Central limit theorem for Hawkes processes with a stochastic and time dependent baseline

Yoann Potiron[1],

[1] *Faculty of Business and Commerce, Keio University, e-mail:* potiron@fbc.keio.ac.jp

Abstract: We develop statistical inference for Hawkes processes with a stochastic and time dependent baseline. The statistical inference procedure is based on the average of the point process. We consider estimation for the average over time of the intensity process. We first show the existence of these point processes. We also show the central limit theorem of the statistical inference procedure. This requires the assumption that the kernel does not have a too fat tail. We also need that the baseline process is Lipschitz continuous with bounded starting value. The main novelty in the proofs is to establish the renewal equation for stochastic processes.

Keywords and phrases: Hawkes mutually exciting processes, central limit theorem, time dependent baseline, stochastic baseline, statistical inference, renewal equation, existence.

1. Introduction

This paper concerns central limit theorem for point processes. The main stylized fact in this strand of literature is the presence of event clustering in time. This motivates to rely on 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 number of events from the starting time 0 to the time t and λ_t its intensity at the time t. A standard definition of Hawkes mutually exciting processes is given by

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

Here, ν is a Poisson baseline of dimension d and h is a kernel 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. 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) and Ogata (1978), Ogata (1988)). 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), 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 (2026)). We can also find some applications in biology (see Reynaud-Bouret and Schbath (2010) and Donnet, Rivoirard and Rousseau (2020)) and epidemiology (see Cheysson and Lang (2022)). See also Liniger (2009) and Hawkes (2018) with the references therein.

There are many theoretical results for Hawkes processes in statistics and applied probability. 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) gives central limit theorem for nonlinear Hawkes processes. Zhu (2015) considers large deviations for Markovian nonlinear Hawkes processes. 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. Potiron and Volkov (2026), Potiron (2025a), Erdemlioglu et al. (2025a), Potiron (2025b) and Potiron (2026) consider parametric inference.

Empirical evidence with financial applications suggests that the baseline is time dependent and stochastic during intraday trading activity. Chen and Hall (2013) show in their empirical study (Section 5.2, pp. 7–10) that goodness-of-fit results are in favor of their time dependent baseline model over a set of alternatives. In Figure 2 (p. 20), they document the time dependent and nonrandom function for both polynomial and exponential kernel. This nonrandom path is also visible in Figure 2 (p. 3488) from Clinet and Potiron (2018). Moreover, the empirical findings in the two aforementioned papers suggest that there is a remaining stochastic effect. Finally, Rambaldi, Pennesi and Lillo (2015) and Rambaldi, Filimonov and Lillo (2018) document that there are frequent intensity bursts in the baseline. In the former paper, the authors focus on a short period around the macroeconomic news announcement and are able to capture the increase in trading activity after the news, both when the news has a significant effect on volatility and when this effect is negligible. In the latter paper, the authors develop a general method to detect intensity bursts and demonstrate empirically that only a relatively small proportion of these bursts can be related to news. The intensity remains locally bounded in this case. Finally, Christensen and Kolokolov (2024) propose a framework for intensity burst detection in which the time dependent and stochastic intensity is potentially unbounded over short time intervals.

There are also theoretical results in statistics of Hawkes processes with a baseline which is time dependent and possibly random. Chen and Hall (2013), Roueff, von Sachs and Sansonnet (2016), Clinet and Potiron (2018), Roueff and Von Sachs (2019), Kwan, Chen and Dunsmuir (2023), Mammen and Müller (2023) and Erdemlioglu et al. (2025b) study locally stationary Hawkes processes. Potiron et al. (2025a) and Potiron et al. (2025b) introduce Hawkes processes with Itô semimartingale baseline.

In this paper, we consider Hawkes processes in which the kernel has a general form. We introduce a baseline ν which is stochastic and time dependent. More specifically, the intensity λ_t of the point process N_t for any time $t \in [0, T]$ follows

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

We consider estimation for the rescaled integral of the intensity process λ_t between the starting time 0 and the final time tT for any time $t \in [0, 1]$, namely

$$\Lambda_{t,T} = \frac{1}{T} \int_0^{tT} \lambda_s ds. \tag{3}$$

We have applications in finance of the target quantity (3) where the intensity of a quote plays an inverse role to the volatility of an asset price. Quote intensity is a central measure in financial economics. Quotes are driven by an underlying stochastic process, which can be interpreted as the flow of news in the mixture of distribution hypothesis (see Clark (1973), Epps and Epps (1976) and Tauchen and Pitts (1983)). This implies that the quote intensity is random and persistent whenever the unobservable arrival of news is random and persistent. Some market microstructure theories suggest that optimal execution strategies are based on time-varying quote intensity (see Kyle (1985), Admati and Pfleiderer (1988) and Almgren and Chriss (2001)). Moreover, the increased volatility observed during distressed market conditions often coincides with abnormal increases in quote intensity.

There are two contributions in this paper. First, we give an existence result in Proposition 1. This complements Theorem 5.1 (p. 3476) from Clinet and Potiron (2018) and Proposition 1 in Erdemlioglu et al. (2025b). Compared to these two papers which are restricted to exponentially decreasing kernels, our framework allows for more general kernels. This is useful for applications as there is empirical evidence in finance that the kernel decays as the power distribution (see Bacry, Dayri and Muzy (2012) and Hardiman, Bercot and Bouchaud (2013)). The arguments used in the proofs are based on the arguments from Brémaud and Massoulié (1996) and Clinet and Potiron (2018). Since we make the classical assumption that the spectral radius of the kernel matrix is smaller than 1, the proof of existence is close from the proof of existence with an exponential kernel.

Second, our main contribution is the central limit theorem of the statistical inference procedure for Hawkes processes with a stochastic and time dependent baseline in Theorem 1. This extends Corollary 1 (p. 2481) from Bacry et al. (2013) which is restricted to nonrandom constant baseline and Corollary 1 from Deschatre, Gruet and Lotz (2025) which is restricted to nonrandom and time dependent baseline. In particular, the functional form and the weak convergence provided in Theorem 1 makes the obtained result very general. Finally, we obtain the stability of the convergence which is useful for statistical applications.

We obtain that the limiting process of the target quantity is stochastic. More specifically, we denote the limiting process of the target quantity $\Lambda_{t,T}$ for any time $t \in [0,1]$ as the final time $T \to \infty$ by

$$\Lambda_t = (I - \|h\|_1)^{-1} \int_0^t \nu_s ds. \tag{4}$$

Here, we define the L^1 norm matrix for the kernel function h of dimension $d \times d$ as

$$\|h\|_1 = \int_0^\infty h(s) ds.$$

The L^1 norm of the kernel $\|h\|_1$ is often called branching ratio. The interpretation of this branching ratio is closely related to the Poisson cluster representation for Hawkes processes (see Hawkes and Oakes (1974)). In terms of population interpretation, this branching ratio is the average number of children of an individual. From a financial perspective, we can interpret the branching ratio as the average proportion of endogenous events. It is an empirical measure of the degree of endogeneity in the market, namely quotes generated by the current participants instead of new participants to the market.

The asymptotic variance becomes stochastic in our central limit theorem. This can be explained by the fact that the limiting process of the target quantity is stochastic. This is classical to obtain stochastic asymptotic variance based on in-fill asymptotics when the final time T is fixed and the number of observations on the interval $[0, T]$ increases as $n \to \infty$. This approach is motivated by financial applications based on high-frequency data (see Aït-Sahalia and Jacod (2014)). Boosting the intensity, namely the use of in-fill asymptotics, boils down to assuming that the intensity is directly proportional to the inverse of the time interval. In other words, the probability of having a quote is proportional to the inverse of the time interval. This seems natural when we consider mutually exciting processes with a stochastic and time dependent intensity applied to quote events. Financial markets reveal submitting many quotes in short time intervals due to a high baseline amplified by the mutual excitation and phases with a lower baseline with fewer quotes in short time intervals. These phases create randomness, which explains why we look at a stochastic baseline and not at a purely deterministic time dependent baseline.

