Estimation of time-dependent latency with locally stationary Hawkes processes

Deniz Erdemlioglu*Yoann Potiron†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 financial data shows that the test for constancy of latency is always rejected.

Keywords: latency matrix; time-dependent; mutually exciting Hawkes processes; local parametric estimation; constancy test; in-fill asymptotics; finance; high-frequency 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

*?. E-mail: ?

†(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 λ its corresponding intensity and d as its dimension, a standard definition of Hawkes mutually exciting processes is given by

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

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

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

We 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 latency estimator is defined as the known function of estimated kernel parameters. 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 mutually exciting Hawkes 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.

In this paper, 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. $\lambda_t = \alpha_n \nu_t^* + \int_0^t h(t-s, \theta^*) dN_s$, is considered where $\alpha_n \to \infty$ is a scaling sequence. [Clinet and Potiron, 2018] introduce a joint boosting of the baseline and the kernel, i.e. $\lambda(t) = n\nu_t^* + \int_0^t n a_s^* \exp(-n b_s^*(t-s)) dN_s$. [Kwan et al., 2023] revisit [Chen and Hall, 2013] with the same in-fill asymptotics as in [Clinet and Potiron, 2018], i.e. $\lambda_t = n\nu_t^* + \int_0^t n a^* \exp(-n b^*(t-s))\, dN_s$. [Kwan, 2023], [Potiron and Volkov, 2025] and [Potiron et al., 2025] also use these in-fill asymptotics.

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 material contains 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 a vector V, we denote its i-th component as $V^{(i)}$. In what follows, we introduce the multidimensional point process N_t. Each component of the point process $N_t^{(i)}$ counts the number of events between 0 and t. For $i = 1, \cdots, d$, 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 $\mathbb{N}$. Here, $\mathcal{B}([0, T])$ is the Borel σ-algebra on the compact space $[0, T]$, $N^{(i)}(C) = \sum_{k \in \mathbb{N}} \mathbf{1}_C(T_k)$ and $\{T_k^{(i)}\}_{k \in \mathbb{N}}$ is a sequence of $\mathbb{R}^+$-valued random event times such that, a.s. $T_0^{(i)} = 0 < T_1^{(i)} < \ldots < T^{(i)}_{N_T^{(i)}} < T < T^{(i)}_{N_T^{(i)}+1}$. 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 $\mathcal{F}_t^X = \sigma\big(X(C), C \in \mathcal{B}([0, T]), C \subset [0, t]\big)$. We assume that, for any $t \in [0, T]$, the canonical filtration of 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 $\mathbb{E}[N((a, b]) \mid \mathcal{F}_a] = \mathbb{E}\big[\int_a^b \lambda_s ds \,\big|\, \mathcal{F}_a\big]$ 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, 2003a], [Daley and Vere-Jones, 2003b], and [Daley and Vere-Jones, 2008].

For a matrix ϕ, 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, ν_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, θ_t^* are the random time-dependent parameters.

The random time-dependent latency is defined as a $d \times d$-dimensional matrix which is a time-invariant known function of the parameters θ_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.

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

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 $P_t^* = (\mu_t^*, \theta_t^*)$. Another goal is to estimate the integral of the parameters, i.e.

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

The problem of the integral of baseline and the kernel parameters is not novel, since it was already considered in [Clinet and Yoshida, 2017].

3 Estimation

In this section, we introduce locally parametric Hawkes processes with in-fill asymptotics, local parametric estimation based on MLE.

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.

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

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.

For any space S such that $0 \in S$, we define the space without zero as S^*. We also 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 Δ, where $T_l = l\Delta$ are the start and end points of each interval. For $l = 1, \dots, M$, we rely on the log likelihood process (see [Ogata, 1978] and [Daley and Vere-Jones, 2003a]) on the l-th interval $[T_{l-1}, T_l)$, i.e.

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

Here, $P = (\mu, \theta)$ are the parameters of the baseline μ and the parameters of the kernel θ, and they belong to the parameter space Θ. 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 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.

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

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

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

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

Here, b corresponds to the bias correction required for local estimation, and is 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 ψ is of polynomial growth in (u, v, w) and in $(\frac{1}{u}, \frac{1}{v}, w)$. 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 time-invariant 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 i-th 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 supplement 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 $\overline{X}_{nT}^{(i)}(P^*, P) \xrightarrow{\mathcal{D}} (\overline{\lambda}_l^{(i)}(P^*, P^*), \overline{\lambda}_l^{(i)}(P^*, P), \partial_\theta \overline{\lambda}_l^{(i)}(P^*, P))$. 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 Γ when the true parameter is time-invariant equal to P^* as

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

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

$$C_{IP}(T) = \int_0^T \Gamma(P_s^*)^{-1} ds$$

For $l = 1, \ldots, M$, 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

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

Here, we define $\overline{N}_t^{(i)} = N_{\frac{t}{n}}^{(i)}$ as the time-transformed point process. Then, we propose local estimation of the inverse Fisher information matrix as

$$\widehat{\Gamma}_l^{-1} = -\partial_P^2 \bar{l}_l\big(\widehat{P}_l - b(\widehat{P}_l), \widehat{P}_l - b(\widehat{P}_l)\big). \tag{14}$$

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

$$\widehat{C}_{IP}(T) = \sum_{l=1}^{M} \widehat{\Gamma}_l^{-1} \Delta. \tag{15}$$

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

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

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

$$Q_T(P^*)_{k,lq} = \sum_{i=1}^{d} -\frac{M^{(i,k)}(P^*)}{T} \int_0^T \frac{\partial_\theta \lambda^{(i,l)}(t, \theta^*) \partial_\theta \lambda^{(i,q)}(t, \theta^*)}{\lambda^{(i)}(t, \theta^*)} dt.$$

Then, we define for any $k \in \{1, \cdots, m\}$ the k-th component of the bias function as

$$b(\theta^*)^{(k)} = \frac{1}{2} \sum_{q=1}^{m} \sum_{l=1}^{m} \sum_{j=1}^{m} \Gamma(\theta^*)^{(j,k)} \Gamma(\theta^*)^{(l,q)} (K(\theta^*)^{(j,l,q)} + 2\{C(\theta^*)_{l,jq} + Q(\theta^*)_{l,jq}\}). \tag{16}$$

4 Theory

In this section, we start with showing an existence result for locally parametric mutually exciting Hawkes processes, where the baseline and the parameters of the kernels are random time-dependent. Then, our main results characterize feasible statistics induced by central limit theory for estimation of the integral of latency and the integral of parameters.

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 θ_t^* from the time t as $\phi_t = \int_0^\infty h(s, \theta_{t+s}^*)ds$. For a matrix ϕ, we denote its spectral radius as $\rho(\phi)$. Let us introduce a set of conditions required for the existence of locally parametric mutually exciting Hawkes processes where the baseline and the parameters of the kernels are time-dependent.

Condition 1. (a) The parameter P_t^* belongs to Θ a.e. a.s., i.e. $\mathbb{P}(P_t^* \in \Theta \ \forall t \in [0, T]) = 1$.

(b) For $i = 1, \cdots, d$, the i-th component of the baseline is positive a.e. a.s., i.e. $\mathbb{P}(\nu_t^{*,(i)} > 0 \ \forall t \in [0, T]) = 1$.

(c) For $i = 1, \cdots, d$, the i-th component of the baseline is integrable a.s., i.e. $\mathbb{P}(\int_0^T \nu_s^{*,(i)} ds < \infty) = 1$.

(d) For any $0 \leq t \leq T$, we have $\mathcal{F}_t = \mathcal{F}_t^{P^*} \vee \mathcal{F}_t^{\overline{N}}$, where the filtration $\mathcal{F}_t^{P^*}$ is independent from the other filtration $\mathcal{F}_t^{\overline{N}}$. We also have $\overline{N}$ is a $2d$-dimensional $\mathcal{F}_t$-adapted Poisson process of intensity 1 that generates N_t, i.e. $N_t^{(i)} = \int_{[0,t]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_s^{(i)}]}(x) \overline{N}^{(2i-1)} * \overline{N}^{(2i)}(ds \times dx)$ for $i = 1, \cdots, d$.

(e) For $i = 1, \cdots, d$ and $j = 1, \cdots, d$, the component (i, j) of the kernel is positive a.e. a.s., i.e. $\mathbb{P}(h(s, \theta_{t+s}^*) \geq 0 \ \forall (t, s) \in [0, T]^2) = 1$.

(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, 2003b]). 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 also 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 Γ. For any $i = 1, \dots, d$ and $j = 1, \dots, d$, we define the component (i, j) for the mixture of generalized gamma kernels as

$$h^{(i,j)}(t, \theta^{(i,j)}) = \sum_{k=1}^{K^{(i,j)}} \alpha_k^{(i,j)} \frac{p_k^{(i,j)} t^{(D_k^{(i,j)}-1)} \exp(-(t/\beta_k^{(i,j)})^{p_k^{(i,j)}})}{(\beta_k^{(i,j)})^{D_k^{(i,j)}} \Gamma(D_k^{(i,j)}/p_k^{(i,j)})}. \tag{17}$$

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

$$\theta = (\theta^{(i,j)})_{1 \leq i,j \leq d} = (\theta^{(1,1)}, \theta^{(1,2)}, \cdots, \theta^{(d,d-1)}, \theta^{(d,d)}) \tag{18}$$

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

For simplicity of exposition, we assume that each term in the sum of Equation (17) 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 material of [Potiron and Volkov, 2025].

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

$$w_p(t,P,s) = \mathbb{E}[\sup_{h \in [0, s \wedge (T-t)]} |P^*_{t+h} - P^*_t|^p | \mathcal{F}_t, P^*_t = P], \; s > 0. \tag{19}$$

We then define the global regularity modulus as

$$w_p(s) = \sup_{(t,P) \in [0,T] \times \Theta} w_p(t,P,s), \; s > 0. \tag{20}$$

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 now introduce a set of conditions required for the CLT of parameter integral.

Condition 2. (a) We assume that Θ is a convex, bounded and open space satisfying the conditions of the Sobolev embedding Theorem (see Theorem 4.12 (p. 85) in [Adams and Fournier, 2003]).

(b) For any $P = (\nu, \theta) \in \Theta$, we have that the kernel parameter θ is of the form (18) and the kernel $h(t, \theta)$ is of the form (17).

(c) There exists a positive real number $\nu_- > 0$ such that for any $P = (\nu, \theta) \in \Theta$ and any $i = 1, \cdots, d$, the i-th component of the baseline is bigger than ν_-, i.e. $\nu^{(i)} > \nu_-$.

(d) There exists a positive real number $p_- > 0$ such that for any $i = 1, \ldots, d$, any $j = 1, \ldots, d$ and any $k = 1, \ldots, K^{(i,j)}$ we have that $p_k^{(i,j)} > p_-$.

(e) There exists a positive real number $D_- > 0$ such that for any $i = 1, \ldots, d$, any $j = 1, \ldots, d$ and any $k = 1, \ldots, K^{(i,j)}$ we have that $D_k^{(i,j)} > D_-$.

(f) There exists a real number strictly between 0 and 1, i.e. $0 < r' < 1$, such that for any $P = (\nu, \theta) \in \Theta$ the spectral norm of the kernel matrix integral when the kernel parameter is time-invariant equal to θ is smaller than r', i.e. $\int_0^\infty h(s, \theta) ds \leq r'$.

(g) There exists a real number $\gamma \in (0, 1]$ such that we have $w_p(s) = O_{\mathbb{P}}(s^{\gamma p})$ when $s \to 0$, .

(h) We assume that there exists a real positive number $\delta > 0$ which satisfies

$$\frac{\Delta}{T} = n^{1/\delta - 1}. \tag{21}$$

(i) δ and γ satisfy the relation $\delta > 1 + \frac{1}{\gamma}$.

(j) δ and γ satisfy the relation $\frac{2\gamma}{2\gamma - 1} < \delta < 3$.

Condition 2 (b) restricts to Hawkes processes with mixture of generalized gamma kernels.

We denote $\xrightarrow{\mathcal{D}-s}$ as the $\mathcal{F}_t$-stable convergence. ξ is defined as an m-dimensional standard normal vector. We now state the CLT for estimation of parameter integral in the following theorem. It also provides feasible statistics induced by the CLT.

Theorem 1. *Under Conditions 1 and 2, we have the CLT and the feasible CLT*

$$\sqrt{n} C_{IP}^{-1/2}(T)(\widehat{IP}(T) - IP(T)) \xrightarrow{\mathcal{D}-s} \xi, \tag{22}$$

$$\sqrt{n} \widehat{C}_{IP}^{-1/2}(T)(\widehat{IP}(T) - IP(T)) \xrightarrow{\mathcal{D}-s} \xi. \tag{23}$$

5 Numerical study

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

6 Empirical application

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,

The code is available online at ???

Supplementary materials

Finally, all proofs of the theory can be found in Appendix **??**.

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.

[Breiman, 1992] Breiman, L. (1992). *Probability, volume 7 of Classics in Applied Mathematics.*

[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, 2003a] Daley, D. and Vere-Jones, D. (2003a). *An Introduction to the Theory of Point Processes*, volume 1. Springer-Verlag New York, 2 edition.

[Daley and Vere-Jones, 2003b] Daley, D. J. and Vere-Jones, D. (2003b). *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 für 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, 2011] Jacod, J. and Protter, P. E. (2011). *Discretization of processes*. Springer Science & Business Media.

[Jacod and Shiryaev, 2003a] Jacod, J. and Shiryaev, A. (2003a). *Limit theorems for stochastic processes*. Springer Berlin, Heidelberg, 2nd edition.

