

Statistical inference for the doubly stochastic self-exciting process

SIMON CLINET^{1,2} and YOANN POTIRON³

¹*Graduate School of Mathematical Sciences, University of Tokyo: 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan. url: <http://www.ms.u-tokyo.ac.jp/~simon/>; E-mail: simon@ms.u-tokyo.ac.jp*

²*CREST, Japan Science and Technology Agency, Japan*

³*Faculty of Business and Commerce, Keio University, 2-15-45 Mita, Minato-ku, Tokyo, 108-8345, Japan. url: <http://www.fbc.keio.ac.jp/~potiron/>; E-mail: potiron@fbc.keio.ac.jp*

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 θ_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; self-exciting process; stochastic; 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, [15] introduced the autoregressive conditional duration (ACD) model. Other references include and are not limited to [35,37], as well as [4,17], and more recently [33] and [32].

The cited work is partly based on the self-exciting Hawkes point process introduced in [20] and [21]. 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, [26] documented such time-clustering property in the order flow of several stocks. Other examples of application can be found in [1,2,14,25,38], 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 [28]. Time-varying parameter extensions have already been considered taking the locally stationary processes approach in [34], and restricting to the baseline time-varying case in [5,18] and [6]. Our approach is much in line with the latter couple of work in that we consider the high-frequency point of view. In [5] the authors allow the background parameter (as they call it) ν 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 [15] 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 [9], and whose numerical computation can be consulted in [30]. 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

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

where $\widehat{\Theta}_{i,n}$ 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 [23] or [27]. Our own recent work includes [7]. The more general literature on local parametric approaches, when not considering the high-frequency data case, includes [16], but also [19], the locally stationary processes of [12], etc.

The first contribution of this paper is to obtain conditions on the stochastic parameter θ_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 [36], 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 Δ_n

will be smaller, and thus the parameter θ_t^* almost constant on each block. In particular, if θ_t^* is not constant, we obtain that a necessary condition is

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

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

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

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

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

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 2κ , 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 Supplement [8].

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 [13] or [24] 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 $T \rightarrow \infty$. [9] 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 θ_t^* is assumed to be 1-dimensional throughout the rest of this section. The parameter θ_t^* is also restricted to belong to a compact set $K = [\underline{\theta}, \bar{\theta}]$, where $\underline{\theta} > 0$. Moreover, θ_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 $\hat{\Theta}_n$ follows [31]. We are interested in assessing the GCLT $\sqrt{n}(\hat{\Theta}_n - \Theta) \rightarrow^d V_T^{\frac{1}{2}} \mathcal{N}(0, 1)$, 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 θ_t^* is smooth, we obtain

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

Consequently, the GCLT will follow if we can prove that

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

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 array and an array of biases. Formally, (3.2) is expressed as

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

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 Theorem 2.2.14, p. 57 of [22]) will hold if¹ uniformly in $i \in \{1, \dots, B_n\}$ we can show that

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

and for some $\kappa > 1$ that

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

If we show the LCLT, that is, 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 2κ , LCLT and no asymptotic bias) in the toy model.

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 2.2.14 of p. 57 in [22]), we obtain the CLT

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

¹The reader can find more details in [8]

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 $p \geq 1$:

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

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

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

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

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

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 2κ . It remains to control the array of biases $S_n^{(B)}$. Calculation gives us

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

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 θ_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

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

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 [10]. 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

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

where $\theta^* = (v^*, a^*, b^*)$ 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 $(nv^*)^{-1}$. To wrap it up, N_t^n is a self-exciting process with parameters (nv^*, na^*, nb^*) . Note that other choices can lead to fairly different asymptotics such as the ones in [29] where authors suggested a model with baseline nv^* 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 θ_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

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

where N_t^n and θ_t^* are adapted to \mathcal{F}_t ,² 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 θ_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.

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 θ_t that ensure the existence of the related doubly stochastic Hawkes process.

²The formal definition of \mathcal{F}_t can be found in Section 5.