At first glance, the asymptotics from this paper as the final time $T \to \infty$ are limited for financial applications based on high-frequency data. However, we can consider a time transformation to rewrite our problem with the final time $T \to \infty$ as a problem with in-fill asymptotics when the final time T is fixed and the number of observations on the interval $[0, T]$ increases as $n \to \infty$. This time transformation was already used in Clinet and Potiron (2018), Kwan, Chen and Dunsmuir (2023), Potiron and Volkov (2026), Erdemlioglu et al. (2025b) and Erdemlioglu et al. (2025a). Then, there is no further technical issues in the proofs after introducing the time transformation. Thus, the theoretical results of this paper easily extend to the in-fill asymptotics case.

In addition, the theoretical results are limited for financial applications based on high-frequency data. This is mainly due to the fact that the limit process Λ_t

in Definition (4) depends on both the integral of the baseline $\int_0^t \nu_s ds$ and the branching ratio $\|h\|_1$. Thus, we cannot identify $\int_0^t \nu_s ds$ in the limiting process as the branching ratio $\|h\|_1$ is unknown to the statistician. However, Potiron et al. (2025a) introduce Hawkes processes with an Itô semimartingale baseline to solve the problem. The reason why the authors rely on an Itô semimartingale rather than a more general process is due to the presence of the branching ratio $\|h\|_1$ in the limit. Namely, estimation for the integral of the baseline is based on estimation for the integral of the baseline volatility. Then, the Itô semimartingale baseline is the most general model where estimation for the integral of the baseline volatility is tractable. Furthermore, the authors obtain five main applications from the theory. First, they derive estimation of the integrated intensity, estimation for the integral of the baseline and estimation for the integral of the baseline volatility. They also provide an hypothesis testing for the absence of a Hawkes component and an hypothesis testing for constant baseline. Moreover, an empirical application on high-frequency data of the E-mini S&P500 futures contracts finds rejection of both the null hypothesis of no Hawkes excitation and that of constant baseline contributing to the formal identification of patterns in high-frequency trading activity. Moreover, Potiron et al. (2025b) develop estimation of the branching ratio $\|h\|_1$ as another application. They also consider an hypothesis testing for a branching ratio equal to a given value. Finally, they also propose a test for a branching ratio value against a branching ratio tending to unity as the number of observations increases.

The proofs of the central limit theorem rely on the use of the triangular array of martingale increments. More specifically, the proofs of Theorem 1 are based on an application of Theorem IX.7.28 (pp. 590-591) in Jacod and Shiryaev (2003). Then, the main novelty in the proofs is to establish the renewal equation for stochastic processes (see Lemma 3). This is required as the baseline ν_t is stochastic and time dependent in our particular case. The renewal equation for stochastic processes allows us to exhibit a martingale representation of the intensity λ_{tT} (see Lemma 4). Thus, establishing the renewal equation for stochastic processes is the most important step in the proofs. The renewal equation for stochastic processes is also used to prove Theorem 1 in Potiron et al. (2025a) who consider the case of Hawkes processes with an Itô semimartingale baseline. In that particular case, the proof of Theorem 1 is based on the same strategy. The establishment of the renewal equation for stochastic processes extends Bacry et al. (2013) and Deschatre, Gruet and Lotz (2025) in which the function from the renewal equation is nonrandom. Jaisson and Rosenbaum (2015) also uses the renewal equation with a different asymptotics. More generally, renewal techniques for Hawkes processes are studied in Costa et al. (2020) and Graham (2021).

We discuss about assumptions required for the central limit theorem in what follows. We first put some restrictions on the kernel shape h. This corresponds exactly to Assumption (A2) in Bacry et al. (2013) (p. 2480). This is required to obtain a martingale form of the intensity process. We also put restrictions on the baseline process ν. More specifically, we rely on a starting point which is bounded almost surely and consider a baseline process ν which is Lipschitz-continuous almost surely. These assumptions are mainly used to obtain a locally bounded baseline process ν. This is needed to exhibit a martingale representation of the intensity in the proofs (see Lemma 4). This relies on establishing the renewal equation. Thus, these assumptions can be relaxed as long as we get a locally bounded baseline process ν. These assumptions are reasonably general assumptions. In particular, the Brownian motion is a particular case allowed by our assumptions. Moreover, a continuous Ito semimartingale with finite drift moments and finite volatility moments also satisfy our assumptions. Finally, the case of an Ito semimartingale baseline which is not necessarily continuous is studied in Potiron et al. (2025a). Finally, our assumptions on the baseline process ν can also be compared to the assumptions from some related work on Hawkes processes with a stochastic and time dependent baseline (see Clinet and Potiron (2018), Erdemlioglu et al. (2025b) and Perrin et al. (2025)).

A simulation study (see Section 4) verifies our results from the central limit theorem. Overall, the standard normal statistics are biased, especially when the final time T gets smaller. However, the bias disappears as the final time T increases. Overall, the variance of the standard normal statistics is close to unity, especially when the final time T increases. We also confirm that the number of observations showing stable results in simulations is compatible with the empirical study from Potiron et al. (2025a).

The remainder of this paper is organized as follows. We introduce Hawkes processes with a stochastic and time dependent baseline and show their existence in Section 2. We prove the central limit theorem for Hawkes processes with a stochastic and time dependent baseline in Section 3. Moreover, we conduct a numerical study in 4. In addition, we provide concluding remarks in Section 5. Finally, the proofs of the theoretical results are gathered in Appendix A.

2. Existence of Hawkes processes with a stochastic and time dependent baseline

In this section, we introduce Hawkes processes with a stochastic and time dependent baseline. We give an existence result in Proposition 1. This complements the framework from Clinet and Potiron (2018) and Erdemlioglu et al. (2025b). In particular, our framework allows for more general kernels, which is useful for applications. The arguments used in the proofs are based on the arguments from Brémaud and Massoulié (1996) and Clinet and Potiron (2018). Since we make the classical assumption that the spectral radius of the kernel matrix is smaller than 1, the proof of existence is close from the proof of existence with an exponential kernel.

We start with the definition of the probabilistic tools in what follows. First, we define $\mathbb{R}_+ = [0, \infty)$ as the space of real nonnegative numbers. We introduce the stochastic basis $\mathcal{B} = \big(\Omega, \mathcal{F}, \overline{\mathcal{F}} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P}\big)$ which is a probability space equipped with a filtration. Namely, we have that Ω is the sample space, $\mathcal{F}$ is the space of events called σ-algebra and $\mathbb{P}$ is a probability function from the sample space Ω to the interval $[0,1]$. The filtration $\overline{\mathcal{F}}$ is a family of σ-algebras $\mathcal{F}_t$ which represents the information available at the time $t \in \mathbb{R}_+$ and satisfies $\mathcal{F}_t \subset \mathcal{F}$. We denote the natural filtration generated by some stochastic process X for any time $t \in \mathbb{R}$ as $\mathcal{F}_t^X = \sigma\{X_s : 0 \leq s \leq t\}$. We assume that $\mathcal{F}_t^N \subset \mathcal{F}_t$ for any time $t \in \mathbb{R}_+$. We also assume that the stochastic basis $\mathcal{B}$ satisfies the usual conditions.

We now give an introduction to the simple point processes. These simple point processes are also called counting processes. We assume that the simple point processes N are of dimension d. Then, each component $N_t^{(i)}$ counts the cumulative number of events between the starting time 0 and the final time t of the ith process for any index $i = 1, \ldots, d$ and any time $t \in \mathbb{R}_+$. Here, we denote the ith component of a vector V by $V^{(i)}$. In the standard literature, a simple point process is typically characterized by the following definition.

Definition 1. We say that the stochastic process N is a simple point process if it satisfies

$$\mathbb{P}\big(N_t^{(i)} \in \{0,1\}\big) = 1 \text{ for any time } t \in \mathbb{R}_+ \text{ and any index } i = 1, \ldots, d\big) = 1.$$

As a consequence of Definition 1, we can deduce some assumptions on the simple point process N. More specifically, we introduce the space of nonnegative integers $\mathbb{N} = \{0, 1, \ldots\}$. We define $N_t^{(i)}$ as a simple point process on the space $\mathbb{R}_+$, namely a family

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

of random variables with values in the space $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$. Here, $\mathcal{B}(S)$ denotes the Borel σ-algebra on the space S for any space S. Moreover,

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

and $\{T_k^{(i)}\}_{k \in \mathbb{N}}$ is a sequence of event times, which are $\mathbb{R}_+$ valued and random.

