

# Statistical inference for the doubly stochastic self-exciting process

Simon Clinet\* and Yoann Potiron\*\*

\* Graduate School of Mathematical Sciences, University of Tokyo: 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan. Email: simon@ms.u-tokyo.ac.jp, website : <http://www.ms.u-tokyo.ac.jp/~simon/>  
 \*\* (Corresponding Author) Faculty of Business and Commerce, Keio University. 2-15-45 Mita, Minato-ku, Tokyo, 108-8345, Japan. Phone: +81-3-5418-6571. Email: potiron@fbc.keio.ac.jp, website: <http://www.fbc.keio.ac.jp/~potiron>

This version: February 27, 2017

## Abstract

We introduce and show the existence of a Hawkes self-exciting point process with exponentially-decreasing kernel and where parameters are time-varying. The quantity of interest is defined as the integrated parameter  $T^{-1} \int_0^T \theta_t^* dt$ , where  $\theta_t^*$  is the time-varying parameter, and we consider the high-frequency asymptotics. To estimate it naïvely, we chop the data into several blocks, compute the maximum likelihood estimator (MLE) on each block, and take the average of the local estimates. The asymptotic bias explodes asymptotically, thus we provide a non-naïve estimator which is constructed as the naïve one when applying a first-order bias reduction to the local MLE. We show the associated central limit theorem. Monte Carlo simulations show the importance of the bias correction and that the method performs well in finite sample, whereas the empirical study discusses the implementation in practice and documents the stochastic behavior of the parameters.

**Keywords:** Hawkes process; high-frequency data; integrated parameter ; time-varying parameter

## 1 Introduction

In high-frequency data, market events are observed more often than ever. As an example, the correlation between the timing of those events and other financial quantities, such as asset price, volatility and microstructure noise has become of special interest. Also, financial agents can model the order book to predict key quantities, such as the volume of trades in the next hour. For all those reasons, models for inter-arrival times, also called duration models, are needed. As a pioneer work, [16] introduced the autoregressive conditional duration (ACD) model. Other references include and are not limited to [38], [40], as well as [4], [18], and more recently [35] and [34].

The cited work is partly based on the self-exciting Hawkes point process introduced in [22] and [23]. In that model, the intensity of the point process  $N_t$  is defined as  $\lambda(t) := \nu + \int_0^t \phi_{t-s} dN_s$ , where the baseline  $\nu > 0$ . Self-exciting processes are very popular to model phenomena mainly because future events can be boosted by past events. In the high-frequency finance literature, [37] documented such time-clustering property in the order flow of several stocks. Other examples of application can be found in [15], [1], [2], [41], [27], etc. Also, [3] offers a general overview of the Hawkes process applications in finance.

We restrict our attention to the case with exponential exciting function  $\phi_t = ae^{-bt}$ , as studied in [30]. Time-varying parameter extensions have already been considered taking the locally stationary

processes approach in [36], and restricting to the baseline time-varying case in [19], [8] and [7]. Our approach is much in line with the latter couple of work in that we consider the high-frequency point of view. In [8] the authors allow the background parameter (as they call it)  $\nu$  to be time-varying to incorporate intraday seasonality and consider the ACD model with time-varying background parameter. They illustrate that on data the ACD performs better when allowing for time-varying background, and that as it was already well-documented in [16] the background parameter is moving a lot intraday.

This calls into question what happens to the other two parameters  $a$  and  $b$  when sampling at the ultra high frequency ? Do they look constant intraday ? In our empirical study, we document that they are moving intraday just as the background parameter does although the intraday seasonality pattern isn't as clear. Indeed from one day to the next, the paths are very much different and although intraday seasonality can definitely be considered as one factor, it seems that it can't solely explain such behavior. Correspondingly we introduce a self-exciting process with stochastic time-varying parameters  $\theta_t^* := (\nu_t^*, a_t^*, b_t^*)$ . The new object of interest is defined as the integrated parameter

$$\Theta := \frac{1}{T} \int_0^T \theta_t^* dt, \quad (1.1)$$

where  $T > 0$  is the horizon time.

To estimate the integrated parameter (1.1), we choose to do locally MLE estimations, which was studied in a parametric context in [10], and whose numerical computation can be consulted in [32]. Specifically, if we consider  $B_n := nh_n^{-1}$  regular non-overlapping blocks of observation with time length  $\Delta_n := Th_n n^{-1}$ , the estimator of (1.1) is defined as

$$\hat{\Theta}_n := \frac{1}{B_n} \sum_{i=1}^{B_n} \hat{\Theta}_{i,n}, \quad (1.2)$$

where  $\hat{\Theta}_{i,n}$  corresponds to the MLE applied to the market events on the  $i$ th block,  $n$  corresponds to the number of events' order between 0 and  $T$  (typically the expected number of events) and the block size  $h_n$  stands for the number of events in a block's order (typically the expected number of events on a block). The idea to use a Riemann sum of local estimates in high-frequency finance problems is very common, and can be found for example in [25] or [28]. Our own recent work includes [9]. The more general literature on local parametric approaches, when not considering the high-frequency data case, includes [17], but also [21], the locally stationary processes of [13], etc.

The first contribution of this paper is to obtain conditions on the stochastic parameter  $\theta_t^*$  and the block size  $h_n$  under which we can show a local central limit theorem (LCLT) in high-frequency asymptotics, and finiteness of moments of order  $2\kappa > 2$ . The technique used, namely Quasi Likelihood Analysis (QLA) whose most general and powerful formulation can be consulted in [39], is not problem-specific and can very much be applied to different models. For this part, blocks with  $h_n$  which goes to infinity very slowly will be preferred, as the block length  $\Delta_n$  will be smaller, and thus the parameter  $\theta_t^*$  almost constant on each block. In particular, if  $\theta_t^*$  is not constant, we obtain that a necessary condition is

$$h_n = o(\sqrt{n}). \quad (1.3)$$

The second issue that this work is addressing is the asymptotic bias generated by  $\hat{\Theta}_n$ . Even in the simple parametric case, note that the bias of the MLE on each block  $\hat{\Theta}_{i,n}$  is of order  $h_n^{-1}$ , and thus that the bias of  $\hat{\Theta}_n$  is also of the same order  $h_n^{-1}$ . The asymptotic bias, i.e. the bias of the scaled error  $\sqrt{n}(\hat{\Theta}_n - \Theta)$ , is thus of order  $\sqrt{n}h_n^{-1}$ . If we want to obtain no asymptotic bias, we thus need to assume that

$$\sqrt{n} = o(h_n). \quad (1.4)$$

Thus, for that part, the block size  $h_n$  should be as large as possible.

In view of the necessary conditions (1.3) and (1.4), there is no hope to obtain any  $h_n$  for which the asymptotic bias of  $\widehat{\Theta}_n$  will vanish. For that reason, we derive the one-order bias-corrected parametric MLE. Correspondingly, we define  $\widehat{\Theta}_{i,n}^{(BC)}$  as the bias-corrected MLE when fitted to the observations on the  $i$ th block. Moreover, the bias-corrected estimator of (1.1) is defined as

$$\widehat{\Theta}_n^{(BC)} := \frac{1}{B_n} \sum_{i=1}^{B_n} \widehat{\Theta}_{i,n}^{(BC)}. \quad (1.5)$$

We provide conditions under which  $\widehat{\Theta}_n^{(BC)}$  has no asymptotic bias. Finally, the global central limit theorem (GCLT) is obtained as an immediate consequence of the finiteness of moments of order  $2\kappa$ , the LCLT and the fact that the asymptotic bias of  $\widehat{\Theta}_n^{(BC)}$  is null.

The following section provides the setup, Section 3 develops the statistical underpinning for the time-varying self-exciting process case and Section 4 introduces the general model. In Section 5, we discuss the main results. We give some practical guidance about the implementation of the statistical procedure in Section 6. We also carry out numerical simulations in Section 7, and give an empirical illustration on real tick-by-tick data in section 8. Finally, Section 9 concludes. Proofs can be found in the Annex.

## 2 The setup

In this work, the terminology "market event" should be understood as possibly corresponding to a time of trade, bid or ask order (limit or market), an order of cancellation, the time of a price change, etc. We need to introduce some notation first, that will be used throughout this work. For any stochastic process  $X_t$ , we define  $\mathbf{F}^X = (\mathcal{F}_t^X)_{t \in [0, T]}$ , where  $\mathcal{F}_t^X = \sigma\{X_s, 0 \leq s \leq t\}$  designates the canonical filtration generated by  $X_t$ . We assume that  $N_t^n$  is a point process, which counts the number of events on  $[0, t]$ . It means that  $dN_t^n = 1$  if there is a market event at time  $t$  and  $dN_t^n = 0$  if not. Moreover, we assume that there is no jump at time 0 and thus that  $dN_0^n = 0$ . Correspondingly, we define the intensity of market events  $\lambda_*^n(t)$ . The intensity process can be thought as the instantaneous expected number of events, i.e.  $\lambda_*^n(t)dt = \mathbb{E}[dN_t^n | \mathcal{F}_t^{(\theta^*, N^n)}]$ , where  $\mathcal{F}_t^{(\theta^*, N^n)}$  is the filtration generated by  $\mathcal{F}_t^{\theta^*}$  and  $\mathcal{F}_t^{N^n}$ . For definitions, the reader can consult [14] or [26] for more general results about the compensator of a point process.

There are commonly two ways to make the number of events go to infinity. The low-frequency asymptotics assume that  $T \rightarrow \infty$ . [10] took this approach in an ergodic framework. On the contrary, the high-frequency point of view (also sometimes called heavy traffic asymptotics) assumes that  $T$  is fixed, and that the number of events explodes on  $[0, T]$ . We adopt the latter approach and further consider a sequence of intensities such that  $\mathbb{E}[\lambda_*^n(t)]$  is exactly of order  $n$ , with  $n \rightarrow \infty$ . This yields a number of observations  $N_T^n$  of order  $n$ , so that we are in the classical framework of the large-sample theory.

## 3 Outline of the problem: an illustrative example

We start our theoretical exposition by the introduction of a point process toy model which provides an insight on the difficulties to overcome when considering the self-exciting model case. For the sake of simplicity, we stay at a heuristic level. The continuous parameter  $\theta_t^*$  is assumed to be 1-dimensional throughout the rest of this section. The parameter  $\theta_t^*$  is also restricted to belong to a compact set

$K = [\underline{\theta}, \bar{\theta}]$ , where  $\underline{\theta} > 0$ . Moreover,  $\theta_t^*$  is assumed to be adapted to some filtration  $\mathcal{F}_t$ , and to satisfy uniformly in  $0 \leq s < t \leq T$  that  $\mathbb{E}_s[|\theta_t^* - \theta_s^*|^p] = O_{\mathbb{P}}((t-s)^p)$ , where  $\mathbb{E}_s[\cdot]$  denotes the conditional expectation with respect to  $\mathcal{F}_s$ . Finally, we assume that the process  $N_t^n$  is adapted to  $\mathcal{F}_t$  and follows the dynamic of a doubly stochastic Poisson process (or Cox process) whose underlying stochastic intensity is assumed to be defined as  $\lambda_*^n(t) = n\sqrt{\theta_t^*}$ .

The estimation procedure  $\widehat{\Theta}_n$  follows [33]. We are interested in assessing the GCLT  $\sqrt{n}(\widehat{\Theta}_n - \Theta) \rightarrow^d V_T^{\frac{1}{2}}\mathcal{N}(0, 1)$ , where the asymptotic random variance  $V_T = T^{-1} \int_0^T v_t dt$  is independent from  $\mathcal{N}(0, 1)$ . Since the parameter  $\theta_t^*$  is smooth, we obtain

$$\Theta = \frac{1}{B_n} \sum_{i=1}^{B_n} \theta_{(i-1)\Delta_n}^* + O_{\mathbb{P}}(\Delta_n). \quad (3.1)$$

Consequently, the GCLT will follow if we can prove that

$$\frac{\sqrt{n}}{B_n} \sum_{i=1}^{B_n} (\widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^*) \rightarrow^d V_T^{\frac{1}{2}}\mathcal{N}(0, 1). \quad (3.2)$$

We focus on how to obtain (3.2) in this simple toy model. To do that, we rewrite the left hand-side of (3.2) as a sum of a martingale triangular arrays and an array of biases. Formally, (3.2) is expressed as

$$\underbrace{\frac{\sqrt{n}}{B_n} \sum_{i=1}^{B_n} M_{i,n}}_{S_n^{(M)}} + \underbrace{\frac{\sqrt{n}}{B_n} \sum_{i=1}^{B_n} B_{i,n}}_{S_n^{(B)}} \rightarrow^d V_T^{\frac{1}{2}}\mathcal{N}(0, 1), \quad (3.3)$$

where  $M_{i,n} = \widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^* - \mathbb{E}_{(i-1)\Delta_n}[\widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^*]$  and  $B_{i,n} = \mathbb{E}_{(i-1)\Delta_n}[\widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^*]$ . Our strategy to show (3.3) relies thus on exploiting the martingale decomposition on the left hand side of (3.3) to show that the covariances between blocks are negligible. More precisely, we want to prove that  $S_n^{(M)} \rightarrow^d V_T^{\frac{1}{2}}\mathcal{N}(0, 1)$  on the one hand, and that  $S_n^{(B)} \rightarrow^{\mathbb{P}} 0$  on the other hand. To show the former statement, classical sufficient conditions (see for instance Corollary 3.1 of pp. 58-59 in [20], or also Theorem VIII.3.33 in [26]) will hold if<sup>1</sup> uniformly in  $i \in \{1, \dots, B_n\}$  we can show that

$$\mathbb{E}_{(i-1)\Delta_n} \left[ \left( \sqrt{h_n} (\widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^*) \right)^2 \right] = v_{(i-1)\Delta_n} + o_{\mathbb{P}}(1), \quad (3.4)$$

and for some  $\kappa > 1$  that

$$\mathbb{E}_{(i-1)\Delta_n} \left[ \left| \sqrt{h_n} (\widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^*) \right|^{2\kappa} \right] = O_{\mathbb{P}}(1). \quad (3.5)$$

If we show the LCLT, i.e. the convergence of  $\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^*) \rightarrow^d v_{(i-1)\Delta_n}^{\frac{1}{2}}\mathcal{N}(0, 1)$  uniformly in the block number  $i \in \{1, \dots, B_n\}$ , we can deduce from (3.5) that (3.4) holds. This will be our strategy to show that  $S_n^{(M)} \rightarrow^d V_T^{\frac{1}{2}}\mathcal{N}(0, 1)$ . Moreover, to obtain the GCLT (3.3), we also need to show that the array of biases vanishes asymptotically. Accordingly, we will look at how to obtain those three conditions (boundedness of local moments of order  $2\kappa$ , LCLT and no asymptotic bias) in the toy model.

---

<sup>1</sup>The reader can find more details in Section 10.5

To fix ideas, we provide one way, which turns out to be helpful when estimating (1.1), to obtain asymptotic properties of the MLE in the parametric case when the intensity of the point process  $N_t^n$  is defined as  $\lambda_*^n(t) := n\sqrt{\theta^*}$ . The log-likelihood of the parametric model can be expressed up to a constant additive term as

$$l_n(\theta) = \log(\sqrt{\theta}) N_T^n - n\sqrt{\theta}T, \quad (3.6)$$

whose maximizer  $\hat{\theta}_n$  admits the explicit form

$$\hat{\theta}_n = \left( \frac{N_T^n}{nT} \right)^2. \quad (3.7)$$

If we introduce the martingale  $\tilde{N}_t^n = N_t^n - n\sqrt{\theta^*}t$ , we can rewrite  $\hat{\theta}_n$  as a function of  $\tilde{N}_T^n$ :

$$\hat{\theta}_n = \theta^* + \frac{2\sqrt{\theta^*}}{nT} \tilde{N}_T^n + \left( \frac{\tilde{N}_T^n}{nT} \right)^2. \quad (3.8)$$

As a consequence of classical limit theorems on martingales (see, e.g., Theorem 2.28 of p. 152 in [29] if we interpolate  $\tilde{N}_t^n$  as a continuous martingale, or the more general Theorem IX.7.3 in p. 584 of [26]), we obtain the CLT

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} T^{-\frac{1}{2}} \Gamma(\theta^*)^{-\frac{1}{2}} \mathcal{N}(0, 1),$$

where the Fisher information has the form  $\Gamma(\theta^*) = \frac{1}{4}(\theta^*)^{-\frac{3}{2}}$ . We also have the stronger statement that for any  $p \geq 1$ :

$$\mathbb{E} \left[ \left( \sqrt{n}(\hat{\theta}_n - \theta^*) \right)^p \right] \rightarrow \mathbb{E} \left[ \left( T^{-\frac{1}{2}} \Gamma(\theta^*)^{-\frac{1}{2}} \xi \right)^p \right], \quad (3.9)$$

where  $\xi$  follows a  $\mathcal{N}(0, 1)$ . Finally, we can also compute in (3.8) the finite-sample bias of the MLE

$$\mathbb{E}[\hat{\theta}_n - \theta^*] = \frac{\sqrt{\theta^*}}{nT}. \quad (3.10)$$

We are now back to the time-varying parameter model case  $\lambda_*^n(t) = n\sqrt{\theta_t^*}$ . In that case, we adapt the definition of the martingale as  $\tilde{N}_t^n = N_t^n - n \int_0^t \sqrt{\theta_s^*} ds$ . Working out from the explicit form (3.7), the local MLE can be expressed as

$$\begin{aligned} \hat{\Theta}_{i,n} &= \left( \frac{\tilde{N}_{i\Delta_n}^n - \tilde{N}_{(i-1)\Delta_n}^n}{h_n T} \right)^2 + \frac{2}{h_n^2 T^2} (\tilde{N}_{i\Delta_n}^n - \tilde{N}_{(i-1)\Delta_n}^n) \int_{(i-1)\Delta_n}^{i\Delta_n} n\sqrt{\theta_s^*} ds \\ &\quad + \frac{1}{h_n^2 T^2} \left( \int_{(i-1)\Delta_n}^{i\Delta_n} n\sqrt{\theta_s^*} ds \right)^2. \end{aligned} \quad (3.11)$$

In view of (3.11) and under the assumption that  $h_n = o(n^{\frac{2}{3}})$ , it is easy to obtain the LCLT with local conditional variance  $v_s = T^{-1} \Gamma(\theta_s^*)^{-1}$  and the boundedness of moments of order  $2\kappa$ . It remains to control the array of biases  $S_n^{(B)}$ . Calculation gives us

$$B_{i,n} = \frac{\sqrt{\theta_{(i-1)\Delta_n}^*}}{h_n T} + O_{\mathbb{P}}(\Delta_n), \quad (3.12)$$

where the residual term  $O_{\mathbb{P}}(\Delta_n)$ , which was not part of the parametric bias (3.10), is due to the deviation of  $\theta_t^*$ . In order to obtain no asymptotic bias, we assume that  $\sqrt{n} = o(h_n)$ . Consequently, if we assume that  $h_n = n^{1/\delta}$  with  $\frac{3}{2} < \delta < 2$ , we can prove the GCLT with asymptotic variance  $V_T = T^{-2} \int_0^T \Gamma(\theta_t^*)^{-1} dt$  in this toy model. This is a simple example where no further bias correction is needed to obtain the GCLT. However, in the time-varying self-exciting model, we will require to bias correct the estimator. This could be done in this simple setting via

$$\widehat{\Theta}_{i,n}^{(BC)} = \widehat{\Theta}_{i,n} - \frac{\sqrt{\widehat{\Theta}_{i,n}}}{h_n T}. \quad (3.13)$$

## 4 The model

We introduce in this section the time-varying self-exciting process, which will also be called the doubly stochastic Hawkes process, in analogy with the doubly stochastic Poisson process introduced in [11]. We first recall the definition of the non time-varying self-exciting point process. In the parametric case, the point process  $N_t^{P,n}$  can be defined via its intensity function

$$\lambda_*^{P,n}(t) = n\nu^* + \int_0^{t^-} na^* e^{-nb^*(t-s)} dN_s^{P,n}, \quad (4.1)$$

where  $\theta^* = (\nu^*, a^*, b^*)$  is the 3-dimensional parameter. The self-excitation property can be read directly from the intensity form  $\lambda_*^{P,n}$  in (4.1). Indeed, a market event arriving at time  $t$  will immediately boost the intensity, with an additional factor of magnitude  $na^*$ , favoring the occurrence of new events in the close future. The excitation then exponentially fades away after a time of order  $(nb^*)^{-1}$ . We explain now our choice regarding the asymptotics. First, we assume that the baseline intensity is proportional to  $n$  to boost the average rate of spontaneous events. Moreover, we assume that the excitation variables are of magnitude  $(na^*, nb^*)$  in order to preserve the proportionality between the typical excitation time after a market event,  $(nb^*)^{-1}$ , and the average inter-arrival time between two spontaneous events  $(n\nu^*)^{-1}$ . To wrap it up,  $N_t^n$  is a self-exciting process with parameters  $(n\nu^*, na^*, nb^*)$ . Note that other choices can lead to fairly different asymptotics such as the ones in [31] where authors suggested a model with baseline  $n\nu^*$  but a constant excitation kernel of the form  $a^* e^{-b^* t}$ .

We consider now the time-varying case. We assume that the 3-dimensional time-varying parameter process  $\theta_t^*$  is component-wise positive and is confined into the interior of a compact space  $K$ . This implies the existence of two non-negative vectors  $\underline{\theta}$  and  $\bar{\theta}$  such that  $0 < \underline{\theta} \leq \theta \leq \bar{\theta}$  for any  $\theta \in K$ , where the inequalities should be read component-wise. Moreover, we assume that  $N_t^n$  admits the  $\mathcal{F}_t$ -stochastic intensity  $\lambda_*^n(t)$  defined as

$$\lambda_*^n(t) = n\nu_t^* + \int_0^{t^-} na_s^* e^{-nb_s^*(t-s)} dN_s^n, \quad t \in (0, T], \quad (4.2)$$

where  $N_t^n$  and  $\theta_t^*$  are adapted to  $\mathcal{F}_t$ <sup>2</sup>, and  $N_0^n = 0$  a.s. The time-varying model (4.2) is a natural time-varying parametric model extension of (4.1). It is constructed in the same spirit as for the doubly stochastic Poisson process, in the sense that conditionally on the path of  $\theta_t^*$ ,  $N_t^n$  is distributed as a standard inhomogeneous Hawkes process. The formal definition of such a property along with the existence of the doubly stochastic Hawkes process can be found in Theorem 5.1. Finally, note that the time-varying parameter model (4.2) is more general than the parametric model (4.1). In particular, the intensity between two market events is not exponentially decreasing, but rather a sum of decreasing exponential functions, each one with its own starting point and decreasing rate.