[Jacod and Shiryaev, 2003b] Jacod, J. and Shiryaev, A. (2003b). *Limit theorems for stochastic processes*. Berlin: Springer-Verlag, 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 time-dependent 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 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

Let us 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 supplement 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 $(N_t^{k,(i)}, \lambda_t^{k,(i)})_{k\geq 0}$ 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 λ_t as $\mathcal{F}_t$-intensity given by Equation (4).

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

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

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

$$\lambda_t^{k+1} = \nu_t^* + \int_0^t h(t-s, \theta_s^*)dN_s^k, \tag{24}$$
$$N_t^{k+1,(i)} = \int_{[0,t]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_s^{k+1,(i)}]}(x)\overline{N}^{(2i-1)} * \overline{N}^{(2i)}(ds \times dx).$$

First, we have that $\lambda^{k,(i)}$ is positive a.e. 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, λ_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 ρ_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) in the supplement of [Clinet and Potiron, 2018] when $\mathcal{G} = \mathcal{F}_T^{P^*}$, with Condition 1 (d) and Equation (25). The second equality is obtained by Tonelli's theorem and the definition of ρ_t^k. If we define Φ_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.

$$\Phi_t^{k+1} = \int_0^t \Big(\int_0^{t-s} h(u, \theta_s^*)du\Big)\rho_s^n ds. \tag{25}$$

Then, Condition 1 (f) implies that $||\Phi_t^{k+1}|| \leq r||\Phi_t^k||$ a.s.. Thus, we can deduce that $F : \Phi_t^k \to \Phi_t^{k+1}$ 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

$$\mathbb{E}\Big[\int_0^t \lambda_s ds \Big| \mathcal{F}_T^{P^*}\Big] \leq \int_0^t \nu_s^* ds + r\mathbb{E}\Big[\int_0^t \lambda_s ds \Big| \mathcal{F}_T^{P^*}\Big]. \tag{26}$$

By rearranging the terms in Expression (26), we get that

$$\mathbb{E}\Big[\int_0^t \lambda_s ds \Big| \mathcal{F}_T^{P^*}\Big] \leq (1-r)^{-1} \int_0^t \nu_s^* ds. \tag{27}$$

Given Condition 1 (c), the expression in the left side of Expression (27) is finite a.s.. Given that its conditional expectation is finite, $\int_0^t \lambda_s ds$ is finite a.s.. Moreover, λ_t is $\mathcal{F}_t$-predictable as a limit of such processes. $N_t^{(i)}$ counts the points of $\overline{N}^{(2i-1)} * \overline{N}^{(2i)}$ under the curve $t \mapsto \lambda_t^{(i)}$ by an application of the monotone convergence theorem. N_t therefore admits λ_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 λ_t is of the form (4). The monotonicity properties of $N_t^{k,(i)}$ and $\lambda_t^{k,(i)}$ ensure that, for any $k \geq 0$ and any $t \in [0,T]$, $\lambda_t^{k,(i)} \leq \nu_t^{*,(i)} + (\int_0^t h(t-s,\theta_s^*)\, dN_s)^{(i)}$ and $\lambda_t^{(i)} \geq \nu_t^{*,(i)} + (\int_0^t h(t-s,\theta_s^*)\, dN_s^k)^{(i)}$, which gives (4) by taking the limit $k \to +\infty$ in both inequalities. □

The following lemma states that K, C and Q admit limit values K_l, C_l and Q_l.

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

Proof of Lemma 1. the three time-averaged quantities Γ_T, K_T and C_T admit deterministic limiting values when $T \to \infty$ because the process N^P is exponentially mixing. Indeed, a slight generalization of Lemma 6.6 in [Clinet and Yoshida, 2017] shows that the vector process $(\lambda(t,\theta^*), \partial_\theta(t,\theta^*), \cdots, \partial_\theta^3(t,\theta^*))$ satisfies the mixing condition [M2] defined on p. 14 in the cited paper, which in turn implies the existence of $\Gamma(\theta^*) \in \mathbb{R}^{3\times3}$, and $K(\theta^*)$, $C(\theta^*) \in \mathbb{R}^{3\times3\times3}$ such that for any $\epsilon \in (0,1)$ and any integer $p \geq 1$,

$$\mathbb{E}|\Gamma_T(\theta^*) - \Gamma(\theta^*)|^p = O(T^{-\epsilon\frac{p}{2}}), \tag{28}$$

$$\mathbb{E}|K_T(\theta^*) - K(\theta^*)|^p = O(T^{-\epsilon\frac{p}{2}}), \tag{29}$$

and

$$\mathbb{E}|C_T(\theta^*) - C(\theta^*)|^p = O(T^{-\epsilon\frac{p}{2}}), \tag{30}$$

where $|x|$ stands for $\sum_i |x_i|$ for any vector or a matrix x. Moreover, it is also an easy consequence of the mixing property along with the fact that $M_T(\theta^*)$ is a martingale that we have the convergence

$$\mathbb{E}[Q_T(\theta^*) - Q(\theta^*)] = O(T^{-\frac{\epsilon}{2}}), \tag{31}$$

for some $Q(\theta^*) \in \mathbb{R}^{3\times3\times3}$. Note that $\Gamma(\theta^*)$ is the asymptotic Fisher information. In particular, in [Clinet and Yoshida, 2017] the authors have shown the convergence of moments of the MLE (see

Theorem 4.6),

$$\mathbb{E}[f(\sqrt{T}(\hat{\theta}_T - \theta^*))] \to \mathbb{E}[f(\Gamma(\theta^*)^{-\frac{1}{2}}\xi)], \tag{32}$$

where f can be any continuous function of polynomial growth, and ξ follows a standard normal distribution. Also, it is easy to see that the convergences in (28)-(32) hold uniformly in $\theta^* \in K$ under a mild change in the proofs of [Clinet and Yoshida, 2017]. □

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

For $i = 1, \cdots, d$, we denote by $\overline{\Lambda}^{(i)}$ the compensating measure of $\overline{N}^{(2i-1)} * \overline{N}^{(2i)}$, i.e. $\overline{\Lambda}^{(i)}(ds, dz) = ds \times dz$. For a predictable function W, we introduce $W * \overline{N}_t^{(i)} = \iint_{[0,t]\times\mathbb{R}} W(s,z)\overline{N}^{(i)}(ds, dz)$ and $W * \overline{\Lambda}_t^{(i)} = \iint_{[0,t]\times\mathbb{R}} W(s,z)\overline{\Lambda}^{(i)}(ds, dz)$. The following lemma is a straightforward adaptation of Lemma I.2.1.5 in [Jacod and Protter, 2011], using also Lemma **??** and (**??**).

Lemma 2. *Let W be a predictable function such that $W^2 * \overline{\Lambda}_t < \infty$ almost surely. Then for any integer $p > 1$, there exists a constant K_p such that*

$$\mathbb{E}[\sup_{t\in[0,T]} |W * (\overline{N} - \overline{\Lambda})_t|^p | \mathcal{F}_T^{\theta}]$$

$$\leq K_p \mathbb{E}[\iint_{[0,T]\times\mathbb{R}} |W(s,z)|^p ds dz + (\iint_{[0,T]\times\mathbb{R}} W(s,z)^2 ds dz)^{\frac{p}{2}} | \mathcal{F}_T^{\theta}]$$

For any random kernel $\chi : (s,t) \to \chi(s,t)$, χ 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 ensures the boundedness of moments of the locally stationary Hawkes process.

Lemma 3. *Under Condition 1, the locally stationary Hawkes process N has moments on $[0,T]$ that can be bounded by values independent from T. Moreover, for any $\mathcal{F}^P$-predictable kernel χ such that $\int_0^t \chi(s,t)ds$ is bounded uniformly in $t \in [0,T]$ independently from T, and for any predictable process ψ*

that has uniformly bounded moments independently from T, we have

$$\sup_{t\in[0,T]} \mathbb{E}[\lambda(t)^p|\mathcal{F}_T^\theta]^{\frac{1}{p}} < Q_p \tag{33}$$

$$\sup_{t\in[0,T]} \mathbb{E}[(\int_0^t \chi(s,t)dN_s)^p|\mathcal{F}_T^\theta]^{\frac{1}{p}} < Q_{p,\chi} \tag{34}$$

where the constants Q_p, $Q_{p,\chi}$ are independent from T.

Proof. We conduct the proof in three steps.

Step 1. We prove that **(i)** holds for $p = 1$. We write $\mathrm{E}[\lambda(t)|\mathcal{F}_T^\theta] = \nu_t + \int_0^{t-} a_s e^{-b_s(t-s)}\mathbb{E}[\lambda(s)|\mathcal{F}_T^\theta]ds$

$\leq \overline{\nu} + \sup_{s\in[0,t]} \mathbb{E}[\lambda(s)|\mathcal{F}_T^\theta] \int_0^{t-} a_s e^{-b_s(t-s)}ds$

$\leq \overline{\nu} + c\sup_{s\in[0,t]} \mathbb{E}[\lambda(s)|\mathcal{F}_T^\theta], whereweusedcondition(\textbf{??})atthelaststep.Takingthesupremumover$[0,T]$onbothsides, weg$

$(1-c)^{-1}\overline{\nu}.Inparticularthisprovesthecase$p=1, $sincetherighthandsideof(8)isindependentfrom$T.

Step 2. We prove that **(i)** holds for any integer $p > 1$. Note that it is sufficient to consider the case $p = 2^q$, $q > 0$. We thus prove our result by induction on q. The initialisation case $q = 0$ has been proved in Step 1. Note that for any $\epsilon > 0$, $\mathrm{E}[\lambda(t)^p|\mathcal{F}_T^\theta] \leq (1+\epsilon^{-1})^{2^q-1}\overline{\nu} + (1+\epsilon)^{2^q-1}\mathbb{E}[.(\int_0^{t-} a_s e^{-b_s(t-s)}dN_s)^p|\mathcal{F}_T^\theta],$

$wherewehaveusedtheinequality$(x+y)$^{2^q} \leq (1+\epsilon)^{2^q-1}x^{2^q} + (1+\epsilon^{-1})^{2^q-1}y^{2^q}$ for any $x, y, \epsilon > 0$. Now, for a fixed $t \in [0,T]$, define $W(s,z) = a_s e^{-b_s(t-s)}\mathbb{1}_{\{0\leq z\leq\lambda(s)\}}$, and note that Eł[ł.ł$(\int_0^{t-} a_s e^{-b_s(t-s)}dN_s)^p|\mathcal{F}_T^\theta] =$

$\mathbb{E}[.(W * \overline{N}_t)^p|\mathcal{F}_T^\theta]$

$\leq (1+\epsilon^{-1})^{2^q-1}\mathbb{E}[.(W * (\overline{N}-\overline{\Lambda})_t)^p|\mathcal{F}_T^\theta]$

$+ (1+\epsilon)^{2^q-1}\mathbb{E}[.(W * \overline{\Lambda}_t)^p|\mathcal{F}_T^\theta].$

$WeapplynowLemma2toget\mathbb{E}[(W*(\overline{N}-\overline{\Lambda})_t)^p|\mathcal{F}_T^\theta] \leq K_p\mathbb{E}[.\iint_{[0,T]\times\mathbb{R}}|W(s,z)|^p dsdz + (\iint_{[0,T]\times\mathbb{R}} W(s,z)^2 dsdz$

$= K_p\mathbb{E}[.\int_0^{t-} a_s^p e^{-pb_s(t-s)}\lambda(s)ds + (\int_0^{t-} a_s^2 e^{-2b_s(t-s)}\lambda(s)ds)^{\frac{p}{2}}|\mathcal{F}_T^\theta].Weeasilyboundthefirsttermbytheinductionhypothesi$

For the second term, an elementary application of Hölder's inequality shows that for any $k > 1$ and any non-negative functions f, g, $(\int fg)^k \leq (\int f^k g)(\int g)^{k-1}$. This along with the induction hypothesis leads to a similar bound for the second term. On the other hand, we have Eł[ł.ł(W $*\overline{\Lambda}_t)^p|\mathcal{F}_T^\theta] = \mathbb{E}[.(\int_0^{t-} a_s e^{-b_s(t-s)}\lambda(s)ds)^p|\mathcal{F}_T^\theta].$

$WeapplyagainthesameHlder'sinequalityasabovewithfunctions$f(s) $= \lambda(s)$ and $g(s) = a_s e^{-b_s(t-s)}$ to get $\mathrm{E}\big[\big|\big(\int_0^{t-} a_s e^{-b_s(t-s)}\lambda(s)ds\big)^p\big|\mathcal{F}_T^\theta\big] \leq c^{p-1}\mathbb{E}\big[\int_0^{t-} a_s e^{-b_s(t-s)}\lambda(s)^p ds\big|\mathcal{F}_T^\theta\big]$

$= c^{p-1}\int_0^{t-} a_s e^{-b_s(t-s)}\mathbb{E}[\lambda(s)^p|\mathcal{F}_T^\theta]ds$

$\leq c^p \sup_{s\in[0,t]}\mathbb{E}[\lambda(s)^p|\mathcal{F}_T^\theta]$ Finally, we have shown that

$\mathrm{E}[\lambda(t)^p|\mathcal{F}_T^\theta] \leq (1+\epsilon^{-1})^{2^q-1}\overline{\nu}+(1+\epsilon)^{2^q-1}(1+\epsilon^{-1})^{2^q-1}A_p+(1+\epsilon)^{2^q}c^p\sup_{s\in[0,t]}\mathbb{E}[\lambda(s)^p|\mathcal{F}_T^\theta]$. This yields, taking suprem

0 small enough so that $(1+\epsilon)^{2^q}c^p < 1$,

$\sup_{t\in[0,T]}\mathbb{E}[\lambda(t)^p|\mathcal{F}_T^\theta](1-(1+\epsilon)^{2^q}c^p) \leq (1+\epsilon^{-1})^{2^q-1}\overline{\nu}+(1+\epsilon)^{2^q-1}(1+\epsilon^{-1})^{2^q-1}A_p$, and dividing by $(1-(1+\epsilon)^{2^q}c^p)$ on both sides we get the result.

