lambda_{\min} (B (x, h))} \| U \left( \frac{X_s - x}{h} \right) \|_{m+1} \mathbf{1}_{|\frac{X_s - x}{h}| \leq 1}
\]
\[
\leq \frac{1}{\lambda_{\min} (B (x, h))} \sqrt{1 + \frac{1}{(1!)^2} + \frac{1}{(2!)^2} + \ldots + \frac{1}{(m!)^2}}
\]
\[
\leq \frac{2}{\lambda_{\min} (B (x, h))}.
\]

Then,
\[
(40)
|w \left( x, h, \frac{X_s - x}{h} \right) \mathbf{1}_{|\frac{X_s - x}{h}| \leq 1} | \leq \frac{|I|}{A_K \nu T}
\]

with
\[
A_K = \frac{K_{\min} \left( \int_R U (u) U^T (u) \mathbf{1}_{0 < u < \frac{\Delta}{2}} du, \int_R U (u) U^T (u) \mathbf{1}_{-\frac{\Delta}{2} < u < 0} du \right)}{2} > 0.
\]

8.3. Proof of Proposition 7.

Lower bound. We suppose for simplicity that \( I = [0, 1] \). As in [35, Section 2.6.1], we consider a real number \( c_0 > 0 \) and
\[
m = \left\lfloor c_0 n^{\frac{1}{m+1}} \right\rfloor + 1, \quad h_n = \frac{1}{m}, \quad x_k = \frac{k - \frac{1}{2}}{m},
\]
\[
\varphi_k (x) = L h_n^2 K \left( \frac{x - x_k}{h_n} \right), \quad k = 1, \ldots, m, \quad x \in [0, 1]
\]

with
\[
K (u) = a \exp \left( -\frac{1}{1 - 4u^2} \right) \mathbf{1}_{|2u| \leq 1}, \quad a > 0.
\]
For a sufficient small, $\mathbf{K} \in \Sigma (\beta, \frac{1}{2}) \cap C^\infty (\mathbb{R})$. According to [35, Equation (2.5)], the functions $\varphi_k$ belongs to $\Sigma (\beta, \frac{1}{2}, I)$ and the set of function

$$\mathcal{C} = \{ q : q (x) = \rho + \sum_{k=1}^{m} w_k \varphi_k (x), \; w_k \in \{0,1\}, \; x \in I \}$$

is included in $\Lambda_{\rho, \beta}$ as the functions $\varphi_k$ have disjoint supports.

Let us suppose that $m \geq 8$. According to [35, Lemma 2.9], there exists a subset $\hat{\mathcal{C}}$ of $\mathcal{C}$ such that for all $f_w = \rho + \sum_{k=1}^{m} w_k \varphi_k (x) \in \hat{\mathcal{C}}$ and all $f_{w'} = \rho + \sum_{k=1}^{m} w'_k \varphi_k (x) \in \hat{\mathcal{C}}$, we have

$$\sum_{k=1}^{m} (w_k - w'_k)^2 \geq \frac{m}{8}$$

and with

$$M \geq \frac{2m}{8}$$

where $M + 1 = |\hat{\mathcal{C}}|$. Now, if we consider two elements $f_w$ and $f_{w'}$ of $\hat{\mathcal{C}}$, we have

$$\| f_w - f_{w'} \|_I = Lh_m + 4 \| \mathbf{K} \| \sqrt{\sum_{k=1}^{m} (w_k - w'_k)^2}$$

$$\geq Lh_m + 4 \| \mathbf{K} \| \sqrt{\frac{m}{16}}$$

$$= \frac{L}{4} \| \mathbf{K} \| m^{-\beta}.$$

Thus, if $n \geq n^* = \left(\frac{2}{\pi}\right)^{2\beta+1}, \; m \geq 8$ and $m^{\beta} \leq (2c_0)^{\beta} n^{\frac{\beta}{\beta+\alpha}}$. Hence,

$$\| f_w - f_{w'} \|_I \geq 2s_n$$

with $s_n = A n^{-\frac{\beta}{\beta+\alpha}}$ and $A = \frac{L}{4} \| \mathbf{K} \| (2c_0)^{-\beta}$.