<sup>2</sup>The formal definition of  $\mathcal{F}_t$  can be found in Section 5.

## 5 Main Results

### 5.1 Preliminary results

We present in this section general results for the doubly stochastic Hawkes process. We start by stating basic conditions on a given parameter process  $\theta_t$  that ensure the existence of the related doubly stochastic Hawkes process.

[E]

(i)

$$r := \sup_{t \in [0, T]} \frac{a_t}{b_t} < 1 \quad \mathbb{P} - a.s. \quad (5.1)$$

(ii)

$$\int_0^T \nu_s ds < +\infty \quad \mathbb{P} - a.s. \quad (5.2)$$

First, note that (5.1) is not harmful. Indeed, the corresponding condition for the existence of the parametric model is  $\frac{a}{b} < 1$ . Moreover, when estimating parameters by local MLE, we need  $\theta_t$  to be contained within a compact set. Thus, (5.2) will be verified automatically in that context. The next theorem shows the existence of the doubly stochastic Hawkes process associated with the process  $\theta_t$ . We recall that  $\mathcal{F}_t^\theta$  designates the canonical filtration associated with  $\theta_t$ . Moreover, the following bigger filtration  $\mathcal{F}_t$  is introduced for the construction of the doubly stochastic Hawkes process. We define the filtration as  $\mathcal{F}_t = \mathcal{F}_t^{(\theta, \bar{N})} = \mathcal{F}_t^\theta \vee \mathcal{F}_t^{\bar{N}}$ , where  $\bar{N}_t = \bar{N}([0, t] \times \mathbb{R})$  is a Poisson process of intensity 1 on  $\mathbb{R}^2$  which is independent from  $\theta_t$ .

**Theorem 5.1.** *(Existence) Under [E], there exists a point process  $N_t$  adapted to  $\mathcal{F}_t$  such that its  $\mathcal{F}_t$ -intensity has the representation*

$$\lambda(t) = \nu_t + \int_0^{t-} a_s e^{-b_s(t-s)} dN_s. \quad (5.3)$$

Moreover, conditionally on the path of  $\theta_t$ ,  $N_t$  is distributed as a standard Hawkes process with inhomogeneous deterministic parameter  $\theta_t$ , that is

$$\mathbb{E} \left[ f(N) \mid \mathcal{F}_T^\theta \right] = \mathbb{E} \left[ f \left( N^{\tilde{\theta}} \right) \right]_{|\tilde{\theta}=\theta}, \quad (5.4)$$

for any continuous bounded function  $f$ , and where  $N_t^{\tilde{\theta}}$  is a doubly stochastic Hawkes process with underlying deterministic process  $\tilde{\theta}_t$ .

From now on we assume that  $\theta_t^*$  satisfies Condition [E]. Under this assumption, since  $N_t^n$  is a time-varying self-exciting process with parameters  $(n\nu_t^*, n a_t^*, n b_t^*)$ ,  $N_t^n$  is well-defined and adapted to  $\mathcal{F}_t$ .

We describe the statistical procedure, provide a formal definition of the local MLE  $\widehat{\Theta}_{i,n}$  as well as its first order bias-corrected version  $\widehat{\Theta}_{i,n}^{(BC)}$ . We state their asymptotic properties, including the main result of this paper which is the GCLT for  $\widehat{\Theta}_n^{(BC)}$  in Theorem 5.4. Recall that we have chopped our

observations into  $B_n$  time blocks of the form  $((i-1)\Delta_n, i\Delta_n]$ . For any  $i \in \{1, \dots, B_n\}$  and any  $\theta \in K$ , we consider the regression family of intensities

$$\lambda^{i,n}(t, \theta) = n\nu + \int_{(i-1)\Delta_n}^{t-} nae^{-nb(t-s)} dN_s^n, \quad (5.5)$$

defined for  $t \in ((i-1)\Delta_n, i\Delta_n]$ . We now define the Quasi Log Likelihood<sup>3</sup> on the  $i$ -th block as

$$l_{i,n}(\theta) = \int_{(i-1)\Delta_n}^{i\Delta_n} \log(\lambda^{i,n}(t, \theta)) dN_t^n - \int_{(i-1)\Delta_n}^{i\Delta_n} \lambda^{i,n}(t, \theta) dt. \quad (5.6)$$

We take the local MLE  $\widehat{\Theta}_{i,n}$  as one maximizer of the Quasi Log Likelihood on the  $i$ -th block defined as

$$l_{i,n}(\widehat{\Theta}_{i,n}) = \max_{\theta \in K} l_{i,n}(\theta). \quad (5.7)$$

Looking at the form of (5.5), (5.6) and (5.7), we can see that  $\lambda^{i,n}$ ,  $l_{i,n}$  and  $\widehat{\Theta}_{i,n}$  are functions of the  $i$ -th block's events<sup>4</sup>. In particular, we don't take account for the possible preexcitation induced by past events in the expression of the candidate intensity (5.5), as the lower bound of the integral is fixed to  $(i-1)\Delta_n$ . Asymptotically, such approximation is valid because the exponential form of the excitation kernel along with the order of the excitation parameters  $(na_t^*, nb_t^*)$  induce a weak-enough influence of the past events on the actual stochastic intensity  $\lambda_*^n(t)$ .

In what follows we specify the form of  $h_n$  and assume the existence of an exponent  $\delta > 1$  such that

$$h_n = n^{1/\delta}. \quad (5.8)$$

We will also have to specify the smoothness of the process  $\theta^*$  using the following quantities. First, define the regularity modulus of order  $p \in \mathbb{N} - \{0\}$ , at time  $t \in [0, T]$  and value  $\theta \in K$  as

$$w_p(t, \theta, r) = \mathbb{E} \left[ \sup_{h \in [0, r \wedge (T-t)]} |\theta_{t+h}^* - \theta_t^*|^p \middle| \mathcal{F}_t, \theta_t^* = \theta \right], \quad r > 0. \quad (5.9)$$

We then define the global regularity modulus as

$$w_p(r) = \sup_{(t, \theta) \in [0, T] \times K} w_p(t, \theta, r), \quad r > 0. \quad (5.10)$$

We introduce the following conditions needed to obtain the LCLT and the boundedness of moments.

[C]

(i) There exists an exponent  $\gamma \in (0, 1]$ , such that for  $r \rightarrow 0$ , we have

$$w_p(r) = O_{\mathbb{P}}(r^{\gamma p}). \quad (5.11)$$

(ii)  $\delta$  and  $\gamma$  satisfy the relation

$$\delta > 1 + \frac{1}{\gamma}. \quad (5.12)$$

---

<sup>3</sup>The model is by definition misspecified and thus  $l_{i,n}$  is not the log likelihood function of the model.

<sup>4</sup>Note that this doesn't mean that  $\widehat{\Theta}_{i,n}$  are uncorrelated.

(iii) The excitation parameters  $a_t^*$  and  $b_t^*$  satisfy

$$c := \sup_{(t,n) \in [0,T] \times \mathbb{N}} \int_0^t n a_s^* e^{-nb_s^*(t-s)} ds < 1 \quad \mathbb{P} - a.s. \quad (5.13)$$

Note that the conditional expectation  $\mathbb{E}[\cdot | \mathcal{F}_s, \theta_s^* = \theta]$  refers to the operator  $\mathbb{E}[\cdot | \mathcal{F}_s]$  conditioned on  $\theta_s^* = \theta$ . By definition, for a  $\mathcal{F}$ -measurable random variable  $X$ , if we write  $G_X(\theta) = \mathbb{E}[X | \mathcal{F}_s, \theta_s^* = \theta]$ , the relationship between both expectations can be expressed as  $\mathbb{E}[X | \mathcal{F}_s] = G_X(\theta_s^*)$ . The justification of the existence of  $\mathbb{E}[\cdot | \mathcal{F}_s, \theta_s^* = \theta]$  can be found in Section 10.3. Condition [C]-(i) quantifies the regularity of the process  $\theta_t^*$  through the regularity exponent  $\gamma$ . A natural example of a process satisfying [C]-(i) is the drift function, i.e. of the form

$$\theta_t^* = \theta_0^* + \int_0^t u_s^* ds, \quad (5.14)$$

where  $u^*$  is a stochastic process that takes its values in a compact subset of  $\mathbb{R}^3$ . Another example is a smoothed version of the Brownian motion that can be obtained as follows. Take some  $\tau > 0$ , a positive vector  $\theta^{(M)} \in \mathbb{R}^3$ , a positive diagonal matrix  $\sigma = \text{diag}(\sigma^\nu, \sigma^a, \sigma^b)$  and consider the process

$$\theta_t^* = \theta^{(M)} + \frac{\sigma}{\tau} \int_{t-\tau}^t W_s ds, \quad (5.15)$$

where  $(W_t)_{t \in [-\tau, T]}$  is a 3-dimensional standard Brownian motion. One can confine  $\theta_t^*$  in a compact space by stopping the process  $W$  when it reaches some critical value. This second example is useful to model the stochastic component of the parameter as a nuisance process, and we use (5.15) in our simulation study. Note that the smaller  $\tau$ , the less auto-correlated  $\theta_t^*$  will be, and that we would be back to a Brownian motion in the limit  $\tau \rightarrow 0$ . For both examples (5.14) and (5.15) we have  $\gamma = 1$ , but note that the correlation structure of  $\theta_t^*$  may be very complex though (to do so, we can take any process  $u_t^*$  which has a complex correlation structure).

Condition [C]-(ii) controls the lower bound of  $h_n$  and is necessary to derive the LCLT and the local boundedness of moments. In particular, as  $\gamma \leq 1$ , [C]-(ii) implies that  $h_n = o(\sqrt{n})$ . This was stated in (1.3). Finally [C]-(iii) is an additional condition that ensures the existence of moments of  $N^n$ . We can see that [C]-(iii) is automatically satisfied if  $\underline{a} \leq a^* \leq \bar{a}$ ,  $\underline{b} \leq b^* \leq \bar{b}$  and  $\bar{a} < \underline{b}$ .

We specify now the value of the exponent

$$\kappa = \gamma(\delta - 1) > 1, \quad (5.16)$$

where the inequality is a direct consequence of [C]-(ii). For  $\theta \in K$ , the positive symmetric matrix  $\Gamma(\theta)$  is defined as the asymptotic Fisher information of a parametric Hawkes process generated by  $\theta$  and can be found in (10.8). The next theorem encompasses the LCLT and the local convergence of moments of order smaller than  $2\kappa$  of the rescaled local MLE  $\sqrt{h_n} (\hat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^*)$ .

**Theorem 5.2.** *(LCLT and boundedness of moments) Let  $L \in [0, 2\kappa]$ . Under [C], we have uniformly in  $i \in \{1, \dots, B_n\}$  that*

$$\mathbb{E}_{(i-1)\Delta_n} \left[ f \left( \sqrt{h_n} (\hat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^*) \right) \right] = \mathbb{E}_{(i-1)\Delta_n} \left[ f \left( T^{-\frac{1}{2}} \Gamma(\theta_{(i-1)\Delta_n}^*)^{-\frac{1}{2}} \xi \right) \right] + o_{\mathbb{P}}(1) \quad (5.17)$$

for any continuous function  $f$  with  $|f(x)| = O(|x|^L)$  when  $|x| \rightarrow \infty$ , and such that  $\xi$  follows a standard normal distribution and is independent of  $\mathcal{F}$ .

We now introduce several observable quantities to correct for the first-order bias of  $\widehat{\Theta}_{i,n}$ . For  $i \in \{1, \dots, B_n\}$ , let

$$\Gamma_{i,n}(\theta) = -\frac{1}{h_n T} \partial_\theta^2 l_{i,n}(\theta) \quad (5.18)$$

and

$$K_{i,n}(\theta) = \frac{1}{h_n T} \partial_\theta^3 l_{i,n}(\theta). \quad (5.19)$$

Moreover, for any  $t \in ((i-1)\Delta_n, i\Delta_n]$ , we define

$$M_{i,n}(t, \theta) = \int_{(i-1)\Delta_n}^t \frac{\partial_\theta \lambda^{i,n}(s, \theta)}{\lambda^{i,n}(s, \theta)} \{dN_s^n - \lambda^{i,n}(s, \theta)ds\}. \quad (5.20)$$

We also define for any  $k, l, m \in \{0, 1, 2\}$ ,

$$C_{i,n}(\theta)_{k,lm} = \frac{1}{h_n T} \int_{(i-1)\Delta_n}^{i\Delta_n} \left\{ \partial_{\theta,k} \lambda^{i,n}(s, \theta) \partial_{\theta,lm}^2 \log \lambda^{i,n}(s, \theta) - [M_{i,n}(s, \theta)]_k \frac{[\partial_\theta \lambda^{i,n}(s, \theta)^{\otimes 2}]_{lm}}{\lambda^{i,n}(s, \theta)} \right\} ds.$$

Finally, for  $k \in \{0, 1, 2\}$ , we define the empirical bias function as

$$b_{i,n}(\theta)_k = \frac{1}{2} \Gamma_{i,n}(\theta)^{jk} \Gamma_{i,n}(\theta)^{lm} (K_{i,n}(\theta)_{jlm} + 2C_{i,n}(\theta)_{l,jm}), \quad (5.21)$$

where repeated indices are implicitly summed and superscripts denote elements of the inverse matrix when it exists, i.e. for a matrix  $A$ ,  $A^{jk}$  stands for the element in position  $(j, k)$  of  $A^{-1}$  if  $A$  admits an inverse,  $A^{jk} = 0$  otherwise. The first-order bias-corrected local MLE is defined for any  $i \in \{1, \dots, B_n\}$  as

$$\widehat{\Theta}_{i,n}^{(BC)} = \widehat{\Theta}_{i,n} - \frac{b_{i,n}(\widehat{\Theta}_{i,n})}{h_n T}. \quad (5.22)$$

Such expression for  $b_{i,n}$  should be compared to its very similar form for the classical i.i.d case, see e.g. [12]. We finally recall the definition of the global bias-corrected estimator that was introduced in (1.5), i.e.

$$\widehat{\Theta}_n^{(BC)} = \frac{1}{B_n} \sum_{i=1}^{B_n} \widehat{\Theta}_{i,n}^{(BC)}. \quad (5.23)$$

In the next theorem, the expression  $x \wedge y$  stands for  $\min\{x, y\}$ .

**Theorem 5.3.** (bias correction) *Let  $\epsilon \in (0, 1)$ . The bias of the estimator  $\widehat{\Theta}_{i,n}$  admits the expansion*

$$\mathbb{E}_{(i-1)\Delta_n} \left[ \widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^* \right] = \frac{b\left(\theta_{(i-1)\Delta_n}^*\right)}{h_n T} + O_{\mathbb{P}}\left(h_n^{-\epsilon(\kappa \wedge \frac{3}{2})}\right), \quad (5.24)$$

uniformly in  $i \in \{1, \dots, B_n\}$ , and where the non-random theoretical bias  $b$  can be found in (10.12). Moreover, the estimator  $\widehat{\Theta}_{i,n}^{(BC)}$  has the uniform bias expansion

$$\mathbb{E}_{(i-1)\Delta_n} \left[ \widehat{\Theta}_{i,n}^{(BC)} - \theta_{(i-1)\Delta_n}^* \right] = O_{\mathbb{P}}\left(h_n^{-\epsilon(\kappa \wedge \frac{3}{2})}\right). \quad (5.25)$$

Now our aim is to combine Theorem 5.2 and Theorem 5.3 to state the asymptotic properties of the global estimator. In the following there are two parts. The main one gives the GCLT when the parameter is assumed to be sufficiently smooth. The second part investigates what happens when the parameter is rough.

## 5.2 Global central limit theorem when parameters are smooth

In this section, we state an additional condition on  $\delta$  and  $\gamma$  so that  $\widehat{\Theta}_n^{(BC)}$  is asymptotically unbiased.

[BC]  $\delta$  and  $\gamma$  satisfy the relation

$$\frac{\gamma}{\gamma - \frac{1}{2}} < \delta < 3. \quad (5.26)$$

Intuitively, the left-hand side inequality in [BC] ensures that the size of each block is not too big so that the bias induced by the parameter process  $\theta_t^*$  itself is negligible. On the contrary, the right-hand side inequality is a sufficient condition to keep under control the finite-sample bias of the local MLE by avoiding too small blocks. More precisely, Condition [BC] implies in particular that the exponent  $\gamma \in (\frac{3}{4}, 1]$ . Note that such condition excludes the class of Itô-processes as a parameter process. Moreover, on  $(\frac{3}{4}, 1]$  we have  $\frac{\gamma}{\gamma - \frac{1}{2}} \geq 1 + \frac{1}{\gamma}$  with equality for  $\gamma = 1$ , and thus [BC] is a stronger condition than [C]-(ii). For instance, in the Lipschitz case  $\gamma = 1$ , [BC] (and thus [C]-(ii)) are satisfied for  $2 < \delta < 3$ . This means by definition of  $\delta$  that  $h_n$  must be taken so that  $n^{\frac{1}{3}} = o(h_n)$  and  $h_n = o(n^{\frac{1}{2}})$ .

We finally state the main result of this work which investigates the limit error of the bias-corrected estimator  $\widehat{\Theta}_n^{(BC)}$ .

**Theorem 5.4.** (GCLT) *Assume that [C] and [BC] hold. Then, stably in law as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\widehat{\Theta}_n^{(BC)} - \Theta) \rightarrow \left( T^{-2} \int_0^T \Gamma(\theta_s^*)^{-1} ds \right)^{\frac{1}{2}} \mathcal{N}(0, 1), \quad (5.27)$$

where  $\mathcal{N}(0, 1)$  is independent from the filtration  $\mathcal{F}$ .

**Remark 5.5.** (convergence rate) *The convergence rate in Theorem (5.4) is the same as in the parametric case. We also conjecture that the asymptotic variance is the nonparametric efficient bound.*

**Remark 5.6.** (robustness to jumps in the parameter process) *We assume that we add a jump component to the parameter process*

$$\theta_t^* = \theta_t^{(C)} + \theta_t^{(J)}, \quad (5.28)$$

where  $\theta_t^{(J)}$  denotes a 3-dimensional finite activity jump process and  $d\theta_t^{(J,k)}$  is either zero (no jump) or a real number indicating the size of the jump at time  $t$  for  $k = 1, 2, 3$ . We further assume that there is no initial jump, i.e.  $J_0 = 0$ . Moreover, we assume that  $J_t$  is a general Poisson process independent from the other quantities. Under similar assumptions, the results of this work can be adapted.

**Remark 5.7.** (mutually exciting process) *The proofs can be adapted to a multidimensional Hawkes process. Investigating the corresponding conditions is beyond the scope of this paper.*

## 5.3 What happens in the rough parameter case ?

In this section, we are interested in the asymptotic properties of our estimators when the regularity condition  $\gamma \in (\frac{3}{4}, 1]$  fails. We first give a theoretical argument to show that the bound 3/4 can be lowered to 1/2, but that the corresponding bias theoretical formula would be too involved to be of any practical interest. Nonetheless the bias can be computed effortlessly with Monte Carlo methods

(see Section 7.1 in our numerical study for more details). We then give what convergence rate can be expected from both the naive and the first order bias-corrected estimators in terms of consistency.

When  $\gamma \notin (\frac{3}{4}, 1]$ , Theorem 5.4 fails in general. This is due to the bias expansion obtained in Theorem 5.3, (5.25), whose order in  $h_n^{-\kappa \wedge \frac{3}{2}}$  can be dominated by  $n^{-\frac{1}{2}}$  only if  $\gamma > 3/4$ . Nevertheless, one may expect that correcting for the bias to a higher order improves the rate of convergence in (5.25) and thus leads to a central limit theorem even for  $\gamma \leq 3/4$ . A closer investigation to the proofs shows that if one conducts the bias correction up to order  $q \in \mathbb{N} - \{0\}$ , conditions [C]-(ii) and [BC] are respectively replaced by  $\delta > 1 + (q+1)/(2\gamma)$  and  $\gamma/(\gamma-1/2) < \delta < 2+q$ , so that the GCLT becomes valid under the weaker condition  $\gamma \in \left(\frac{1}{2} + \frac{1}{2(1+q)}, 1\right]$ . For  $q \rightarrow +\infty$ , the asymptotic admissible interval becomes thus  $(\frac{1}{2}, 1]$ , so it is theoretically possible to construct an asymptotically normal estimator for any  $\gamma > 1/2$ . However, the expression of the successive terms in the bias expansion being very involved, such option is not realistic from a practical point of view.

When  $\gamma \in (0, \frac{1}{2}]$ , it is no longer possible to carry out the martingale approach that we have taken in this paper. In general, the bias induced by the parameter process in the expansion (5.25) cannot be corrected without further knowledge on the structure of  $\theta_t^*$ . In that case, there is no hope to render such bias negligible compared to  $n^{-1/2}$ , because this would imply a choice of  $\delta$  such that  $\gamma < \delta(\gamma-1/2)$  which is clearly impossible if  $\gamma$  reaches the critical value  $\gamma = 1/2$ . Investigating if other approaches yield a better estimate of the bias under additional specification on the structure of  $\theta_t^*$  is beyond the scope of this work.

We now turn to the convergence rate of our estimators when the central limit theorem fails. Both estimators  $\widehat{\Theta}_n$  and  $\widehat{\Theta}_n^{(BC)}$  prove to be consistent. Indeed, it turns out that the first order bias corrected version is  $n^\alpha$ -consistent<sup>5</sup> for any  $\alpha \in (0, \frac{\gamma}{1+\frac{2}{3}\gamma})$ , whereas the naive estimator is only  $n^\alpha$ -consistent for any  $\alpha \in (0, \frac{\gamma}{1+\gamma})$ . More specifically, we have the following result, that can be easily showed following a similar path as for the proof of Theorem 5.4.