[E] (i)

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

(ii)

$$\int_0^T v_s ds < +\infty, \quad \mathbb{P}\text{-a.s.} \tag{5.2}$$

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 θ_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 θ_t . We recall that \mathcal{F}_t^θ designates the canonical filtration associated with θ_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 θ_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*

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

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

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

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 θ_t^* satisfies Condition [E]. Under this assumption, since N_t^n is a time-varying self-exciting process with parameters (nv_t^*, na_t^*, nb_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 $\hat{\Theta}_{i,n}$ as well as its first order bias-corrected version $\hat{\Theta}_{i,n}^{(BC)}$. We state their asymptotic properties, including the main result of this paper which is the GCLT for $\hat{\Theta}_n^{(BC)}$ 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

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

defined for $t \in ((i - 1)\Delta_n, i\Delta_n]$. We now define the Quasi Log Likelihood³ on the i th block as

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

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

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

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.⁴ 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_i^* , nb_i^*) 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

$$h_n = n^{1/\delta}. \tag{5.8}$$

We will also have to specify the smoothness of the process θ^* 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

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

We then define the global regularity modulus as

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

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

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

(ii) δ and γ satisfy the relation

$$\delta > 1 + \frac{1}{\gamma}. \tag{5.12}$$

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

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

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

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

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 [8]. Condition [C](i) quantifies the regularity of the process θ_t^* through the regularity exponent γ . A natural example of a process satisfying [C](i) is the drift function, that is, of the form

$$\theta_t^* = \theta_0^* + \int_0^t u_s^* ds, \tag{5.14}$$

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 $\sigma = \text{diag}(\sigma^v, \sigma^a, \sigma^b)$ and consider the process

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

where $(W_t)_{t \in [-\tau, T]}$ is a 3-dimensional standard Brownian motion. One can confine θ_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 τ , the less auto-correlated θ_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 θ_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 $\bar{a} < \bar{b}$.

We specify now the value of the exponent

$$\kappa = \gamma(\delta - 1) > 1, \tag{5.16}$$

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 θ and can be found in (10.9) from [8]. The next theorem encompasses the LCLT and the local convergence of moments of order smaller than 2κ 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*

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

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

We now introduce the first-order bias-corrected local MLE for any $i \in \{1, \dots, B_n\}$ as

$$\widehat{\Theta}_{i,n}^{(BC)} = \widehat{\Theta}_{i,n} - \frac{b(\widehat{\Theta}_{i,n})}{h_n T}, \quad (5.18)$$

where b is defined on equation (10.14) in [8], and should be compared to its very similar form for the classical i.i.d. case, see, for example, [11]. We finally recall the definition of the global bias-corrected estimator that was introduced in (1.5), i.e.

$$\widehat{\Theta}_n^{(BC)} = \frac{1}{B_n} \sum_{i=1}^{B_n} \widehat{\Theta}_{i,n}^{(BC)}. \quad (5.19)$$

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*

$$\mathbb{E}_{(i-1)\Delta_n} [\widehat{\Theta}_{i,n} - \theta_{(i-1)\Delta_n}^*] = \frac{b(\theta_{(i-1)\Delta_n}^*)}{h_n T} + O_{\mathbb{P}}(h_n^{-\epsilon(\kappa \wedge \frac{3}{2})}), \quad (5.20)$$

uniformly in $i \in \{1, \dots, B_n\}$. Moreover, the estimator $\widehat{\Theta}_{i,n}^{(BC)}$ has the uniform bias expansion

$$\mathbb{E}_{(i-1)\Delta_n} [\widehat{\Theta}_{i,n}^{(BC)} - \theta_{(i-1)\Delta_n}^*] = O_{\mathbb{P}}(h_n^{-\epsilon(\kappa \wedge \frac{3}{2})}). \quad (5.21)$$

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 δ and γ so that $\widehat{\Theta}_n^{(BC)}$ is asymptotically unbiased.