Step 3. It remains to show **(ii)** and **(iii)**. But note that they are direct consequences of the boundedness of moments of λ along with Lemma 2.

□

We focus on asymptotic properties of the local maximum likelihood estimator $\widehat{\Theta}_{i,n}$ of our model on each block $i \in \{1, \cdots, B_n\}$. Recall that we are given the global filtration $\mathcal{F}_t = \mathcal{F}_t^{(\theta^*,\overline{N})}$ that bears a sequence of doubly stochastic Hawkes processes $(N_t^n)_{t\in[0,T]}$. We perform maximum likelihood estimation on each time block $\big((i-1)\Delta_n T, i\Delta_n T\big]$, $i \in \{1, \cdots, B_n\}$ on the regression family of a parametric Hawkes process and show the local central limit theorem for every local estimator $\widehat{\Theta}_{i,n}$ of $\theta^*_{(i-1)\Delta_n}$, uniformly in the block index i. In addition, we show that all moments up to order $2\kappa > 2$ of the rescaled estimators $\sqrt{h_n}\big(\widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n}\big)$ are convergent uniformly in i.

Instead of deriving the limit theorems directly on each block, we show that by a well-chosen time change it is possible to reduce our statistical problem to a long-run framework. Such procedure is based on the following elementary lemma. Let $(N_t)_t$ be a point process adapted to a filtration $\mathcal{F}_t$, with $\mathcal{F}_t$-stochastic intensity $\lambda(t)$. For $\gamma > 0$, consider $N_t^\gamma = N_{\gamma t}$, which is adapted to $\mathcal{F}_t^\gamma = \mathcal{F}_{\gamma t}$. Then, N_t^γ admits $\lambda^\gamma(t) = \gamma\lambda(\gamma t)$ as a $\mathcal{F}_t^\gamma$-stochastic intensity. Moreover, if N_t is a doubly stochastic Hawkes process with parameter process $(\theta_s)_s$, N_t^γ has the distribution of a Hawkes process of parameter $(\gamma\theta_{\gamma s})_s$, that is, $\lambda^\gamma(t) = \gamma\nu_{\gamma t} + \int_0^{t-}\gamma a_{\gamma s}e^{-\gamma b_{\gamma s}(t-s)}dN_s^\gamma$.

Proof. First note that $N_t^\gamma = N_{\gamma t}$ is compensated by $\int_0^{\gamma t}\lambda(s)ds$. By a simple change of variable

$u = \gamma^{-1}s$ this integral can be written as $\int_0^t \gamma\lambda(\gamma u)du$ which proves the first part of the lemma. In the doubly stochastic Hawkes case, let us write the integral form of the time-changed intensity and apply once again the change of variable $u = \gamma^{-1}s$, $\lambda^\gamma(t) = \gamma\lambda(\gamma t)$

$= \gamma\nu_{\gamma t} + \int_0^{\gamma t-} \gamma a_s e^{-b_s(\gamma t - s)} dN_s$

$= \gamma\nu_{\gamma t} + \int_0^{t-} \gamma a_{\gamma u} e^{-\gamma b_{\gamma u}(t-u)} dN_u^\gamma,$

$and we are done.$ □

By virtue of Lemma 8, for any block index $i \in \{1, \cdots, B_n\}$, we consider the time change $\tau_i^n : t \mapsto n^{-1}t + (i-1)\Delta_n$ and the point process $(N_s^n)_{\{s\in((i-1)\Delta_n, i\Delta_n]\}}$ in order to get a time changed point process $N^{i,n}$ defined on the time set $[0, h_nT]$ by the formula $N_t^{i,n} = N_{\tau_i^n(t)}^n - N^n$. Such process is adapted to the filtration $\mathcal{F}_t^{i,n} = \mathcal{F}_{\tau_i^n(t)}$, for $t \in [0, h_nT]$. The parameter processes are now $(\theta_t^{i,n,*})_{\{t\in[0,h_nT]\}} = (\theta^*_{\tau_i^n(t)})_{\{t\in[0,h_nT]\}}$ whose canonical filtration can be expressed as $\mathcal{F}_t^{\theta^{i,n,*}} = \sigma\{\theta_s^{i,n,*} | 0 \le s \le t\}$, for $t \in [0, h_nT]$. Finally note that the $\mathcal{F}_t^{i,n}$-stochastic intensities are now of the form $\lambda_*^{i,n}(t) = \nu_t^{i,n,*} + \int_0^{t-} a_s^{i,n,*} e^{-b_s^{i,n,*}(t-s)} dN_s^{i,n} + R_{i,n}(t), where \mathrm{R}_{i,n}(t)$ is the $\mathcal{F}_0^{i,n}$-measurable residual process defined by the relation $\mathrm{R}_{i,n}(t) = \int_0^{(i-1)\Delta_n -} n a_s^* e^{-nb_s^*(\tau_i^n(t)-s)} dN_s^n . \mathrm{R}_{i,n}(t)$ should be interpreted as the pre-excitation induced by the preceding blocks. Note that in view of the exponential form of the kernel $\phi_t = ae^{-bt}$ assumption, $R_{i,n}(t)$ can be bounded by $\mathrm{R}_{i,n}(t) \le e^{-\underline{b}t} R_{i,n}(0) Note that all the processes \mathrm{N}^{i,n}$ can be represented as integrals over a sequence of Poisson processes $\overline{N}^{i,n}$ of intensity 1 on $\mathbb{R}^2$ as follows: $\mathrm{N}_t^{i,n} = \iint_{[0,t]\times\mathbb{R}_+} \mathbb{1}_{\{0\le z\le\lambda_*^{i,n}(s)\}} \overline{N}^{i,n}(ds, dz). Indeed, \overline{N}^{i,n}$ is the time-space changed version of the initial Poisson process $\overline{N}$ defined by $\overline{N}^{i,n}(A \times B) = \overline{N}(\tau_i^n(A) \times nB\Sigma)$ for A and B any two Borel sets of $\mathbb{R}$. In the time-changed representation, we define the regression family of stochastic intensities ${}^{i,n}(t, \theta) = \nu + \int_0^{t-} ae^{-b(t-s)} dN_s^{i,n}, which is related to \lambda^{i,n}$ (see (**??**)) by $\tilde{\lambda}^{i,n}(t, \theta) = n^{-1}\lambda^{i,n}(\tau_i^n(t), \theta)$. Also, the Quasi Log Likelihood process defined in (**??**) on the i-th block has now the representation (up to the constant term $\log(n)N_{h_nT}^{i,n}$) $\mathrm{l}_{i,n}(\theta) = \int_0^{h_nT} \log(\tilde{\lambda}^{i,n}(t, \theta)) dN_t^{i,n} - \int_0^{h_nT} \tilde{\lambda}^{i,n}(t, \theta)dt, Note that in our case, the true underlying intensit$ does not belong to the regression family $(\tilde{\lambda}^{i,n}(., \theta))_{\theta\in K}$ for two reasons : the parameter process θ^* is not constant on the i-th block, and the regression family does not take into account the existence of a pre-excitation term in (8). We are in a mispecified case, but we wish to take advantage of the continuity of the process θ^* to show that the asymptotic theory still holds, that is, the MLE tends to the value $\theta_0^{i,n,*} = \theta^*_{(i-1)\Delta_n}$ which is the value of the process θ^* at the beginning of the i-th block. The procedure is thus asymptotically equivalent to performing the MLE on the model whose stochastic intensity is in the regression family with true value $\theta = \theta^*_{(i-1)\Delta_n}$. To formalize such idea, we introduce an auxiliary model corresponding to the parametric case generated by the true value $\theta^*_{(i-1)\Delta_n}$. More precisely, we introduce the constant parameter Hawkes process $N^{i,n,c}$ generated by $\overline{N}^{i,n}$ and the initial value $\theta_0^{i,n,*}$, whose stochastic intensity satisfies $\lambda^{i,n,c}(t) = \nu_0^{i,n,*} + \int_0^{t-} a_0^{i,n,*} e^{-b_0^{i,n,*}(t-s)} dN_s^{i,n,c} . Moreover, we assume that \mathrm{N}_t^{i,n,c}$ has the representation $\mathrm{N}_t^{i,n,c} = \iint_{[0,t]\times\mathbb{R}_+} \mathbb{1}_{\{0\le z\le\lambda^{i,n,c}(s)\}} \overline{N}^{i,}$ is unobserved and just used as an intermediary to derive the asymptotic properties of the MLE, by showing systematically that any variable $N^{i,n}$, $\tilde{\lambda}^{i,n}$, $l_{i,n}$, etc. is asymptotically very close to its counterpart that is generated by the constant parameter model.

For reasons that will become apparent later, it is crucial to localize the pre-excitation $R_{i,n}(0)$ and bound it by some deterministic value M_n that depends solely on n and such that $M_n = O(n^q)$ for some $q > 1$. To reduce our local estimation problem to the case of a parametric Hawkes process, we will also need to condition with respect to the initial value of the parameter process. We will thus use extensively the conditional expectations $\mathbb{E}[.\mathbf{1}_{\{R_{i,n}(0)\le M_n\}}|\mathcal{F}_0^{i,n}, \theta_0^{i,n,*} = \theta_0]$, that we denote by , and whose existences are justified by a classical regular distribution argument[1] (see for instance Section 4.3 (pp. $77-80$) in [Breiman, 1992]). In the same spirit, for a measurable set $A \in \mathcal{F}$, $[A]$ should be understood as $[\mathbf{1}_A]$. Finally we will need frequently to take supremum over the quadruplet (θ_0, i, n, t). For that reason we introduce the notation $= \{.(\theta_0, i, n, t) \in K \times \mathbb{N}^2 \times \mathbb{R}_+ | 1 \le i \le B_n$ and $0 \le t \le h_n T\}$. When $n \in \mathbb{N}$ is fixed, we define $_n$ the subset of as $_n = \{.(\theta_0, i, t) \in K \times \mathbb{N} \times \mathbb{R}_+ | 1 \le i \le B_n$ and $0 \le t \le h_n T\}$. In the same spirit, it is also useful when truncation arguments appear, to consider in the previous equation the subset of $_n$ for which we have the stronger condition $h_n^\alpha T \le t \le h_n T$ where $\alpha \in (0,1)$ that we denote by $^\alpha_n$. The next lemma states the uniform boundedness of the moments of $\lambda_*^{i,n}$ and $\lambda^{i,n,c}$, along with $\mathbb{L}^p$ estimates for stochastic integrals over $N^{i,n}$ and $N^{i,n,c}$.

[1]This is a consequence to the fact that $K \subset \mathbb{R}^3$ is a Borel space.

We have, for any integer $p \geq 1$ and any $^{\theta^{i,n,*}}$-predictable kernel χ such that $\int_0^t \chi(s,t)ds$ is bounded uniformly in $t \in [0, h_n T]$ independently from T and n, [**(i)**] $\sup_{(\theta_0,i,n,t)\in} |\lambda_*^{i,n}(t)|^p \leq M_p$ $\mathbb{P}$-a.s. [**(ii)**] $\sup_{(\theta_0,i,n,t)\in} |\int_0^t \chi(s,t)dN_s^{i,n}|^p < M_{p,\chi}$ $\mathbb{P}$-a.s. [**(iii)**] $\sup_{(\theta_0,i,n,t)\in} |\lambda^{i,n,c}(t)|^p < M_p$ $\mathbb{P}$-a.s. [**(iv)**] $\sup_{(\theta_0,i,n,t)\in} |\int_0^t \chi(s,t)dN_s^{i,}$ $M_{p,\chi}$ $\mathbb{P}$-a.s.

where M_p and $M_{p,\chi}$ are finite constants depending respectively solely on p and on p and χ.