We assume that the time of the first event $T_0^{(i)}$ is bigger than 0 and smaller than the real positive number $T_0^+ \in \mathbb{R}_+^*$ almost surely and the following times are increasing for each process almost surely. Namely, we assume that

$$\mathbb{P}(0 < T_0^{(i)} < T_0^+ \text{ and } T_k^{(i)} < T_{k+1}^{(i)} \text{ for } k \in \mathbb{N}^* \text{ and } i = 1, \cdots, d) = 1. \tag{5}$$

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 almost surely, namely

$$\mathbb{P}(T_k^{(i)} \neq T_l^{(j)} \text{ for } k, l \in \mathbb{N}^* \text{ and } i, j = 1, \cdots, d \text{ such that } i \neq j) = 1.$$

We now introduce the definition of the intensity with respect to the filtration $\overline{\mathcal{F}}$ for the simple point process N.

Definition 2. Any stochastic process λ defined on the space of real positive numbers $\mathbb{R}_+$ and satisfying the following properties is called an intensity with respect to the filtration $\overline{\mathcal{F}}$ of the point process N. First, we have that

$$\mathbb{E}[N((a,b]) \mid \mathcal{F}_a] = \mathbb{E}\Big[\int_a^b \lambda_s ds \mid \mathcal{F}_a\Big] \text{ almost surely} \tag{6}$$

for all intervals $(a,b] \subset \mathbb{R}^+$. Second, the stochastic process λ is progressively measurable with respect to the filtration $\overline{\mathcal{F}}$, of dimension d where each component $\lambda_t^{(i)}$ takes its values in the space of nonnegative real numbers $\mathbb{R}^+$.

Intuitively, the intensity corresponds to the expected number of events given the past information, namely

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

Moreover, we have that the compensated point process defined as

$$M_t = N_t - \int_0^t \lambda_s ds \tag{7}$$

is a martingale with respect to the filtration $\overline{\mathcal{F}}$ almost surely. Finally, we note that $N((a,b])$ is almost surely finite if and only if $\int_a^b \lambda_s ds$ is almost surely finite. For background on simple 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 Hawkes processes featuring a stochastic and time dependent baseline. More specifically, the intensity λ_t of the point process N_t for any time $t \in [0,T]$ follows

$$\lambda_t = b_t + \int_0^t h(t-s)\, dN_s. \tag{8}$$

Here, 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 b_t is a stochastic process of dimension d.

We introduce a rescaled baseline which satisfies $b_t = \nu_{t/T}$ for any time $t \in [0,T]$. Here, ν_t is a stochastic process of dimension d defined on the time interval $[0,1]$. Then, the intensity λ_{tT} of the point process with rescaled baseline N_{tT} for any time $t \in [0,1]$ and any final time $T > 0$ follows

$$\lambda_{tT} = \nu_t + \int_0^{tT} h(tT - s)\, dN_s. \tag{9}$$

Here, the point process N_{tT} and its intensity λ_{tT} implicitly depend on the final time T. Moreover, the baseline process ν_t is rescaled from the time interval $[0,T]$ to the time interval $[0,1]$.

We denote the spectral radius of any matrix ϕ by $\rho(\phi)$. We have by definition that $\rho(\phi)$ is the square root of the largest eigenvalue of the matrix $\phi^T\phi$, namely

$$\rho(\phi) = \sqrt{\lambda_{\max}\left(\phi^T\phi\right)}.$$

Here ϕ^T denotes the transpose matrix of the matrix ϕ. Then, we define the L^1 norm matrix for the kernel h of dimension $d \times d$ as

$$\|h\|_1 = \int_0^\infty h(s)ds.$$

Finally, we denote the product measure by $\mu_1 * \mu_2$ for two measures μ_1 and μ_2.

We first introduce assumptions required for the existence of Hawkes processes with a stochastic and time dependent baseline.

Assumption 1. (a) For any index $i = 1, \dots, d$, the ith component of the baseline process is almost surely positive on the time interval $[0,1]$ almost everywhere, namely

$$\mathbb{P}\big(\nu_t^{(i)} > 0 \ \forall t \in [0,1] \text{ almost everywhere}\big) = 1.$$

(b) For any index $i = 1, \dots, d$, the ith component of the baseline process is almost surely integrable on the time interval $[0,1]$, namely

$$\mathbb{P}\Big(\int_0^1 \nu_s^{(i)} ds < \infty\Big) = 1.$$

(c) We have that the point process N is generated by a stochastic process $\underline{N}$, which is an Poisson process adapted to the filtration $\overline{\mathcal{F}}$ of intensity 1 and dimension $2d$. Namely, we have for any index $i = 1, \dots, d$, any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ that

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

Moreover, we have that the baseline process ν is independent from the Poisson process $\underline{N}$. Finally, we have that the filtration is equal to $\mathcal{F}_{tT} = \mathcal{F}_t^{\nu} \vee \mathcal{F}_{tT}^{\underline{N}}$ for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$.

(d) For any index $i = 1, \dots, d$, any index $j = 1, \dots, d$ and any time $t \in \mathbb{R}^+$, the component with ith raw and jth column of the kernel is nonnegative at time t, namely $h^{(i,j)}(t) \geq 0$.

(e) The spectral radius of the L^1 norm matrix for the kernel is strictly less than one, namely $\rho(\|h\|) < 1$.

Assumption 1 (a) implies that the point process is well-defined. Assumption 1 (b) takes its roots in the simpler case of heterogeneous Poisson processes without a kernel (see Daley and Vere-Jones (2003)). Assumption 1 (c) introduces Poisson imbedding and is already required with traditional Hawkes processes (see Brémaud and Massoulié (1996), Section 3, pp. 1571-1572). In particular, the point process N is generated by a Poisson process $\underline{N}$. More specifically, the stochastic process N is defined as the point process counting the points of the Poisson process $\underline{N}$ below the intensity curve $t \to \lambda_t$. Assumption 1 (c) also considers independence between the the baseline process ν and the Poisson process $\underline{N}$. Moreover, we can deduce from Assumption 1 (c) that for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ the natural filtration generated by the point process N is included in the main filtration, namely $\mathcal{F}_{tT}^N \subset \mathcal{F}_{tT}$. Assumption 1 (d) is restrictive for kernels with inhibitory effects. Finally, Assumption 1 (e) 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)).

Overall, the constraints on the kernel shape are weaker than the constraints on the kernel shape from Theorem 5.1 in Clinet and Potiron (2018). More specifically, our framework only requires the nonnegativity of the kernel whereas Clinet and Potiron (2018) considers exponential kernels, which are very restrictive for applications. However, Clinet and Potiron (2018) consider locally parametric Hawkes processes, where the baseline and the parameters of the kernels are stochastic and time dependent. See also Condition 1 in Erdemlioglu et al. (2025b) for the extension to generalized gamma kernels.

In the proposition that follows, we state the existence of Hawkes processes with a stochastic and time dependent baseline. The kernel has a general form. This complements Theorem 5.1 in Clinet and Potiron (2018) and Proposition 1 in Erdemlioglu et al. (2025b). In particular, our kernel framework allows for more general kernels, which is useful for applications. The arguments used in the proofs are based on the arguments from Brémaud and Massoulié (1996) and Clinet and Potiron (2018). Since we make the classical assumption that the spectral radius of the kernel matrix is smaller than 1, the proof of existence is close from the proof of existence with an exponential kernel.

Proposition 1. *We assume that Assumption 1 holds. Then, there exists a point process $(N_{tT})_{t\in[0,1]}$ adapted to the filtration $(\mathcal{F}_{tT})_{t\in[0,1]}$ with intensity with respect to the filtration $(\mathcal{F}_{tT})_{t\in[0,1]}$ of the form (9) for any final time $T \in \mathbb{R}^+$. Moreover, the intensity process $(\lambda_{tT})_{t\in[0,1]}$ is almost surely integrable.*

3. Central limit theorem of Hawkes processes with a stochastic and time dependent baseline

In this section, we develop statistical inference for Hawkes processes with a stochastic and time dependent baseline. The inference procedure is based on the average of the point process. As this is useful for applications, we extend the framework from Bacry et al. (2013) and Deschatre, Gruet and Lotz (2025) which is restricted to nonrandom baseline. We show the central limit theorem of the statistical inference procedure in Theorem 1. This requires the assumption that the kernel does not have a too fat tail. We also need that the baseline process is Lipschitz continuous with bounded starting value. The main novelty in the proofs is to establish the renewal equation for stochastic processes. This extends Bacry et al. (2013) and Deschatre, Gruet and Lotz (2025) in which the function from the renewal equation is nonrandom.

