Annex of *Statistical inference for the doubly stochastic self-exciting process*

10 Appendix: Proofs

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 $T \to \infty$. 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

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

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

$$\lambda_*^P(t) = \lambda(t, \theta^*). \tag{10.2}$$

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

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

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

$$\Gamma_T(\theta^*) = -\frac{1}{T} \partial_\theta^2 l_T(\theta^*) \in \mathbb{R}^{3\times3}, \tag{10.4}$$

$$K_T(\theta^*) = \frac{1}{T} \partial_\theta^3 l_T(\theta^*) \in \mathbb{R}^{3\times3\times3}, \tag{10.5}$$

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

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

$$C_T(\theta^*)_{k,lm} = \frac{1}{T} \int_0^T \partial_{\theta,k} \lambda(t, \theta^*) \partial_{\theta,lm}^2 \log(\lambda(t, \theta^*))\, dt, \tag{10.7}$$

and

$$Q_T(\theta^*)_{k,lm} = -\frac{M_T(\theta^*)_k}{T} \int_0^T \frac{\partial_\theta \lambda(t, \theta^*)_l \partial_\theta \lambda(t, \theta^*)_m}{\lambda(t, \theta^*)} dt. \tag{10.8}$$

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

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

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

and

$$\mathbb{E}\left|C_T(\theta^*) - C(\theta^*)\right|^p = O\left(T^{-\epsilon\frac{p}{2}}\right), \tag{10.11}$$

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

$$\mathbb{E}\left[Q_T(\theta^*) - Q(\theta^*)\right] = O\left(T^{-\frac{\epsilon}{2}}\right), \tag{10.12}$$

for some $Q(\theta^*) \in \mathbb{R}^{3\times 3\times 3}$. Note that $\Gamma(\theta^*)$ is the asymptotic Fisher information. In particular, in [3] the authors have shown the convergence of moments of the MLE (see Theorem 4.6),

$$\mathbb{E}\left[f\left(\sqrt{T}(\hat{\theta}_T - \theta^*)\right)\right] \to \mathbb{E}\left[f\left(\Gamma(\theta^*)^{-\frac{1}{2}}\xi\right)\right], \tag{10.13}$$

where f can be any continuous function of polynomial growth, and ξ follows a standard normal distribution. Also, it is easy to see that the convergences in (10.9)-(10.13) hold uniformly in $\theta^* \in K$ under a mild change in the proofs of [3]. The result (10.13) should be compared to Theorem 5.2. Finally, from Γ, K, C and Q we define for any $k \in \{0,1,2\}$

$$b(\theta^*)_k = \frac{1}{2}\Gamma(\theta^*)^{jk}\Gamma(\theta^*)^{lm}\left(K(\theta^*)_{jlm} + 2\left\{C(\theta^*)_{l,jm} + Q(\theta^*)_{l,jm}\right\}\right) \tag{10.14}$$

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 [2] 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 $\overline{N}$ of intensity 1 on $\mathbb{R}^2$ which is independent of θ. In other words, $\mathcal{F}_t = \mathcal{F}_t^{(\theta,\overline{N})}$. 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 σ-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. □

We now show the existence of the doubly stochastic Hawkes process associated to θ.

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

$$N_t^0 = \iint_{[0,t]\times\mathbb{R}} \mathbb{1}_{\{0\leq x\leq\lambda^0(s)\}} \overline{N}(ds, dx). \tag{10.15}$$

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 λ^n as

$$\lambda^{n+1}(t) = \nu_t + \int_0^{t-} a_s e^{-b_s(t-s)} dN_s^n, \tag{10.16a}$$

$$N_t^{n+1} = \iint_{[0,t]\times\mathbb{R}} \mathbb{1}_{\{0\leq x\leq\lambda^{n+1}(s)\}} \overline{N}(ds, dx). \tag{10.16b}$$

Note that both λ^n and N^n are increasing with n and thus both converge point-wise to some limiting values λ and N that take their values on $[0, +\infty]$. Moreover, N counts the points of $\overline{N}$ 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 ρ^n defined as $\rho_t^n = \mathbb{E}\left[\lambda^n(t) - \lambda^{n-1}(t) \mid \mathcal{F}_T^\theta\right]$. Then

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

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 $\overline{N}(ds, dz) - ds \otimes dz$ and the independence between $\mathcal{F}_T^{\overline{N}}$ and $\mathcal{F}_T^\theta$. Thus, setting $\Phi_t^n = \int_0^t \rho_s^n ds$, we have by Fubini's theorem

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

Note that $\int_0^{t-s} a_s e^{-b_s u} du \leq \frac{a_s}{b_s} \leq r < 1$ 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 $\left(\sum_{k=0}^n \Phi_t^k\right)_n$ yields

$$\mathbb{E}\left[\int_0^t \lambda(s) ds \middle| \mathcal{F}_T^\theta\right] \leq \int_0^t \nu_s ds + r\mathbb{E}\left[\int_0^t \lambda(s) ds \middle| \mathcal{F}_T^\theta\right]. \tag{10.17}$$

A straightforward rearrangement of the terms in (10.17) gives us that

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

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 mononicity, we deduce by taking the limit $n \to +\infty$ in (10.16a) that

$$\lambda(t) = \nu_t + \int_0^{t-} a_s e^{-b_s(t-s)} dN_s. \tag{10.18}$$

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

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 $\overline{\Lambda}$ the compensating measure of $\overline{N}$, that is $\overline{\Lambda}(ds,dz) = ds \otimes dz$. Given a predictable function W, write $W * \overline{N}_t = \iint_{[0,t]\times\mathbb{R}} W(s,z)\overline{N}(ds,dz)$, and the associated definition for $W * \overline{\Lambda}_t$. Predictable function and integral with respect to random measures definitions can be consulted in [6], paragraph II.1. The following lemma is a straightforward adaptation of Lemma I.2.1.5 in [5], using also Lemma 10.1 and (5.4).

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

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

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

For any (random) kernel $\chi : (s,t) \to \chi(s,t)$, we say that χ 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 χ such that $\int_0^t \chi(s,t)ds$ is bounded uniformly in $t \in [0,T]$ independently from T, and for any predictable process ψ that has uniformly bounded moments independently from T, we have*

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

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

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

Proof. We conduct the proof in three steps.
Step 1. We prove that **(i)** holds for $p = 1$. We write

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

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

$$\sup_{t\in[0,T]} \mathbb{E}[\lambda(t)|\mathcal{F}_T^\theta] \leq (1-c)^{-1}\overline{\nu}. \tag{10.19}$$

In particular this proves the case $p = 1$, since the right hand side of (10.19) 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 $\epsilon > 0$,

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

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 $W(s, z) = a_s e^{-b_s(t-s)}\mathbb{1}_{\{0\leq z\leq\lambda(s)\}}$, and note that

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

We apply now Lemma 10.2 to get

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

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, $(\int fg)^k \leq (\int f^k g)(\int g)^{k-1}$. This along with the induction hypothesis leads to a similar bound for the second term. On the other hand, we have

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

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

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

Finally, we have shown that

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

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

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

and dividing by $\left(1-(1+\epsilon)^{2^q}c^p\right)$ on both sides we get the result.
Step 3. It remains to show **(ii)** and **(iii)**. But note that they are direct consequences of the boundedness of moments of λ along with Lemma 10.2.