Proof. This is a straightforward adaptation of the proof of Lemma 3, with the conditional expectation $\mathbb{E}[.\mathbf{1}_{\{R_{i,n}(0)\leq M_n\}}|\mathcal{F}_0^{i,n} \vee \mathcal{F}_{h_n T}^{\theta^{i,n,*}}, \theta_0^{i,n,*} = \theta_0]$. The presence of $\mathbf{1}_{\{R_{i,n}(0)\leq M_n\}}$ along with the exponential decay in (8) show clearly that the result still holds, uniformly in the quadruplet (θ_0, i, n, t). By an immediate application of Jensen's inequality, this is still true replacing $\mathcal{F}_0^{i,n} \vee \mathcal{F}_{h_n T}^{\theta^{i,n,*}}$ by the smaller filtration $\mathcal{F}_0^{i,n}$, that is, for the operator . □

Before we turn to estimating the distance between the two models, we state a technical lemma. Let $h : s \mapsto ae^{-bs}$, and let f,g be two non-negative functions satisfying the inequality $f \leq g + f * h$ where $(f * h)(t) = \int_0^t f(t-s)h(s)ds$ is the usual convolution. Then we have the majoration for any $t \geq 0$

$$f(t) \leq g(t) + a(g * e^{(a-b).})(t)$$

Proof. Iterating the inequality we get for any $n \in \mathbb{N}^*$ f $\leq g + g * \sum_{k=1}^n h^{*(k)} + f * h^{*(n+1)}.We fix$t ≥ 0, and note that by a straightforward computation, for any integer $k \geq 1$ we have $h^{*(k)}(t) = \frac{t^{k-1}}{(k-1)!}a^k e^{-bt}$. We deduce that f $*h^{*(n+1)}(t) = \int_0^t f(t-s)\frac{s^n}{n!}a^{n+1}e^{-bs}ds$

$\leq \frac{t^n}{n!}a^{n+1}\int_0^t f(s)ds \to 0 as$n $\to +\infty$. We also have for any integer $n \geq 1$ $\sum_{k=1}^n h^{*(k)}(t) = \sum_{k=1}^n \frac{t^{k-1}}{(k-1)!}a^k e^{-bt}$

$\leq ae^{(a-b)t} and thus we get the result by taking the limit$n $\to +\infty$ in (8) evaluated at any point $t \geq 0$. □

In what follows, we quantify the local error between the doubly stochastic model and its constant parameter approximation. We recall the value of the key exponent $\kappa = \gamma(\delta - 1)$ that has been introduced in (**??**), and which plays an important role in the next results as it proves to be the rate of convergence of one model to the other in power of h_n^{-1}, where h_n is proportional to the typical size of one block after our time change. Recall that γ represents the regularity exponent in time of θ while δ controls the size of small blocks compared to n by the relation $h_n = n^{1/\delta}$. Note that by (**??**) we have $\kappa > 1$. The next lemma shows that the models $(N^{i,n,c}, \lambda^{i,n,c})$ and $(N^{i,n}, \lambda_*^{i,n})$ are asymptotically close in the $\mathbb{L}^p$ sense. The proof follows the same path as the proof of Lemma 3.

Let $\alpha \in (0,1)$ be a truncation exponent, and $\epsilon \in (0,1)$. We have, for any $p \geq 1$, any deterministic kernel χ 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 : [**(i)**] $\sup_{(\theta_0,i,t)\in_n^\alpha} |\lambda^{i,n,c}(t) - \lambda_*^{i,n}(t)|^p = O_{\mathbb{P}}(h_n^{-\kappa})$ [**(ii)**] $\sup_{\theta_0\in K, 1\leq i\leq B_n} | \int_{h_n^\alpha T}^{h_nT} \psi_s\{dN_s^{i,n,c} - dN_s^{i,n}\}|^p = O_{\mathbb{P}}(h_n^{p-\epsilon\kappa})$ [**(iii)**] $\sup_{\theta_0\in K, 1\leq i\leq B_n} | \int_{h_n^\alpha T}^{h_nT} \chi(s,h_nT)\{dN_s^{i,n,c} - dN_s^{i,n}\}|^p = O_{\mathbb{P}}(h_n^{-\kappa})$

For $p = 1$, if we recall that $\Delta_n = h_n n^{-1}T$ and $\kappa = \gamma(\delta - 1)$, we get a typical deviation in $h_n^{-\kappa} = T^{-\gamma}\Delta_n^\gamma$ between the real model and its constant parameter approximation. This is not very surprising since on one block the parameter process θ^* has exactly a deviation of that order. For $p > 1$, the situation is fairly different. One would expect a deviation of the same order of that of the parameter process, that is of order $h_n^{-\kappa p} = T^{-\gamma p}\Delta_n^{\gamma p}$. But as it is shown in the previous lemma, deviations between the two models are quite weaker since the deviation remains of order $h_n^{-\kappa} = T^{-\gamma}\Delta_n^\gamma$ for any p. This loss is due to the point process structure and the shape of its related Burkholder-Davis-Gundy type inequality (see Lemma 2). This is the same phenomenon as in the following fact. For a Poisson process N of intensity λ, we have $\mathbb{E}[|N_t - \lambda t|^p] \sim \alpha_p t$ when $t \to 0$, i.e. a rate of convergence which is linear regardless of the moment chosen.

Proof. We will show by recurrence on $q \in \mathbb{N}$ that for every p of the form $p = 2^q$, we have the majoration for $n \in \mathbb{N}$, $t \in [0, h_nT]$ and uniformly in (θ_0, i), ł$|\lambda^{i,n,c}(t) - \lambda_*^{i,n}(t)|^{2^q} \leq L_{n,q} + M_{n,q}e^{-\underline{b}(1-r)t}, where \mathrm{L}_{n,q}$ and $M_{n,q}$ depend on n and q only, $L_{n,q} = O_{\mathbb{P}}(h_n^{-\kappa})$, and $M_{n,q}$ is of polynomial growth in n. Note that then **(i)** will be automatically proved since by taking the supremum over the set $[h_n^\alpha T, h_nT]$ and using the estimate $M_{n,q}e^{-\underline{b}(1-r)h_n^\alpha T} = o_{\mathbb{P}}(h_n^{-\kappa})$ we get

$$|\lambda^{n,c}(t) - \lambda^n_*(t)|^p = O_{\mathbb{P}}(h_n^{-\kappa})$$

uniformly over the set $^{\alpha}_{n}$.

Step 1. We show our claim in the case $q = 0$, that is $p = 1$. Write $|\lambda^{i,n}_*(t) - \lambda^{i,n,c}(t)| \leq$
$|\nu^{i,n,*}_t - \nu^{i,n,*}_0| + |\int_0^{t-} (a^{i,n,*}_s e^{-b^{i,n,*}_s(t-s)} - a^{i,n,*}_0 e^{-b^{i,n,*}_0(t-s)}) dN^{i,n}_s|$
$+ |\int_0^{t-} a^{i,n,*}_0 e^{-b^{i,n,*}_0(t-s)}(dN^{i,n,c}_s - dN^{i,n}_s)| + R_{i,n}(t)$
$\leq A_{i,n}(t) + B_{i,n}(t) + C_{i,n}(t) + R_{i,n}(t)$

The (uniform) majoration $A_{i,n}(t) = O_{\mathbb{P}}(h_n^{-\kappa})$ is an immediate consequence of [C]-**(i)**. By the inequality $|\mathrm{a}e^{-bt} - a' e^{-b' t}| \leq (|a - a'| + |b - b'|)e^{-\underline{b}t} for any (\nu, a, b), (\nu', a', b') \in K$, we can write
$\mathrm{B}_{i,n}(t) \leq \int_0^{t-} (|a^{i,n,*}_s - a_0| + |b^{i,n,*}_s - b_0|)e^{-\underline{b}(t-s)} dN^{i,n}_s$
$\leq \sqrt{|\sup_{s\in[0,t]} (|a^{i,n,*}_s - a_0| + |b^{i,n,*}_s - b_0|)^2| \int_0^{t-} e^{-\underline{b}(t-s)} dN^{i,n}_s|^2}, where Cauchy-Schwartz inequality was applied in the$
$\mathrm{B}_{i,n}(t) = O_{\mathbb{P}}(h_n^{-\kappa})$ follows from [C]-**(i)**. Finally, for $C_{i,n}(t)$, write $\mathrm{C}_{i,n}(t) \leq \int_0^{t-} a_0 e^{-b_0(t-s)} d|N^{i,n,c} - N^{i,n}|_s where d|\mathrm{N}^{i,n}$
$N^{i,n}|_s$ is the integer measure which counts the jumps that don't belong to both $dN^{i,n,c}$ and $dN^{i,n}$, i.e. the points of $\overline{N}^{i,n}$ that lay between the curves $t \to \lambda^{i,n}_*(t)$ and $t \to \lambda^{i,n,c}(t)$. A short calculation shows that this counting process admits $|\lambda^{i,n,c}(s) - \lambda^{i,n}_*(s)|$ as stochastic intensity. We compute now:
$\mathrm{C}_{i,n}(t) \leq \int_0^{t-} a_0 e^{-b_0(t-s)}|\lambda^{i,n,c}(s) - \lambda^{i,n}_*(s)|ds$
$= \int_0^{t-} a_0 e^{-b_0(t-s)}|\lambda^{i,n,c}(s) - \lambda^{i,n}_*(s)|ds.$
$So far we have shown that there exists a sequence \mathrm{L}_n$ such that $L_n = O(h_n^{-\kappa})$ and such that the function $f(t) = |\lambda^{i,n,c}(t) - \lambda^{i,n}_*(t)|$ satisfies the inequality $\mathrm{f(t)} \leq L_n + R_{i,n}(t) + f * h(t), where h is the kernel defined as$ h:
$\mathrm{t} \mapsto a_0 e^{-b_0 t}$. By Lemma 8, this yields $\mathrm{f(t)} \leq L_n + R_{i,n}(t) + \int_0^t \{L_n + R_{i,n}(s)\} a_0 e^{(a_0 - b_0)(t-s)} ds. Now recall that \mathrm{b}_0 -$
$a_0 > \underline{b}(1 - r)$ and that on the set $\{R_{i,n}(0) \leq M_n\}$, we have $R_{i,n}(s) \leq M_n e^{-\underline{b}s} < M_n e^{-\underline{b}(1-r)s}$ to get
$\mathrm{f(t)} \leq (1 + (1 - r)^{-1})L_n + R_{i,n}(t) + \int_0^t \{M_n e^{-\underline{b}(1-r)s}\} a_0 e^{\underline{b}(1-r)(t-s)} ds$
$\leq (1 + (1 - r)^{-1})L_n + (1 + \overline{a}t)M_n e^{-\underline{b}(1-r)t}. If we recall that in the above expression \mathrm{f(t)} stands for |\lambda^{i,n,c}(t) -$
$\lambda^{i,n}_*(t)|$, such uniform estimate clearly proves (8) in the case $q = 1$.

Step 2. We prove the result for any $q \in \mathbb{N}^*$. Let the expression $f(t)$ stands for $|\lambda^{i,n,c}(t) - \lambda^{i,n}_*(t)|^p$. With similar notations as for the previous step, we have for any $\eta > 0$ $\mathrm{f(t)} = |\lambda^{i,n,c}(t) - \lambda^{i,n}_*(t)|^p \leq$

$|A_{i,n}(t) + B_{i,n}(t) + C_{i,n}(t) + R_{i,n}(t)|^p$

$\leq (1+\eta^{-1})^{2^q-1}|A_{i,n}(t) + B_{i,n}(t) + R_{i,n}(t)|^p$

$+(1+\eta)^{2^q-1}C_{i,n}(t)^p$ *Itisstraightforwardtoseethatsimilarargumentstothepreviouscaseleadtotheuniformestimate* A_i

$B_{i,n}(t)^p = O_{\mathbb{P}}(h_n^{-\kappa p})$. *Now, define* W(s,z)= $a_0 e^{-b_0(t-s)}|\mathbb{1}_{\{0\leq z\leq \lambda^{i,n,c}(s)\}} - \mathbb{1}_{\{0\leq z\leq \lambda_*^{i,n}(s)\}}|$ to get $[C_{i,n}(t)^p] =$

$[(W * \overline{N}_t)^p]$

$\leq (1+\eta^{-1})^{2^q-1}[(W*(\overline{N}-\overline{\Lambda})_t)^p] + (1+\eta)^{2^q-1}[(W*\overline{\Lambda}_t)^p],$

andapplyLemma2toget $[(W*(\overline{N}-\overline{\Lambda})_t)^p] \leq K_p([\iint_{[0,T]\times\mathbb{R}} |W(s,z)|^p ds dz]$

$+ .[(\iint_{[0,T]\times\mathbb{R}} W(s,z)^2 ds dz)^{\frac{p}{2}}])$

$= K_p([\int_0^{t-} a_0^p e^{-pb_0(t-s)}|\lambda^{i,n,c}(s) - \lambda_*^{i,n}(s)|ds]$

$+.[(\int_0^{t-} a_0^2 e^{-2b_0(t-s)}|\lambda^{i,n,c}(s) - \lambda_*^{i,n}(s)|ds)^{\frac{p}{2}}])$, *whichiseasilyboundedasin*(8)*usingtheinductionhypothesis.Notethathe*

$\lambda_*^{i,n}(s)|$ is the major obstacle to getting the stronger estimate $O_{\mathbb{P}}(h_n^{-\kappa p})$ that one would expect. Finally the term

$$[(W * \overline{\Lambda}_t)^p] = [(\int_0^{t-} a_0 e^{-b_0(t-s)}|\lambda^{i,n,c}(s) - \lambda_*^{i,n}(s)|ds)^p]$$

is treated exactly in the same way as for the proof of Lemma 3, to get the bound [(W $*\overline{\Lambda}_t)^p] \leq$ $c_q f * h(t)$, *whereagain* h : s $\mapsto a_0 e^{-b_0 s}$, and $c_q < 1$ if η is taken small enough. We have thus shown that f satisfies a similar convolution inequality as for the case $q = 1$ and we can apply Lemma 8 to conclude.

