PARAMETRIC INFERENCE FOR NONLINEAR HAWKES PROCESSES

BY YOANN POTIRON[1,a],

[1]*Faculty of Business and Commerce, Keio University,* [a]*potiron@fbc.keio.ac.jp*

We develop inference for Hawkes processes. We assume that the intensity has a parametric form and is nonlinear. We also assume that the kernel is parametric and has a general form. The inference procedure is based on maximum likelihood estimation. We show the central limit theorem of the inference procedure for the Hawkes processes. In particular, we allow for nonlinear intensity under some Lipschitz continuity assumptions on the intensity function. The main novelty in the proofs is to adapt the arguments from the linear intensity case to the nonlinear intensity case.

1. Introduction. This paper concerns parametric inference for point processes. The main stylized fact in this strand of literature, the presence of event clustering in time, motivates the so-called Hawkes mutually exciting processes (see Hawkes (1971a) and Hawkes (1971b)). We define the point process N_t of dimension d as the cumulative number of events from the starting time 0 to the final time t and λ_t its intensity. A standard definition of Hawkes mutually exciting processes with nonlinear intensity is given by

$$\lambda_t = f\Big(\int_0^t h(t-s)\,dN_s\Big). \tag{1}$$

Here, the function f of dimension d is nonlinear and nonrandom. In addition, the exciting kernel h is a matrix of dimension $d \times d$. The point processes are mutually exciting in the sense that the diagonal components $h^{(i,i)}$ are self-exciting terms for the ith process and nondiagonal components $h^{(i,j)}$ are cross-exciting terms for the ith process made by events from the jth process. The particular case $h = 0$ corresponds to a classical Poisson process, so that we can view Hawkes processes as a natural extension of Poisson processes.

The main application of Hawkes processes lies in seismology (see Rubin (1972), Ozaki (1979), Vere-Jones and Ozaki (1982) Ogata (1978), Ogata (1988) and Ikefuji et al. (2022)). There are also applications in quantitative finance (see Chavez-Demoulin, Davison and McNeil (2005), Embrechts, Liniger and Lin (2011), Bacry et al. (2013), Jaisson and Rosenbaum (2015), Jaisson and Rosenbaum (2016) and Clinet and Yoshida (2017)). Some applications are also in financial econometrics (see Chen and Hall (2013), Clinet and Potiron (2018), Kwan, Chen and Dunsmuir (2023), Potiron and Volkov (2025)). We can also find some applications in biology (see Reynaud-Bouret and Schbath (2010), Donnet, Rivoirard and Rousseau (2020) and Cai, Zhang and Guan (2024)). Finally, there are applications in social studies (see Fox et al. (2016) and Fang et al. (2024)).

There are many theoretical results for Hawkes processes in statistics. Hawkes and Oakes (1974) provide a Poisson cluster process representation for the Hawkes process. Brémaud and Massoulié (1996) study stability of nonlinear Hawkes processes. Zhu (2013) and Zhu (2015) consider central limit theorem and large deviations for Markovian nonlinear Hawkes processes. Roueff, von Sachs and Sansonnet (2016), Roueff and Von Sachs (2019), Cheysson and Lang (2022), Mammen and Müller (2023) and Erdemlioglu et al. (2025) study locally stationary Hawkes processes. Potiron et al. (2025a) and Potiron et al. (2025b) introduce Hawkes

MSC2020 subject classifications: Primary 60G55, 62M09; secondary 60F05.

Keywords and phrases: point processes, Hawkes mutually exciting processes, parametric inference, nonlinear intensity, general kernel, central limit theorem.

processes with Itô semimartingale baseline. Potiron (2025a) introduces a more general baseline. The microstructure of stochastic volatility models with self-excitation is investigated in Horst and Xu (2022). Horst and Xu (2021) and Horst and Xu (2024) give functional limit theorems for Hawkes processes. Xu (2024) studies diffusion approximations for self-excited systems. Karim, Laeven and Mandjes (2025) introduce compound multivariate Hawkes processes.

In this paper, we consider Hawkes processes in which the intensity is parametric and nonlinear. We also assume that the kernel is parametric and has a general form. The inference procedure is based on maximum likelihood estimation. Ogata (1978) shows the central limit theorem of the inference procedure for an ergodic stationary point process. However, the definition of ergodicity is vague in that paper. Most of the papers on parametric inference for Hawkes processes make this ergodicity assumption (see Bowsher (2007), Large (2007) and Cavaliere et al. (2023), Assumption 1(b) and Remark 2.1).

Clinet and Yoshida (2017) exhibit the conditions required, i.e. ergodicity of the Hawkes intensity process and its derivative jointly. They consider general point processes and derive the central limit theorem of the inference procedure in Theorem 3.11 (p. 1809) under these ergodicity assumptions. They also show these ergodicity assumptions in the case of a Hawkes process with exponential kernel in Theorem 4.6 (p. 1821). The proofs rely heavily on the Markov property of the exponential distribution.

Kwan (2023) and Kwan, Chen and Dunsmuir (2024) consider the non-exponential kernel case but the authors mention that such case is challenging since the Hawkes intensity process is non-Markovian. Thus, this renders standard Markov tools inapplicable. Consequently, the authors can only show the ergodicity for the Hawkes intensity process and for its derivative (see Theorem 4.3.2 (p. 91) in Kwan (2023) and Theorem 2.1 (p. 4) in Kwan, Chen and Dunsmuir (2024)), but not jointly. Thus, they can only show the consistency of the inference procedure in Theorem 3.4.3 (p. 73) from Kwan (2023) and Theorem 3.2 (p. 9) from Kwan, Chen and Dunsmuir (2024). When the kernel follows a generalized gamma distribution, Potiron and Volkov (2025) can show that the ergodicity assumptions are satisfied and also obtain the central limit theorem of the inference procedure. This is due to the exponentially decreasing nature of the kernel. Potiron (2025b) weakens the assumptions from the point process theory in Clinet and Yoshida (2017) by considering a different approach in the proofs. Under ergodicity of the point process intensity and its derivative, he shows the central limit theorem of the inference procedure (see Theorem 1). Second, he shows the ergodicity of the Hawkes intensity process and its derivative, in case of a general kernel. Moreover, he shows the central limit theorem of the inference procedure (see Theorem 2) for Hawkes processes with general kernel. With a general kernel, Costa et al. (2020) (Theorem 1.2, p. 884) and Graham (2021) (Theorem 5.4, p. 2856) shows the ergodicity of the Hawkes processes, but not its intensity. See also Section 3.2 (p. 893) in Reynaud-Bouret and Roy (2007).

All these results are useful, but the obtained central limit theorems for the inference procedure are restricted to a linear intensity. In finance, there is empirical evidence that the intensity is not linear (see Blanc, Donier and Bouchaud (2017)). Consequently, we consider Hawkes processes in which the intensity is nonlinear. We show the central limit theorem of the inference procedure for the Hawkes processes (see Theorem 1). This is the main result of this paper. In particular, we allow for nonlinear intensity, under some Lipschitz continuity assumptions on the intensity function. This extends Clinet and Yoshida (2017) (Theorem 4.6), Kwan (2023) (Theorem 3.4.3), Kwan, Chen and Dunsmuir (2024) (Theorem 3.2), Potiron and Volkov (2025) (Theorem 1) and Potiron (2025b) (Theorem 2), who are restricted to linear Hawkes processes. The proofs are based on an application of Theorem 1 in Potiron (2025b). The main novelty in the proofs is to adapt the arguments from the linear intensity case to the nonlinear intensity case.

2. Parametric inference for point processes. In this section, we develop inference for point processes in which its intensity has a parametric form. The inference procedure is based on maximum likelihood estimation.

We start with an introduction to the point process. We assume that the point process N_t is of dimension d. For any index $i = 1, \dots, d$, each component of the point process $N_t^{(i)}$ counts the cumulative number of events between the starting time 0 and the final time t for the ith process. Here, we denote the ith component of a vector V by $V^{(i)}$. We define $N_t^{(i)}$ as a point process on the space of nonnegative real numbers $\mathbb{R}^+$, i.e. a family

$$(N^{(i)}(C))_{C \in \mathcal{B}(\mathbb{R}^+)}$$

of random variables with values in the space of natural numbers $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$. Here, $\mathcal{B}(S)$ denotes the Borel σ-algebra on the space S for any space S. Moreover, the point process $N^{(i)}(C)$ can be expressed as

$$N^{(i)}(C) = \sum_{k \in \mathbb{N}} \mathbf{1}_C(T_k^{(i)}).$$

Finally, the sequence of event times $(T_k^{(i)})_{k \in \mathbb{N}}$ takes its values in the space of nonnegative real numbers $\mathbb{R}^+$ and is random.

The definition of point process requires some specific assumptions on the event times. We assume that the time of the first event $T_0^{(i)}$ is equal to 0 a.s. and the following times are increasing for each process a.s. Namely, we assume that

$$\mathbb{P}(T_0^{(i)} = 0 \text{ and } T_k^{(i)} < T_{k+1}^{(i)} \text{ for } k \in \mathbb{N}_* \text{ and } i = 1, \dots, d) = 1. \tag{2}$$

Here, we define for any space S such that $0 \in S$ the space without zero as S_*. We also assume that no events happen at the same time for different processes a.s., i.e.

$$\mathbb{P}(T_k^{(i)} \neq T_l^{(j)} \text{ for } k, l \in \mathbb{N}_* \text{ and } i, j = 1, \dots, d \text{ s.t. } i \neq j) = 1.$$

We define the probabilistic tools in what follows. We introduce the stochastic basis $\mathcal{B} = \big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}^+}, \mathbb{P}\big)$, namely a probability space equipped with a filtration. The filtration $\mathcal{F}_t$ represents the information available at the time $t \in \mathbb{R}^+$. We assume that the stochastic basis $\mathcal{B}$ satisfies the usual conditions. We first introduce the definition of the $\mathcal{F}_t$ intensity for the point process N_t.

DEFINITION 1. Any stochastic process λ_t defined on the space of real nonnegative numbers $\mathbb{R}^+$ and satisfying the following properties is called an $\mathcal{F}_t$ intensity of the point process N_t. First, the stochastic process λ_t is $\mathcal{F}_t$ progressively measurable. Secondly, the stochastic process λ_t is of dimension d where each component $\lambda_t^{(i)}$ takes its values in the space of nonnegative real numbers $\mathbb{R}^+$. Moreover, we have for any interval $(a, b] \subset \mathbb{R}^+$ that

$$\mathbb{E}[N((a, b]) \mid \mathcal{F}_a] = \mathbb{E}\Big[\int_a^b \lambda_s ds \mid \mathcal{F}_a\Big] \text{ a.s.} \tag{3}$$

Intuitively, the intensity corresponds to the expected number of events given the past information, i.e.

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

Moreover, we have that the compensated point process defined as

$$M_t = N_t - \int_0^t \lambda_s ds \tag{4}$$

is an $\mathcal{F}_t$ martingale a.s. Finally, we note that $N((a,b])$ is a.s. finite if and only if $\int_a^b \lambda_s ds$ is a.s. finite. For background on point processes, the reader can consult Jacod (1975), Jacod and Shiryaev (2003), Daley and Vere-Jones (2003), and Daley and Vere-Jones (2008).