[BC] δ and γ satisfy the relation

$$\frac{\gamma}{\gamma - \frac{1}{2}} < \delta < 3. \quad (5.22)$$

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 θ_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 δ 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, $\mathcal{F}_T^{\theta^*}$ -stably in law as $n \rightarrow \infty$,*

$$\sqrt{n}(\widehat{\Theta}_n^{(BC)} - \Theta) \rightarrow \left(T^{-2} \int_0^T \Gamma(\theta_s^*)^{-1} ds \right)^{\frac{1}{2}} \mathcal{N}(0, 1), \tag{5.23}$$

where $\mathcal{N}(0, 1)$ is independent from the σ -field $\mathcal{F}_T^{\theta^*}$.

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 non-parametric efficient bound.

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

$$\theta_t^* = \theta_t^{(C)} + \theta_t^{(J)}, \tag{5.24}$$

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, that is, $J_0 = 0$. 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 $1/2$, although the some of the corresponding bias theoretical formula terms would be too involved to be of any practical interest. Nonetheless the bias can be computed with Monte Carlo methods (see Section 7.1 in our numerical study for more details). We then provide the expected convergence rate of the consistency for both the naive and the first order bias-corrected estimators.

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, (5.21), whose order in $h_n^{-\kappa \wedge \frac{3}{2}}$ can be dominated by $n^{-\frac{1}{2}}$ only if $\gamma > 3/4$. Nevertheless, we can expect that correcting for the bias to a higher order improves the rate of convergence in (5.21). Thus we would obtain a corresponding 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 $\gamma \in (\frac{1}{2} + \frac{1}{2(1+q)}, 1]$. 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 $\gamma > 1/2$.

When $\gamma \in (0, \frac{1}{2}]$, we can't use the same martingale approach. In general, the bias induced by the parameter process in the expansion (5.21) cannot be corrected without some information on the distribution of θ_t^* . We can't show that the bias is of the right order, because this would imply a choice of δ such that $\gamma < \delta(\gamma - 1/2)$ which is not possible if γ 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 θ_t^* is beyond the scope of this work.

We now turn to the convergence rate of our estimators when the central limit theorem fails. We can prove that both estimators $\widehat{\Theta}_n$ and $\widehat{\Theta}_n^{(BC)}$ are consistent. Indeed, it turns out that the first order bias corrected version is n^α -consistent⁵ for any $\alpha \in (0, \frac{\gamma}{1+\frac{2}{3}\gamma})$, whereas the naive estimator is only n^α -consistent for any $\alpha \in (0, \frac{\gamma}{1+\gamma})$. More specifically, we have the following result. The proposition can be showed following a similar reasoning as in the proof of Theorem 5.4.

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

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

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

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

In particular, we can see 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 α given in Proposition 5.8 are optimal is beyond the scope of this paper.

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

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 Θ . When properly divided by n , this yields the same estimate as the non feasible procedure, that is, we have $n\widehat{\Theta} = n\widehat{\Theta}$, 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

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

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 ξ 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

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

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*

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

Moreover, we have the convergence in distribution

$$\widehat{C}_n^{-1/2} (n\widehat{\Theta}_n^{(BC)} - n\Theta) \rightarrow \mathcal{N}(0, 1). \tag{6.3}$$

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 θ_t^* so that the expected number of events between 0 and T is roughly one when parameters are equal to θ_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

$$Z_n^{(\text{BC})} = \widehat{C}_n^{-1/2} (n\widehat{\Theta}_n^{(\text{BC})} - n\Theta) \rightarrow \mathcal{N}(0, 1)$$