The following part of the proof differs from [35]. $\hat{\mathcal{C}}$ can be written $\{q_0, q_1, ..., q_M\}$. Let us denote by $\mathbb{P}_j$ the probability measure associated to the intensity $nq_j (X_s), \; s \in [0,T]$. We consider the Kullback divergence between $\mathbb{P}_0$ and $\mathbb{P}_j$ conditionally on $D(I, \nu)$, that is $\mathbb{E}_{\mathbb{P}_0} \left( \log \frac{\mathbb{P}_j}{\mathbb{P}_0} \right | D(I, \nu))$, denoted by $K(\mathbb{P}_0, \mathbb{P}_j)$. We have

$$K(\mathbb{P}_0, \mathbb{P}_j) = \mathbb{E} \left( \int_{0}^{T} n (q_j (X_s) - q_0 (X_s) - q_0 (X_s) \log \frac{q_j (X_s)}{q_0 (X_s)}) \; ds \bigg | D(I, \nu) \right)$$

$$= \mathbb{E} \left( \int_{0}^{T} nq_0 (X_s) \left( q_j (X_s) - q_0 (X_s) \right) - \log \left( 1 + \frac{q_j (X_s) - q_0 (X_s)}{q_0 (X_s)} \right) \; ds \bigg | D(I, \nu) \right)$$

$$\leq \mathbb{E} \left( \int_{0}^{T} nq_0 (X_s) \left( q_j (X_s) - q_0 (X_s) \right) - \log \left( 1 + \frac{q_j (X_s) - q_0 (X_s)}{q_0 (X_s)} \right) \; ds \bigg | D(I, \nu) \right)$$

$$\leq \mathbb{E} \left( \int_{0}^{T} n (q_j (X_s) - q_0 (X_s))^2 \; ds \bigg | D(I, \nu) \right)$$

$$\leq \mathbb{E} \left( \int_{0}^{T} n (q_j (X_s) - q_0 (X_s))^2 \; ds \bigg | D(I, \nu) \right)$$

(42)
using the fact that for \( x > -1 \), \( \log(1 + x) \geq \frac{x}{1 + x} \). Continuing from (42), we have:

\[
    K(P_0, P_j) \leq \frac{nh^{2\beta} L^2 T \|K\|_{\infty}^2}{\rho} = \frac{L^2 T \|K\|_{\infty}^2 c_0^{-2(\beta+1)m}}{\rho}.
\]

Let \( \alpha \in (0, \frac{1}{8}) \). As \( \log(M) \geq \frac{\log(2)m}{8} \), we choose

\[
    c_0 = \left( \frac{8L^2 \|K\|_{\infty} T}{\alpha \log(2) \rho} \right)^{\frac{1}{2\beta+1}}
\]

and we have

\[
    \frac{1}{M} \sum_{j=1}^M K(P_0, P_j) \leq \alpha \log(M).
\]

We have according to (41) and (43)

(i) \( \|\lambda_j - \lambda_k\|_I > 2s_n \) for all \( j \neq k \) with \( s_n = An^{-\frac{\beta}{2(\beta+1)}} \) and

(ii) \( \frac{1}{M} \sum_{j=1}^M K(P_0, P_j) \leq \alpha \log(M) \).

We can conclude using [35, Theorem 2.5]:

\[
    \liminf_{n \to \infty} \inf_{T_n} R(T_n, \Lambda_{\rho, \beta}, n^{-\frac{\beta}{2(\beta+1)}}) > 0.
\]

**Upper bound.** Let us assume that \( l \leq m \). Let \( h \) in \( \mathcal{H} \). If \( q \in \Sigma(L, \beta, I) \), the bias part of \( \mathbb{E}(\|q_h - q\|_I^2 | D(I, \nu)) \), \( \mathbb{E}(\|q - q_h\|_I^2 | D(I, \nu)) \), is equal to, using Proposition 2,

\[
    E \left( \int_I \left( \int_0^T w \left( x, h, \frac{X_s - x}{h} \right) K_h(X_s - x) 1_{X_s \in I} (q(X_s) - q(x)) \, ds \right)^2 \, dx | D(I, \nu) \right).
\]

