# Estimation of time-dependent latency with locally stationary Hawkes processes

Deniz Erdemlioglu<sup>\*</sup>Yoann Potiron<sup>†</sup>and Vladimir Volkov

#### Abstract

We consider estimation of latency, i.e. the time to learn an event and respond. We assume that the latency depends on time. We consider locally parametric Hawkes processes, where the baseline and the parameters of the kernels are time-dependent. We define latency as a known function of kernel parameters. We propose local estimation based on maximum likelihood. We characterize feasible statistics induced by central limit theory for the estimation procedure. We propose a test statistic for constancy of latency. The results are obtained with in-fill asymptotics. A numerical simulation corroborates the asymptotic theory. An empirical application to news data shows that the test for constancy of latency is always rejected and ???.

**Keywords**: latency matrix; time-dependent; Hawkes mutually exciting processes; local parametric estimation; constancy test; in-fill asymptotics; news data

## 1 Introduction

This paper concerns estimation of a latency matrix, i.e. the time to learn an event and respond. The latency can also be called a delay. We assume that the latency is a  $d \times d$  dimensional matrix. In the finance

<sup>\*?.</sup> E-mail: ?

<sup>&</sup>lt;sup>†</sup>(CONTACT) Faculty of Business and Commerce, Keio University. 2-15-45 Mita, Minato-ku, Tokyo, 108-8345, Japan. E-mail: potiron@fbc.keio.ac.jp website: http://www.fbc.keio.ac.jp/~potiron

literature, a common definition of latency is based on datasets that are not necessarily available to the statistician (see [Hasbrouck and Saar, 2013]). An alternative definition of latency is using a statistical model based on point processes which characterizes the event times (see [Potiron and Volkov, 2025]). The main stylized fact is the presence of event clustering in time. A popular specification targeting this relies on the so-called Hawkes mutually exciting processes (see [Hawkes, 1971b] and [Hawkes, 1971a]). If we define the point process as  $N_t$ , with  $\lambda$  its corresponding intensity and d as its dimension, a standard definition of Hawkes mutually exciting processes is given by

$$\lambda_t = \nu^* + \int_0^t h(t-s) \, dN_s.$$
 (1)

Here,  $\nu^*$  is a *d* dimensional Poisson baseline and *h* is a  $d \times d$  dimensional kernel matrix. If we define  $\theta^*$  as the parameters of the kernel, we restrict to a parametric specification

$$\lambda_t = \nu^* + \int_0^t h(t - s, \theta^*) \, dN_s. \tag{2}$$

Then, we can define the  $d \times d$  dimensional latency matrix as a known function F of kernel parameters

$$L = F(\theta^*). \tag{3}$$

Since latency is not well-defined with an exponential kernel, we consider generalized gamma kernels. The main novelty in this paper is that the latency matrix depends on time.

The main application of latency lies in finance. [Gagnon and Karolyi, 2010] show that price parity deviations relate positively to proxies for holding costs that can limit arbitrage. The empirical application from [Hasbrouck and Saar, 2013] suggests that high-frequency trading is beneficial to market quality. In [Hoffmann, 2014], fast traders can revise their quotes quickly after news arrivals to reduce market risks. [Budish et al., 2015], [Biais et al., 2015], [Foucault et al., 2016] and [Pagnotta and Philippon, 2018] also consider trading speed. [Potiron and Volkov, 2025] propose estimation of latency.

When seen as a delay, there are also applications in management. [Dong et al., 2019] investigate the impact of delay on the coordination within hospitals. [Gámiz et al., 2022] and [Gámiz et al., 2023] consider nonparametric local estimation of Hawkes processes and applications to pandemic. There are also

applications in seismology (see [Nolet and Dahlen, 2000]), insurance (see [Lesage et al., 2022]), criminology (see [Nagin and Pogarsky, 2004]), sociology (see [Lahad, 2012]) and medicine (see [Harris, 1990]).

The main application of Hawkes processes lies in seismology (see [Rubin, 1972], [Vere-Jones, 1978], [Ozaki, 1979], [Vere-Jones and Ozaki, 1982], [Ogata, 1978]). [Ikefuji et al., 2022] analyze the impact of earthquake risk based on marked Hawkes processes. There are also applications in financial econometrics (see [Yu, 2004], [Bowsher, 2007], [Embrechts et al., 2011], [Aït-Sahalia et al., 2014]), finance (see [Large, 2007], [Aït-Sahalia et al., 2015] and [Fulop et al., 2015]) and quantitative finance (see [Chavez-Demoulin et al., 2005], [Bacry et al., 2013], [Jaisson and Rosenbaum, 2015]). See the references in [Liniger, 2009] and [Hawkes, 2018]. More recently, [Corradi et al., 2020] develop a test for conditional independence in quadratic variation of jumps. A bootstrap approach is developed in [Cavaliere et al., 2023].

To allow latency to depend on time, we introduce locally parametric Hawkes mutually exciting processes

$$\lambda_t = \nu_t^* + \int_0^t h(t - s, \theta_s^*) \, dN_s. \tag{4}$$

Here, the baseline and the parameters of the kernel are time-dependent. Then, time-dependent latency is defined as a known function F of the kernel parameters

$$L_t = F(\theta_t^*). \tag{5}$$

The model (1) defines a class of locally stationary processes (see [Fan, 1993] and [Dahlhaus, 1996]). There are some examples. [Chen and Hall, 2013], [Kwan et al., 2023] and [Kwan, 2023] allow for a time-dependent parametric baseline, with time-invariant kernel parameters. [Clinet and Potiron, 2018] consider random time-dependent baseline and random time-dependent kernel parameters, in the exponential kernel case. There are also some other related papers on locally stationary Hawkes processes. [Roueff et al., 2016] and [Roueff and Von Sachs, 2019] propose nonparametric estimation based on local Bartlett spectrum. [Omi et al., 2017] study a Bayesian method with time-dependent parametric baseline. Spectral parametric estimation for misobserved Hawkes processes with a setting also covering a time-dependent baseline is given in [Cheysson and Lang, 2022]. Nonparametric estimation based on B-splines is given by [Mammen and Müller, 2023]. [Potiron et al., 2025] propose nonparametric estimation of Ito semimartingale baseline.

We focus on in-fill asymptotics, i.e. when T is fixed and the number of observations on [0, T] increases as  $n \to \infty$ . These asymptotics are popular with financial applications based on high-frequency data (see [Aït-Sahalia and Jacod, 2014]). The main reason why we use these asymptotics is that we observe time-dependent latency during the day (see our empirical study), and between different days (see Figures 1 and 2 in [Potiron and Volkov, 2025]). There already exists work to accommodate for in-fill asymptotics with Hawkes processes. In-fill asymptotic results from [Chen and Hall, 2013] are based on random observation times of order n. A single boosting of the baseline, i.e.  $\lambda_t = \alpha \nu_t^* + \int_0^t h(t-s, \theta^*) dN_s$ , is considered where  $\alpha \to \infty$  is a scaling sequence. [Clinet and Potiron, 2018] introduce a joint boosting of the baseline and the kernel, i.e.  $\lambda(t) = n\nu_t^* + \int_0^t na_s^* \exp(-nb_s^*(t-s)) dN_s$ . [Kwan et al., 2023] revisit [Chen and Hall, 2013] with the same in-fill asymptotics as in [Clinet and Potiron, 2018], i.e.  $\lambda_t = n\nu_t^* + \int_0^t na^* \exp(-nb^*(t-s)) dN_s$ . [Kwan, 2023], [Potiron and Volkov, 2025] and [Potiron et al., 2025] also use these in-fill asymptotics.

We propose local estimation based on maximum likelihood estimation (MLE). The latency estimator is defined as the known function of estimated kernel parameters. When the point process is stationary and ergodic, [Ogata, 1978] shows the central limit theory (CLT) for MLE. However, the definition of ergodicity is vague in that paper. Most of the papers on inference for Hawkes processes with parametric kernel make this ergodicity assumption (see, e.g., [Cavaliere et al., 2023], Assumption 1(b) and Remark 2.1). In fact, [Clinet and Yoshida, 2017] exhibit the conditions required, i.e. ergodicity of the Hawkes intensity process and its derivative. They consider general point processes and derive the CLT for MLE in Theorem 3.11 (p. 1809) under these ergodicity assumptions. They also show these ergodicity assumptions in the case of a Hawkes process with exponential kernel in Theorem 4.6 (p. 1821). The proofs rely heavily on the Markov property of the exponential distribution. [Kwan, 2023] considers the non-exponential kernel case but the author mentions that such case is challenging since the Hawkes intensity process is non-Markovian, thus rendering standard Markov tools inapplicable. Consequently, the author can only show the ergodicity for the Hawkes intensity process itself but not for its derivative. Thus, he can only show the consistency of the MLE in Theorem 3.4.3 (p. 73). When the kernel follows a generalized gamma distribution, [Potiron and Volkov, 2025] can show that the ergodicity assumptions are satisfied and also obtain the CLT of the MLE and latency estimation. This is due to the exponentially decreasing nature of the kernel. In the absence of latency and when the kernel is exponential, Theorem 5.4 (p. 3480) in [Clinet and Potiron, 2018] and Theorem 3.2 (p. 78) in [Kwan et al., 2023] give the CLT for MLE for locally stationary Hawkes processes. [Kwan et al., 2023] also provide a test for baseline constancy in Theorem 4.1 (p. 79).

All these results are useful, but none of them consider locally parametric Hawkes processes with generalized gamma kernels, and estimation of latency with in-fill asymptotics. In our Theorem 1, we give the CLT for MLE of the integral of parameter. We provide the CLT for MLE of the integral of latency in our Theorem 2. We also provide feasible statistics induced by the CLT. We finally introduce a Wald test for constancy of the latency matrix. This test compares the estimation of latency between two consecutive intervals. Corollary 1 shows that the Wald test statistic converges in distribution to a chi-squared distribution under the null hypothesis and is consistent under the alternative hypothesis. Our proof strategy follows the general machinery of [Clinet and Yoshida, 2017], which consider large-T asymptotics when  $T \to \infty$ . To rewrite our problem with in-fill asymptotics as a problem with large-T asymptotics, we consider a time transformation (see [Clinet and Potiron, 2018], [Kwan et al., 2023] and [Potiron and Volkov, 2025]). The main novelty in the proofs is in showing that the local approximation from [Clinet and Potiron, 2018] stays robust to the generalized gamma kernel.

???

The rest of this paper is organized as follows. The setting is introduced in Section 2. Estimation and tests are given in Section 3. The theory is developed in Section 4. Our numerical study is carried in Section 5. Our empirical application is provided in Section 6. We conclude in Section 7. The supplementary materials contain all the proofs of the manuscript.

## 2 Setting

In this section, we introduce locally parametric Hawkes processes, where the baseline and the parameters of the kernels are random time-dependent. We also introduce the random time-dependent latency when the horizon T is finite.

For any space S such that  $0 \in S$ , we define the space without zero as  $S_*$ . For a vector V, we denote its i-th component as  $V^{(i)}$ . In what follows, we introduce the multidimensional point process  $N_t$ . For  $i = 1, \dots, d$ , each component of the point process  $N_t^{(i)}$  counts the number of events between 0 and t. We define  $N^{(i)}$  as a simple point process on [0,T], i.e., a family  $\{N^{(i)}(C)\}_{C\in\mathcal{B}([0,T])}$  of random variables with values in the space of natural integers N. Here,  $\mathcal{B}([0,T])$  is the Borel  $\sigma$ -algebra on the compact space [0,T],  $N^{(i)}(C) = \sum_{k \in \mathbb{N}} \mathbf{1}_C(\tau_k^{(i)})$  and  $\{\tau_k^{(i)}\}_{k \in \mathbb{N}}$  is a sequence of  $\mathbb{R}^+$ -valued event times, which are random. We assume that the first time is equal to 0 and the following times are increasing for each process a.s., i.e.  $\mathbb{P}(\tau_0^{(i)} = 0 < \tau_1^{(i)} < \ldots < \tau_{N_T^{(i)}}^{(i)} < T < \tau_{N_T^{(i)}+1}^{(i)}$  for  $i = 1, \cdots, d$  = 1. We also assume that no events happen at the same time for different processes a.s., i.e.  $\mathbb{P}(\tau_k^{(i)} \neq \tau_l^{(j)} \text{ for } k, l \in \mathbb{P}(\tau_k^{(i)} \neq \tau_l^{(j)})$  $\mathbb{N}_*$  and  $i, j = 1, \cdots, d$  s.t.  $i \neq j$  = 1. Let  $\mathcal{B} = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0,T]}, \mathbb{P})$  be a filtered probability space which satisfies the usual conditions. For any process  $X_t$ , the canonical filtration of  $X_t$  is defined as  $\mathcal{F}_t^X = \sigma\big(X(C), C \in \mathcal{B}([0,T]), C \subset [0,t]\big). \text{ We assume that, for any } t \in [0,T], \text{ the canonical filtration of } C \in \mathcal{B}([0,T]), C \subset [0,t]\big).$  $N_t$  included in the main filtration, i.e.  $\mathcal{F}_t^N \subset \mathcal{F}_t$ . Any nonnegative  $\mathcal{F}_t$ -progressively measurable process  $\{\lambda_t\}_{t\in[0,T]}$ , which is *d*-dimensional, such that  $\mathbb{E}[N((a,b]) \mid \mathcal{F}_a] = \mathbb{E}\left[\int_a^b \lambda_s ds \mid \mathcal{F}_a\right]$  a.s. for all intervals  $(a,b] \subset [0,T]$ , is called an  $\mathcal{F}_t$ -intensity of  $N_t$ . Intuitively, the intensity corresponds to the expected number of events given the past information, i.e.,

$$\lambda_t = \lim_{u \to 0} \mathbb{E} \Big[ \frac{N_{t+u} - N_t}{u} \mid \mathcal{F}_t \Big] \text{ a.s.}$$

For background on point processes, the reader can consult [Jacod, 1975], [Jacod and Shiryaev, 2003], [Daley and Vere-Jones, 2003], and [Daley and Vere-Jones, 2008].

For a matrix  $\phi$ , we denote its component (i, j) as  $\phi^{(i,j)}$ . The present work is concerned with locally parametric mutually exciting Hawkes processes, i.e. point processes N admitting an  $\mathcal{F}_t$ -intensity equal to

$$\lambda_t = \nu_t^* + \int_0^t h(t - s, \theta_s^*) \, dN_s. \tag{6}$$

Here,  $\nu_t$  is a *d* dimensional random time-dependent baseline. Moreover, the parametric kernel  $h(t, \theta)$  is a  $d \times d$  dimensional matrix. Its diagonal components  $h^{(i,i)}(t, \theta)$  are raising the probability of observing events from the *i*th process when there are events of the *i*th process, while non-diagonal components  $h^{(i,j)}(t, \theta)$  are raising the probability of observing events from the *i*th process when there are events of the *j*th process. Finally,  $\theta_t^*$  are the kernel parameters, which are time-dependent and random.

The random time-dependent latency is defined as a  $d \times d$  dimensional matrix which is a timeinvariant known function of the kernel parameters  $\theta_t^*$ , i.e.

$$L_t = F(\theta_t^*). \tag{7}$$

With a latency matrix, we can study the latency of each individual process, but also the latency between two different processes. More specifically, a latency between events from the *j*th process and its impact on events from the *i*th process is introduced at time *t* if  $L_t^{(i,j)} > 0$ . In this paper, we set *F* such that the latency  $L_t^{(i,j)}$  is equal to the time required to reach the pick of the kernel  $h^{(i,j)}(t, \theta_t^*)$ , i.e. the mode. This definition of latency is in agreement with the finance literature, which defines latency as the time it takes to learn and generate response to a trading event (see [Hasbrouck and Saar, 2013]). An advantage of this definition is that latency can be characterized by parameters  $\theta_t^{(i,j)}$  associated with factors affecting latency. Such a structural approach permits identification of different aspects of latency.

This paper targets estimation of the integral of latency on [0, T] where T is finite, i.e.

$$IL(T) = \int_0^T L_t dt.$$
(8)

As far as we know, the problem of integral of latency (8) is novel to the literature. It echoes the so-called integrated variance problem. We also define the couple of baseline parameters and kernel parameters as  $P_t^* = (\nu_t^*, \theta_t^*)$ . Another goal is to estimate the integral of the parameters, i.e.

$$IP(T) = \int_0^T P_t^* dt.$$
(9)

The problem for the integral of the parameters is not novel to the literature, since it was already considered in [Clinet and Potiron, 2018].

### 3 Estimation

In this section, we introduce locally parametric Hawkes processes with in-fill asymptotics, local parametric estimation based on MLE, latency estimation, and the test for constancy of latency.