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

$$Z_n = \widehat{C}_n^{-1/2} (n\widehat{\Theta}_n - n\Theta).$$

Finally, we compare the performance of $\widehat{\Theta}_n^{(\text{BC})}$ 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 [5] (CH) that assumes that $v_t^* = f(t, \theta)$ with f being a polynomial of order 3. More specifically in this setting the MLE estimates (θ, 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 [30]. To compute $\widehat{\Theta}_n^{(\text{BC})}$, we can either implement the function defined in (5.18) or carry out Monte-Carlo simulations to compute $b_{i,n}(\theta)$ for any θ 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 θ and a grid of block length Δ 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 10 am to 4 pm. 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 [30] (Section 4, pp. 148–149). The integrated parameter is set to $n\Theta \approx (0.8, 11, 30)$, 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 $\theta^{(M)} = (0.8, 11, 30)$ and the amplitude is set to $A = (0.5, 4, 10)$. This means that θ_t^* takes values in $[0.3, 1.3] \times [7, 15] \times [20, 40]$, which is comparable to the daily variation in our empirical results. In Model II θ_t^* oscillates around $\theta^{(M)} = (0.8, 11, 30)$ 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 taken directly from the literature. We keep a_t^* and b_t^* constant whereas v_t^* follows a usual intraday pattern, so that CH is well specified for this model. As pointed out in [15] (see discussions in Sections 5–6 and Figure 2), the expected duration before the next trade tends to follow a U-shape intraday pattern. This diurnal effect motivated [5] (see Section 5, pp. 1011–1017) to model a Hawkes process where v_t^* is time-varying with a quadratic form. The model is written as $v_t^* = e^{\beta_1} + \{e^{\beta_2} + e^{\beta_3}\}^2 (t/T - e^{\beta_2} / (e^{\beta_2} + e^{\beta_3}))^2$. We fit the model to the empirical intraday mean and find $\beta_1 \approx -0.84$, $\beta_2 \approx -0.26$ and $\beta_3 \approx -0.39$, which implies that $T^{-1} \int_0^T v_t^* dt \approx 0.61$. 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 more realistic considerations where a_t^* and b_t^* also feature intraday seasonality. In addition, we allow for additive stochastic component in the three parameters. We assume that

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

where $(\beta_1^v, \beta_2^v, \beta_3^v) \approx (-0.84, -0.26, -0.39)$, $(\beta_1^a, \beta_2^a, \beta_3^a) \approx (2.35, -0.05, 0.40)$, 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 $(\sigma^v, \sigma^a, \sigma^b) = (0.8/(6T^{1/2}), 11/(6T^{1/2}), 30/(6T^{1/2}))$ and $\tilde{W}_t = (\tilde{W}_t^v, \tilde{W}_t^a, \tilde{W}_t^b) = \int_{t-1}^t W_s ds$ with $W_t = (W_t^v, W_t^a, W_t^b)$ 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 v_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 addition, Figure 1 provides the associated QQ-plot. In this case, the sample mean is very close to 0 indicating that our proposed

Table 1. Finite sample properties of Z_n for several models[†]

Param.	Mean	Stdv.	RMSE	0.5%	2.5%	5%	95%	97.5%	99.5%
Model I									
ν^*	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									
ν^*	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									
ν^*	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									
ν^*	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

[†]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 $M = 1000$ Monte-Carlo simulations.

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. In Model III, CH performs a MLE with no misspecification. Although CH performs much better than in the other models, it still doesn't outperform the local approach, which indicates that the local approach can perform better even on standard parametric models. Both estimators are badly biased in case of misspecification for them. More surprisingly, although Model III follows a model included in [5] 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).

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.

Table 2. Finite sample properties of $Z_n^{(BC)}$ for several models[†]

Param.	Mean	Stdv.	RMSE	0.5%	2.5%	5%	95%	97.5%	99.5%
Model I									
ν^*	-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									
ν^*	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									
ν^*	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									
ν^*	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

[†]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 $M = 1000$ Monte-Carlo simulations.