We consider estimation for the rescaled integral of the intensity process λ between the starting time 0 and the final time tT for any time $t \in [0,1]$, namely

$$\Lambda_{t,T} = \frac{1}{T}\int_0^{tT} \lambda_s ds. \tag{10}$$

As discussed in the introduction, we have applications in finance of the target quantity (10) where the intensity of a quote plays an inverse role to the volatility of an asset price. This is due to the fact that quote intensity is a central measure in financial economics.

This estimation procedure is in the sense of a stochastic process starting from the time interval $[0,1]$. In the particular case when $t=1$, the target quantity (10) corresponds to estimation for the average of the intensity process λ between the starting time 0 and the final time T. We denote the limit process of $\Lambda_{t,T}$ for any time $t \in [0,1]$ as the time T increases by

$$\Lambda_t = (I - \|h\|_1)^{-1}\int_0^t \nu_s ds. \tag{11}$$

We propose estimation of the limit process Λ_t for any time $t \in [0,1]$ by

$$\widehat{\Lambda}_t = \frac{N_{tT}}{T}. \tag{12}$$

Then, we introduce some quantities required to establish the form of the asymptotic covariance matrix. We define w_t as the stochastic process which is a diagonal matrix of dimension $d \times d$ for any time $t \in [0,1]$. More specifically, we have that the ith diagonal component of the stochastic process w_t at the time $t \in [0,1]$ is equal to

$$w_t^{(i,i)} = \big((I - \|h\|_1)^{-1}\nu_t\big)^{(i)}.$$

Then, we define c_t as the stochastic process of dimension $d \times d$ for any time $t \in [0,1]$ which satisfies

$$c_t = (I - \|h\|_1)^{-1} w_t^{1/2}.$$

We have now all the ingredients to derive the form of the asymptotic covariance matrix. We define the asymptotic covariance matrix for any time $t \in [0,1]$ as

$$\Sigma_t^2 = \int_0^t c_s c_s^T ds.$$

We deliver in what follows the assumptions used for the central limit theorem of Hawkes processes with a stochastic and time dependent baseline.

Assumption 2. (a) The kernel satisfies $\int_0^\infty th(t)dt < \infty$.

(b) The starting point of the baseline is almost surely bounded, namely there is a nonrandom constant satisfying $C_0 \geq 0$ and

$$\mathbb{P}\big(\nu_0 \leq C_0\big) = 1.$$

(c) The baseline process is almost surely Lipschitz-continuous with nonrandom constant satisfying $C > 0$ on the time interval $[0, 1]$, namely

$$\mathbb{P}\big(|\nu_t - \nu_s| \leq C|t - s| \forall (t, s) \in [0, 1]^2\big) = 1.$$

Assumption 2 (a) puts some restrictions on the kernel shape h. This corresponds exactly to Assumption (A2) in Bacry et al. (2013) (p. 2480). This is required to obtain a martingale form of the intensity process. Assumption 2 (a) is also used in Jaisson and Rosenbaum (2015). Overall, the assumptions on the kernel shape from this paper are exactly the same as the assumptions used for Corollary 1 in Bacry et al. (2013).

Assumptions 2 (b) and (c) put restrictions on the baseline process ν. More specifically, Assumption 2 (b) relies on a starting point which is bounded almost surely. Assumption 2 (c) considers a baseline process ν which is Lipschitz-continuous almost surely. Assumptions 2 (b) and (c) are mainly used to obtain a locally bounded baseline process ν. This is needed to exhibit a martingale representation of the intensity in the proofs (see Lemma 4). This relies on establishing the renewal equation. Thus, Assumptions 2 (b) and (c) can be relaxed as long as we get a locally bounded baseline process ν. However, we were not able to find weaker assumptions with a reasonably tractable form.

Assumptions 2 (b) and (c) are reasonably general assumptions. In particular, we restrict to the baseline process ν which is almost surely Lipschitz-continuous on the time interval $[0, 1]$. By definition, this implies that baseline process ν is Holder continuous with Holder condition 1 on the time interval $[0, 1]$. When $0 < \alpha \leq \beta \leq 1$, then all Holder continuous functions with Holder condition β on the time interval $[0, 1]$ are also Holder continuous with Holder condition α. This also includes $\beta = 1$ and therefore all Lipschitz continuous functions on the time interval $[0, 1]$ are also Holder continuous with Holder condition α. As an example, the Brownian motion is almost surely Holder continuous with Holder condition $1/2 - \epsilon$ for any $\epsilon \in \mathbb{R}_+^*$. Thus, the Brownian motion is a particular case allowed by our assumptions. Moreover, a continuous Ito semimartingale with finite drift moments and finite volatility moments also satisfy our assumptions. Finally, the case of an Ito semimartingale baseline which is not necessarily continuous is studied in Potiron et al. (2025a).

Our assumptions on the baseline process ν can also be compared to the assumptions from some related work on Hawkes processes with a stochastic and time dependent baseline. Assumption [C] (i) from Clinet and Potiron (2018) (p. 3477) is related to the global regularity modulus of the baseline process ν. Thus, their assumption is stronger than the assumptions from our paper and they do not allow for the simple case of Brownian motion. The same assumption appears in Condition 2 (e) from Erdemlioglu et al. (2025b). The reason is that they need smoother baseline processes to approximate locally with maximum likelihood estimation. Perrin et al. (2025) consider Hawkes process with a diffusion-driven baseline. More specifically, the baseline process ν is modeled as the solution of the stochastic differential equation. Assumption 1 in Perrin et al. (2025) restricts to coefficient functions which are continuous and globally Lipschitz. However, they allow for baseline process ν to be discontinuous. Thus, they are more general than us since we require the continuity of the baseline process ν. The reason why Perrin et al. (2025) can allow dor discontinuous baseline processes ν is that their proofs do not rely on the same martingale representation of the intensity process λ.

In the theorem that follows, we state the central limit theorem for Hawkes processes with a stochastic and time dependent baseline. The kernel has a general form. We consider asymptotics when the final time diverges to infinity, namely $T \to +\infty$. This is the main result of this paper. This extends Corollary 1 (p. 2481) from Bacry et al. (2013) which is restricted to nonrandom constant baseline and Corollary 1 from Deschatre, Gruet and Lotz (2025) which is restricted to nonrandom time-dependent baseline. The main novelty in the proofs is to establish the renewal equation for stochastic processes. This extends Bacry et al. (2013) and Deschatre, Gruet and Lotz (2025) in which the function from the renewal equation is nonrandom. Moreover, the convergence rate is $\sqrt{T}$.

We denote the collection of cadlag functions starting from the space $[0,1]$ to the space $\mathbb{R}$ by $\mathbb{D}([0,1],\mathbb{R})$, which is referred as Skorokhod space. Moreover, we denote $\xrightarrow{\mathcal{D}-s}$ as the stable weak convergence with respect to the filtration $\overline{\mathcal{F}}$ for the Skorokhod space $\mathbb{D}([0,1],\mathbb{R})$ equipped with the Skorokhod topology.

Theorem 1. *We assume that Assumptions 1 and 2 hold. Then, there is a canonical d-dimensional standard Brownian extension of the stochastic basis $\mathcal{B}$. This extension includes the canonical standard Brownian motion W_t which satisfies as $T \to \infty$ that*

$$\sqrt{T}\big(\widehat{\Lambda}_t - \Lambda_t\big) \xrightarrow{\mathcal{D}-s} \int_0^t c_s dW_s. \tag{13}$$

4. Simulation study

In this section, we conduct a simulation study to document how the estimator behaves in finite samples.