We prefer most of the time not to write explicitly the dependence on n, and any limit theorem refers to the convergence when  $n \to \infty$ . For inference purposes, we consider in-fill asymptotics with joint boosting of the baseline and the kernel, i.e.

$$\lambda_t = n\nu_t^* + \int_0^t nh(n(t-s), \theta_s^*) \, dN_s.$$
(10)

Here, in-fill asymptotics are based on random observation times of order n within the time interval [0, T] for a finite horizon time T. They extend the asymptotic analysis of [Clinet and Potiron, 2018], [Kwan et al., 2023], [Kwan, 2023], [Potiron and Volkov, 2025] and [Potiron et al., 2025], also based on joint boosting, by not imposing an exponential or time-invariant parameters. They are different from [Chen and Hall, 2013] in-fill asymptotics which considers no boosting of the kernel. Here, in-fill asymptotics are desirable because we can incorporate random features of the baseline and the parameters into asymptotic variances in the CLT.

We denote the floor function by  $\lfloor \cdot \rfloor$ . For a finite horizon T, we consider  $M = \lfloor T/\Delta \rfloor$  intervals  $[T_{l-1}, T_l)$  with equal length  $\Delta$ , where  $T_l = l\Delta$  are the start and end points of each interval. For  $l = 1, \ldots, M$ , we rely on the log likelihood process (see [Ogata, 1978] and [Daley and Vere-Jones, 2003]) on the *l*-th interval  $[T_{l-1}, T_l)$ , i.e.

$$l_l(P) = \sum_{i=1}^d \int_{T_{l-1}}^{T_l} \log(\lambda_t^{(i)}(P)) dN_t^{(i)} - \sum_{i=1}^d \int_{T_{l-1}}^{T_l} \lambda_t^{(i)}(P) dt.$$

Here,  $P = (\nu, \theta)$  are the parameters of the baseline  $\nu$  and the parameters of the kernel  $\theta$ , and they belong to the parameter space  $\Theta = (\Theta_{\nu}, \Theta_{h})$ . Also, the intensity of the point process at the parameter P is defined as

$$\lambda_t(P) = n\nu + \int_0^t nh(n(t-s),\theta) \, dN_s.$$

We denote the total number of parameters by m, thus  $\Theta \subset \mathbb{R}^m$ . Since each baseline has exactly one parameter, the number of parameters from the multidimensional baseline  $\Theta_h$  is equal to d. We naturally assume that  $2d \leq m$ . Then, the local MLE is defined as a maximizer of the local log likelihood process, i.e.

$$\widehat{P}_l = (\widehat{\nu}_l, \widehat{\theta}_l) \in \operatorname{argmax}_{P \in \Theta} l_l(P).$$

Finally, we propose an estimator for the integral of latency and the integral of the parameter as

$$\widehat{IL}(T) = \sum_{l=1}^{M} F(\widehat{\theta}_l - b_h(\widehat{\theta}_l)) \Delta, \qquad (11)$$

$$\widehat{IP}(T) = \sum_{l=1}^{M} \left( \widehat{P}_l - b(\widehat{P}_l) \right) \Delta.$$
(12)

Here, b and  $b_h$  correspond to the bias corrections required for local estimation, and are defined in what follows.

We define the space E as  $E = \mathbb{R}^+_* \times \mathbb{R}^+_* \times \mathbb{R}^m$ . We also define as  $C_{\uparrow}(E, \mathbb{R})$  the set of continuous functions  $\psi : (u, v, w) \to \psi(u, v, w)$  from E to  $\mathbb{R}$  that satisfy  $\psi$  is of polynomial growth in  $u, v, w, \frac{1}{u}$  and  $\frac{1}{v}$ . For any  $P^* \in \Theta$  and any  $P \in \Theta$ , we define the intensity process at the time-invariant parameter P when the true parameter is time-invariant equal to  $P^*$  as  $\lambda_t(P^*, P)$ . We also define the rescaled time-transformed intensity process at the time-invariant parameter P when the true parameter is timeinvariant equal to  $P^*$  as  $\overline{\lambda}_t(P^*, P) = \frac{\lambda_{t/n}(P^*, P)}{n}$ . For any  $i = 1, \ldots, d$ , we define the triplet of the ith process as  $\overline{X}_t^{(i)}(P^*, P) = (\overline{\lambda}_t^{(i)}(P^*, P^*), \overline{\lambda}_t^{(i)}(P^*, P), \partial_P \overline{\lambda}_t^{(i)}(P^*, P))$ . Propositions C1 and C2 from the supplementary materials of [Potiron and Volkov, 2025] state that  $\overline{X}_t^{(i)}(P^*, P)$  is stable, i.e. there exists an  $\mathbb{R}^*_+$ -valued random variable  $\overline{\lambda}_l^{(i)}(P^*, P)$  such that  $\overline{X}_{nT}^{(i)}(P^*, P) \stackrel{\mathcal{D}}{\to} (\overline{\lambda}_l^{(i)}(P^*, P^*), \overline{\lambda}_l^{(i)}(P^*, P))$ . They also state that the triplet is ergodic, i.e. there exists a mapping  $\pi_{P^*}^{(i)} : C_{\uparrow}(E, \mathbb{R}) \times \Theta \to \mathbb{R}$  such that for any  $(\psi, P) \in C_{\uparrow}(E, \mathbb{R}) \times \Theta$  we have  $\frac{1}{nT} \int_{0}^{nT} \psi(X_{s}^{(i)}(P^{*}, P)) ds \xrightarrow{\mathbb{P}} \pi_{P^{*}}^{(i)}(\psi, P)$ , where  $\pi_{P^{*}}^{(i)}(\psi, P) = \mathbb{E}[\psi(\overline{\lambda}_{l}^{(i)}(P^{*}, P^{*}), \overline{\lambda}_{l}^{(i)}(P^{*}, P), \partial_{\theta}\overline{\lambda}_{l}^{(i)}(P^{*}, P))]$ . Finally, they state that there exists a probability measure  $\Pi_{P^{*}}^{(i)}$  on  $(E, \mathbf{B}(E))$  such that for any  $\psi \in C_{\uparrow}(E, \mathbb{R})$ , we have  $\pi_{P^{*}}^{(i)}(\psi, \theta) = \int_{E} \psi(u, v, w) \Pi_{P^{*}}^{(i)}(du, dv, dw)$ . If we consider a vector  $z \in \mathbb{R}^{m}$ , we define the tensor product as  $z^{\otimes 2} = z \times z^{T} \in \mathbb{R}^{m \times m}$ . Thus, we can define the  $m \times m$  dimensional Fisher information matrix  $\Gamma$  when the true parameter is time-invariant equal to  $P^{*}$  as

$$\Gamma(P^*) = \sum_{i=1}^d \int_E w^{\otimes 2} \frac{1}{u} \Pi_{P^*}^{(i)}(du, dv, dw).$$
(13)

This means that  $\Gamma^{-1}(P^*)$  is the asymptotic covariance matrix when the true parameter is time-invariant equal to  $P^*$ . We can naturally define the asymptotic covariance matrix for estimation of parameter integral as

$$c_{IP}(t)c_{IP}(t)^{T} = \Gamma(P_{t}^{*})^{-1} \text{ and } C_{IP}(T) = \int_{0}^{T} c_{IP}(t)c_{IP}(t)^{T}dt.$$
 (14)

For l = 1, ..., M, we define the rescaled time-transformed likelihood on the *l*th interval at the time-invariant parameter *P* when the true parameter is time-invariant equal to  $P^*$  as

$$\bar{l}_{l}(P^{*},P) = \sum_{i=1}^{d} \int_{T_{l-1}n}^{T_{l}n} \log(\bar{\lambda}_{t}^{(i)}(P^{*},P)) d\bar{N}_{t}^{(i)} - \sum_{i=1}^{d} \int_{T_{l-1}n}^{T_{l}n} \bar{\lambda}_{t}^{(i)}(P^{*},P) dt.$$

Here, we define  $\overline{N}_t^{(i)} = N_t^{(i)}$ , for  $t \in [0, nT]$ , as the time-transformed point process. Then, we propose local estimation of the inverse Fisher information matrix as

$$\widehat{\Gamma}_{l}^{-1} = -\partial_{P}^{2} \overline{l}_{l} \big( \widehat{P}_{l} - b(\widehat{P}_{l}), \widehat{P}_{l} - b_{h}(\widehat{\theta}_{l}) \big).$$
(15)

Here,  $\partial_P^2 \bar{l}_l(P^*, P)$  is the  $m \times m$  dimensional Hessian matrix of  $\bar{l}_l(P^*, P)$ . Finally, we propose estimation for the asymptotic covariance matrix of the parameter integral as

$$\widehat{C}_{IP}(T) = \sum_{l=1}^{M} \widehat{\Gamma}_l^{-1} \Delta.$$
(16)

For any  $P^* \in \Theta$ , we define the rescaled time-transformed point process when the true parameter is time-invariant equal to  $P^*$  as  $\overline{N}_t(P^*)$ . We also define the rescaled time-transformed likelihood as

$$\bar{l}(P^*, P) = \sum_{i=1}^d \int_0^{Tn} \log(\bar{\lambda}_t^{(i)}(P^*, P)) d\overline{N}_t^{(i)}(P^*) - \sum_{i=1}^d \int_0^{Tn} \bar{\lambda}_t^{(i)}(P^*, P) dt.$$

We introduce for any  $i = 1, \dots, d \; M^{(i)}(P^*) = \int_0^{Tn} \frac{\partial_P \lambda_t^{(i)}(P^*, P^*)}{\lambda_t^{(i)}(P^*, P^*)} (dN_t^{(i)}(P^*) - \lambda_t^{(i)}(P^*, P^*) dt)$  and  $K(P^*) = \frac{1}{Tn} \partial_P^3 \overline{l}(P^*, P^*) \in \mathbb{R}^{m \times m \times m}$ . We introduce, for indices  $k, l, q \in \{1, \dots, m\}, \; C(P^*)_{k, lq} = \sum_{i=1}^d \frac{1}{Tn} \int_0^{Tn} \partial_P \lambda_t^{(i,k)}(P^*, P^*) \partial_P^2 \log(\lambda_t^{(i,l,q)}(P^*, P^*)) dt$  and

$$Q(P^*)_{k,lq} = \sum_{i=1}^d -\frac{M^{(i,k)}(P^*)}{Tn} \int_0^{Tn} \frac{\partial_P \lambda_t^{(i,l)}(P^*, P^*) \partial_P \lambda_t^{(i,q)}(P^*, P^*)}{\lambda_t^{(i)}(P^*, P^*)} dt$$

Lemma 1 in the state that K, C and Q converge in probability to limit values  $K_l$ ,  $C_l$  and  $Q_l$ . Then, we define for any  $k \in \{1, \dots, m\}$  the k-th component of the bias function as

$$b(P^*)^{(k)} = \frac{1}{2} \sum_{q=1}^m \sum_{l=1}^m \sum_{j=1}^m \Gamma(P^*)^{(j,k)} \Gamma(P^*)^{(l,q)} (K_l(P^*)^{(j,l,q)} + 2\{C_l(P^*)_{l,jq} + Q_l(P^*)_{l,jq}\}).$$
(17)

For convenience we rewrite the  $d \times d$  dimensional matrix of latencies  $L_t$  and its integral IL(T) as a  $d^2$  dimensional vector of latencies  $\overline{L}_t = (L_t^{(1,1)}, L_t^{(1,2)}, \cdots, L_t^{(d,d)})^T$  and

$$\overline{IL}(T) = (IL(T)^{(1,1)}, IL(T)^{(1,2)}, \cdots, IL(T)^{(d,d)})^T.$$

We also rewrite the  $d \times d$  dimensional matrix for estimation of latency integrals  $(\widehat{IL}(T))$  as a  $d^2$  dimensional vector  $\widehat{TL}(T) = (\widehat{IL}^{(1,1)}(T), \widehat{IL}^{(1,2)}(T), \cdots, \widehat{IL}^{(d,d)}(T))^T$ . We denote the bias function restricted to the kernel parameter  $\theta$  by  $b_h(\theta)$ . We also denote the Fisher information matrix restricted to the kernel parameter  $\theta$  by  $\Gamma_h(\theta)$ . For any  $i = 1, \ldots, d$  and  $j = 1, \ldots, d$ , we define the differential vector of  $F^{(i,j)}$  at the kernel parameter  $\theta$ , which is (m-d) dimensional, as  $dF^{(i,j)}(\theta) = (dF^{(i,j,1)}(\theta), \cdots, dF^{(i,j,m-d)}(\theta))$ . We introduce the  $d^2 \times d^2$  dimensional asymptotic covariance matrix  $\overline{\Gamma}(\theta^*)^{-1}$  satisfying

$$(\overline{\Gamma}(\theta^*)^{-1})^{((i-1)d+j,(k-1)d+l)}$$

$$= \sum_{q=1}^{m-d} \Big( \sum_{r=1}^{m-d} dF^{(i,j,r)}(\theta^*) \big( \Gamma_h(\theta^*)^{-1/2} \big)^{(r,q)} \Big) \Big( \sum_{r=1}^{m-d} dF^{(k,l,r)}(\theta^*) \big( \Gamma_h(\theta^*)^{-1/2} \big)^{(r,q)} \Big),$$

$$(18)$$

for any i = 1, ..., d, j = 1, ..., d, k = 1, ..., d and l = 1, ..., d. We can naturally define the asymptotic covariance matrix for estimation of latency integral as

$$c_{\overline{IL}}(t)c_{\overline{IL}}(t)^T = \overline{\Gamma}(\theta_t^*)^{-1} \text{ and } C_{\overline{IL}}(T) = \int_0^T c_{\overline{IL}}(t)c_{\overline{IL}}(t)^T dt.$$
(19)

Then, we propose local estimation of the asymptotic covariance matrix as

$$(\widehat{\Gamma}_{l}^{-1})^{((i-1)d+j,(k-1)d+l)}$$

$$= \sum_{q=1}^{m-d} \Big( \sum_{r=1}^{m-d} dF^{(i,j,r)}(\widehat{\theta}_{l}^{*}) \big( \Gamma_{h}(\widehat{\theta}_{l}^{*})^{-1/2} \big)^{(r,q)} \Big) \Big( \sum_{r=1}^{m-d} dF^{(k,l,r)}(\widehat{\theta}_{l}^{*}) \big( \Gamma_{h}(\widehat{\theta}_{l}^{*})^{-1/2} \big)^{(r,q)} \Big),$$

$$(20)$$

for any i = 1, ..., d, j = 1, ..., d, k = 1, ..., d and l = 1, ..., d. Finally, we propose estimation for the asymptotic covariance matrix of the latency integral as

$$\widehat{C}_{\overline{IL}}(T) = \sum_{l=1}^{M} \widehat{\overline{\Gamma}}_{l}^{-1} \Delta.$$
(21)

We finally introduce a Wald test of constancy for a linear hypotheses on the  $d^2$  dimensional latency vector. This test compares the estimation of latency between two consecutive intervals. This test is based on the  $a \times d^2$  dimensional matrix A. We define the null hypothesis as  $H_0$ :  $\{A\overline{L}_t \text{ is constant for all } t \in [0,T]\}$  and the alternative hypothesis as  $H_1 : \{A\overline{L}_t \text{ is not constant for all } t \in [0,T]\}$ . For  $l = 1, \dots, M$ , we define the local estimation for the  $d^2$  dimensional vector of latency as  $\widehat{L}_l = (F(\widehat{\theta}_l - b_h(\widehat{\theta}_l))^{(1,1)}, F(\widehat{\theta}_l - b_h(\widehat{\theta}_l))^{(1,2)}, \dots, F(\widehat{\theta}_l - b_h(\widehat{\theta}_l))^{(d,d)})^T$ . We let our test statistic be

$$S(T) = \frac{n}{2} \sum_{l=1}^{M-1} \left( A(\widehat{\overline{L}}_l - \widehat{\overline{L}}_{l+1}) \right)^T \left( A\widehat{\overline{\Gamma}}_l^{-1} A^T \right)^{-1} \left( A(\widehat{\overline{L}}_l - \widehat{\overline{L}}_{l+1}) \right) \Delta.$$
(22)

#### 4 Theory

In this section, we start with showing an existence result for locally parametric Hawkes mutually exciting processes, where the baseline and the parameters of the kernels are random time-dependent. Then, we characterize feasible statistics induced by CLT for MLE of the integral of parameters and the integral of latency. Finally, we show the Wald test statistic asymptotic properties.

For any t > T, we define the kernel parameter fixed to its value in T as  $\theta_t^* = \theta_T^*$ . Then, we define the integral of the kernel matrix h for a time-dependent kernel parameter  $\theta_t^*$  from the time t as  $\phi_t = \int_0^\infty h(s, \theta_{t+s}^*) ds$ . For a matrix  $\phi$ , we denote its spectral radius as  $\rho(\phi)$ . Let us introduce a set of conditions required for the existence of locally parametric Hawkes mutually exciting processes where the baseline and the parameters of the kernels are time-dependent.