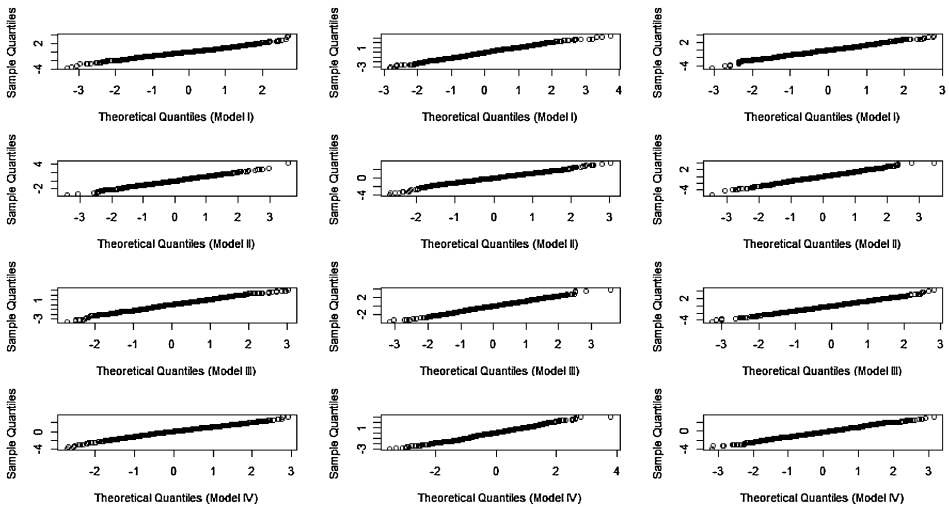


Figure 1. QQ-plots benchmarked to the $N(0, 1)$ distribution for the feasible Z-statistics related to the bias-corrected estimator with $h_n = 273$ (which corresponds to a 216 second block length). The left column corresponds to ν_t^* , the column in the middle for a_t^* and the right column for b_t^* . The simulation design is Model I–IV with $M = 1000$ Monte-Carlo simulations.

Table 3. Performance of 8 estimators for several models

Est.	ν^*		a^*		b^*	
	Mean	Stdv.	Mean	Stdv.	Mean	Stdv.
Model I						
naive 2 m	0.009	0.007	0.286	0.196	1.239	0.630
BC 2 m	0.000	0.007	-0.012	0.201	-0.124	0.658
naive 4 m	0.005	0.007	0.134	0.189	0.546	0.495
BC 4 m	0.000	0.007	0.002	0.192	0.005	0.503
naive 7 m	0.002	0.007	0.068	0.188	0.271	0.477
BC 7 m	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 2 m	0.009	0.007	0.287	0.213	1.346	0.734
BC 2 m	0.000	0.007	-0.018	0.218	-0.078	0.744
naive 4 m	0.005	0.007	0.126	0.198	0.538	0.516
BC 4 m	0.000	0.007	-0.009	0.201	-0.019	0.525
naive 7 m	0.002	0.007	0.057	0.196	0.245	0.490
BC 7 m	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 2 m	0.009	0.006	0.348	0.235	1.474	0.734
BC 2 m	0.000	0.006	-0.009	0.241	-0.108	0.742
naive 4 m	0.005	0.005	0.158	0.227	0.645	0.568
BC 4 m	0.000	0.005	-0.003	0.241	-0.015	0.578
naive 7 m	0.002	0.006	0.074	0.225	0.316	0.543
BC 7 m	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 2 m	0.009	0.006	0.624	0.332	3.533	2.579
BC 2 m	0.000	0.006	-0.011	0.311	-0.199	2.333
naive 4 m	0.005	0.006	0.276	0.274	1.389	1.022
BC 4 m	0.000	0.006	-0.006	0.270	-0.008	0.967
naive 7 m	0.003	0.006	0.133	0.265	0.655	0.889
BC 7 m	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

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), 216 s (roughly 4 m) and 432 s (roughly 7 m)), the MLE and the CH. The simulation design is Model I-IV with $M = 1000$ Monte-Carlo simulations.