□

10.3 LCLT and boundedness of moments of order 2κ

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

$$\lambda^\gamma(t)=\gamma\nu_{\gamma t}+\int_0^{t-}\gamma a_{\gamma s}e^{-\gamma b_{\gamma s}(t-s)}dN_s^\gamma. \tag{10.20}$$

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. □

By virtue of Lemma 10.4, for any block index $i \in \{1, \cdots, B_n\}$, we consider the time change $\tau_i^n : t \mapsto n^{-1}t + (i-1)\Delta_n$ and the point process $(N_s^n)_{\{s\in((i-1)\Delta_n, i\Delta_n]\}}$ in order to get a time changed point process $N^{i,n}$ defined on the time set $[0, h_nT]$ by the formula $N_t^{i,n} = N_{\tau_i^n(t)}^n - N_{(i-1)\Delta_n}^n$. Such process is adapted to the filtration $\mathcal{F}_t^{i,n} = \mathcal{F}_{\tau_i^n(t)}$, for $t \in [0, h_nT]$. The parameter processes are now $(\theta_t^{i,n,*})_{\{t\in[0,h_nT]\}} = (\theta^*_{\tau_i^n(t)})_{\{t\in[0,h_nT]\}}$ whose canonical filtration can be expressed as $\mathcal{F}_t^{\theta^{i,n,*}} = \sigma\{\theta_s^{i,n,*} | 0 \le s \le t\}$, for $t \in [0, h_nT]$. Finally note that the $\mathcal{F}_t^{i,n}$-stochastic intensities are now of the form

$$\lambda_*^{i,n}(t) = \nu_t^{i,n,*} + \int_0^{t-} a_s^{i,n,*} e^{-b_s^{i,n,*}(t-s)} dN_s^{i,n} + R_{i,n}(t), \tag{10.21}$$

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

$$R_{i,n}(t) = \int_0^{(i-1)\Delta_n-} n a_s^* e^{-nb_s^*(\tau_i^n(t)-s)} dN_s^n. \tag{10.22}$$

$R_{i,n}(t)$ 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

$$R_{i,n}(t) \le e^{-\underline{b}t} R_{i,n}(0) \tag{10.23}$$

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

$$N_t^{i,n} = \iint_{[0,t]\times\mathbb{R}_+} \mathbb{1}_{\{0\le z\le \lambda_*^{i,n}(s)\}} \overline{N}^{i,n}(ds, dz). \tag{10.24}$$

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

$$\tilde{\lambda}^{i,n}(t,\theta) = \nu + \int_0^{t-} ae^{-b(t-s)} dN_s^{i,n}, \tag{10.25}$$

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 $\log(n)N_{h_nT}^{i,n}$)

$$l_{i,n}(\theta) = \int_0^{h_nT} \log(\tilde{\lambda}^{i,n}(t,\theta)) dN_t^{i,n} - \int_0^{h_nT} \tilde{\lambda}^{i,n}(t,\theta) dt, \tag{10.26}$$

Note that in our case, the true underlying intensity, $\lambda_*^{i,n}$ does not belong to the regression family $(\tilde{\lambda}^{i,n}(.,\theta))_{\theta\in K}$ for two reasons : the parameter process θ^* is not constant on the i-th block, and the regression family does not take into account the existence of a pre-excitation term in (10.21). We are in a mispecified case, but we wish to take advantage of the continuity of the process θ^* to show that the asymptotic theory still holds, that is, the MLE tends to the value $\theta_0^{i,n,*} = \theta^*_{(i-1)\Delta_n}$ which is the value of the process θ^* at the beginning of the i-th block. The procedure is thus asymptotically equivalent to performing the MLE on the model whose stochastic intensity is in the regression family with true value $\theta = \theta^*_{(i-1)\Delta_n}$. To formalize such idea, we introduce an auxiliary model corresponding to the parametric case generated by the true value $\theta^*_{(i-1)\Delta_n}$. More precisely, we introduce the constant parameter Hawkes process $N^{i,n,c}$ generated by $\overline{N}^{i,n}$ and the initial value $\theta_0^{i,n,*}$, whose stochastic intensity satisfies

$$\lambda^{i,n,c}(t) = \nu_0^{i,n,*} + \int_0^{t-} a_0^{i,n,*} e^{-b_0^{i,n,*}(t-s)} dN_s^{i,n,c}. \tag{10.27}$$

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

$$N_t^{i,n,c} = \iint_{[0,t]\times\mathbb{R}_+} \mathbb{1}_{\{0\leq z\leq \lambda^{i,n,c}(s)\}} \overline{N}^{i,n}(ds,dz). \tag{10.28}$$

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[1] (see for instance Section 4.3 (pp. 77 − 80) in [1]). 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 (θ_0, i, n, t). For that reason we introduce the notation $\mathbf{E} = \{ (\theta_0, i, n, t) \in K \times \mathbb{N}^2 \times \mathbb{R}_+ \big| 1 \leq i \leq B_n \text{ and } 0 \leq t \leq h_n T\}$. When $n \in \mathbb{N}$ is fixed, we define $\mathbf{E}_n$ the subset of $\mathbf{E}$ as $\mathbf{E}_n = \{ (\theta_0, i, t) \in K \times \mathbb{N} \times \mathbb{R}_+ \big| 1 \leq i \leq B_n \text{ and } 0 \leq t \leq h_n T\}$. 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 χ such that $\int_0^t \chi(s,t)ds$ is bounded uniformly in $t \in [0, h_n T]$ independently from T and n,*

(i) $\sup_{(\theta_0,i,n,t)\in\mathbf{E}} \mathbb{E}_{\theta_0,i,n} \left|\lambda_*^{i,n}(t)\right|^p \leq M_p$ *$\mathbb{P}$-a.s.*

(ii) $\sup_{(\theta_0,i,n,t)\in\mathbf{E}} \mathbb{E}_{\theta_0,i,n} \left|\int_0^t \chi(s,t) dN_s^{i,n}\right|^p < M_{p,\chi}$ *$\mathbb{P}$-a.s.*

(iii) $\sup_{(\theta_0,i,n,t)\in\mathbf{E}} \mathbb{E}_{\theta_0,i,n} \left|\lambda^{i,n,c}(t)\right|^p < M_p$ *$\mathbb{P}$-a.s.*

(iv) $\sup_{(\theta_0,i,n,t)\in\mathbf{E}} \mathbb{E}_{\theta_0,i,n} \left|\int_0^t \chi(s,t) dN_s^{i,n,c}\right|^p < M_{p,\chi}$ *$\mathbb{P}$-a.s.*

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

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

Proof. This is a straightforward adaptation of the proof of Lemma 10.3, with the conditional expectation $\mathbb{E}[.\mathbf{1}_{\{R_{i,n}(0)\le M_n\}}|\mathcal{F}_0^{i,n}\vee\mathcal{F}_{h_nT}^{\theta^{i,n,*}},\theta_0^{i,n,*}=\theta_0]$. The presence of $\mathbf{1}_{\{R_{i,n}(0)\le M_n\}}$ along with the exponential decay in (10.23) show clearly that the result still holds, uniformly in the quadruplet (θ_0,i,n,t). By an immediate application of Jensen's inequality, this is still true replacing $\mathcal{F}_0^{i,n}\vee\mathcal{F}_{h_nT}^{\theta^{i,n,*}}$ by the smaller filtration $\mathcal{F}_0^{i,n}$, that is, for the operator $\mathbb{E}_{\theta_0,i,n}$. □

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 $f\le g+f*h$ where $(f*h)(t)=\int_0^t f(t-s)h(s)ds$ is the usual convolution. Then we have the majoration for any $t\ge 0$*

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

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

$$f\le g+g*\sum_{k=1}^{n}h^{*(k)}+f*h^{*(n+1)}. \tag{10.29}$$

We fix $t\ge 0$, and note that by a straightforward computation, for any integer $k\ge 1$ we have $h^{*(k)}(t)=\frac{t^{k-1}}{(k-1)!}a^ke^{-bt}$. We deduce that