- Condition 1. (a) The parameter  $P_t^*$  belongs to  $\Theta$  a.s., i.e.  $\mathbb{P}(P_t^* \in \Theta \ \forall t \in [0,T]) = 1$ .
  - (b) For  $i = 1, \dots, d$ , the ith component of the baseline is positive a.s., i.e.  $\mathbb{P}(\nu_t^{*,(i)} > 0 \ \forall t \in [0,T]) = 1$ .
  - (c) For  $i = 1, \dots, d$ , the ith component of the baseline is integrable a.s., i.e.  $\mathbb{P}(\int_0^T \nu_s^{*,(i)} ds < \infty) = 1$ .
  - (d) For any  $0 \le t \le T$ , we have  $\mathcal{F}_t = \mathcal{F}_t^{P^*} \lor \mathcal{F}_t^{\underline{N}}$ , where the filtration  $\mathcal{F}_t^{P^*}$  is independent from the filtration  $\mathcal{F}_t^{\underline{N}}$ . We also have  $\underline{N}$  is a 2*d* dimensional  $\mathcal{F}_t$ -adapted Poisson process of intensity 1 that generates  $N_t$ , i.e.  $N_t^{(i)} = \int_{[0,t]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_s^{(i)}]}(x)\underline{N}^{(2i-1)} * \underline{N}^{(2i)}(ds \times dx)$  for  $i = 1, \cdots, d$ .
  - (e) For  $i = 1, \dots, d$  and  $j = 1, \dots, d$ , the (i, j)th component of the kernel is positive a.s., i.e.  $\mathbb{P}(h^{(i,j)}(s, \theta^*_{t+s}) \ge 0 \ \forall (t, s) \in [0, T]^2) = 1.$
  - (f) There exists a real number strictly between 0 and 1, i.e. 0 < r < 1, such that the spectral norm of the kernel matrix integral from the time t is smaller than r a.e. a.s., i.e.  $\mathbb{P}(\rho(\phi_t) \leq r \ \forall t \in [0,T]) = 1$ .

Condition 1 (b) implies that the point processes are well-defined, and is a generalization of Assumption 1 (a) in [Potiron et al., 2025] to the multidimensional case. Condition 1 (c) is also required in the simpler case of heterogeneous Poisson processes without a kernel (see [Daley and Vere-Jones, 2003]). This is a generalization of [Clinet and Potiron, 2018] (see Assumption E (ii), p. 3476) and also [Potiron et al., 2025] to the multidimensional case. Condition 1 (d) is a generalization of Poisson imbedding (see [Brémaud and Massoulié, 1996], Section 3, pp. 1571-1572), [Clinet and Potiron, 2018] (see the last sentence before Theorem 5.1, p. 3476), [Potiron et al., 2025]) to the multidimensional case. Condition 1 (e) restricts to the case of Hawkes processes with exhibition. Finally, Condition 1 (f) is a generalization of the assumptions used in [Clinet and Yoshida, 2017] (Proposition 4.4, pp. 1819-1820) and [Clinet and Potiron, 2018] (see Assumption E (i), p. 3476) to the multidimensional and time-dependent kernel case.

We provide now our existence result bringing new theory for multidimensional point processes. It is obtained by extending the proof machinery of Poisson imbedding for time-invariant two-dimensional Hawkes processes (see Theorem 7 (p. 1585) in [Brémaud and Massoulié, 1996]) to the time-dependent case. It also complements Theorem 5.1 (p. 3476) in [Clinet and Potiron, 2018] in which the kernel is exponential, and Proposition 4.1 in [Potiron et al., 2025] in which the kernel parameters are time-invariant.

**Proposition 1.** Under Condition 1, there exists an  $\mathcal{F}_t$ -adapted multidimensional point process  $N_t$  with an  $\mathcal{F}_t$ -intensity of the form (4).

We denote the gamma function by  $\gamma$ . For any i = 1, ..., d and j = 1, ..., d, we define the (i, j)th component for the mixture of generalized gamma kernels as

$$h^{(i,j)}(t,\theta^{(i,j)}) = \sum_{k=1}^{K^{(i,j)}} \alpha_k^{(i,j)} \frac{p_k^{(i,j)} t^{(D_k^{(i,j)}-1)} \exp(-(t/\beta_k^{(i,j)})^{p_k^{(i,j)}})}{(\beta_k^{(i,j)})^{D_k^{(i,j)}} \gamma(D_k^{(i,j)}/p_k^{(i,j)})}.$$
(23)

Here,  $\alpha_k^{(i,j)} \in \mathbb{R}^*_+$  is the size of the jump,  $\beta_k^{(i,j)} \in \mathbb{R}^*_+$  is the scale parameter,  $D_k^{(i,j)} \in \mathbb{R}^*_+$  and  $p_k^{(i,j)} \in \mathbb{R}^*_+$ are shape parameters. Moreover,  $K^{(i,j)}$  is the known number of terms. We assume that the kernel parameter is of the form

$$\theta = (\theta^{(i,j)})_{1 \le i,j \le d} = (\theta^{(1,1)}, \theta^{(1,2)}, \cdots, \theta^{(d,d-1)}, \theta^{(d,d)})$$

$$\theta^{(i,j)} = (\alpha^{(i,j)}, \beta^{(i,j)}, D^{(i,j)}, p^{(i,j)}) \in (\mathbb{R}^*_+)^{K^{(i,j)}} \times (\mathbb{R}^*_+)^{K^{(i,j)}} \times (\mathbb{R}^*_+)^{K^{(i,j)}} \times (\mathbb{R}^*_+)^{K^{(i,j)}}.$$
(24)

For simplicity of exposition, we assume that each term in the sum of Equation (23) is generalized gamma kernel. However, all the theory of this paper also holds when some of parameters  $\theta^{(i,j)}$  are fixed to a value or equal to each other. In particular, the kernel can be exponential, gamma or Weibull. Several examples covered by this framework are discussed in Appendix B from the supplementary materialsof [Potiron and Volkov, 2025].

For a vector or a matrix V of dimension k, we denote its  $L^1$  norm as  $|V| = \sum_{i=1}^k |V^{(i)}|$ . For two real numbers a and b, we denote the infimum between them as  $a \wedge b$ . We also define the regularity modulus of order  $p \in \mathbb{N}^*$ , at time  $t \in [0, T]$  and parameter  $P \in \Theta$  as

$$w_p(t, P, s) = \mathbb{E}\Big[\sup_{h \in [0, s \land (T-t)]} |P_{t+h}^* - P_t^*|^p |\mathcal{F}_t, P_t^* = P\Big], \text{ with } s > 0.$$
(25)

We then define the global regularity modulus as

$$w_p(s) = \sup_{(t,P) \in [0,T] \times \Theta} w_p(t,P,s), \text{ with } s > 0.$$
(26)

We denote the big O in probability by  $O_{\mathbb{P}}$ . It is defined through  $X = O_{\mathbb{P}}(\alpha) \iff \frac{X}{\alpha}$  is stochastically bounded. We define  $\overline{\Theta}$  as the closure space of  $\Theta$ . We now introduce a set of conditions required for the CLT of parameter integral.

- Condition 2. (a) We have that  $\Theta$  is such that its closure  $\overline{\Theta}$  is a compact space and which satisfies the assumptions from the Sobolev embedding theorem (see Theorem 4.12 (p. 85) in [Adams and Fournier, 2003]).
- (b) For any P = (ν, θ) ∈ Θ, we have that the kernel parameter θ is of the form (24) and the kernel h(t, θ) is of the form (23).
- (c) There exists a positive real number  $p_- > 0$  such that for any i = 1, ..., d, any j = 1, ..., d and any  $k = 1, ..., K^{(i,j)}$  we have that  $p_k^{(i,j)} > p_-$ .
- (d) There exists a positive real number  $D_- > 0$  such that for any i = 1, ..., d, any j = 1, ..., d and any  $k = 1, ..., K^{(i,j)}$  we have that  $D_k^{(i,j)} > D_-$ .
- (e) There exists a real number  $\gamma \in (0,1]$  such that we have  $w_p(s) = O_{\mathbb{P}}(s^{\gamma p})$  when  $s \to 0$ , .
- (f) We assume that there exists a real positive number  $\delta > 0$  which satisfies  $\frac{\Delta}{T} = n^{1/\delta 1}$ .
- (g)  $\delta$  and  $\gamma$  satisfy the relation  $\delta > 1 + \frac{1}{\gamma}$ .
- (h)  $\delta$  and  $\gamma$  satisfy the relation  $\frac{2\gamma}{2\gamma-1} < \delta < 3$ .

Condition 2 (a) is about the parameter space, and already appears in [Potiron and Volkov, 2025] (Condition [A] (vi)). Condition 2 (b) restricts to Hawkes processes with mixture of generalized gamma kernels and corresponds to Condition [A] (ii) in [Potiron and Volkov, 2025]. Conditions 2 (c) and (d) requires more parameters restrictions and can be compared to Conditions [A] (iii) and (iv) in [Potiron and Volkov, 2025]. Conditions 2 (e), (f) and (g) are required for local estimation and are Conditions [C] (i) and (ii) in [Clinet and Potiron, 2018]. Finally, Condition 2 (h) is used for bias correction and corresponds to Condition [BC] in [Clinet and Potiron, 2018].

We denote  $\xrightarrow{\mathcal{D}-s}$  as the  $\mathcal{F}_t$ -stable convergence.  $\xi$  is defined as an m-dimensional standard normal vector. We now state the CLT for MLE of parameter integral in the following theorem. It also provides feasible statistics induced by the CLT. The results are obtained with in-fill asymptotics. This extends Theorem 5.4 (p. 3480) in [Clinet and Potiron, 2018] to the case of non exponential kernels.

**Theorem 1.** We assume that Conditions 1 and 2 hold. There is an extension of  $\mathcal{B}$  on which is defined a standard Brownian motion W, which is of dimension m, such that we have the CLT and the feasible CLT

$$\sqrt{n}(\widehat{IP}(T) - IP(T)) \xrightarrow{\mathcal{D}-s} \int_0^T c_{IP}(t) dW_t,$$
(27)

$$\sqrt{n}\widehat{C}_{IP}^{-1/2}(T)(\widehat{IP}(T) - IP(T)) \xrightarrow{\mathcal{D}-s} \xi.$$
(28)

We introduce the  $d^2 \times (m-d)$  dimensional matrix

$$M_t^{((i-1)d+j,q)} = \sum_{r=1}^{m-d} dF^{(i,j,r)}(\theta_t^*) \big( \Gamma_h(\theta_t^*)^{-1/2} \big)^{(r,q)},$$

for any i = 1, ..., d, j = 1, ..., d and q = 1, ..., m - d. We now introduce a set of conditions required for the CLT of latency integral.

Condition 3. (a) The latency function  $F: \Theta_h \to \mathbb{R}^{d \times d}_+$  is continuously differentiable twice.

(b) The matrix  $M_t$  has full rank a.s., i.e.  $\mathbb{P}(M_t$  has full rank  $\forall t \in [0, T]) = 1$ .

Conditions 3 (a) and (b) are natural and already appear in [Potiron and Volkov, 2025] (Conditions [B] and [C]).

 $\overline{\xi}$  is defined as an m-dimensional standard normal vector. In what follows, we give the CLT for estimation of latency integral. It also provides feasible statistics induced by the CLT. The results are obtained with in-fill asymptotics. This extends Corollary 3 in [Potiron and Volkov, 2025] to the case of random time-dependent latency. **Theorem 2.** We assume that Conditions 1, 2 and 3 hold. There is an extension of  $\mathcal{B}$  on which is defined a standard Brownian motion W, which is of dimension  $d^2$ , such that we have the CLT and the feasible CLT

$$\sqrt{n}(\widehat{\overline{IL}}(T) - \overline{IL}(T)) \xrightarrow{\mathcal{D}-s} \int_0^T c_{\overline{IL}}(t) dW_t,$$
(29)

$$\sqrt{n}\widehat{C}_{\overline{IL}}^{-1/2}(T)(\widehat{\overline{IL}}(T) - \overline{IL}(T)) \xrightarrow{\mathcal{D}-s} \overline{\xi}.$$
(30)

We make a final condition required for the test of latency constancy under the alternative.

Condition 4. (a) The linear latency  $A\overline{L}_t$  is continuously differentiable on [0,T] a.s. and there exists  $L_-$  such that the integral of the squared derivatives is bigger than  $L_-$  a.s., namely  $\mathbb{P}\left(\int_0^T \left(\frac{d}{dt}(A\overline{L}_t)^{(i)}\right)^2 dt \ge L_-\right) = 1$  for any  $i = 1, \cdots, a$ .

This condition is novel to the literature since this is the first test on constancy of latency. We define Q(u) as the quantile function of the chi-squared distribution with q degrees of freedom. Finally, the following corollary shows that the Wald test statistic converges in distribution to a chi-squared distribution with q degrees of freedom under the null hypothesis and is consistent under the alternative hypothesis. This is an application of Theorem 2. This extends the test for baseline constancy in Theorem 4.1 (p. 79) from [Kwan et al., 2023] to the case of latency.

**Corollary 1.** We assume that Conditions 1, 2 and 3 hold. Then, the test statistic S converges in distribution to a chi-squared random variable with q degrees of freedom under the null hypothesis  $H_0$ . If we also assume Condition 4, the test statistic S is consistent under the alternative hypothesis  $H_1$ , i.e. we have  $\mathbb{P}(S > Q(u) \mid H_1) \rightarrow 1$  for any 0 < u < 1.

## 5 Numerical study

will require pre-simulation of the bias for a grid of parameter values since the formula is too hard to implement

Please follow the simulation study of [Potiron et al., 2025] for the tables and figures to report. I think we can copy/paste the style of them directly from there. That numerical study is more professional than the one from [Potiron and Volkov, 2025]. However, there will be a lot in common with [Potiron and Volkov, 2025], so I advice you base your work on it.

## 6 Empirical application

#### 6.1 News data

News data were collected from the Thomson Reuters News Analytics database. A news stream is split up into news items. Each news item is an atomic piece of news and may physically represent, for example, a single line alert, a full news article, or an updated news article. The exact nature of a news item is determined by the feed handler.

The data represent all news headlines for the stocks included in the S&P 500 (the United States) from the 5th of January 2015 to 31st of December 2024. Each news message provides a relevance, sentiment, sentiment position and an item type. Relevance is represented by a number in the [0,1] interval, sentiment takes values 1, 0, and -1 for a positive, neutral and negative tone of the story, respectively. The sentiment positions are three values that can be interpreted as probabilities of the positive, negative, and neutral tones the sum of which is equal to one. The item type allows for the identification of alerts, articles, updates, or corrections.

To do:

1) implementation of the test for constancy of latency - intuitively will be always rejected

2) behavior of latency intraday, is there U-shape?

3) Show that there is some residual stochastic component, so that this corroborates our stochastic model of latency

## 7 Conclusion

In this paper, we have studied estimation of latency, when it depends on time. We have considered locally parametric Hawkes processes, where the baseline and the parameters of the kernels are timedependent. We have proposed local estimation based on MLE. We have derived CLT for MLE of parameter integral and latency integral. We have proposed a test statistic for constancy of latency. A numerical simulation have corroborated the asymptotic theory. An empirical application to news data showed that the test for constancy of latency were always rejected and ???.

The code is available online at ???

## Supplementary materials

All proofs of the theory can be found in the supplementary materials. These proofs are based on [Brémaud and Massoulié, 1996], [Jacod, 1997], [Jacod and Shiryaev, 2003], [Jacod and Protter, 2012], [Yoshida, 2011], [Clinet and Yoshida, 2017], [Clinet and Potiron, 2018], [Potiron and Mykland, 2020] [Kwan et al., 2023], [Potiron and Volkov, 2025], [Potiron et al., 2025],

## Funding

Financial support from the Japanese Society for the Promotion of Science under grant 23H00807 (Potiron) is gratefully acknowledged.

## References

- [Adams and Fournier, 2003] Adams, R. A. and Fournier, J. J. (2003). Sobolev spaces, volume 140. Academic press.
- [Aït-Sahalia et al., 2015] Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117:585–606.
- [Aït-Sahalia and Jacod, 2014] Aït-Sahalia, Y. and Jacod, J. (2014). High-frequency financial econometrics. Princeton University Press.

- [Aït-Sahalia et al., 2014] Aït-Sahalia, Y., Laeven, R., and Pelizzon, L. (2014). Mutual excitation in eurozone sovereign CDS. *Journal of Econometrics*, 183:151–167.
- [Bacry et al., 2013] Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499.
- [Biais et al., 2015] Biais, B., Foucault, T., and Moinas, S. (2015). Equilibrium fast trading. Journal of Financial Economics, 116(2):292–313.
- [Bowsher, 2007] Bowsher, C. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141:876–912.
- [Brémaud and Massoulié, 1996] Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear hawkes processes. Annals of Probability, pages 1563–1588.
- [Budish et al., 2015] Budish, E., Cramton, P., and Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621.
- [Cavaliere et al., 2023] Cavaliere, G., Lu, Y., Rahbek, A., and Stærk-Østergaard, J. (2023). Bootstrap inference for Hawkes and general point processes. *Journal of Econometrics*, 235:133–165.
- [Chavez-Demoulin et al., 2005] Chavez-Demoulin, V., Davison, A., and McNeil, A. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance*, 5:227–234.
- [Chen and Hall, 2013] Chen, F. and Hall, P. (2013). Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. *Journal of Applied Probability*, 50(4):1006–1024.
- [Cheysson and Lang, 2022] Cheysson, F. and Lang, G. (2022). Spectral estimation of Hawkes processes from count data. *Annals of Statistics*, 50(3):1722–1746.