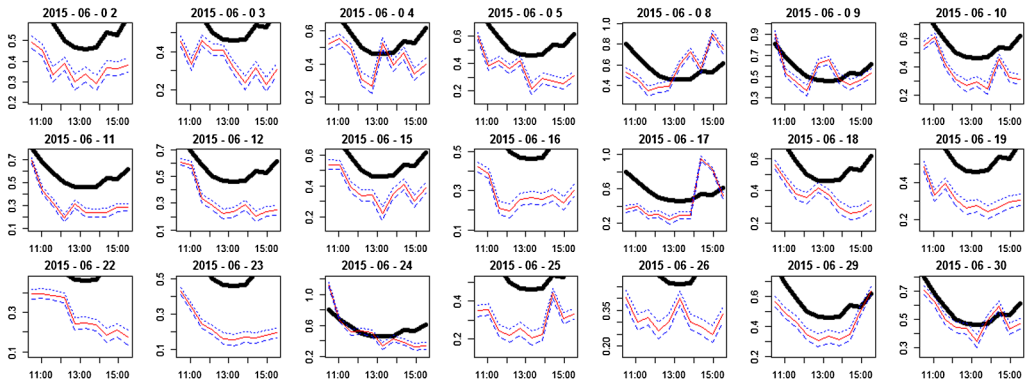


Figure 2. Local estimated ν^* parameter on 30 minute long blocks in June 2015. The two dashed lines correspond to the 95% confidence intervals. The thick line stands for the seasonality intraday effect, estimated as a temporal local mean across all the trading days in 2015.

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 NASDAQ in 2015. Our aim is twofold. First, using relatively large (30 minute) local blocks, we document about seasonality and intraday variability 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 9:30 am and 3:30 pm, 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 3000 for the least active days.

In Figures 2–4, 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. 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 parameter⁶ 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.

⁶Probably even better if we add a “day effect” in the model.

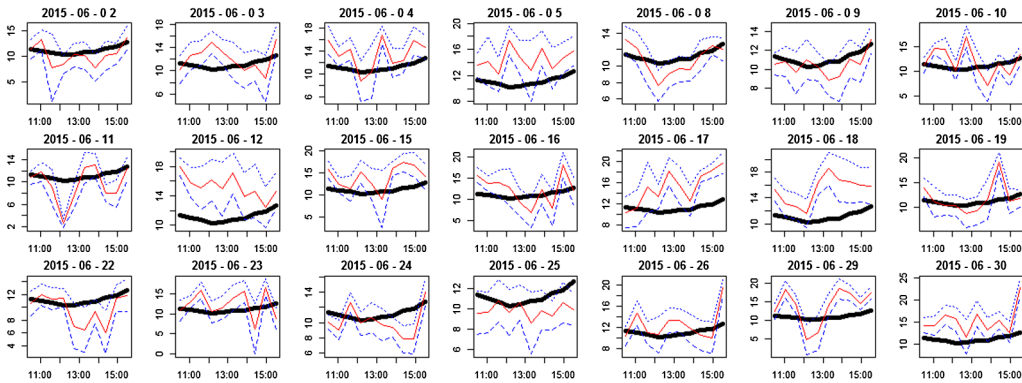


Figure 3. Local estimated a^* parameter on 30 minute long blocks in June 2015. The two dashed lines correspond to the 95% confidence intervals. The thick line stands for the seasonality intraday effect, estimated as a temporal local mean across all the trading days in 2015.

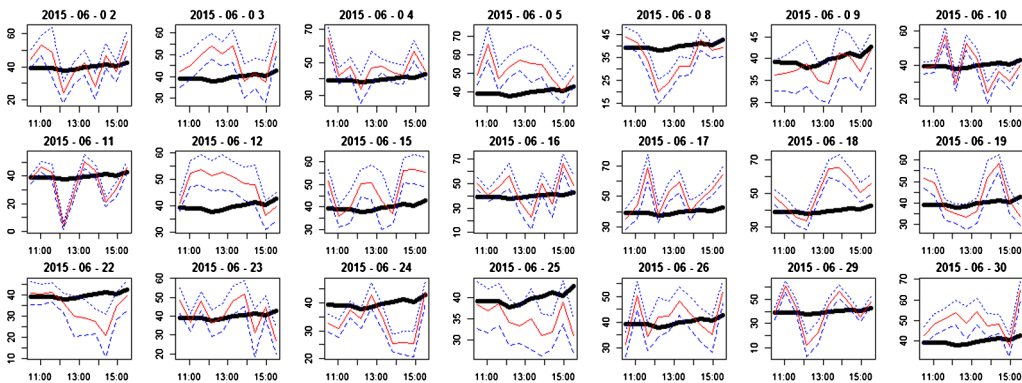


Figure 4. Local estimated b^* parameter on 30 minute long blocks in June 2015. The two dashed lines correspond to the 95% confidence intervals. The thick line stands for the seasonality intraday effect, estimated as a temporal local mean across all the trading days in 2015.