The present work is concerned with point processes N_t admitting an $\mathcal{F}_t$ intensity which has a parametric form. More specifically, we introduce the parameter space Θ which consists of n parameters. We also introduce the family of intensities $\lambda_t(\theta)$ for any parameter $\theta \in \Theta$. We assume that the intensity process $\lambda_t(\theta)$ is of dimension d where each component $\lambda_t^{(i)}(\theta)$ takes its value in the space of nonnegative real numbers $\mathbb{R}^+$ for any parameter $\theta \in \Theta$ and any $\omega \in \Omega$. Finally, we assume the existence of the true parameter $\theta^* \in \Theta$ such that

$$\lambda_t = \lambda_t(\theta^*). \tag{5}$$

For any parameter $\theta \in \Theta$, we rely on the log likelihood process (see Ogata (1978) and Daley and Vere-Jones (2003))

$$l_T(\theta) = \sum_{i=1}^{d} \int_0^T \log(\lambda_t^{(i)}(\theta)) dN_t^{(i)} - \sum_{i=1}^{d} \int_0^T \lambda_t^{(i)}(\theta) dt. \tag{6}$$

Here, 0 is the starting time and T is the final time. Then, the maximum likelihood estimator is defined as a maximizer of the log likelihood process between the starting time 0 and the final time T, i.e.

$$\widehat{\theta}_T \in \operatorname{argmax}_{\theta\in\Theta} l_T(\theta).$$

The form of the asymptotic covariance matrix relies on the ergodicity conditions in parametric inference for point processes. More specifically, we focus on the stochastic process $X_t = (\lambda_t(\theta^*), \lambda_t(\theta), \partial_\theta \lambda_t(\theta))$ taking values in the space E^d where $E = \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^n$. We denote by $X_t^{(i)} \in E$ the ith component of the stochastic process X_t. Here, $\partial_\theta G(\theta)$ denotes the vector of partial derivatives for any function $G(\theta)$, i.e. $\partial_\theta G(\theta) = \frac{\partial G}{\partial \theta}(\theta)$. Moreover, we denote by $\mathcal{C}_b(E,F)$ the space of bounded and continuous functions from the starting space E to the final space F. In what follows, we provide the definition of ergodicity. This corresponds to Definition 3.1 (p. 1805) in Clinet and Yoshida (2017). See also Definition C1 in Supplement C of Potiron and Volkov (2025) and Definition 1 in Potiron (2025b).

DEFINITION 2. We say that the stochastic process X is ergodic if for any index $i = 1, \dots, d$ there exists a limit function $\pi^{(i)} : C_b(E,\mathbb{R}) \to \mathbb{R}$ such that for any function $\psi \in C_b(E,\mathbb{R})$ we have as the final time $T \to \infty$ that

$$\frac{1}{T} \int_0^T \psi(X_s^{(i)}) ds \xrightarrow{\mathbb{P}} \pi^{(i)}(\psi).$$

Lemma 7 states that the stochastic process X_t is stable, i.e. for any parameter $\theta \in \Theta$ and any index $i = 1, \dots, d$ there exists an $\mathbb{R}_+^*$ valued random variable $\lambda_l^{(i)}(\theta)$ such that as the final time $T \to \infty$ we have

$$X_T^{(i)} \xrightarrow{\mathcal{D}} \big(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta)\big).$$

From Lemma 8, the stochastic process X_t is also ergodic in the sense of Definition 2 for any parameter $\theta \in \Theta$. Moreover, we have the more explicit expression of the limit function for any index $i = 1, \dots, d$ as

$$\pi^{(i)}(\psi) = \mathbb{E}\big[\psi(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta))\big].$$

Since the space of bounded functions is not large enough to establish the central limit theorem, we introduce a bigger space in the following definition. We denote this bigger space by $C_\uparrow(E,\mathbb{R})$. This corresponds to Definition 3.7 (p. 1806) in Clinet and Yoshida (2017) and Definition 2 in Potiron (2025b).

DEFINITION 3. We denote by $C_{\uparrow}(E,\mathbb{R})$ the set of continuous functions $\psi:(u,v,w)\rightarrow\psi(u,v,w)$ from the starting space E to the final space $\mathbb{R}$ that satisfy

(a) ψ is continuous on $\mathbb{R}_*^+\times\mathbb{R}_*^+\times\mathbb{R}^n$.
(b) ψ is of polynomial growth in u, v, w, $\frac{\mathbf{1}_{\{u>0\}}}{u}$ and $\frac{\mathbf{1}_{\{v>0\}}}{v}$.
(c) For any $(u,v,w)\in E$, we have $\psi(0,v,w)=\psi(u,0,w)=0$.

Lemma 10 extends the starting space of the limit function π from $C_b(E,\mathbb{R})$ to $C_{\uparrow}(E,\mathbb{R})$ and gives a more explicit form. More specifically, it shows that for any index $i=1,\dots,d$ and any parameter $\theta\in\Theta$ there exists a probability measure $\pi_\theta^{(i)}$ on the space E such that for any function $\psi\in C_{\uparrow}(E,\mathbb{R})$ we have

$$\pi^{(i)}(\psi)=\int_E\psi(u,v,w)\pi_\theta^{(i)}(du,dv,dw).$$

Moreover, Lemma 11 ensures that the family of intensities $\lambda_t(\theta)$ does not explode on any compact space based on these ergodicity assumptions.

We have now all the ingredients to derive the form of the asymptotic covariance matrix. Namely, we define the asymptotic Fisher information matrix Γ of dimension $n\times n$ as

$$\Gamma=\sum_{i=1}^{d}\int_E w^{\otimes 2}\frac{1}{u}\pi_{\theta^*}^{(i)}(du,dv,dw). \tag{7}$$

Here, we define the tensor product of a vector $z\in\mathbb{R}^n$ as $z^{\otimes 2}=zz^T\in\mathbb{R}^{n\times n}$. The Fisher information matrix measures the amount of information that the intensity λ_t carries about the parameter θ^*. Formally, it is the expected value of the observed information. We use the Fisher information matrix to calculate the covariance matrices associated with maximum likelihood estimation. Namely, the inverse information matrix Γ^{-1} is the asymptotic covariance matrix. We show in the proof of Theorem 1 from Potiron (2025b) that we can reexpress the asymptotic Fisher information matrix as

$$\Gamma=-\lim_{T\rightarrow\infty}\frac{1}{T}\mathbb{E}\big[\partial_\theta^2 l_T(\theta^*)\big].$$

Here, $\partial_\theta^2 G(\theta)$ denotes the Hessian matrix for any function $G(\theta)$, i.e. $\partial_\theta^2 G(\theta)=\frac{\partial^2 G}{\partial^2\theta}(\theta)$.

3. Parametric inference for nonlinear Hawkes processes. In this section, we consider inference for Hawkes mutually exciting processes. We assume that the intensity is parametric and nonlinear. We also assume that the kernel is parametric and has a general form. We show the central limit theorem of the inference procedure for the Hawkes processes. This is the main result of this paper. In particular, we allow for nonlinear intensity, under some Lipschitz continuity assumptions on the intensity function. This extends Clinet and Yoshida (2017) (Theorem 4.6), Kwan (2023) (Theorem 3.4.3), Kwan, Chen and Dunsmuir (2024) (Theorem 3.2), Potiron and Volkov (2025) (Theorem 1) and Potiron (2025b) (Theorem 2), who are restricted to linear Hawkes processes. The proofs are based on an application of Theorem 1 in Potiron (2025b). The main novelty in the proofs is to adapt the arguments from the linear intensity case to the nonlinear intensity case.

We consider Hawkes mutually exciting processes in which the intensity is parametric and nonlinear. Also, the kernel is parametric and has a general form. More specifically, we introduce for any parameter $\theta\in\Theta$ the family of intensities

$$\lambda_t(\theta)=f\Big(\nu,\int_0^t h(t-s,\kappa)\,dN_s\Big). \tag{8}$$

Here, the function f of dimension d is nonlinear and nonrandom. Also, the kernel h is a matrix of dimension $d \times d$. The point processes are mutually exciting in the sense that the diagonal components $h^{(i,i)}$ are self exciting terms for the ith process and non diagonal components $h^{(i,j)}$ are cross exciting terms for the ith process made by events from the jth process. Moreover, the baseline ν consists of q parameters, while κ consists of $n-q$ kernel parameters. We assume that the parameter θ has the form $\theta = (\nu, \kappa)$, and that they belong to the parameter space $\Theta = (\Theta_\nu, \Theta_\kappa)$. We also assume that $n \geq q + d$. Finally, we assume the existence of the true parameter $\theta^* \in \Theta$ such that

$$\lambda_t = \lambda_t(\theta^*). \tag{9}$$

Here, we assume that the parameter θ^* has the form $\theta^* = (\nu^*, \kappa^*)$ in which $\nu^* \in \Theta_\nu$ and $\kappa^* \in \Theta_\kappa$.

Before introducing the assumptions, we first need to introduce some notation. First, we define the space $\overline{\Theta}$ as the closure space of Θ. We also define the sum of the absolute values of its components as $|x| = \sum_i |x_i|$ when x is a real number, a vector, a matrix or a tensor. We denote the spectral radius of any matrix ϕ as $\rho(\phi)$. Then, we denote by κ_t^+ the maximum argument parameter of the spectral radius function $\rho(h(t,\kappa))$ for any time $t \in \mathbb{R}^+$. More specifically, the kernel parameter κ_t^+ is defined implicitly for any time $t \in \mathbb{R}^+$ as

$$\rho\big(h(t, \kappa_t^+)\big) = \sup_{\kappa \in \Theta_\kappa} \big(h(t,\kappa)\big). \tag{10}$$

Moreover, we define the matrix ϕ of dimension $d \times d$ as the integral of $h(t, \kappa_t^+)$ over time, i.e.

$$\phi = \int_0^\infty h(t, \kappa_t^+) dt.$$

In addition, we denote by $\kappa_{t,2}^+$ the maximum argument parameter of the spectral radius function $\rho(h^2(t,\kappa))$ for any time $t \in \mathbb{R}^+$. More specifically, the kernel parameter $\kappa_{t,2}^+$ is defined implicitly for any time $t \in \mathbb{R}^+$ as

$$\rho\big(h^2(t, \kappa_{t,2}^+)\big) = \sup_{\kappa \in \Theta_\kappa} \rho\big(h^2(t,\kappa)\big). \tag{11}$$

Then, we define the matrix ϕ_2 of dimension $d \times d$ as the integral of $h^2(t, \kappa_{t,2}^+)$ over time, i.e.

$$\phi_2 = \int_0^\infty h^2(t, \kappa_{t,2}^+) dt.$$

We denote by $k_{t,3}^{(i,j)}$ the maximum argument of $\big|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k)}\big|$ for any index $i = 1, \dots, d$, any index $j = 1, \dots, d$ and any time $t \in \mathbb{R}^+$. More specifically, the kernel index $k_{t,3}^{(i,j)}$ is defined implicitly for any index $i = 1, \dots, d$, any index $j = 1, \dots, d$ and any time $t \in \mathbb{R}^+$ as

$$\big|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k_{t,3}^{(i,j)})}\big| = \sup_{k=1,\dots,n-q} \big|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k)}\big|. \tag{12}$$

Moreover, we define the matrix $\phi_3(\kappa)$ of dimension $d \times d$ for any kernel parameter $\kappa \in \Theta_\kappa$, any index $i = 1, \dots, d$ and any index $j = 1, \dots, d$ as the integral of $\big|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k_{t,3}^{(i,j)})}\big|$ over time, i.e.

$$\phi_3^{(i,j)}(\kappa) = \int_0^\infty \big|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k_{t,3}^{(i,j)})}\big| dt.$$

In addition, we denote by $(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})$ the maximum argument of $\big|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k,l)}\big|$ for any kernel parameter $\kappa \in \Theta_\kappa$, any index $i = 1, \dots, d$, any index $j = 1, \dots, d$ and any time

$t \in \mathbb{R}^+$. More specifically, the kernel indices $(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})$ are defined implicitly for any index $i = 1, \dots, d$, any index $j = 1, \dots, d$ and any time $t \in \mathbb{R}^+$ as

$$\big|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})}\big| = \sup_{k,l=1,\dots,n-q} \big|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k,l)}\big|. \tag{13}$$

Moreover, we define the matrix $\phi_4(\kappa)$ of dimension $d \times d$ for any kernel parameter $\kappa \in \Theta_\kappa$, any index $i = 1, \dots, d$ and any index $j = 1, \dots, d$ as the integral of $\big|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})}\big|$ over time, i.e.

$$\phi_4^{(i,j)}(\kappa) = \int_0^\infty \big|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})}\big| dt.$$

We define the matrix ϕ_5 of dimension $d \times d$ for any index $i = 1, \dots, d$ and any index $j = 1, \dots, d$ as the integral of $\big|\partial_\kappa h^{(i,j)}(t,\kappa^*)^{(k_{t,3}^{(i,j)})}\big|^2$ over time, i.e.

$$\phi_5^{(i,j)} = \int_0^\infty \big|\partial_\kappa h^{(i,j)}(t,\kappa^*)^{(k_{t,3}^{(i,j)})}\big|^2 dt.$$

Finally, we define the matrix ϕ_6 of dimension $d \times d$ for any index $i = 1, \dots, d$ and any index $j = 1, \dots, d$ as the integral of $\big|\partial_\kappa^2 h^{(i,j)}(t,\kappa^*)^{(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})}\big|^2$ over time, i.e.

$$\phi_6^{(i,j)} = \int_0^\infty \big|\partial_\kappa^2 h^{(i,j)}(t,\kappa^*)^{(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})}\big|^2 ds.$$

We now introduce a set of assumptions required for the central limit theorem of the parametric inference procedure for Hawkes processes. In particular, we allow for nonlinear intensity under some Lipschitz continuity assumptions on the intensity function.