Using Proposition 2 and Taylor’s expansion, there exists \( (\tau_s)_{0 \leq s \leq T} \) such that the integral inside the expectation in (44) is equal to, on the event \( D(I, \nu) \),

\[
    \int_I \left( \int_0^T w \left( x, h, \frac{X_s - x}{h} \right) K_h(X_s - x) 1_{X_s \in I} \left( q^{(l)}(x + \tau_s (X_s - x)) - q^{(l)}(x) \right) \frac{(X_s - x)^l}{l!} \, ds \right)^2 \, dx.
\]
The variance part is equal to

$$q = l > m$$

depending on $q$, depending on $|h| \leq \hat{C} (K, |I|)$ is a constant depending only on $K$ and $|I|$ and $\mathbb{E} (\|q - \hat{q}_h\|_2^2 | D (I, \nu))$ is of order $n^{-2\beta}$,

$$\mathbb{E} (\|q - \hat{q}_h\|_2^2 | D (I, \nu)) \leq C_2 (q, \mathbb{E} (\|l_T\|_{1, \infty} | D (I, \nu)), K, T, \beta, |I|) n^{-2\beta}$$

with $C_1 (q, \mathbb{E} (\|l_T\|_{1, \infty} | D (I, \nu)), K, T, \beta, |I|)$ and $C_2 (q, \mathbb{E} (\|l_T\|_{1, \infty} | D (I, \nu)), K, T, \beta, |I|)$ constants depending on $q$, $\mathbb{E} (\|l_T\|_{1, \infty} | D (I, \nu))$, $K, T, \beta$ and $|I|$. Finally,

$$\lim_{n \to \infty} \mathbb{R} (\hat{q}_n, \Lambda_{\rho, \beta}, n^{-2\beta}) < \infty.$$  

In the case where $l > m$, we apply the Taylor expansion formula of the bias up to the order $m$ and we find that the bias is of order $n^{m}$ as $q^{(m)}$ is bounded. The convergence rate is then of order $n^{m/2}$.

8.4. Proof of Proposition 8. For simplicity, the proof is done in the case where $\Theta \subset \mathbb{R}$, that is when $d = 1$. The proof is similar for any $d \geq 1$. In the following, let us use the following notation:

$$\hat{K} (x, h, z) = w (x, h, z) K (z), \text{ for } x \in I, h > 0, z \in \mathbb{R}.$$  

As $X$ is an ancillary statistic, one can work as if $X$ was deterministic. We also work on the event $D (I, \nu)$. 

Convergence of $\hat{\theta}_n$. First, let us study the convergence of $M_n(\theta)$ for $\theta \in \Theta$. We can write $M_n(\theta)$ as the sum of

\begin{equation}
\frac{2}{n^2h_n^2} \int_0^T \int_0^{s_n} \int_I \tilde{K} \left( x, h_n, \frac{X_s - x}{h_n} \right) \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) 1_{X_s \in I} 1_{X_u \in I} dx \, ds \, dx,
\end{equation}

\begin{equation}
\frac{1}{h_n^2} \int_0^T \int_I \tilde{K} \left( x, h_n, \frac{X_s - x}{h_n} \right) 1_{X_s \in I} \left[ q(X_s) - g_\theta(X_s) \right] ds \, dx
\end{equation}

and

\begin{equation}
\frac{2}{n^2h_n^4} \int_0^T \int_0^{s_n} \int_I \int_I \tilde{K} \left( x, h_n, \frac{X_s - x}{h_n} \right) \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) (q(X_s) - g_\theta(X_s)) 1_{X_s \in I} 1_{X_u \in I} ds \, dx.
\end{equation}