In Table 4, we report statistics of the implemented estimators. As our method is non parametric, the assumption of any particular parametric model for the time-varying parameter is not required. 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

Table 4. Summary statistics for 12 estimators[†]

Est.	ν^*			a^*			b^*		
	Mean	Stdv.	Corr.(,BC 3)	Mean	Stdv.	Corr.(,BC 3)	Mean	Stdv.	Corr.(,BC 3)
naive 1	0.57	0.24	≈1	12.48	2.18	0.94	49.80	10.33	0.78
BC 1	0.56	0.24	≈1	11.16	2.17	≈1	40.24	9.54	0.98
naive 2	0.57	0.24	≈1	11.78	2.00	0.98	44.74	8.20	0.89
BC 2	0.56	0.24	≈1	11.12	2.00	≈1	39.92	7.79	0.99
naive 3	0.56	0.24	≈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	≈1	11.22	1.96	≈1	40.66	7.40	0.96
BC 4	0.56	0.24	≈1	11.07	1.96	≈1	39.72	7.36	0.94
naive 5	0.56	0.24	≈1	11.07	1.97	0.99	39.60	7.50	0.94
BC 5	0.56	0.24	≈1	11.01	1.97	0.98	39.14	7.50	0.92
MLE	0.55	0.23	≈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

[†]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.

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 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 variability 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 in 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.

Acknowledgments

The research of Simon Clinet is in part supported by CREST Japan Science and Technology Agency. The research of Yoann Potiron is supported by the National Science Foundation [DMS 14-07812], Japanese Society for the Promotion of Science Grant-in-Aid for Young Scientists (60781119) and a private grant from 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 advice.

Supplementary Material

Supplement to “Statistical inference for the doubly stochastic self-exciting process” (DOI: [10.3150/17-BEJ966SUPP](https://doi.org/10.3150/17-BEJ966SUPP); .pdf). The Appendix can be found in the Appendix referenced as [8] below. This material is comprised of one Appendix, which contains all the proofs.

References

- [1] Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quant. Finance* **13** 65–77. [MR3005350](#)
- [2] Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Process. Appl.* **123** 2475–2499. [MR3054533](#)
- [3] Bacry, E., Mastromatteo, I. and Muzy, J.-F. (2015). Hawkes processes in finance. *Market Microstructure and Liquidity* **1** 1550005.
- [4] Bowsher, C.G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *J. Econometrics* **141** 876–912. [MR2413490](#)
- [5] Chen, F. and Hall, P. (2013). Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. *J. Appl. Probab.* **50** 1006–1024. [MR3161370](#)
- [6] Chen, F. and Hall, P. (2016). Nonparametric estimation for self-exciting point processes—a parsimonious approach. *J. Comput. Graph. Statist.* **25** 209–224. [MR3474044](#)
- [7] Clinet, S. and Potiron, Y. (2017). Efficient asymptotic variance reduction when estimating volatility in high frequency data. Preprint. Available at [arXiv:1701.01185](https://arxiv.org/abs/1701.01185).
- [8] Clinet, S. and Potiron, Y. (2017). Supplement to “Statistical inference for the doubly stochastic self-exciting process.” DOI:[10.3150/17-BEJ966SUPP](https://doi.org/10.3150/17-BEJ966SUPP).
- [9] Clinet, S. and Yoshida, N. (2017). Statistical inference for ergodic point processes and application to limit order book. *Stochastic Process. Appl.* **127** 1800–1839. [MR3646432](#)
- [10] Cox, D.R. (1955). Some statistical methods connected with series of events. *J. Roy. Statist. Soc. Ser. B.* **17** 129–157; discussion, 157–164. [MR0092301](#)
- [11] Cox, D.R. and Snell, E.J. (1968). A general definition of residuals. *J. Roy. Statist. Soc. Ser. B* **30** 248–275. [MR0237052](#)
- [12] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. [MR1429916](#)
- [13] Daley, D.J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*, 2nd ed. *Probability and Its Applications* (New York). New York: Springer. [MR2371524](#)