ASSUMPTION 1. (a) The family of nonlinear functions $f : \Theta_\nu \times \mathbb{R}_+^{d\times d} \to \mathbb{R}_+^d$ is $\mathcal{B}(\Theta_\nu) \otimes \mathbb{R}_+^{d\times d}$ measurable.

(b) There exists a positive real number $f_- > 0$ such that for any index $i = 1, \dots, d$, any kernel parameter $\nu \in \Theta_\nu$ and any matrix $x \in \mathbb{R}_+^{d\times d}$ we have $f^{(i)}(\nu, x) > f_-$.

(c) For any kernel parameter $\kappa \in \Theta_\kappa$, any time $t \in \mathbb{R}^+$ any index $i = 1, \dots, d$ and any index $j = 1, \dots, d$, we have that the kernel function is positive, i.e. $h^{(i,j)}(t,\kappa) > 0$.

(d) The parameter space $\Theta \subset \mathbb{R}^n$ is such that its closure $\overline{\Theta}$ is a compact space.

(e) We have $f(0,0) = 0$. The nonlinear function f is Lipschitz continuous with a real constant such that $f_+ > 0$, namely for any kernel parameter $\nu_1 \in \Theta_\nu$, any kernel parameter $\nu_2 \in \Theta_\nu$, any matrix $x_1 \in \mathbb{R}_+^{d\times d}$ and any matrix $x_2 \in \mathbb{R}_+^{d\times d}$ we have

$$|f(\nu_1, x_1) - f(\nu_2, x_2)| \le f_+\big(|\nu_1 - \nu_2| + |x_1 - x_2|\big).$$

(f) We have $f_+ \rho(\phi) < 1$.

(g) We have $\rho(\phi_2) < +\infty$.

(h) We have the nonlinear function $f : \Theta_\nu \times \mathbb{R}_+^{d\times d} \to \mathbb{R}_+^{d\times d}$ is twice continuously differentiable and there exists a continuous extension to the space $\overline{\Theta}_\nu \times \mathbb{R}^+$.

(i) We have $df(0,0) = 0$ and $d^2 f(0,0) = 0$. The nonlinear function derivatives df and $d^2 f$ are Lipschitz continuous with real constants such that $f_{1,+} > 0$ and $f_{2,+} > 0$, namely for any kernel parameter $\nu_1 \in \Theta_\nu$, any kernel parameter $\nu_2 \in \Theta_\nu$, any matrix $x_1 \in \mathbb{R}_+^{d\times d}$ and any matrix $x_2 \in \mathbb{R}_+^{d\times d}$ we have

$$|df(\nu_1, x_1) - df(\nu_2, x_2)| \le f_{1,+}\big(|\nu_1 - \nu_2| + |x_1 - x_2|\big),$$

$$|d^2 f(\nu_1, x_1) - d^2 f(\nu_2, x_2)| \le f_{2,+}\big(|\nu_1 - \nu_2| + |x_1 - x_2|\big).$$

(j) We have $\rho(\phi_5) < +\infty$ and $\rho(\phi_6) < +\infty$.

(k) There exists a positive real number $\phi_+ \in (0,1)$ such that for any kernel parameter $\kappa \in \Theta_\kappa$ we have $f_{1,+}\rho(\phi_3(\kappa)) < \phi_+$ and $f_{2,+}\rho(\phi_4(\kappa)) < \phi_+$.
(l) We have $\mathbb{P}\big(\lambda_l(\theta^*) = \lambda_l(\theta)\big) = 1$ implies that $\theta^* = \theta$.
(m) For any time $t \in \mathbb{R}^+$ a.e., we have the kernel function $\kappa \to h(t,\kappa)$ is twice continuously differentiable from the kernel parameter space Θ_κ to the space $\mathbb{R}_+^{d\times d}$ and there exists a continuous extension to the space $\overline{\Theta}_\kappa$.

Assumption 1 (a) is a novel assumption of measurability on the nonlinear function f. Assumption 1 (b) is also a novel assumption to bound below the nonlinear function f. In particular, Assumption 1 (b) imply that the point processes are well-defined and is also required in the simpler case of heterogeneous Poisson processes without a kernel (see Daley and Vere-Jones (2003)). In addition, Assumption 1 (c) is restrictive for kernels with inhibitory effects and already appears in Assumption 2 (b) from Potiron (2025b). These three assumptions are used to prove Assumption 1 (a) in Potiron (2025b). Moreover, Assumption 1 (d) corresponds exactly to Assumption 1 (b) in Potiron (2025b).

Assumption 1 (e) is a novel assumption which states that the nonlinear function f is Lipschitz continuous. This is the most important assumption to adapt the proofs from the linear intensity to the nonlinear intensity case. A similar assumption is used in Brémaud and Massoulié (1996) and Assumption 1 from Zhu (2013) in a different framework. Assumption 1 (f) states that the spectral radius of the kernel integral when evaluated at the maximum argument parameter of $\rho(h(t,\kappa))$ is strictly smaller than the Lipschitz constant inverse $1/f_+$. This is slightly stronger than the assumption which is necessary to obtain a stationary intensity with finite first moment (see Lemma 1 (p. 495) in Hawkes and Oakes (1974) and Theorem 1 (p. 1567) in Brémaud and Massoulié (1996)). In the case when $f_+ = 1$, this corresponds exactly to the case $\rho(\phi) < 1$ in Assumption 2 (d) from Potiron (2025b). Assumption 1 (g) ensures that the spectral radius of the kernel integral, when squared and evaluated at the maximum argument parameter of $\rho(h^2(t,\kappa))$, is finite. This is exactly the case $\rho(\phi_2) < +\infty$ in Assumption 2 (d) from Potiron (2025b). These three assumptions are mainly used to prove Assumption 1 (d) in Potiron (2025b).

Assumption 1 (h) is a novel regularity assumption on the nonlinear function f which is twice continuously differentiable. This implies Assumption 1 (g) in Potiron (2025b) which requires that the intensity is twice continuously differentiable. Assumption 1 (i) assumes that the nonlinear function derivatives df and d^2f are Lipschitz continuous. This is a novel assumption which is mainly used to get Assumption 1 (h) in Potiron (2025b). In addition, the case $\rho(\phi_5) < +\infty$ in Assumption 1 (j) ensures that the spectral radius of the kernel derivative integral, when squared and evaluated at the maximum argument parameter of $\big|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k)}\big|$, is finite. When $\rho(\phi_6) < +\infty$ in Assumption 1 (j), we have that the spectral radius of the kernel second derivative integral, when squared and evaluated at the maximum argument parameter of $\big|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k,l)}\big|$, is finite. This corresponds to Assumption 2 (g) in Potiron (2025b). Assumption 1 (j) implies Assumption 1 (h) in Potiron (2025b). It is necessary since Assumption 1 (h) in Potiron (2025b) considers the product of the intensity derivatives $\partial_\theta \lambda_t(\theta^*)$ and the intensity Hessian matrix $\partial_\theta^2 \lambda_t(\theta^*)$.

The case $f_{1,+}\rho(\phi_3(\kappa)) < \phi_+$ in Assumption 1 (k) states that the spectral radius of the kernel derivative integral when evaluated at the maximum argument of $\big|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k)}\big|$ is strictly smaller than $\phi_+/f_{1,+}$ uniformly in the space parameter value. The case $f_{2,+}\rho(\phi_4(\kappa)) < \phi_+$ in Assumption 1 (k) ensures that the spectral radius of the kernel second derivative integral when evaluated at the maximum argument of $\big|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k,l)}\big|$ is strictly smaller than $\phi_+/f_{1,+}$ uniformly in the space parameter value. This extends Assumption 2 (g) in Potiron (2025b). Assumption 1 (k) yields Assumption 1 (g) in Potiron (2025b). It is required since Assumption 1 (g) in Potiron (2025b) considers the intensity derivatives $\partial_\theta \lambda_t(\theta)$ and

Hessian matrix of the intensity, i.e. $\partial^2_\theta \lambda_t(\theta)$. Moreover, Assumption 1 (l) is required for the non-degeneracy of the parametric inference procedure and gives Assumption 1 (e) in Potiron (2025b). This is exactly Assumption 2 (h) in Potiron (2025b). Finally, Assumption 1 (m) requires some smoothness assumptions on the kernel shape and is used to show Assumption 1 (f) in Potiron (2025b). This is Assumption 2 (e) in Potiron (2025b).

In the theorem that follows, we state the central limit theorem of the parametric inference procedure for Hawkes processes. The intensity is parametric and nonlinear. The kernel is parametric and has a general form. The inference procedure is based on maximum likelihood estimation. We consider asymptotics when the final time diverges to infinity, i.e. $T \to +\infty$. This is the main result of this paper. In particular, we allow for nonlinear intensity under some Lipschitz continuity assumptions on the intensity function. This extends Clinet and Yoshida (2017) (Theorem 4.6), Kwan (2023) (Theorem 3.4.3), Kwan, Chen and Dunsmuir (2024) (Theorem 3.2), Potiron and Volkov (2025) (Theorem 1) and Potiron (2025b) (Theorem 2), who are restricted to linear Hawkes processes. The proofs are based on an application of Theorem 1 in Potiron (2025b). The main novelty in the proofs is to adapt the arguments from the linear intensity case to the nonlinear intensity case. In the theorem and what follows, ξ is defined as a standard normal vector of dimension an n.

THEOREM 1. *We assume that Assumption 1 holds. As the final time $T \to +\infty$, we have the central limit theorem of the inference procedure for Hawkes processes in which the intensity is parametric and nonlinear, i.e.*

$$\sqrt{T}(\widehat{\theta}_T - \theta^*) \xrightarrow{\mathcal{D}} \Gamma^{-1/2}\xi. \tag{14}$$

4. Proofs. In this section, we give the proofs of the central limit theorem for Hawkes processes in which the intensity is parametric and nonlinear, i.e. Theorem 1. The proofs are based on an application of Theorem 1 in Potiron (2025b). The main novelty in the proofs is to adapt the arguments from the linear intensity case to the nonlinear intensity case.

In what follows, the constant C refers to a generic constant, which can differ from line to line. We first show the following lemma, which corresponds to Assumption 1 (a) in Potiron (2025b).

LEMMA 1. *We assume that Assumptions 1 (a), (b) and (c) hold. Then, the family of intensities $\lambda : \Omega \times \mathbb{R}^+ \times \Theta \to \mathbb{R}^d_+$ defined in Equation (8) is $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{B}(\Theta)$ measurable.*

PROOF OF LEMMA 1. First, we get by Definition (8), Assumptions 1 (b) and (c) that the intensity process is nonnegative, namely $\lambda_t \geq 0$ for any time $t \in \mathbb{R}^+$ and any $\omega \in \Omega$. Then, we can deduce that the intensity process λ is $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{B}(\Theta)$ measurable by Definition (8) and Assumption 1 (a). □

We now prove the following lemma, which corresponds to Assumption 1 (b) in Potiron (2025b).