- [Clinet and Potiron, 2018] Clinet, S. and Potiron, Y. (2018). Statistical inference for the doubly stochastic self-exciting process. *Bernoulli*, 24(4B):3469–3493.
- [Clinet and Yoshida, 2017] Clinet, S. and Yoshida, N. (2017). Statistical inference for ergodic point processes and application to limit order book. *Stochastic Processes and their Applications*, 127:1800– 1839.
- [Corradi et al., 2020] Corradi, V., Distaso, W., and Fernandes, M. (2020). Testing for jump spillovers without testing for jumps. *Journal of the American Statistical Association*, 115:1214–1226.
- [Dahlhaus, 1996] Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. Stochastic Processes and their Applications, 62(1):139–168.
- [Daley and Vere-Jones, 2003] Daley, D. J. and Vere-Jones, D. (2003). An introduction to the theory of point processes: Elementary theory and methods, volume 1. Springer New York, NY, 2nd edition.
- [Daley and Vere-Jones, 2008] Daley, D. J. and Vere-Jones, D. (2008). An introduction to the theory of point processes: General theory and structure, volume 2. Springer New York, NY, 2nd edition.
- [Dong et al., 2019] Dong, J., Yom-Tov, E., and Yom-Tov, G. B. (2019). The impact of delay announcements on hospital network coordination and waiting times. *Management Science*, 65(5):1969–1994.
- [Embrechts et al., 2011] Embrechts, P., Liniger, T., and Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378.
- [Fan, 1993] Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. Annals of Statistics, pages 196–216.
- [Foucault et al., 2016] Foucault, T., Hombert, J., and Roşu, I. (2016). News trading and speed. The Journal of Finance, 71(1):335–382.
- [Fulop et al., 2015] Fulop, A., Li, J., and Yu, J. (2015). Self-exciting jumps, learning, and asset pricing implications. The Review of Financial Studies, 28(3):876–912.

- [Gagnon and Karolyi, 2010] Gagnon, L. and Karolyi, G. (2010). Multi-market trading and arbitrage. Journal of Financial Economics, 97:53–80.
- [Gámiz et al., 2022] Gámiz, M., Mammen, E., Martínez-Miranda, M., and Nielsen, J. (2022). Missing link survival analysis with applications to available pandemic data. *Computational Statistics & Data Analysis*, 169:107405.
- [Gámiz et al., 2023] Gámiz, M., Mammen, E., Martínez-Miranda, M., and Nielsen, J. (2023). Monitoring a developing pandemic with available data. arXiv preprint arXiv:2308.09919.
- [Harris, 1990] Harris, J. (1990). Reporting delays and the incidence of aids. Journal of the American Statistical Association, 85:915–924.
- [Hasbrouck and Saar, 2013] Hasbrouck, J. and Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, 16:646–679.
- [Hawkes, 2018] Hawkes, A. (2018). Hawkes processes and their applications to finance: a review. Quantitative Finance, 18(2):193–198.
- [Hawkes, 1971a] Hawkes, A. G. (1971a). Point spectra of some mutually exciting point processes. Journal of the Royal Statistical Society. Series B (Methodological), pages 438–443.
- [Hawkes, 1971b] Hawkes, A. G. (1971b). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.
- [Hoffmann, 2014] Hoffmann, P. (2014). A dynamic limit order market with fast and slow traders. Journal of Financial Economics, 113:156–169.
- [Ikefuji et al., 2022] Ikefuji, M., Laeven, R., Magnus, J., and Yue, Y. (2022). Earthquake risk embedded in property prices: Evidence from five japanese cities. *Journal of the American Statistical* Association, 117:82–93.

- [Jacod, 1975] Jacod, J. (1975). Multivariate point processes: predictable projection, radon-nikodym derivatives, representation of martingales. Zeitschrift f
  ür Wahrscheinlichkeitstheorie und verwandte Gebiete, 31(3):235–253.
- [Jacod, 1997] Jacod, J. (1997). On continuous conditional gaussian martingales and stable convergence in law. Séminaire de Probabilités de Strasbourg, 31:232–246.
- [Jacod and Protter, 2012] Jacod, J. and Protter, P. E. (2012). Discretization of processes. Springer Berlin, Heidelberg.
- [Jacod and Shiryaev, 2003] Jacod, J. and Shiryaev, A. (2003). Limit theorems for stochastic processes. Springer Berlin, Heidelberg, 2nd edition.
- [Jaisson and Rosenbaum, 2015] Jaisson, T. and Rosenbaum, M. (2015). Limit theorems for nearly unstable hawkes processes. *Annals of Applied Probability*, 25(2):600–631.
- [Kwan, 2023] Kwan, T. (2023). Asymptotic analysis and ergodicity of the Hawkes process and its extensions. PhD thesis, UNSW Sydney.
- [Kwan et al., 2023] Kwan, T., Chen, F., and Dunsmuir, W. (2023). Alternative asymptotic inference theory for a nonstationary Hawkes process. *Journal of Statistical Planning and Inference*, 227:75–90.
- [Lahad, 2012] Lahad, K. (2012). Singlehood, waiting, and the sociology of time 1. In Sociological Forum, volume 27, pages 163–186. Wiley Online Library.
- [Large, 2007] Large, J. (2007). Measuring the resiliency of an electronic limit order book. Journal of Financial Markets, 10:1–25.
- [Lesage et al., 2022] Lesage, L., Deaconu, M., Lejay, A., Meira, J., Nichil, G., and State, R. (2022). Hawkes processes framework with a gamma density as excitation function: application to natural disasters for insurance. *Methodology and Computing in Applied Probability*, 24:2509–2537.

[Liniger, 2009] Liniger, T. (2009). Multivariate Hawkes processes. PhD thesis, ETH Zurich.

- [Mammen and Müller, 2023] Mammen, E. and Müller, M. (2023). Nonparametric estimation of locally stationary Hawkes processes. *Bernoulli*, 29(3):2062–2083.
- [Nagin and Pogarsky, 2004] Nagin, D. S. and Pogarsky, G. (2004). Time and punishment: Delayed consequences and criminal behavior. *Journal of Quantitative Criminology*, 20:295–317.
- [Nolet and Dahlen, 2000] Nolet, G. and Dahlen, F. (2000). Wave front healing and the evolution of seismic delay times. *Journal of Geophysical Research: Solid Earth*, 105(B8):19043–19054.
- [Ogata, 1978] Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. Annals of the Institute of Statistical Mathematics, 30(1):243–261.
- [Omi et al., 2017] Omi, T., Hirata, Y., and Aihara, K. (2017). Hawkes process model with a timedependent background rate and its application to high-frequency financial data. *Physical Review E*, 96(1):012303.
- [Ozaki, 1979] Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. Annals of the Institute of Statistical Mathematics, 31(1):145–155.
- [Pagnotta and Philippon, 2018] Pagnotta, E. S. and Philippon, T. (2018). Competing on speed. Econometrica, 86(3):1067–1115.
- [Potiron and Mykland, 2020] Potiron, Y. and Mykland, P. A. (2020). Local parametric estimation in high frequency data. Journal of Business & Economic Statistics, 38(3):679–692.
- [Potiron et al., 2025] Potiron, Y., Scaillet, O., Volkov, V., and Yu, S. (2025). High-frequency estimation of Itô semimartingale baseline for Hawkes processes. Working paper available at https://www.fbc.keio.ac.jp/~ potiron/Potiron2025workingpaper.pdf.
- [Potiron and Volkov, 2025] Potiron, Y. and Volkov, V. (2025+). Mutually exciting point processes with latency. To appear in Journal of the American Statistical Association.

- [Roueff and Von Sachs, 2019] Roueff, F. and Von Sachs, R. (2019). Time-frequency analysis of locally stationary Hawkes processes. *Bernoulli*, 25(2):1355–1385.
- [Roueff et al., 2016] Roueff, F., von Sachs, R., and Sansonnet, L. (2016). Locally stationary Hawkes processes. Stochastic Processes and their Applications, 126(6):1710–1743.
- [Rubin, 1972] Rubin, I. (1972). Regular point processes and their detection. IEEE Transactions on Information Theory, 18:547–557.
- [Vere-Jones, 1978] Vere-Jones, D. (1978). Earthquake prediction-a statistician's view. Journal of Physics of the Earth, 26:129–146.
- [Vere-Jones and Ozaki, 1982] Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical estimation applied to earthquake data: I. cyclic poisson and self-exciting models. Annals of the Institute of Statistical Mathematics, 34:189–207.
- [Yoshida, 2011] Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. Annals of the Institute of Statistical Mathematics, 63(3):431–479.
- [Yu, 2004] Yu, J. (2004). Empirical characteristic function estimation and its applications. *Econometric Reviews*, 23(2):93–123.

## Supplementary materials

This part corresponds to the supplementary materials of "Estimation of time-dependent latency with locally stationary Hawkes processes" by Deniz Erdemlioglu, Yoann Potiron and Vladimir Volkov submitted to the Journal of American Statistical Association. All the proofs of the theory can be found in Section 8.

## 8 Proofs

We use C for any constant, and the value of the constant can change from one line to the next. Any operation with two vectors of the same size means the operation component by component. We begin with the proof of the existence of locally parametric Hawkes processes where the baseline and the parameters of the kernels are time-dependent. It extends the proof of Theorem 7 (pp. 1585-1587) in [Brémaud and Massoulié, 1996], the proof of Theorem 5.1 (pp. 3-4) in the supplementary materials of [Clinet and Potiron, 2018], and the proof of Proposition 4.1 in [Potiron et al., 2025], to the general kernel with time-varying parameters case.

Proof of Proposition 1. The strategy of the proof consists in defining a suitable sequence of simple point processes and intensity  $(N_t^k, \lambda_t^k)_{k\geq 0}$  such that their limit defined as  $(N_t, \lambda_t) = \lim_{k\to\infty} (N_t^k, \lambda_t^k)$ exists and  $N_t$  admits  $\lambda_t$  as  $\mathcal{F}_t$ -intensity given by Equation (4).

We first define, for any  $t \in [0, T]$  and any  $i = \cdots, d$ ,  $\lambda^{0,(i)}(t) = \nu_t^{*,(i)}$  and  $N_t^{0,(i)}$  the simple point process counting the points of  $\underline{N}^{(2i-1)} * \underline{N}^{(2i)}$  below the curve  $t \to \lambda_t^{0,(i)}$  as

$$N_t^{0,(i)} = \int_{[0,t]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_s^{0,(i)}]}(x)\underline{N}^{(2i-1)} * \underline{N}^{(2i)}(ds \times dx).$$

We then define recursively the sequence of  $(N^{k,(i)}_t,\lambda^{k,(i)}_t)_{k\geq 1}$  as

$$\lambda_t^{k+1} = \nu_t^* + \int_0^t h(t-s,\theta_s^*) dN_s^k, \tag{31}$$
$$N_t^{k+1,(i)} = \int_{[0,t]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_s^{k+1,(i)}]}(x) \underline{N}^{(2i-1)} * \underline{N}^{(2i)}(ds \times dx) \text{ for any } i = \cdots, d.$$

First, we have that  $\lambda^{k,(i)}$  is positive on [0, T] a.s. as an application of Conditions 1 (b) and (e). Thus,  $\lambda^{k,(i)}$  is a well-defined intensity. Then, an extension to the time-dependent case of the arguments from Lemma 3 and Example 4 (pp. 1571-1572) in [Brémaud and Massoulié, 1996] yields that  $N_t^k$  is  $\mathcal{F}_t$ -adapted,  $\lambda_t^k$  is  $\mathcal{F}_t$ -predictable and an  $\mathcal{F}_t$ -intensity of  $N_t^k$ . Moreover, Condition 1 (e) implies that  $(N_t^{k,(i)}, \lambda_t^{k,(i)})$  is componentwise increasing with k and thus converges to some limit  $(N_t^{(i)}, \lambda_t^{(i)})$  a.s. for any  $t \in [0, T]$ .

We now introduce the sequence of vector processes  $\rho_t^k$  defined as  $\rho_t^k = \mathbb{E}[\lambda_t^k - \lambda_t^{k-1} | \mathcal{F}_T^{P^*}]$ . Then

$$\rho_t^{k+1} = \mathbb{E}\Big[\int_0^t h(t-s,\theta_s^*)(\lambda_s^k - \lambda_s^{k-1})ds\Big|\mathcal{F}_T^{P^*}\Big] = \int_0^t h(t-s,\theta_s^*)\rho_s^k ds$$

Here, the first equality is obtained by Lemma 10.1 (p. 2) from the supplementary materials of [Clinet and Potiron, 2018] when  $\mathcal{G} = \mathcal{F}_T^{P^*}$ , with Condition 1 (d) and Equation (32). The second equality is obtained by Tonelli's theorem and the definition of  $\rho_t^k$ . If we define  $\Phi_t^k$  as  $\Phi_t^k = \int_0^t \rho_s^k ds$ , we have by another application of Tonelli's theorem that a.s.

$$\Phi_t^{k+1} = \int_0^t \Big( \int_0^{t-s} h(u, \theta_s^*) du \Big) \rho_s^k ds.$$
(32)

Then, Condition 1 (f) implies that  $|\Phi_t^{k+1}| \leq r |\Phi_t^k|$  a.s.. Thus, we can deduce that  $G : \Phi_t^k \to \Phi_t^{k+1}$  is a.s. a contraction function. It turns out that the limit of the telescopic series  $(\sum_{l=0}^k \Phi_t^l)_{k\geq 1}$  exists by arguments used in Banach fixed-point theorem. Working with the telescopic series and applying the monotone convergence theorem to the series yields

$$\mathbb{E}\Big[\int_0^t \lambda_s ds \Big| \mathcal{F}_T^{P^*}\Big] \le \int_0^t \nu_s^* ds + r \mathbb{E}\Big[\int_0^t \lambda_s ds \Big| \mathcal{F}_T^{P^*}\Big].$$
(33)

By rearranging the terms in Expression (33), we get that

$$\mathbb{E}\left[\int_0^t \lambda_s ds \left| \mathcal{F}_T^{P^*} \right] \le (1-r)^{-1} \int_0^t \nu_s^* ds.$$
(34)

Given Condition 1 (c), the expression in the left side of Expression (34) is finite a.s.. Given that its conditional expectation is finite,  $\int_0^t \lambda_s ds$  is finite a.s.. Moreover,  $\lambda_t$  is  $\mathcal{F}_t$ -predictable as a limit of such processes.  $N_t^{(i)}$  counts the points of  $\underline{N}^{(2i-1)} * \underline{N}^{(2i)}$  under the curve  $t \mapsto \lambda_t^{(i)}$  by an application of

the monotone convergence theorem.  $N_t$  therefore admits  $\lambda_t$  as an  $\mathcal{F}_t$ -intensity by an extension to the time-dependent case of the arguments from Lemma 3 (p. 1571) in [Brémaud and Massoulié, 1996]. It implies that  $N_t$  is finite a.s.. Finally, it remains to show that  $\lambda_t$  is of the form (4). The monotonicity properties of  $N_t^{k,(i)}$  and  $\lambda_t^{k,(i)}$  ensure that, for any  $k \ge 0$ , any  $t \in [0,T]$  and any  $i = \cdots, d, \lambda_t^{k,(i)} \le \nu_t^{*,(i)} + (\int_0^t h(t-s,\theta_s^*) dN_s)^{(i)}$  and  $\lambda_t^{(i)} \ge \nu_t^{*,(i)} + (\int_0^t h(t-s,\theta_s^*) dN_s)^{(i)}$ , which gives Equation (4) by taking the limit  $k \to +\infty$  in both inequalities.

The following lemma states that K, C and Q converge in probability to limit values  $K_l$ ,  $C_l$  and  $Q_l$ . This extends Section 10.1 (pp. 1-2) in the supplementary materials of [Clinet and Potiron, 2018].

**Lemma 1.** For any  $P^* \in \Theta$  and any indices  $k, l, q \in \{1, \dots, m\}$   $K(P^*)$ ,  $C(P^*)_{k,lq}$  and  $Q(P^*)_{k,lq}$ converge in probability to limit values  $K_l(P^*)$ ,  $C_l(P^*)_{k,lq}$  and  $Q_l(P^*)_{k,lq}$ .

Proof of Lemma 1. We can prove the lemma by extending the arguments from the proof of Lemma A.6 (p. 1834) in [Clinet and Yoshida, 2017].  $\hfill \Box$ 

Our proof strategy follows the general machinery of [Clinet and Yoshida, 2017], which consider large-T asymptotics. To rewrite our problem with in-fill asymptotics as a problem with large-T asymptotics, we consider a time transformation as in [Clinet and Potiron, 2018], [Kwan et al., 2023] and [Potiron and Volkov, 2025]. More specifically, we define the time-transformed filtration as  $(\overline{\mathcal{F}}_t)_{t\in[0,nT]}$ , where  $\overline{\mathcal{F}}_t = \mathcal{F}_{\frac{t}{n}}$ . For any  $i = 1, \ldots, d$  the *i*th process of the time-transformed point process  $\overline{N}_t^{(i)}$  has events at times  $(\overline{\tau}_1^{(i)}, \ldots, \overline{\tau}_{N^{(i)}}^{(i)})$ , defined as  $\overline{\tau}_k^{(i)} = n\tau_k^{(i)}$  for  $k = 1, \cdots, N^{(i)}$ . We also define the rescaled time-transformed stochastic  $\overline{\mathcal{F}}_t$ -intensity process as  $\overline{\lambda}_t = \frac{\lambda_{t/n}}{n}$  for  $t \in [0, nT]$ . Finally, we define the time-transformed parameter as  $\overline{\mathcal{F}}_t^* = P_{t/n}$  for  $t \in [0, nT]$ , where  $\overline{\mathcal{F}}_t^* = (\overline{\nu}_t^*, \overline{\theta}_t^*)$ . In this first lemma, we rewrite the rescaled time-transformed intensity in terms of the time-transformed point process. This extends Lemma C1 in the supplementary materials of [Potiron and Volkov, 2025].