We consider the following simulation design to replicate features observed in financial markets. All the models we introduce satisfy the conditions of the theory discussed in the previous sections. We interpret the interval $[0,1]$ as 6.5-hour-long day of trading. The order of the observation number n varies is fixed to $50,000$ for checking the asymptotic approximation. Moreover, the final time T varies from 1 to 20. With these realistic values, the simulation design allows for moderately traded stocks. The number of replications is equal to $1,000$. We use the Python package tick (see Bacry et al. (2017)) for the generation of the point process.

We define the intensity process as

$$\lambda_t \quad = \quad n(1 - \|\phi\|_1)\nu_t + \textstyle\int_0^t nh(n(t-s))dN_s. \tag{14}$$

Here, n is the order of the observation number which is fixed to $50,000$. It means that we have around $50,000$ quotes for a trading day, which is what we get for moderately traded stocks in practice. We also have that the component of the baseline ν satisfies a square root process (SRP)

$$d\nu_t \quad = \quad 30(b_t - \nu_t)dt + 3\sqrt{\nu_t}dW_t. \tag{15}$$

The nonrandom function b is a solution of the ordinary differential equation $dr_t = 30(b_t - r_t)dt$ with U-shape r_t defined as $r_0 = \nu_0$ and

$$r_t = 20\left((t - 0.53)^4 + \frac{1}{24}\right).$$

The tune parameter 0.53 centers the U-shaped component in the middle of the trading day. The drift term in Equation (15) ensures mean reversion of the process ν to the function b. Moreover, the function b pushes the process ν to follow the U-shape nonrandom term r. In Equation (15), the diffusion term $\sqrt{\nu_t}dW_t$ captures the random fluctuation. The Feller condition (see Feller (1951)) is satisfied with $30 \times b_t \geq 3^2$ for any time $t \in [0, T]$, thus the stochastic process ν is positive. Compared to the simulation design from Potiron et al. (2025a), we consider a continuous baseline with no jumps. With this choice, the intensity process remains continuous, locally bounded and the conditions of the theory hold.

Kernels are specified as follows. An exponential kernel is defined as $\phi_e(t) = 1.6e^{-2t}$ and a power kernel defined as $\phi_p(t) = 1.6(1+t)^{-3}$. With these kernel values, the L^1 norm is equal to $\|\phi\|_1 = 0.8$, which is the average value that we obtain in the empirical application in Potiron et al. (2025a) and in the results of Filimonov and Sornette (2012). Moreover, these kernels correspond exactly to the kernels chosen in the numerical study from Potiron et al. (2025a).

We consider the following model variants to disentangle the effects. First, we set Model 1 as a null kernel and a constant baseline, namely $\lambda_t = n$. Second, we set Model 2 as a null kernel and a U-shape baseline, namely $\lambda_t = 20\big((t-0.53)^4 + \frac{1}{24}\big)n$. Third, we set Model 3 as a null kernel and a U-shape + SRP, namely $\lambda_t = n\nu_t$. Then, we set Model 4 as an exponential kernel and a constant baseline, namely $\lambda_t = n + \int_0^t n\phi_e(n(t-s))dN_s$. We also set Model 5 as an exponential kernel and a U-shape baseline, namely $\lambda_t = n\mu_t + \int_0^t n\phi_e(n(t-s))dN_s$ in which $\mu_t = 20(1 - \|\phi\|_1)\big((t-0.53)^4 + \frac{1}{24}\big)$. We set Model 6 as an exponential kernel and a U-shape + SRP, namely $\lambda_t = n(1 - \|\phi\|_1)\nu_t + \int_0^t n\phi_e(n(t-s))dN_s$. We set Model 7 as a power kernel and a constant baseline, namely $\lambda_t = n + \int_0^t n\phi_p(n(t-s))dN_s$. We set Model 8 as a power kernel and a U-shape baseline, namely $\lambda_t = n\mu_t + \int_0^t n\phi_p(n(t-s))dN_s$. Finally, we set Model 9 as a power kernel and a U-shape + SRP, namely $\lambda_t = n(1-\|\phi\|_1)\nu_t + \int_0^t n\phi_p(n(t-s))dN_s$. These models are summarized in Table 1.

Figure 1 provides a comparison between simulated intensity from Model 9 (left panel) and intensity based on AAPL (Apple) data on April 1st 2016 (right panel). The intensity is obtained from one-minute intervals. The simulated process captures the U-shaped pattern and intensity burst well. It also exhibits some random fluctuation of the baseline intensity. These patterns can also be seen in the data that justify our simulation design being realistic.

TABLE 1
Summary of models.

	Baseline Model (ν)		
Kernel	Constant	U-shape	U-shape + SRP
Null	Model 1	Model 2	Model 3
Exponential	Model 4	Model 5	Model 6
Power	Model 7	Model 8	Model 9

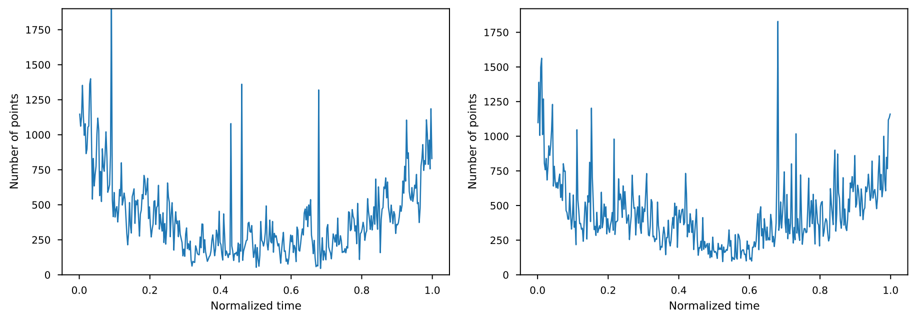

Fig 1: Comparison between simulated intensity with Model 9 (left panel) and intensity based on AAPL data on April 1st 2016 (right panel).

Table 2 reports the summary for the standard normal statistics of the baseline integral using Models 1 to 9. The order of the observation number is $n = 50,000$, the final time T varies from 1 to 20 days of trading and the number of replications is $1,000$. The absolute value of the mean ranges from 1% to 27%, with an average of 10%. Overall, the statistics are biased, especially when the final time T gets smaller. However, the bias disappears as the final time T increases. The variance ranges from 99% to 108%, with an average of 103%. Overall, the variance is close to unity, especially when the final time T increases.

5. Conclusion

We have developed statistical inference for Hawkes processes with a stochastic and time dependent baseline. The statistical inference procedure was based on the average of the point process. We have considered estimation for the average over time of the intensity process. We first showed the existence of these point processes. We also showed the central limit theorem of the statistical inference procedure. This required the assumption that the kernel does not have a too fat tail. We also needed that the baseline process was Lipschitz continuous with bounded starting value. The main novelty in the proofs was to establish the renewal equation for stochastic processes.

There are two recent papers who are extending the theoretical work from this paper. First, Potiron et al. (2025a) introduce Hawkes processes with an Itô semi-

TABLE 2
Summary for the standard normal statistics of the baseline integral using Models 1 to 9. The order of the observation number is $n = 50,000$, the final time T varies from 1 to 20 days of trading and the number of replications is 1,000.

n	50,000							
T	1		5		10		20	
Model	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Model 1	-0.0217	1.0548	-0.0178	1.0365	0.0109	1.0056	0.0088	1.0067
Model 2	0.0638	1.0683	0.0420	1.0406	-0.0136	1.0318	-0.0097	1.0175
Model 3	-0.1456	1.0713	-0.0836	1.0418	-0.0762	1.0167	0.0423	1.0178
Model 4	0.1549	1.0194	0.0945	1.0156	-0.0567	1.0552	0.0379	1.0553
Model 5	0.1230	1.0777	0.0803	1.0569	0.0638	1.0201	0.0442	1.0145
Model 6	0.0862	1.0450	0.0534	1.0428	0.0383	1.0088	0.0326	1.0068
Model 7	0.2726	0.9854	0.1683	0.9934	0.1225	1.0090	0.0748	1.0089
Model 8	0.2553	1.0182	0.1592	1.0124	0.1223	1.0213	0.0773	1.0170
Model 9	0.2012	1.0824	0.1185	1.0436	0.0732	1.0694	0.4122	1.0221

martingale baseline. The authors obtain five main applications from the theory. They derive estimation of the integrated intensity, estimation for the integral of the baseline and estimation for the integral of the baseline volatility. They also provide an hypothesis testing for the absence of a Hawkes component and an hypothesis testing for constant baseline. Moreover, an empirical application on high-frequency data finds rejection of both the null hypothesis of no Hawkes excitation and that of constant baseline contributing to the formal identification of patterns in high-frequency trading activity. Second, Potiron et al. (2025b) develop estimation of the branching ratio $\|h\|_1$ as another application. They also consider an hypothesis testing for a branching ratio equal to a given value. Finally, they also propose a test for a branching ratio value against a branching ratio tending to unity as the number of observations increases.