Step 3. It remains to show **(ii)** and **(iii)**. They are just consequences of the application of Lemma 2 to the case $W_\psi(s,z) = \psi_s|\mathbb{1}_{\{0\leq z\leq \lambda^{n,c}(s)\}} - \mathbb{1}_{\{0\leq z\leq \lambda_*^n(s)\}}|$ and $W_\chi(s,z) = \chi(s,t)|\mathbb{1}_{\{0\leq z\leq \lambda^{n,c}(s)\}} - \mathbb{1}_{\{0\leq z\leq \lambda_*^n(s)\}}|$ along with Hölder's inequality. □

We are now ready to show the uniform asymptotic normality of the MLE by proving that any quantity related to the estimation is asymptotically very close to its counterpart for the constant parameter model $(N^{i,n,c}, \lambda^{i,n,c})$. To this end we introduce the fake candidate intensity family and the fake log-likelihood process, as $\lambda^{i,n,c}(t,\theta) = \nu + \int_0^{t-} a e^{-b(t-s)} dN_s^{i,n,c}$ *and* $l_{i,n}^c(\theta) = \int_0^{h_n T} \log(\lambda^{i,n,c}(t,\theta)) dN_t^{i,n,c} -$

$\int_0^{h_nT} \lambda^{i,n,c}(t,\theta)dt, for any \theta \in K$. Note that $\lambda^{i,n,c}(t,\theta_0^{i,n,*}) = \lambda^{i,n,c}(t)$ by definition. Those quantities, which are all related to $(N^{i,n,c}, \lambda^{i,n,c})$, are unobserved.

As a consequence of the previous lemma we state the uniform $\mathbb{L}^p$ boundedness of the candidate intensity families, along with estimates of their relative deviations. Let $\alpha \in (0,1)$. We have for any integer $p \geq 1$ and any $j \in \mathbb{N}$ that

[**(i)**] $\sup_{(\theta_0,i,n,t)\in} \sup_{\theta\in K} |\partial_\theta^j \tilde{\lambda}^{i,n}(t,\theta)|^p \leq K_j$ $\mathbb{P}$-a.s. [**(ii)**] $\sup_{(\theta_0,i,n,t)\in} \sup_{\theta\in K} |\partial_\theta^j \lambda^{i,n,c}(t,\theta)|^p \leq K_j$ $\mathbb{P}$-a.s. [**(iii)**] $\sup_{(\theta_0,i,t)\in_n^\alpha} \sup_{\theta\in K} |\partial_\theta^j \tilde{\lambda}^{i,n}(t,\theta) - \partial_\theta^j \lambda^{i,n,c}(t,\theta)|^p = O_{\mathbb{P}}(h_n^{-\kappa})$

where the constants K_j depend solely on j.

Proof. Note that the derivatives of $\tilde{\lambda}^{i,n}(t,\theta)$ can be all bounded uniformly in θ by linear combinations of terms of the form $\overline{\nu}$ or $\int_0^{t-} (t-s)^j e^{-\underline{b}(t-s)} dN_s^{i,n}$, $j \in \mathbb{N}$. The boundedness of moments of those terms uniformly in $n \in \mathbb{N}$ and in the time interval $[0, h_nT]$ is thus the consequence of Lemma 3 **(ii)** with $\chi(s,t) = (t-s)^j e^{-\underline{b}(t-s)}$, and consequently **(i)** follows. **(ii)** is proved in the same way. Finally we show **(iii)**. Note that $\sup_{\theta\in K} |\partial_\theta^j \tilde{\lambda}^{i,n}(t,\theta) - \partial_\theta^j \lambda^{i,n,c}(t,\theta)|$ can be bounded by linear combinations of terms of the form $\int_0^{t-} (t-s)^j e^{-\underline{b}(t-s)} d|N^{i,n} - N^{i,n,c}|_s$. The $\mathbb{L}^p$ estimate of such expression is then easily derived by a truncation argument and Lemma 8 **(iii)**. □

We now follow similar notations to the ones introduced in [?], and consider the main quantities of interest to derive the properties of the MLE. We define for any $(\theta,\theta_0) \in K^2$, $\mathrm{Y}_{i,n}(\theta,\theta_0) = \frac{1}{h_nT}(l_{i,n}(\theta) - l_{i,n}(\theta_0))$, $\Delta_{i,n}(\theta_0) = \frac{1}{\sqrt{h_nT}}\partial_\theta l_{i,n}(\theta_0), and finally \Gamma_{i,n}(\theta_0) = -\frac{1}{h_nT}\partial_\theta^2 l_{i,n}(\theta_0). We define in the same way \mathrm{Y}_{i,n}^c$, $\Delta_{i,n}^c$, and $\Gamma_{i,n}^c$. We introduce for the next lemma the set $= \{(\theta_0,i,n) \in K \times \mathbb{N}^2 | 1 \leq i \leq B_n\}$.

Let $\epsilon \in (0,1)$, and $L \in (0,2\kappa)$. For any $p \in \mathbb{N}^*$, for any $\epsilon \in (0,1)$, we have the estimates $\sup_{\theta_0\in K, 1\leq i\leq B_n} |\Delta_{i,n}(\theta_0) - \Delta_{i,n}^c(\theta_0)|^L \to^{\mathbb{P}} 0, \sup_{\theta_0\in K, 1\leq i\leq B_n} [\sup_{\theta\in K} |\mathbb{Y}_{i,n}(\theta,\theta_0) - \mathbb{Y}_{i,n}^c(\theta,\theta_0)|^p] = O_{\mathbb{P}}(h_n^{-\epsilon\kappa}), \sup_{\theta_0\in K, 1\leq i\leq}$ $\Gamma_{i,n}^c(\theta_0)|^p = O_{\mathbb{P}}(h_n^{-\epsilon\kappa}), \sup_{(\theta_0,i,n)\in} \left|h_n^{-1} \sup_{\theta\in K} |\partial_\theta^3 l_{i,n}(\theta)|\right|^p < K$ $\mathbb{P}$-a.s.

Proof. Let us show (8). We can express the equation in (8) and its counterpart for the constant model as $\Delta_{i,n}(\theta_0) = \frac{1}{\sqrt{h_nT}}\{\int_0^{h_nT} \frac{\partial_\theta\tilde{\lambda}^{i,n}(s,\theta_0)}{\tilde{\lambda}^{i,n}(s,\theta_0)} dN_s^{i,n} - \int_0^{h_nT} \partial_\theta\tilde{\lambda}^{i,n}(s,\theta_0)ds\} and \Delta_{i,n}^c(\theta_0) = \frac{1}{\sqrt{h_nT}}\{\int_0^{h_nT} \frac{\partial_\theta\lambda^{i,n,c}(s,\theta_0)}{\lambda^{i,n,c}(s,\theta_0)} dN_s^{i,n,c} -$

$\int_0^{h_nT} \partial_\theta \lambda^{i,n,c}(s,\theta_0)ds\Big\}. By Lemma 8(\mathbf{i}) and (\mathbf{iii}), and Lemma 8(\mathbf{i}) and (\mathbf{ii}) and the presence of the factor 1\frac{}{\sqrt{h_nT}}$,

it is possible to replace the lower bounds of those integrals by $h_n^\alpha T$ for some $\alpha \in (0, \frac{1}{2})$. Thus the difference $\sqrt{h_nT}(\Delta_{i,n}(\theta_0) - \Delta_{i,n}^c(\theta_0))$ is equivalent to the sum of the three terms

$\int_{h_n^\alpha T}^{h_nT} \frac{\partial_\theta \tilde{\lambda}^{i,n}(s,\theta_0)}{\tilde{\lambda}^{i,n}(s,\theta_0)}(dN_s^{i,n} - dN_s^{i,n,c}) + \int_{h_n^\alpha T}^{h_nT} \{\frac{\partial_\theta \tilde{\lambda}^{i,n}(s,\theta_0)}{\tilde{\lambda}^{i,n}(s,\theta_0)} - \frac{\partial_\theta \lambda^{i,n,c}(s,\theta_0)}{\lambda^{i,n,c}(s,\theta_0)}\Big\}dN_s^{i,n,c}$

$+\int_{h_n^\alpha T}^{h_nT} \{\partial_\theta \tilde{\lambda}^{i,n}(s,\theta_0) - \partial_\theta \lambda^{i,n,c}(s,\theta_0)\}ds. We therefore apply Lemmas 8(\mathbf{ii}) and 8(\mathbf{i}) to the first term, Lemmas 8(\mathbf{iii}) and 8(\mathbf{i}$

$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} |\Delta_{i,n}(\theta_0) - \Delta_{i,n}^c(\theta_0)\Big|^L = O_{\mathbb{P}}(h_n^{\frac{L}{2}-\epsilon\kappa}\Big), for any \epsilon \in (0,1)$. This tends to 0 if we can find an ϵ such that $\frac{L}{2} - \epsilon\kappa < 0$, and this can be done by taking ϵ sufficiently close to 1 since $L < 2\kappa$. Equations (8), (8) and (8) are proved similarly. □

For any integer $p \geq 1$, there exists a constant M such that $\sup_{(\theta_0,i,n)\in} |\Delta_n^c(\theta_0)\Big|^p < M$ $\mathbb{P}$-a.s.$Furthermore, there exis$ $\mathbb{Y}(\theta,\theta_0)$ such that for any $\epsilon \in (0,1)$, $\sup_{\theta_0 \in K, 1 \leq i \leq B_n} [\sup_{\theta \in K} |\mathbb{Y}_{i,n}^c(\theta,\theta_0) - \mathbb{Y}(\theta,\theta_0)\Big|= O(h_n^{-\epsilon\frac{p}{2}}\Big)$ $\mathbb{P}$-a.s.$Finally, for any \theta_0 \in$ K, and for any $\epsilon \in (0,1)$,

$$\sup_{\theta_0 \in K, 1 \leq i \leq \Delta_n^{-1}} |\Gamma_{i,n}^c(\theta_0) - \Gamma(\theta_0)\Big|^p = O(h_n^{-\epsilon\frac{p}{2}}\Big) \ \mathbb{P}\text{-a.s.} \tag{35}$$

where $\Gamma(\theta_0)$ is the asymptotic Fisher information matrix of the parametric Hawkes process regression model with parameter θ_0 as introduced in (28).

Proof. Note that when $\theta_0^{i,n,*} = \theta_0$, the constant model $N^{i,n,c}$ is simply a parametric Hawkes process with parameter θ_0, and is independent of the filtration $\mathcal{F}_0^{i,n}$. Thus, by a regular distribution argument the operator acts as the simple operator $\mathbb{E}$ for $N^{i,n,c}$ distributed as a Hawkes with true value θ_0. It is straightforward to see that under a mild change in the proofs of Lemma 3.15 and Theorem 4.6 in [?] those estimates hold uniformly in $\theta_0 \in K$ and in the block index. □

Let $L \in (0, 2\kappa)$. We have $\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \{\Big[f(\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta_0)\Big)\Big] - \mathbb{E}[f(T^{-\frac{1}{2}}\Gamma(\theta_0)^{-\frac{1}{2}}\xi)\Big]\Big\} \to^{\mathbb{P}} 0, for any continuous fu$ = O(|x|^L) when $|x| \to \infty$, and such that ξ follows a standard normal distribution.

Proof. By (8) and (8), we can define some number $\epsilon \in (0,1)$ such that

$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} h_n^{\epsilon(\frac{p}{2}\wedge\kappa)}[\sup_{\theta \in K} |\mathbb{Y}_{i,n}(\theta,\theta_0) - \mathbb{Y}(\theta,\theta_0)|^p\Big] \to^{\mathbb{P}} 0, and as \widehat{\Theta}_{i,n}$ is also a maximizer of $\theta \to \mathbb{Y}_{i,n}(\theta,\theta_0)$, (8) implies the uniform consistency in the block index i and the initial value of $\widehat{\Theta}_{i,n}$ to $\theta_0^{i,n,*}$, i.e. $\sup_{\theta_0 \in K, 1 \le i \le B_n} [\widehat{\Theta}_{i,n} - \theta_0] \to^{\mathbb{P}} 0, since Y satisfies the non-degeneracy condition [A4] in [?]. From (8) and (35) we deduc$

$\Gamma(\theta_0)|^p \to^{\mathbb{P}} 0.$

By (8), $\Delta_{i,n}(\theta_0)$ and $\Delta^c_{i,n}(\theta_0)$ have the same asymptotic distribution, which is of the form $\Gamma(\theta_0)^{\frac{1}{2}}\xi$, where ξ follows a standard normal distribution. Following the proof of Theorem 3.11 in [?], we deduce that $\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta_0)$ converges uniformly in distribution to $T^{-\frac{1}{2}}\Gamma(\theta_0)^{-\frac{1}{2}}\xi$ when $\theta_0^{i,n,*} = \theta_0$, i.e. $\sup_{\theta_0 \in K, 1 \le i \le B_n} \{ \left[f(\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta_0)) \right] - \mathbb{E}[f(T^{-\frac{1}{2}}\Gamma(\theta_0)^{-\frac{1}{2}}\xi)]\} \to^{\mathbb{P}} 0, for any bounded continuous function$ f.

Finally, we extend (8) to the case of a function of polynomial growth of order smaller than L. First note that by (8) and (8) we have for any $L' \in (L, 2\kappa)$ $\sup_{\theta_0 \in K, 1 \le i \le B_n} |\Delta_{i,n}(\theta_0)|^{L'} = O_{\mathbb{P}}(1). We now adopt the notations of$ [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 (8), (8), (8) and finally (8), conditions $[A1'']$, $[A4']$, $[A6]$, $[B1]$ and $[B2]$ in [Yoshida, 2011] are satisfied. It is straightforward that we can apply a conditional version (with respect to the operator) of Theorem 3 and Proposition 1 from [Yoshida, 2011] to get that for any $p \le L$, $\sup_{\theta_0 \in K, 1 \le i \le \Delta_n^{-1}} |\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta_0)|^p = O_{\mathbb{P}}(1). Such stochastic boundedness of conditional moments along with the converge$

□

So far we have focused on the case where $R_{i,n}(0)$ is bounded by the sequence M_n. Nonetheless, the time-varying parameter Hawkes process has a residual which is a priori not bounded at the beginning of a block. In Theorem **??**, we relax this assumption. In addition, we use regular conditional distribution techniques (see for instance Section 4.3 (pp. $77 - 80$) in [Breiman, 1992]) to obtain (8) when not conditioning by any particular starting value of θ_t^*. We provide the formal proof in what follows. Recall that $\mathbb{E}_{(i-1)\Delta_n}$ stands for $\mathbb{E}[.|\mathcal{F}_0^{i,n}]$.