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

as $n\to+\infty$. We also have for any integer $n\ge 1$

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

and thus we get the result by taking the limit $n\to+\infty$ in (10.29) evaluated at any point $t\ge 0$. □

In what follows, we quantify the local error between the doubly stochastic model and its constant parameter approximation. We recall the value of the key exponent $\kappa=\gamma(\delta-1)$ that has been introduced in (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 γ represents the regularity exponent in time of θ while δ controls the size of small blocks compared to n by the relation $h_n=n^{1/\delta}$. Note that by (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 $p\ge 1$, any deterministic kernel χ such that $\int_0^t\chi(s,t)ds$ is bounded uniformly in $t\in\mathbb{R}_+$, and any predictable process $(\psi_s)_{s\in\mathbb{R}_+}$ whose moments are bounded :*

(i) $\sup_{(\theta_0,i,t)\in\mathbf{E}_n^\alpha} \mathbb{E}_{\theta_0,i,n}\left|\lambda^{i,n,c}(t)-\lambda_*^{i,n}(t)\right|^p = O_{\mathbb{P}}\left(h_n^{-\kappa}\right)$

(ii) $\sup_{\theta_0\in K,1\le i\le B_n} \mathbb{E}_{\theta_0,i,n}\left|\int_{h_n^\alpha T}^{h_nT}\psi_s\{dN_s^{i,n,c}-dN_s^{i,n}\}\right|^p = O_{\mathbb{P}}\left(h_n^{p-\epsilon\kappa}\right)$

(iii) $\sup_{\theta_0\in K,1\le i\le B_n} \mathbb{E}_{\theta_0,i,n}\left|\int_{h_n^\alpha T}^{h_nT}\chi(s,h_nT)\{dN_s^{i,n,c}-dN_s^{i,n}\}\right|^p = O_{\mathbb{P}}\left(h_n^{-\kappa}\right)$

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 θ^* has exactly a deviation of that order. For $p>1$, the situation is fairly different. One would expect a deviation of the same order of that of the parameter process, that is of order $h_n^{-\kappa p} = T^{-\gamma p}\Delta_n^{\gamma p}$. But as it is shown in the previous lemma, deviations between the two models are quite weaker since the deviation remains of order $h_n^{-\kappa} = T^{-\gamma}\Delta_n^\gamma$ for any p. This loss is due to the point process structure and the shape of its related Burkholder-Davis-Gundy type inequality (see Lemma 10.2). This is the same phenomenon as in the following fact. For a Poisson process N of intensity λ, we have $\mathbb{E}[|N_t-\lambda t|^p]\sim\alpha_p t$ when $t\to 0$, i.e. a rate of convergence which is linear regardless of the moment chosen.*

Proof. We will show by recurrence on $q\in\mathbb{N}$ that for every p of the form $p=2^q$, we have the majoration for $n\in\mathbb{N}$, $t\in[0,h_nT]$ and uniformly in (θ_0,i),

$$\mathbb{E}_{\theta_0,i,n}\left|\lambda^{i,n,c}(t)-\lambda_*^{i,n}(t)\right|^{2^q} \le L_{n,q}+M_{n,q}e^{-\underline{b}(1-r)t}, \tag{10.30}$$

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

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

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

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

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

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

$$|ae^{-bt}-a'e^{-b't}| \le \left(|a-a'|+|b-b'|\right)e^{-\underline{b}t} \tag{10.31}$$

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

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

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

$$\mathbb{E}_{\theta_0,i,n}C_{i,n}(t) \leq \mathbb{E}_{\theta_0,i,n}\int_0^{t-} a_0 e^{-b_0(t-s)} d\left|N^{i,n,c} - N^{i,n}\right|_s \tag{10.32}$$

where $d\left|N^{i,n,c} - N^{i,n}\right|_s$ is the integer measure which counts the jumps that don't belong to both $dN^{i,n,c}$ and $dN^{i,n}$, i.e. the points of $\overline{N}^{i,n}$ that lay between the curves $t \to \lambda_*^{i,n}(t)$ and $t \to \lambda^{i,n,c}(t)$. A short calculation shows that this counting process admits $|\lambda^{i,n,c}(s) - \lambda_*^{i,n}(s)|$ as stochastic intensity. We compute now:

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

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

$$f(t) \leq L_n + R_{i,n}(t) + f * h(t), \tag{10.33}$$

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

$$f(t) \leq L_n + R_{i,n}(t) + \int_0^t \{L_n + R_{i,n}(s)\}a_0 e^{(a_0-b_0)(t-s)}ds. \tag{10.34}$$

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

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

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 (10.30) 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 $\eta > 0$

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

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

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

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

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

and apply Lemma 10.2 to get

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

which is easily bounded as in (10.30) 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 $O_{\mathbb{P}}\left(h_n^{-\kappa p}\right)$ that one would expect. Finally the term

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

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

$$\mathbb{E}_{\theta_0,i,n}\left[\left(W*\overline{\Lambda}_t\right)^p\right]\leq c_q f*h(t), \tag{10.35}$$

where again $h: s\mapsto a_0e^{-b_0 s}$, and $c_q<1$ if η is taken small enough. We have thus shown that f satisfies a similar convolution inequality as for the case $q=1$ and we can apply Lemma 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 $W_\psi(s,z)=\psi_s|\mathbb{1}_{\{0\leq z\leq\lambda^{n,c}(s)\}}-\mathbb{1}_{\{0\leq z\leq\lambda_*^n(s)\}}|$ and $W_\chi(s,z)=\chi(s,t)|\mathbb{1}_{\{0\leq z\leq\lambda^{n,c}(s)\}}-\mathbb{1}_{\{0\leq z\leq\lambda_*^n(s)\}}|$ along with Hölder's inequality. □

We are now ready to show the uniform asymptotic normality of the MLE by proving that any quantity related to the estimation is asymptotically very close to its counterpart for the constant parameter model $(N^{i,n,c},\lambda^{i,n,c})$. To this end we introduce the fake candidate intensity family and the fake log-likelihood process, as

$$\lambda^{i,n,c}(t,\theta)=\nu+\int_0^{t-}ae^{-b(t-s)}dN_s^{i,n,c} \tag{10.36}$$

and

$$l_{i,n}^c(\theta)=\int_0^{h_nT}\log(\lambda^{i,n,c}(t,\theta))dN_t^{i,n,c}-\int_0^{h_nT}\lambda^{i,n,c}(t,\theta)dt, \tag{10.37}$$

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) $\sup_{(\theta_0,i,n,t)\in\mathbf{E}} \mathbb{E}_{\theta_0,i,n} \sup_{\theta\in K} \left|\partial_\theta^j \tilde{\lambda}^{i,n}(t,\theta)\right|^p \leq K_j$ $\mathbb{P}$-*a.s.*

(ii) $\sup_{(\theta_0,i,n,t)\in\mathbf{E}} \mathbb{E}_{\theta_0,i,n} \sup_{\theta\in K} \left|\partial_\theta^j \lambda^{i,n,c}(t,\theta)\right|^p \leq K_j$ $\mathbb{P}$-*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 θ by linear combinations of terms of the form $\overline{\nu}$ or $\int_0^{t-} (t-s)^j e^{-\underline{b}(t-s)} dN_s^{i,n}$, $j \in \mathbb{N}$. The boundedness of moments of those terms uniformly in $n \in \mathbb{N}$ and in the time interval $[0, h_n T]$ 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 $\int_0^{t-} (t-s)^j e^{-\underline{b}(t-s)} d|N^{i,n} - N^{i,n,c}|_s$. The $\mathbb{L}^p$ estimate of such expression is then easily derived by a truncation argument and Lemma 10.7 **(iii)**. □

We now follow similar notations to the ones introduced in [3], and consider the main quantities of interest to derive the properties of the MLE. We define for any $(\theta, \theta_0) \in K^2$,

$$\mathbb{Y}_{i,n}(\theta,\theta_0) = \frac{1}{h_n T}(l_{i,n}(\theta) - l_{i,n}(\theta_0)), \tag{10.38}$$

$$\Delta_{i,n}(\theta_0) = \frac{1}{\sqrt{h_n T}} \partial_\theta l_{i,n}(\theta_0), \tag{10.39}$$

and finally

$$\Gamma_{i,n}(\theta_0) = -\frac{1}{h_n T} \partial_\theta^2 l_{i,n}(\theta_0). \tag{10.40}$$

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 $\mathbf{I} = \{(\theta_0, i, n) \in K \times \mathbb{N}^2 | 1 \leq i \leq B_n\}$.