LEMMA 2. *We assume that Assumption 1 (d) holds. Then, the parameter space $\Theta \subset \mathbb{R}^n$ is such that its closure $\overline{\Theta}$ is a compact space.*

PROOF OF LEMMA 2. The statement of the lemma corresponds exactly to Assumption 1 (d). □

The next lemma is Lemma A.2 (p. 1825) from Clinet and Yoshida (2017) and Lemma 9 in Potiron (2025b). Although the general statement holds for a stochastic process g_t, we will only require the particular case when g_t is a nonrandom function.

LEMMA 3. *We assume that Assumptions 1 (a), (b) and (c) hold. Let an integer $p \in \mathbb{N}_*$ and g_t be a stochastic process such that $g_t^{2^p}$ is a.s. locally integrable on the space $\mathbb{R}^+$. Then, we have for any index $i = 1, \dots, d$ that*

$$\mathbb{E}\Big[\Big|\int_0^T g_t dN_t^{(i)}\Big|^{2^p}\Big] \leq \mathbb{E}\Big[\int_0^T g_t^{2^p} \lambda_t^{(i)} dN_t^{(i)}\Big] + \mathbb{E}\Big[\Big|\int_0^T g_t^2 \lambda_t^{(i)} dN_t^{(i)}\Big|^{2^{p-1}}\Big].$$

PROOF OF LEMMA 3. This is a direct application of Lemma A.2 (p. 1825) from Clinet and Yoshida (2017). □

We define the L^p norm of a random variable Y as $||Y||_p = \mathbb{E}[|Y|^p]^{1/p}$ for any positive real number p. We now show the following lemma, which corresponds to Assumption 1 (d) in Potiron (2025b). This complements Lemma C4 in Supplement C of Potiron and Volkov (2025) and Lemma 10 in Potiron (2025b).

LEMMA 4. *We assume that Assumptions 1 (a), (b), (c), (d), (e), (f) and (g) hold. Then, we have*

$$\sup_{t \in \mathbb{R}^+} \Big|\Big| \sup_{\theta \in \Theta} |\lambda_t(\theta)| \Big|\Big|_2 < +\infty.$$

PROOF OF LEMMA 4. We first prove that

$$\sup_{t \in \mathbb{R}^+} \Big|\Big| \sup_{\theta \in \Theta} |\lambda_t(\theta)| \Big|\Big|_1 < +\infty. \tag{15}$$

We have

$$\begin{aligned}
\Big|\Big| \sup_{\theta \in \Theta} |\lambda_t(\theta)| \Big|\Big|_1 &= \mathbb{E}\Big[\sup_{\theta \in \Theta} |\lambda_t(\theta)|\Big] \\
&= \mathbb{E}\Big[\sup_{\theta \in \Theta} \Big\{ \sum_{i=1}^d \lambda_t^{(i)}(\theta) \Big\}\Big] \\
&= \mathbb{E}\Big[\sup_{\theta \in \Theta} \Big\{ \sum_{i=1}^d f^{(i)}\Big(\nu, \sum_{j=1}^d \int_0^t h^{(i,j)}(t-s,\kappa)\, dN_s^{(j)}\Big) \Big\}\Big].
\end{aligned}$$

Here, we use the definition of the norm $|| \, ||_1$ in the first equality, the definition of $| \, |$ in the second equality, and Definition (8) in the third equality.

Then, we have that

$$\begin{aligned}
\Big|\Big| \sup_{\theta \in \Theta} |\lambda_t(\theta)| \Big|\Big|_1 &= \mathbb{E}\Big[\sup_{\theta \in \Theta} \Big\{ \sum_{i=1}^d f^{(i)}\Big(\nu, \sum_{j=1}^d \int_0^t h^{(i,j)}(t-s,\kappa)\, dN_s^{(j)}\Big) \Big\}\Big] \\
&\leq \mathbb{E}\Big[\sup_{\theta \in \Theta} \Big\{ f_+\Big(|\nu| + \sum_{j=1}^d \int_0^t h^{(i,j)}(t-s,\kappa) dN_s^{(j)}\Big) \Big\}\Big] \\
&= f_+ \mathbb{E}\Big[\sup_{\theta \in \Theta} \Big\{ |\nu| + \sum_{j=1}^d \int_0^t h^{(i,j)}(t-s,\kappa) dN_s^{(j)} \Big\}\Big].
\end{aligned}$$

Here, we use Assumption 1 (e) in the inequality and expectation properties in the second equality.

Moreover, we have

$$\begin{aligned}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 &\leq f_+\mathbb{E}\Big[\sup_{\theta\in\Theta}\Big\{\big|\nu\big|+\sum_{j=1}^{d}\int_0^t h^{(i,j)}(t-s,\kappa)dN_s^{(j)}\Big\}\Big]\\ &\leq f_+\mathbb{E}\Big[C+\sup_{\theta\in\Theta}\Big\{\sum_{j=1}^{d}\int_0^t h^{(i,j)}(t-s,\kappa)dN_s^{(j)}\Big\}\Big]\\ &= C+f_+\mathbb{E}\Big[\sup_{\theta\in\Theta}\Big\{\sum_{j=1}^{d}\int_0^t h^{(i,j)}(t-s,\kappa)dN_s^{(j)}\Big\}\Big].\end{aligned}$$

Here, we use Assumption 1 (d) in the inequality and expectation properties in the equality.

Then, we obtain

$$\begin{aligned}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 &\leq C+f_+\mathbb{E}\Big[\sup_{\theta\in\Theta}\Big\{\sum_{j=1}^{d}\int_0^t h^{(i,j)}(t-s,\kappa)dN_s^{(j)}\Big\}\Big]\\ &= C+f_+\mathbb{E}\Big[\sup_{\kappa\in\Theta_\kappa}\Big\{\sum_{j=1}^{d}\int_0^t h^{(i,j)}(t-s,\kappa)dN_s^{(j)}\Big\}\Big]\\ &\leq C+f_+\mathbb{E}\Big[\sum_{j=1}^{d}\int_0^t h^{(i,j)}(t-s,\kappa_s^+)dN_s^{(j)}\Big].\end{aligned}$$

Here, we use the fact that the kernel depends only on the parameter κ in the equality, and Definition (10) in the second inequality.

In addition, we obtain

$$\begin{aligned}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 &\leq C+f_+\mathbb{E}\Big[\sum_{j=1}^{d}\int_0^t h^{(i,j)}(t-s,\kappa_s^+)dN_s^{(j)}\Big]\\ &= C+f_+\sum_{j=1}^{d}\int_0^t h^{(i,j)}(t-s,\kappa_s^+)\Big|\Big|\lambda_t^{(j)}\Big|\Big|_1 ds\\ &\leq C+f_+\sum_{j=1}^{d}\sup_{t\in\mathbb{R}^+}\Big|\Big|\lambda_t^{(j)}\Big|\Big|_1\int_0^t h^{(i,j)}(t-s,\kappa_s^+)ds.\end{aligned}$$

Here, we use martingale properties in the equality and supremum properties in the second inequality.

By Assumption 1 (f), there exists a real positive number h_+ which satisfies $0<f_+h_+<1$. In particular, this real positive number h_+ satisfies for any $i=1,\dots,d$ and any $j=1,\dots,d$ that

$$\int_0^t h^{(i,j)}(t-s,\kappa_s^+)ds \leq h_+. \tag{16}$$

Then, we have

$$\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 \leq C+f_+\sum_{j=1}^{d}\sup_{t\in\mathbb{R}^+}\Big|\Big|\lambda_t^{(j)}\Big|\Big|_1\int_0^t h^{(i,j)}(t-s,\kappa_s^+)ds$$

$$\leq C + \sum_{j=1}^{d} \sup_{t\in\mathbb{R}^+} \left|\left| \lambda_t^{(j)} \right|\right|_1 f_+ h_+$$

$$\leq C + \sup_{t\in\mathbb{R}^+} \left|\left| \sup_{\theta\in\Theta} |\lambda_t(\theta)| \right|\right|_1 f_+ h_+.$$

Here, we use Expression (16) in the second inequality, the definition of $|\ |$ and supremum properties in the third inequality. By taking the supremum over the time $t \in \mathbb{R}^+$ on the left side of the expression, we can deduce that

$$\sup_{t\in\mathbb{R}^+} \left|\left| \sup_{\theta\in\Theta} |\lambda_t(\theta)| \right|\right|_1 \leq C + f_+ h_+ \sup_{t\in\mathbb{R}^+} \left|\left| \sup_{\theta\in\Theta} |\lambda_t(\theta)| \right|\right|_1. \tag{17}$$

Since $0 < f_+ h_+ < 1$, Expression (17) implies

$$\sup_{t\in\mathbb{R}^+} \left|\left| \sup_{\theta\in\Theta} |\lambda_t(\theta)| \right|\right|_1 \leq \frac{C}{1 - f_+ h_+}. \tag{18}$$

Thus, we have shown Expression (15). Finally, the lemma can be shown by an application of Lemma 3 in the case $g_t = h^{(i,j)}(t-s, \kappa_s^+)$ with Assumption 1 (g). □

We define the function F for any index $i = 1, \dots, d$, any parameter $\theta \in \Theta$ and any time $t \in \mathbb{R}^+$ as

$$F_t^{(i)}(\theta) = \frac{(\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta)}{\lambda_t^{(i)}(\theta)^2}.$$

We now show the following lemma, which corresponds to Assumption 1 (h) in Potiron (2025b). This extends Lemma 13 in Potiron (2025b) to the nonlinear intensity case.

LEMMA 5. *We assume that Assumptions 1 (a),(b), (c), (e), (h), (i) and (j) hold. Then, we have for any index $i = 1, \dots, d$ that*

$$\sup_{t\in\mathbb{R}^+} \left|\left| \left| \partial_\theta \Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \Big) \right| \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} \right|\right|_2 < +\infty,$$

$$\sup_{t\in\mathbb{R}^+} \left|\left| \left| \partial_\theta F_t^{(i)}(\theta^*) \right| \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} \right|\right|_1 < +\infty,$$

$$\sup_{t\in\mathbb{R}^+} \left|\left| \left| (\partial_\theta^2 \lambda_t^{(i)})(\theta^*) \right| \left| \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \right| \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} \right|\right|_1 < +\infty.$$

PROOF OF LEMMA 5. We define I for any index $i = 1, \dots, d$ as

$$I = \sup_{t\in\mathbb{R}^+} \left|\left| \left| \partial_\theta \Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \Big) \right| \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} \right|\right|_2.$$

By Assumptions 1 (b) and (c), we can deduce for any index $i = 1, \dots, d$ that

$$I = \sup_{t\in\mathbb{R}^+} \left|\left| \left| \partial_\theta \Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \Big) \right| \lambda_t^{(i)}(\theta^*) \right|\right|_2.$$

By derivative formula, we obtain for any index $i = 1, \dots, d$ that

$$I = \sup_{t\in\mathbb{R}^+} \left|\left| \left| \frac{\partial_\theta^2 \lambda_t^{(i)}(\theta^*) - (\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*)}{\lambda_t^{(i)}(\theta^*)^2} \right| \lambda_t^{(i)}(\theta^*) \right|\right|_2.$$

This can be reexpressed for any index $i = 1, \dots, d$ as

$$I = \sup_{t \in \mathbb{R}^+} \Big|\Big| \Big| \frac{\partial_\theta^2 \lambda_t^{(i)}(\theta^*) - (\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \Big| \Big|\Big|_2.$$