Proof of Theorem **??**. We can decompose $\mathbb{E}_{(i-1)\Delta_n}\left[f\left(\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n})\right)\right]$ as

$$\mathbb{E}_{(i-1)\Delta_n}[f(\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n}))\mathbf{1}_{\{R_{i,n}(0) \le M_n\}}] \tag{36}$$

$$+ \mathbb{E}_{(i-1)\Delta_n}[f(\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n}))\mathbf{1}_{\{R_{i,n}(0) > M_n\}}]. \tag{37}$$

Let ξ as in Theorem **??**. On the one hand by a regular conditional distribution argument, if we define $G(\theta_0) = \Big[f\big(\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta_0)\big)\Big] - \mathbb{E}\Big[f\big(T^{-\frac{1}{2}}\Gamma(\theta_0)^{-\frac{1}{2}}\xi\big)\Big]$, we can express uniformly in $i \in \{1, \cdots, B_n\}$ the quantity $\mathrm{E}_{(i-1)\Delta_n}[f(\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n}))\mathbf{1}_{\{R_{i,n}(0) \leq M_n\}} - f(T^{-\frac{1}{2}}\Gamma(\theta^*_{(i-1)\Delta_n})^{-\frac{1}{2}}\xi)]$ as $G(\theta^*_{(i-1)\Delta_n})$ by definition of and because $\xi \mathbf{1} \mathcal{F}$. We note that $|G(\theta^*_{(i-1)\Delta_n})| \leq \sup_{\theta_0 \in K} | \Big[f(\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta_0))\Big] - \mathbb{E}[f(T^{-\frac{1}{2}}\Gamma(\theta_0)^{-\frac{1}{2}}\xi)]|$,

take the sup over i in (8), and in view of Theorem 8, we have shown that (8) is uniformly of order $o_{\mathbb{P}}(1)$.

On the other hand, (37) is bounded by $h_n^L Q \mathbf{1}_{\{R_{i,n}(0) > M_n\}}$ for some $Q > 0$, where we have used that $\widehat{\Theta}_{i,n}$ takes its values in a compact space. By a straightforward computation it is easy to see that $\mathbb{P}[R_{i,n}(0) > M_n] \leq \mathbb{P}[\lambda_*^n((i-1)\Delta_n) > M_n]$, which in turn can be dominated easily with Markov's inequality by $M_n^{-1}\mathbb{E}[\lambda_*^n((i-1)\Delta_n)] = O(nM_n^{-1})$. We recall that M_n is of the form n^q where q can be taken arbitrarily big, and we have thus shown that (37) vanishes asymptotically. □

8.1 Bias reduction of the local MLE

We go one step further and study the properties of the asymptotic conditional bias of the local MLE, i.e. the quantity $\mathbb{E}_{(i-1)\Delta_n}[\widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n}]$. We then derive the expression of a bias-corrected estimator $\widehat{\Theta}_{i,n}^{(BC)}$ whose expectation tends faster to $\theta^*_{(i-1)\Delta_n}$.

We start by estimating the order of the bias of the local MLE. As the reader can see, the following computations are very involved. Therefore, in this section only, we adopt the following notation conventions. First, we drop the index reference i. Consequently, all the variables N^n,λ_*^n,l_n, , etc. should be read $N^{i,n}$,$\lambda_*^{i,n}$,$l_{i,n}$, , etc. All the results are implicitly stated uniformly in the block index. Second, for a random variable Z that admits a first order moment for the operator , we denote by $\overline{Z}$ its centered version, i.e. the random variable $Z - [Z]$. We adopt Einstein's summation convention, i.e. any indice that is repeated in an expression is implicitly summed. For example the expression $a_{ij}b_j$ should be read $\sum_j a_{ij}b_j$. Finally, as in Section **??**, for a matrix M, we use superscripts to designate elements of its inverse, i.e. M^{ij} stands for the element in position (i,j) of M^{-1} when it is well-defined,

$M^{ij} = 0$ otherwise.

By a Taylor expansion of the score function around the maximizer of the likelihood function, it is immediate to see that there exists $\xi_n \in [\widehat{\Theta}_n, \theta_0]$ such that $0 = \partial_\theta l_n(\widehat{\Theta}_n) = \partial_\theta l_n(\theta_0) + \partial_\theta^2 l_n(\theta_0)(\widehat{\Theta}_n - \theta_0) + \frac{1}{2}\partial_\theta^3 l_n(\xi_n)(\widehat{\Theta}_n - \theta_0)^{\otimes 2}, where \partial_\theta^3 l_n(\xi_n)(\widehat{\Theta}_n - \theta_0)^{\otimes 2}$ is a compact expression for the vector whose i-th component is $\partial^3_{\theta,ijk} l_n(\xi_n)(\widehat{\Theta}_n - \theta_0)_j(\widehat{\Theta}_n - \theta_0)_k$. Let $\epsilon \in (0,1)$. By application of Lemmas 8 and 8, it still holds that $\partial_\theta l_n^c(\theta_0) + \partial_\theta^2 l_n^c(\theta_0)(\widehat{\Theta}_n - \theta_0) + \frac{1}{2}\partial_\theta^3 l_n^c(\xi_n)(\widehat{\Theta}_n - \theta_0)^{\otimes 2} = O_{\mathbb{P}}(h_n^{1-\epsilon\kappa}), where the residual term O_{\mathbb{P}}(h_n^{1-\epsilon\kappa})$ admits clearly moments of any order with respect to . We now apply the operator , divide by $h_n T$ and obtain $[\overline{\Gamma}_n^c(\theta_0)(\widehat{\Theta}_n - \theta_0)] + [\Gamma_n^c(\theta_0)][\widehat{\Theta}_n - \theta_0] - [\frac{\partial_\theta^3 l_n^c(\xi_n)}{2h_nT}(\widehat{\Theta}_n - \theta_0)^{\otimes 2}] = O_{\mathbb{P}}(h_n^{-\epsilon\kappa}), where the expectation of the first term has$ in (8). The term $[\Gamma_n^c(\theta_0)][\widehat{\Theta}_n - \theta_0]$ is of interest since it contains the quantity we want to evaluate. The first and the third terms have thus to be evaluated to derive an expansion of the bias. We start by the first term, i.e. the covariance between our estimator and $\Gamma_n^c(\theta_0)$. To compute the limiting value of such covariance, we consider the martingale $M_n^c(t,\theta_0) = \int_0^t \frac{\partial_\theta \lambda^{n,c}(s,\theta_0)}{\lambda^{n,c}(s,\theta_0)}\{dN_s^{n,c} - \lambda^{n,c}(s,\theta_0)ds\}$, and we define the empirical covariance processes $C_n^c(\theta_0)$ and $Q_n^c(\theta_0)$ whose components are, for any triplet $(i,j,k) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3$,

$$\mathrm{C}_n^c(\theta_0)_{i,jk} = \frac{1}{h_nT}\int_0^{h_nT} \partial_{\theta,i}\lambda^{n,c}(s,\theta_0)\partial^2_{\theta,jk}\log\lambda^{n,c}(s,\theta_0)ds, and Q_n^c(\theta_0)_{i,jk} = -\frac{M_n^c(T,\theta_0)_i}{h_nT}\int_0^{h_nT}\frac{\partial_\theta\lambda^{n,c}(s,\theta_0)_j\partial_\theta\lambda^{n,c}(s,\theta_0)_k}{\lambda^{n,c}(s,\theta_0)}ds,$$

We define in a similar way $C_n(\theta_0)$ and $Q_n(\theta_0)$. The next lemma clarifies the role of $C_n^c(\theta_0) + Q_n^c(\theta_0)$ and is a straightforward calculation.

We have $[\mathrm{C}_n^c(\theta_0)_{i,jk} + Q_n^c(\theta_0)_{i,jk}] = -\sqrt{h_nT}[\Delta_n^c(\theta_0)_i\Gamma_n^c(\theta_0)_{jk}]$.

Proof. Note that for two $\mathbb{L}_2$ bounded processes $(u_s)_s$, $(v_s)_s$, we have $\langle\int_0^\cdot u_s\{dN_s^{n,c} - \lambda^{n,c}(s,\theta_0)ds\}, \int_0^\cdot v_s\{dN_s^{n,c} - \lambda^{n,c}(s$

Now, by the same argument as for the proof of (8), we have for any integer $p \geq 1$ and any $\epsilon \in (0,1)$, $\sup_{\theta_0\in K} h_n^{\epsilon\frac{p}{2}}\mathbb{E}_{\theta_0,n}|C_n^c(\theta_0) - C(\theta_0)|^p \to^{\mathbb{P}} 0, and \sup_{\theta_0\in K} h_n^{\epsilon\frac{p}{2}}|\mathbb{E}_{\theta_0,n}[Q_n^c(\theta_0) - Q(\theta_0)]|^p \to^{\mathbb{P}} 0 where C and Q were defined respe$ $\theta_0)]_i in our expansion of the bias in terms of \mathrm{C}(\theta_0) + Q(\theta_0)$, we need to control the convergence of $\Gamma_n^c(\theta_0)^{-1}$ toward $\Gamma(\theta_0)^{-1}$. We define $c_0 = \min_{\theta_0\in K}\min\{c \in \mathbb{R}_+ | \forall x \in \mathbb{R}^3 - \{0\}, x^T\Gamma(\theta_0)x \geq c|x|_2^2 > 0\}$, the smallest eigenvalue of all the matrices $\Gamma(\theta_0)$. We consider the sequence of events $\mathbb{B}_n(\theta_0) = \{\forall x \in \mathbb{R}^3 - \{0\}, x^T\Gamma_n^c(\theta_0)x \geq \frac{c_0}{2}|x|_2^2\}$, and their complements $\mathbb{B}_n(\theta_0)^c$. We have, for any integer $p \geq 1$ and any $\epsilon \in (0,1)$ that [**(i)**] $\sup_{\theta_0\in K}[\mathbb{B}_n(\theta_0)^c] = O_{\mathbb{P}}(h_n^{-\epsilon\frac{p}{2}})$. [**(ii)**] $\sup_{\theta_0\in K} h_n^{\epsilon\frac{p}{2}}[|\Gamma_n^c(\theta_0)^{-1} - \Gamma(\theta_0)^{-1}|\mathbf{1}_{\mathbb{B}_n}] \to^{\mathbb{P}} 0$.

Proof. We start by showing **(i)**. We recall that in our notation convention, the symbol $|x|$ stands for $\sum_i |x_i|$ for any vector or matrix. Clearly, we have that ł[$\mathrm{B}_n(\theta_0)^c$] $\leq \{\forall x \in \mathbb{R}^3 - \{0\}, \frac{|x^T(\Gamma_n^c(\theta_0)-\Gamma(\theta_0))x|}{|x|_2^2} > \frac{c_0}{2}\}$, $andbyequivalenceofthenorms$ $|\mathrm{M}|$ and $\sup_{x\in\mathbb{R}^3-\{0\}} \frac{|x^TMx|}{|x|_2^2}$ on the space of symmetric matrices of $\mathbb{R}^3$, (8.1) implies the existence of some constant $\eta > 0$ such that ł[$\mathrm{B}_n(\theta_0)^c$] $\leq [|\Gamma_n^c(\theta_0) - \Gamma(\theta_0)| > \eta c_0]$

$\leq (\eta c_0)^{-p}|\Gamma_n^c(\theta_0) - \Gamma(\theta_0)|^p$, $whereMarkov'sinequalitywasusedatthelaststep.$ $(\mathbf{i})$ $thusfollowsfrom(8).Moreover,$ $(\mathbf{ii})$ ise $B^{-1}| = |B^{-1}(B-A)A^{-1}| \leq |A^{-1}|_\infty |B^{-1}|_\infty |B-A|$ applied to $\Gamma_n^c(\theta_0)$ and $\Gamma(\theta_0)$ on the set $\mathbb{B}_n(\theta_0)$. □

Let $\epsilon \in (0,1)$ and $i \in \{0,1,2\}$. The following expansion holds. $[\overline{\Gamma}_n^c(\theta_0)(\widehat{\Theta}_n - \theta_0)]_i = -\frac{\Gamma(\theta_0)^{jk}\{C(\theta_0)_{k,ij}+Q(\theta_0)_{k,ij}\}}{h_nT} + O_{\mathbb{P}}(h_n^{-\epsilon(\kappa\wedge\frac{3}{2})})$.