The first term (45) has expectation 0 and variance equal to

\begin{equation}
\frac{4}{n^4h_n^4} \int_0^T \int_0^{s_n} \left( \int_I \tilde{K} \left( x, h_n, \frac{X_s - x}{h_n} \right) \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) 1_{X_s \in I} 1_{X_u \in I} dx \right)^2 n^2 q(X_u) q(X_u) \, ds \, duds
\end{equation}

which is equal to

\begin{equation}
\frac{2}{n^2h_n^4} \int_0^T \int_0^{T} \left( \int_I \tilde{K} \left( x, h_n, \frac{X_s - x}{h_n} \right) \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) 1_{X_s \in I} 1_{X_u \in I} dx \right)^2 q(X_u) q(X_u) \, duds
\end{equation}

Using the occupation time formula, (48) is equal to

\begin{equation}
\frac{2}{n^2h_n^4} \int_0^T \int_0^{T} \left( \int_I \tilde{K} \left( x, h_n, \frac{r - x}{h_n} \right) \tilde{K} \left( x, h_n, \frac{y - x}{h_n} \right) dx \right)^2 q(r) l_T q(y) l_T \, dr \, dy.
\end{equation}

By writing $I = [1, 1]$, (48) is equal to

\begin{equation}
\frac{2}{n^2h_n^4} \int_0^T \int_0^{T} \left( \int \tilde{K} \left( y - uh_n, h_n, u \right) \tilde{K} \left( y - uh_n, h_n, u + p \right) du \right)^2 q(y + h_n p) l_T q(y) l_T \, dp \, dy.
\end{equation}

Using the continuity properties of $x \mapsto l_T^\nu$ from Proposition 1, several times the dominated convergence theorem and the fact that $\inf_{x \in I} l_T^\nu \geq \nu > 0$ and $\|l_T\|_{L^\infty} < \infty$, we find that

\begin{equation}
\tilde{K} \left( x, h_n, z \right) \to \frac{1}{n^2h_n^4} \left( l_T^\nu \right)^{-1} w(z) K(z), \text{ for } x \in I, \ z \in \mathbb{R}
\end{equation}

with

\begin{equation}
w(u) = U^T(0) \left( \int_{\mathbb{R}} U(z) U^T(z) K(z) \, dz \right)^{-1} U(u)
\end{equation}

and that (48) is equivalent to

\begin{equation}
\frac{2}{n^2h_n^4} \int \left( \int w(u) w(u + p) K(u) K(u + p) \, du \right)^2 dp \int \left( \frac{q(y)}{l_T^\nu} \right)^2 \, dy.
\end{equation}

Thus, (45) is of order $O_p \left( \frac{1}{n^{1/2}h_n} \right)$ and goes to 0 in probability. Concerning the second term (46), using the dominated convergence theorem again, we find that it converges to

\begin{equation}
\int w(u) K(u) \, du \int \left( q(x) - g_\theta(x) \right)^2 \, dx
\end{equation}
and
\[ \int_{\mathbb{R}} w(u) K(u) \, du = 1 \]
as it is the limit of \( \int_{0}^{T} w \left( x, h_n, \frac{X_u - x}{h_n} \right) K_{h_n} \left( \frac{X_u - x}{h_n} \right) \, ds \) which is equal to 1 according to Proposition 2. The last term (47) has mean 0 and variance equal to \( O \left( \frac{1}{h_n} \right) \) and goes to 0 when \( n \to \infty \). Thus,
\[ M_n (\theta) \overset{p}{\to} M (\theta) = \| q - g_\theta \|_2^2. \]

Under Assumption 3, we easily show that
\[ |M_n (\theta_1) - M (\theta_1) - (M_n (\theta_2) - M (\theta_2))| \lesssim |\theta_1 - \theta_2|. \]

According to Kolmogorov continuity criterion, the convergence is then uniform in \( \theta \). Furthermore, under \( H_0 \), the minimum of \( M (\theta) \) is achieved for \( \theta = \theta_0 \) which is the unique minimum under Assumption 3. We then have, under \( H_0 \),
\[ \hat{\theta}_n \overset{p}{\to} \theta_0. \]