**Proposition 5.8.** *(consistency) For any  $\alpha \in (0, \frac{\gamma}{1+\gamma})$ , the choice  $\delta \in \left(1 + \frac{1}{\gamma}, \frac{1}{\alpha}\right)$  gives*

$$n^\alpha(\widehat{\Theta}_n - \Theta) \rightarrow^{\mathbb{P}} 0.$$

*Moreover, if  $\gamma \in (0, 3/4]$ , for any  $\alpha \in (0, \frac{\gamma}{1+\frac{2}{3}\gamma})$ , the choice  $\delta \in \left((1 + \frac{1}{\gamma}) \vee \frac{\gamma}{\gamma-\alpha}, \frac{3}{2\alpha}\right)$  gives*

$$n^\alpha(\widehat{\Theta}_n^{(BC)} - \Theta) \rightarrow^{\mathbb{P}} 0.$$

In particular it is interesting to note that  $\widehat{\Theta}_n$  is already almost  $\sqrt{n}$ -consistent when  $\gamma = 1$  without any bias correction. In a similar way, the case  $\gamma = 3/4$  also yields an almost  $\sqrt{n}$ -consistent bias corrected estimator  $\widehat{\Theta}_n^{(BC)}$  as was expected. Again, knowing if the bounds for  $\alpha$  given in Proposition 5.8 are optimal is beyond the scope of this paper.

## 6 Statistical implementation

In this section, we give some practical guidance to the above theory including a studentized version of the GCLT. Actually, on real data, the quantity of interest is  $n\Theta$  whereas  $n$  is (usually) unknown. This doesn't prevent us from obtaining a studentized version of the GCLT. A feasible procedure consists in estimating directly  $n\Theta$  in place of estimating  $\Theta$ . When properly divided by  $n$ , this yields the same

---

<sup>5</sup>An estimator  $\bar{\Theta}_n$  is said to be  $a_n$ -consistent if  $a_n(\bar{\Theta}_n - \Theta) = O_{\mathbb{P}}(1)$ .

estimate as the non feasible procedure, i.e. we have  $n\widehat{\Theta} = \widehat{n\Theta}$ , where  $\widehat{\Theta}$  is the naive or the bias-corrected estimator. Indeed note that a maximizer  $\widehat{\Theta}_{i,n}$  of  $l_{i,n}(\theta)$  is equal to  $n^{-1}\tilde{\Theta}_{i,n}$ , where  $\tilde{\Theta}_{i,n}$  is a maximizer of  $l_{i,n}(n^{-1}\theta)$ , which corresponds to the ordinary quasi-likelihood (i.e. with disregard for the actual value of  $n$ ).

Now we provide an estimator (up to a scaling factor) of the asymptotic variance  $V_T = T^{-2} \int_0^T \Gamma(\theta_s^*)^{-1} ds$ , which also requires no information on the value of  $n$ . For any  $i \in \{1, \dots, B_n\}$ , we estimate the contribution of the  $i$ -th block by the formula

$$\widehat{C}_{i,n} := - \left[ \partial_\xi^2 l_{i,n} (n^{-1}\xi) \Big|_{\xi=n\widehat{\Theta}_{i,n}} \right]^{-1}. \quad (6.1)$$

The term  $\partial_\xi^2 l_{i,n} (n^{-1}\xi)$  doesn't depend on  $n$  (when  $h_n$  is chosen) and corresponds precisely to the Hessian matrix at point  $\xi$  of the likelihood function of a Hawkes model when one disregards the value of  $n$ . In particular, this implies that  $\widehat{C}_{i,n}$  can be computed. The asymptotic variance is then estimated, up to a scaling factor, as the weighted sum

$$\widehat{C}_n = \frac{1}{B_n^2} \sum_{i=1}^{B_n} \widehat{C}_{i,n}. \quad (6.2)$$

The next proposition states the consistency of  $n^{-1}\widehat{C}_n$  towards  $V_T$  along with a corresponding studentized version of Theorem 5.4, which is a corollary to the stable convergence in the GCLT.

**Proposition 6.1.** *We have*

$$n^{-1}\widehat{C}_n \xrightarrow{\mathbb{P}} V_T.$$

Moreover, we have the convergence in distribution

$$\frac{n\widehat{\Theta}_n^{(BC)} - n\Theta}{\sqrt{\widehat{C}_n}} \xrightarrow{\mathcal{N}(0, 1)}. \quad (6.3)$$

Note that  $\widehat{C}_n$  is the asymptotic variance of the dispersion between the estimated value of the scaled integrated parameter  $n\widehat{\Theta}_n^{(BC)}$  and the target  $n\Theta$  itself. In particular feasible asymptotic confidence intervals can be constructed from the data.

As the value of  $n$  is unknown, which value to choose for  $h_n$ ? One idea is to normalize the value of  $\theta_t^*$  so that the expected number of events between 0 and  $T$  is roughly one when parameters are equal to  $\theta_t^*$  (by analogy with other models in high frequency data where  $n$  corresponds exactly to the size of the sample data, as when estimating volatility from log-price returns observed regularly at times  $iT/n$ ). This amounts to taking  $n \approx N_T$  in practice. Although not perfect this provides guidance to the choice of  $h_n$ , which is assumed to be  $n^{1/3} = o(h_n)$  and  $h_n = o(n^{1/2})$  for a regular process ( $\gamma = 1$ ). In our numerical study, we have  $N_T \approx 27,300$  which amounts to taking  $n = 27,300$ . This gives us  $n^{1/3} \approx 30$  and  $n^{1/2} \approx 165$ , and correspondingly we look at different  $h_n = 136.5, 273, 546$  which are of the same order. In our empirical study, we consider  $h_n = \sqrt{n}, 2\sqrt{n}, 4\sqrt{n}, 8\sqrt{n}, 16\sqrt{n}$ .

## 7 Numerical simulations

### 7.1 Goal of the study

In this section, we report the numerical results which assess the central limit theory of

$$Z_n^{(BC)} = \frac{n\widehat{\Theta}_n^{(BC)} - n\Theta}{\sqrt{\widehat{C}_n}} \xrightarrow{\mathcal{N}(0, 1)}$$

in a finite sample context for several time-varying parameter models. In addition, we report the behavior of the studentized naive estimator

$$Z_n = \frac{n\widehat{\Theta}_n - n\Theta}{\sqrt{\widehat{C}_n}}.$$

Finally, we compare the performance of  $\widehat{\Theta}_n^{(BC)}$  and  $\widehat{\Theta}_n$  with two concurrent methods which are

1. The MLE on  $[0, T]$  when considering that the parameters are not time-varying on  $[0, T]$ .
2. The time-varying baseline intensity MLE from [8] (CH) that assumes that  $\nu_t^* = f(t, \theta)$  with  $f$  being a polynomial of order 3. More specifically in this setting the MLE estimates  $(\theta, a, b)$  where  $a$  and  $b$  are assumed to be constant over time.

The local log-likelihood functions and local variance estimators are computed implementing the formula obtained in [32]. To compute the bias of  $\widehat{\Theta}_n^{(BC)}$ , we can either implement the function defined in (5.22) or carry out Monte-Carlo simulations to compute  $b_{i,n}(\theta)$  for any  $\theta$  prior to the numerical study. We choose the latter option as this allows to get also rid of bias terms which appears in the Taylor expansion in a higher order than 1. Indeed, although those terms vanish asymptotically, they can pop up in a finite sample context. To be more specific, we first compute the sample mean for a grid of parameter values  $\theta$  and a grid of block length  $\Delta$  with 100,000 Monte Carlo paths of the parametric model, that we denote  $b(\theta, \Delta)$ . Then on each block, we estimate the bias by  $b(\widehat{\Theta}_{i,n}, \Delta_{i,n})$ .

## 7.2 Model design

We consider that  $T = 21,600$  seconds, which corresponds to the period of activity for one working day from 10am to 4pm. The market events are chosen to correspond to trades. The  $N_t^n$  process is generated using a time-varying version of the algorithm described in [32] (Section 4, pp. 148-149). The integrated parameter is set to  $n\Theta \approx (.8, 11, 30)$ , which are comparable values to our empirical results. This yields on average  $N_T^n \approx 27,300$  trades a day.

We consider three deterministic and one stochastic models for the time-varying parameter. The first two settings are toy models. Model I is a linear trend with  $\theta_t^* = \theta^{(M)} + A(-1 + 2\frac{t}{T})$ , where the non-random target value  $\theta^{(M)} = (.8, 11, 30)$  and the amplitude is set to  $A = (.5, 4, 10)$ . This means that  $\theta_t^*$  takes values in  $[.3, 1.3] \times [7, 15] \times [20, 40]$ , which is comparable to the daily variation in our empirical results. In Model II  $\theta_t^*$  oscillates around  $\theta^{(M)} = (.8, 11, 30)$  and has the form  $\theta_t^* = \theta^{(M)} + A \cos(\frac{t}{T}2\pi)$ , in particular implying that the range of taken values is the same as in Model I.

Model III is based on more realistic considerations and is taken directly from the literature. As pointed out in [16] (see discussions in Section 5-6 and Figure 2), the expected duration before the next trade tends to follow a U-shape intraday pattern. This diurnal effect motivated [8] (see Section 5, pp. 1011-1017) to model a Hawkes process where  $\nu_t^*$  is time-varying with a quadratic form. The model is written as  $\nu_t^* = e^{\beta_1} + \{e^{\beta_2} + e^{\beta_3}\}^2(t/T - e^{\beta_2}/(e^{\beta_2} + e^{\beta_3}))^2$ . We fit the model to the empirical intraday mean and find  $\beta_1 \approx -.84$ ,  $\beta_2 \approx -.26$  and  $\beta_3 \approx -.39$ , which implies that  $T^{-1} \int_0^T \nu_t^* dt \approx .61$ . The other two parameters  $(a_t^*, b_t^*) = (11, 30)$  are assumed to be constant.

Model IV is an extension of Model III based on our empirical findings. In this model, the three parameters are allowed to feature intraday seasonality with an additive stochastic component. We assume that

$$\begin{aligned} \nu_t^* &= e^{\beta_1^\nu} + \{e^{\beta_2^\nu} + e^{\beta_3^\nu}\}^2(t/T - e^{\beta_2^\nu}/(e^{\beta_2^\nu} + e^{\beta_3^\nu}))^2 + \sigma^\nu \tilde{W}_t^\nu, \\ a_t^* &= e^{\beta_1^a} + \{e^{\beta_2^a} + e^{\beta_3^a}\}^2(t/T - e^{\beta_2^a}/(e^{\beta_2^a} + e^{\beta_3^a}))^2 + \sigma^a \tilde{W}_t^a, \\ b_t^* &= e^{\beta_1^b} + \{e^{\beta_2^b} + e^{\beta_3^b}\}^2(t/T - e^{\beta_2^b}/(e^{\beta_2^b} + e^{\beta_3^b}))^2 + \sigma^b \tilde{W}_t^b, \end{aligned}$$

where  $(\beta_1^\nu, \beta_2^\nu, \beta_3^\nu) \approx (-0.84, -0.26, -0.39)$ ,  $(\beta_1^a, \beta_2^a, \beta_3^a) \approx (2.35, -0.05, 0.40)$ , and  $(\beta_1^b, \beta_2^b, \beta_3^b) \approx (3.66, -0.33, 0.67)$  were obtained when fitted to the respective parameter intraday mean. We also set  $(\sigma^\nu, \sigma^a, \sigma^b) = (0.8/(6T^{1/2}), 11/(6T^{1/2}), 30/(6T^{1/2}))$  and  $\tilde{W}_t = (\tilde{W}_t^\nu, \tilde{W}_t^a, \tilde{W}_t^b) = \int_{t-1}^t W_s ds$  with  $W_t = (W_t^\nu, W_t^a, W_t^b)$  a standard 3-dimensional Brownian motion. This means that the standard deviation of the noise factor is roughly equal to  $1/6$  the value of the parameter at time  $T$ . Also, we cap the possible value taken by  $\nu_t^*$  so that the intensity parameter stays bigger than 0.2. It is clear that in all those models the parameter is smooth enough to satisfy the assumptions of the GCLT.

Finally, we look at several values for  $h_n = 136.5, 273, 546$ , which correspond respectively to block lengths of size  $Th_n/n = 108, 216$  and  $432$  seconds. We have  $N_T \approx 27,300$  which means that we should take  $n = 27,300$  as explained in Section 6. According to the theory we would expect to need that  $h_n$  is of the same order as  $n^{1/3}$  and  $n^{1/2}$ , which are approximately equal to 30 and 165, thus our choices for  $h_n$  seem coherent with the theory.

### 7.3 Results

Table 1 shows the Monte Carlo results for the feasible statistic of the naive estimator. For all the models, the value of the bias is striking as it is of the same magnitude as the standard deviation. This indicates that the bias do play a crucial role in finite sample too.

Table 2 shows the result for the bias-corrected estimator. In this case, the sample mean is very close to 0 indicating that our proposed reduction method is working well. The standard deviation obtained for the intensity parameter is close to 1, but it is bigger for the other parameters  $a^*$  and  $b^*$ . Correspondingly, the asymptotics are slightly underestimating the mass of the distribution in the tails. The reason for this is probably that it is difficult to accurately estimate the variance of the parametric model on the small blocks.

Table 3 shows the performance of the estimators with concurrent approaches. It is clear that regardless of the model at hand the bias-corrected local approach performs better than the MLE and the CH. Both estimators are badly biased in case of misspecification for them. More surprisingly, although Model III follows a model included in [8] and thus CH performs a MLE with no misspecification in that specific case, the estimates are still biased (although very good for  $a^*$  and with a smaller standard deviation). This is most likely explained by the fact that the model tries to fit a 3-degree polynomial on a linear model for  $\nu^*$ . In practice, the model will a priori not be linear and thus this is the reason why we haven't equipped the CH with a linear model.

Finally, we can see that the naive estimator is more biased when  $h_n$  is smaller, which is in line with what we expected. The bias-correction is performed better with bigger  $h_n$ , although too big of a  $h_n$  will tend to bias-correct less efficiently (due to the fact that parameters are moving too much on a bigger block). This can be seen in Table 3 as the bias-corrected estimator seems to perform slightly better with 4 minute block than 7 minute block.

## 8 Empirical study

In this section, we implement local MLE on intraday transaction (corresponding to trade) times of Apple (APPL) shares carried out on the National Association of Securities Dealers Automated Quotation (NASDAQ) in 2015. Our aim is twofold. First, using relatively large (30 minute) local blocks, we document about seasonality and intraday stochasticity in the parameters. Second, we implement the naive and the bias-corrected estimator. We exclude January 1, the day after Thanksgiving and December 24 which are less active. This leaves us with 251 trading days of data. To prevent from opening and closing effect, we consider transactions that were carried out between 9:30am and 3:30pm,

Table 1: Finite sample properties of  $Z_n$  for several models<sup>†</sup>

Param.	Mean	Stdv.	RMSE	0.5%	2.5%	5%	95%	97.5%	99.5%
Model I									
$\nu^*$	0.69	1.02	1.23	0.00	0.40	1.00	82.60	89.70	97.00
$a^*$	0.77	1.10	1.34	0.40	1.00	1.50	78.00	86.50	95.90
$b^*$	1.37	1.24	1.85	0.20	0.60	1.00	58.20	67.60	83.90
Model II									
$\nu^*$	0.71	1.02	1.24	0.00	0.60	1.10	81.20	89.40	97.60
$a^*$	0.71	1.14	1.34	0.30	1.40	2.10	79.50	86.70	94.30
$b^*$	1.33	1.27	1.83	0.10	0.40	1.10	60.00	69.60	85.40
Model III									
$\nu^*$	0.80	1.05	1.32	0.10	0.30	0.90	78.40	86.30	95.30
$a^*$	0.78	1.14	1.38	0.00	1.20	1.70	78.10	85.70	94.10
$b^*$	1.43	1.24	1.89	0.00	0.20	0.40	55.70	66.10	80.60
Model IV									
$\nu^*$	0.83	0.99	1.29	0.00	0.20	0.50	79.70	87.10	96.30
$a^*$	1.05	1.04	1.48	0.00	0.60	0.90	71.60	80.90	93.00
$b^*$	1.62	1.10	1.96	0.00	0.10	0.20	52.20	63.20	79.70

<sup>†</sup>This table shows summary statistics and empirical quantiles benchmarked to the  $N(0,1)$  distribution for the feasible Z-statistics related to the naive estimator with  $h_n = 273$  (which corresponds to a 216 second block length). The simulation design is Model I-IV with  $M = 1,000$  Monte-Carlo simulations.

 Table 2: Finite sample properties of  $Z_n^{(BC)}$  for several models<sup>†</sup>

Param.	Mean	Stdv.	RMSE	0.5%	2.5%	5%	95%	97.5%	99.5%
Model I									
$\nu^*$	-0.01	1.02	1.02	0.40	2.80	5.70	94.70	97.20	99.30
$a^*$	0.00	1.12	1.12	1.10	4.00	7.60	93.50	96.70	99.10
$b^*$	-0.02	1.30	1.30	2.60	6.60	10.70	90.10	93.40	98.60
Model II									
$\nu^*$	0.02	1.02	1.02	0.90	2.70	5.20	95.70	98.10	99.40
$a^*$	-0.07	1.16	1.16	2.00	5.00	9.00	92.10	95.70	99.00
$b^*$	-0.08	1.32	1.33	4.20	7.60	11.50	90.50	94.30	98.20
Model III									
$\nu^*$	0.00	1.05	1.05	0.10	2.90	6.20	94.30	96.90	99.50
$a^*$	-0.02	1.15	1.15	1.90	4.80	8.80	91.00	95.90	99.00
$b^*$	-0.06	1.29	1.30	4.00	7.30	11.00	90.80	94.50	98.30
Model IV									
$\nu^*$	0.07	0.99	1.00	0.30	2.30	4.60	94.20	97.60	99.70
$a^*$	-0.04	1.05	1.05	1.20	3.50	5.40	94.90	97.30	99.30
$b^*$	-0.07	1.15	1.15	1.40	5.70	9.10	92.00	95.80	99.30

<sup>†</sup>This table shows summary statistics and empirical quantiles benchmarked to the  $N(0,1)$  distribution for the feasible Z-statistics related to the bias-corrected estimator with  $h_n = 273$  (which corresponds to a 216 second block length). The simulation design is Model I-IV with  $M = 1,000$  Monte-Carlo simulations.

Table 3: Performance of 8 estimators for several models<sup>†</sup>

Est.	$\nu^*$		$a^*$		$b^*$	
	Mean	Stdv.	Mean	Stdv.	Mean	Stdv.
Model I						
naive 2m	0.009	0.007	0.286	0.196	1.239	0.630
BC 2m	0.000	0.007	-0.012	0.201	-0.124	0.658
naive 4m	0.005	0.007	0.134	0.189	0.546	0.495
BC 4m	0.000	0.007	0.002	0.192	0.005	0.503
naive 7m	0.002	0.007	0.068	0.188	0.271	0.477
BC 7m	0.000	0.007	0.004	0.189	0.010	0.481
MLE	-0.011	0.006	0.489	0.198	0.485	0.494
CH	0.018	0.010	0.424	0.438	1.378	0.942
Model II						
naive 2m	0.009	0.007	0.287	0.213	1.346	0.734
BC 2m	0.000	0.007	-0.018	0.218	-0.078	0.744
naive 4m	0.005	0.007	0.126	0.198	0.538	0.516
BC 4m	0.000	0.007	-0.009	0.201	-0.019	0.525
naive 7m	0.002	0.007	0.057	0.196	0.245	0.490
BC 7m	0.000	0.007	-0.009	0.197	-0.022	0.494
MLE	-0.017	0.006	0.708	0.214	0.666	0.516
CH	-0.063	0.012	0.294	0.443	-0.265	1.097
Model III						
naive 2m	0.009	0.006	0.348	0.235	1.474	0.734
BC 2m	0.000	0.006	-0.009	0.241	-0.108	0.742
naive 4m	0.005	0.005	0.158	0.227	0.645	0.568
BC 4m	0.000	0.005	-0.003	0.241	-0.015	0.578
naive 7m	0.002	0.006	0.074	0.225	0.316	0.543
BC 7m	0.000	0.006	-0.004	0.227	-0.020	0.548
MLE	-0.004	0.006	-0.081	0.220	-0.572	0.519
CH	-0.009	0.005	-0.002	0.213	-0.082	0.454
Model IV						
naive 2m	0.009	0.006	0.624	0.332	3.533	2.579
BC 2m	0.000	0.006	-0.011	0.311	-0.199	2.333
naive 4m	0.005	0.006	0.276	0.274	1.389	1.022
BC 4m	0.000	0.006	-0.006	0.270	-0.008	0.967
naive 7m	0.003	0.006	0.133	0.265	0.655	0.889
BC 7m	0.000	0.006	0.000	0.265	0.013	0.876
MLE	-0.004	0.006	0.005	0.286	0.924	0.959
CH	-0.012	0.009	-0.857	0.699	-3.802	2.430

<sup>†</sup>This table shows the statistic  $\widehat{\Theta} - \Theta$  where  $\widehat{\Theta}$  is equal to the naive estimator and the bias-corrected estimator with  $h_n = 136.5, 273, 546$  (this corresponds respectively to a 108 second block (roughly 2 minutes), 216s (roughly 4m) and 432s (roughly 7m)), the MLE and the CH. The simulation design is Model I-IV with  $M = 1,000$  Monte-Carlo simulations.

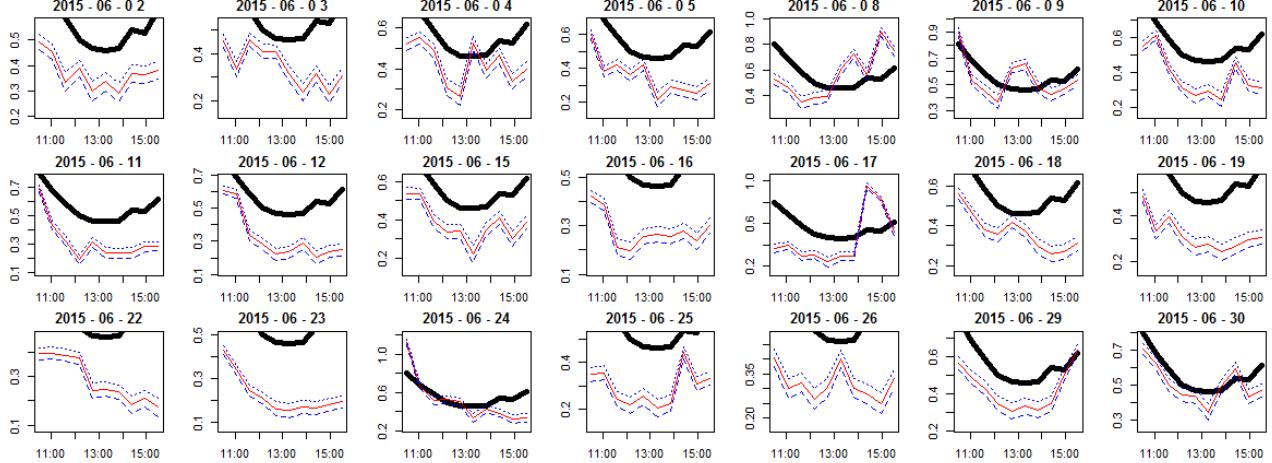


Figure 1: Local estimated  $\nu^*$  parameter on 30 minute long blocks in June 2015. The two dashed lines correspond to the 95% confidence intervals. The thick line stands for the seasonality intraday effect, estimated as a temporal local mean across all the trading days in 2015.

which corresponds to 5 full hours of trading. The number of daily trades is on average 15,000 with more than 50,000 trades for the most active days and slightly more than 3,000 for the least active days.