Lemma 2. Under Condition 1, we have that

$$\overline{\lambda}_t = \overline{\nu}_t^* + \int_0^t h(t - s, \overline{\theta}_s^*) d\overline{N}_s.$$
(35)

Proof of Lemma 2. If we substitute the definitions of  $\overline{\lambda}_t$ ,  $\overline{P}_t^*$  and  $\overline{N}_t$  into Definition (4), we obtain the lemma.

We define the compensated rescaled time-changed point process as

$$\overline{M}_t = \overline{N}_t - \int_0^t \overline{\lambda}_s ds.$$
(36)

The next lemma shows that  $\overline{M}_t$  is an  $\overline{\mathcal{F}}_t$ -local martingale and that  $\overline{N}_t$  is a locally parametric Hawkes process with the same kernel, and parameters  $\overline{P}_t^*$ . This extends Lemma C2 in [Potiron and Volkov, 2025].

**Lemma 3.** Under Condition 1, we have that  $\overline{M}_t$  is an  $\overline{\mathcal{F}}_t$ -local martingale and that  $\overline{N}_t$  is a locally parametric Hawkes process with the same kernel h, parameters  $\overline{P}_t^*$ , and  $\overline{\mathcal{F}}_t$ -intensity  $\overline{\lambda}_t$ .

Proof of Lemma 3. By definition of a compensator, we have that

$$M_t = N_t - \int_0^t \lambda_s ds \tag{37}$$

is a  $\mathcal{F}_t$ -local martingale. First, we will show that  $\overline{M}_t$  is an  $\overline{\mathcal{F}}_t$ -local martingale. We have

$$\begin{split} \overline{M}_t &= \overline{N}_t - \int_0^t \overline{\lambda}_s ds \\ &= N_{t/n} - \int_0^t \frac{\lambda_{s/n}}{n} ds \\ &= N_{t/n} - \int_0^{\frac{t}{n}} \lambda_y dy \\ &= M_{t/n}. \end{split}$$

Here, we use Equation (36) in the first equality, the definitions of  $\overline{N}_t$  and  $\overline{\lambda}_t$  in the second equality, integral change of variable in the third equality and Equation (37) in the fourth equality. As  $M_t$  is an  $\mathcal{F}_t$ -local martingale, we also have that the time-transformed local martingale  $M_{t/n}$  is an  $\overline{\mathcal{F}}_t$ -local martingale. Since  $\overline{M}_t = M_{t/n}$ , it means that  $\overline{M}_t$  is an  $\overline{\mathcal{F}}_t$ -local martingale. Then, we can deduce that  $\overline{N}_t$  is a locally parametric Hawkes process with the same kernel h, parameters  $\overline{\mathcal{P}}_t^*$ , and  $\overline{\mathcal{F}}_t$ -intensity  $\overline{\lambda}_t$ by Theorem 3.17 (p. 32) in [Jacod and Shiryaev, 2003]. For any  $P \in \Theta$ , we define  $\overline{\lambda}_t(P)$  as  $\overline{\lambda}_t(P) = \frac{\lambda_{t/n}(P)}{n}$  for  $t \in [0, nT]$ . For  $l = 0, \dots, M$ , we define the time-transformed times as  $\overline{T}_l = nT_l$ . We also define the log likelihood process of the rescaled time-transformed point process  $\overline{N}$  on the *l*-th interval  $[\overline{T}_{l-1}, \overline{T}_l)$  as

$$\bar{l}_l(P) = \sum_{i=1}^d \int_{\overline{T}_{l-1}}^{\overline{T}_l} \log(\overline{\lambda}_t^{(i)}(P)) d\overline{N}_t^{(i)} - \sum_{i=1}^d \int_{\overline{T}_{l-1}}^{\overline{T}_l} \overline{\lambda}_t^{(i)}(P) dt.$$
(38)

Then, the local MLE is defined as a maximizer of the local log likelihood process, i.e.

$$\widehat{\overline{P}}_l = (\widehat{\overline{\nu}}_l, \widehat{\overline{\theta}}_l) \in \operatorname{argmax}_{P \in \Theta} \overline{l}_l(P).$$

The following lemma states that a.s. the MLE of the point process on the *l*-th interval  $[T_{l-1}, T_l)$  is equal to the MLE of the rescaled time-changed point process on the *l*-th interval  $[\overline{T}_{l-1}, \overline{T}_l)$ . This extends Lemma C3 in [Potiron and Volkov, 2025].

Lemma 4. Under Condition 1, we have that

$$\mathbb{P}(\widehat{P}_l = \widehat{\overline{P}}_l \text{ for } l = 1, \cdots, M) = 1$$

*Proof of Lemma 4.* By Definition (38), the definition of  $\widehat{\overline{P}}_l$  and Lemma 3, the lemma follows.

For  $i = 1, \dots, d$ , we define the rescaled time-changed Poisson processes  $\overline{\underline{N}}^{(2i-1)} * \overline{\underline{N}}^{(2i)}$  as  $\overline{\underline{N}}^{(2i-1)} * \overline{\underline{N}}^{(2i)}(ds \times dx) = \underline{\underline{N}}^{(2i-1)} * \underline{\underline{N}}^{(2i)}(\frac{ds}{n} \times ndx)$ . For  $i = 1, \dots, d$ , we denote by  $\overline{\underline{\Lambda}}^{(i)}$  the compensating measure of  $\overline{\underline{N}}^{(2i-1)} * \overline{\underline{N}}^{(2i)}$ , i.e.  $\overline{\underline{\Lambda}}^{(i)}(ds, dz) = \frac{ds}{n} \times ndz$ . We define a *d*-dimensional predictable function by W. We introduce

$$W * \overline{\underline{N}}_{t}^{(i)} = \iint_{[0,t] \times \mathbb{R}} W^{(i)}(s,z) \overline{\underline{N}}^{(2i-1)} * \overline{\underline{N}}^{(2i)}(ds,dz)$$

and  $W * \overline{\Delta}_t^{(i)} = \iint_{[0,t] \times \mathbb{R}} W^{(i)}(s,z) \overline{\Delta}^{(i)}(ds,dz)$ . The following lemma corresponds to Burkholder-Davis-Gundy inequality. This is an application of Lemma 2.1.5 (p. 41) in [Jacod and Protter, 2012].

**Lemma 5.** We assume that Condition 1 holds and that  $\mathbb{P}(W^2 * \overline{\Lambda}_t^{(i)} < \infty \text{ for } i = 1, \cdots, d) = 1$ . For any integer p > 1 and any  $i = 1, \cdots, d$ , there exists a constant C such that

$$\mathbb{E}\Big[\sup_{t\in[0,nT]} |W*(\overline{\underline{N}}-\overline{\underline{\Lambda}})_t^{(i)}|^p |\overline{\mathcal{F}}_{nT}^{\overline{P}^*}\Big] \leq C \mathbb{E}\Big[\iint_{[0,nT]\times\mathbb{R}} |W^{(i)}(s,z)|^p dsdz \\ + \Big(\iint_{[0,nT]\times\mathbb{R}} W^{(i)}(s,z)^2 dsdz\Big)^{\frac{p}{2}} \Big|\overline{\mathcal{F}}_{nT}^{\overline{P}^*}\Big].$$

Proof of Lemma 5. We get the proof of the lemma by an application of Lemma 2.1.5 (p. 41) in [Jacod and Protter, 2012].  $\hfill \Box$ 

For any random kernel  $\chi : (s,t) \to \chi(s,t), \chi$  is  $\mathcal{G}_t$ -predictable for some filtration  $\mathcal{G}_t$  if for any  $t \in [0,T]$  the process  $\chi(.,t)$  is  $\mathcal{G}_t$ -predictable. We introduce the following lemma, which gives the boundedness of moments for the rescaled time-changed point process  $\overline{N}$ . This extends Lemma 10.3 (p. 4) in the supplementary materials of [Clinet and Potiron, 2018].

**Lemma 6.** We assume that Condition 1, Conditions 2 (a) and (b) hold. The intensity of the rescaled time-changed point process  $\overline{N}^{(i)}$  has moments on [0, nT] for  $i = 1, \dots, d$  that can be bounded by values independent from T, i.e.

$$\sup_{t \in [0,nT]} \mathbb{E}\left[ (\overline{\lambda}_t^{(i)})^p \big| \overline{\mathcal{F}}_{nT}^{\overline{P}^*} \right] \le C.$$
(39)

For any  $\overline{\mathcal{F}}_t^{\overline{P}^*}$ -predictable kernel  $\chi$  such that  $\int_0^t \chi(s,t) ds$  is bounded uniformly in  $t \in [0, nT]$  and  $n \in \mathbb{N}$ independently from T, we have

$$\sup_{t \in [0,nT], n \in \mathbb{N}} \mathbb{E}\left[\left(\int_0^t \chi(s,t) d\overline{N}_s^{(i)}\right)^p \middle| \overline{\mathcal{F}}_{nT}^{\overline{P}^*} \right] \le C.$$
(40)

*Proof of Lemma 6.* We first prove that Expression (39) holds for p = 1. We have

$$\mathbb{E}\left[\overline{\lambda}_{t} \middle| \overline{\mathcal{F}}_{nT}^{\overline{P}^{*}}\right] = \overline{\nu}_{t}^{*} + \int_{0}^{t} h(t - s, \overline{\theta}_{s}^{*}) \mathbb{E}\left[\overline{\lambda}_{s} \middle| \overline{\mathcal{F}}_{nT}^{\overline{P}^{*}}\right] ds$$
$$\leq C + \sup_{s \in [0, t]} \mathbb{E}\left[\overline{\lambda}_{s} \middle| \overline{\mathcal{F}}_{nT}^{\overline{P}^{*}}\right] \int_{0}^{t} h(t - s, \overline{\theta}_{s}^{*}) ds$$
$$\leq C + r \sup_{s \in [0, t]} \mathbb{E}\left[\overline{\lambda}_{s} \middle| \overline{\mathcal{F}}_{nT}^{\overline{P}^{*}}\right].$$

Here, we use Lemma 2 in the equality, Condition 2 (a) in the first inequality and Condition 1 (f) in the second inequality. Taking the supremum over [0, T] and  $n \in \mathbb{N}$  on both sides, we get

$$\sup_{t \in [0,T] , n \in \mathbb{N}} \mathbb{E}\left[\overline{\lambda}_s \middle| \overline{\mathcal{F}}_{nT}^{\overline{P}^*} \right] \le (1-r)^{-1}C.$$
(41)

We prove now that Expression (39) holds for any integer p > 1. It is sufficient to consider the case  $p = 2^q$ , where q > 0. We thus prove our result by induction on q. First, we have for any  $\epsilon > 0$  that

$$\mathbb{E}\left[\overline{\lambda}_{t}^{p} \middle| \overline{\mathcal{F}}_{nT}^{\overline{P}^{*}} \right] \leq (1+\epsilon^{-1})^{2^{q}-1} C + (1+\epsilon)^{2^{q}-1} \mathbb{E}\left[\left(\int_{0}^{t} h(t-s,\overline{\theta}_{s}^{*}) d\overline{N}_{s}\right)^{p} \middle| \overline{\mathcal{F}}_{nT}^{\overline{P}^{*}} \right].$$
(42)

Here, we use the inequality  $(x+y)^{2^q} \leq (1+\epsilon)^{2^q-1}x^{2^q} + (1+\epsilon^{-1})^{2^q-1}y^{2^q}$  for any  $x, y, \epsilon > 0$ . For any  $t \in [0, nT]$  and any  $i = 1, \dots, d$ , we define  $W^{(i)}(s, z)$  as  $W^{(i)}(s, z) = \sum_{j=1}^d h^{(i,j)}(t-s, \overline{\theta}_s^*) \mathbf{1}_{\{0 \leq z \leq \lambda^{(i)}(s)\}}$ . We obtain that

$$\mathbb{E}\Big[\Big(\int_{0}^{t}h(t-s,\overline{\theta}_{s}^{*})d\overline{N}_{s}\Big)^{p}\Big|\overline{\mathcal{F}}_{nT}^{\overline{P}^{*}}\Big]^{(i)} = \mathbb{E}\Big[(W*\underline{\overline{N}}_{t}^{(i)})^{p}\Big|\overline{\mathcal{F}}_{nT}^{\overline{P}^{*}}\Big]$$

$$\leq (1+\epsilon^{-1})^{2^{q}-1}\mathbb{E}\Big[(W*(\underline{\overline{N}}-\underline{\overline{\Lambda}})_{t}^{(i)})^{p}\Big|\overline{\mathcal{F}}_{nT}^{\overline{P}^{*}}\Big]$$

$$+ (1+\epsilon)^{2^{q}-1}\mathbb{E}\Big[(W*\underline{\overline{\Lambda}}_{t}^{(i)})^{p}\Big|\overline{\mathcal{F}}_{nT}^{\overline{P}^{*}}\Big].$$

$$(43)$$

We define I as  $I = \mathbb{E}\left[ (W * (\overline{\underline{N}} - \overline{\underline{\Lambda}})_t^{(i)})^p \middle| \overline{\mathcal{F}}_{nT}^{\overline{P}^*} \right]$ . We have

$$I \leq C\mathbb{E} \Big[ \iint_{[0,nT]\times\mathbb{R}} |W^{(i)}(s,z)|^p dsdz + \Big( \iint_{[0,nT]\times\mathbb{R}} W^{(i)}(s,z)^2 dsdz \Big)^{\frac{p}{2}} \Big| \overline{\mathcal{F}}_{nT}^{\overline{P}^*} \Big]$$
$$= C\mathbb{E} \Big[ \iint_{[0,nT]\times\mathbb{R}} \Big| \sum_{j=1}^d h^{(i,j)} (t-s,\overline{\theta}_s^*) \mathbf{1}_{\{0\leq z\leq\lambda^{(i)}(s)\}} \Big|^p dsdz$$
$$+ C \Big( \iint_{[0,nT]\times\mathbb{R}} \sum_{j=1}^d h^{(i,j)} (t-s,\overline{\theta}_s^*) \mathbf{1}_{\{0\leq z\leq\lambda^{(i)}(s)\}}^2 dsdz \Big)^{\frac{p}{2}} \Big| \overline{\mathcal{F}}_{nT}^{\overline{P}^*} \Big]$$
$$\leq C. \tag{44}$$

Here, the first inequality is an application of Lemma 5, the equality is obtained by the definition of  $W^{(i)}(s, z)$ , and the second inequality is obtained by Condition 2 (b) with Holder's inequality. We can also show with similar arguments that

$$\mathbb{E}\left[ (W * \underline{\overline{\Delta}}_{t}^{(i)})^{p} \middle| \overline{\mathcal{F}}_{nT}^{\overline{P}^{*}} \right] \leq C.$$
(45)

From Expressions (42), (43), (44) and (45), we can deduce that

$$\mathbb{E}\left[\overline{\lambda}_{t}^{p} | \overline{\mathcal{F}}_{nT}^{\overline{P}^{*}}\right] \leq C.$$

We can show Expression (40) with the same arguments.

We introduce the sequence of stochastic processes R which represents the difference between the rescaled time-changed intensity and rescaled time-changed intensity when assuming that it starts from the baseline value at each beginning of interval. It is defined through

$$R(t) = \overline{\lambda}_t - \overline{\nu}_t^* - \int_{\overline{T}_l}^t h(t - s, \overline{\theta}_s^*) d\overline{N}_s \text{ for } t \in [\overline{T}_l, \overline{T}_{l+1}).$$
(46)

The next lemma shows that R is exponentially decreasing uniformly on each interval of approximation. This extends Expression (10.23) (p. 7) in the supplementary materials of [Clinet and Potiron, 2018].

**Lemma 7.** We assume that Condition 1, Conditions 2 (a), (b), (c) and (d) hold. We have that R is exponentially decreasing uniformly on each interval of approximation  $[\overline{T}_{l-1}, \overline{T}_l)$  for  $l = 1, \ldots, M$ .