Proof. Note first that in view of Lemma 8.1 **(i)** along with Hölder's inequality, we have that $[\overline{\Gamma}_n^c(\theta_0)(\widehat{\Theta}_n - \theta_0)] = [\overline{\Gamma}_n^c(\theta_0)(\widehat{\Theta}_n - \theta_0)\mathbf{1}_{\mathbb{B}_n(\theta_0)}] + o_{\mathbb{P}}(h_n^{-\frac{3}{2}})$. Thus we can assume without loss of generality the presence of the indicator of the event $\mathbb{B}_n(\theta_0)$ in the expectation of the left-hand side of (8.1). Take $\epsilon \in (0,1)$ and $\tilde{\epsilon} \in (\epsilon, 1)$. As a consequence of (8.1), we have the representation, $\widehat{\Theta}_n - \theta_0 = \frac{1}{\sqrt{h_nT}}\Gamma_n^c(\theta_0)^{-1}\Delta_n^c(\theta_0) + \Gamma_n^c(\theta_0)^{-1}\frac{\partial_\theta^3 l_n^c(\xi_n)(\widehat{\Theta}_n-\theta_0)^{\otimes 2}}{2h_nT} + O_{\mathbb{P}}(h_n^{-\tilde{\epsilon}\kappa})$, $ontheset$ $\mathrm{B}_n(\theta_0)$, where the residual term $O_{\mathbb{P}}(h_n^{-\tilde{\epsilon}\kappa})$ admits moments of any order with respect to the operator . We inject (8.1) in the expectation and get $[\overline{\Gamma}_n^c(\theta_0)(\widehat{\Theta}_n - \theta_0)] = \frac{1}{\sqrt{h_nT}}[\overline{\Gamma}_n^c(\theta_0)\Gamma_n^c(\theta_0)^{-1}\Delta_n^c(\theta_0)\mathbf{1}_{\mathbb{B}_n(\theta_0)}]$
$+ [\overline{\Gamma}_n^c(\theta_0)\Gamma_n^c(\theta_0)^{-1}\frac{\partial_\theta^3 l_n^c(\xi_n)(\widehat{\Theta}_n-\theta_0)^{\otimes 2}}{2h_nT}\mathbf{1}_{\mathbb{B}_n(\theta_0)}]$
$+ O_{\mathbb{P}}(h_n^{-\epsilon\kappa})$, $wheretheresidualterm$ $\mathrm{O}_{\mathbb{P}}(h_n^{-\epsilon\kappa})$ is obtained by Hölder's inequality using the fact that $\epsilon < \tilde{\epsilon}$. By Lemma 8.1 **(ii)**, the first term admits the expansion $1\frac{}{\sqrt{h_nT}[\overline{\Gamma}_n^c(\theta_0)\Gamma(\theta_0)^{-1}\Delta_n^c(\theta_0)] + O_{\mathbb{P}}(h_n^{-\frac{3\epsilon}{2}})}$, $whereweusedHlder'sinequalityto$ and we neglected the effect of the indicator function by Lemma 8.1 **(i)**. For any $i \in \{0,1,2\}$, we develop the matrix product in (8.1), use Lemma 8.1 along with (8.1), and this leads to the estimate $1\frac{}{\sqrt{h_nT}[\overline{\Gamma}_n^c(\theta_0)\Gamma(\theta_0)^{-1}\Delta_n^c(\theta_0)]_i = \frac{\Gamma(\theta_0)^{jk}\{C(\theta_0)_{k,ij}+Q(\theta_0)_{k,ij}\}}{h_nT} + O_{\mathbb{P}}(h_n^{-\frac{3\epsilon}{2}})}$ $.Itremainstocontroltheterm$ ł$[\overline{\Gamma}_n^c(\theta_0)\Gamma_n^c(\theta_0)^{-1}\frac{\partial_\theta^3 l_n^c(\xi_n)(\widehat{\Theta}_n-\theta_0)^{\otimes 2}}{2h_nT}\mathbf{1}_{\mathbb{B}_n}$
Take $L \in (2, 2\kappa)$. By boundedness of moments of $h_n^{\frac{\epsilon}{2}}\overline{\Gamma}_n^c(\theta_0)_{ij}\Gamma_n^c(\theta_0)^{jk}\frac{\partial_{\theta,klm}^3 l_n^c(\theta)}{2h_nT}\mathbf{1}_{\mathbb{B}_n(\theta_0)}$, for any (i,j,k,l,m) and uniformly in $\theta_0 \in K$, we have

$$[\overline{\Gamma}_n^c(\theta_0)_{ij}\Gamma_n^c(\theta_0)^{jk}\frac{\partial_{\theta,klm}^3 l_n^c(\xi_n)(\widehat{\Theta}_n - \theta_0)_l(\widehat{\Theta}_n - \theta_0)_m}{2h_nT}\mathbf{1}_{\mathbb{B}_n(\theta_0)}]$$

$\leq Kh_n^{-\frac{\epsilon}{2}}[|(\widehat{\Theta}_n - \theta_0)_l(\widehat{\Theta}_n - \theta_0)_m|^{\frac{L}{2}}]^{\frac{2}{L}}$

$= O_{\mathbb{P}}(h_n^{-\frac{3\epsilon}{2}}\Big), where H lder' s inequality was applied for the first inequality, and Theorem 8 was used with the function$ f

: x $\to (x_l x_m\Big)^{\frac{L}{2}}$, which is of polynomial growth of order L, to get the final estimate. □

Finally, we derive the expansion of $\frac{1}{2h_nT}[\partial_\theta^3 l_n^c(\xi_n)(\widehat{\Theta}_n - \theta_0)^{\otimes 2}]$. First note that for any integer $p \geq 1$ and any $\epsilon \in (0,1)$, $\sup_{\theta_0 \in K} h_n^{\epsilon\frac{p}{2}} \mathbb{E}_{\theta_0,n}|\frac{1}{h_nT}\partial_\theta^3 l_n^c(\theta_0) - K(\theta_0)\Big|^p \to^{\mathbb{P}} 0, where$ K(θ_0) was introduced in (29). The next lemma is proved the same way as for Lemma 8.1.

Let $\epsilon \in (0,1)$ and $i \in \{0,1,2\}$. We have the expansion $1\overline{2h_nT[\partial_\theta^3 l_n^c(\xi_n)(\widehat{\Theta}_n - \theta_0)^{\otimes 2}]_i = \frac{\Gamma(\theta_0)^{jk}K(\theta_0)_{ijk}}{2h_nT} + O_{\mathbb{P}}(h_n^{-\epsilon(\kappa\wedge\frac{3}{2}\Big)}\Big)}.$

Proof. Consider three indices $i,j,k \in \{0,1,2\}$ and $\epsilon \in (0,1)$. We have the decomposition $1\overline{2h_nT[\partial_{\theta,ijk}^3 l_n^c(\xi_n)(\widehat{\Theta}_n - \theta_0)_j(\widehat{\Theta}_n - \theta_0}$

$\theta_0)_j(\widehat{\Theta}_n - \theta_0)_k]$ and $\frac{1}{2h_nT}[\partial_{\theta,ijk}^3 l_n^c(\xi_n)]$ by their estimates $[(\widehat{\Theta}_n - \theta_0)_j(\widehat{\Theta}_n - \theta_0)_k] = \frac{\Gamma(\theta_0)^{jk}}{h_nT} + O_{\mathbb{P}}(h_n^{-\epsilon(\kappa\wedge\frac{3}{2}\Big)}\Big), and \frac{1}{2h_nT}[\partial_{\theta,ijk}^3 l_n^c$

$K(\theta_0)_{ijk} + O_{\mathbb{P}}(h_n^{-\frac{\epsilon}{2}}\Big).(8.1) is obtained by injecting the expansion of \widehat{\Theta}_n - \theta_0$ in (8.1) up to the first order only, and (8.1) is a consequence of (8.1) and the uniform boundedness of moments of $\frac{\partial_\theta^4 l_n^c(\theta_0)}{h_nT}$ in $\theta_0 \in K$ by Lemma 8 **(ii)**. Note that the expansion (8.1) is not a direct consequence of Theorem 8 applied to $x \to x_jx_k$ since this would lead to the weaker estimate $\frac{\Gamma(\theta_0)^{jk}}{h_nT} + o_{\mathbb{P}}(h_n^{-1})$ instead. Finally, the second term is of order $O_{\mathbb{P}}(h_n^{-\frac{3\epsilon}{2}}\Big)$ by Hölder's inequality along with Theorem 8, and thus we are done. □

Before we turn to the final theorem, we recall for any $j \in \{0,1,2\}$ the expression $\mathrm{b}(\theta_0)_j = \frac{1}{2}\Gamma(\theta_0)^{ij}\Gamma(\theta_0)^{kl}(K(\theta_0)_{ikl} + 2\{C(\theta_0)_{k,il} + Q(\theta_0)_{k,il}\Big\}), which was defined in (16). We are now ready to state the general theore$

Let $\epsilon \in (0,1)$. The bias of the estimator $\widehat{\Theta}_{i,n}$ has the expansion $\text{ł}[\widehat{\Theta}_{i,n} - \theta_0\Big] = \frac{b(\theta_0)}{h_nT} + O_{\mathbb{P}}(h_n^{-\epsilon(\kappa\wedge\frac{3}{2}\Big)}\Big), uniformly in$ i $\in \{1,\cdots,B_n\}$ and in $\theta_0 \in K$. Moreover, the bias-corrected estimator $\widehat{\Theta}_{i,n}^{(BC)}$ defined in (**??**) has the (uniform) bias expansion $\text{ł}[\widehat{\Theta}_{i,n}^{BC} - \theta_0\Big] = O_{\mathbb{P}}(h_n^{-\epsilon(\kappa\wedge\frac{3}{2}\Big)}\Big).$

Proof. We drop the index i in this proof. Take $\epsilon \in (0,1)$ and some $j \in \{0,1,2\}$. By Lemma 8.1 and Lemma 8.1, we have $\text{ł}[\Gamma_n^c(\theta_0)\Big]_{jk}[\widehat{\Theta}_n - \theta_0\Big]_k = \frac{\Gamma(\theta_0)^{kl}(K(\theta_0)_{jkl} + 2\{C(\theta_0)_{l,jk} + Q(\theta_0)_{l,jk}\Big\}\Big)}{2h_nT} + O_{\mathbb{P}}(h_n^{-\epsilon(\kappa\wedge\frac{3}{2}\Big)}\Big), which is a set of simultan$

$\in \{0,1,2\}$,

$\text{ł}[\widehat{\Theta}_n - \theta_0\Big]_j = \frac{\Gamma(\theta_0)^{ij}\Gamma(\theta_0)^{kl}(K(\theta_0)_{ikl} + 2\{C(\theta_0)_{k,il} + Q(\theta_0)_{k,il}\Big\}\Big)}{2h_nT} + O_{\mathbb{P}}(h_n^{-\epsilon(\kappa\wedge\frac{3}{2}\Big)}\Big), which is exactly (8.1). Finally, a calculation sin$

$\mathrm{b}\text{ł}(\widehat{\Theta}_n\Big) = b(\theta_0\Big) + O_{\mathbb{P}}(h_n^{-\frac{\epsilon}{2}}\Big) so that we have (8.1) and this concludes the proof.$ □

We conclude by showing the version of the preceding theorem in terms of $\mathbb{E}_{(i-1)\Delta_n}$.

Proof of Theorem **??**. This follows exactly the same argument as for the proof of Theorem **??**. □

8.2 Proof of the GCLT

In this section we present the proof of Theorem **??** using a similar martingale approach as in [**?**]. Using a different decomposition than (34) on p. 22 of the cited work, we obtain following the same line of reasoning as in the proof of (37) on p. 47-48 that a sufficient condition to show that the GCLT holds is We have uniformly in $i \in \{1, \cdots, B_n\}$ that there exists $\epsilon > 0$ such that

$$\mathrm{Var}_{(i-1)\Delta_n}\left[\sqrt{h_n}\left(\widehat{\Theta}_{i,n}^{(BC)} - \theta^*_{(i-1)\Delta_n}\right)\right] = T^{-1}\Gamma\left(\theta^*_{(i-1)\Delta_n}\right)^{-1} + o_{\mathbb{P}}(1), \tag{38}$$

$$\mathbb{E}_{(i-1)\Delta_n}\left[\left|\sqrt{h_n}\left(\widehat{\Theta}_{i,n}^{(BC)} - \theta^*_{(i-1)\Delta_n}\right)\right|^{2+\epsilon}\right] = O_{\mathbb{P}}(1), \tag{39}$$

$$\mathbb{E}_{(i-1)\Delta_n}\left[\widehat{\Theta}_{i,n}^{(BC)} - \theta^*_{(i-1)\Delta_n}\right] = o_{\mathbb{P}}\left(n^{-1/2}\right), \tag{40}$$

where for any $t \in [0, T]$ and any random variable X, $\mathrm{Var}_t[X] = \mathbb{E}_t\big[(X - \mathbb{E}_t[X])^2\big]$.

The above-mentioned approach is based on techniques introduced in [**?**], but it is much different and deeper. Indeed, [**?**] provides conditions which in this specific case are hard to verify due to the past correlation of the model. We choose to go through a different path. More specifically, the cited author uses a different decomposition than (**??**). We thus obtain different conditions which are hard to verify, and this is the main goal of the proofs.

Proof of Theorem **??** *under* [C*]. We split the proof into two parts.