In Figure 1-3, we document the intraday variation of the three parameters. To do that, we divide the 5 hours of trading into 10 blocks of 30 minutes. On each block, we fit the MLE and obtain the corresponding estimates. We also estimate the standard deviation, which allows us to build 95% confidence intervals. As reported in Table 1-2, the theoretical estimate tends to underestimate the standard deviation, and thus we could also use bootstrapping to correct for the finite sample effect, but the picture would be very similar. Given how volatile the estimates are with respect to their own confidence interval, it is clear that neither the parametric model nor the seasonal component model can be satisfactory to fit such data. This time-varying tendency of parameter intraday values was consistently observed across most of the trading days in 2015. The behavior is heterogeneous in the three parameters. The seasonal model seems to do a decent job for the intensity parameter<sup>6</sup> although the shape of the parameter is very particular for each different day. The seasonal tendency is less clear for the other two parameters.  $a^*$  tends to oscillate not too far around the seasonal path with a behavior which is day specific, whereas  $b^*$  can really go far off from one side or the other with no specific pattern. For all those reasons we believe that including in the model both a seasonal and a stochastic effect is more realistic.

In Table 4, we report statistics of the implemented estimators. Overall we find that the daily estimates are on average roughly equal to  $(0.56, 11, 40)$ , with a standard deviation around  $(0.24, 2, 8)$ . The results are in line with the numerical study. We implemented five levels of  $h_n = \sqrt{n}, 2\sqrt{n}, 4\sqrt{n}, 8\sqrt{n}, 16\sqrt{n}$  that we denote respectively the corresponding bias-corrected estimators BC 1-5. We can see that BC 2-4 are highly correlated, whereas BC 1 and BC 5 are slightly less correlated. This is probably due to the fact that  $h_n$  can be too small on non active days in the case of BC 1 and too big when considering BC 5. This shows that the local method seems robust to a wide range of possible tuning parameter  $h_n$ . Furthermore, the mean of the MLE and the CH are very different from the one of BC. This is most likely explained by the strong bias obtained in our numerical study. Among those two estimators it is not surprising to find that the MLE is more in line with the local estimates than the

<sup>6</sup>probably even better if we add a "day effect" in the model

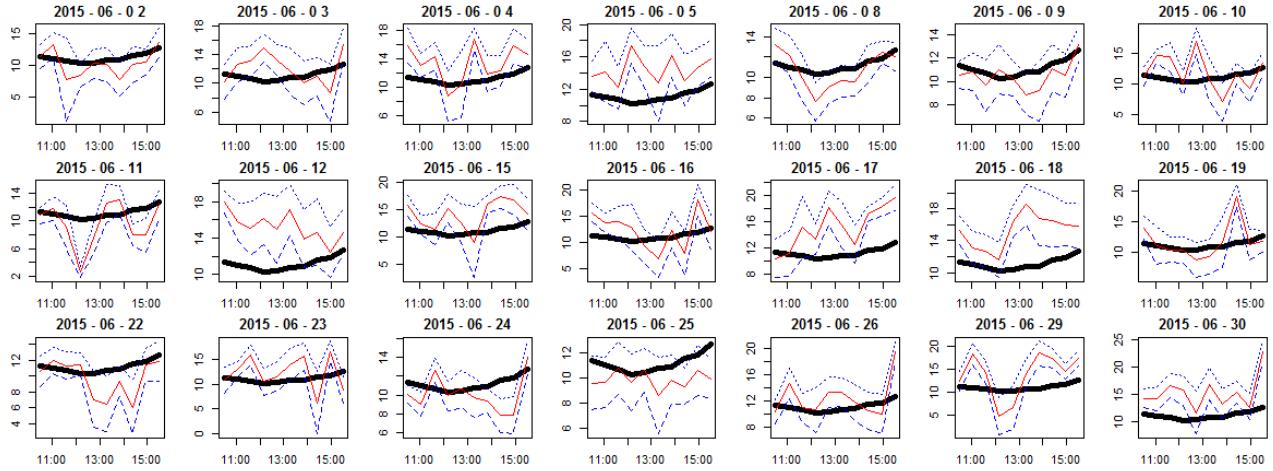


Figure 2: Local estimated  $a^*$  parameter on 30 minute long blocks in June 2015. The two dashed lines correspond to the 95% confidence intervals. The thick line stands for the seasonality intraday effect, estimated as a temporal local mean across all the trading days in 2015.

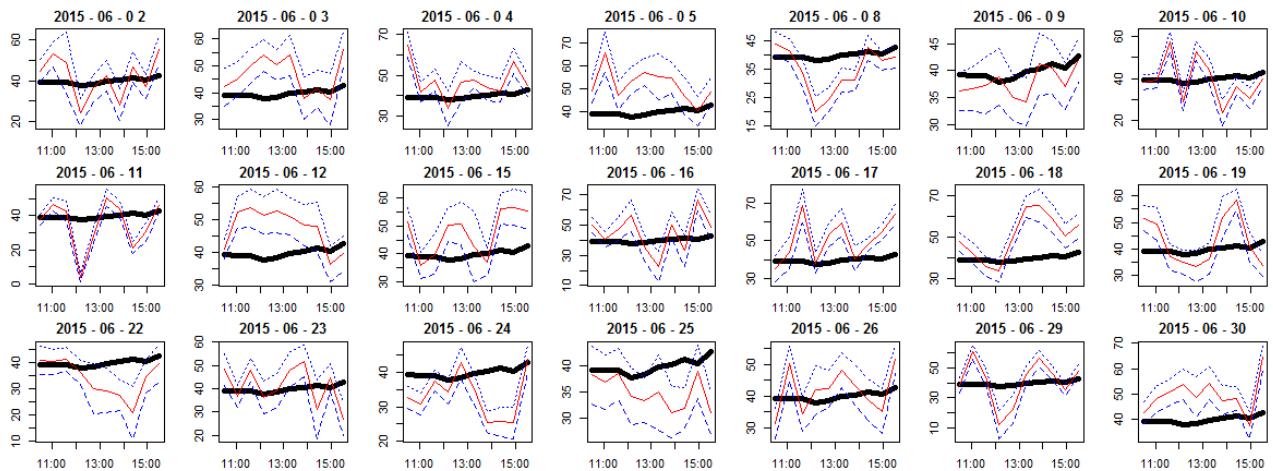


Figure 3: Local estimated  $b^*$  parameter on 30 minute long blocks in June 2015. The two dashed lines correspond to the 95% confidence intervals. The thick line stands for the seasonality intraday effect, estimated as a temporal local mean across all the trading days in 2015.

Table 4: Summary statistics for 12 estimators<sup>†</sup>

Est.	$\nu^*$			$a^*$			$b^*$		
	Mean	Stdv.	Corr.(,BC 3)	Mean	Stdv.	Corr.(,BC 3)	Mean	Stdv.	Corr.(,BC 3)
naive 1	0.57	0.24	$\approx 1$	12.48	2.18	0.94	49.80	10.33	0.78
BC 1	0.56	0.24	$\approx 1$	11.16	2.17	$\approx 1$	40.24	9.54	0.98
naive 2	0.57	0.24	$\approx 1$	11.78	2.00	0.98	44.74	8.20	0.89
BC 2	0.56	0.24	$\approx 1$	11.12	2.00	$\approx 1$	39.92	7.79	0.99
naive 3	0.56	0.24	$\approx 1$	11.42	1.96	0.99	42.14	7.58	0.92
BC 3	0.56	0.24	1	11.10	1.96	1	40.16	7.48	1
naive 4	0.56	0.24	$\approx 1$	11.22	1.96	$\approx 1$	40.66	7.40	0.96
BC 4	0.56	0.24	$\approx 1$	11.07	1.96	$\approx 1$	39.72	7.36	0.94
naive 5	0.56	0.24	$\approx 1$	11.07	1.97	0.99	39.60	7.50	0.94
BC 5	0.56	0.24	$\approx 1$	11.01	1.97	0.98	39.14	7.50	0.92
MLE	0.55	0.23	$\approx 1$	10.78	2.11	0.95	36.77	8.55	0.91
CH	0.55	0.22	0.99	11.69	1.44	0.63	40.50	4.85	0.63

<sup>†</sup>Sample mean, standard deviation and correlation with BC 3 for the naive estimators (naive 1-5) and the bias-corrected estimators (BC 1-5) with respectively  $h_n = \sqrt{n}, 2\sqrt{n}, 4\sqrt{n}, 8\sqrt{n}, 16\sqrt{n}$ , the MLE and the CH implemented for APPL in 2015.

CH as the MLE is a "local estimate" in the degenerate case  $h_n = n$ .

## 9 Conclusion

We have introduced a time-varying parameter extension of the Hawkes process with exponential exciting function. We have also provided an estimator, along with its central limit theorem, of the integrated parameter. We have seen on numerical simulations that this is of particular interest to the practitioner because some concurrent methods (e.g. MLE applied to all the observations) are biased. Finally, our empirical study points out the possible presence of stochastic variation in the parameter in addition to seasonal effects.

There are some questions left to explore such as what would happen to the local MLE in the case of a kernel with a fatter tail, such as a polynomial decreasing kernel. As far as the authors know, no convergence of moments of the rescaled MLE has been investigated even under the parametric case. Also, optimality of the tuning parameter  $h_n$  could be investigated, and we could potentially allow for time-varying tuning parameter.

Finally, we point out that the method can be extended to estimate more general key quantities than the integrated parameter, such as functional of the parameter  $T^{-1} \int_0^T f_s(\theta_s^*) ds$ . In particular, the GCLT for weighted versions  $T^{-1} \int_0^T \theta_s^* w_s ds$  where  $w_s$  is a weight process chosen by the practitioner may be derived by a similar reasoning.

## 10 Appendix

### 10.1 The standard MLE for the parametric Hawkes process

We briefly introduce the standard maximum likelihood estimation procedure for the parametric Hawkes process with exponential kernel  $\phi_t = ae^{-bt}$  in the long run (also called low-frequency) asymptotics, that is when we consider observations of a Hawkes process  $N^P$  on the time interval  $[0, T]$  with  $T \rightarrow \infty$ .

We define several deterministic key quantities, such as the Fisher information matrix, as time average limits of quantities which depend on the point process  $N_t^P$ .

The regression family is defined for each  $\theta \in K$  as

$$\lambda(t, \theta) = \nu + \int_0^{t-} ae^{-b(t-s)} dN_s^P. \quad (10.1)$$

We assume that there exists an unknown parameter  $\theta^* \in K$  such that the  $\mathcal{F}_t^{N^P}$ -intensity of  $N_t^P$  is expressed as

$$\lambda_*^P(t) = \lambda(t, \theta^*). \quad (10.2)$$

The log-likelihood process is, up to a constant term,

$$l_T(\theta) = \int_0^T \log(\lambda(t, \theta)) dN_t^P - \int_0^T \lambda(t, \theta) dt. \quad (10.3)$$

The MLE  $\hat{\theta}_T$  is a maximizer of  $l_T(\theta)$ . We define

$$\Gamma_T(\theta^*) = -\frac{1}{T} \partial_\theta^2 l_T(\theta^*) \in \mathbb{R}^{3 \times 3}, \quad (10.4)$$

$$K_T(\theta^*) = \frac{1}{T} \partial_\theta^3 l_T(\theta^*) \in \mathbb{R}^{3 \times 3 \times 3}, \quad (10.5)$$

$$M_T(\theta^*) = \int_0^T \frac{\partial_\theta \lambda(t, \theta^*)}{\lambda(t, \theta^*)} \{dN_t^P - \lambda(t, \theta^*) dt\}, \quad (10.6)$$

and for any indices  $k, l, m \in \{0, 1, 2\}$ ,

$$C_T(\theta^*)_{k,lm} = \frac{1}{T} \int_0^T \left\{ \partial_{\theta,k} \lambda(t, \theta^*) \partial_{\theta,lm}^2 \log(\lambda(t, \theta^*)) - [M_t(\theta^*)]_k \frac{[\partial_\theta \lambda(t, \theta^*)^{\otimes 2}]_{lm}}{\lambda(t, \theta^*)} \right\} dt, \quad (10.7)$$

where  $x^{\otimes 2}$  stands for the product  $xx^T$ . The three time-averaged quantities  $\Gamma_T$ ,  $K_T$  and  $C_T$  admit deterministic limiting values when  $T \rightarrow \infty$  because the process  $N^P$  is exponentially mixing. Indeed, a slight generalization of Lemma 6.6 in [10] shows that the vector process  $(\lambda(t, \theta^*), \partial_\theta(t, \theta^*), \dots, \partial_\theta^3(t, \theta^*))$  satisfies the mixing condition [M2] defined on p. 14 in the cited paper, which in turn implies the existence of  $\Gamma(\theta^*) \in \mathbb{R}^{3 \times 3}$ , and  $K(\theta^*)$ ,  $C(\theta^*) \in \mathbb{R}^{3 \times 3 \times 3}$  such that for any  $\epsilon \in (0, 1)$  and any integer  $p \geq 1$ ,

$$\mathbb{E} |\Gamma_T(\theta^*) - \Gamma(\theta^*)|^p = O\left(T^{-\epsilon \frac{p}{2}}\right), \quad (10.8)$$

$$\mathbb{E} |K_T(\theta^*) - K(\theta^*)|^p = O\left(T^{-\epsilon \frac{p}{2}}\right), \quad (10.9)$$

and

$$\mathbb{E} |C_T(\theta^*) - C(\theta^*)|^p = O\left(T^{-\epsilon \frac{p}{2}}\right), \quad (10.10)$$

where  $|x|$  stands for  $\sum_i |x_i|$  for any vector or a matrix  $x$ . Note that  $\Gamma(\theta^*)$  is the asymptotic Fisher information. In particular, [10] have shown the convergence of moments of the MLE (see Theorem 4.6),

$$\mathbb{E} \left[ f \left( \sqrt{T}(\hat{\theta}_T - \theta^*) \right) \right] \rightarrow \mathbb{E} \left[ f \left( \Gamma(\theta^*)^{-\frac{1}{2}} \xi \right) \right], \quad (10.11)$$

where  $f$  can be any continuous function of polynomial growth, and  $\xi$  follows a standard normal distribution. Also, it is easy to see that the convergences in (10.8), (10.9), (10.10) and finally (10.11) hold uniformly in  $\theta^* \in K$  under a mild change in the proofs of [10]. The result (10.11) should be compared to Theorem 5.2. Finally, from  $\Gamma$ ,  $K$  and  $C$  we define for any  $k \in \{0, 1, 2\}$

$$b(\theta^*)_k = \frac{1}{2} \Gamma(\theta^*)^{jk} \Gamma(\theta^*)^{lm} (K(\theta^*)_{jlm} + 2C(\theta^*)_{l,jm}) \quad (10.12)$$

with implicit summation of repeated indices. The function  $b$  appears in the expression of the expansion of the bias of the local MLE in Section 10.4.

## 10.2 Construction of the doubly stochastic Hawkes process

We establish the existence of the doubly stochastic self-exciting process under very general conditions on the parameter process. We also provide the boundedness of moments of various stochastic integrals with respect to such point process when the parameter is assumed to take its values in a compact space. We follow the same procedure as in [6] for the construction of a Hawkes process, that is, we show the existence of the doubly stochastic Hawkes process by a fixed point argument. In what follows we let  $\mathcal{B} = (\Omega, \mathcal{F}, \mathbf{F}, \mathbb{P})$ ,  $\mathbf{F} = (\mathcal{F}_t)_{t \in [0, T]}$ ,  $\mathcal{F} = \mathcal{F}_T$  be a stochastic basis such that the filtration  $\mathbf{F}$  is generated by the three-dimensional predictable process  $(\theta_s)_{s \in [0, T]} = (\nu_s, a_s, b_s)_{s \in [0, T]}$  which is component-wise non-negative, and by a Poisson process  $\bar{N}$  of intensity 1 on  $\mathbb{R}^2$  which is independent of  $\theta$ . In other words,  $\mathcal{F}_t = \mathcal{F}_t^{(\theta, \bar{N})}$ . In the following, properties such as predictability or adaptivity will automatically refer to  $\mathbf{F}$ . Before we turn to the existence of the self-exciting doubly stochastic process, we recall a key result for martingales.

**Lemma 10.1.** *Let  $\mathbf{F} = (\mathcal{F}_t)_{t \in [0, T]}$ ,  $\mathcal{F} = \mathcal{F}_T$  be a filtration and  $\mathcal{G}$  a  $\sigma$ -field that is independent of  $\mathcal{F}$ . Consider also the extended filtration defined by  $\mathcal{H}_t = \mathcal{F}_t \vee \mathcal{G}$ . Then any square integrable  $\mathcal{F}_t$ -martingale  $M$  is also a  $\mathcal{H}_t$ -martingale. In particular, for any  $\mathcal{H}_t$ -predictable process  $u$  such that  $\int_0^T u_s d\langle M, M \rangle_s$  is integrable,  $\mathbb{E}[\int_0^T u_s dM_s | \mathcal{G}] = 0$ .*

*Proof.* Let  $M$  defined as in the lemma and write for  $0 \leq s \leq t \leq T$ ,

$$\begin{aligned} \mathbb{E}[M_t | \mathcal{H}_s] &= \mathbb{E}[M_t | \mathcal{F}_s \vee \mathcal{G}] \\ &= \mathbb{E}[M_t | \mathcal{F}_s] \\ &= M_s, \end{aligned}$$

since  $\mathcal{G} \perp\!\!\!\perp M_t$  and  $\mathcal{G} \perp\!\!\!\perp \mathcal{F}_s$ . It follows that  $\int_0^t u_s dM_s$  is a  $\mathcal{H}_t$ -martingale, the second part of the lemma follows.  $\square$

We now show the existence of the doubly stochastic Hawkes process associated to  $\theta$ .

*Proof of Theorem 5.1.* We apply a fixed point argument using integrals over the two-dimensional integer measure  $\bar{N}(dt, dx)$ . Let us first define  $\lambda^0(t) = \nu_t$  and  $N^0$  the point process defined as

$$N_t^0 = \iint_{[0, t] \times \mathbb{R}} 1_{\{0 \leq x \leq \lambda^0(s)\}} \bar{N}(ds, dx). \quad (10.13)$$

It is immediate to see that  $\lambda^0(t)$  is the  $\mathcal{F}_t$ -intensity of  $N_t^0$ . We then define recursively the sequence of  $\mathcal{F}_t$ -adapted point processes  $N^n$  along with their stochastic intensities  $\lambda^n$  as

$$\lambda^{n+1}(t) = \nu_t + \int_0^{t^-} a_s e^{-b_s(t-s)} dN_s^n, \quad (10.14a)$$

$$N_t^{n+1} = \iint_{[0,t] \times \mathbb{R}} 1_{\{0 \leq x \leq \lambda^{n+1}(s)\}} \bar{N}(ds, dx). \quad (10.14b)$$

Note that both  $\lambda^n$  and  $N^n$  are increasing with  $n$  and thus both converge point-wise to some limiting values  $\lambda$  and  $N$  that take their values on  $[0, +\infty]$ . Moreover,  $N$  counts the points of  $\bar{N}$  which belong to the positive domain under the curve  $t \mapsto \lambda(t)$  by an immediate application of the monotone convergence theorem. Let's now introduce the sequence of processes  $\rho^n$  defined as  $\rho_t^n = \mathbb{E} [\lambda^n(t) - \lambda^{n-1}(t) | \mathcal{F}_T^\theta]$ . Then

$$\begin{aligned} \rho_t^{n+1} &= \mathbb{E} \left[ \int_0^t a_s e^{-b_s(t-s)} (\lambda^n(s) - \lambda^{n-1}(s)) ds \middle| \mathcal{F}_T^\theta \right] \\ &= \int_0^t a_s e^{-b_s(t-s)} \mathbb{E} [\lambda^n(s) - \lambda^{n-1}(s) | \mathcal{F}_T^\theta] ds \\ &= \int_0^t a_s e^{-b_s(t-s)} \rho_s^n ds, \end{aligned}$$

where we used Fubini's theorem in the second equality. Also, the first equality is obtained by Lemma 10.1 applied to the compensated measure  $\bar{N}(ds, dz) - ds \otimes dz$  and the independence between  $\mathcal{F}_T^{\bar{N}}$  and  $\mathcal{F}_T^\theta$ . Thus, setting  $\Phi_t^n = \int_0^t \rho_s^n ds$ , we have by Fubini's theorem

$$\Phi_t^{n+1} = \int_0^t \left\{ \int_0^{t-s} a_u e^{-b_u u} du \right\} \rho_s^n ds.$$

Note that  $\int_0^{t-s} a_u e^{-b_u u} du \leq \frac{a_s}{b_s} \leq r < 1$  by condition (5.1). Therefore,  $\Phi_t^{n+1} \leq r \Phi_t^n$ , and thus the application of the monotone convergence theorem to the sequence  $(\sum_{k=0}^n \Phi_t^k)_n$  yields

$$\mathbb{E} \left[ \int_0^t \lambda(s) ds \middle| \mathcal{F}_T^\theta \right] \leq \int_0^t \nu_s ds + r \mathbb{E} \left[ \int_0^t \lambda(s) ds \middle| \mathcal{F}_T^\theta \right]. \quad (10.15)$$

A straightforward rearrangement of the terms in (10.15) gives us that

$$\mathbb{E} \left[ \int_0^t \lambda(s) ds \middle| \mathcal{F}_T^\theta \right] \leq (1-r)^{-1} \int_0^t \nu_s ds < \infty \text{ } \mathbb{P} - a.s.$$

where the last inequality is a consequence of condition (5.2). In particular, we deduce that  $\int_0^t \lambda(s) ds$  and  $N_t$  are both finite almost surely. We need to show that  $\lambda(t)$  satisfies (5.3). By monotonicity, we deduce by taking the limit  $n \rightarrow +\infty$  in (10.14a) that

$$\lambda(t) = \nu_t + \int_0^{t^-} a_s e^{-b_s(t-s)} dN_s. \quad (10.16)$$

Finally, we show how to obtain (5.4). As  $\bar{N}$  and  $\mathcal{F}_T^\theta$  are independent, it still holds that conditioned on  $\mathcal{F}_T^\theta$ ,  $\bar{N}$  is a Poisson process of intensity 1. From the representation of  $N$  as an integral over  $\bar{N}$  we conclude that (5.4) holds, and this completes the proof.  $\square$

We now adapt well-known results on point processes to the case of the doubly stochastic Hawkes process, and derive some useful moments estimates for stochastic integrals with respect to  $N$ . Write  $\bar{\Lambda}$  the compensating measure of  $\bar{N}$ , that is  $\bar{\Lambda}(ds, dz) = ds \otimes dz$ . Given a predictable function  $W$ , write  $W * \bar{N}_t = \iint_{[0,t] \times \mathbb{R}} W(s, z) \bar{N}(ds, dz)$ , and the associated definition for  $W * \bar{\Lambda}_t$ . Predictable function and integral with respect to random measures definitions can be consulted in [26], paragraph II.1. The following lemma is a straightforward adaptation of Lemma I.2.1.5 in [24], using also Lemma 10.1 and (5.4).