Proof of Lemma 7. By Condition 2 (b), we can deduce that the kernel  $h(t, \theta)$  is exponentially decreasing for any  $\theta \in \Theta_h$ . By Conditions 2 (a), (c) and (d), we can use the same arguments as in the proof of Lemma C4 in the supplementary materials of [Potiron and Volkov, 2025] to show that R is exponentially decreasing uniformly on each interval of approximation  $[\overline{T}_{l-1}, \overline{T}_l)$  for  $l = 1, \ldots, M$ .  $\Box$ 

We introduce the deterministic sequence K that bounds the pre-excitation  $R(\overline{T}_l)$  for  $l = 0, \ldots, M - 1$ . 1. We assume that  $K = O(n^q)$  for some q > 1. We denote  $\mathbb{E}[\mathbf{1}_{\{R(\overline{T}_{l-1}) \leq K\}} | \overline{\mathcal{F}}_{\overline{T}_{l-1}}, \overline{\mathcal{P}}_{\overline{T}_{l-1}}^* = P]$  by  $\mathbb{E}_{P,l}$ for  $l = 1, \ldots, M$ . For a measurable set  $A \in \mathcal{F}$ , we also denote  $\mathbb{E}_{P,l}[\mathbf{1}_A]$  by  $\mathbb{P}_{P,l}[A]$ . Finally, we introduce the notation  $\mathbf{E} = \{(P, l, n, t) \in \Theta \times \mathbb{N}^2 \times \mathbb{R}^+ | l = 1, \ldots, M \text{ and } 0 \leq t \leq \overline{T}\}$ . When  $n \in \mathbb{N}$  is fixed, we define  $\mathbf{E}_n$  the subset of  $\mathbf{E}$  as  $\mathbf{E}_n = \{(P, l, t) \in \Theta \times \mathbb{N} \times \mathbb{R}^+ | l = 1, \ldots, M \text{ and } 0 \leq t \leq \overline{T}\}$ . For  $\alpha \in (0, 1)$ , we denote by  $\mathbf{E}_n^{\alpha}$  the subset of  $\mathbf{E}_n$  for which we have the stronger condition  $(\Delta T)^{-\alpha}T \leq t \leq \Delta^{-1}$ . We define the rescaled time-changed Hawkes processes with constant parameters on each interval as

$$\overline{N}_{t,c}^{(i)} = \iint_{[0,t]\times\mathbb{R}_+} \mathbb{1}_{\{0\leq z\leq\overline{\lambda}_{s,c}\}} \overline{\underline{N}}^{(2i-1)} * \overline{\underline{N}}^{(2i)}(ds, dz) \text{ for } i = 1, \cdots, d,$$
$$\overline{\lambda}_{t,c} = \overline{\nu}_{\overline{T}_l}^* + \int_{\overline{T}_l}^t h(t-s, \overline{\theta}_{\overline{T}_l}^*) d\overline{N}_{s,c} \text{ for } t \in [\overline{T}_l, \overline{T}_{l+1}).$$

The next lemma states the uniform boundedness of the moments of  $\overline{\lambda}$  and  $\overline{\lambda}_c$ , along with stochastic integrals with respect to  $\overline{N}$  and  $\overline{N}_c$ . This extends Lemma 10.5 (p. 8) in the supplementary materials of [Clinet and Potiron, 2018].

**Lemma 8.** We assume that Condition 1, Conditions 2 (a), (b), (c) and (d) hold. We have for any integer  $p \ge 1$ , any  $i = 1, \dots, d$  and any  $\overline{\mathcal{F}}^{\overline{P}^*}$ -predictable kernel  $\chi$  such that  $\int_0^t \chi(s, t) ds$  is bounded uniformly in  $t \in [0, nT]$  independently from T and n a.s.

$$\sup_{(P,l,n,t)\in\mathbf{E}} \mathbb{E}_{P,l} |\overline{\lambda}_{\overline{T}_l}^{(i)}|^p \le C, \tag{47}$$

$$\sup_{(P,l,n,t)\in\mathbf{E}} \mathbb{E}_{P,l} \left| \int_{\overline{T}_{l-1}}^{t\wedge T_l} \chi(s,t) d\overline{N}_s^{(i)} \right|^p \le C,$$
(48)

$$\sup_{(P,l,n,t)\in\mathbf{E}} \mathbb{E}_{P,l} |\overline{\lambda}_{\overline{T}_{l,c}}|^{p} \le C,$$
(49)

$$\sup_{(P,l,n,t)\in\mathbf{E}} \mathbb{E}_{P,l} \left| \int_{\overline{T}_{l-1}}^{t\wedge\overline{T}_l} \chi(s,t) d\overline{N}_{s,c}^{(i)} \right|^p \le C.$$
(50)

Proof of Lemma 8. This is an application of Lemma 6 with Jensen's inequality in the case when the conditional expectation is equal to  $\mathbb{E}_{P,l}$ . The presence of  $\mathbf{1}_{\{R(\overline{T}_{l-1})\leq K\}}$  with Lemma 7 shows the result uniformly in the quadruplet (P, l, n, t).

We define  $\kappa$  as  $\kappa = \gamma(\delta - 1)$ . In the lemma that follows, we quantify the error between the locally stationary Hawkes processes and the Hawkes processes with constant parameters on each interval. This extends Lemma 10.7 (p. 9) in the supplementary materials of [Clinet and Potiron, 2018].

**Lemma 9.** We assume that Condition 1, Conditions 2 (a), (b), (c), (d), (e), (f) and (g) hold. Let  $\alpha \in (0,1)$  be a truncation exponent, and  $\epsilon \in (0,1)$ . We have for any  $p \ge 1$ , any  $i = 1, \dots, d$ , any deterministic kernel  $\chi$  such that  $\int_0^t \chi(s,t) ds$  is bounded uniformly in  $t \in \mathbb{R}_+$ , and any predictable process  $(\psi_s)_{s \in \mathbb{R}_+}$  whose moments are bounded that

$$\sup_{(P,l,t)\in\mathbf{E}_{n}^{\alpha}}\mathbb{E}_{P,l}|\overline{\lambda}_{t,c}^{(i)}-\overline{\lambda}_{t}^{(i)}|^{p}=O_{\mathbb{P}}(\Delta^{\kappa}),$$
(51)

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{E}_{P,l} \left| \int_{\overline{T}_{l-1} + \Delta^{-\alpha} T^{1-\alpha}}^{\overline{T}_l} \psi_s \{ d\overline{N}_{s,c}^{(i)} - d\overline{N}_s^{(i)} \} \right|^p = O_{\mathbb{P}}((\Delta T)^{\epsilon \kappa - p}), \tag{52}$$

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{E}_{P,l} \left| \int_{\overline{T}_{l-1} + \Delta^{-\alpha} T^{1-\alpha}}^{T_l} \chi(s, \overline{T}_l) \{ d\overline{N}_{s,c}^{(i)} - d\overline{N}_s^{(i)} \} \right|^p = O_{\mathbb{P}}((\Delta T)^{\kappa}).$$
(53)

Proof of Lemma 9. To prove Equation (51), it is sufficient to show by recurrence on  $q \in \mathbb{N}$  that for every p of the form  $p = 2^q$ , we have for any  $n \in \mathbb{N}$ , any  $t \in [\overline{T}_{l-1}, \overline{T}_l]$  any  $i = 1, \dots, d$ , and uniformly in (P, l)

$$\mathbb{E}_{P,l}|\overline{\lambda}_{t,c}^{(i)} - \overline{\lambda}_t^{(i)}|^{2^q} \le L_q + M_q e^{-\left(\frac{t}{\beta^+}\right)^{p^-}}.$$
(54)

Here,  $L_q$  and  $M_q$  depend on n and q only,  $L_q = O_{\mathbb{P}}((\Delta T)^{\kappa})$ , and  $M_q$  is of polynomial growth in n. Also,  $\beta^+$  is an upper bound of the parameter  $\beta$  which exists by Conditions 2 (a) and (b). By taking the supremum over the set  $[\overline{T}_{l-1} + \Delta^{-\alpha}T^{1-\alpha}, \overline{T}_l]$  and the fact that  $M_q e^{-\left(\frac{t}{\beta^+}\right)^{p^-}\Delta^{-\alpha}T^{1-\alpha}} = o_{\mathbb{P}}((\Delta T)^{\kappa})$ , we get

$$\mathbb{E}_{P,l}|\overline{\lambda}_{t,c}^{(i)} - \overline{\lambda}_{t}^{(i)}|^{p} = O_{\mathbb{P}}((\Delta T)^{\kappa}),$$

uniformly on the set  $\mathbf{E}_n^{\alpha}$ .

In what follows, we show Expression (54) in the case q = 0, i.e. p = 1. We obtain

$$\begin{split} |\overline{\lambda}_{t,c}^{(i)} - \overline{\lambda}_{t}^{(i)}| &\leq |\overline{\nu}_{t}^{*} - \overline{\nu}_{\overline{T}_{l-1}}^{*}|^{(i)} + \Big| \int_{\overline{T}_{l-1}}^{t} \left( h(t-s,\overline{\theta}_{s}^{*}) - h(t-s,\overline{\theta}_{\overline{T}_{l-1}}^{*}) \right) d\overline{N}_{s} \Big|^{(i)} \\ &+ \Big| \int_{\overline{T}_{l-1}}^{t} h(t-s,\overline{\theta}_{\overline{T}_{l-1}}^{*}) (d\overline{N}_{s,c} - d\overline{N}_{s}) \Big|^{(i)} + R^{(i)}(t). \end{split}$$

We can deduce that  $\mathbb{E}_{P,l}|\overline{\nu}_t^* - \overline{\nu}_{\overline{T}_{l-1}}^*|^{(i)} = O_{\mathbb{P}}((\Delta T)^{\kappa})$  as a consequence of Condition 2 (a). We define  $I_l$  as

$$I_l = \int_{\overline{T}_{l-1}}^t \left( h(t-s,\overline{\theta}_s^*) - h(t-s,\overline{\theta}_{\overline{T}_{l-1}}^*) \right) d\overline{N}_s.$$

By Conditions 2 (a), (b), (c) and (d), we get

$$|h(t-s,\overline{\theta}_s^*) - h(t-s,\overline{\theta}_{\overline{T}_{l-1}}^*)| \le C_t e^{-\left(\frac{t}{\beta^+}\right)^p} .$$
(55)

Then, we obtain

$$\mathbb{E}_{P,l}|I_l| \leq \mathbb{E}_{P,l} \left| \int_0^t C_t e^{-\left(\frac{t}{\beta^+}\right)^{p^-}} d\overline{N}_s \right| \\
\leq \sqrt{\mathbb{E}_{P,l} \left[ \sup_{s \in [0,t]} C_t^2 \right] \mathbb{E}_{P,l} \left| \int_0^t e^{-\left(\frac{t}{\beta^+}\right)^{p^-}} d\overline{N}_s \right|^2}.$$
(56)

Here, we use Expression (55) and the definition of  $I_l$  in the first equality, and Cauchy-Schwartz inequality in the second inequality. We obtain that the left term from the multiplication in the right side of Expression (56) is finite by Conditions 2 (a), (b), (c) and (d). We also obtain that the right term from the multiplication in the right side of Expression (56) is a.s. finite by Lemma 8. Thus, we can deduce that  $\mathbb{E}_{P,l}|I_l| = O_{\mathbb{P}}((T\Delta)^{\kappa})$  follows from Conditions 2(e). Finally, we define  $II_l$  as

$$II_{l} = \int_{\overline{T}_{l-1}}^{t} h(t-s,\overline{\theta}_{\overline{T}_{l-1}}^{*})(d\overline{N}_{s,c} - d\overline{N}_{s}).$$

We have

$$\mathbb{E}_{P,l}|II_l| \le \mathbb{E}_{P,l} \Big[ \int_0^t h(t-s,\overline{\theta}_{\overline{T}_{l-1}}^*) d|\overline{N}_{s,c} - d\overline{N}_s| \Big].$$
(57)

Here,  $d|\overline{N}_{s,c} - d\overline{N}_s|$  is the integer measure which counts the events that do not belong to both  $d\overline{N}_{s,c}$ and  $d\overline{N}_s$ , i.e. the points of  $\underline{\overline{N}}^{i,n}$  that are between the curves  $s \to \overline{\lambda}_s$  and  $s \to \overline{\lambda}_{s,c}$ . We can show that the stochastic intensity of this point process is equal to  $|\overline{\lambda}_{s,c} - \overline{\lambda}_s|$ . We obtain

$$\mathbb{E}_{P,l}|II_l| \leq \mathbb{E}_{P,l} \Big[ \int_0^t h(t-s,\overline{\theta}_{\overline{T}_{l-1}}^*) |\overline{\lambda}_{s,c} - \overline{\lambda}_s| ds \Big] \\ = \int_0^t h(t-s,\overline{\theta}_{\overline{T}_{l-1}}^*) \mathbb{E}_{P,l} |\overline{\lambda}_{s,c} - \overline{\lambda}_s| ds.$$

So far we have shown that there exists a sequence L such that  $L = O((\Delta T)^{\kappa})$  and such that the function  $f(t) = \mathbb{E}_{P,l}|\overline{\lambda}_{t,c} - \overline{\lambda}_t|$  satisfies the inequality

$$f(t) \le L + R(\overline{\theta}_{\overline{T}_{l-1}}^*) + f * h(t-s, \overline{\theta}_{\overline{T}_{l-1}}^*).$$

By an extension of Lemma 10.6 (p. 9) from the supplementary material [Clinet and Potiron, 2018], we obtain Expression (54) in the case q = 1. The case for any  $q \in \mathbb{N}^*$  can be proven with similar arguments. Finally, Equations (52) and (53) are an application of Lemma 5 to the case  $W_{\psi}(s, z) =$  $\psi_s |\mathbb{1}_{\{0 \leq z \leq \overline{\lambda}_{s,c}\}} - \mathbb{1}_{\{0 \leq z \leq \overline{\lambda}_s\}}|$  and  $W_{\chi}(s, z) = \chi(s, t) |\mathbb{1}_{\{0 \leq z \leq \overline{\lambda}_{s,c}\}} - \mathbb{1}_{\{0 \leq z \leq \overline{\lambda}_s\}}|$  with Hölder's inequality.  $\Box$ 

We define the rescaled time-changed intensity with constant parameters equal to P on each interval as

$$\overline{\lambda}_{t,c}(P) = \nu + \int_{\overline{T}_l}^t h(t-s,\theta) d\overline{N}_{s,c} \text{ for } t \in [\overline{T}_l, \overline{T}_{l+1}).$$

We also define another rescaled time-changed intensity with constant parameters equal to P on each interval as

$$\widetilde{\lambda}_{t,c}(P) = \nu + \int_{\overline{T}_l}^t h(t-s,\theta) d\overline{N}_s \text{ for } t \in [\overline{T}_l, \overline{T}_{l+1}).$$

We state in the following lemma the uniform boundedness for moments of intensities with their derivatives. This extends Lemma 10.9 (p. 12) in the supplementary materials of [Clinet and Potiron, 2018].

**Lemma 10.** We assume that Condition 1, Conditions 2 (a), (b), (c), (d), (e), (f) and (g) hold. Let  $\alpha \in (0,1)$ . We have for any integer  $p \ge 1$ , any  $i = 1, \dots, d$ , and any  $j \in \mathbb{N}$  that a.s.

$$\sup_{(P,l,n,t)\in\mathbf{E}} \mathbb{E}_{P,l} \Big[ \sup_{P\in\Theta} |\partial_{\theta}^{j} \widetilde{\lambda}_{t,c}^{(i)}(P)|^{p} \Big] \le C,$$
(58)

$$\sup_{(P,l,n,t)\in\mathbf{E}} \mathbb{E}_{P,l} \Big[ \sup_{P\in\Theta} |\partial_{\theta}^{j} \overline{\lambda}_{t,c}^{(i)}(P)|^{p} \Big] \le C,$$
(59)

$$\sup_{(P,l,t)\in\mathbf{E}_{n}^{\alpha}}\mathbb{E}_{P,l}\left[\sup_{P\in\Theta}|\partial_{\theta}^{j}\widetilde{\lambda}_{t,c}^{(i)}(P)-\partial_{\theta}^{j}\overline{\lambda}_{t,c}^{(i)}(P)|^{p}\right]=O_{\mathbb{P}}((\Delta T)^{\kappa}).$$
(60)

Here, the constants C depend solely on j.

Proof of Lemma 10. We have that the derivatives of  $\overline{\lambda}_{t}^{(i)}(P)$  can be all bounded uniformly in  $P \in \Theta$  by linear combinations of terms of the form  $\overline{\nu}$  or  $\int_{0}^{t-} (t-s)^{j} e^{-\left(\frac{t-s}{\beta^{+}}\right)^{p^{-}}} dN_{s}^{i,n}$ , for  $j \in \mathbb{N}$ . The boundedness of moments of these terms uniformly in  $n \in \mathbb{N}$  and in the time t is the consequence of Expression (40) in Lemma 6 with  $\chi(s,t) = (t-s)^{j} e^{-\left(\frac{t-s}{\beta^{+}}\right)^{p^{-}}}$ . Thus, Expression (58) follows. Expression (59) is proved with similar arguments. Finally, we show Expression (60) in what follows. We have that  $\sup_{P \in \Theta} |\partial_{P}^{j} \widetilde{\lambda}_{t,c}^{(i)}(P) - \partial_{P}^{j} \overline{\lambda}_{t,c}^{(i)}(P)|$  can be bounded by linear combinations of terms of the form  $\int_{0}^{t} (t-s)^{j} e^{-\left(\frac{t-s}{\beta^{+}}\right)^{p^{-}}} d|N^{(i)} - N_{c}^{(i)}|_{s}$ . The  $\mathbb{L}^{p}$  estimate of such expression is then easily derived by a truncation argument and Expression (53) in Lemma 9.