Step 1. The first part of the proof consists in showing that

$$\Theta = \frac{1}{B_n}\sum_{i=1}^{B_n} \theta^*_{(i-1)\Delta_n} + o_{\mathbb{P}}\left(n^{-1/2}\right). \tag{41}$$

Note that (41) is to be compared to (**??**) for the toy model. Moreover, (41) was also shown in (35) on pp. 46-47 in [**?**], but the parameter process was restricted to follow a continuous Itô-process. To show (41), it is sufficient to show that

$$\frac{\sqrt{n}}{B_n}\sum_{i=1}^{B_n}\left|\theta^*_{(i-1)\Delta_n} - \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} \theta^*_s ds\right| = o_{\mathbb{P}}(1). \tag{42}$$

We can bound (42) by

$$\frac{\sqrt{n}}{B_n}\sum_{i=1}^{B_n}\Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n}\underbrace{\left|\theta^*_{(i-1)\Delta_n}-\theta^*_s\right|}_{O_{\mathbb{P}}(\Delta_n^{\gamma})}ds=o_{\mathbb{P}}(1), \tag{43}$$

where we used [C]-**(i)** to obtain the order in (43). Thus, we deduce that the left-hand side in (43) is of order $O_{\mathbb{P}}(h_n^{\gamma}n^{\frac{1}{2}-\gamma})$. In view of the left inequality in [BC] and the fact that $\gamma>\frac{1}{2}$, this vanishes asymptotically. Thus, we have proved (41).

Step 2. We keep here the techniques and notations introduced in Section **??**, and replace $\widehat{\Theta}_{i,n}$ by the local estimator $\widehat{\Theta}_{i,n}^{(BC)}$ in the definitions of $M_{i,n}$ and $B_{i,n}$. To show the GCLT, we will show that $S_n^{(B)}\to^{\mathbb{P}}0$ and we will prove the existence of some V_T such that $\mathcal{F}_T^{\theta^*}$-stably in law, $S_n^{(M)}\to V_T^{\frac{1}{2}}\mathcal{N}(0,1)$. Note that the former is a straightforward consequence of (40). To show the latter $S_n^{(M)}\to V_T^{\frac{1}{2}}\mathcal{N}(0,1)$, we will use Theorem 3.2 of p. 244 in [Jacod, 1997]. First, we show the conditional Lindeberg condition (3.13), i.e. in our case that for any $\eta>0$ we have

$$\frac{n}{B_n^2}\sum_{i=1}^{B_n}\mathbb{E}_{(i-1)\Delta_n}\Big[M_{i,n}^2\mathbf{1}_{\{\frac{\sqrt{n}}{B_n}M_{i,n}>\eta\}}\Big]\to^{\mathbb{P}}0. \tag{44}$$

Let $\eta>0$. First, note that $\frac{n}{B_n}=h_n$. Using Hölder's inequality, we obtain that

$$h_n\mathbb{E}_{(i-1)\Delta_n}\Big[M_{i,n}^2\mathbf{1}_{\{\frac{\sqrt{n}}{B_n}M_{i,n}>\eta\}}\Big]\leq\underbrace{\Big(\mathbb{E}_{(i-1)\Delta_n}\Big[\big(\sqrt{h_n}M_{i,n}\big)^{2+\epsilon}\Big]\Big)^{\frac{2}{2+\epsilon}}}_{a_{i,n}}\underbrace{\Big(\mathbb{E}_{(i-1)\Delta_n}\Big[\mathbf{1}_{\{\frac{\sqrt{n}}{B_n}M_{i,n}>\eta\}}\Big]\Big)^{\frac{\epsilon}{2+\epsilon}}}_{b_{i,n}}.$$

On the one hand we have that $a_{i,n}$ is uniformly bounded in view of (39) from [C*]. On the other hand, using also (39) along with [C]-**(ii)**, we have that $b_{i,n}$ goes uniformly to 0. We have thus proved (44). We now prove the conditional variance condition (3.11), i.e. that

$$\frac{n}{B_n^2}\sum_{i=1}^{B_n}\mathbb{E}_{(i-1)\Delta_n}\big[M_{i,n}^2\big]\to^{\mathbb{P}}V_T:=T^{-2}\int_0^T\Gamma(\theta_s^*)^{-1}ds. \tag{45}$$

We have that

$$\frac{n}{B_n^2}\sum_{i=1}^{B_n}\mathbb{E}_{(i-1)\Delta_n}\big[M_{i,n}^2\big]=\frac{1}{T}\sum_{i=1}^{B_n}h_n\mathbb{E}_{(i-1)\Delta_n}\big[M_{i,n}^2\big]\Delta_n.$$

We use Proposition I.4.44 on p.51 in [Jacod and Shiryaev, 2003b] along with (38) from [C*] to show (45). Now, conditions (3.10) and (3.12) are automatically satisfied because $M_{i,n}$ is a martingale increment and since we consider the reference continuous martingale $\boldsymbol{M} = 0$. Finally we show condition (3.14) to get the stable convergence. We thus consider a bounded ${}^{\theta^*}$-martingale Z, and we show that

$\sqrt{n}\frac{}{B_n \sum_{i=1}^{B_n} \mathbb{E}_{(i-1)\Delta_n}\left[M_{i,n}\Delta Z_{i,n}\right] \to^{\mathbb{P}} 0, where} \Delta Z_{i,n} := Z_{i\Delta_n} - Z_{(i-1)\Delta_n}$. Using the Taylor expansion (8.1) and the boundedness of Z, by a similar calculation as in Lemma 8.1, we have

$\sqrt{n}\frac{}{B_n \sum_{i=1}^{B_n} \mathbb{E}_{(i-1)\Delta_n}\left[M_{i,n}\Delta Z_{i,n}\right] = \frac{h_n}{\sqrt{n}} \sum_{i=1}^{B_n} \Gamma(\theta^*_{(i-1)\Delta_n})^{-1} \mathbb{E}_{(i-1)\Delta_n}[\partial_\theta l^c_{i,n}(\theta^*_{(i-1)\Delta_n})\Delta Z_{i,n}] + o_{\mathbb{P}}(1). Note now that} \mathrm{l}^c_{i,n}(\theta^*_{(i-1)\Delta_n})$ can be written as an integral over the canonical Poisson martingale :

$\mathrm{l}^c_{i,n}(\theta^*_{(i-1)\Delta_n}) = \int_0^{h_n T} \int_{\mathbb{R}_+} \frac{\partial_\theta \lambda^{i,n,c}(s,\theta^*_{(i-1)\Delta_n})}{\lambda^{i,n,c}(s,\theta^*_{(i-1)\Delta_n})} \mathbb{1}_{\{0 \leq z \leq \lambda^{i,n,c}(s,\theta^*_{(i-1)\Delta_n})\}} \{\overline{N}^{i,n}(ds,dz) - \overline{\Lambda}^{i,n}(ds,dz)\}, with \overline{\Lambda}^{i,n}(ds,dz) = ds \otimes dz$. We deduce from the above representation that $\mathbb{E}_{(i-1)\Delta_n}[\partial_\theta l^c_{i,n}(\theta^*_{(i-1)\Delta_n})\Delta Z_{i,n}] = 0$, since both σ-fields $\mathcal{F}_T^{\theta^*}$ and $\mathcal{F}_T^{\overline{N}}$ are independent, so that Z and $\overline{N}^{i,n} - \overline{\Lambda}^{i,n}$ are orthogonal. Thus (8.2) holds. Thus, by Theorem 3.2 of [Jacod, 1997], we have the $\mathcal{F}_T^{\theta^*}$-stable convergence in law of $S_n^{(M)}$ toward an $\mathcal{F}_T^{\theta^*}$-conditional Gaussian limit with random variance V_T. In particular, we have that V_T and $\mathcal{N}(0,1)$ in Theorem **??** are independent from each other. □

We prove now that we can obtain (38), (39) and (40) in Condition $[C^*]$. First note that for any $L \in (0, 2\kappa)$, a calculation gives $\mathrm{E}_{(i-1)\Delta_n}|\sqrt{h_n}(\widehat{\Theta}^{(BC)}_{i,n} - \widehat{\Theta}_{i,n})|^L = h_n^{-\frac{L}{2}} T^{-L} \mathbb{E}_{(i-1)\Delta_n}|b(\widehat{\Theta}_{i,n})|^L = O_{\mathbb{P}}(h_n^{-\frac{L}{2}}) uniformly in i \in \{1, ..., B_n\}$. Thus, combining the previous estimate with Theorem **??**, we have shown that Theorem **??** remains true if $\widehat{\Theta}_{i,n}$ is replaced by $\widehat{\Theta}^{(BC)}_{i,n}$. We will use this fact in the following. If we decompose the conditional variance in (38) as

$$\mathbb{E}_{(i-1)\Delta_n}[(\sqrt{h_n}(\widehat{\Theta}^{(BC)}_{i,n} - \theta^*_{(i-1)\Delta_n}))^2] - \mathbb{E}_{(i-1)\Delta_n}[\sqrt{h_n}(\widehat{\Theta}^{(BC)}_{i,n} - \theta^*_{(i-1)\Delta_n})]^2,$$

then (38) follows from Theorem **??**. Moreover, (39) is a direct consequence of Theorem **??**. Finally, in view of (**??**) in Theorem **??**, (40) holds if there exists $\epsilon \in (0,1)$ such that $\sqrt{n} = o_{\mathbb{P}}(h_n^{\epsilon(\kappa \wedge \frac{3}{2})})$. From the relation $\sqrt{n} = h_n^{\frac{\delta}{2}}$, this can be reexpressed as $\frac{\delta}{2} < \kappa \wedge \frac{3}{2}$. If we replace κ by its expression, we get the two conditions $\frac{\delta}{2} < \gamma(\delta - 1)$ and $\frac{\delta}{2} < \frac{3}{2}$, that is $\frac{\gamma}{\gamma - \frac{1}{2}} < \delta < 3$. This is exactly condition [BC].

8.3 Proof of Proposition ??

Proof. Let $\gamma \in (0,1]$ and $\alpha \in (0, \frac{\gamma}{1+\gamma})$ and finally $\delta \in (1 + \frac{1}{\gamma}, \frac{1}{\alpha})$. We follow the proof of Theorem ??. (38) and (39) are true since $\delta > 1 + \frac{1}{\gamma}$. Moreover, by assumption on δ and α, (40) is replaced by $\mathbb{E}_{(i-1)\Delta_n}\left[\widehat{\Theta}_{i,n}^{(BC)} - \theta^*_{(i-1)\Delta_n}\right] = O_{\mathbb{P}}(n^{-\gamma(1-\delta^{-1})\wedge\delta^{-1}}) = o_{\mathbb{P}}(n^{-\alpha})$. writing the decomposition

$$\mathrm{n}^{\alpha}\frac{}{B_n\sum_{i=1}^{B_n}(\widehat{\Theta}_{i,n}-\theta^*_{(i-1)\Delta_n})=n^{\alpha-\frac{1}{2}}\{S_n^{(B)}+S_n^{(M)}\},wehave}\mathrm{n}^{\alpha-\frac{1}{2}}S_n^{(M)} \to^{\mathbb{P}} 0$$ since the central limit theorem for $S_n^{(M)}$ is still valid and $\alpha < \frac{1}{2}$. Finally $n^{\alpha-\frac{1}{2}}S_n^{(B)} = o_{\mathbb{P}}(1)$. This concludes the proof for $\widehat{\Theta}_n$. The proof for the bias corrected case follows the same path using $\mathbb{E}_{(i-1)\Delta_n}\left[\widehat{\Theta}_{i,n}^{(BC)} - \theta^*_{(i-1)\Delta_n}\right] = O_{\mathbb{P}}(n^{-\gamma(1-\delta^{-1})\wedge\frac{3}{2}\delta^{-1}})$ in lieu of the previous estimate. □

8.4 Proof of Proposition ??

Note that for any $\theta \in K$, we have $\partial_\xi^2 l_{i,n}(n^{-1}\xi)_{|\xi=n\theta} = n^{-2}\partial_\theta^2 l_{i,n}(\theta), andthusn^{-1}\widehat{C}_n = \frac{1}{B_n}\sum_{i=1}^{B_n}\partial_\theta^2 l_{i,n}(\widehat{\Theta}_{i,n})^{-1}h_n$

$= \frac{1}{TB_n}\sum_{i=1}^{B_n}\Gamma_{i,n}(\widehat{\Theta}_{i,n})^{-1}$,

$sothatitissufficienttoproveuniformlyin \mathrm{i} \in \{1,...,B_n\}$ the estimates $\Gamma_{i,n}(\widehat{\Theta}_{i,n})^{-1} = \Gamma(\theta^*)^{-1} + o_{\mathbb{P}}(1) and \Gamma(\theta^*)^{-1} = \Delta_n^{-1}\int\Gamma(\theta_t^*)^{-1}dt + o_{\mathbb{P}}(1). To show (8.4), we consider the decomposition \Gamma_{i,n}(\widehat{\Theta}_{i,n})^{-1} - \Gamma(\theta^*)^{-1} = \underbrace{\Gamma_{i,n}(\widehat{\Theta}_{i,n})^{-1} - \Gamma_{i,n}(\theta^*)^{-1}}_{a_{i,n}}$

$\sup_{\theta\in K}\frac{1}{h_n}|\partial_\theta(\partial_\theta^2 l_{i,n}(\theta))^{-1}||\widehat{\Theta}_{i,n} - \theta^*|. Bysomealgebraiccalculusitisstraightforwardtoshowthattheterm \sup_{\theta\in K}\frac{1}{h_n}|\partial_\theta($ is $\mathbb{L}_p$ bounded by virtue of Lemma 8 **(i)** and Lemma 8.1 **(i)**. By uniform consistency of $\widehat{\Theta}_{i,n}$, this yields $a_{i,n} = o_{\mathbb{P}}(1)$. Moreover, we have that $b_{i,n} = o_{\mathbb{P}}(1)$ as a direct consequence of Lemma 8.1 **(ii)**. Thus (8.4) holds. Finally the approximation (8.4) is a straightforward consequence of Lemma 8 **(i)** and Lemma 8.1 **(i)** along with assumption [C]-**(i)**.