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*

$$\sup_{\theta_0\in K, 1\leq i\leq B_n} \mathbb{E}_{\theta_0,i,n} \left|\Delta_{i,n}(\theta_0) - \Delta_{i,n}^c(\theta_0)\right|^L \to^{\mathbb{P}} 0, \tag{10.41}$$

$$\sup_{\theta_0\in K, 1\leq i\leq B_n} \mathbb{E}_{\theta_0,i,n} \left[\sup_{\theta\in K} |\mathbb{Y}_{i,n}(\theta,\theta_0) - \mathbb{Y}_{i,n}^c(\theta,\theta_0)|^p\right] = O_{\mathbb{P}}\left(h_n^{-\epsilon\kappa}\right), \tag{10.42}$$

$$\sup_{\theta_0\in K, 1\leq i\leq B_n} \mathbb{E}_{\theta_0,i,n} \left|\Gamma_{i,n}(\theta_0) - \Gamma_{i,n}^c(\theta_0)\right|^p = O_{\mathbb{P}}\left(h_n^{-\epsilon\kappa}\right), \tag{10.43}$$

$$\sup_{(\theta_0,i,n)\in\mathbf{I}} \mathbb{E}_{\theta_0,i,n} \left|h_n^{-1} \sup_{\theta\in K} |\partial_\theta^3 l_{i,n}(\theta)|\right|^p < K \ \mathbb{P}\text{-}a.s. \tag{10.44}$$

Proof. Let us show (10.41). We can express the equation in (10.39) and its counterpart for the constant model as

$$\Delta_{i,n}(\theta_0) = \frac{1}{\sqrt{h_n T}} \left\{ \int_0^{h_n T} \frac{\partial_\theta \tilde{\lambda}^{i,n}(s,\theta_0)}{\tilde{\lambda}^{i,n}(s,\theta_0)} dN_s^{i,n} - \int_0^{h_n T} \partial_\theta \tilde{\lambda}^{i,n}(s,\theta_0) ds \right\} \tag{10.45}$$

and

$$\Delta^c_{i,n}(\theta_0) = \frac{1}{\sqrt{h_n T}} \left\{ \int_0^{h_n T} \frac{\partial_\theta \lambda^{i,n,c}(s,\theta_0)}{\lambda^{i,n,c}(s,\theta_0)} dN_s^{i,n,c} - \int_0^{h_n T} \partial_\theta \lambda^{i,n,c}(s,\theta_0) ds \right\}. \tag{10.46}$$

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 $\sqrt{h_n T}(\Delta_{i,n}(\theta_0) - \Delta^c_{i,n}(\theta_0))$ is equivalent to the sum of the three terms

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

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

$$\sup_{\theta_0 \in K, 1 \le i \le B_n} \mathbb{E}_{\theta_0,i,n} \left| \Delta_{i,n}(\theta_0) - \Delta^c_{i,n}(\theta_0) \right|^L = O_{\mathbb{P}} \left(h_n^{\frac{L}{2} - \epsilon\kappa} \right), \tag{10.47}$$

for any $\epsilon \in (0,1)$. This tends to 0 if we can find an ϵ such that $\frac{L}{2} - \epsilon\kappa < 0$, and this can be done by taking ϵ sufficiently close to 1 since $L < 2\kappa$. Equations (10.42), (10.43) and (10.44) are proved similarly. ☐

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

$$\sup_{(\theta_0,i,n) \in \mathbf{I}} \mathbb{E}_{\theta_0,i,n} |\Delta^c_n(\theta_0)|^p < M \ \mathbb{P}\text{-a.s.} \tag{10.48}$$

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

$$\sup_{\theta_0 \in K, 1 \le i \le B_n} \mathbb{E}_{\theta_0,i,n} \left[\sup_{\theta \in K} |\mathbb{Y}^c_{i,n}(\theta,\theta_0) - \mathbb{Y}(\theta,\theta_0)| \right] = O\left(h_n^{-\epsilon \frac{p}{2}} \right) \ \mathbb{P}\text{-a.s.} \tag{10.49}$$

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

$$\sup_{\theta_0 \in K, 1 \le i \le \Delta_n^{-1}} \mathbb{E}_{\theta_0,i,n} \left| \Gamma^c_{i,n}(\theta_0) - \Gamma(\theta_0) \right|^p = O\left(h_n^{-\epsilon \frac{p}{2}} \right) \ \mathbb{P}\text{-a.s.} \tag{10.50}$$

where $\Gamma(\theta_0)$ is the asymptotic Fisher information matrix of the parametric Hawkes process regression model with parameter θ_0 as introduced in (10.9).

Proof. Note that when $\theta_0^{i,n,*} = \theta_0$, the constant model $N^{i,n,c}$ is simply a parametric Hawkes process with parameter θ_0, and is independent of the filtration $\mathcal{F}_0^{i,n}$. Thus, by a regular distribution argument the operator $\mathbb{E}_{\theta_0,i,n}$ acts as the simple operator $\mathbb{E}$ for $N^{i,n,c}$ distributed as a Hawkes with true value θ_0. It is straightforward to see that under a mild change in the proofs of Lemma 3.15 and Theorem 4.6 in [3] 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*

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \left\{ \mathbb{E}_{\theta_0,i,n}\left[f\left(\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta_0) \right) \right] - \mathbb{E}\left[f\left(T^{-\frac{1}{2}} \Gamma(\theta_0)^{-\frac{1}{2}} \xi \right) \right] \right\} \to^{\mathbb{P}} 0, \tag{10.51}$$

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

Proof. By (10.42) and (10.49), we can define some number $\epsilon \in (0,1)$ such that

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} h_n^{\epsilon(\frac{p}{2} \wedge \kappa)} \mathbb{E}_{\theta_0,i,n}\left[\sup_{\theta \in K} |\mathbb{Y}_{i,n}(\theta, \theta_0) - \mathbb{Y}(\theta, \theta_0)|^p \right] \to^{\mathbb{P}} 0, \tag{10.52}$$

and as $\widehat{\Theta}_{i,n}$ is also a maximizer of $\theta \to \mathbb{Y}_{i,n}(\theta, \theta_0)$, (10.52) 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.

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{P}_{\theta_0,i,n}\left[\widehat{\Theta}_{i,n} - \theta_0 \right] \to^{\mathbb{P}} 0, \tag{10.53}$$

since $\mathbb{Y}$ satisfies the non-degeneracy condition [A4] in [3]. From (10.43) and (10.50) we deduce

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} h_n^{\epsilon(\frac{p}{2} \wedge \kappa)} \mathbb{E}_{\theta_0,i,n} |\Gamma_{i,n}(\theta_0) - \Gamma(\theta_0)|^p \to^{\mathbb{P}} 0. \tag{10.54}$$

By (10.41), $\Delta_{i,n}(\theta_0)$ 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 ξ follows a standard normal distribution. Following the proof of Theorem 3.11 in [3], 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.