**Lemma 10.2.** *Let  $W$  be a predictable function such that  $W^2 * \bar{\Lambda}_t < \infty$  almost surely. Then for any integer  $p > 1$ , there exists a constant  $K_p$  such that*

$$\begin{aligned} & \mathbb{E} \left[ \sup_{t \in [0, T]} |W * (\bar{N} - \bar{\Lambda})_t|^p \middle| \mathcal{F}_T^\theta \right] \\ & \leq K_p \mathbb{E} \left[ \iint_{[0, T] \times \mathbb{R}} |W(s, z)|^p ds dz + \left( \iint_{[0, T] \times \mathbb{R}} W(s, z)^2 ds dz \right)^{\frac{p}{2}} \middle| \mathcal{F}_T^\theta \right] \end{aligned}$$

For any (random) kernel  $\chi : (s, t) \rightarrow \chi(s, t)$ , we say that  $\chi$  is  $\mathbf{G}$ -predictable for some filtration  $\mathbf{G}$  if for any  $t \in [0, T]$  the process  $\chi(., t)$  is. For example the kernel  $\chi(s, t) = a_s e^{-b_s(t-s)}$  is  $\mathbf{F}^\theta$ -predictable. Nonetheless, we will also need to deal with other kernels in the course of the proofs. Consequently, we introduce the following lemma, which ensures the boundedness of moments of the doubly stochastic Hawkes process under the condition (5.13).

**Lemma 10.3.** *Under the condition  $c := \sup_{t \in [0, T]} \int_0^t a_s e^{-b_s(t-s)} ds < 1$   $\mathbb{P} - a.s.$ , the counting process  $N$  defined through (5.3) admits moments on  $[0, T]$  that can be bounded by values independent from  $T$ . Moreover, for any  $\mathbf{F}^\theta$ -predictable kernel  $\chi$  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  $\psi$  that has uniformly bounded moments independently from  $T$ , we have*

- (i)  $\sup_{t \in [0, T]} \mathbb{E} [\lambda(t)^p | \mathcal{F}_T^\theta]^{\frac{1}{p}} < Q_p$
- (ii)  $\sup_{t \in [0, T]} \mathbb{E} \left[ \left( \int_0^t \chi(s, t) dN_s \right)^p | \mathcal{F}_T^\theta \right]^{\frac{1}{p}} < Q_{p, \chi}$
- (iii)  $\sup_{t \in [0, T]} \left\| \int_0^t \psi_s dN_s \right\|_p < Q_{p, \psi} T$

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

*Proof.* We conduct the proof in three steps.

**Step 1.** We prove that (i) holds for  $p = 1$ . We write

$$\begin{aligned} \mathbb{E}[\lambda(t) | \mathcal{F}_T^\theta] &= \nu_t + \int_0^{t-} a_s e^{-b_s(t-s)} \mathbb{E}[\lambda(s) | \mathcal{F}_T^\theta] ds \\ &\leq \bar{\nu} + \sup_{s \in [0, t]} \mathbb{E}[\lambda(s) | \mathcal{F}_T^\theta] \int_0^{t-} a_s e^{-b_s(t-s)} ds \\ &\leq \bar{\nu} + c \sup_{s \in [0, t]} \mathbb{E}[\lambda(s) | \mathcal{F}_T^\theta], \end{aligned}$$

where we used condition (5.13) at the last step. Taking the supremum over  $[0, T]$  on both sides, we get

$$\sup_{t \in [0, T]} \mathbb{E}[\lambda(t) | \mathcal{F}_T^\theta] \leq (1 - c)^{-1} \bar{\nu}. \quad (10.17)$$

In particular this proves the case  $p = 1$ , since the right hand side of (10.17) is independent from  $T$ .

**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  $\epsilon > 0$ ,

$$\mathbb{E}[\lambda(t)^p | \mathcal{F}_T^\theta] \leq (1 + \epsilon^{-1})^{2^q-1} \bar{\nu} + (1 + \epsilon)^{2^q-1} \mathbb{E} \left[ \left( \int_0^{t-} a_s e^{-b_s(t-s)} dN_s \right)^p \middle| \mathcal{F}_T^\theta \right],$$

where we have used the inequality  $(x + y)^{2^q} \leq (1 + \epsilon)^{2^q-1} x^{2^q} + (1 + \epsilon^{-1})^{2^q-1} y^{2^q}$  for any  $x, y, \epsilon > 0$ . Now, for a fixed  $t \in [0, T]$ , define  $W(s, z) = a_s e^{-b_s(t-s)} 1_{\{0 \leq z \leq \lambda(s)\}}$ , and note that

$$\begin{aligned} \mathbb{E} \left[ \left( \int_0^{t-} a_s e^{-b_s(t-s)} dN_s \right)^p \middle| \mathcal{F}_T^\theta \right] &= \mathbb{E} \left[ (W * \bar{N}_t)^p \middle| \mathcal{F}_T^\theta \right] \\ &\leq (1 + \epsilon^{-1})^{2^q-1} \mathbb{E} \left[ (W * (\bar{N} - \bar{\Lambda})_t)^p \middle| \mathcal{F}_T^\theta \right] \\ &\quad + (1 + \epsilon)^{2^q-1} \mathbb{E} \left[ (W * \bar{\Lambda}_t)^p \middle| \mathcal{F}_T^\theta \right]. \end{aligned}$$

We apply now Lemma 10.2 to get

$$\begin{aligned} \mathbb{E} \left[ (W * (\bar{N} - \bar{\Lambda})_t)^p \middle| \mathcal{F}_T^\theta \right] &\leq K_p \mathbb{E} \left[ \iint_{[0, T] \times \mathbb{R}} |W(s, z)|^p ds dz + \left( \iint_{[0, T] \times \mathbb{R}} W(s, z)^2 ds dz \right)^{\frac{p}{2}} \middle| \mathcal{F}_T^\theta \right] \\ &= K_p \mathbb{E} \left[ \int_0^{t-} a_s^p e^{-pb_s(t-s)} \lambda(s) ds + \left( \int_0^{t-} a_s^2 e^{-2b_s(t-s)} \lambda(s) ds \right)^{\frac{p}{2}} \middle| \mathcal{F}_T^\theta \right]. \end{aligned}$$

We easily bound the first term by the induction hypothesis by some constant  $\frac{A_p}{2}$ . For the second term, an elementary application of Hölder's inequality shows that for any  $k > 1$  and any non-negative functions  $f, g$ ,  $(\int f g)^k \leq (\int f^k g)(\int g)^{k-1}$ . This along with the induction hypothesis leads to a similar bound for the second term. On the other hand, we have

$$\mathbb{E} \left[ (W * \bar{\Lambda}_t)^p \middle| \mathcal{F}_T^\theta \right] = \mathbb{E} \left[ \left( \int_0^{t-} a_s e^{-b_s(t-s)} \lambda(s) ds \right)^p \middle| \mathcal{F}_T^\theta \right].$$

We apply again the same Hölder's inequality as above with functions  $f(s) = \lambda(s)$  and  $g(s) = a_s e^{-b_s(t-s)}$  to get

$$\begin{aligned} \mathbb{E} \left[ \left( \int_0^{t-} a_s e^{-b_s(t-s)} \lambda(s) ds \right)^p \middle| \mathcal{F}_T^\theta \right] &\leq c^{p-1} \mathbb{E} \left[ \int_0^{t-} a_s e^{-b_s(t-s)} \lambda(s)^p ds \middle| \mathcal{F}_T^\theta \right] \\ &= c^{p-1} \int_0^{t-} a_s e^{-b_s(t-s)} \mathbb{E} \left[ \lambda(s)^p \middle| \mathcal{F}_T^\theta \right] ds \\ &\leq c^p \sup_{s \in [0, t]} \mathbb{E} \left[ \lambda(s)^p \middle| \mathcal{F}_T^\theta \right] \end{aligned}$$

Finally, we have shown that

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

This yields, taking supremum over the set  $[0, T]$  and taking  $\epsilon > 0$  small enough so that  $(1 + \epsilon)^{2^q} c^p < 1$ ,

$$\sup_{t \in [0, T]} \mathbb{E}[\lambda(t)^p | \mathcal{F}_T^\theta] (1 - (1 + \epsilon)^{2^q} c^p) \leq (1 + \epsilon^{-1})^{2^q-1} \bar{\nu} + (1 + \epsilon)^{2^q-1} (1 + \epsilon^{-1})^{2^q-1} A_p,$$

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  $\lambda$  along with Lemma 10.2.  $\square$

### 10.3 LCLT and boundedness of moments of order $2\kappa$

We focus on asymptotic properties of the local maximum likelihood estimator  $\hat{\Theta}_{i,n}$  of our model on each block  $i \in \{1, \dots, B_n\}$ . Recall that we are given the global filtration  $\mathcal{F}_t = \mathcal{F}_t^{(\theta^*, \bar{N})}$  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  $((i-1)\Delta_n T, i\Delta_n T]$ ,  $i \in \{1, \dots, B_n\}$  on the regression family of a parametric Hawkes process and show the local central limit theorem for every local estimator  $\hat{\Theta}_{i,n}$  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  $\sqrt{h_n}(\hat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^*)$  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.

**Lemma 10.4.** *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,  $N_t^\gamma$  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  $(\theta_s)_s$ ,  $N_t^\gamma$  has the distribution of a Hawkes process of parameter  $(\gamma\theta_{\gamma s})_s$ , that is,*

$$\lambda^\gamma(t) = \gamma\nu_{\gamma t} + \int_0^{t-} \gamma a_{\gamma s} e^{-\gamma b_{\gamma s}(t-s)} dN_s^\gamma. \quad (10.18)$$

*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  $u = \gamma^{-1}s$ ,

$$\begin{aligned} \lambda^\gamma(t) &= \gamma\lambda(\gamma t) \\ &= \gamma\nu_{\gamma t} + \int_0^{\gamma t-} \gamma a_s e^{-b_s(\gamma t-s)} dN_s \\ &= \gamma\nu_{\gamma t} + \int_0^{t-} \gamma a_{\gamma u} e^{-\gamma b_{\gamma u}(t-u)} dN_u^\gamma, \end{aligned}$$

and we are done.  $\square$

By virtue of Lemma 10.4, for any block index  $i \in \{1, \dots, B_n\}$ , 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  $[0, h_n T]$  by the formula  $N_t^{i,n} = N_{\tau_i^n(t)}^n - N_{(i-1)\Delta_n}^n$ . Such process is adapted to the filtration  $\mathcal{F}_t^{i,n} = \mathcal{F}_{\tau_i^n(t)}$ , for  $t \in [0, h_n T]$ . The parameter processes are now  $(\theta_t^{i,n,*})_{\{t \in [0, h_n T]\}} = (\theta_{\tau_i^n(t)}^*)_{\{t \in [0, h_n T]\}}$  whose canonical filtration can be expressed as  $\mathcal{F}_t^{\theta^{i,n,*}} = \sigma\{\theta_s^{i,n,*} | 0 \leq s \leq t\}$ , for  $t \in [0, h_n T]$ . Finally note that the  $\mathcal{F}_t^{i,n}$ -stochastic intensities are now of the form

$$\lambda_*^{i,n}(t) = \nu_t^{i,n,*} + \int_0^{t-} a_s^{i,n,*} e^{-b_s^{i,n,*}(t-s)} dN_s^n + R_{i,n}(t), \quad (10.19)$$

where  $R_{i,n}(t)$  is the  $\mathcal{F}_0^{i,n}$ -measurable residual process defined by the relation

$$R_{i,n}(t) = \int_0^{(i-1)\Delta_n} n a_s^* e^{-nb_s^*(\tau_i^n(t)-s)} dN_s^n. \quad (10.20)$$

$R_{i,n}(t)$  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

$$R_{i,n}(t) \leq e^{-bt} R_{i,n}(0) \quad (10.21)$$

Note that all the processes  $N^{i,n}$  can be represented as integrals over a sequence of Poisson processes  $\bar{N}^{i,n}$  of intensity 1 on  $\mathbb{R}^2$  as follows:

$$N_t^{i,n} = \iint_{[0,t] \times \mathbb{R}_+} 1_{\{0 \leq z \leq \lambda_*^{i,n}(s)\}} \bar{N}^{i,n}(ds, dz). \quad (10.22)$$

Indeed,  $\bar{N}^{i,n}$  is the time-space changed version of the initial Poisson process  $\bar{N}$  defined by  $\bar{N}^{i,n}(A \times B) = \bar{N}(\tau_i^n(A) \times nB)$  for  $A$  and  $B$  any two Borel sets of  $\mathbb{R}$ . In the time-changed representation, we define the regression family of stochastic intensities

$$\tilde{\lambda}^{i,n}(t, \theta) = \nu + \int_0^{t-} a e^{-b(t-s)} dN_s^{i,n}, \quad (10.23)$$

which is related to  $\lambda^{i,n}$  (see (5.5)) 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 (5.6) on the  $i$ -th block has now the representation (up to the constant term  $\log(n) N_{h_n T}^{i,n}$ )

$$l_{i,n}(\theta) = \int_0^{h_n T} \log(\tilde{\lambda}^{i,n}(t, \theta)) dN_t^{i,n} - \int_0^{h_n T} \tilde{\lambda}^{i,n}(t, \theta) dt, \quad (10.24)$$

Note that in our case, the true underlying intensity,  $\lambda_*^{i,n}$  does not belong to the regression family  $(\tilde{\lambda}^{i,n}(\cdot, \theta))_{\theta \in K}$  for two reasons : the parameter process  $\theta^*$  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 (10.19). We are in a misspecified case, but we wish to take advantage of the continuity of the process  $\theta^*$  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  $\theta^*$  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  $\bar{N}^{i,n}$  and the initial value  $\theta_0^{i,n,*}$ , whose stochastic intensity satisfies

$$\lambda^{i,n,c}(t) = \nu_0^{i,n,*} + \int_0^{t-} a_0^{i,n,*} e^{-b_0^{i,n,*}(t-s)} dN_s^{i,n,c}. \quad (10.25)$$

Moreover, we assume that  $N_t^{i,n,c}$  has the representation

$$N_t^{i,n,c} = \iint_{[0,t] \times \mathbb{R}_+} 1_{\{0 \leq z \leq \lambda^{i,n,c}(s)\}} \bar{N}^{i,n}(ds, dz). \quad (10.26)$$

Note that  $N_t^{i,n,c}$  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  $\mathbb{E}[\mathbf{1}_{\{R_{i,n}(0) \leq M_n\}} | \mathcal{F}_0^{i,n}, \theta_0^{i,n,*} = \theta_0]$ , that we denote by  $\mathbb{E}_{\theta_0, i, n}$ , and whose existences are justified by a classical regular distribution argument<sup>7</sup> (see for instance Section 4.3 (pp. 77–80) in [5]). In the same spirit, for a measurable set  $A \in \mathcal{F}$ ,  $\mathbb{P}_{\theta_0, i, n}[A]$  should be understood as  $\mathbb{E}_{\theta_0, i, n}[\mathbf{1}_A]$ . Finally we will need frequently to take supremum over the quadruplet  $(\theta_0, i, n, t)$ . For that reason we introduce the notation  $\mathbf{E} = \{(\theta_0, i, n, t) \in K \times \mathbb{N}^2 \times \mathbb{R}_+ \mid 1 \leq i \leq B_n \text{ and } 0 \leq t \leq h_n T\}$ . When  $n \in \mathbb{N}$  is fixed, we define  $\mathbf{E}_n$  the subset of  $\mathbf{E}$  as  $\mathbf{E}_n = \{(\theta_0, i, t) \in K \times \mathbb{N} \times \mathbb{R}_+ \mid 1 \leq i \leq B_n \text{ and } 0 \leq t \leq h_n T\}$ . In the same spirit, it is also useful when truncation arguments appear, to consider in the previous equation the subset of  $\mathbf{E}_n$  for which we have the stronger condition  $h_n^\alpha T \leq t \leq h_n T$  where  $\alpha \in (0, 1)$  that we denote by  $\mathbf{E}_n^\alpha$ . 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  $N^{i,n,c}$ .

**Lemma 10.5.** *We have, for any integer  $p \geq 1$  and any  $\mathbf{F}^{\theta^{i,n,*}}$ -predictable kernel  $\chi$  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)  $\sup_{(\theta_0, i, n, t) \in \mathbf{E}} \mathbb{E}_{\theta_0, i, n} |\lambda_*^{i,n}(t)|^p \leq M_p \mathbb{P}\text{-a.s.}$
- (ii)  $\sup_{(\theta_0, i, n, t) \in \mathbf{E}} \mathbb{E}_{\theta_0, i, n} \left| \int_0^t \chi(s, t) dN_s^{i,n} \right|^p < M_{p,\chi} \mathbb{P}\text{-a.s.}$
- (iii)  $\sup_{(\theta_0, i, n, t) \in \mathbf{E}} \mathbb{E}_{\theta_0, i, n} |\lambda^{i,n,c}(t)|^p < M_p \mathbb{P}\text{-a.s.}$
- (iv)  $\sup_{(\theta_0, i, n, t) \in \mathbf{E}} \mathbb{E}_{\theta_0, i, n} \left| \int_0^t \chi(s, t) dN_s^{i,n,c} \right|^p < M_{p,\chi} \mathbb{P}\text{-a.s.}$

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

*Proof.* This is a straightforward adaptation of the proof of Lemma 10.3, 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 (10.21) show clearly that the result still holds, uniformly in the quadruplet  $(\theta_0, i, n, t)$ . 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  $\mathbb{E}_{\theta_0, i, n}$ .  $\square$

<sup>7</sup>This is a consequence to the fact that  $K \subset \mathbb{R}^3$  is a Borel space.

Before we turn to estimating the distance between the two models, we state a technical lemma.

**Lemma 10.6.** *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  $t \geq 0$*

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

*Proof.* Iterating the inequality we get for any  $n \in \mathbb{N}^*$

$$f \leq g + g * \sum_{k=1}^n h^{*(k)} + f * h^{*(n+1)}. \quad (10.27)$$

We fix  $t \geq 0$ , 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

$$\begin{aligned} f * h^{*(n+1)}(t) &= \int_0^t f(t-s) \frac{s^n}{n!} a^{n+1} e^{-bs} ds \\ &\leq \frac{t^n}{n!} a^{n+1} \int_0^t f(s) ds \rightarrow 0 \end{aligned}$$

as  $n \rightarrow +\infty$ . We also have for any integer  $n \geq 1$

$$\begin{aligned} \sum_{k=1}^n h^{*(k)}(t) &= \sum_{k=1}^n \frac{t^{k-1}}{(k-1)!} a^k e^{-bt} \\ &\leq a e^{(a-b)t} \end{aligned}$$

and thus we get the result by taking the limit  $n \rightarrow +\infty$  in (10.27) evaluated at any point  $t \geq 0$ .  $\square$

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 (5.16), 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  $\gamma$  represents the regularity exponent in time of  $\theta$  while  $\delta$  controls the size of small blocks compared to  $n$  by the relation  $h_n = n^{1/\delta}$ . Note that by (5.16) 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 10.3.