Funding

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

References

ADMATI, A. R. and PFLEIDERER, P. (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies* **1** 3–40.

AÏT-SAHALIA, Y. and JACOD, J. (2014). *High-frequency financial econometrics.* Princeton University Press.

ALMGREN, R. and CHRISS, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk* **3** 5–40.

BACRY, E., DAYRI, K. and MUZY, J.-F. (2012). Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B* **85** 1–12.

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.

Bacry, E., Bompaire, M., Deegan, P., Gaïffas, S. and Poulsen, S. V. (2017). tick: A Python library for statistical learning, with an emphasis on Hawkes processes and time-dependent models. *The Journal of Machine Learning Research* **18** 7937–7941.

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

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.

Christensen, K. and Kolokolov, A. (2024). An unbounded intensity model for point processes. *Journal of Econometrics* **244** 105840.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 135–155.

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.

Deschatre, T., Gruet, P. and Lotz, A. (2025). Some limit theorems for locally stationary Hawkes processes. *arXiv preprint arXiv:2501.17245*.

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.

Epps, T. W. and Epps, M. L. (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis. *Econometrica* 305–321.

Erdemlioglu, D., Potiron, Y., Xu, T. and Volkov, V. (2025a).

Estimation of latency for Hawkes processes with a polynomial periodic kernel. *Working paper available at https://www.fbc.keio.ac.jp/∼potiron/Erdemlioglu2025workingpaperestimationlatency.pdf*.
ERDEMLIOGLU, D., POTIRON, Y., VOLKOV, V. and XU, T. (2025b). Estimation of time-dependent latency with locally stationary Hawkes processes. *Working paper available at https://www.fbc.keio.ac.jp/∼potiron/Erdemlioglu2025workingpaper.pdf*.
FELLER, W. (1951). Two singular diffusion problems. *Annals of Mathematics* 173–182.
FILIMONOV, V. and SORNETTE, D. (2012). Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E* **85** 056108.
GRAHAM, C. (2021). Regenerative properties of the linear Hawkes process with unbounded memory. *Annals of Applied Probability* **31** 2844–2863.
HARDIMAN, S. J., BERCOT, N. and BOUCHAUD, J.-P. (2013). Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B* **86** 1–9.
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. (2018). Hawkes processes and their applications to finance: a review. *Quantitative Finance* **18** 193–198.
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.*
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 PROTTER, P. E. (2012). *Discretization of processes.* Springer Berlin, Heidelberg.
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., Chen, F. and Dunsmuir, W. (2023). Alternative asymptotic inference theory for a nonstationary Hawkes process. *Journal of Statistical Planning and Inference* **227** 75–90.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica* 1315–1335.

Liniger, T. (2009). Multivariate Hawkes processes, PhD thesis, ETH Zurich.

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.

Perrin, M. S., Bonnet, A., Dion-Blanc, C. and Samson, A. (2025). Hawkes process with a diffusion-driven baseline: long-run behavior, inference, statistical tests. *arXiv preprint arXiv:2512.01447*.

Potiron, Y. (2025a). Parametric inference for Hawkes processes with a general kernel. *Working paper available at https://www.fbc.keio.ac.jp/∼ potiron/Potiron2025inferenceworkingpaper.pdf*.

Potiron, Y. (2025b). Parametric inference for nonlinear Hawkes processes. *Working paper available at https://www.fbc.keio.ac.jp/∼ potiron/Potiron2025nonlinearworkingpaper.pdf*.

Potiron, Y. (2026). Statistical inference for general ergodic point processes. *Working paper available at https://www.fbc.keio.ac.jp/∼ potiron/Potiron2026statisticalworkingpaper.pdf*.

Potiron, Y. and Volkov, V. (2026). 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/∼ 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/∼ potiron/Potiron2025estimationworkingpaper.pdf*.

Rambaldi, M., Filimonov, V. and Lillo, F. (2018). Detection of intensity bursts using Hawkes processes: An application to high-frequency financial data. *Physical Review E* **97** 032318.

Rambaldi, M., Pennesi, P. and Lillo, F. (2015). Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Physical Review E* **91** 012819.

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.
Tauchen, G. E. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica* 485–505.
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.

Appendix A: Proofs

We begin this appendix with some general guidelines that we use extensively during the proofs. First, we use C for any generic constant, and the value of the constant can change from one line to the next. In addition, any operation with two vectors of the same dimension means the operation component by component.

A.1. Proof of existence

In this part, we focus on the proof of the existence of Hawkes processes with a rescaled stochastic and time dependent baseline. This corresponds to the proof of Proposition 1.

Prior to the first lemma, we introduce some notation. First, we define the point process N_{tT} conditioned by the information from the baseline process for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ as

$$N_{tT,\nu} = \mathbb{E}\big[N_{tT}\big|\mathcal{F}_1^{\nu}\big]. \tag{16}$$

Here, $\mathbb{E}[X|\mathcal{F}]$ denotes the conditional expectation of the random variable X with respect to the σ-algebra $\mathcal{F}$. Then, $\mathbb{E}[X|\mathcal{F}]$ is a random variable which is measurable with respect to the σ-algebra $\mathcal{F}$. We also define the intensity process λ_{tT} conditioned by the information from the baseline process for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ as

$$\lambda_{tT,\nu} = \mathbb{E}\big[\lambda_{tT}\big|\mathcal{F}_1^{\nu}\big]. \tag{17}$$

Finally, we define the filtration $\mathcal{F}_{tT}$ conditioned by the information from the baseline process for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ as

$$\mathcal{F}_{tT,\nu} = \mathbb{E}\big[\mathcal{F}_{tT}\big|\mathcal{F}_1^{\nu}\big]. \tag{18}$$

This first lemma shows that the stochastic process $(\lambda_{tT,\nu})_{t\in[0,1]}$ is the intensity with respect to the filtration $(\mathcal{F}_{tT,\nu})_{t\in[0,1]}$ of the point process $(N_{tT,\nu})_{t\in[0,1]}$ in the sense of Definition 2 for any final time $T \in \mathbb{R}_+$. This extends Lemma 10.1 (p. 2) from the supplementary materials of Clinet and Potiron (2018).

Lemma 1. *We assume that Assumption 1 (c) hold. Then, the stochastic process $(\lambda_{tT,\nu})_{t\in[0,1]}$ is the intensity with respect to the filtration $(\mathcal{F}_{tT,\nu})_{t\in[0,1]}$ of the point process $(N_{tT,\nu})_{t\in[0,1]}$ in the sense of Definition 2 for any final time $T \in \mathbb{R}_+$.*

Proof of Lemma 1. To prove the lemma, we verify that the properties from Definition 2 are satisfied. First, we have for any interval $(a,b] \subset [0,T]$, any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ by Definitions (16) and (18) that

$$\mathbb{E}\big[N_{tT,\nu}((a,b])\big|\mathcal{F}_{a,\nu}\big] = \mathbb{E}\Big[\mathbb{E}\big[N((a,b])\big|\mathcal{F}_1^{\nu}\big]\Big|\mathbb{E}\big[\mathcal{F}_a\big|\mathcal{F}_1^{\nu}\big]\Big]. \tag{19}$$

Then, we can rewrite the right side of Equation (19) by conditional expectation properties for any interval $(a,b] \subset [0,T]$, any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ as

$$\mathbb{E}\Big[\mathbb{E}\big[N((a,b])\big|\mathcal{F}_1^{\nu}\big]\Big|\mathbb{E}\big[\mathcal{F}_a\big|\mathcal{F}_1^{\nu}\big]\Big] = \mathbb{E}\Big[\mathbb{E}\big[N((a,b])\big|\mathcal{F}_a\big]\Big|\mathcal{F}_1^{\nu}\Big]. \tag{20}$$