- [14] Embrechts, P., Liniger, T. and Lin, L. (2011). Multivariate Hawkes processes: An application to financial data. *J. Appl. Probab.* **48A** 367–378. [MR2865638](#)
- [15] Engle, R.F. and Russell, J.R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* **66** 1127–1162. [MR1639411](#)
- [16] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. London: Chapman & Hall. [MR1383587](#)
- [17] Fernandes, M. and Grammig, J. (2006). A family of autoregressive conditional duration models. *J. Econometrics* **130** 1–23. [MR2208899](#)
- [18] Fox, E.W., Short, M.B., Schoenberg, F.P., Coronges, K.D. and Bertozzi, A.L. (2016). Modeling e-mail networks and inferring leadership using self-exciting point processes. *J. Amer. Statist. Assoc.* **111** 564–584. [MR3538687](#)
- [19] Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55** 757–796. [MR1229881](#)
- [20] Hawkes, A.G. (1971). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B* **33** 438–443. [MR0358976](#)
- [21] Hawkes, A.G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. [MR0278410](#)
- [22] Jacod, J. and Protter, P. (2012). *Discretization of Processes. Stochastic Modelling and Applied Probability* **67**. Heidelberg: Springer. [MR2859096](#)
- [23] Jacod, J. and Rosenbaum, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *Ann. Statist.* **41** 1462–1484. [MR3113818](#)
- [24] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Berlin: Springer. [MR1943877](#)
- [25] Jaisson, T. and Rosenbaum, M. (2015). Limit theorems for nearly unstable Hawkes processes. *Ann. Appl. Probab.* **25** 600–631. [MR3313750](#)
- [26] Muni Toke, I. (2011). “Market making” in an order book model and its impact on the spread. In *Econophysics of Order-Driven Markets. New Econ. Windows* 49–64. Milan: Springer. [MR3220487](#)
- [27] Mykland, P.A. and Zhang, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* **77** 1403–1445. [MR2561071](#)
- [28] Oakes, D. (1975). The Markovian self-exciting process. *J. Appl. Probab.* **12** 69–77. [MR0362522](#)
- [29] Ogihara, T. and Yoshida, N. (2015). Quasi likelihood analysis of point processes for ultra high frequency data. Preprint. Available at [arXiv:1512.01619](#).
- [30] Ozaki, T. (1979). Maximum likelihood estimation of Hawkes’ self-exciting point processes. *Ann. Inst. Statist. Math.* **31** 145–155. [MR0541960](#)
- [31] Potiron, Y. and Mykland, P.A. (2016). Local parametric estimation in high frequency data. Preprint. Available at [arXiv:1603.05700](#). Version 3.
- [32] Potiron, Y. and Mykland, P.A. (2017). Estimation of integrated quadratic covariation with endogenous sampling times. *J. Econometrics* **197** 20–41. [MR3598643](#)
- [33] Renault, E., van der Heijden, T. and Werker, B.J.M. (2014). The dynamic mixed hitting-time model for multiple transaction prices and times. *J. Econometrics* **180** 233–250. [MR3197795](#)
- [34] Roueff, F., von Sachs, R. and Sansonnet, L. (2016). Locally stationary Hawkes processes. *Stochastic Process. Appl.* **126** 1710–1743. [MR3483734](#)
- [35] Van den Berg, G.J. (2001). Duration models: Specification, identification and multiple durations. In *Handbook of Econometrics* **5** 3381–3460. Elsevier.
- [36] Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Statist. Math.* **63** 431–479. [MR2786943](#)

- [37] Zhang, M.Y., Russell, J.R. and Tsay, R.S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *J. Econometrics* **104** 179–207. [MR1862032](#)
- [38] Zheng, B., Roueff, F. and Abergel, F. (2014). Modelling bid and ask prices using constrained Hawkes processes: Ergodicity and scaling limit. *SIAM J. Financial Math.* **5** 99–136. [MR3164121](#)

Received February 2017 and revised June 2017