**Lemma 10.7.** *Let  $\alpha \in (0, 1)$  be a truncation exponent, and  $\epsilon \in (0, 1)$ . We have, for any  $p \geq 1$ , any deterministic kernel  $\chi$  such that  $\int_0^t \chi(s, t) ds$  is bounded uniformly in  $t \in \mathbb{R}_+$ , and any predictable process  $(\psi_s)_{s \in \mathbb{R}_+}$  whose moments are bounded :*

- (i)  $\sup_{(\theta_0, i, t) \in \mathbf{E}_n^\alpha} \mathbb{E}_{\theta_0, i, n} \left| \lambda^{i,n,c}(t) - \lambda_*^{i,n}(t) \right|^p = O_{\mathbb{P}}(h_n^{-\kappa})$
- (ii)  $\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{E}_{\theta_0, i, n} \left| \int_{h_n^\alpha T}^{h_n T} \psi_s \{dN_s^{i,n,c} - dN_s^{i,n}\} \right|^p = O_{\mathbb{P}}(h_n^{p-\epsilon\kappa})$
- (iii)  $\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{E}_{\theta_0, i, n} \left| \int_{h_n^\alpha T}^{h_n T} \chi(s, h_n T) \{dN_s^{i,n,c} - dN_s^{i,n}\} \right|^p = O_{\mathbb{P}}(h_n^{-\kappa})$

**Remark 10.8.** 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  $\theta^*$  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 10.2). This is the same phenomenon as in the following fact. For a Poisson process  $N$  of intensity  $\lambda$ , we have  $\mathbb{E}[|N_t - \lambda t|^p] \sim \alpha_p t$  when  $t \rightarrow 0$ , 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  $n \in \mathbb{N}$ ,  $t \in [0, h_n T]$  and uniformly in  $(\theta_0, i)$ ,

$$\mathbb{E}_{\theta_0, i, n} |\lambda^{i, n, c}(t) - \lambda_*^{i, n}(t)|^{2^q} \leq L_{n, q} + M_{n, q} e^{-\underline{b}(1-r)t}, \quad (10.28)$$

where  $L_{n, q}$  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  $[h_n^\alpha T, h_n T]$  and using the estimate  $M_{n, q} e^{-\underline{b}(1-r)h_n^\alpha T} = o_{\mathbb{P}}(h_n^{-\kappa})$  we get

$$\mathbb{E}_{\theta_0, i, n} |\lambda^{n, c}(t) - \lambda_*^n(t)|^p = O_{\mathbb{P}}(h_n^{-\kappa})$$

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

**Step 1.** We show our claim in the case  $q = 0$ , that is  $p = 1$ . Write

$$\begin{aligned} |\lambda_*^i(t) - \lambda^{i, n, c}(t)| &\leq |\nu_t^{i, n, *} - \nu_0^{i, n, *} + \left| \int_0^{t^-} (a_s^{i, n, *} e^{-b_s^{i, n, *}(t-s)} - a_0^{i, n, *} e^{-b_0^{i, n, *}(t-s)}) dN_s^{i, n} \right| \\ &\quad + \left| \int_0^{t^-} a_0^{i, n, *} e^{-b_0^{i, n, *}(t-s)} (dN_s^{i, n, c} - dN_s^{i, n}) \right| + R_{i, n}(t) \\ &\leq A_{i, n}(t) + B_{i, n}(t) + C_{i, n}(t) + R_{i, n}(t) \end{aligned}$$

The (uniform) majoration  $\mathbb{E}_{\theta_0, i, n} A_{i, n}(t) = O_{\mathbb{P}}(h_n^{-\kappa})$  is an immediate consequence of [C]-**(i)**. By the inequality

$$|ae^{-bt} - a'e^{-b't}| \leq (|a - a'| + |b - b'|) e^{-bt} \quad (10.29)$$

for any  $(\nu, a, b), (\nu', a', b') \in K$ , we can write

$$\begin{aligned} \mathbb{E}_{\theta_0, i, n} B_{i, n}(t) &\leq \mathbb{E}_{\theta_0, i, n} \int_0^{t^-} (|a_s^{i, n, *} - a_0| + |b_s^{i, n, *} - b_0|) e^{-\underline{b}(t-s)} dN_s^{i, n} \\ &\leq \sqrt{\mathbb{E}_{\theta_0, i, n} \left| \sup_{s \in [0, t]} (|a_s^{i, n, *} - a_0| + |b_s^{i, n, *} - b_0|) \right|^2 \mathbb{E}_{\theta_0, i, n} \left| \int_0^{t^-} e^{-\underline{b}(t-s)} dN_s^{i, n} \right|^2}, \end{aligned}$$

where Cauchy-Schwartz inequality was applied in the last inequality. Note that the right term is almost surely dominated by a constant by Lemma 10.5 and thus the uniform majoration  $\mathbb{E}_{\theta_0, i, n} B_{i, n}(t) = O_{\mathbb{P}}(h_n^{-\kappa})$  follows from [C]-**(i)**. Finally, for  $C_{i, n}(t)$ , write

$$\mathbb{E}_{\theta_0, i, n} C_{i, n}(t) \leq \mathbb{E}_{\theta_0, i, n} \int_0^{t^-} a_0 e^{-b_0(t-s)} d|N^{i, n, c} - N^{i, n}|_s \quad (10.30)$$

where  $d|N^{i,n,c} - N^{i,n}|_s$  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  $t \rightarrow \lambda_*^{i,n}(t)$  and  $t \rightarrow \lambda^{i,n,c}(t)$ . A short calculation shows that this counting process admits  $|\lambda^{i,n,c}(s) - \lambda_*^{i,n}(s)|$  as stochastic intensity. We compute now:

$$\begin{aligned}\mathbb{E}_{\theta_0,i,n} C_{i,n}(t) &\leq \mathbb{E}_{\theta_0,i,n} \int_0^{t^-} a_0 e^{-b_0(t-s)} |\lambda^{i,n,c}(s) - \lambda_*^{i,n}(s)| ds \\ &= \int_0^{t^-} a_0 e^{-b_0(t-s)} \mathbb{E}_{\theta_0,i,n} |\lambda^{i,n,c}(s) - \lambda_*^{i,n}(s)| ds.\end{aligned}$$

So far we have shown that there exists a sequence  $L_n$  such that  $L_n = O(h_n^{-\kappa})$  and such that the function  $f(t) = \mathbb{E}_{\theta_0,i,n} |\lambda^{i,n,c}(t) - \lambda_*^{i,n}(t)|$  satisfies the inequality

$$f(t) \leq L_n + R_{i,n}(t) + f * h(t), \quad (10.31)$$

where  $h$  is the kernel defined as  $h : t \mapsto a_0 e^{-b_0 t}$ . By Lemma 10.6, this yields

$$f(t) \leq L_n + R_{i,n}(t) + \int_0^t \{L_n + R_{i,n}(s)\} a_0 e^{(a_0 - b_0)(t-s)} ds. \quad (10.32)$$

Now recall that  $b_0 - a_0 > \underline{b}(1-r)$  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

$$\begin{aligned}f(t) &\leq (1 + (1-r)^{-1}) L_n + R_{i,n}(t) + \int_0^t \{M_n e^{-\underline{b}(1-r)s}\} a_0 e^{\underline{b}(1-r)(t-s)} ds \\ &\leq (1 + (1-r)^{-1}) L_n + (1 + \bar{a}t) M_n e^{-\underline{b}(1-r)t}.\end{aligned}$$

If we recall that in the above expression  $f(t)$  stands for  $\mathbb{E}_{\theta_0,i,n} |\lambda^{i,n,c}(t) - \lambda_*^{i,n}(t)|$ , such uniform estimate clearly proves (10.28) in the case  $q = 1$ .

**Step 2.** We prove the result for any  $q \in \mathbb{N}^*$ . Let the expression  $f(t)$  stands for  $\mathbb{E}_{\theta_0,i,n} |\lambda^{i,n,c}(t) - \lambda_*^{i,n}(t)|^p$ . With similar notations as for the previous step, we have for any  $\eta > 0$

$$\begin{aligned}f(t) &= \mathbb{E}_{\theta_0,i,n} |\lambda^{i,n,c}(t) - \lambda_*^{i,n}(t)|^p \leq \mathbb{E}_{\theta_0,i,n} |A_{i,n}(t) + B_{i,n}(t) + C_{i,n}(t) + R_{i,n}(t)|^p \\ &\leq (1 + \eta^{-1})^{2^q-1} \mathbb{E}_{\theta_0,i,n} |A_{i,n}(t) + B_{i,n}(t) + R_{i,n}(t)|^p \\ &\quad + (1 + \eta)^{2^q-1} \mathbb{E}_{\theta_0,i,n} C_{i,n}(t)^p\end{aligned}$$

It is straightforward to see that similar arguments to the previous case lead to the uniform estimate

$$\mathbb{E}_{\theta_0,i,n} A_{i,n}(t)^p + \mathbb{E}_{\theta_0,i,n} B_{i,n}(t)^p = O_{\mathbb{P}}(h_n^{-\kappa p}).$$

Now, define  $W(s, z) = a_0 e^{-b_0(t-s)} |1_{\{0 \leq z \leq \lambda^{i,n,c}(s)\}} - 1_{\{0 \leq z \leq \lambda_*^{i,n}(s)\}}|$  to get

$$\begin{aligned}\mathbb{E}_{\theta_0,i,n} [C_{i,n}(t)^p] &= \mathbb{E}_{\theta_0,i,n} [(W * \overline{N}_t)^p] \\ &\leq (1 + \eta^{-1})^{2^q-1} \mathbb{E}_{\theta_0,i,n} [(W * (\overline{N} - \overline{\Lambda})_t)^p] + (1 + \eta)^{2^q-1} \mathbb{E}_{\theta_0,i,n} [(W * \overline{\Lambda}_t)^p],\end{aligned}$$

and apply Lemma 10.2 to get

$$\begin{aligned}
\mathbb{E}_{\theta_0, i, n} [(W * (\bar{N} - \bar{\Lambda})_t)^p] &\leq K_p \left( \mathbb{E}_{\theta_0, i, n} \left[ \iint_{[0, T] \times \mathbb{R}} |W(s, z)|^p ds dz \right] \right. \\
&\quad \left. + \mathbb{E}_{\theta_0, i, n} \left[ \left( \iint_{[0, T] \times \mathbb{R}} W(s, z)^2 ds dz \right)^{\frac{p}{2}} \right] \right) \\
&= K_p \left( \mathbb{E}_{\theta_0, i, n} \left[ \int_0^{t-} a_0^p e^{-pb_0(t-s)} |\lambda^{i, n, c}(s) - \lambda_*^{i, n}(s)| ds \right] \right. \\
&\quad \left. + \mathbb{E}_{\theta_0, i, n} \left[ \left( \int_0^{t-} a_0^2 e^{-2b_0(t-s)} |\lambda^{i, n, c}(s) - \lambda_*^{i, n}(s)| ds \right)^{\frac{p}{2}} \right] \right),
\end{aligned}$$

which is easily bounded as in (10.28) using the induction hypothesis. Note that here the presence of the integral term in  $|\lambda^{i, n, c}(s) - \lambda_*^{i, n}(s)|$  is the major obstacle to getting the stronger estimate  $O_{\mathbb{P}}(h_n^{-\kappa p})$  that one would expect. Finally the term

$$\mathbb{E}_{\theta_0, i, n} [(W * \bar{\Lambda}_t)^p] = \mathbb{E}_{\theta_0, i, n} \left[ \left( \int_0^{t-} a_0 e^{-b_0(t-s)} |\lambda^{i, n, c}(s) - \lambda_*^{i, n}(s)| ds \right)^p \right]$$

is treated exactly in the same way as for the proof of Lemma 10.3, to get the bound

$$\mathbb{E}_{\theta_0, i, n} [(W * \bar{\Lambda}_t)^p] \leq c_q f * h(t), \quad (10.33)$$

where again  $h : s \mapsto a_0 e^{-b_0 s}$ , and  $c_q < 1$  if  $\eta$  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 10.6 to conclude.

**Step 3.** It remains to show (ii) and (iii). They are just consequences of the application of Lemma 10.2 to the case  $W_\psi(s, z) = \psi_s |1_{\{0 \leq z \leq \lambda^{n, c}(s)\}} - 1_{\{0 \leq z \leq \lambda_*^n(s)\}}|$  and  $W_\chi(s, z) = \chi(s, t) |1_{\{0 \leq z \leq \lambda^{n, c}(s)\}} - 1_{\{0 \leq z \leq \lambda_*^n(s)\}}|$  along with Hölder's inequality.  $\square$

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

$$\lambda^{i, n, c}(t, \theta) = \nu + \int_0^{t-} a e^{-b(t-s)} dN_s^{i, n, c} \quad (10.34)$$

and

$$l_{i, n}^c(\theta) = \int_0^{h_n T} \log(\lambda^{i, n, c}(t, \theta)) dN_t^{i, n, c} - \int_0^{h_n T} \lambda^{i, n, c}(t, \theta) dt, \quad (10.35)$$

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.

**Lemma 10.9.** *Let  $\alpha \in (0, 1)$ . We have for any integer  $p \geq 1$  and any  $j \in \mathbb{N}$  that*

- (i)  $\sup_{(\theta_0, i, n, t) \in \mathbf{E}} \mathbb{E}_{\theta_0, i, n} \sup_{\theta \in K} \left| \partial_\theta^j \tilde{\lambda}^{i, n}(t, \theta) \right|^p \leq K_j \text{ } \mathbb{P}\text{-a.s.}$
- (ii)  $\sup_{(\theta_0, i, n, t) \in \mathbf{E}} \mathbb{E}_{\theta_0, i, n} \sup_{\theta \in K} \left| \partial_\theta^j \lambda^{i, n, c}(t, \theta) \right|^p \leq K_j \text{ } \mathbb{P}\text{-a.s.}$
- (iii)  $\sup_{(\theta_0, i, t) \in \mathbf{E}_n^\alpha} \mathbb{E}_{\theta_0, i, n} \sup_{\theta \in K} \left| \partial_\theta^j \tilde{\lambda}^{i, n}(t, \theta) - \partial_\theta^j \lambda^{i, n, c}(t, \theta) \right|^p = O_{\mathbb{P}}(h_n^{-\kappa})$

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  $\theta$  by linear combinations of terms of the form  $\bar{\nu}$  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  $[0, h_n T]$  is thus the consequence of Lemma 10.3 (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  $\int_0^{t-} (t-s)^j e^{-\underline{b}(t-s)} d|N_s^{i, n} - N_s^{i, n, c}|$ . The  $\mathbb{L}^p$  estimate of such expression is then easily derived by a truncation argument and Lemma 10.7 (iii).  $\square$

We now follow similar notations to the ones introduced in [10], and consider the main quantities of interest to derive the properties of the MLE. We define for any  $(\theta, \theta_0) \in K^2$ ,

$$\mathbb{Y}_{i, n}(\theta, \theta_0) = \frac{1}{h_n T} (l_{i, n}(\theta) - l_{i, n}(\theta_0)), \quad (10.36)$$

$$\Delta_{i, n}(\theta_0) = \frac{1}{\sqrt{h_n T}} \partial_\theta l_{i, n}(\theta_0), \quad (10.37)$$

and finally recall the expression

$$\Gamma_{i, n}(\theta_0) = -\frac{1}{h_n T} \partial_\theta^2 l_{i, n}(\theta_0), \quad (10.38)$$

that we introduced in (5.18). We define in the same way  $\mathbb{Y}_{i, n}^c$ ,  $\Delta_{i, n}^c$ , and  $\Gamma_{i, n}^c$ . We introduce for the next lemma the set  $\mathbf{I} = \{(\theta_0, i, n) \in K \times \mathbb{N}^2 \mid 1 \leq i \leq B_n\}$ .

**Lemma 10.10.** *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*

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{E}_{\theta_0, i, n} |\Delta_{i, n}(\theta_0) - \Delta_{i, n}^c(\theta_0)|^L \rightarrow^{\mathbb{P}} 0, \quad (10.39)$$

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{E}_{\theta_0, i, n} \left[ \sup_{\theta \in K} |\mathbb{Y}_{i, n}(\theta, \theta_0) - \mathbb{Y}_{i, n}^c(\theta, \theta_0)|^p \right] = O_{\mathbb{P}}(h_n^{-\epsilon\kappa}), \quad (10.40)$$

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{E}_{\theta_0, i, n} |\Gamma_{i, n}(\theta_0) - \Gamma_{i, n}^c(\theta_0)|^p = O_{\mathbb{P}}(h_n^{-\epsilon\kappa}), \quad (10.41)$$

$$\sup_{(\theta_0, i, n) \in \mathbf{I}} \mathbb{E}_{\theta_0, i, n} \left| h_n^{-1} \sup_{\theta \in K} |\partial_\theta^3 l_{i, n}(\theta)| \right|^p < K \text{ } \mathbb{P}\text{-a.s.} \quad (10.42)$$

*Proof.* Let us show (10.39). We can express the equation in (10.37) and its counterpart for the constant model as

$$\Delta_{i,n}(\theta_0) = \frac{1}{\sqrt{h_n T}} \left\{ \int_0^{h_n T} \frac{\partial_\theta \tilde{\lambda}^{i,n}(s, \theta_0)}{\tilde{\lambda}^{i,n}(s, \theta_0)} dN_s^{i,n} - \int_0^{h_n T} \partial_\theta \tilde{\lambda}^{i,n}(s, \theta_0) ds \right\} \quad (10.43)$$

and

$$\Delta_{i,n}^c(\theta_0) = \frac{1}{\sqrt{h_n T}} \left\{ \int_0^{h_n T} \frac{\partial_\theta \lambda^{i,n,c}(s, \theta_0)}{\lambda^{i,n,c}(s, \theta_0)} dN_s^{i,n,c} - \int_0^{h_n T} \partial_\theta \lambda^{i,n,c}(s, \theta_0) ds \right\}. \quad (10.44)$$

By Lemma 10.5 (i) and (iii), and Lemma 10.9 (i) and (ii) and the presence of the factor  $\frac{1}{\sqrt{h_n T}}$ , 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  $\sqrt{h_n T}(\Delta_{i,n}(\theta_0) - \Delta_{i,n}^c(\theta_0))$  is equivalent to the sum of the three terms

$$\begin{aligned} & \int_{h_n^\alpha T}^{h_n T} \frac{\partial_\theta \tilde{\lambda}^{i,n}(s, \theta_0)}{\tilde{\lambda}^{i,n}(s, \theta_0)} (dN_s^{i,n} - dN_s^{i,n,c}) + \int_{h_n^\alpha T}^{h_n T} \left\{ \frac{\partial_\theta \tilde{\lambda}^{i,n}(s, \theta_0)}{\tilde{\lambda}^{i,n}(s, \theta_0)} - \frac{\partial_\theta \lambda^{i,n,c}(s, \theta_0)}{\lambda^{i,n,c}(s, \theta_0)} \right\} dN_s^{i,n,c} \\ & \quad + \int_{h_n^\alpha T}^{h_n T} \{\partial_\theta \tilde{\lambda}^{i,n}(s, \theta_0) - \partial_\theta \lambda^{i,n,c}(s, \theta_0)\} ds. \end{aligned}$$

We therefore apply Lemmas 10.7 (ii) and 10.9 (i) to the first term, Lemmas 10.5 (iii) and 10.9 (iii) to the second term, and finally Lemma 10.9 (iii) to the last term to obtain the overall estimate

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{E}_{\theta_0, i, n} |\Delta_{i,n}(\theta_0) - \Delta_{i,n}^c(\theta_0)|^L = O_{\mathbb{P}} \left( h_n^{\frac{L}{2} - \epsilon \kappa} \right), \quad (10.45)$$

for any  $\epsilon \in (0, 1)$ . This tends to 0 if we can find an  $\epsilon$  such that  $\frac{L}{2} - \epsilon \kappa < 0$ , and this can be done by taking  $\epsilon$  sufficiently close to 1 since  $L < 2\kappa$ . Equations (10.40), (10.41) and (10.42) are proved similarly.  $\square$

**Lemma 10.11.** *For any integer  $p \geq 1$ , there exists a constant  $M$  such that*

$$\sup_{(\theta_0, i, n) \in \mathbf{I}} \mathbb{E}_{\theta_0, i, n} |\Delta_n^c(\theta_0)|^p < M \text{ } \mathbb{P}\text{-a.s.} \quad (10.46)$$

Furthermore, there exists a mapping  $(\theta, \theta_0) \rightarrow \mathbb{Y}(\theta, \theta_0)$  such that for any  $\epsilon \in (0, 1)$ ,

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{E}_{\theta_0, i, n} \left[ \sup_{\theta \in K} |\mathbb{Y}_{i,n}^c(\theta, \theta_0) - \mathbb{Y}(\theta, \theta_0)| \right] = O \left( h_n^{-\epsilon \frac{p}{2}} \right) \text{ } \mathbb{P}\text{-a.s.} \quad (10.47)$$

Finally, for any  $\theta_0 \in K$ , and for any  $\epsilon \in (0, 1)$ ,

$$\sup_{\theta_0 \in K, 1 \leq i \leq \Delta_n^{-1}} \mathbb{E}_{\theta_0, i, n} |\Gamma_{i,n}^c(\theta_0) - \Gamma(\theta_0)|^p = O \left( h_n^{-\epsilon \frac{p}{2}} \right) \text{ } \mathbb{P}\text{-a.s.} \quad (10.48)$$

where  $\Gamma(\theta_0)$  is the asymptotic Fisher information matrix of the parametric Hawkes process regression model with parameter  $\theta_0$  as introduced in (10.8).

*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  $\theta_0$ , and is independent of the filtration  $\mathcal{F}_0^{i,n}$ . Thus, by a regular distribution argument the operator  $\mathbb{E}_{\theta_0,i,n}$  acts as the simple operator  $\mathbb{E}$  for  $N^{i,n,c}$  distributed as a Hawkes with true value  $\theta_0$ . It is straightforward to see that under a mild change in the proofs of Lemma 3.15 and Theorem 4.6 in [10] those estimates hold uniformly in  $\theta_0 \in K$  and in the block index.  $\square$

**Theorem 10.12.** *Let  $L \in (0, 2\kappa)$ . We have*

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \left\{ \mathbb{E}_{\theta_0,i,n} \left[ f \left( \sqrt{h_n} (\widehat{\Theta}_{i,n} - \theta_0) \right) \right] - \mathbb{E} \left[ f \left( T^{-\frac{1}{2}} \Gamma(\theta_0)^{-\frac{1}{2}} \xi \right) \right] \right\} \rightarrow^{\mathbb{P}} 0, \quad (10.49)$$

for any continuous function  $f$  with  $|f(x)| = O(|x|^L)$  when  $|x| \rightarrow \infty$ , and such that  $\xi$  follows a standard normal distribution.

*Proof.* By (10.40) and (10.47), we can define some number  $\epsilon \in (0, 1)$  such that

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} h_n^{\epsilon(\frac{p}{2} \wedge \kappa)} \mathbb{E}_{\theta_0,i,n} \left[ \sup_{\theta \in K} |\mathbb{Y}_{i,n}(\theta, \theta_0) - \mathbb{Y}(\theta, \theta_0)|^p \right] \rightarrow^{\mathbb{P}} 0, \quad (10.50)$$

and as  $\widehat{\Theta}_{i,n}$  is also a maximizer of  $\theta \rightarrow \mathbb{Y}_{i,n}(\theta, \theta_0)$ , (10.50) 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.

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{P}_{\theta_0,i,n} \left[ \widehat{\Theta}_{i,n} - \theta_0 \right] \rightarrow^{\mathbb{P}} 0, \quad (10.51)$$

since  $\mathbb{Y}$  satisfies the non-degeneracy condition [A4] in [10]. From (10.41) and (10.48) we deduce

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} h_n^{\epsilon(\frac{p}{2} \wedge \kappa)} \mathbb{E}_{\theta_0,i,n} |\Gamma_{i,n}(\theta_0) - \Gamma(\theta_0)|^p \rightarrow^{\mathbb{P}} 0. \quad (10.52)$$

By (10.39),  $\Delta_{i,n}(\theta_0)$  and  $\Delta_{i,n}^c(\theta_0)$  have the same asymptotic distribution, which is of the form  $\Gamma(\theta_0)^{\frac{1}{2}} \xi$ , where  $\xi$  follows a standard normal distribution. Following the proof of Theorem 3.11 in [10], 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.