In addition, we obtain by Equation (6) from Definition 2 for any interval $(a,b] \subset [0,T]$, any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ that

$$\mathbb{E}\Big[\mathbb{E}\big[N((a,b])\big|\mathcal{F}_a\big]\Big|\mathcal{F}_1^{\nu}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\int_a^b \lambda_s ds\Big|\mathcal{F}_a\Big]\Big|\mathcal{F}_1^{\nu}\Big] \text{ a.s.} \tag{21}$$

Moreover, we get by conditional expectation properties for any interval $(a,b] \subset [0,T]$, any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ that

$$\mathbb{E}\Big[\mathbb{E}\Big[\int_a^b \lambda_s ds\Big|\mathcal{F}_a\Big]\Big|\mathcal{F}_1^{\nu}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\int_a^b \lambda_s ds\Big|\mathcal{F}_1^{\nu}\Big]\Big|\mathbb{E}\big[\mathcal{F}_a\big|\mathcal{F}_1^{\nu}\big]\Big]. \tag{22}$$

Finally, we deduce by Tonelli's theorem, Definitions (17) and (18) for any interval $(a,b] \subset [0,T]$, any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ that

$$\mathbb{E}\Big[\mathbb{E}\Big[\int_a^b \lambda_s ds\Big|\mathcal{F}_1^{\nu}\Big]\Big|\mathbb{E}\big[\mathcal{F}_a\big|\mathcal{F}_1^{\nu}\big]\Big] = \mathbb{E}\Big[\int_a^b \lambda_{s,\nu} ds\Big|\mathcal{F}_{a,\nu}\Big]. \tag{23}$$

Thus, Equations (19), (20), (21), (22) and (23) for any interval $(a,b] \subset [0,T]$, any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ yield

$$\mathbb{E}\big[N_{tT,\nu}((a,b])\big|\mathcal{F}_{a,\nu}\big] = \mathbb{E}\Big[\int_a^b \lambda_{s,\nu} ds\Big|\mathcal{F}_{a,\nu}\Big] \text{ a.s.} \tag{24}$$

This means that we have shown Equation (6). Second, the stochastic process $(\lambda_{tT,\nu})_{t\in[0,1]}$ is progressively measurable with respect to the filtration

$$(\mathcal{F}_{tT,\nu})_{t\in[0,1]},$$

of dimension d where each component $\lambda_{t,\nu}^{(i)}$ takes its values in the space of non-negative real numbers $\mathbb{R}^+$. Thus, we have shown Definition 2. □

We now give the proof of the existence of Hawkes processes with a stochastic and time dependent baseline. It extends the proof of Theorem 7 (pp. 1585-1587) in Brémaud and Massoulié (1996). It complements the proof of Theorem 5.1 (pp. 3-4) in the supplement of Clinet and Potiron (2018) and the proof of Proposition 1 in Erdemlioglu et al. (2025b).

Proof of Proposition 1. The strategy of the proof consists in defining a suitable sequence of point processes and intensity $(N_{tT,k}, \lambda_{tT,k})_{k\in\mathbb{N}}$ for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$. Then, we show that their limit defined as $(N_{tT}, \lambda_{tT}) = \lim_{k\to\infty}(N_{tT,k}, \lambda_{tT,k})$ exists and that the point process N_{tT} admits $(\lambda_{tT})_{t\in[0,1]}$ as an intensity with respect to the filtration $(\mathcal{F}_{tT})_{t\in[0,1]}$ given by Equation (9).

We first introduce for any time $t \in [0,1]$, any final time $T \in \mathbb{R}^+$ and any index $i = 1, \ldots, d$ the ith component of the initial intensity process

$$\lambda_{tT,0}^{(i)} = \nu_t^{(i)}.$$

We also introduce the ith component of the initial point process $N_{tT,0}^{(i)}$ for any time $t \in [0,1]$, any final time $T \in \mathbb{R}^+$ and any index $i = 1, \ldots, d$. It is defined as the point process counting the points of the Poisson process $\underline{N}^{(2i-1)} * \underline{N}^{(2i)}$ below the curve $t \to \lambda_{tT,0}^{(i)}$, namely

$$N_{tT,0}^{(i)} = \int_{[0,tT]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_{s,0}^{(i)}]}(x)\underline{N}^{(2i-1)} * \underline{N}^{(2i)}(ds \times dx).$$

We define recursively the sequence of point process and its intensity $(N_{tT,k}^{(i)}, \lambda_{tT,k}^{(i)})$ for any time $t \in [0,1]$, any final time $T \in \mathbb{R}^+$ and any $k \in \mathbb{N}$ as

$$\begin{aligned} \lambda_{tT,k+1} &= \nu_t + \int_0^{tT} h(tT-s) dN_{s,k}, \qquad (25) \\ N_{tT,k+1}^{(i)} &= \int_{[0,tT]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_{s,k+1}^{(i)}]}(x)\underline{N}^{(2i-1)} * \underline{N}^{(2i)}(ds \times dx) \text{ for any } i = 1, \ldots, d. \end{aligned}$$

First, we have that the stochastic process $\lambda^{(i)}_{tT,k}$ is positive for any time $t \in [0,1]$, any final time $T \in \mathbb{R}^+$ and any index $i = 1, \ldots, d$ almost surely by Assumptions 1 (a) and (d). Thus, the stochastic process $(\lambda_{tT,k})_{t\in[0,1]}$ is a well-defined intensity. Then, an extension to the stochastic and time dependent baseline case of the arguments from Lemma 3 and Example 4 (pp. 1571-1572) in Brémaud and Massoulié (1996) yields that the point process $(N_{tT,k})_{t\in[0,1]}$ is adapted to the filtration $(\mathcal{F}_{tT})_{t\in[0,1]}$. It also gives that the stochastic process $(\lambda_{tT,k})_{t\in[0,1]}$ is predictable with respect to the filtration $(\mathcal{F}_{tT})_{t\in[0,1]}$ and an intensity with respect to the filtration $(\mathcal{F}_{tT})_{t\in[0,1]}$ of $(N_{tT,k})_{t\in[0,1]}$ in the sense of Definition 2. Moreover, Assumption 1 (d) implies that $(N^{(i)}_{tT,k}, \lambda^{(i)}_{tT,k})$ is componentwise increasing with k and thus converges to some limit $(N^{(i)}_{tT}, \lambda^{(i)}_{tT})$ almost surely for any time $t \in [0,1]$, any final time $T \in \mathbb{R}^+$ and any index $i = 1, \ldots, d$.

We now introduce the sequence of vector processes $\rho_{tT,k}$ for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ defined as

$$\rho_{tT,k} = \mathbb{E}[\lambda_{tT,k} - \lambda_{tT,k-1} | \mathcal{F}^{\nu}_1]. \tag{26}$$

First, we get by Definition (26) for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ that

$$\rho_{tT,k+1} = \mathbb{E}[\lambda_{tT,k+1} - \lambda_{tT,k} | \mathcal{F}^{\nu}_1]. \tag{27}$$

Then, we obtain by inserting the intensity definition (9) into Equation (27) for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ that

$$\rho_{tT,k+1} = \mathbb{E}\Big[\int_0^{tT} h(tT-s)(dN_{s,k+1} - dN_{s,k})\Big|\mathcal{F}^{\nu}_1\Big]. \tag{28}$$

Moreover, we get by an application of Lemma 1 with point process properties for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ that

$$\rho_{tT,k+1} = \mathbb{E}\Big[\int_0^{tT} h(tT-s)(\lambda_{s,k+1} - \lambda_{s,k})ds\Big|\mathcal{F}^{\nu}_1\Big]. \tag{29}$$