By Assumptions 1 (b) and (c), we can deduce for any index $i = 1, \dots, d$ that

$$I < \frac{1}{f_-^2} \sup_{t \in \mathbb{R}^+} \Big|\Big| \Big| \partial_\theta^2 \lambda_t^{(i)}(\theta^*) - (\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*) \Big| \Big|\Big|_2.$$

Moreover, we get by Definition (8) for any index $i = 1, \dots, d$ that

$$I < \frac{1}{f_-^2} \sup_{t \in \mathbb{R}^+} \Big|\Big| \Big| \partial_\theta^2 \Big(f^{(i)} \Big(\nu^*, \int_0^t h(t-s, \kappa^*) dN_s \Big) \Big)$$

$$-\Big\{ \partial_\theta \Big(f^{(i)} \Big(\nu^*, \int_0^t h(t-s, \kappa^*) dN_s \Big) \Big) \Big\}^{\otimes 2} (\theta^*) \Big| \Big|\Big|_2. \tag{19}$$

By Assumptions 1 (e), (i) and (j), we obtain for any index $i = 1, \dots, d$ that

$$I < +\infty. \tag{20}$$

We define II for any index $i = 1, \dots, d$ as

$$II = \sup_{t \in \mathbb{R}^+} \Big|\Big| \Big| \partial_\theta F_t^{(i)}(\theta^*) \Big| \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} \Big|\Big|_1.$$

By definition of $F_t^{(i)}(\theta^*)$, we can deduce for any index $i = 1, \dots, d$ that

$$II = \sup_{t \in \mathbb{R}^+} \Big|\Big| \Big| \partial_\theta \Big(\frac{(\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*)}{\lambda_t^{(i)}(\theta^*)^2} \Big) \Big| \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} \Big|\Big|_1.$$

By Assumptions 1 (b) and (c), we get for any index $i = 1, \dots, d$ that

$$II = \sup_{t \in \mathbb{R}^+} \Big|\Big| \Big| \partial_\theta \Big(\frac{(\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*)}{\lambda_t^{(i)}(\theta^*)^2} \Big) \Big| \lambda_t^{(i)}(\theta^*) \Big|\Big|_1.$$

Moreover, we get by Definition (8) for any index $i = 1, \dots, d$ that

$$II = \sup_{t \in \mathbb{R}^+} \Bigg|\Bigg| \Bigg| \partial_\theta \Bigg(\frac{\Big\{ \partial_\theta \Big(f^{(i)} \Big(\nu^*, \int_0^t h(t-s, \kappa^*) dN_s \Big) \Big) \Big\}^{\otimes 2} (\theta^*)}{f^{(i)} \Big(\nu^*, \int_0^t h(t-s, \kappa^*) dN_s \Big)^2} \Bigg) \Bigg|$$

$$\times f^{(i)} \Big(\nu^*, \int_0^t h(t-s, \kappa^*) dN_s \Big) \Bigg|\Bigg|_1.$$

By extending the arguments from the proof of the case $I < +\infty$ with Assumption 1 (b), we obtain for any index $i = 1, \dots, d$ that

$$II < +\infty. \tag{21}$$

We define III for any index $i = 1, \dots, d$ as

$$III = \sup_{t \in \mathbb{R}^+} \Big|\Big| \Big| (\partial_\theta^2 \lambda_t^{(i)})(\theta^*) \Big| \Big| \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \Big| \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} \Big|\Big|_1.$$

By Assumptions 1 (b) and (c), we get for any index $i = 1, \dots, d$ that

$$III = \sup_{t \in \mathbb{R}^+} \Big|\Big| \left|(\partial_\theta^2 \lambda_t^{(i)})(\theta^*)\right| \left|\frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\right| \Big|\Big|_1.$$

By Assumptions 1 (b) and (c), we obtain for any index $i = 1, \dots, d$ that

$$III < \frac{1}{f_-} \sup_{t \in \mathbb{R}^+} \Big|\Big| \left|(\partial_\theta^2 \lambda_t^{(i)})(\theta^*)\right| \left|\partial_\theta \lambda_t^{(i)}(\theta^*)\right| \Big|\Big|_1.$$

Moreover, we get by Definition (8) for any index $i = 1, \dots, d$ that

$$III < \frac{1}{f_-} \sup_{t \in \mathbb{R}^+} \Big|\Big| \left|\partial_\theta^2 \Big(f^{(i)}\Big(\nu^*, \int_0^t h(t-s, \kappa^*) dN_s\Big)\Big)\right| \times \left|\partial_\theta \Big(f^{(i)}\Big(\nu^*, \int_0^t h(t-s, \kappa^*) dN_s\Big)\Big)\right| \Big|\Big|_1.$$

By extending the arguments from the proof of the case $I < +\infty$, we obtain for any index $i = 1, \dots, d$ that

$$III < +\infty. \tag{22}$$

Finally, we can prove the lemma with Expressions (20), (21) and (22). □

We introduce the covariance supremum for any index $i = 1, \dots, d$ and any time $T > 0$ as

$$\mu_T^{(i)} = \sup_{s \in \mathbb{R}^+} \left| \mathrm{Cov}\left[\phi(X_s^{(i)}), \psi(X_{s+T}^{(i)})\right]\right|. \tag{23}$$

The following definition introduces the notion of mixing for stochastic processes. This corresponds to the definition from Section 3.4 in Clinet and Yoshida (2017). See also Definition C2 in Supplement C of Potiron and Volkov (2025) and Definition 3 in Potiron (2025b).

DEFINITION 4. We say that the stochastic process X_t is mixing if for any function ϕ, any function $\psi \in C_b(E, \mathbb{R})$ and any index $i = 1, \dots, d$ we have as the final time $T \to \infty$ that

$$\mu_T^{(i)} \to 0.$$

The following lemma states that the stochastic process X_t is mixing in the sense of Definition 4. This extends Lemma A.6 (p. 1834) in Clinet and Yoshida (2017), Proposition C1 (i) in Supplement C of Potiron and Volkov (2025) and Lemma 14 in Potiron (2025b) to the nonlinear intensity case.

LEMMA 6. *We assume that Assumptions 1 (a), (b), (c), (d), (e), (f), (g), (h), (i), (j) and (k) hold. Then, the stochastic process X_t is mixing in the sense of Definition 4 for any parameter $\theta \in \Theta$.*

PROOF OF LEMMA 6. We first define the truncation of the stochastic process $X_T^{(i)}$ at the time $t \leq T$ for any index $i = 1, \dots, d$ as

$$\widetilde{X}_{t,T}^{(i)} = \Big(\lambda_t^{(i)}(\theta^*), \sum_{j=1}^d \int_t^T h^{(i,j)}(T-u, \theta) dN_u^{(i)}, \sum_{j=1}^d \int_t^T \partial_\theta\big(h^{(i,j)}(T-u,\theta)\big) dN_u^{(i)}\Big).$$

Then, we can reexpress the covariance supremum $\mu_T^{(i)}$ for any index $i = 1, \dots, d$ as

$$\begin{aligned}\mu_T^{(i)} &= \sup_{s\in\mathbb{R}^+} \left| \operatorname{Cov}\left[\phi(X_s^{(i)}), \psi(X_{s+T}^{(i)})\right]\right| \\ &= \sup_{s\in\mathbb{R}^+} \left| \operatorname{Cov}\left[\phi(X_s^{(i)}), \psi(X_{s+T}^{(i)}) - \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)}) + \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\right]\right|.\end{aligned}$$

Here, we use Definition (23) in the first equality and an algebraic manipulation in the second equality. Using the triangular inequality, covariance and supremum properties, we can bound the covariance supremum $\mu_T^{(i)}$ for any index $i = 1, \dots, d$ as

$$\begin{aligned}\mu_T^{(i)} \leq &\sup_{s\in\mathbb{R}^+} \left| \operatorname{Cov}\left[\phi(X_s^{(i)}), \psi(X_{s+T}^{(i)}) - \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\right]\right| \\ &+ \sup_{s\in\mathbb{R}^+} \left| \operatorname{Cov}\left[\phi(X_s^{(i)}), \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\right]\right|.\end{aligned} \tag{24}$$

We define $I_T^{(i)}$ for any index $i = 1, \dots, d$ and any time $T > 0$ as

$$I_T^{(i)} = \sup_{s\in\mathbb{R}^+} \left| \operatorname{Cov}\left[\phi(X_s^{(i)}), \psi(X_{s+T}^{(i)}) - \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\right]\right|.$$

We also define $II_T^{(i)}$ for any index $i = 1, \dots, d$ and any time $T > 0$ as

$$II_T^{(i)} = \sup_{s\in\mathbb{R}^+} \left| \operatorname{Cov}\left[\phi(X_s^{(i)}), \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\right]\right|.$$

By the definition of $I_T^{(i)}$, Cauchy-Schwarz inequality and supremum properties, we can deduce for any index $i = 1, \dots, d$ and any time $T > 0$ that

$$I_T^{(i)} \leq \sup_{s\in\mathbb{R}^+} \operatorname{Var}\left[\phi(X_s^{(i)})\right] \sup_{s\in\mathbb{R}^+} \operatorname{Var}\left[\psi(X_{s+T}^{(i)}) - \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\right]. \tag{25}$$

By an extension to Lemma 4 with Assumptions 1 (i), (k) and (j), we get for any index $i = 1, \dots, d$ that

$$\sup_{s\in\mathbb{R}^+} \operatorname{Var}\left[\phi(X_s^{(i)})\right] \leq C. \tag{26}$$

Then, we obtain since $\sqrt{T} \to \infty$ as the final time $T \to \infty$ for any index $i = 1, \dots, d$ that

$$\sup_{s\in\mathbb{R}^+} \operatorname{Var}\left[\psi(X_{s+T}^{(i)}) - \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\right] \to 0. \tag{27}$$

By Expressions (25), (26) and (27), we can deduce for any index $i = 1, \dots, d$ that

$$I_T^{(i)} \to 0. \tag{28}$$

Moreover, we obtain by similar arguments to the case $I_T^{(i)} \to 0$ as the final time $T \to \infty$ for any index $i = 1, \dots, d$ that

$$II_T^{(i)} \to 0. \tag{29}$$

By Expressions (24), (28) and (29), we can deduce as the final time $T \to \infty$ for any index $i = 1, \dots, d$ that

$$\mu_T^{(i)} \to 0.$$

□

The following lemma states that the stochastic process X_t is stable. This extends Lemma A.6 (p. 1834) in Clinet and Yoshida (2017), Proposition C1 (ii) in Supplement C of Potiron and Volkov (2025) and Lemma 15 in Potiron (2025b) to the nonlinear intensity case.

LEMMA 7. *We assume that Assumptions 1 (a), (b), (c), (d), (e) (f), (h), (i), (j), (k) and (m) hold. Then, the stochastic process X_t is stable, i.e. for any index $i = 1, \ldots, d$ and any parameter $\theta \in \Theta$ there exists an $\mathbb{R}_+^*$ valued random variable $\lambda_l^{(i)}(\theta)$ such that we have as the final time $T \to \infty$ that*

$$X_T^{(i)} \xrightarrow{\mathcal{D}} \big(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta)\big).$$

PROOF OF LEMMA 7. The proof is obtained by extending the arguments from the proof of Theorem 1 in Brémaud and Massoulié (1996) and Lemma 4 in Brémaud and Massoulié (1996). □

The following lemma states that the stochastic process X_t is ergodic in the sense of Definition 2. Moreover, it delivers a more explicit expression of the limit function $\pi(\psi)$. This shows Assumption 1 (c) in Potiron (2025b). This extends Lemma 3.16 (p. 1815) in Clinet and Yoshida (2017), Proposition C1 (iii) in Supplement C of Potiron and Volkov (2025) and Lemma 16 in Potiron (2025b) to the nonlinear intensity case.