Using Taylor’s formula, on the event \( \{ M_n^{(2)} (\theta_0) + \frac{1}{2} (\theta_0 - \theta_n)^2 M_n^{(3)} (\hat{\theta}_n) \neq 0 \} \),
\[ \sqrt{n} (\hat{\theta}_n - \theta_0) = \frac{-\sqrt{n} M_n' (\theta_0)}{M_n^{(2)} (\theta_0) + \frac{1}{2} (\theta_0 - \theta_n)^2 M_n^{(3)} (\hat{\theta}_n)} \]
with \( \hat{\theta}_n \) between \( \theta_0 \) and \( \hat{\theta}_n \). \( M_n' (\theta) \) is equal to
\[ -\frac{2}{nh_n^2} \int_{0}^{T} f (h_n, \theta, X_s) (dN_s - n g_\theta (X_s) \) ds
with
\[ f (h_n, \theta, X_s) = \int_{I} \int_{0}^{T} \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) 1_{X_s \in I} \tilde{g}_{\theta g_\theta} (X_s) \, du \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) 1_{X_s \in I} \, dx, \]
\( M_n^{(2)} (\theta) \) to
\[ -\frac{2}{nh_n^2} \int_{0}^{T} \partial_\theta f (h_n, \theta, X_s) (dN_s - n g_\theta (X_s) \) ds
\[ + \frac{2}{h_n^2} \int_{I} \left( \int_{0}^{T} \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) 1_{X_s \in I} \tilde{g}_{\theta g_\theta} (X_s) \, du \right)^2 \, dx \]
and \( M_n^{(3)} (\theta) \) to
\[ -\frac{2}{nh_n^2} \int_{0}^{T} \partial_{\theta, \theta} f (h_n, \theta, X_s) (dN_s - n g_\theta (X_s) \) ds
\[ + \frac{6}{h_n^2} \int_{I} \int_{0}^{T} \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) 1_{X_s \in I} \tilde{g}_{\theta g_\theta} (X_s) \, ds \int_{0}^{T} \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) 1_{X_s \in I} \partial_{\theta, \theta} g_\theta (X_s) \, ds \, dx. \]

Under \( H_0 \), \( M_n = N_s - n \int_{0}^{s} g_\theta (X_s) \) ds. Thus, \( \sqrt{n} M_n' (\theta_0) \) has expectation 0 and variance
\[ \frac{4}{h_n^4} \int_{0}^{T} \left( \int_{I} \int_{0}^{T} \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) \partial_\theta g_\theta (X_s) 1_{X_s \in I} \, ds \tilde{K} \left( x, h_n, \frac{X_u - x}{h_n} \right) 1_{X_s \in I} \right)^2 g_\theta (X_u) \, du. \]
that converges to
\[ 4 \int_I \frac{(\partial_0 g_{\theta_0}(u))^2}{l_T} g_{\theta_0}(u) du. \]

We can easily show that
\[ \frac{8}{n^3 h_n^6} \int_0^T (f(h_n, \theta, X_s))^3 n g_{\theta_0}(X_u) du = O \left( \frac{1}{\sqrt{n}} \right) \]
and converges to 0. Thus, according to [29, Theorem 3], we have the following central limit theorem
\[ (49) \]
\[ - \sqrt{n} M_n' (\theta_0) \xrightarrow{L} \frac{1}{4} \int_I \frac{(\partial_0 g_{\theta_0}(u))^2}{l_T} g_{\theta_0}(u) du N(0, 1). \]

Under $H_0$, the first term of $M_n^{(2)} (\theta_0)$
\[ (50) \]
\[ - \frac{2}{n h_n^2} \int_0^T \partial_0 f(h_n, \theta, X_s) dM_s \]
has mean 0 and variance
\[ \frac{4}{n^2 h_n^4} \int_0^T (\partial_0 f(h_n, \theta, X_s))^2 n g_{\theta_0}(X_s) ds = O \left( \frac{1}{n} \right). \]