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \left\{ \mathbb{E}_{\theta_0,i,n}\left[f\left(\sqrt{h_n}(\widehat{\Theta}_{i,n} - \theta_0) \right) \right] - \mathbb{E}\left[f\left(T^{-\frac{1}{2}} \Gamma(\theta_0)^{-\frac{1}{2}} \xi \right) \right] \right\} \to^{\mathbb{P}} 0, \tag{10.55}$$

for any bounded continuous function f .

Finally, we extend (10.55) to the case of a function of polynomial growth of order smaller than L. First note that by (10.41) and (10.48) we have for any $L' \in (L, 2\kappa)$

$$\sup_{\theta_0 \in K, 1 \leq i \leq B_n} \mathbb{E}_{\theta_0,i,n} |\Delta_{i,n}(\theta_0)|^{L'} = O_{\mathbb{P}}(1). \tag{10.56}$$

We now adopt the notations of [8] 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 (10.52), (10.54), (10.56) and finally (10.44), conditions $[A1'']$, $[A4']$,

$[A6]$, $[B1]$ and $[B2]$ in [8] 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 [8] to get that for any $p \leq L$,

$$\sup_{\theta_0 \in K, 1 \leq i \leq \Delta_n^{-1}} \mathbb{E}_{\theta_0,i,n} \left| \sqrt{h_n} \left(\widehat{\Theta}_{i,n} - \theta_0 \right) \right|^p = O_{\mathbb{P}}(1). \tag{10.57}$$

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

□

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 [1]) to obtain (10.51) when not conditioning by any particular starting value of θ_t^*. We provide the formal proof in what follows. Recall that $\mathbb{E}_{(i-1)\Delta_n}$ stands for $\mathbb{E}[.|\mathcal{F}_0^{i,n}]$.

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

$$\mathbb{E}_{(i-1)\Delta_n} \left[f\left(\sqrt{h_n} \left(\widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n} \right) \right) \mathbf{1}_{\{R_{i,n}(0) \leq M_n\}} \right] \tag{10.58}$$

$$+ \mathbb{E}_{(i-1)\Delta_n} \left[f\left(\sqrt{h_n} \left(\widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n} \right) \right) \mathbf{1}_{\{R_{i,n}(0) > M_n\}} \right]. \tag{10.59}$$

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

$$\mathbb{E}_{(i-1)\Delta_n} \left[f\left(\sqrt{h_n} \left(\widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n} \right) \right) \mathbf{1}_{\{R_{i,n}(0) \leq M_n\}} - f\left(T^{-\frac{1}{2}} \Gamma\left(\theta^*_{(i-1)\Delta_n} \right)^{-\frac{1}{2}} \xi \right) \right] \tag{10.60}$$

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

$$\left| G\left(\theta^*_{(i-1)\Delta_n}\right) \right| \leq \sup_{\theta_0 \in K} \left| \mathbb{E}_{\theta_0,i,n} \left[f\left(\sqrt{h_n} \left(\widehat{\Theta}_{i,n} - \theta_0 \right) \right) \right] - \mathbb{E}\left[f\left(T^{-\frac{1}{2}} \Gamma(\theta_0)^{-\frac{1}{2}} \xi \right) \right] \right|, \tag{10.61}$$

take the sup over i in (10.61), and in view of Theorem 10.12, we have shown that (10.60) is uniformly of order $o_{\mathbb{P}}(1)$.

On the other hand, (10.59) 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 $\mathbb{P}\left[R_{i,n}(0) > M_n\right] \leq \mathbb{P}\left[\lambda_*^n((i-1)\Delta_n) > M_n\right]$, which in turn can be dominated easily with Markov's inequality by $M_n^{-1}\mathbb{E}\left[\lambda_*^n((i-1)\Delta_n)\right] = O(nM_n^{-1})$. We recall that M_n is of the form n^q where q can be taken arbitrarily big, and we have thus shown that (10.59) vanishes asymptotically. □

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,λ_*^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 $\overline{Z}$ 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 $\xi_n \in [\widehat{\Theta}_n, \theta_0]$ such that

$$0 = \partial_\theta l_n(\widehat{\Theta}_n) = \partial_\theta l_n(\theta_0) + \partial_\theta^2 l_n(\theta_0)(\widehat{\Theta}_n - \theta_0) + \frac{1}{2}\partial_\theta^3 l_n(\xi_n)(\widehat{\Theta}_n - \theta_0)^{\otimes 2}, \tag{10.62}$$

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

$$\partial_\theta l_n^c(\theta_0) + \partial_\theta^2 l_n^c(\theta_0)(\widehat{\Theta}_n - \theta_0) + \frac{1}{2}\partial_\theta^3 l_n^c(\xi_n)(\widehat{\Theta}_n - \theta_0)^{\otimes 2} = O_{\mathbb{P}}\left(h_n^{1-\epsilon\kappa}\right), \tag{10.63}$$

where the residual term $O_{\mathbb{P}}\left(h_n^{1-\epsilon\kappa}\right)$ 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_nT and obtain

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

where the expectation of the first term has vanished because of the martingale form of $\Delta_n^c(\theta_0)$ in (10.46). The term $\mathbb{E}_{\theta_0,n}[\Gamma_n^c(\theta_0)]\mathbb{E}_{\theta_0,n}[\widehat{\Theta}_n-\theta_0]$ is of interest since it contains the quantity we want to evaluate. The first and the third terms have thus to be evaluated to derive an expansion of the bias. We start by the first term, i.e. the covariance between our estimator and $\Gamma_n^c(\theta_0)$. To compute the limiting value of such covariance, we consider the martingale $M_n^c(t,\theta_0) = \int_0^t \frac{\partial_\theta \lambda^{n,c}(s,\theta_0)}{\lambda^{n,c}(s,\theta_0)}\{dN_s^{n,c} - \lambda^{n,c}(s,\theta_0)ds\}$, and we define the empirical covariance processes $C_n^c(\theta_0)$ and $Q_n^c(\theta_0)$ whose components are, for any triplet $(i,j,k) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3$,

$$C_n^c(\theta_0)_{i,jk} = \frac{1}{h_nT}\int_0^{h_nT} \partial_{\theta,i}\lambda^{n,c}(s,\theta_0)\partial_{\theta,jk}^2 \log\lambda^{n,c}(s,\theta_0)ds,$$

and

$$Q_n^c(\theta_0)_{i,jk} = -\frac{M_n^c(T,\theta_0)_i}{h_nT}\int_0^{h_nT} \frac{\partial_\theta\lambda^{n,c}(s,\theta_0)_j\partial_\theta\lambda^{n,c}(s,\theta_0)_k}{\lambda^{n,c}(s,\theta_0)}ds,$$

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

Lemma 10.13. *We have*

$$\mathbb{E}_{\theta_0,n}\left[C_n^c(\theta_0)_{i,jk} + Q_n^c(\theta_0)_{i,jk}\right] = -\sqrt{h_nT}\,\mathbb{E}_{\theta_0,n}\left[\Delta_n^c(\theta_0)_i\Gamma_n^c(\theta_0)_{jk}\right]. \tag{10.64}$$

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

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

Formula (10.64) is then obtained directly from the expression of $\Gamma_n^c(\theta_0)$ and $\Delta_n^c(\theta_0)$. □