LEMMA 8. *We assume that Assumptions 1 (a), (b), (c), (d), (e), (f), (g), (h), (i), (j) and (k) hold. Then, the stochastic process X_t is ergodic in the sense of Definition 2 for any parameter $\theta \in \Theta$. Moreover, we have for any index $i = 1, \ldots, d$ and any parameter $\theta \in \Theta$ that*

$$\pi^{(i)}(\psi) = \mathbb{E}\big[\psi(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta))\big].$$

PROOF OF LEMMA 8. For any function $\psi \in C_b(E, \mathbb{R})$, any index $i = 1, \ldots, d$ and any final time $T > 0$, we define $V^{(i)}(\psi)$ as

$$V^{(i)}(\psi) = \frac{1}{T} \int_0^T \psi(X_s^{(i)}) ds. \tag{30}$$

To show that the stochastic process X_t is ergodic, it is sufficient to show for any function $\psi \in C_b(E, \mathbb{R})$ and any index $i = 1, \ldots, d$ as the final time $T \to \infty$ that $V^{(i)}(\psi) \xrightarrow{\mathbb{P}} \pi^{(i)}(\psi)$ in which

$$\pi^{(i)}(\psi) = \mathbb{E}\big[\psi(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta))\big].$$

Since L^2 convergence implies convergence in probability, it is sufficient to show L^2 convergence. Since for any random variable X and any nonrandom real number $a \in \mathbb{R}$ we have $\mathbb{E}[(X-a)^2] = \mathrm{Var}[X] + (\mathbb{E}[X] - a)^2$, we can deduce for any function $\psi \in C_b(E, \mathbb{R})$, any index $i = 1, \ldots, d$ and any final time $T > 0$ that

$$\mathbb{E}\big[(V^{(i)}(\psi) - \pi^{(i)}(\psi))^2\big] = \mathrm{Var}[V^{(i)}(\psi)] + (\mathbb{E}[V^{(i)}(\psi)] - \pi^{(i)}(\psi))^2. \tag{31}$$

We define $I^{(i)}$ for any function $\psi \in C_b(E, \mathbb{R})$, any index $i = 1, \ldots, d$ and any final time $T > 0$ as

$$I^{(i)} = \mathrm{Var}[V^{(i)}(\psi)].$$

We also define $II^{(i)}$ for any function $\psi \in C_b(E, \mathbb{R})$, any index $i = 1, \ldots, d$ and any final time $T > 0$ as

$$II^{(i)} = (\mathbb{E}[V^{(i)}(\psi)] - \pi^{(i)}(\psi))^2.$$

We have for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ that

$$\begin{aligned} I^{(i)} &= \mathrm{Var}[V^{(i)}(\psi)] \\ &= \mathrm{Var}\Big[\frac{1}{T}\int_0^T \psi(X_s^{(i)})ds\Big] \\ &= \frac{1}{T^2}\mathrm{Var}\Big[\int_0^T \psi(X_s^{(i)})ds\Big]. \end{aligned}$$

Here, we use the definition of $I^{(i)}$ in the first equality, the definition of $V^{(i)}(\psi)$ in the second equality. We also use the fact that for any nonrandom real number $a \in \mathbb{R}$ and any random variable X we have $\mathrm{Var}[aX] = a^2\,\mathrm{Var}[X]$ in the third equality.

Then, we have for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ that

$$\begin{aligned} I^{(i)} &= \frac{1}{T^2}\mathrm{Var}\Big[\int_0^T \psi(X_s^{(i)})ds\Big] \\ &= \frac{1}{T^2}\lim_{K\to\infty}\mathrm{Var}\Big[\frac{T}{K}\sum_{k=0}^{K-1}\psi(X_{kT/K}^{(i)})\Big] \\ &= \frac{1}{T^2}\lim_{K\to\infty}\frac{T^2}{K^2}\mathrm{Var}\Big[\sum_{k=0}^{K-1}\psi(X_{kT/K}^{(i)})\Big]. \end{aligned}$$

Here, we use the approximation of the Riemann sum in the second equality as the random function $\psi(X_s^{(i)})$ is Riemann integrable for any $\omega \in \Omega$ and an application of the dominated convergence theorem in the third equality.

Then, we have for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ that

$$\begin{aligned} I^{(i)} &= \frac{1}{T^2}\lim_{K\to\infty}\frac{T^2}{K^2}\mathrm{Var}\Big[\sum_{k=0}^{K-1}\psi(X_{kT/K}^{(i)})\Big] \\ &= \frac{1}{T^2}\lim_{K\to\infty}\frac{T^2}{K^2}\sum_{k=0}^{K-1}\sum_{l=0}^{K-1}\mathrm{Cov}\Big[\psi(X_{kT/K}^{(i)}),\psi(X_{lT/K}^{(i)})\Big] \\ &= \frac{1}{T^2}\int_0^T\int_0^T \mathrm{Cov}\Big[\psi(X_s^{(i)}),\psi(X_u^{(i)})\Big]dsdu. \end{aligned}$$

Here, we use Bienayme's identity in the second equality.

By Definition (23), we obtain for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ that

$$I^{(i)} \le \frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)}dsdu.$$

A split of the integral into two terms leads for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ to

$$\begin{aligned} I^{(i)} \le{}& \frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)}\mathbf{1}_{\{|s-u|\le\sqrt{T}\}}dsdu \\ &+\frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)}\mathbf{1}_{\{|s-u|>\sqrt{T}\}}dsdu. \end{aligned} \tag{32}$$

For any function $\psi \in C_b(E,\mathbb{R})$ and any index $i = 1,\dots,d$, there exists by Lemma 6 a positive real number $\mu_+^{(i)} > 0$ such that for any time $t \geq 0$ we have $\mu_t^{(i)} \leq \mu_+^{(i)}$.

Then, we obtain for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ that

$$\frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)} \mathbf{1}_{\{|s-u|\leq\sqrt{T}\}} dsdu \leq \frac{\mu_+^{(i)}}{T^2}\int_0^T\int_0^T \mathbf{1}_{\{|s-u|\leq\sqrt{T}\}} dsdu. \tag{33}$$

Moreover, we can deduce for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ that

$$\frac{\mu_+^{(i)}}{T^2}\int_0^T\int_0^T \mathbf{1}_{\{|s-u|\leq\sqrt{T}\}} dsdu \to 0. \tag{34}$$

By Expressions (33) and (34), we can deduce for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ that

$$\frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)} \mathbf{1}_{\{|s-u|\leq\sqrt{T}\}} dsdu \to 0. \tag{35}$$

We also have for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ that

$$\frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)} \mathbf{1}_{\{|s-u|>\sqrt{T}\}} dsdu \leq \sup_{y>\sqrt{T}} \mu_y^{(i)} \frac{1}{T^2}\int_0^T\int_0^T \mathbf{1}_{\{|s-u|>\sqrt{T}\}} dsdu. \tag{36}$$

Then, we obtain for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ that

$$\sup_{y>\sqrt{T}} \mu_y^{(i)} \frac{1}{T^2}\int_0^T\int_0^T \mathbf{1}_{\{|s-u|>\sqrt{T}\}} dsdu \leq \sup_{y>\sqrt{T}} \mu_y^{(i)}. \tag{37}$$

Since $\mu_T^{(i)} \to 0$ as the final time $T \to \infty$ by an application of Lemma 6, we can also deduce for any function $\psi \in C_b(E,\mathbb{R})$ and any index $i = 1,\dots,d$ as the final time $T \to \infty$ that

$$\sup_{y>\sqrt{T}} \mu_y^{(i)} \to 0. \tag{38}$$

Expressions (36), (37) and (38) imply for any function $\psi \in C_b(E,\mathbb{R})$ and any index $i = 1,\dots,d$ as the final time $T \to \infty$ that

$$\frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)} \mathbf{1}_{\{|s-u|>\sqrt{T}\}} dsdu \to 0. \tag{39}$$

Expressions (32), (35) and (39) yield for any function $\psi \in C_b(E,\mathbb{R})$ and any index $i = 1,\dots,d$ as the final time $T \to \infty$ that

$$I^{(i)} \to 0. \tag{40}$$

By the definitions of $II^{(i)}$ and $V^{(i)}$, we have for any function $\psi \in C_b(E,\mathbb{R})$, any index $i = 1,\dots,d$ and any final time $T > 0$ that

$$II^{(i)} = \Big(\mathbb{E}\Big[\frac{1}{T}\int_0^T \psi(X_s^{(i)})ds\Big] - \pi^{(i)}(\psi)\Big)^2.$$

By Fubini theorem with Lemmas 4 and 5, we obtain for any function $\psi \in C_b(E, \mathbb{R})$, any index $i = 1, \dots, d$ and any final time $T > 0$ that

$$II^{(i)} = \Big(\frac{1}{T}\int_0^T \mathbb{E}\Big[\psi(X_s^{(i)})\Big]ds - \pi^{(i)}(\psi)\Big)^2. \tag{41}$$

By Lemma 7, we have for any function $\psi \in C_b(E, \mathbb{R})$ and any index $i = 1, \dots, d$ as the final time $T \to \infty$ that

$$X_T^{(i)} \xrightarrow{\mathcal{D}} \big(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta)\big).$$

Since convergence in distribution implies convergence in expectation of any bounded function, we obtain for any function $\psi \in C_b(E, \mathbb{R})$ and any index $i = 1, \dots, d$ as the final time $T \to \infty$ that

$$\mathbb{E}[\psi(X_T^{(i)})] \to \mathbb{E}\big[\psi(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta))\big].$$

By the definition of $\pi^{(i)}(\psi)$, we can deduce for any function $\psi \in C_b(E, \mathbb{R})$ and any index $i = 1, \dots, d$ as the final time $T \to \infty$ that

$$\mathbb{E}[\psi(X_T^{(i)})] \to \pi^{(i)}(\psi). \tag{42}$$

Moreover, Expressions (41) and (42) imply for any function $\psi \in C_b(E, \mathbb{R})$ and any index $i = 1, \dots, d$ as the final time $T \to \infty$ that

$$II^{(i)} \to 0. \tag{43}$$

Finally, we can deduce by Expressions (31), (40) and (43) for any function $\psi \in C_b(E, \mathbb{R})$ and any index $i = 1, \dots, d$ as the final time $T \to \infty$ that

$$\mathbb{E}\big[(V^{(i)}(\psi) - \pi^{(i)}(\psi))^2\big].$$

□

We define the limit of the normalized deviation between the log likelihood at the parameter value $\theta \in \Theta$ and the log likelihood at the true parameter value θ^* as

$$Y(\theta) = \sum_{i=1}^d \int_E \Big(\log\Big(\frac{v}{u}\Big)u - (v-u)\Big)\mathbf{1}_{\{u>0,v>0\}}\pi_\theta^{(i)}(du, dv, dw). \tag{44}$$

As the true parameter value θ^* is a maximum argument of the log likelihood limit, we have by definition that $Y(\theta) \leq 0$ for any parameter $\theta \in \Theta$.

We now show the following lemma, which corresponds to Assumption 1 (e) in Potiron (2025b). This complements Lemma A.7 (p. 1836) in Clinet and Yoshida (2017). This extends Lemma C6 in Supplement C of Potiron and Volkov (2025) and Lemma 17 in Potiron (2025b) to the nonlinear intensity case.