We define the log likelihood process of the rescaled time-changed intensity with constant parameters equal to P on the *l*-th interval  $[\overline{T}_{l-1}, \overline{T}_l)$  as

$$\overline{l}_{l,c}(P) = \sum_{i=1}^d \int_{\overline{T}_{l-1}}^{\overline{T}_l} \log(\overline{\lambda}_{t,c}^{(i)}(P)) d\overline{N}_{t,c}^{(i)} - \sum_{i=1}^d \int_{\overline{T}_{l-1}}^{\overline{T}_l} \overline{\lambda}_{t,c}^{(i)}(P) dt.$$

We introduce for any  $(P, P_0) \in \Theta^2$  the rescaled difference between the two log likelihood values

$$\overline{\mathbb{Y}}_l(P, P_0) = \Delta(\overline{l}_l(P) - \overline{l}_l(P_0)).$$
(61)

We also introduce for any  $(P, P_0) \in \Theta^2$  the rescaled difference between the two log likelihood values with constant parameters

$$\overline{\mathbb{Y}}_{l,c}(P,P_0) = \Delta(\overline{l}_{l,c}(P) - \overline{l}_{l,c}(P_0)).$$

We define the Hessian matrices of  $\bar{l}_l(P)$  and  $\bar{l}_{l,c}(P)$  as

$$\overline{\Gamma}_{l}(P) = -\Delta \partial_{P}^{2} \overline{l}_{l}(P),$$
  
$$\overline{\Gamma}_{l,c}(P) = -\Delta \partial_{P}^{2} \overline{l}_{l,c}(P).$$

Finally, we introduce the set  $\mathbf{I} = \{(P, l, n) \in \Theta \times \mathbb{N}^2 \text{ s.t. } 1 \leq l \leq M\}$ . In the following lemma, we show how close are the log likelihoods with their derivatives. This extends Lemma 10.10 (p. 13) in the supplementary materials of [Clinet and Potiron, 2018].

**Lemma 11.** We assume that Condition 1, Conditions 2 (a), (b), (c), (d), (e), (f) and (g) hold. Let  $\epsilon \in (0,1)$ , and  $L \in (0,2\kappa)$ . For any  $p \in \mathbb{N}^*$ , we have

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{E}_{P,l} |\sqrt{\Delta} \partial_P \bar{l}_l(P) - \sqrt{\Delta} \partial_P \bar{l}_{l,c}(P)|^L \xrightarrow{\mathbb{P}} 0,$$
(62)

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{E}_{P,l} \Big[ \sup_{P_0 \in \Theta} |\overline{\mathbb{Y}}_l(P_0, P) - \overline{\mathbb{Y}}_{l,c}(P_0, P)|^p \Big] = O_{\mathbb{P}}((\Delta T)^{\epsilon \kappa}),$$
(63)

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{E}_{P,l} |\overline{\Gamma}_l(P_0) - \overline{\Gamma}_{l,c}(P)|^p = O_{\mathbb{P}}((\Delta T)^{\epsilon \kappa}),$$
(64)

$$\sup_{(P,l,n)\in\mathbf{I}} \mathbb{E}_{P,l} \left| \Delta T \sup_{P_0\in\Theta} \left| \partial_P^3 \bar{l}_l(P_0) \right| \right|^p \le C \ a.s..$$
(65)

Proof of Lemma 11. We first show Expression (62). We can reexpress the expressions as

$$\sqrt{\Delta}\partial_P \bar{l}_l(P) = \sum_{i=1}^d \sqrt{\Delta} \Big( \int_{\overline{T}_{l-1}}^{\overline{T}_{l-1}+1/\Delta} \frac{\partial_P \overline{\lambda}_s^{(i)}(P)}{\overline{\lambda}_s^{(i)}(P)} dN_s^{(i)} - \int_{\overline{T}_{l-1}}^{\overline{T}_{l-1}+1/\Delta} \partial_P \overline{\lambda}_s^{(i)}(P) ds \Big)$$

and

$$\sqrt{\Delta}\partial_{P}\bar{l}_{l,c}(P) = \sum_{i=1}^{d} \sqrt{\Delta} \Big( \int_{\overline{T}_{l-1}}^{\overline{T}_{l-1}+1/\Delta} \frac{\partial_{P}\overline{\lambda}_{s,c}^{(i)}(P)}{\overline{\lambda}_{s,c}^{(i)}(P)} dN_{s,c}^{(i)} - \int_{0}^{1/\Delta} \partial_{P}\overline{\lambda}_{s,c}^{(i)}(P) ds \Big).$$

By Lemma 8 Expressions (47) and (49), and Lemma 10 Expressions (58) and (59), we can replace the lower bounds of those integrals by  $(T\Delta)^{-\alpha}T$  for some  $\alpha \in (0, 1/2)$ . Thus, the expression  $\partial_P \bar{l}_l(P) - \partial_P \bar{l}_{l,c}(P)$  is equivalent to the sum of the three terms

$$\begin{split} \sum_{i=1}^{d} \Big( \int_{\overline{T}_{l-1}+T^{1-\alpha}\Delta^{-\alpha}}^{\overline{T}_{l-1}+1/\Delta} \frac{\partial_{P}\overline{\lambda}_{s}^{(i)}(P)}{\overline{\lambda}_{s}^{(i)}(P)} (dN_{s}^{(i)} - dN_{s,c}^{(i)}) + \int_{\overline{T}_{l-1}+T^{1-\alpha}\Delta^{-\alpha}}^{\overline{T}_{l-1}+1/\Delta} \Big( \frac{\partial_{P}\overline{\lambda}_{s}^{(i)}(P)}{\overline{\lambda}_{s}^{(i)}(P)} - \frac{\partial_{P}\overline{\lambda}_{s,c}^{(i)}(P)}{\overline{\lambda}_{s,c}^{(i)}(P)} \Big) dN_{s,c}^{(i)} + \int_{\overline{T}_{l-1}+T^{1-\alpha}\Delta^{-\alpha}}^{\overline{T}_{l-1}+1/\Delta} \Big( \partial_{P}\overline{\lambda}_{s}^{(i)}(P) - \partial_{P}\lambda_{s}^{(i)}(P) \Big) ds \Big). \end{split}$$

We then apply Lemma 9 Expression (52) and Lemma 10 Expression (58) to the first term, Lemma 8 Expression (49) and Lemma 10 Expression (60) to the second term, and Lemma 10 Expression (60) to the last term, to get

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{E}_{P,l} |\sqrt{\Delta} \partial_P \bar{l}_l(P) - \sqrt{\Delta} \partial_P \bar{l}_{l,c}(P)|^L = O_{\mathbb{P}}((\Delta T)^{-\frac{L}{2} + \epsilon\kappa}), \tag{66}$$

for any  $\epsilon \in (0,1)$ . If we take  $\epsilon$  sufficiently close to 1, this expression tends to 0. We can prove Expressions (63), (64) and (65) with similar arguments.

In the following lemma, we show that the Hessian matrix of the log likelihood can be approximated by the Hessian matrix from the stationary case. This extends Lemma 10.10 (p. 13) in the supplementary materials of [Clinet and Potiron, 2018].

**Lemma 12.** We assume that Condition 1, Conditions 2 (a), (b), (c), (d), (e), (f) and (g) hold. For any integer  $p \ge 1$ , there exists a constant C such that a.s.

$$\sup_{(P,l,n)\in\mathbf{I}} \mathbb{E}_{P,l} |\sqrt{\Delta}\partial_P \bar{l}_{l,c}(P)|^p \le C.$$
(67)

Furthermore, there exists a function  $(P, P_0) \to \mathbb{Y}(P, P_0)$  such that for any  $\epsilon \in (0, 1)$ , we have

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{E}_{P,l} \Big[ \sup_{P_0 \in \Theta} |\mathbb{Y}_{l,c}(P_0, P) - \mathbb{Y}_l(P_0, P)| \Big] = O_{\mathbb{P}}((\Delta T)^{\epsilon \frac{p}{2}}).$$
(68)

Finally, for any  $P_0 \in \Theta$  and for any  $\epsilon \in (0, 1)$ , we have

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{E}_{P,l} |\overline{\Gamma}_{l,c}(P) - \Gamma(P)|^p = O_{\mathbb{P}}((\Delta T)^{\epsilon \frac{p}{2}}).$$
(69)

Proof of Lemma 12. When  $\overline{P}_{\overline{T}_{l-1}}^* = P$ , the point process  $\overline{N}_c$  is simply a stationary parametric Hawkes process with parameter P. By a regular distribution argument, the operator  $\mathbb{E}_{P,l}$  acts as the simple operator  $\mathbb{E}$  for  $\overline{N}_c$  distributed as a Hawkes with parameter P. We can conclude by extending the arguments in the proofs of Lemma 3.15 (p. 1813) and Theorem 4.6 (p. 1821) in [Clinet and Yoshida, 2017].  $\Box$ 

In what follows, we show that the functional of the rescaled estimator can be approximated by the stationary case. This extends Theorem 10.12 (p. 15) in the supplementary materials of [Clinet and Potiron, 2018].

**Lemma 13.** We assume that Condition 1, Conditions 2 (a), (b), (c), (d), (e), (f) and (g) hold. Let  $L \in (0, 2\kappa)$ . We have

$$\sup_{P \in \Theta, 1 \le l \le M} \left\{ \mathbb{E}_{P,l}[f(\sqrt{\Delta/T}(\widehat{P}_l - P))] - \mathbb{E}[f(T^{-\frac{1}{2}}\Gamma(P)^{-\frac{1}{2}}\xi)] \right\} \xrightarrow{\mathbb{P}} 0,$$
(70)

for any continuous function f with  $|f(x)| = O(|x|^L)$  when  $|x| \to \infty$ .

Proof of Lemma 13. By Lemma 11 Equation (63) and Lemma 12 Equation (68), we can define some real number  $\epsilon \in (0, 1)$  such that

$$\sup_{P \in \Theta, 1 \le l \le M} (\Delta T)^{-\epsilon(\frac{p}{2} \land \kappa)} \mathbb{E}_{P,l} \Big[ \sup_{P_0 \in \Theta} |\mathbb{Y}_l(P_0, P) - \mathbb{Y}(P_0, P)|^p \Big] \xrightarrow{\mathbb{P}} 0.$$
(71)

As  $\hat{P}_l$  is a maximizer of  $P_0 \to \mathbb{Y}_l(P_0, P)$  and  $\mathbb{Y}$  satisfies the non-degeneracy condition [A4] used in [Clinet and Yoshida, 2017], Expression (71) implies the uniform consistency in the block index l and the initial value of  $\hat{P}_l$  to  $P^*_{T_{l-1}}$ , i.e.

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{P}_{P,l} \left[ \widehat{P}_l - P \right] \xrightarrow{\mathbb{P}} 0.$$
(72)

From Lemma 11 Equation (64) and Lemma 12 Equation (69), we can deduce

$$\sup_{P \in \Theta, 1 \le l \le M} (\Delta T)^{-\epsilon(\frac{p}{2} \wedge \kappa)} \mathbb{E}_{P,l} |\overline{\Gamma}_l(P) - \Gamma(P)|^p \xrightarrow{\mathbb{P}} 0.$$
(73)

By Lemma 11 Equation (62),  $\sqrt{\Delta}\partial_P \bar{l}_l(P)$  and  $\sqrt{\Delta}\partial_P \bar{l}_{l,c}(P)$  have the same asymptotic distribution  $\Gamma(P)^{\frac{1}{2}}\xi$ . With the same arguments as in the proof of Theorem 3.11 in [Clinet and Yoshida, 2017],

we can deduce that  $\sqrt{(\Delta T)^{-1}}(\hat{P}_l - P)$  converges uniformly in distribution to  $T^{-\frac{1}{2}}\Gamma(P)^{-\frac{1}{2}}\xi$  when  $P_{T_{l-1}}^* = P$ . More specifically, we have

$$\sup_{P \in \Theta, 1 \le l \le M} \left\{ \mathbb{E}_{P,l} \left[ f \left( \sqrt{(\Delta T)^{-1}} (\widehat{P}_l - P) \right) \right] - \mathbb{E} \left[ f \left( T^{-\frac{1}{2}} \Gamma(P)^{-\frac{1}{2}} \xi \right) \right] \right\} \xrightarrow{\mathbb{P}} 0, \tag{74}$$

for any bounded continuous function f.

Finally, we extend Expression (74) to the case of a function of polynomial growth of order smaller than L. First, we have by Lemma 11 Equation (62) and Lemma 12 Equation (67) for any  $L' \in (L, 2\kappa)$ that

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{E}_{P,l} |\sqrt{\Delta} \partial_P \bar{l}_l(P)|^{L'} = O_{\mathbb{P}}(1).$$
(75)

We now adopt the notations of [Yoshida, 2011] and define  $\beta_1 = \frac{\epsilon}{2}$ ,  $\beta_2 = \frac{1}{2} - \beta_1$ ,  $\rho = 2$ ,  $0 < \rho_2 < 1 - 2\beta_2$ ,  $0 < \alpha < \frac{\rho_2}{2}$ , and  $0 < \rho_1 < \min\{1, \frac{\alpha}{1-\alpha}, \frac{2\beta_1}{1-\alpha}\}$  all sufficiently small so that  $M_1 = L(1-\rho_1)^{-1} < L'$ ,  $M_4 = \beta_1 L(\frac{2\beta_1}{1-\alpha} - \rho_1)^{-1} < 2\frac{\gamma(\delta-1)}{2} = \kappa$ ,  $M_2 = (\frac{1}{2} - \beta_2)L(1 - 2\beta_2 - \rho_2)^{-1} < \kappa$  and finally  $M_3 = L(\frac{\alpha}{1-\alpha} - \rho_1)^{-1} < \infty$ . Then, by Expressions (65), (71), (73) and (75), the conditions [A1''], [A4'], [A6], [B1] and [B2] in [Yoshida, 2011] are satisfied. By extending the arguments of the proofs, we can use Theorem 3 and Proposition 1 from [Yoshida, 2011] to get for any  $p \leq L$  that

$$\sup_{P \in \Theta, 1 \le l \le M} \mathbb{E}_{P,l} |\sqrt{(\Delta T)^{-1}} (\widehat{P}_l - P)|^p = O_{\mathbb{P}}(1).$$
(76)

We denote  $\mathbb{E}[.|\overline{\mathcal{F}}_{\overline{T}_{l-1}}]$  by  $\mathbb{E}_{l-1}$ . So far we have focused on the case where  $R(\overline{T}_l)$  is bounded by the sequence K. In the following lemma, we relax this assumption. This extends Theorem 5.2 (p. 3479) in [Clinet and Potiron, 2018].

**Lemma 14.** We assume that Condition 1, Conditions 2 (a), (b), (c), (d), (e), (f) and (g) hold. Let  $L \in [0, 2\kappa)$ . We have uniformly in  $l \in \{1, \dots, M\}$  that

$$\mathbb{E}_{l-1}\left[f(\sqrt{(\Delta T)^{-1}}(\widehat{P}_l - \overline{P}_{\overline{T}_{l-1}}^*))\right] = \mathbb{E}_{l-1}\left[f(T^{-1/2}\Gamma(\overline{P}_{\overline{T}_{l-1}}^*)^{-\frac{1}{2}}\xi)\right] + o_{\mathbb{P}}(1), \tag{77}$$

for any continuous function f with  $|f(x)| = O(|x|^L)$  when  $|x| \to \infty$ .