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

$$\sup_{\theta_0 \in K} h_n^{\epsilon\frac{p}{2}} \mathbb{E}_{\theta_0,n} \left|C_n^c(\theta_0) - C(\theta_0)\right|^p \to^{\mathbb{P}} 0, \tag{10.65}$$

and

$$\sup_{\theta_0 \in K} h_n^{\epsilon\frac{p}{2}} \left|\mathbb{E}_{\theta_0,n}\left[Q_n^c(\theta_0) - Q(\theta_0)\right]\right|^p \to^{\mathbb{P}} 0 \tag{10.66}$$

where C and Q were defined respectively in (10.11) and (10.12). Before we turn to the limiting expression of the term

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

in our expansion of the bias in terms of $C(\theta_0) + Q(\theta_0)$, we need to control the convergence of $\Gamma_n^c(\theta_0)^{-1}$ toward $\Gamma(\theta_0)^{-1}$. We define $c_0 = \min_{\theta_0 \in K} \min\{c \in \mathbb{R}_+ | \forall x \in \mathbb{R}^3 - \{0\}, x^T\Gamma(\theta_0)x \geq c|x|_2^2 > 0\}$, the smallest eigenvalue of all the matrices $\Gamma(\theta_0)$. We consider the sequence of events $\mathbb{B}_n(\theta_0) = \{\forall x \in \mathbb{R}^3 - \{0\}, x^T\Gamma_n^c(\theta_0)x \geq \frac{c_0}{2}|x|_2^2\}$, and their complements $\mathbb{B}_n(\theta_0)^c$.

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

(i) $\sup_{\theta_0 \in K} \mathbb{P}_{\theta_0,n}\left[\mathbb{B}_n(\theta_0)^c\right] = O_{\mathbb{P}}\left(h_n^{-\epsilon\frac{p}{2}}\right)$.

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

Proof. We start by showing **(i)**. We recall that in our notation convention, the symbol $|x|$ stands for $\sum_i |x_i|$ for any vector or matrix. Clearly, we have that

$$\mathbb{P}_{\theta_0,n}\left[\mathbb{B}_n(\theta_0)^c\right] \leq \mathbb{P}_{\theta_0,n}\left\{\forall x \in \mathbb{R}^3 - \{0\}, \frac{|x^T(\Gamma_n^c(\theta_0) - \Gamma(\theta_0))x|}{|x|_2^2} > \frac{c_0}{2}\right\}, \tag{10.67}$$

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 $\mathbb{R}^3$, (10.67) implies the existence of some constant $\eta > 0$ such that

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

where Markov's inequality was used at the last step. **(i)** thus follows from (10.54). 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 $\mathbb{B}_n(\theta_0)$. □

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

$$\mathbb{E}_{\theta_0,n}[\overline{\Gamma}_n^c(\theta_0)(\widehat{\Theta}_n - \theta_0)]_i = -\frac{\Gamma(\theta_0)^{jk}\left\{C(\theta_0)_{k,ij} + Q(\theta_0)_{k,ij}\right\}}{h_n T} + O_{\mathbb{P}}\left(h_n^{-\epsilon\left(\kappa \wedge \frac{3}{2}\right)}\right). \tag{10.68}$$

Proof. Note first that in view of Lemma 10.14 **(i)** along with Hölder's inequality, we have that $\mathbb{E}_{\theta_0,n}[\overline{\Gamma}_n^c(\theta_0)(\widehat{\Theta}_n - \theta_0)] = \mathbb{E}_{\theta_0,n}[\overline{\Gamma}_n^c(\theta_0)(\widehat{\Theta}_n - \theta_0)\mathbf{1}_{\mathbb{B}_n(\theta_0)}] + o_{\mathbb{P}}\left(h_n^{-\frac{3}{2}}\right)$. 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 (10.63), we have the representation,

$$\widehat{\Theta}_n - \theta_0 = \frac{1}{\sqrt{h_n T}}\Gamma_n^c(\theta_0)^{-1}\Delta_n^c(\theta_0) + \Gamma_n^c(\theta_0)^{-1}\frac{\partial_\theta^3 l_n^c(\xi_n)(\widehat{\Theta}_n - \theta_0)^{\otimes 2}}{2h_n T} + O_{\mathbb{P}}\left(h_n^{-\tilde{\epsilon}\kappa}\right), \tag{10.69}$$

on the set $\mathbb{B}_n(\theta_0)$, where the residual term $O_{\mathbb{P}}\left(h_n^{-\tilde{\epsilon}\kappa}\right)$ admits moments of any order with respect to the operator $\mathbb{E}_{\theta_0,n}$. We inject (10.69) in the expectation and get

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

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

$$\frac{1}{\sqrt{h_n T}}\mathbb{E}_{\theta_0,n}[\overline{\Gamma}_n^c(\theta_0)\Gamma(\theta_0)^{-1}\Delta_n^c(\theta_0)] + O_{\mathbb{P}}\left(h_n^{-\frac{3\epsilon}{2}}\right), \tag{10.70}$$

where we used Hölder's inequality to control $\frac{1}{\sqrt{h_n T}}\mathbb{E}_{\theta_0,n}\left[\overline{\Gamma}_n^c(\theta_0)(\Gamma_n^c(\theta_0)^{-1} - \Gamma(\theta_0)^{-1})\Delta_n^c(\theta_0)\right]$ 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 (10.70), use Lemma 10.13 along with (10.65), and this leads to the estimate

$$\frac{1}{\sqrt{h_n T}}\mathbb{E}_{\theta_0,n}[\overline{\Gamma}_n^c(\theta_0)\Gamma(\theta_0)^{-1}\Delta_n^c(\theta_0)]_i = \frac{\Gamma(\theta_0)^{jk}\left\{C(\theta_0)_{k,ij} + Q(\theta_0)_{k,ij}\right\}}{h_n T} + O_{\mathbb{P}}\left(h_n^{-\frac{3\epsilon}{2}}\right). \tag{10.71}$$

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

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

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

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

$$\sup_{\theta_0\in K} h_n^{\epsilon\frac{p}{2}}\mathbb{E}_{\theta_0,n}\left|\frac{1}{h_nT}\partial^3_\theta l^c_n(\theta_0)-K(\theta_0)\right|^p \to^{\mathbb{P}} 0, \tag{10.72}$$

where $K(\theta_0)$ was introduced in (10.10). 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*

$$\frac{1}{2h_nT}\mathbb{E}_{\theta_0,n}[\partial^3_\theta l^c_n(\xi_n)(\widehat{\Theta}_n-\theta_0)^{\otimes 2}]_i = \frac{\Gamma(\theta_0)^{jk}K(\theta_0)_{ijk}}{2h_nT}+O_{\mathbb{P}}\left(h_n^{-\epsilon\left(\kappa\wedge\frac{3}{2}\right)}\right). \tag{10.73}$$

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

$$\begin{aligned}\frac{1}{2h_nT}\mathbb{E}_{\theta_0,n}[\partial^3_{\theta,ijk}l^c_n(\xi_n)(\widehat{\Theta}_n-\theta_0)_j(\widehat{\Theta}_n-\theta_0)_k] &= \frac{1}{2h_nT}\mathbb{E}_{\theta_0,n}[\partial^3_{\theta,ijk}l^c_n(\xi_n)]\mathbb{E}_{\theta_0,n}[(\widehat{\Theta}_n-\theta_0)_j(\widehat{\Theta}_n-\theta_0)_k]\\ &+\frac{1}{2h_nT}\mathbb{E}_{\theta_0,n}[\overline{\partial^3_{\theta,ijk}l^c_n(\xi_n)}(\widehat{\Theta}_n-\theta_0)_j(\widehat{\Theta}_n-\theta_0)_k].\end{aligned}$$