LEMMA 9. *We assume that Assumptions 1 (a), (b), (c), (d), (e) (f), (h), (i), (j), (k), (m) and (l) hold. Then, we have for any parameter $\theta \in \overline{\Theta} - \theta^*$ that $Y(\theta) \neq 0$.*

PROOF OF LEMMA 9. We assume that the parameter $\theta \in \overline{\Theta}$ and that $Y(\theta) = 0$. By Definition (44), we can deduce that

$$0 = \sum_{i=1}^d \int_E \Big(\log\Big(\frac{v}{u}\Big)u - (v-u)\Big)\pi_{\theta^*}^{(i)}(du, dv, dw).$$

By Lemma 7, this can be reexpressed as

$$0 = \sum_{i=1}^{d} \mathbb{E}\Big[\log\Big(\frac{\lambda_l^{(i)}(\theta)}{\lambda_l^{(i)}(\theta^*)}\Big) \lambda_l^{(i)}(\theta^*) - \big(\lambda_l^{(i)}(\theta) - \lambda_l^{(i)}(\theta^*)\big)\Big]. \tag{45}$$

We also have by definition for any index $i = 1, \dots, d$ and any $\omega \in \Omega$ that

$$0 \geq \log\Big(\frac{\lambda_l^{(i)}(\theta)}{\lambda_l^{(i)}(\theta^*)}\Big) \lambda_l^{(i)}(\theta^*) - \big(\lambda_l^{(i)}(\theta) - \lambda_l^{(i)}(\theta^*)\big). \tag{46}$$

Moreover, Expressions (45) and (46) yield a.s.

$$0 = \log\Big(\frac{\lambda_l^{(i)}(\theta)}{\lambda_l^{(i)}(\theta^*)}\Big) \lambda_l^{(i)}(\theta^*) - \big(\lambda_l^{(i)}(\theta) - \lambda_l^{(i)}(\theta^*)\big).$$

Thus, we can deduce that a.s.

$$\lambda_l(\theta^*) = \lambda_l(\theta).$$

Finally, we obtain $\theta^* = \theta$ by Assumption 1 (l). □

We denote by $L^1(\mu)$ the space of functions that are integrable with respect to μ for any measure μ. The functions that we will be using in the proofs which follow will not necessarily be bounded. Thus, we extend from the space $C_b(E, \mathbb{R})$ to the space $C_\uparrow(E, \mathbb{R})$ the space of functions in which the ergodicity assumption holds. We also give a more explicit form to the function $\pi(\psi)$. The following lemma is Proposition 3.8 (pp. 1806-1807) in Clinet and Yoshida (2017). The proof follows the arguments from the proof of Proposition 3.8 (pp. 1822-1824) in Clinet and Yoshida (2017). See also Lemma 1 in Potiron (2025b).

LEMMA 10. *We assume that Assumptions 1 (a), (b), (c), (d), (e), (f), (g), (h), (i), (j) and (k) hold. Then, we have for any parameter $\theta \in \Theta$ that*

(a) *The ergodicity assumption in the sense of Definition 2 still holds for any function $\psi \in C_\uparrow(E, \mathbb{R})$. In particular, the function $\pi^{(i)}(\psi)$ can be extended to $C_\uparrow(E, \mathbb{R})$ for any index $i = 1, \dots, d$. Moreover, the convergence is uniform in the parameter $\theta \in \Theta$ for any function $\psi \in C_\uparrow(E, \mathbb{R})$.*
(b) *For any index $i = 1, \dots, d$, there exists a probability measure $\pi_\theta^{(i)}$ on the space E which satisfies what follows. More specifically, we have $\pi^{(i)}(\psi) = \int_E \psi(u, v, w) \pi_\theta^{(i)}(du, dv, dw)$ for any function $\psi \in C_\uparrow(E, \mathbb{R})$. In particular, we have $C_\uparrow(E, \mathbb{R}) \subset L^1(\pi_\theta^{(i)})$.*

PROOF OF LEMMA 10. We can use the arguments from the proof of Proposition 3.8 (pp. 1822-1824) in Clinet and Yoshida (2017). □

We introduce now the following lemma, which ensures that the family of intensities does not explode on any compact space. Its proof is a direct consequence to Lemma 10. This corresponds to Lemma 2 in Potiron (2025b).

LEMMA 11. *We assume that Assumptions 1 (a), (b), (c), (d), (e), (f), (g), (h), (i), (j) and (k) hold. Then, the intensity process $\lambda_t(\theta)$ is a.s. locally integrable on the space $\mathbb{R}^+$ for any parameter $\theta \in \Theta$.*

PROOF OF LEMMA 11. This is a direct consequence to Lemma 10 with the function $\psi(u, v, w) = v$ for any time $(u, v, w) \in E$. □

We now show the following lemma, which corresponds to Assumption 1 (f) in Potiron (2025b). This extends Lemma 11 in Potiron (2025b) to the nonlinear intensity case.

LEMMA 12. *We assume that Assumptions 1 (a), (b), (c), (d), (e), (f), (g), (h), (i), (j), (k) and (m) hold. Then, we have a.s. that the intensity $\theta \to \lambda_t(\theta)$ is twice continuously differentiable from the parameter space Θ to the space $\mathbb{R}_+^d$ for any time $t \in \mathbb{R}^+$ a.e. Moreover, there exists a continuous extension to $\overline{\Theta}$.*

PROOF OF LEMMA 12. By Definition (8), we have for any time $t \in \mathbb{R}^+$ and any parameter $\theta \in \Theta$ that

$$\lambda_t(\theta) = f\Big(\nu, \int_0^t h(t-s,\kappa)\, dN_s\Big).$$

First, the nonlinear function $f : \Theta_\nu \times \mathbb{R}^+ \to \mathbb{R}_+^{d\times d}$ is twice continuously differentiable and there exists a continuous extension to $\overline{\Theta}_\nu \times \mathbb{R}^+$ by Assumption 1 (h). By function composition properties, it remains to show the lemma for any time $t \in \mathbb{R}^+$ a.e. and any kernel parameter $\kappa \in \Theta_\kappa$ with

$$\lambda_{t,h}(\kappa) = \int_0^t h(t-s,\kappa)\, dN_s.$$

The intensity can be rewritten for any index $i = 1, \dots, d$, any time $t \in \mathbb{R}^+$ and any kernel parameter $\kappa \in \Theta_\kappa$ as

$$\lambda_{t,h}^{(i)}(\kappa) = \sum_{j=1}^d \int_0^t h^{(i,j)}(t-s,\kappa)\, dN_s^{(j)}.$$

By the assumption on the times of the point process (2), the intensity $\lambda_{t,h}^{(i)}(\kappa)$ can be reexpressed for any index $i = 1, \dots, d$, any time $t \in \mathbb{R}^+$, any kernel parameter $\kappa \in \Theta_\kappa$ and a.s. as

$$\lambda_{t,h}^{(i)}(\kappa) = \sum_{j=1}^d \sum_{k\in\mathbb{N}_* \text{ s.t. } 0<T_k^{(j)}<t} h^{(i,j)}\big(t - T_k^{(j)}, \kappa\big).$$

By Lemma 11 and compacity of the segment $[0,T]$, we have that the number of terms in the sum and each term are a.s. finite. Then, we can deduce that $\lambda_{t,h}^{(i)}(\kappa)$ is a.s. finite. As the kernel is differentiable a.e. by Assumption 1 (m), we can deduce that $\lambda_{t,h}^{(i)}(\kappa)$ is a.s. differentiable and

$$\partial_\kappa \lambda_{t,h}^{(i)}(\kappa) = \sum_{j=1}^d \sum_{k\in\mathbb{N}_* \text{ s.t. } 0<T_k^{(j)}<t} \partial_\kappa h^{(i,j)}\big(t - T_k^{(j)}, \kappa\big).$$

As the kernel is differentiable twice a.e. by Assumption 1 (m), we can deduce that $\lambda_{t,h}^{(i)}(\kappa)$ is a.s. differentiable and

$$\partial_\kappa^2 \lambda_{t,h}^{(i)}(\kappa) = \sum_{j=1}^d \sum_{k\in\mathbb{N}_* \text{ s.t. } 0<T_k^{(j)}<t} \partial_\kappa^2 h^{(i,j)}(t - T_k^{(j)}, \kappa).$$

Thus, we have a.s. that the intensity process $\theta \to \lambda_s(\theta)$ for any time $s \in \mathbb{R}^+$ a.e. is twice continuously differentiable from the parameter space Θ to the space $\mathbb{R}_+^d$, and there exists a continuous extension to $\overline{\Theta}$. $\square$

We now show the following lemma, which corresponds to Assumption 1 (g) in Potiron (2025b). This extends Lemma 12 in Potiron (2025b) to the nonlinear intensity case.

LEMMA 13. *We assume that Assumptions 1 (a), (b), (c), (d), (e), (f), (g), (h), (i), (j), (k) and (m) hold. Then, we have* $\mathbb{P}(\int_0^T |\partial_\theta \lambda_t(\theta)|dt < \infty) = 1$ *and* $\mathbb{P}(\int_0^T |\partial_\theta^2 \lambda_t(\theta)|dt < \infty) = 1$ *for any parameter* $\theta \in \Theta$ *and any final time* $T > 0$.

PROOF OF LEMMA 13. First, we have by Definition (8) for any parameter $\theta \in \Theta$ and any time $t > 0$ that

$$\lambda_t(\theta) = f\Big(\nu, \int_0^t h(t-s,\kappa)\, dN_s\Big).$$

Then, the ith component of the intensity process $\lambda_t(\theta)$ can be rewritten for any parameter $\theta \in \Theta$, any time $t > 0$ and any index $i = 1, \dots, d$ as

$$\lambda_t^{(i)}(\theta) = f^{(i)}\Big(\nu, \sum_{j=1}^{d} \int_0^t h^{(i,j)}(t-s,\kappa)\, dN_s^{(j)}\Big).$$

By the assumption on the times of the point process (2), the ith component of the intensity process $\lambda_t(\theta)$ can be reexpressed for any parameter $\theta \in \Theta$, any time $t > 0$ and any index $i = 1, \dots, d$ as

$$\lambda_t^{(i)}(\theta) = f^{(i)}\Big(\nu, \sum_{j=1}^{d} \sum_{k \in \mathbb{N}_*} \mathbf{1}_{\{0 < T_k^{(j)} < t\}} h^{(i,j)}\big(t - T_k^{(j)}, \kappa\big)\Big).$$

By Lemma 11 and compacity of the segment $[0,T]$, we have that the number of terms in the sum and each term are a.s. finite for any parameter $\theta \in \Theta$, any time $t > 0$ and any index $i = 1, \dots, d$. Then, we can deduce that $\lambda_t^{(i)}(\theta)$ is a.s. finite for any parameter $\theta \in \Theta$, any time $t > 0$ and any index $i = 1, \dots, d$. As the kernel is differentiable a.e. by Assumption 1 (m), we can deduce that $\lambda_t^{(i)}(\theta)$ is a.s. differentiable for any parameter $\theta \in \Theta$, any time $t > 0$ and any index $i = 1, \dots, d$ such that

$$\partial_\theta \lambda_t^{(i)}(\theta) = \Big(1, \sum_{j=1}^{d} \sum_{k \in \mathbb{N}_*} \mathbf{1}_{\{0 < T_k^{(j)} < t\}} \partial_\theta h^{(i,j)}\big(t - T_k^{(j)}, \kappa\big)\Big)$$
$$* df^{(i)}\Big(\nu, \sum_{j=1}^{d} \sum_{k \in \mathbb{N}_*} \mathbf{1}_{\{0 < T_k^{(j)} < t\}} h^{(i,j)}\big(t - T_k^{(j)}, \kappa\big)\Big).$$

By the assumption on the times of the point process (2), the intensity derivative $\partial_\theta \lambda_{t,h}^{(i)}(\theta)$ can be reexpressed for any parameter $\theta \in \Theta$, any time $t > 0$ and any index $i = 1, \dots, d$ as

$$\partial_\theta \lambda_t^{(i)}(\theta) = \Big(1, \sum_{j=1}^{d} \int_0^t \partial_\theta h^{(i,j)}(t-s,\kappa) dN_s^{(j)}\Big) df^{(i)}\Big(\nu, \sum_{j=1}^{d} \int_0^t h^{(i,j)}(t-s,\kappa) dN_s^{(j)}\Big).$$

This can be rewritten in multidimensional notation for any parameter $\theta \in \Theta$ and any time $t > 0$ as

$$\partial_\theta \lambda_t(\theta) = \Big(1, \int_0^t \partial_\theta h(t-s,\kappa) dN_s\Big) df\Big(\nu, \int_0^t h(t-s,\kappa) dN_s\Big). \tag{47}$$

In addition, we obtain by Assumption 1 (e) for any parameter $\theta \in \Theta$ and any time $t > 0$ that

$$\left|\partial_\theta \lambda_t(\theta)\right| \leq f_+ \left|1 + \int_0^t \partial_\theta h(t-s,\kappa) dN_s\right|.$$

Moreover, we can deduce by the triangular inequality for any parameter $\theta \in \Theta$ and any time $t > 0$ that

$$\left|\partial_\theta \lambda_t(\theta)\right| \leq f_+ \Big(1 + \Big|\int_0^t \partial_\theta h(t-s,\kappa) dN_s\Big|\Big).$$

Furthermore, the triangular inequality with the assumption on the times of the point process (2) yield for any parameter $\theta \in \Theta$ and any time $t > 0$ that

$$\left|\partial_\theta \lambda_t(\theta)\right| \leq f_+ \Big(1 + \int_0^t \left|\partial_\theta h(t-s,\kappa)\right| dN_s\Big).$$

Finally, we obtain $\mathbb{P}(\int_0^T |\partial_\theta \lambda_t(\theta)| dt < \infty) = 1$ for any parameter $\theta \in \Theta$ and any time $t > 0$ by an extension of Lemma 4 with Assumptions 1 (i), (j) and (k).