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \left\{ \mathbb{E}_{\theta_0,i,n} \left[ f \left( \sqrt{h_n} (\widehat{\Theta}_{i,n} - \theta_0) \right) \right] - \mathbb{E} \left[ f \left( T^{-\frac{1}{2}} \Gamma(\theta_0)^{-\frac{1}{2}} \xi \right) \right] \right\} \rightarrow^{\mathbb{P}} 0, \quad (10.53)$$

for any bounded continuous function  $f$ .

Finally, we extend (10.53) to the case of a function of polynomial growth of order smaller than  $L$ . First note that by (10.39) and (10.46) we have for any  $L' \in (L, 2\kappa)$

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{E}_{\theta_0,i,n} |\Delta_{i,n}(\theta_0)|^{L'} = O_{\mathbb{P}}(1). \quad (10.54)$$

We now adopt the notations of [39] 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  $M_3 = L\left(\frac{\alpha}{1-\alpha} - \rho_1\right)^{-1} < \infty$ . Then, by (10.50), (10.52), (10.54) and finally (10.42), conditions [A1''], [A4'], [A6], [B1] and [B2] in [39] are satisfied. It is straightforward that we can apply a conditional version

(with respect to the operator  $\mathbb{E}_{\theta_0, i, n}$ ) of Theorem 3 and Proposition 1 from [39] to get that for any  $p \leq L$ ,

$$\sup_{\theta_0 \in K, 1 \leq i \leq \Delta_n^{-1}} \mathbb{E}_{\theta_0, i, n} \left| \sqrt{h_n} (\widehat{\Theta}_{i, n} - \theta_0) \right|^p = O_{\mathbb{P}}(1). \quad (10.55)$$

Such stochastic boundedness of conditional moments along with the convergence in distribution is clearly sufficient to imply the theorem.  $\square$

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 5.2, we relax this assumption. In addition, we use regular conditional distribution techniques (see for instance Section 4.3 (pp. 77 – 80) in [5]) to obtain (10.49) when not conditioning by any particular starting value of  $\theta_t^*$ . We provide the formal proof in what follows. Recall that  $\mathbb{E}_{(i-1)\Delta_n}$  stands for  $\mathbb{E}[\cdot | \mathcal{F}_0^{i, n}]$ .

*Proof of Theorem 5.2.* We can decompose  $\mathbb{E}_{(i-1)\Delta_n} \left[ f(\sqrt{h_n} (\widehat{\Theta}_{i, n} - \theta_{(i-1)\Delta_n}^*)) \right]$  as

$$\mathbb{E}_{(i-1)\Delta_n} \left[ f \left( \sqrt{h_n} (\widehat{\Theta}_{i, n} - \theta_{(i-1)\Delta_n}^*) \right) \mathbf{1}_{\{R_{i, n}(0) \leq M_n\}} \right] \quad (10.56)$$

$$+ \mathbb{E}_{(i-1)\Delta_n} \left[ f \left( \sqrt{h_n} (\widehat{\Theta}_{i, n} - \theta_{(i-1)\Delta_n}^*) \right) \mathbf{1}_{\{R_{i, n}(0) > M_n\}} \right]. \quad (10.57)$$

Let  $\xi$  as in Theorem 5.2. On the one hand by a regular conditional distribution argument, if we define  $G(\theta_0) = \mathbb{E}_{\theta_0, i, n} \left[ f(\sqrt{h_n} (\widehat{\Theta}_{i, n} - \theta_0)) \right] - \mathbb{E} \left[ f(T^{-\frac{1}{2}} \Gamma(\theta_0)^{-\frac{1}{2}} \xi) \right]$ , we can express uniformly in  $i \in \{1, \dots, B_n\}$  the quantity

$$\mathbb{E}_{(i-1)\Delta_n} \left[ f \left( \sqrt{h_n} (\widehat{\Theta}_{i, n} - \theta_{(i-1)\Delta_n}^*) \right) \mathbf{1}_{\{R_{i, n}(0) \leq M_n\}} - f \left( T^{-\frac{1}{2}} \Gamma \left( \theta_{(i-1)\Delta_n}^* \right)^{-\frac{1}{2}} \xi \right) \right] \quad (10.58)$$

as  $G(\theta_{(i-1)\Delta_n}^*)$  by definition of  $\mathbb{E}_{\theta_0, i, n}$  and because  $\xi \perp\!\!\!\perp \mathcal{F}$ . We note that

$$\left| G(\theta_{(i-1)\Delta_n}^*) \right| \leq \sup_{\theta_0 \in K} \left| \mathbb{E}_{\theta_0, i, n} \left[ f \left( \sqrt{h_n} (\widehat{\Theta}_{i, n} - \theta_0) \right) \right] - \mathbb{E} \left[ f \left( T^{-\frac{1}{2}} \Gamma(\theta_0)^{-\frac{1}{2}} \xi \right) \right] \right|, \quad (10.59)$$

take the sup over  $i$  in (10.59), and in view of Theorem 10.12, we have shown that (10.58) is uniformly of order  $o_{\mathbb{P}}(1)$ .

On the other hand, (10.57) 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 (10.57) vanishes asymptotically.  $\square$

## 10.4 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  $\mathbb{E}_{(i-1)\Delta_n} \left[ \widehat{\Theta}_{i, n} - \theta_{(i-1)\Delta_n}^* \right]$ . 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  $N^n, \lambda_*^n, l_n, \mathbb{E}_{\theta_0, n}$ , etc. should be read  $N^{i,n}, \lambda_*^{i,n}, l_{i,n}, \mathbb{E}_{\theta_0, i, n}$ , 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  $\mathbb{E}_{\theta_0, n}$ , we denote by  $\bar{Z}$  its centered version, i.e. the random variable  $Z - \mathbb{E}_{\theta_0, n}[Z]$ . 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 5, 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  $\xi_n \in [\hat{\Theta}_n, \theta_0]$  such that

$$0 = \partial_\theta l_n(\hat{\Theta}_n) = \partial_\theta l_n(\theta_0) + \partial_\theta^2 l_n(\theta_0)(\hat{\Theta}_n - \theta_0) + \frac{1}{2} \partial_\theta^3 l_n(\xi_n)(\hat{\Theta}_n - \theta_0)^{\otimes 2}, \quad (10.60)$$

where  $\partial_\theta^3 l_n(\xi_n)(\hat{\Theta}_n - \theta_0)^{\otimes 2}$  is a compact expression for the vector whose  $i$ -th component is  $\partial_{\theta,ijk}^3 l_n(\xi_n)(\hat{\Theta}_n - \theta_0)_j(\hat{\Theta}_n - \theta_0)_k$ . Let  $\epsilon \in (0, 1)$ . By application of Lemmas 10.7 and 10.9, it still holds that

$$\partial_\theta l_n^c(\theta_0) + \partial_\theta^2 l_n^c(\theta_0)(\hat{\Theta}_n - \theta_0) + \frac{1}{2} \partial_\theta^3 l_n^c(\xi_n)(\hat{\Theta}_n - \theta_0)^{\otimes 2} = O_{\mathbb{P}}(h_n^{1-\epsilon\kappa}), \quad (10.61)$$

where the residual term  $O_{\mathbb{P}}(h_n^{1-\epsilon\kappa})$  admits clearly moments of any order with respect to  $\mathbb{E}_{\theta_0, n}$ . We now apply the operator  $\mathbb{E}_{\theta_0, n}$ , divide by  $h_n T$  and obtain

$$\mathbb{E}_{\theta_0, n}[\bar{\Gamma}_n^c(\theta_0)(\hat{\Theta}_n - \theta_0)] + \mathbb{E}_{\theta_0, n}[\Gamma_n^c(\theta_0)]\mathbb{E}_{\theta_0, n}[\hat{\Theta}_n - \theta_0] - \mathbb{E}_{\theta_0, n}\left[\frac{\partial_\theta^3 l_n^c(\xi_n)}{2h_n T}(\hat{\Theta}_n - \theta_0)^{\otimes 2}\right] = O_{\mathbb{P}}(h_n^{-\epsilon\kappa}),$$

where the expectation of the first term has vanished because of the martingale form of  $\Delta_n^c(\theta_0)$  in (10.44). The term  $\mathbb{E}_{\theta_0, n}[\Gamma_n^c(\theta_0)]\mathbb{E}_{\theta_0, n}[\hat{\Theta}_n - \theta_0]$  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 process  $C_n^c(\theta_0)$  whose components are, for any triplet  $(i, j, k) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3$ ,

$$C_n^c(\theta_0)_{i,jk} = \frac{1}{h_n T} \int_0^{h_n T} \left\{ \partial_{\theta,i} \lambda^{n,c}(s, \theta_0) \partial_{\theta,jk}^2 \log \lambda^{n,c}(s, \theta_0) - [M_n^c(s, \theta_0)]_i \frac{[\partial_\theta \lambda^{n,c}(s, \theta_0)^{\otimes 2}]_{jk}}{\lambda^{n,c}(s, \theta_0)} \right\} ds$$

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

**Lemma 10.13.** *We have*

$$\mathbb{E}_{\theta_0, n}[C_n^c(\theta_0)_{i,jk}] = -\sqrt{h_n T} \mathbb{E}_{\theta_0, n}[\Delta_n^c(\theta_0)_i \Gamma_n^c(\theta_0)_{jk}]. \quad (10.62)$$

*Proof.* Note that for two  $\mathbb{L}_2$  bounded processes  $(u_s)_s, (v_s)_s$ , we have

$$\left\langle \int_0^{\cdot} u_s \{dN_s^{n,c} - \lambda^{n,c}(s, \theta_0)ds\}, \int_0^{\cdot} v_s \{dN_s^{n,c} - \lambda^{n,c}(s, \theta_0)ds\} \right\rangle_t = \int_0^t u_s v_s \lambda^{n,c}(s, \theta_0) ds$$

Taking expectation, this yields

$$\mathbb{E}_{\theta_0, n} \left[ \int_0^t u_s \{dN_s^{n,c} - \lambda^{n,c}(s, \theta_0)ds\} \int_0^t v_s \{dN_s^{n,c} - \lambda^{n,c}(s, \theta_0)ds\} \right] = \mathbb{E}_{\theta_0, n} \left[ \int_0^t u_s v_s \lambda^{n,c}(s, \theta_0)ds \right]$$

We also have the by-part integration formula

$$\mathbb{E}_{\theta_0, n} \left[ \int_0^t u_s ds M_n^c(t, \theta_0) \right] = \mathbb{E}_{\theta_0, n} \left[ \int_0^t u_s M_n^c(s, \theta_0) ds \right] \quad (10.63)$$

Formula (10.62) is then obtained directly from the expression of  $\Gamma_n^c(\theta_0)$  and  $\Delta_n^c(\theta_0)$ .  $\square$

Now, by the same argument as for the proof of (10.11), we have for any integer  $p \geq 1$  and any  $\epsilon \in (0, 1)$ ,

$$\sup_{\theta_0 \in K} h_n^{\epsilon \frac{p}{2}} \mathbb{E}_{\theta_0, n} |C_n^c(\theta_0) - C(\theta_0)|^p \rightarrow^{\mathbb{P}} 0 \quad (10.64)$$

where  $C$  was defined in (10.10). Before we turn to the limiting expression of the term

$$\mathbb{E}_{\theta_0, n} [\bar{\Gamma}_n^c(\theta_0) (\hat{\Theta}_n - \theta_0)]_i$$

in our expansion of the bias in terms of  $C(\theta_0)$ , 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  $\mathbb{B}_n(\theta_0)^c$ .

**Lemma 10.14.** *We have, for any integer  $p \geq 1$  and any  $\epsilon \in (0, 1)$  that*

- (i)  $\sup_{\theta_0 \in K} \mathbb{P}_{\theta_0, n} [\mathbb{B}_n(\theta_0)^c] = O_{\mathbb{P}} \left( h_n^{-\epsilon \frac{p}{2}} \right)$ .
- (ii)  $\sup_{\theta_0 \in K} h_n^{\epsilon \frac{p}{2}} \mathbb{E}_{\theta_0, n} [|\Gamma_n^c(\theta_0)^{-1} - \Gamma(\theta_0)^{-1}| \mathbf{1}_{\mathbb{B}_n}] \rightarrow^{\mathbb{P}} 0$ .

*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

$$\mathbb{P}_{\theta_0, n} [\mathbb{B}_n(\theta_0)^c] \leq \mathbb{P}_{\theta_0, n} \left\{ \forall x \in \mathbb{R}^3 - \{0\}, \frac{|x^T (\Gamma_n^c(\theta_0) - \Gamma(\theta_0)) x|}{|x|_2^2} > \frac{c_0}{2} \right\}, \quad (10.65)$$

and by equivalence of the norms  $|M|$  and  $\sup_{x \in \mathbb{R}^3 - \{0\}} \frac{|x^T M x|}{|x|_2^2}$  on the space of symmetric matrices of  $\mathbb{R}^3$ , (10.65) implies the existence of some constant  $\eta > 0$  such that

$$\begin{aligned} \mathbb{P}_{\theta_0, n} [\mathbb{B}_n(\theta_0)^c] &\leq \mathbb{P}_{\theta_0, n} [|\Gamma_n^c(\theta_0) - \Gamma(\theta_0)| > \eta c_0] \\ &\leq (\eta c_0)^{-p} \mathbb{E}_{\theta_0, n} |\Gamma_n^c(\theta_0) - \Gamma(\theta_0)|^p, \end{aligned}$$

where Markov's inequality was used at the last step. (i) thus follows from (10.52). Moreover, (ii) is easily obtained using the elementary result  $|A^{-1} - B^{-1}| = |B^{-1}(B - A)A^{-1}| \leq |A^{-1}|_{\infty} |B^{-1}|_{\infty} |B - A|$  applied to  $\Gamma_n^c(\theta_0)$  and  $\Gamma(\theta_0)$  on the set  $\mathbb{B}_n(\theta_0)$ .  $\square$

**Lemma 10.15.** *Let  $\epsilon \in (0, 1)$  and  $i \in \{0, 1, 2\}$ . The following expansion holds.*

$$\mathbb{E}_{\theta_0, n} [\bar{\Gamma}_n^c(\theta_0) (\hat{\Theta}_n - \theta_0)]_i = -\frac{\Gamma(\theta_0)^{jk} C(\theta_0)_{k,ij}}{h_n T} + O_{\mathbb{P}} \left( h_n^{-\epsilon(\kappa \wedge \frac{3}{2})} \right). \quad (10.66)$$

*Proof.* Note first that in view of Lemma 10.14 (i) along with Hölder's inequality, we have that  $\mathbb{E}_{\theta_0,n}[\bar{\Gamma}_n^c(\theta_0)(\hat{\Theta}_n - \theta_0)] = \mathbb{E}_{\theta_0,n}[\bar{\Gamma}_n^c(\theta_0)(\hat{\Theta}_n - \theta_0)\mathbf{1}_{\mathbb{B}_n(\theta_0)}] + o_{\mathbb{P}}(h_n^{-\frac{3}{2}})$ . 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 (10.15). Take  $\epsilon \in (0, 1)$  and  $\tilde{\epsilon} \in (\epsilon, 1)$ . As a consequence of (10.61), we have the representation,

$$\hat{\Theta}_n - \theta_0 = \frac{1}{\sqrt{h_n T}} \Gamma_n^c(\theta_0)^{-1} \Delta_n^c(\theta_0) + \Gamma_n^c(\theta_0)^{-1} \frac{\partial_{\theta}^3 l_n^c(\xi_n)(\hat{\Theta}_n - \theta_0)^{\otimes 2}}{2h_n T} + O_{\mathbb{P}}(h_n^{-\tilde{\epsilon}\kappa}), \quad (10.67)$$

on the set  $\mathbb{B}_n(\theta_0)$ , where the residual term  $O_{\mathbb{P}}(h_n^{-\tilde{\epsilon}\kappa})$  admits moments of any order with respect to the operator  $\mathbb{E}_{\theta_0,n}$ . We inject (10.67) in the expectation and get

$$\begin{aligned} \mathbb{E}_{\theta_0,n}[\bar{\Gamma}_n^c(\theta_0)(\hat{\Theta}_n - \theta_0)] &= \frac{1}{\sqrt{h_n T}} \mathbb{E}_{\theta_0,n}[\bar{\Gamma}_n^c(\theta_0)\Gamma_n^c(\theta_0)^{-1} \Delta_n^c(\theta_0)\mathbf{1}_{\mathbb{B}_n(\theta_0)}] \\ &\quad + \mathbb{E}_{\theta_0,n} \left[ \bar{\Gamma}_n^c(\theta_0)\Gamma_n^c(\theta_0)^{-1} \frac{\partial_{\theta}^3 l_n^c(\xi_n)(\hat{\Theta}_n - \theta_0)^{\otimes 2}}{2h_n T} \mathbf{1}_{\mathbb{B}_n(\theta_0)} \right] \\ &\quad + O_{\mathbb{P}}(h_n^{-\epsilon\kappa}), \end{aligned}$$

where the residual term  $O_{\mathbb{P}}(h_n^{-\epsilon\kappa})$  is obtained by Hölder's inequality using the fact that  $\epsilon < \tilde{\epsilon}$ . By Lemma 10.14 (ii), the first term admits the expansion

$$\frac{1}{\sqrt{h_n T}} \mathbb{E}_{\theta_0,n}[\bar{\Gamma}_n^c(\theta_0)\Gamma(\theta_0)^{-1} \Delta_n^c(\theta_0)] + O_{\mathbb{P}}(h_n^{-\frac{3\epsilon}{2}}), \quad (10.68)$$

where we used Hölder's inequality to control  $\frac{1}{\sqrt{h_n T}} \mathbb{E}_{\theta_0,n}[\bar{\Gamma}_n^c(\theta_0)(\Gamma_n^c(\theta_0)^{-1} - \Gamma(\theta_0)^{-1})\Delta_n^c(\theta_0)]$  and we neglected the effect of the indicator function by Lemma 10.14 (i). For any  $i \in \{0, 1, 2\}$ , we develop the matrix product in (10.68), use Lemma 10.13 along with (10.64), and this leads to the estimate

$$\frac{1}{\sqrt{h_n T}} \mathbb{E}_{\theta_0,n}[\bar{\Gamma}_n^c(\theta_0)\Gamma(\theta_0)^{-1} \Delta_n^c(\theta_0)]_i = \frac{\Gamma(\theta_0)^{jk} C(\theta_0)_{k,ij}}{h_n T} + O_{\mathbb{P}}(h_n^{-\frac{3\epsilon}{2}}). \quad (10.69)$$

It remains to control the term  $\mathbb{E}_{\theta_0,n} \left[ \bar{\Gamma}_n^c(\theta_0)\Gamma_n^c(\theta_0)^{-1} \frac{\partial_{\theta}^3 l_n^c(\xi_n)(\hat{\Theta}_n - \theta_0)^{\otimes 2}}{2h_n T} \mathbf{1}_{\mathbb{B}_n(\theta_0)} \right]$ . Take  $L \in (2, 2\kappa)$ . By boundedness of moments of  $h_n^{\frac{\epsilon}{2}} \bar{\Gamma}_n^c(\theta_0)_{ij} \Gamma_n^c(\theta_0)^{jk} \frac{\partial_{\theta}^3 l_n^c(\theta)}{2h_n T} \mathbf{1}_{\mathbb{B}_n(\theta_0)}$ , for any  $(i, j, k, l, m)$  and uniformly in  $\theta_0 \in K$ , we have

$$\begin{aligned} \mathbb{E}_{\theta_0,n} \left[ \bar{\Gamma}_n^c(\theta_0)_{ij} \Gamma_n^c(\theta_0)^{jk} \frac{\partial_{\theta}^3 l_n^c(\xi_n)(\hat{\Theta}_n - \theta_0)_l (\hat{\Theta}_n - \theta_0)_m}{2h_n T} \mathbf{1}_{\mathbb{B}_n(\theta_0)} \right] \\ \leq K h_n^{-\frac{\epsilon}{2}} \mathbb{E}_{\theta_0,n} \left[ \left| (\hat{\Theta}_n - \theta_0)_l (\hat{\Theta}_n - \theta_0)_m \right|^{\frac{L}{2}} \right]^{\frac{2}{L}} \\ = O_{\mathbb{P}}(h_n^{-\frac{3\epsilon}{2}}), \end{aligned}$$

where Hölder's inequality was applied for the first inequality, and Theorem 10.12 was used with the function  $f : x \rightarrow (x_l x_m)^{\frac{L}{2}}$ , which is of polynomial growth of order  $L$ , to get the final estimate.  $\square$

Finally, we derive the expansion of  $\frac{1}{2h_nT} \mathbb{E}_{\theta_0, n} [\partial_\theta^3 l_n^c(\xi_n)(\widehat{\Theta}_n - \theta_0)^{\otimes 2}]$ . First note that for any integer  $p \geq 1$  and any  $\epsilon \in (0, 1)$ ,

$$\sup_{\theta_0 \in K} h_n^{\frac{\epsilon p}{2}} \mathbb{E}_{\theta_0, n} \left| \frac{1}{h_n T} \partial_\theta^3 l_n^c(\theta_0) - K(\theta_0) \right|^p \rightarrow^{\mathbb{P}} 0, \quad (10.70)$$

where  $K(\theta_0)$  was introduced in (10.9). The next lemma is proved the same way as for Lemma 10.15.