We now remark that the first term admits the expansion

$$\frac{\Gamma(\theta_0)^{jk}K(\theta_0)_{ijk}}{2h_nT}+O_{\mathbb{P}}\left(h_n^{-\epsilon\left(\kappa\wedge\frac{3}{2}\right)}\right), \tag{10.74}$$

by replacing $\mathbb{E}_{\theta_0,n}[(\widehat{\Theta}_n-\theta_0)_j(\widehat{\Theta}_n-\theta_0)_k]$ and $\frac{1}{2h_nT}\mathbb{E}_{\theta_0,n}[\partial^3_{\theta,ijk}l^c_n(\xi_n)]$ by their estimates

$$\mathbb{E}_{\theta_0,n}[(\widehat{\Theta}_n-\theta_0)_j(\widehat{\Theta}_n-\theta_0)_k] = \frac{\Gamma(\theta_0)^{jk}}{h_nT}+O_{\mathbb{P}}\left(h_n^{-\epsilon\left(\kappa\wedge\frac{3}{2}\right)}\right), \tag{10.75}$$

and

$$\frac{1}{2h_nT}\mathbb{E}_{\theta_0,n}\left[\partial^3_{\theta,ijk}l^c_n(\xi_n)\right] = K(\theta_0)_{ijk}+O_{\mathbb{P}}\left(h_n^{-\frac{\epsilon}{2}}\right). \tag{10.76}$$

(10.75) is obtained by injecting the expansion of $\widehat{\Theta}_n-\theta_0$ in (10.69) up to the first order only, and (10.76) is a consequence of (10.72) and the uniform boundedness of moments of $\frac{\partial^4_\theta l^c_n(\theta_0)}{h_nT}$ in $\theta_0\in K$ by Lemma 10.9 **(ii)**. Note that the expansion (10.75) is not a direct consequence of Theorem 10.12 applied to $x\to x_jx_k$ since this would lead to the weaker estimate $\frac{\Gamma(\theta_0)^{jk}}{h_nT}+o_{\mathbb{P}}(h_n^{-1})$ instead. Finally, the second term is of order $O_{\mathbb{P}}\left(h_n^{-\frac{3\epsilon}{2}}\right)$ by Hölder's inequality along with Theorem 10.12, and thus we are done. □

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

$$b(\theta_0)_j = \frac{1}{2}\Gamma(\theta_0)^{ij}\Gamma(\theta_0)^{kl}(K(\theta_0)_{ikl}+2\left\{C(\theta_0)_{k,il}+Q(\theta_0)_{k,il}\right\}), \tag{10.77}$$

which was defined in (10.14). 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*

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

uniformly in $i \in \{1,\cdots,B_n\}$ and in $\theta_0 \in K$. Moreover, the bias-corrected estimator $\widehat{\Theta}_{i,n}^{(BC)}$ defined in (5.18) has the (uniform) bias expansion

$$\mathbb{E}_{\theta_0,i,n}\left[\widehat{\Theta}_{i,n}^{BC} - \theta_0\right] = O_{\mathbb{P}}\left(h_n^{-\epsilon\left(\kappa\wedge\frac{3}{2}\right)}\right). \tag{10.79}$$

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

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

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

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

which is exactly (10.78). Finally, a calculation similar to the proofs of Lemmas 10.15 and 10.16 shows that

$$\mathbb{E}_{\theta_0,n} b\left(\widehat{\Theta}_n\right) = \mathbb{E}_{\theta_0,n} b\left(\theta_0\right) + O_{\mathbb{P}}\left(h_n^{-\frac{\epsilon}{2}}\right) \tag{10.80}$$

so that we have (10.79) and this concludes the proof. □

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

Proof of Theorem 5.3. This follows exactly the same argument as for the proof of Theorem 5.2. □

10.5 Proof of the GCLT

In this section we present the proof of Theorem 5.4 using a similar martingale approach as in [7]. 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 $i \in \{1,\cdots,B_n\}$ that there exists $\epsilon > 0$ such that*

$$\mathrm{Var}_{(i-1)\Delta_n}\left[\sqrt{h_n}\left(\widehat{\Theta}_{i,n}^{(BC)} - \theta^*_{(i-1)\Delta_n}\right)\right] = T^{-1}\Gamma\left(\theta^*_{(i-1)\Delta_n}\right)^{-1} + o_{\mathbb{P}}(1), \tag{10.81}$$

$$\mathbb{E}_{(i-1)\Delta_n}\left[\left|\sqrt{h_n}\left(\widehat{\Theta}_{i,n}^{(BC)} - \theta^*_{(i-1)\Delta_n}\right)\right|^{2+\epsilon}\right] = O_{\mathbb{P}}(1), \tag{10.82}$$

$$\mathbb{E}_{(i-1)\Delta_n}\left[\widehat{\Theta}_{i,n}^{(BC)} - \theta^*_{(i-1)\Delta_n}\right] = o_{\mathbb{P}}\left(n^{-1/2}\right), \tag{10.83}$$

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

The above-mentioned approach is based on techniques introduced in [7], but it is much different and deeper. Indeed, [7] 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

$$\Theta = \frac{1}{B_n}\sum_{i=1}^{B_n} \theta^*_{(i-1)\Delta_n} + o_{\mathbb{P}}\left(n^{-1/2}\right). \tag{10.84}$$

Note that (10.84) is to be compared to (3.1) for the toy model. Moreover, (10.84) was also shown in (35) on pp. 46-47 in [7], but the parameter process was restricted to follow a continuous Itô-process. To show (10.84), it is sufficient to show that

$$\frac{\sqrt{n}}{B_n}\sum_{i=1}^{B_n}\left|\theta^*_{(i-1)\Delta_n} - \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n}\theta^*_s ds\right| = o_{\mathbb{P}}(1). \tag{10.85}$$

We can bound (10.85) by

$$\frac{\sqrt{n}}{B_n}\sum_{i=1}^{B_n}\Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n}\underbrace{\left|\theta^*_{(i-1)\Delta_n} - \theta^*_s\right|}_{O_{\mathbb{P}}(\Delta_n^{\gamma})} ds = o_{\mathbb{P}}(1), \tag{10.86}$$

where we used [C]-**(i)** to obtain the order in (10.86). Thus, we deduce that the left-hand side in (10.86) 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 (10.84).

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 $S_n^{(B)} \to^{\mathbb{P}} 0$ and we will prove the existence of some V_T such that $\mathcal{F}_T^{\theta^*}$-stably in law, $S_n^{(M)} \to V_T^{\frac{1}{2}}\mathcal{N}(0,1)$. Note that the former is a straightforward consequence of (10.83). To show the latter $S_n^{(M)} \to V_T^{\frac{1}{2}}\mathcal{N}(0,1)$, we will use Theorem 3.2 of p. 244 in [4]. First, we show the conditional Lindeberg condition (3.13), i.e. in our case that for any $\eta > 0$ we have

$$\frac{n}{B_n^2}\sum_{i=1}^{B_n}\mathbb{E}_{(i-1)\Delta_n}\left[M_{i,n}^2\mathbf{1}_{\left\{\frac{\sqrt{n}}{B_n}M_{i,n}>\eta\right\}}\right] \to^{\mathbb{P}} 0. \tag{10.87}$$

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 (10.82) from [C*]. On the other hand, using also (10.82) along with [C]-**(ii)**, we have that $b_{i,n}$ goes uniformly to 0. We have thus proved