As the kernel is differentiable twice a.e. by Assumption 1 (m), we can deduce that the intensity process $\lambda_t^{(i)}(\theta)$ is a.s. differentiable twice for any parameter $\theta \in \Theta$, any time $t > 0$ and any index $i = 1, \dots, d$. Then, we obtain by Equation (47) for any parameter $\theta \in \Theta$ and any time $t > 0$ that

$$\partial_\theta^2 \lambda_t(\theta) = \partial_\theta \Big\{ \Big(1, \int_0^t \partial_\theta h(t-s,\kappa) dN_s\Big) df\Big(\nu, \int_0^t h(t-s,\kappa) dN_s\Big)\Big\}.$$

In addition, we obtain by Assumptions 1 (e) and (i) for any parameter $\theta \in \Theta$ and any time $t > 0$ that

$$\left|\partial_\theta^2 \lambda_t(\theta)\right| \leq C \left|1 + \int_0^t \partial_\theta h(t-s,\kappa) dN_s + \int_0^t \partial_\theta^2 h(t-s,\kappa) dN_s\right|.$$

Moreover, we can deduce by the triangular inequality for any parameter $\theta \in \Theta$ and any time $t > 0$ that

$$\left|\partial_\theta^2 \lambda_t(\theta)\right| \leq C \Big(1 + \Big|\int_0^t \partial_\theta h(t-s,\kappa) dN_s\Big| + \Big|\int_0^t \partial_\theta^2 h(t-s,\kappa) dN_s\Big|\Big).$$

Furthermore, the triangular inequality with the assumption on the times of the point process (2) yield for any parameter $\theta \in \Theta$ and any time $t > 0$ that

$$\left|\partial_\theta \lambda_t(\theta)\right| = C \Big(1 + \int_0^t \left|\partial_\theta h(t-s,\kappa)\right| dN_s + \int_0^t \left|\partial_\theta^2 h(t-s,\kappa)\right| dN_s\Big).$$

Finally, we obtain $\mathbb{P}(\int_0^T |\partial_\theta^2 \lambda_t(\theta)| dt < \infty) = 1$ for any parameter $\theta \in \Theta$ and any time $t > 0$ by an extension of Lemma 4 with Assumptions 1 (i), (j) and (k). □

We now give the proof of Theorem 1. This is based on an application of Theorem 1 in Potiron (2025b) with the previous lemmas.

PROOF OF THEOREM 1. The proof is an application of Theorem 1 in Potiron (2025b) with Lemmas 1, 4 , 5, 8, 9, 12 and 13. □

Funding. The author was supported in part by Japanese Society for the Promotion of Science Grants-in-Aid for Scientific Research (B) 23H00807.

REFERENCES

BACRY, E., DELATTRE, S., HOFFMANN, M. and MUZY, J.-F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications* **123** 2475–2499.

BLANC, P., DONIER, J. and BOUCHAUD, J.-P. (2017). Quadratic Hawkes processes for financial prices. *Quantitative Finance* **17** 171–188.

BOWSHER, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics* **141** 876–912.

BRÉMAUD, P. and MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Annals of Probability* 1563–1588.

CAI, B., ZHANG, J. and GUAN, Y. (2024). Latent network structure learning from high-dimensional multivariate point processes. *Journal of the American Statistical Association* **119** 95–108.

CAVALIERE, G., LU, Y., RAHBEK, A. and STÆRK-ØSTERGAARD, J. (2023). Bootstrap inference for Hawkes and general point processes. *Journal of Econometrics* **235** 133–165.

CHAVEZ-DEMOULIN, V., DAVISON, A. and MCNEIL, A. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance* **5** 227–234.

CHEN, F. and HALL, P. (2013). Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. *Journal of Applied Probability* **50** 1006–1024.

CHEYSSON, F. and LANG, G. (2022). Spectral estimation of Hawkes processes from count data. *Annals of Statistics* **50** 1722–1746.

CLINET, S. and POTIRON, Y. (2018). Statistical inference for the doubly stochastic self-exciting process. *Bernoulli* **24** 3469–3493.

CLINET, S. and YOSHIDA, N. (2017). Statistical inference for ergodic point processes and application to limit order book. *Stochastic Processes and their Applications* **127** 1800–1839.

COSTA, M., GRAHAM, C., MARSALLE, L. and TRAN, V. C. (2020). Renewal in Hawkes processes with self-excitation and inhibition. *Advances in Applied Probability* **52** 879–915.

DALEY, D. J. and VERE-JONES, D. (2003). *An introduction to the theory of point processes: Elementary theory and methods* **1**, 2nd ed. Springer New York, NY.

DALEY, D. J. and VERE-JONES, D. (2008). *An introduction to the theory of point processes: General theory and structure* **2**, 2nd ed. Springer New York, NY.

DONNET, S., RIVOIRARD, V. and ROUSSEAU, J. (2020). Nonparametric Bayesian estimation for multivariate Hawkes processes. *Annals of Statistics* **48** 2698–2727.

EMBRECHTS, P., LINIGER, T. and LIN, L. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability* **48** 367–378.

ERDEMLIOGLU, D., POTIRON, Y., VOLKOV, V. and XU, T. (2025). Estimation of time-dependent latency with locally stationary Hawkes processes. *Working paper available at https://www.fbc.keio.ac.jp/∼* $potiron/Erdemlioglu2025workingpaper.pdf$.

FANG, G., XU, G., XU, H., ZHU, X. and GUAN, Y. (2024). Group network Hawkes process. *Journal of the American Statistical Association* **119** 2328–2344.

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. *Journal of the American Statistical Association* **111** 564–584.

GRAHAM, C. (2021). Regenerative properties of the linear Hawkes process with unbounded memory. *Annals of Applied Probability* **31** 2844–2863.

HAWKES, A. G. (1971a). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90.

HAWKES, A. G. (1971b). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)* 438–443.

HAWKES, A. G. and OAKES, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability* **11** 493–503.

HORST, U. and XU, W. (2021). Functional limit theorems for marked Hawkes point measures. *Stochastic Processes and their Applications* **134** 94–131.

HORST, U. and XU, W. (2022). The microstructure of stochastic volatility models with self-exciting jump dynamics. *Annals of Applied Probability* **32** 4568–4610.

HORST, U. and XU, W. (2024). Functional limit theorems for Hawkes processes. *To appear in Probability Theory and Related Fields*.

IKEFUJI, M., LAEVEN, R., MAGNUS, J. and YUE, Y. (2022). Earthquake risk embedded in property prices: Evidence from five Japanese cities. *Journal of the American Statistical Association* **117** 82–93.

JACOD, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **31** 235–253.

JACOD, J. and SHIRYAEV, A. (2003). *Limit theorems for stochastic processes*, 2nd ed. Springer Berlin, Heidelberg.

JAISSON, T. and ROSENBAUM, M. (2015). Limit theorems for nearly unstable Hawkes processes. *Annals of Applied Probability* **25** 600–631.

JAISSON, T. and ROSENBAUM, M. (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *Annals of Applied Probability* **26** 2860–2882.

KARIM, R. S., LAEVEN, R. J. and MANDJES, M. (2025). Compound multivariate Hawkes processes: Large deviations and rare event simulation. *Bernoulli* **31** 3113–3138.

KWAN, T. (2023). Asymptotic analysis and ergodicity of the Hawkes process and its extensions, PhD thesis, UNSW Sydney.

KWAN, T., CHEN, F. and DUNSMUIR, W. (2023). Alternative asymptotic inference theory for a nonstationary Hawkes process. *Journal of Statistical Planning and Inference* **227** 75–90.

KWAN, T.-K. J., CHEN, F. and DUNSMUIR, W. T. (2024). Ergodic properties of the Hawkes process with a general excitation kernel. *Journal of Applied Probability* 1–24.

LARGE, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets* **10** 1-25.

MAMMEN, E. and MÜLLER, M. (2023). Nonparametric estimation of locally stationary Hawkes processes. *Bernoulli* **29** 2062–2083.

OGATA, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics* **30** 243–261.

OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association* **83** 9–27.

OZAKI, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics* **31** 145–155.

POTIRON, Y. (2025a). Nonparametric inference for Hawkes processes with a stochastic time-dependent baseline. *Working paper available at https://www.fbc.keio.ac.jp/$\sim$ potiron/$Potiron2025nonparametricworkingpaper.pdf$*.

POTIRON, Y. (2025b). Parametric inference for Hawkes processes with a general kernel. *Revise and Resubmit for the Annals of the Institute of Statistical Mathematics*.

POTIRON, Y. and VOLKOV, V. (2025). Mutually exciting point processes with latency. *To appear in Journal of the American Statistical Association*.

POTIRON, Y., SCAILLET, O., VOLKOV, V. and YU, S. (2025a). High-frequency estimation of Itô semimartingale baseline for Hawkes processes. *Working paper available at https://www.fbc.keio.ac.jp/$\sim$ potiron/$Potiron2025workingpaper.pdf$*.

POTIRON, Y., SCAILLET, O., VOLKOV, V. and YU, S. (2025b). Estimation of branching ratio for Hawkes processes with Itô semimartingale baseline. *Working paper available at https://www.fbc.keio.ac.jp/$\sim$ potiron/$Potiron2025estimationworkingpaper.pdf$*.

REYNAUD-BOURET, P. and ROY, E. (2007). Some non asymptotic tail estimates for Hawkes processes. *Bulletin Belgium Mathematical Society Simon Stevin* **13** 883–896.

REYNAUD-BOURET, P. and SCHBATH, S. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. *Annals of Statistics* **38** 2781–2822.

ROUEFF, F., VON SACHS, R. and SANSONNET, L. (2016). Locally stationary Hawkes processes. *Stochastic Processes and their Applications* **126** 1710–1743.

ROUEFF, F. and VON SACHS, R. (2019). Time-frequency analysis of locally stationary Hawkes processes. *Bernoulli* **25** 1355–1385.

RUBIN, I. (1972). Regular point processes and their detection. *IEEE Transactions on Information Theory* **18** 547–557.

VERE-JONES, D. and OZAKI, T. (1982). Some examples of statistical estimation applied to earthquake data: I. Cyclic Poisson and self-exciting models. *Annals of the Institute of Statistical Mathematics* **34** 189–207.

XU, W. (2024). Diffusion approximations for self-excited systems with applications to general branching processes. *Annals of Applied Probability* **34** 2650–2713.

ZHU, L. (2013). Central limit theorem for nonlinear Hawkes processes. *Journal of Applied Probability* **50** 760–771.

ZHU, L. (2015). Large deviations for Markovian nonlinear Hawkes processes. *Annals of Applied Probability* **25** 548–581.