Then, (50) converges to 0 in probability. The second term of $M^{(2)} (\theta_0)$
\[ \frac{2}{h_n^2} \int_I \left( \int_0^T \hat{K} \left( x, h_n, X_s - x \right) 1_{X_s \in \partial_0 g_{\theta_0}(X_s) ds} \right)^2 dx \]
converges in probability to
\[ 2 \int_I (\partial_0 g_{\theta_0}(x))^2 dx. \]

Then,
\[ (51) \]
\[ M_n^{(2)} (\theta_0) \xrightarrow{P} 2 \int_I (\partial_0 g_{\theta_0}(u))^2 du. \]

The absolute value of $M^{(3)} (\hat{\theta}_n)$ has mean equal to $O(1)$. As $\hat{\theta}_n - \theta \xrightarrow{P} 0$,
\[ (52) \]
\[ (\hat{\theta}_n - \theta) M_n^{(3)} (\hat{\theta}_n) \xrightarrow{P} 0. \]

Using (49), (51) and (52),
\[ \sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{P} \sqrt{\int_I \frac{(\partial_0 g_{\theta_0}(u))^2}{l_T} g_{\theta_0}(u) du} \int_I (\partial_0 g_{\theta_0}(u))^2 du N(0, 1), \]
achieving the proof of (i).
Convergence of $M_n\left(\hat{\theta}_n\right)$ under $H_0$. $M_n\left(\hat{\theta}_n\right)$ is the sum of (45), (46) and (47), replacing $\theta$ by $\hat{\theta}_n$. Concerning (45), we have

$$\frac{16}{n^2 h_n^2} \int_0^T \int_{\mathbb{R}} \left( \int_I \bar{K} \left( x, h_n, \frac{X_s - x}{h_n} \right) \bar{K} \left( x, h_n, \frac{X_{u} - x}{h_n} \right) 1_{X_s \in I} 1_{X_{u} \in I} dx \right)^4 n^2 q \left( X_u \right) q \left( X_s \right) duds$$

which is equivalent to

$$\frac{8}{h_n n^2} \int_{\mathbb{R}} \left( \int_0^T w \left( u \right) w \left( u + p \right) K \left( u \right) K \left( u + p \right) du \right)^4 dp \int_I \frac{\left( q \left( y \right) \right)^2}{\left( \int_{\mathbb{R}} w \left( u \right) w \left( u + p \right) K \left( u \right) K \left( u + p \right) du \right)^2} dy = O_p \left( \frac{1}{n} \right)$$

and converges to 0. According to [29, Theorem 3],

$$\frac{2 \sqrt{h_n} n}{n^2 h_n^2} \int_0^T \int_{\mathbb{R}} \left( \int_I \bar{K} \left( x, h_n, \frac{X_s - x}{h_n} \right) \bar{K} \left( x, h_n, \frac{X_{u} - x}{h_n} \right) dx 1_{X_s \in I} 1_{X_{u} \in I} dM_u dM_s \right)^2 \left( q \left( X_u \right) - g_{\hat{\theta}_n} \left( X_u \right) \right) dx = O_p \left( \frac{\sqrt{h_n}}{n} \right) \overset{p}{\rightarrow} 0.$$

The last term corresponding to (47) is bounded by

$$\frac{2}{n^2 h_n^2} M \left( \left| \theta_0 - \hat{\theta}_n \right| \right) \int_0^T \int_I \bar{K} \left( x, h_n, \frac{X_s - x}{h_n} \right) 1_{X_s \in I} dM_s \int_0^T \bar{K} \left( x, h_n, \frac{X_{u} - x}{h_n} \right) 1_{X_{u} \in I} ds$$

equal to $O_p \left( \frac{1}{n^2} \right)$. Thus,

$$\frac{2 \sqrt{h_n} n}{nh_n^2} \int_0^T \int_I \bar{K} \left( x, h_n, \frac{X_s - x}{h_n} \right) 1_{X_s \in I} dM_s \int_0^T \bar{K} \left( x, h_n, \frac{X_{u} - x}{h_n} \right) \left( q \left( X_u \right) - g_{\hat{\theta}_n} \left( X_u \right) \right) 1_{X_{u} \in I} ds$$