**Lemma 10.16.** *Let  $\epsilon \in (0, 1)$  and  $i \in \{0, 1, 2\}$ . We have the expansion*

$$\frac{1}{2h_n T} \mathbb{E}_{\theta_0, n} [\partial_\theta^3 l_n^c(\xi_n)(\widehat{\Theta}_n - \theta_0)^{\otimes 2}]_i = \frac{\Gamma(\theta_0)^{jk} K(\theta_0)_{ijk}}{2h_n T} + O_{\mathbb{P}} \left( h_n^{-\epsilon(\kappa \wedge \frac{3}{2})} \right). \quad (10.71)$$

*Proof.* Consider three indices  $i, j, k \in \{0, 1, 2\}$  and  $\epsilon \in (0, 1)$ . We have the decomposition

$$\begin{aligned} \frac{1}{2h_n T} \mathbb{E}_{\theta_0, n} [\partial_{\theta, ijk}^3 l_n^c(\xi_n)(\widehat{\Theta}_n - \theta_0)_j(\widehat{\Theta}_n - \theta_0)_k] &= \frac{1}{2h_n T} \mathbb{E}_{\theta_0, n} [\partial_{\theta, ijk}^3 l_n^c(\xi_n)] \mathbb{E}_{\theta_0, n} [(\widehat{\Theta}_n - \theta_0)_j(\widehat{\Theta}_n - \theta_0)_k] \\ &\quad + \frac{1}{2h_n T} \mathbb{E}_{\theta_0, n} [\overline{\partial_{\theta, ijk}^3 l_n^c(\xi_n)}(\widehat{\Theta}_n - \theta_0)_j(\widehat{\Theta}_n - \theta_0)_k]. \end{aligned}$$

We now remark that the first term admits the expansion

$$\frac{\Gamma(\theta_0)^{jk} K(\theta_0)_{ijk}}{2h_n T} + O_{\mathbb{P}} \left( h_n^{-\epsilon(\kappa \wedge \frac{3}{2})} \right), \quad (10.72)$$

by replacing  $\mathbb{E}_{\theta_0, n} [(\widehat{\Theta}_n - \theta_0)_j(\widehat{\Theta}_n - \theta_0)_k]$  and  $\frac{1}{2h_n T} \mathbb{E}_{\theta_0, n} [\partial_{\theta, ijk}^3 l_n^c(\xi_n)]$  by their estimates

$$\mathbb{E}_{\theta_0, n} [(\widehat{\Theta}_n - \theta_0)_j(\widehat{\Theta}_n - \theta_0)_k] = \frac{\Gamma(\theta_0)^{jk}}{h_n T} + O_{\mathbb{P}} \left( h_n^{-\epsilon(\kappa \wedge \frac{3}{2})} \right), \quad (10.73)$$

and

$$\frac{1}{2h_n T} \mathbb{E}_{\theta_0, n} [\partial_{\theta, ijk}^3 l_n^c(\xi_n)] = K(\theta_0)_{ijk} + O_{\mathbb{P}} \left( h_n^{-\frac{\epsilon}{2}} \right). \quad (10.74)$$

(10.73) is obtained by injecting the expansion of  $\widehat{\Theta}_n - \theta_0$  in (10.67) up to the first order only, and (10.74) is a consequence of (10.70) and the uniform boundedness of moments of  $\frac{\partial_\theta^4 l_n^c(\theta_0)}{h_n T}$  in  $\theta_0 \in K$  by Lemma 10.9 (ii). Note that the expansion (10.73) is not a direct consequence of Theorem 10.12 applied to  $x \rightarrow x_j x_k$  since this would lead to the weaker estimate  $\frac{\Gamma(\theta_0)^{jk}}{h_n T} + o_{\mathbb{P}}(h_n^{-1})$  instead. Finally, the second term is of order  $O_{\mathbb{P}} \left( h_n^{-\frac{3\epsilon}{2}} \right)$  by Hölder's inequality along with Theorem 10.12, and thus we are done.  $\square$

Before we turn to the final theorem, we recall for any  $j \in \{0, 1, 2\}$  the expression

$$b(\theta_0)_j = \frac{1}{2} \Gamma(\theta_0)^{ij} \Gamma(\theta_0)^{kl} (K(\theta_0)_{ikl} + 2C(\theta_0)_{k,il}), \quad (10.75)$$

which was defined in (10.12) and its empirical counterpart

$$b_n(\theta_0)_j = \frac{1}{2} \Gamma_n(\theta_0)^{ij} \Gamma_n(\theta_0)^{kl} (K_n(\theta_0)_{ikl} + 2C_n(\theta_0)_{k,il}), \quad (10.76)$$

where we have  $K_n(\theta_0) = \frac{1}{h_n T} \partial_\theta^3 l_n(\theta_0)$ . We are now ready to state the general theorem on bias correction of the local MLE, which we formulate with the block index  $i$ .

**Theorem 10.17.** Let  $\epsilon \in (0, 1)$ . The bias of the estimator  $\widehat{\Theta}_{i,n}$  has the expansion

$$\mathbb{E}_{\theta_0, i, n} [\widehat{\Theta}_{i,n} - \theta_0] = \frac{b(\theta_0)}{h_n T} + O_{\mathbb{P}} \left( h_n^{-\epsilon(\kappa \wedge \frac{3}{2})} \right), \quad (10.77)$$

uniformly in  $i \in \{1, \dots, B_n\}$  and in  $\theta_0 \in K$ . Moreover, the bias-corrected estimator  $\widehat{\Theta}_{i,n}^{(BC)}$  defined in (5.22) has the (uniform) bias expansion

$$\mathbb{E}_{\theta_0, i, n} [\widehat{\Theta}_{i,n}^{(BC)} - \theta_0] = O_{\mathbb{P}} \left( h_n^{-\epsilon(\kappa \wedge \frac{3}{2})} \right). \quad (10.78)$$

*Proof.* We drop the index  $i$  in this proof. Take  $\epsilon \in (0, 1)$  and some  $j \in \{0, 1, 2\}$ . By Lemma 10.15 and Lemma 10.16, we have

$$\mathbb{E}_{\theta_0, n} [\Gamma_n^c(\theta_0)]_{jk} \mathbb{E}_{\theta_0, n} [\widehat{\Theta}_n - \theta_0]_k = \frac{\Gamma(\theta_0)^{kl} (K(\theta_0)_{jkl} + 2C(\theta_0)_{l,jk})}{2h_n T} + O_{\mathbb{P}} \left( h_n^{-\epsilon(\kappa \wedge \frac{3}{2})} \right), \quad (10.79)$$

which is a set of simultaneous linear equations. After inversion of this system of equations and application of Lemma 10.14, the expression of the bias becomes for  $j \in \{0, 1, 2\}$ ,

$$\mathbb{E}_{\theta_0, n} [\widehat{\Theta}_n - \theta_0]_j = \frac{\Gamma(\theta_0)^{ij} \Gamma(\theta_0)^{kl} (K(\theta_0)_{ikl} + 2C(\theta_0)_{k,il})}{2h_n T} + O_{\mathbb{P}} \left( h_n^{-\epsilon(\kappa \wedge \frac{3}{2})} \right), \quad (10.80)$$

which is exactly (10.77). Finally we show (10.78). A calculation similar to the proofs of Lemmas 10.15 and 10.16 shows that

$$\mathbb{E}_{\theta_0, n} b_n(\widehat{\Theta}_n) = \mathbb{E}_{\theta_0, n} b_n(\theta_0) + O_{\mathbb{P}} \left( h_n^{-\frac{\epsilon}{2}} \right) \quad (10.81)$$

and

$$\mathbb{E}_{\theta_0, n} b_n(\theta_0) = b(\theta_0) + O_{\mathbb{P}} \left( h_n^{-\frac{\epsilon}{2}} \right) \quad (10.82)$$

and this concludes the proof.  $\square$

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

*Proof of Theorem 5.3.* This follows exactly the same argument as for the proof of Theorem 5.2.  $\square$

## 10.5 Proof of the GCLT

In this section we present the proof of Theorem 5.4 using a similar martingale approach as in [33]. 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

[C\*]. We have uniformly in  $i \in \{1, \dots, B_n\}$  that there exists  $\epsilon > 0$  such that

$$\text{Var}_{(i-1)\Delta_n} \left[ \sqrt{h_n} \left( \widehat{\Theta}_{i,n}^{(BC)} - \theta_{(i-1)\Delta_n}^* \right) \right] = T^{-1} \Gamma \left( \theta_{(i-1)\Delta_n}^* \right)^{-1} + o_{\mathbb{P}}(1), \quad (10.83)$$

$$\mathbb{E}_{(i-1)\Delta_n} \left[ \left| \sqrt{h_n} \left( \widehat{\Theta}_{i,n}^{(BC)} - \theta_{(i-1)\Delta_n}^* \right) \right|^{2+\epsilon} \right] = O_{\mathbb{P}}(1), \quad (10.84)$$

$$\mathbb{E}_{(i-1)\Delta_n} \left[ \widehat{\Theta}_{i,n}^{(BC)} - \theta_{(i-1)\Delta_n}^* \right] = o_{\mathbb{P}} \left( n^{-1/2} \right), \quad (10.85)$$

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

The above-mentioned approach is based on techniques introduced in [33], but it is much different and deeper. Indeed, [33] 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 (3.3). We thus obtain different conditions which are hard to verify, and this is the main goal of the proofs.

*Proof of Theorem 5.4 under [C\*].* We split the proof into two parts.

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

$$\Theta = \frac{1}{B_n} \sum_{i=1}^{B_n} \theta_{(i-1)\Delta_n}^* + o_{\mathbb{P}}\left(n^{-1/2}\right). \quad (10.86)$$

Note that (10.86) is to be compared to (3.1) for the toy model. Moreover, (10.86) was also shown in (35) on pp. 46-47 in [33], but the parameter process was restricted to follow a continuous Itô-process. To show (10.86), it is sufficient to show that

$$\frac{\sqrt{n}}{B_n} \sum_{i=1}^{B_n} \left| \theta_{(i-1)\Delta_n}^* - \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} \theta_s^* ds \right| = o_{\mathbb{P}}(1). \quad (10.87)$$

We can bound (10.87) by

$$\frac{\sqrt{n}}{B_n} \sum_{i=1}^{B_n} \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} \underbrace{\left| \theta_{(i-1)\Delta_n}^* - \theta_s^* \right| ds}_{O_{\mathbb{P}}(\Delta_n^\gamma)} = o_{\mathbb{P}}(1), \quad (10.88)$$

where we used [C]-**(i)** to obtain the order in (10.88). Thus, we deduce that the left-hand side in (10.88) 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 (10.86).

**Step 2.** We keep here the techniques and notations introduced in Section 3, 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  $S_n^{(B)} \rightarrow^{\mathbb{P}} 0$  and we will prove the existence of some  $V_T$  such that  $S_n^{(M)} \rightarrow^d V_T^{\frac{1}{2}} \mathcal{N}(0, 1)$ . Note that the former is a straightforward consequence of (10.85). To show the latter  $S_n^{(M)} \rightarrow^d V_T^{\frac{1}{2}} \mathcal{N}(0, 1)$ , we provide a proof which is based on the proof of (37) on p. 47-48 in [33]. We will use Corollary 3.1 of pp. 58-59 in [20]. We verify now the three conditions of the corollary. First, note that the assumption (3.21) on nested filtration holds in our case. We show now the conditional Lindeberg condition (3.7), i.e. that for any  $\eta > 0$  we have

$$\frac{n}{B_n^2} \sum_{i=1}^{B_n} \mathbb{E}_{(i-1)\Delta_n} \left[ M_{i,n}^2 \mathbf{1}_{\left\{ \frac{\sqrt{n}}{B_n} M_{i,n} > \eta \right\}} \right] \rightarrow^{\mathbb{P}} 0. \quad (10.89)$$

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

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

On the one hand we have that  $a_{i,n}$  is uniformly bounded in view of (10.84) from [C\*]. On the other hand, using also Theorem (10.84) along with [C]-**(ii)**, we have that  $b_{i,n}$  goes uniformly to 0. We have thus proved (10.89). We now prove the conditional variance condition, i.e. that

$$\frac{n}{B_n^2} \sum_{i=1}^{B_n} \mathbb{E}_{(i-1)\Delta_n} [M_{i,n}^2] \xrightarrow{\mathbb{P}} V_T := T^{-2} \int_0^T \Gamma(\theta_s^*)^{-1} ds. \quad (10.90)$$

We have that

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

We use Proposition I.4.44 on p. 51 in [26] along with (10.83) from [C\*] to show (10.90). Finally, note that the convergence  $S_n^{(M)} \xrightarrow{d} V_T^{\frac{1}{2}} \mathcal{N}(0, 1)$  is mixing as a simple consequence of Corollary 3.3 in p.64 of [20]. In particular, we have that  $V_T$  and  $\mathcal{N}(0, 1)$  are independent from each other.  $\square$

We prove now that we can obtain (10.83), (10.84) and (10.85) in Condition [C\*]. First note that for any  $L \in (0, 2\kappa)$ , a calculation gives

$$\mathbb{E}_{(i-1)\Delta_n} \left| \sqrt{h_n} \left( \widehat{\Theta}_{i,n}^{(BC)} - \widehat{\Theta}_{i,n} \right) \right|^L = h_n^{-\frac{L}{2}} T^{-L} \mathbb{E}_{(i-1)\Delta_n} \left| b_{i,n} \left( \widehat{\Theta}_{i,n} \right) \right|^L = O_{\mathbb{P}} \left( h_n^{-\frac{L}{2}} \right)$$

uniformly in  $i \in \{1, \dots, B_n\}$ . Thus, combining the previous estimate with Theorem 5.2, we have shown that Theorem 5.2 remains true if  $\widehat{\Theta}_{i,n}$  is replaced by  $\widehat{\Theta}_{i,n}^{(BC)}$ . We will use this fact in the following. If we decompose the conditional variance in (10.83) as

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

then (10.83) follows from Theorem 5.2. Moreover, (10.84) is a direct consequence of Theorem 5.2. Finally, in view of (5.25) in Theorem 5.3, (10.85) holds if there exists  $\epsilon \in (0, 1)$  such that  $\sqrt{n} = o_{\mathbb{P}} \left( h_n^{\epsilon(\kappa \wedge \frac{3}{2})} \right)$ . 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  $\kappa$  by its expression, we get the two conditions  $\frac{\delta}{2} < \gamma(\delta - 1)$  and  $\frac{\delta}{2} < \frac{3}{2}$ , that is  $\frac{\gamma}{\gamma-1} < \delta < 3$ . This is exactly condition [BC].

## 10.6 Proof of Proposition 5.8

*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 5.4. (10.83) and (10.84) are true since  $\delta > 1 + \frac{1}{\gamma}$ . Moreover, by assumption on  $\delta$  and  $\alpha$ , (10.85) is replaced by  $\mathbb{E}_{(i-1)\Delta_n} \left[ \widehat{\Theta}_{i,n}^{(BC)} - \theta_{(i-1)\Delta_n}^* \right] = O_{\mathbb{P}} \left( n^{-\gamma(1-\delta^{-1}) \wedge \delta^{-1}} \right) = o_{\mathbb{P}}(n^{-\alpha})$ . Writing the decomposition

$$\frac{n^\alpha}{B_n} \sum_{i=1}^{B_n} \left( \widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^* \right) = n^{\alpha - \frac{1}{2}} \left\{ S_n^{(B)} + S_n^{(M)} \right\}, \quad (10.91)$$

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

## 10.7 Proof of Proposition 6.1

Note that for any  $\theta \in K$ , we have

$$\partial_\xi^2 l_{i,n}(n^{-1}\xi)_{|\xi=n\theta} = n^{-2} \partial_\theta^2 l_{i,n}(\theta), \quad (10.92)$$

and thus

$$\begin{aligned} n^{-1} \widehat{C}_n &= \frac{1}{B_n} \sum_{i=1}^{B_n} \partial_\theta^2 l_{i,n} \left( \widehat{\Theta}_{i,n} \right)^{-1} h_n \\ &= \frac{1}{TB_n} \sum_{i=1}^{B_n} \Gamma_{i,n} \left( \widehat{\Theta}_{i,n} \right)^{-1}, \end{aligned}$$

so that it is sufficient to prove uniformly in  $i \in \{1, \dots, B_n\}$  the estimates

$$\Gamma_{i,n} \left( \widehat{\Theta}_{i,n} \right)^{-1} = \Gamma \left( \theta_{(i-1)\Delta_n}^* \right)^{-1} + o_{\mathbb{P}}(1) \quad (10.93)$$

and

$$\Gamma \left( \theta_{(i-1)\Delta_n}^* \right)^{-1} = \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} \Gamma(\theta_t^*)^{-1} dt + o_{\mathbb{P}}(1). \quad (10.94)$$

To show (10.93), we consider the decomposition

$$\Gamma_{i,n} \left( \widehat{\Theta}_{i,n} \right)^{-1} - \Gamma \left( \theta_{(i-1)\Delta_n}^* \right)^{-1} = \underbrace{\Gamma_{i,n} \left( \widehat{\Theta}_{i,n} \right)^{-1} - \Gamma_{i,n} \left( \theta_{(i-1)\Delta_n}^* \right)^{-1}}_{a_{i,n}} + \underbrace{\Gamma_{i,n} \left( \theta_{(i-1)\Delta_n}^* \right)^{-1} - \Gamma \left( \theta_{(i-1)\Delta_n}^* \right)^{-1}}_{b_{i,n}}.$$

We have that

$$|a_{i,n}| \leq \sup_{\theta \in K} \frac{1}{h_n} \left| \partial_\theta (\partial_\theta^2 l_{i,n}(\theta))^{-1} \right| \left| \widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^* \right|. \quad (10.95)$$

By some algebraic calculus it is straightforward to show that the term  $\sup_{\theta \in K} \frac{1}{h_n} \left| \partial_\theta (\partial_\theta^2 l_{i,n}(\theta))^{-1} \right|$  is  $\mathbb{L}_p$  bounded by virtue of Lemma 10.9 (i) and Lemma 10.14 (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 10.14 (ii). Thus (10.93) holds. Finally the approximation (10.94) is a straightforward consequence of Lemma 10.9 (i) and Lemma 10.14 (i) along with assumption [C]- (i).

## Acknowledgments

The research of Simon Clinet is in part supported by CREST Japan Science and Technology Agency : Grants-in-Aid for Scientific Research No. 24340015 (Scientific Research), No. 26540011 (Challenging Exploratory Research). The research of Yoann Potiron is supported by the National Science Foundation [DMS 14-07812] and Keio University. All financial data is provided by the Chair of Quantitative Finance of the Ecole Centrale Paris. We would like to thank Per Mykland, Nakahiro Yoshida, Frederic Abergel, Feng Chen, Holger Dette (the Editor), an anonymous Associate Editor and two anonymous referees for helpful discussions and advices.

## References

- [1] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.
- [2] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Some limit theorems for hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499, 2013.
- [3] E. Bacry, I. Mastromatteo, and J.-F. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.
- [4] C. G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.
- [5] L. Breiman. *Probability, volume 7 of Classics in Applied Mathematics*. 1992.
- [6] P. Brémaud and L. Massoulié. Stability of nonlinear hawkes processes. *The Annals of Probability*, pages 1563–1588, 1996.
- [7] F. Chen and P. Hall. Nonparametric estimation for self-exciting point processes -a parsimonious approach. *Journal of Computational and Graphical Statistics*, 25(1):209–224, 2016.
- [8] F. Chen, P. Hall, et al. 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, 2013.
- [9] S. Clinet and Y. Potiron. Efficient asymptotic variance reduction when estimating volatility in high frequency data. *arXiv preprint arXiv:1701.01185*, 2017.
- [10] S. Clinet and N. Yoshida. Statistical inference for ergodic point processes and application to limit order book. *Stochastic Processes and their Applications*, 2016.
- [11] D. R. Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 129–164, 1955.
- [12] D. R. Cox and E. J. Snell. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 248–275, 1968.
- [13] R. Dahlhaus. Fitting time series models to nonstationary processes. *The annals of Statistics*, 25(1):1–37, 1997.
- [14] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.
- [15] P. Embrechts, T. Liniger, L. Lin, et al. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.
- [16] R. F. Engle and J. R. Russell. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162, 1998.
- [17] J. Fan and I. Gijbels. *Local polynomial modelling and its applications: monographs on statistics and applied probability*, volume 66. CRC Press, 1996.

- [18] M. Fernandes and J. Grammig. A family of autoregressive conditional duration models. *Journal of Econometrics*, 130(1):1–23, 2006.
- [19] E. W. Fox, M. B. Short, F. P. Schoenberg, K. D. Coronges, and A. L. Bertozzi. Modeling e-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association*, 111(514):564–584, 2016.
- [20] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic press, 1980.
- [21] T. Hastie and R. Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 757–796, 1993.
- [22] A. G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 438–443, 1971.
- [23] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.
- [24] J. Jacod and P. E. Protter. *Discretization of processes*, volume 67. Springer Science & Business Media, 2011.
- [25] J. Jacod, M. Rosenbaum, et al. Quarticity and other functionals of volatility: efficient estimation. *The Annals of Statistics*, 41(3):1462–1484, 2013.
- [26] J. Jacod and A. Shiryaev. *Limit theorems for stochastic processes (2nd ed.)*, volume 288. Berlin: Springer-Verlag, 2003.
- [27] T. Jaisson, M. Rosenbaum, et al. Limit theorems for nearly unstable hawkes processes. *The Annals of Applied Probability*, 25(2):600–631, 2015.
- [28] P. A. Mykland and L. Zhang. Inference for continuous semimartingales observed at high frequency. *Econometrica*, 77(5):1403–1445, 2009.
- [29] P. A. Mykland and L. Zhang. The econometrics of high frequency data. *Statistical methods for stochastic differential equations*, 124:109, 2012.
- [30] D. Oakes. The markovian self-exciting process. *Journal of Applied Probability*, pages 69–77, 1975.
- [31] T. Ogihara and N. Yoshida. Quasi likelihood analysis for point process regression models. *preprint*, 2015.
- [32] T. Ozaki. Maximum likelihood estimation of hawkes’ self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.
- [33] Y. Potiron. Estimating the integrated parameter of the locally parametric model in high-frequency data. *arXiv preprint arXiv:1603.05700*, 2016.
- [34] Y. Potiron and P. A. Mykland. Estimation of integrated quadratic covariation with endogenous sampling times. *Journal of Econometrics*, 197(1):20–41, 2017.
- [35] E. Renault, T. Van der Heijden, and B. J. Werker. The dynamic mixed hitting-time model for multiple transaction prices and times. *Journal of Econometrics*, 180(2):233–250, 2014.
- [36] F. Roueff, R. von Sachs, and L. Sansonnet. Locally stationary hawkes processes. *Stochastic Processes and their Applications*, 126(6):1710–1743, 2016.

- [37] I. M. Toke. "market making" in an order book model and its impact on the spread. In *Econophysics of Order-driven Markets*, pages 49–64. Springer, 2011.
- [38] G. J. Van den Berg. Duration models: specification, identification and multiple durations. *Handbook of econometrics*, 5:3381–3460, 2001.
- [39] N. Yoshida. Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Annals of the Institute of Statistical Mathematics*, 63(3):431–479, 2011.
- [40] M. Y. Zhang, J. R. Russell, and R. S. Tsay. A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, 104(1):179–207, 2001.
- [41] B. Zheng, F. Roueff, and F. Abergel. Modelling bid and ask prices using constrained hawkes processes: Ergodicity and scaling limit. *SIAM Journal on Financial Mathematics*, 5(1):99–136, 2014.