Proof of Lemma 14. We can decompose  $\mathbb{E}_{l-1}\left[f(\sqrt{(\Delta T)^{-1}}(\widehat{P}_l - \overline{P}^*_{\overline{T}_{l-1}}))\right]$  as

$$\mathbb{E}_{l-1}\left[f(\sqrt{(\Delta T)^{-1}}(\widehat{P}_l - \overline{P}^*_{\overline{T}_{l-1}}))\mathbf{1}_{\{R(\overline{T}_{l-1}) \le K\}}\right]$$
(78)

$$+ \mathbb{E}_{l-1} \Big[ f(\sqrt{(\Delta T)^{-1}} (\widehat{P}_l - \overline{P}_{\overline{T}_{l-1}}^*)) \mathbf{1}_{\{R(\overline{T}_{l-1}) > K\}} \Big].$$

$$\tag{79}$$

We define G as  $G(P) = \mathbb{E}_{P,l} \left[ f\left(\sqrt{(\Delta T)^{-1}}(\widehat{P}_l - P)\right) - \mathbb{E} \left[ f\left(T^{-\frac{1}{2}}\Gamma(P)^{-\frac{1}{2}}\xi\right) \right]$ . By a regular conditional distribution argument, we can express uniformly in  $l \in \{1, \dots, M\}$ 

$$\mathbb{E}_{l-1}\left[f(\sqrt{(\Delta T)^{-1}}(\widehat{P}_l - \overline{P}^*_{\overline{T}_{l-1}}))\mathbf{1}_{\{R(\overline{T}_{l-1}) \le K\}} - f(T^{-\frac{1}{2}}\Gamma(\overline{P}^*_{\overline{T}_{l-1}})^{-\frac{1}{2}}\xi)\right]$$
(80)

as  $G(\overline{P}^*_{\overline{T}_{l-1}})$ . We have that

$$\left|G(\overline{P}_{\overline{T}_{l-1}}^{*})\right| \leq \sup_{P \in \Theta} \left|\mathbb{E}_{P,l}\left[f(\sqrt{(\Delta T)^{-1}}(\widehat{P}_{l} - P))\right] - \mathbb{E}\left[f(T^{-\frac{1}{2}}\Gamma(P)^{-\frac{1}{2}}\xi)\right]\right|,\tag{81}$$

If we take the supremum over l in Expression (81) and use Theorem 13, we can show that the term (80) is uniformly equal to  $o_{\mathbb{P}}(1)$ . Moreover, the term (79) is bounded by  $(\Delta T)^{-L}Q\mathbf{1}_{\{R(\overline{T}_{l-1})>K\}}$  for some Q > 0 by Condition 2 (a). We can show that  $\mathbb{P}[R(\overline{T}_{l-1}) > K] \leq \mathbb{P}[\lambda_{\overline{T}_{l-1}} > K]$ , which then can be dominated by  $K^{-1}\mathbb{E}[\lambda_{\overline{T}_{l-1}}] = O(nK^{-1})$  using Markov's inequality. By definition of K, we can deduce that the term (79) goes to 0 asymptotically.

In the following lemma, we consider the bias correction of the estimation procedure. This extends Theorem 5.3 (p. 3479) in [Clinet and Potiron, 2018].

**Lemma 15.** We assume that Condition 1, Conditions 2 (a), (b), (c), (d), (e), (f) and (g) hold. Let  $\epsilon \in (0,1)$ . The bias of the estimator  $\hat{P}_l$  has the expansion

$$\mathbb{E}_{l-1}\left[\widehat{P}_l - \overline{P}^*_{\overline{T}_{l-1}}\right] = O_{\mathbb{P}}((\Delta T)^{\epsilon(\kappa \wedge \frac{3}{2})}),\tag{82}$$

uniformly in  $l \in \{1, \cdots, M\}$ .

Proof of Lemma 15. We can prove the lemma by extending the arguments from the proof of Theorem 5.3 (p. 21) in the supplementary materials of [Clinet and Potiron, 2018].  $\Box$ 

We introduce the bias corrected local estimator  $\hat{P}_l^{(BC)} = \hat{P}_l - b(\hat{P}_l)$ . We also define for any random variable X the conditional variance as  $\operatorname{Var}_l[X] = \mathbb{E}_l[(X - \mathbb{E}_l[X])^2]$ . The following lemma gives conditions for the proof of Theorem 1. This extends Conditions  $[C^*]$  (p. 21) in the supplementary material of [Clinet and Potiron, 2018].

**Lemma 16.** We assume that Conditions 1 and 2 hold. We have uniformly in  $l \in \{1, \dots, M\}$  that there exists  $\epsilon > 0$  such that

$$\operatorname{Var}_{l-1}\left[\sqrt{(\Delta T)^{-1}} \left(\widehat{P}_{l}^{(BC)} - P_{T_{l-1}}^{*}\right)\right] = T^{-1} \Gamma(P_{T_{l-1}}^{*})^{-1} + o_{\mathbb{P}}(1),$$
(83)

$$\mathbb{E}_{l-1}\left[\left|\sqrt{(\Delta T)^{-1}}(\hat{P}_{l}^{(BC)} - P_{T_{l-1}}^{*})\right|^{2+\epsilon}\right] = O_{\mathbb{P}}(1), \tag{84}$$

$$\mathbb{E}_{l-1}\left[\widehat{P}_{l}^{(BC)} - P_{T_{l-1}}^{*}\right] = o_{\mathbb{P}}(n^{-1/2}).$$
(85)

Proof of Lemma 16. First, we have for any  $L \in (0, 2\kappa)$  and uniformly in  $l \in \{1, \dots, M\}$  that

$$\mathbb{E}_{l-1} \left| \sqrt{(\Delta T)^{-1}} \left( \widehat{P}_l^{(BC)} - \widehat{P}_l \right) \right|^L = (\Delta T)^{\frac{L}{2}} T^{-L} \mathbb{E}_{l-1} \left| b(\widehat{P}_l) \right|^L = O_{\mathbb{P}}((\Delta T)^{\frac{L}{2}}).$$

Then, we can show that Lemma 14 still holds if  $\hat{P}_l$  is replaced by  $\hat{P}_l^{(BC)}$ . We decompose the conditional variance in Equation (83) as

$$\mathbb{E}_{l-1}\Big[\big(\sqrt{(\Delta T)^{-1}}\big(\widehat{P}_{l}^{(BC)}-P_{T_{l-1}}^{*}\big)\big)^{2}\Big]-\mathbb{E}_{l-1}\Big[\sqrt{(\Delta T)^{-1}}\big(\widehat{P}_{l}^{(BC)}-P_{T_{l-1}}^{*}\big)\Big]^{2}.$$

Then, we can show Equations (83) and (84) by an application of Theorem 14. Finally, Equation (85) holds if there exists  $\epsilon \in (0, 1)$  such that  $\sqrt{n} = o_{\mathbb{P}}((\Delta/T)^{\epsilon(\kappa \wedge \frac{3}{2})})$  by Lemma 15. From the relation  $\sqrt{n} = (\Delta T)^{-\frac{\delta}{2}}$ , this can be reexpressed as  $\frac{\delta}{2} < \kappa \wedge \frac{3}{2}$ . By the definition of  $\kappa$ , we get the two conditions  $\frac{\delta}{2} < \gamma(\delta - 1)$  and  $\frac{\delta}{2} < \frac{3}{2}$ , i.e.  $\frac{\gamma}{\gamma - \frac{1}{2}} < \delta < 3$ . This corresponds to Condition 2 (h).

In what follows, we give the proof of Theorem 1. This is based on Theorem 3.2 (p. 244) in [Jacod, 1997]. See also [Potiron and Mykland, 2020]. This extends the arguments from the proof of Theorem 5.4 (pp. 21-24) in the supplementary materials of [Clinet and Potiron, 2018].

Proof of Theorem 1. By the definition (12), we have

$$\sqrt{n}(\widehat{IP}(T) - IP(T)) = \sqrt{n} \Big( \sum_{l=1}^{M} \widehat{P}_l^{(BC)} \Delta - IP(T) \Big).$$
(86)

We introduce the bias increments  $B_l = \mathbb{E}_{l-1} \left[ \hat{P}_l - P^*_{T_{l-1}} \right]$  and the martingale increments

$$M_{l} = \hat{P}_{l} - P_{T_{l-1}}^{*} - B_{l}.$$

Then, we define the rescaled sum of the bias increments as

$$S^{(B)} = \sqrt{n} \sum_{l=1}^{M} B_l.$$

Finally, we define the rescaled sum of the martingale increments as

$$S^{(M)} = \sqrt{n} \sum_{l=1}^{M} M_l.$$

We can rewrite the right side of Equation (86) as

$$\sqrt{n} \Big( \sum_{l=1}^{M} \widehat{P}_{l}^{(BC)} \Delta - IP(T) \Big) = \sqrt{n} \sum_{l=1}^{M} \Big( \widehat{P}_{l}^{(BC)} - P_{T_{l-1}}^{*} \Big) \Delta + \sqrt{n} \Big( \sum_{l=1}^{M} P_{T_{l-1}}^{*} \Delta - IP(T) \Big).$$
(87)

We first show that

$$\sqrt{n} \Big( \sum_{l=1}^{M} P_{T_{l-1}}^* \Delta - IP(T) \Big) = o_{\mathbb{P}}(1).$$
(88)

To show Equation (88), it is sufficient to show that

$$\sqrt{n} \sum_{l=1}^{M} \left| P_{T_{l-1}}^* \Delta - \int_{T_{l-1}}^{T_l} P_s^* ds \right| = o_{\mathbb{P}}(1).$$
(89)

By the triangular inequality, we can deduce that

$$\sqrt{n}\sum_{l=1}^{M} \left| P_{T_{l-1}}^* \Delta - \int_{T_{l-1}}^{T_l} P_s^* ds \right| \le \sqrt{n}\sum_{l=1}^{M} \int_{T_{l-1}}^{T_l} \left| P_{T_{l-1}}^* - P_s^* \right| ds.$$
(90)

By Condition 2 (e), we can deduce that the right side in Expression (90) is equal to  $O_{\mathbb{P}}((\Delta T)^{-\gamma}n^{\frac{1}{2}-\gamma})$ . By Condition 2 (h) and since  $\gamma > \frac{1}{2}$ , we can deduce that the right side in Expression (90) converges to 0 in probability. Then, we can prove Equation (88) by Expressions (89) and (90). Moreover, we can show that  $S^{(B)} \xrightarrow{\mathbb{P}} 0$  by the condition (85) from Lemma 16.

To show that  $S^{(M)} \xrightarrow{\mathcal{D}-s} \int_0^T c_{IP}(t) dW_t$ , we use Theorem 3.2 (p. 244) in [Jacod, 1997]. First, we consider the conditional Lindeberg condition (3.13), i.e. for any  $\eta > 0$  we have

$$n\sum_{l=1}^{M} \mathbb{E}_{l-1} \Big[ M_l^2 \mathbf{1}_{\{\sqrt{n}M_l > \eta\}} \Big] \xrightarrow{\mathbb{P}} 0.$$
(91)

Let  $\eta > 0$ . Using Hölder's inequality, we obtain that

$$n\mathbb{E}_{l-1}\left[M_l^2\mathbf{1}_{\{\sqrt{n}M_l>\eta\}}\right] \le \left(\mathbb{E}_{l-1}\left[(\sqrt{n}M_l)^{2+\epsilon}\right]\right)^{\frac{2}{2+\epsilon}} \left(\mathbb{E}_{l-1}\left[\mathbf{1}_{\{\sqrt{n}M_l>\eta\}}\right]\right)^{\frac{\epsilon}{2+\epsilon}}$$

First, we have that

$$\left(\mathbb{E}_{l-1}\left[(\sqrt{n}M_l)^{2+\epsilon}\right]\right)^{\frac{2}{2+\epsilon}}$$

is uniformly bounded by the condition (84) from Lemma 16. Second, we have that

$$\left(\mathbb{E}_{l-1}\left[\mathbf{1}_{\{\sqrt{n}M_l > \eta\}}\right]\right)^{\frac{\epsilon}{2+\epsilon}}$$

goes uniformly to 0 by the condition (84) from Lemma 16 and Condition 2 (g). Then, we can deduce that the Lindeberg condition (91) holds. We now prove the conditional variance condition (3.11), i.e. that

$$n\sum_{l=1}^{M} \mathbb{E}_{l-1}\left[M_l^2\right] \xrightarrow{\mathbb{P}} \int_0^T c_{IP}(t)c_{IP}(t)^T dt.$$
(92)

We have that

$$n\sum_{l=1}^{M} \mathbb{E}_{l-1}[M_l^2] = \frac{1}{T}\sum_{l=1}^{M} (\Delta T)^{-1} \mathbb{E}_{l-1}[M_l^2] \Delta.$$

Then, we use Proposition I.4.44 (p. 51) in [Jacod and Shiryaev, 2003] with the condition (83) from Lemma 16 to show Expression (92). The conditions (3.10) and (3.12) are satisfied because  $M_l$  is a martingale increment and since we consider the reference continuous martingale  $\mathbf{M} = 0$ . Finally, we show the condition (3.14). We thus consider a bounded  $\mathcal{F}^{P^*}$ -martingale Z and we show that

$$\sqrt{n} \sum_{l=1}^{M} \mathbb{E}_{l-1} \left[ M_l \Delta Z_l \right] \xrightarrow{\mathbb{P}} 0, \tag{93}$$

where  $\Delta Z_l = Z_{T_l} - Z_{T_{l-1}}$ . Using the arguments from the proof of Lemma 15, we obtain

$$\sqrt{n} \sum_{l=1}^{M} \mathbb{E}_{l-1} \left[ M_l \Delta Z_l \right] = \sqrt{n} \sum_{l=1}^{M} \Gamma(P_{T_{l-1}}^*)^{-1} \mathbb{E}_{l-1} \left[ \partial_P \bar{l}_{T_l,c}(P_{T_{l-1}}^*) \Delta Z_l \right] + o_{\mathbb{P}}(1)$$

We have that  $\bar{l}_{T_{l,c}}(P^*_{T_{l-1}})$  can be rewritten as an integral over the canonical Poisson martingale

$$\bar{l}_{T_{l,c}}(P_{T_{l-1}}^{*}) = \sum_{i=1}^{d} \int_{0}^{\Delta^{-1}} \int_{\mathbb{R}_{+}} \frac{\partial_{\theta} \lambda_{s,c}^{(i)}(P_{T_{l-1}}^{*})}{\lambda_{s,c}^{(i)}(P_{T_{l-1}}^{*})} \mathbb{1}_{\{0 \le z \le \lambda_{s,c}^{(i)}(P_{T_{l-1}}^{*})\}} \\ \{ \underline{\overline{N}}^{(2i-1)} * \underline{\overline{N}}^{(2i)}(ds, dz) - \underline{\overline{\Lambda}}^{(2i-1)} * \underline{\overline{\Lambda}}^{(2i)}(ds, dz) \}.$$

We can deduce that  $\mathbb{E}_{l-1}\left[\partial_P \overline{l}_{T_{l,c}}(P_{T_{l-1}}^*)\Delta Z_l\right] = 0$ , since both  $\sigma$ -fields  $\mathcal{F}_T^{P^*}$  and  $\mathcal{F}_T^{\overline{N}}$  are independent. Thus, Z and  $\overline{N}^{(2i-1)} * \overline{N}^{(2i)} - \overline{\Lambda}^{(2i-1)} * \overline{\Lambda}^{(2i)}$  are orthogonal. This implies that Expression (93) holds. Thus, we can deduce that  $S^{(M)} \xrightarrow{\mathcal{D}} \int_0^T c_{IP}(t) dW_t$  and that Expression (27) holds. From the definition of the variance estimator (15) and the variance (19), we can deduce that the estimator of variance is consistent. Finally, we can deduce Expression (28) by Slutsky's theorem.

In what follows, we give the proof of Theorem 2. This is an application of Theorem 1.

Proof of Theorem 2. The proof of the theorem can be obtained by an application of Theorem 1 with a Taylor expansion of the latency function F.

In what follows, we consider estimation of the integral of linear latency  $A\overline{IL}(T)$ . We consider the natural estimation procedure  $\widehat{AIL} = A\widehat{\overline{IL}}(T)$ . We can also define the asymptotic covariance matrix for estimation of linear latency integral as

$$c_{A\overline{IL}}(t)c_{A\overline{IL}}(t)^{T} = A\overline{\Gamma}(\theta_{t}^{*})^{-1}A^{T},$$

$$C_{A\overline{IL}}(T) = \int_{0}^{T} c_{A\overline{IL}}(t)c_{A\overline{IL}}(t)^{T}dt.$$
(94)

Finally, we propose estimation for the asymptotic covariance matrix for the integral of linear latency as

$$\widehat{C}_{A\overline{IL}}(T) = \sum_{l=1}^{M} A \widehat{\overline{\Gamma}}_{l}^{-1} A^{T} \Delta.$$
(95)

In the corollary that follows, we extend Theorem 2 to the case of integral of linear latency. This is an application of Theorem 2.

**Corollary 2.** We assume that Conditions 1, 2 and 3 hold. There is an extension of  $\mathcal{B}$  on which is defined a standard Brownian motion W, which is of dimension a, such that we have the CLT and the feasible CLT

$$\sqrt{n} \left( A \widehat{\overline{IL}}(T) - A \overline{IL}(T) \right) \xrightarrow{\mathcal{D}-s} \int_0^T c_{A \overline{IL}}(t) dW_t, \tag{96}$$

$$\sqrt{n}\widehat{C}_{A\overline{IL}}^{-1/2}(T)(\widehat{A\overline{IL}}(T) - A\overline{IL}(T)) \xrightarrow{\mathcal{D}-s} \overline{\xi}.$$
(97)

*Proof of Corollary 2.* The proof of the corollary can be obtained by an application of Theorem 2.  $\Box$ 

Finally, we give the proof of Corollary 1 in what follows. This is an application of Corollary 2.

Proof of Corollary 1. The proof that the test statistic S converges in distribution to a chi-squared random variable with one degree of freedom under the null hypothesis  $H_0$  can be obtained by an application of Corollary 2 with a Taylor expansion. If we also assume Condition 4, we can make a Taylor expansion of the linear latency. By Condition 2 (f), we obtain that the test statistic S is consistent under the alternative hypothesis  $H_1$ , i.e. we have  $\mathbb{P}(S > Q(u) \mid H_1) \to 1$  for any 0 < u < 1.