is equal to $O_p \left( \sqrt{h_n} \right)$ and converges to 0 in probability. Finally,

$$n \sqrt{h_n} M_n \left( \hat{\theta}_n \right) \overset{p}{\rightarrow} \left[ 2 \int_{\mathbb{R}} \left( \int_0^T w \left( u \right) w \left( u + p \right) K \left( u \right) K \left( u + p \right) du \right)^2 dp \int_I \frac{\left( q \left( y \right) \right)^2}{\left( \int_{\mathbb{R}} w \left( u \right) w \left( u + p \right) K \left( u \right) K \left( u + p \right) du \right)^2} dy \right] \mathcal{N} \left( 0, 1 \right).$$

Convergence of $M_n\left(\hat{\theta}_n\right)$ under $H_1$. As the convergence of $M_n \left( \theta \right)$ is uniform in $\theta$, $M_n \left( \hat{\theta}_n \right)$ converges to $\inf_{\theta \in \Theta} \left\| q - g_{\theta} \right\|^2$ which is strictly positive under $H_1$. Thus, under $H_1$,

$$\left| n \sqrt{h_n} M_n \left( \hat{\theta}_n \right) \right| \rightarrow \infty.$$

Acknowledgements. I am grateful to Olivier Féron and Marc Hoffmann for helpful discussion and comments. This research is supported by the department OSIRIS (Optimization, Simulation, Risk and Statistics for Energy Markets) of EDF in the context of a CIFRE contract and by FiME (Finance for Energy Markets) Research Initiative.
References

[1] Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726, 1978.
[2] Yacine Aït-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 9(2):385–426, 1996.
[3] Yacine Aït-Sahalia, Julio Cacho-Diaz, and Roger JA Laeven. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606, 2015.
[4] Martin T Barlow and Marc Yor. Semi-martingale inequalities via the garsia-rodemich-rumsey lemma, and applications to local times. *Journal of Functional Analysis*, 49(2):198–229, 1982.
[5] Fred Espen Benth, Rüdiger Kiesel, and Anna Nazarova. A critical empirical study of three electricity spot price models. *Energy Economics*, 34(5):1589–1616, 2012.
[6] Fred Espen Benth, Nina Lange, and Tor Age Myklebust. Pricing and hedging quanto options in energy markets. *Journal of Energy Markets*, 2015.
[7] Fred Espen Benth and Jūratė Šaltytė Benth. Weather derivatives and stochastic modelling of temperature. *International Journal of Stochastic Analysis*, 2011, 2011.
[8] Maria Mori Brooks and J Stephen Marron. Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Applications*, 38(1):157–165, 1991.
[9] Alvaro Cartea and Marcelo G Figueroa. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4):313–335, 2005.
[10] G Castellan and F Letué. Estimation of the cox regression function via model selection. *F. Letué's PhD Thesis*, 2000.
[11] Feng Chen, Paul SF Yip, and KF Lam. On the local polynomial estimators of the counting process intensity function and its derivatives. *Scandinavian Journal of Statistics*, 38(4):631–649, 2011.
[12] Fabienne Comte, Stéphane Gaïffas, and Agathe Guilleux. Adaptive estimation of the conditional intensity of marker-dependent counting processes. In *Annales de l'institut Henri Poincaré (B)*, volume 47, pages 1171–1196, 2011.
[13] David R Cox. Regression models and life-tables. In *Breakthroughs in statistics*, pages 527–541. Springer, 1992.
[14] Sylvain Delattre, Christian Y Robert, and Mathieu Rosenbaum. Estimating the efficient price from the order flow: a brownian cox process approach. *Stochastic Processes and their Applications*, 123(7):2603–2619, 2013.
[15] Thomas Deschatre, Olivier Férон, and Marc Hoffmann. Estimating fast mean-reverting jumps in electricity market models. arXiv preprint arXiv:1803.03803, 2018.
[16] Peter Diggle. A kernel method for smoothing point process data. *Applied statistics*, pages 138–147, 1985.
[17] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, pages 1608–1632, 2011.
[18] Marc Hoffmann. On estimating the diffusion coefficient: parametric versus nonparametric. In *Annales de l'IHP Probabilités et statistiques*, volume 37, pages 339–372, 2001.
[19] Christian Houdré and Patricia Reynaud-Bouret. Exponential inequalities, with constants, for u-statistics of order two. In *Stochastic inequalities and applications*, pages 55–69. Springer, 2003.
[20] Alan Karr. *Point processes and their statistical inference*, volume 7. CRC press, 1991.
[21] Claire Lacour, Pascal Massart, and Vincent Rivoirard. Estimator selection: a new method with applications to kernel density estimation. *Sankhya A*, 70(2):298–335, 2017.
[22] Sarah Lemler et al. Oracle inequalities for the lasso in the high-dimensional aalen multiplicative intensity model. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 52, pages 981–1008. Institut Henri Poincaré, 2016.
[23] Matthieu Lerasle, Nelo Molter Magalhães, and Patricia Reynaud-Bouret. Optimal kernel selection for density estimation. In *High Dimensional Probability VII*, pages 425–460. Springer, 2016.
[24] Thilo Meyer-Brandis and Peter Tankov. Multi-factor jump-diffusion models of electricity prices. *International Journal of Theoretical and Applied Finance*, 11(05):503–528, 2008.
[25] Susan Albritton Murphy and Pranab Kumar Sen. Time-dependent coefficients in a cox-type regression model. *Stochastic Processes and their Applications*, 39(1):153–180, 1991.
[26] Jens P Nielsen and Oliver B Linton. Kernel estimation in a nonparametric marker dependent hazard model. *The Annals of Statistics*, pages 1735–1748, 1995.
[27] Yoshihiko Ogata and Koichi Katsura. Point-process models with linearly parametrized intensity for application to earthquake data. *Journal of applied probability*, 25(A):291–310, 1986.
[28] Finbarr O'Sullivan. Nonparametric estimation in the cox model. *The Annals of Statistics*, pages 124–145, 1993.
38 THOMAS DESCHATURE