(10.87). We now prove the conditional variance condition (3.11), i.e. that

$$\frac{n}{B_n^2}\sum_{i=1}^{B_n}\mathbb{E}_{(i-1)\Delta_n}\left[M_{i,n}^2\right]\to^{\mathbb{P}} V_T := T^{-2}\int_0^T \Gamma(\theta_s^*)^{-1}ds. \tag{10.88}$$

We have that

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

We use Proposition I.4.44 on p.51 in [6] along with (10.81) from [C*] to show (10.88). Now, conditions (3.10) and (3.12) are automatically satisfied because $M_{i,n}$ is a martingale increment and since we consider the reference continuous martingale $\boldsymbol{M} = 0$. Finally we show condition (3.14) to get the stable convergence. We thus consider a bounded $\mathbf{F}^{\theta^*}$-martingale Z, and we show that

$$\frac{\sqrt{n}}{B_n}\sum_{i=1}^{B_n}\mathbb{E}_{(i-1)\Delta_n}\left[M_{i,n}\Delta Z_{i,n}\right]\to^{\mathbb{P}} 0, \tag{10.89}$$

where $\Delta Z_{i,n} := Z_{i\Delta_n} - Z_{(i-1)\Delta_n}$. Using the Taylor expansion (10.63) and the boundedness of Z, by a similar calculation as in Lemma 10.15, we have

$$\frac{\sqrt{n}}{B_n}\sum_{i=1}^{B_n}\mathbb{E}_{(i-1)\Delta_n}\left[M_{i,n}\Delta Z_{i,n}\right] = \frac{h_n}{\sqrt{n}}\sum_{i=1}^{B_n}\Gamma\left(\theta^*_{(i-1)\Delta_n}\right)^{-1}\mathbb{E}_{(i-1)\Delta_n}\left[\partial_\theta l_{i,n}^c\left(\theta^*_{(i-1)\Delta_n}\right)\Delta Z_{i,n}\right] + o_{\mathbb{P}}(1).$$

Note now that $l_{i,n}^c\left(\theta^*_{(i-1)\Delta_n}\right)$ can be written as an integral over the canonical Poisson martingale :

$$l_{i,n}^c\left(\theta^*_{(i-1)\Delta_n}\right) = \int_0^{h_nT}\int_{\mathbb{R}_+}\frac{\partial_\theta\lambda^{i,n,c}(s,\theta^*_{(i-1)\Delta_n})}{\lambda^{i,n,c}(s,\theta^*_{(i-1)\Delta_n})}\mathbb{1}_{\{0\leq z\leq\lambda^{i,n,c}(s,\theta^*_{(i-1)\Delta_n})\}}\left\{\overline{N}^{i,n}(ds,dz) - \overline{\Lambda}^{i,n}(ds,dz)\right\},$$

with $\overline{\Lambda}^{i,n}(ds,dz) = ds\otimes dz$. We deduce from the above representation that $\mathbb{E}_{(i-1)\Delta_n}\left[\partial_\theta l_{i,n}^c\left(\theta^*_{(i-1)\Delta_n}\right)\Delta Z_{i,n}\right] = 0$, since both σ-fields $\mathcal{F}_T^{\theta^*}$ and $\mathcal{F}_T^{\overline{N}}$ are independent, so that Z and $\overline{N}^{i,n} - \overline{\Lambda}^{i,n}$ are orthogonal. Thus (10.89) holds. Thus, by Theorem 3.2 of [4], we have the $\mathcal{F}_T^{\theta^*}$-stable convergence in law of $S_n^{(M)}$ toward an $\mathcal{F}_T^{\theta^*}$-conditional Gaussian limit with random variance V_T. In particular, we have that V_T and $\mathcal{N}(0,1)$ in Theorem 5.4 are independent from each other. □

We prove now that we can obtain (10.81), (10.82) and (10.83) in Condition $[C^*]$. First note that for any $L\in(0,2\kappa)$, a calculation gives

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

uniformly in $i\in\{1,...,B_n\}$. 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 (10.81) 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 (10.81) follows from Theorem 5.2. Moreover, (10.82) is a direct consequence of Theorem 5.2. Finally, in view of (5.21) in Theorem 5.3, (10.83) 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 κ by its expression, we get the two conditions $\frac{\delta}{2} < \gamma(\delta - 1)$ and $\frac{\delta}{2} < \frac{3}{2}$, that is $\frac{\gamma}{\gamma - \frac{1}{2}} < \delta < 3$. This is exactly condition [BC].

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. (10.81) and (10.82) are true since $\delta > 1 + \frac{1}{\gamma}$. Moreover, by assumption on δ and α, (10.83) is replaced by $\mathbb{E}_{(i-1)\Delta_n}\left[\widehat{\Theta}_{i,n}^{(BC)} - \theta^*_{(i-1)\Delta_n}\right] = O_{\mathbb{P}}\left(n^{-\gamma(1-\delta^{-1}) \wedge \delta^{-1}}\right) = o_{\mathbb{P}}\left(n^{-\alpha}\right)$. writing the decomposition

$$\frac{n^\alpha}{B_n} \sum_{i=1}^{B_n} \left(\widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n}\right) = n^{\alpha - \frac{1}{2}} \left\{S_n^{(B)} + S_n^{(M)}\right\}, \tag{10.90}$$

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

10.7 Proof of Proposition 6.1

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

$$\partial_\xi^2 l_{i,n}\left(n^{-1}\xi\right)_{|\xi = n\theta} = n^{-2} \partial_\theta^2 l_{i,n}(\theta), \tag{10.91}$$

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 $i \in \{1, ..., B_n\}$ the estimates

$$\Gamma_{i,n}\left(\widehat{\Theta}_{i,n}\right)^{-1} = \Gamma\left(\theta^*_{(i-1)\Delta_n}\right)^{-1} + o_{\mathbb{P}}(1) \tag{10.92}$$

and

$$\Gamma\left(\theta^*_{(i-1)\Delta_n}\right)^{-1} = \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} \Gamma\left(\theta_t^*\right)^{-1} dt + o_{\mathbb{P}}(1). \tag{10.93}$$

To show (10.92), 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

$$|a_{i,n}| \leq \sup_{\theta \in K} \frac{1}{h_n} \left| \partial_\theta \left(\partial_\theta^2 l_{i,n}(\theta) \right)^{-1} \right| \left| \widehat{\Theta}_{i,n} - \theta^*_{(i-1)\Delta_n} \right|. \tag{10.94}$$

By some algebraic calculus it is straightforward to show that the term $\sup_{\theta \in K} \frac{1}{h_n} \left| \partial_\theta \left(\partial_\theta^2 l_{i,n}(\theta) \right)^{-1} \right|$ 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 (10.92) holds. Finally the approximation (10.93) is a straightforward consequence of Lemma 10.9 **(i)** and Lemma 10.14 **(i)** along with assumption [C]-**(i)**.

References

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

[2] P. Brémaud and L. Massoulié. Stability of nonlinear hawkes processes. *The Annals of Probability*, pages 1563–1588, 1996.

[3] S. Clinet and N. Yoshida. Statistical inference for ergodic point processes and application to limit order book. *Stochastic Processes and their Applications*, 127(6):1800–1839, 2017.

[4] J. Jacod. On continuous conditional gaussian martingales and stable convergence in law. *Séminaire de probabilités de Strasbourg*, 31:232–246, 1997.

[5] J. Jacod and P. E. Protter. *Discretization of processes*. Springer Science & Business Media, 2011.

[6] J. Jacod and A. Shiryaev. *Limit theorems for stochastic processes (2nd ed.)*. Berlin: Springer-Verlag, 2003.

[7] Y. Potiron and P. A. Mykland. Local parametric estimation in high frequency data. *arXiv preprint arXiv:1603.05700*, 2016.

[8] 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.