[29] Giovanni Peccati, Murad S Taqqu, et al. Central limit theorems for double poisson integrals. Bernoulli, 14(3):791–821, 2008.

[30] Daniel Revuz and Marc Yor. Continuous martingales and Brownian motion, volume 293. Springer Science & Business Media, 2013.

[31] Patricia Reynaud-Bouret. Adaptive estimation of the intensity of inhomogeneous poisson processes via concentration inequalities. Probability Theory and Related Fields, 126(1):103–153, 2003.

[32] Patricia Reynaud-Bouret. Concentration inequalities, counting processes and adaptive statistics. In ESAIM: Proceedings, volume 44, pages 79–98. EDP Sciences, 2014.

[33] Peter Tankov. Financial modelling with jump processes, volume 2. CRC press, 2003.

[34] Wilson Truccolo, Uri T Eden, Matthew R Fellows, John P Donoghue, and Emery N Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. Journal of neurophysiology, 93(2):1074–1089, 2005.

[35] Alexandre B Tsybakov. Introduction to nonparametric estimation, 2009.

[36] Klaus J Utikal. Nonparametric inference for a doubly stochastic poisson process. Stochastic processes and their applications, 45(2):331–349, 1993.

[37] Suzanne Varet, Claire Lacour, Pascal Massart, and Vincent Rivoirard. Performances numériques du choix de fenêtre par pco lors de l’estimation de densités multivariées par méthode à noyau, 2017.

[38] Ju-Yi Yen and Marc Yor. Local times and excursion theory for brownian motion a tale of wiener and ito measures preface, 2013.

[39] Tingting Zhang and SC Kou. Nonparametric inference of doubly stochastic poisson process data via the kernel method. The annals of applied statistics, 4(4):1913, 2010.

THOMAS DESCHATURE, EDF Lab, Boulevard Gaspard Monge, Palaiseau, France.
E-mail address: thomas.deschatre@gmail.com