In addition, we have that the integrand in the integral from Equation (29) is nonnegative as we have that $(N^{(i)}_{tT,k}, \lambda^{(i)}_{tT,k})$ is componentwise increasing with k. We have already seen that this monotonicity property is due to Assumption 1 (d) and Definition (25). Thus, we can deduce by Tonelli's theorem and Definition (26) that

$$\rho_{tT,k+1} = \int_0^{tT} h(tT-s)\rho_{s,k}ds. \tag{30}$$

We define $\Phi_{tT,k}$ as the integral of the stochastic process $\rho_{s,k}$ from the starting time 0 to the final time tT for any time $t \in [0,1]$ and any time $T \in \mathbb{R}^+$, namely $\Phi_{tT,k} = \int_0^{tT} \rho_{s,k}ds$. Then, we have by another application of Tonelli's theorem that for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ almost surely that

$$\Phi_{tT,k+1} = \int_0^{tT} \Big(\int_0^{tT-s} h(u)du\Big)\rho_{s,k}ds. \tag{31}$$

Then, Assumption 1 (e) implies that $|\Phi_{tT,k+1}| \leq r|\Phi_{tT,k}|$ almost surely in which $r = \rho(\|h\|_1)$. It turns out that the limit of the telescopic series $(\sum_{l=0}^{k} \Phi_{tT,l})_{k\in\mathbb{N}}$ exists by arguments used in Banach fixed-point theorem. Working with the telescopic series and applying the monotone convergence theorem to the series yield for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ that

$$\mathbb{E}\Big[\int_0^{tT} \lambda_s ds \Big| \mathcal{F}_1^{\nu}\Big] \quad \leq \quad \int_0^{t} \nu_s ds + r\mathbb{E}\Big[\int_0^{tT} \lambda_s ds \Big| \mathcal{F}_1^{\nu}\Big]. \tag{32}$$

By rearranging the terms in Expression (32), we get for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ that

$$\mathbb{E}\Big[\int_0^{tT} \lambda_s ds \Big| \mathcal{F}_1^{\nu}\Big] \leq (1-r)^{-1} \int_0^{t} \nu_s ds. \tag{33}$$

Given Condition 1 (b), the expression in the right side of Expression (33) is finite almost surely. As its conditional expectation is finite, we can deduce that $\int_0^{tT} \lambda_s ds$ is finite almost surely for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$. Moreover, the stochastic process $(\lambda_{tT})_{t\in[0,1]}$ is predictable with respect to the filtration $(\mathcal{F}_{tT})_{t\in[0,1]}$ as a limit of such processes. The point process $N_{tT}^{(i)}$ counts the points of the Poisson process $\underline{N}^{(2i-1)} * \underline{N}^{(2i)}$ under the curve $t \mapsto \lambda_{tT}^{(i)}$ for any time $t \in [0,1]$ by an application of the monotone convergence theorem. Thus, the point process $(N_{tT})_{t\in[0,1]}$ admits the stochastic process $(\lambda_{tT})_{t\in[0,1]}$ as an intensity with respect to the filtration $(\mathcal{F}_{tT})_{t\in[0,1]}$ in the sense of Definition 2. This is obtained by an extension to the stochastic time dependent baseline case of the arguments from Lemma 3 (p. 1571) in Brémaud and Massoulié (1996). Then, it implies that the point process N_{tT} is finite almost surely for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$.

Finally, it remains to show that the intensity process λ_{tT} is of the form (9) for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$. The monotonicity properties of the point process $N_{tT,k}^{(i)}$ and the intensity process $\lambda_{tT,k}^{(i)}$ ensure for any index $k \in \mathbb{N}$, any time $t \in [0,1]$, any final time $T \in \mathbb{R}^+$ and any index $i = 1,\dots,d$ that

$$\begin{aligned}
\lambda_{tT,k}^{(i)} \quad &\leq \quad \nu_t^{(i)} + \Big(\int_0^{tT} h(tT-s)dN_s\Big)^{(i)}, \\
\lambda_{tT}^{(i)} \quad &\geq \quad \nu_t^{(i)} + \Big(\int_0^{tT} h(tT-s)dN_{s,k}\Big)^{(i)}.
\end{aligned} \tag{34}$$

This gives Equation (9) by taking the limit $k \to +\infty$ in both inequalities. □

A.2. Proof of the central limit theorem

In this part, we focus on the proof of the central limit theorem for Hawkes processes with a stochastic and time dependent baseline. This corresponds to the proof of Theorem 1.

We start with the discretization in time of the statistical problem. For any final time $T \in \mathbb{R}_+$, we consider $M = \lfloor 1/\Delta \rfloor$ intervals with equal length $\Delta = 1/T$ such that $\bigcup_{i=1}^{M} [(i-1)\Delta, i\Delta) \subset [0,1)$. Here, $\lfloor \cdot \rfloor$ denotes the floor function. For any index $i = 1, \dots, M$, we define an estimator for local Poisson estimates on the ith interval $[(i-1)\Delta, i\Delta)$ as

$$\widehat{\lambda}_i = \frac{1}{\Delta}(N_{i\Delta-} - N_{(i-1)\Delta}). \tag{35}$$

Before giving the first lemma in the proof of Theorem 1, we introduce some definition. First, we define the Laplace transform of the kernel h of dimension $d \times d$ at the frequency $s \in \mathbb{R}_+$ as

$$\widehat{h}(s) = \int_0^\infty e^{-st} h(t) dt. \tag{36}$$

Then, we define the convolution of f and g for f and g two integrable functions defined on the space of positive real numbers $\mathbb{R}_+$ at time $t \in \mathbb{R}_+$ as

$$f * g_t = \int_0^t f(t-s) g(s)\, ds. \tag{37}$$

Moreover, we define recursively f^{*k} for any $k \in \mathbb{N}$ as $f^{*1} = f$ and f^{*k} is the convolution product of $f^{*(k-1)}$ and the function f for any $k \geq 2$. Similarly, we define the convolution of f and X for an integrable function f and a stochastic process X defined on the space of positive real numbers $\mathbb{R}_+$ at time $t \in \mathbb{R}_+$ as

$$f * dX_t = \int_0^t f(t-s)\, dX_s. \tag{38}$$

We denote by $\psi \colon \mathbb{R}_+ \to \mathbb{R}_+^d$ the resolvent kernel at the time $t \in \mathbb{R}^+$ of the kernel h which satisfies

$$\psi(t) = h(t) + h * \psi_t. \tag{39}$$

In mathematics, the resolvent formalism is a technique for applying concepts from complex analysis to the study of the spectrum of operators on Banach spaces and more general spaces. The resolvent captures the spectral properties of an operator in the analytic structure of the functional. In particular, the resolvent may be used to solve the inhomogeneous Fredholm integral equations. Then, we define the integral of the resolvent kernel ψ from the starting time 0 to the final time t as

$$\Psi(t) = \int_0^t \psi(s) ds. \tag{40}$$

We also define the integral of $\psi(s - tT)$ between the starting time $(i-1)\Delta$ and the final time $i\Delta$ for any time $t \in [0,1]$, any final time $T \in \mathbb{R}^+$ and any index $i = 1, \dots, M$ as

$$\Delta_i \Psi(-tT) = \int_{(i-1)\Delta}^{i\Delta} \psi(s - tT)\, ds. \tag{41}$$

Moreover, we denote the uniform big O by $\underline{O}$. It is defined through

$$f(tT) = \underline{O}\big(g(tT)\big) \iff |f(tT)| \leq Cg(tT)$$

for any time $t \in [0,1]$, any final time $T \in \mathbb{R}^+$ and some constant $C \in \mathbb{R}_+$ which does not depend on the final time T and the time t. Finally, we introduce $a \wedge b$ which is the minimum between two real numbers $a \in \mathbb{R}$ and $b \in \mathbb{R}$.

The first lemma gives the asymptotic properties of the resolvent kernel, which can be expressed as a Laplace transform of the kernel.

Lemma 2. *We assume that Assumption 1 holds. Then, we have for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ that*

$$\psi(tT) \quad \geq \quad 0. \tag{42}$$

We also assume that Assumption 2 (a) holds. Then, we have that

$$\Psi(tT) \quad = \quad \widehat{h}(0) + \underline{O}\bigg(1 \wedge \frac{1}{tT}\bigg). \tag{43}$$

Moreover, we have for any index $i = 1, \ldots, M$ that

$$\Delta_i \Psi(-tT) \quad = \quad \underline{O}\bigg(1 \wedge \frac{1}{((i-1)\Delta - tT)}\bigg). \tag{44}$$

Proof of Lemma 2. Since $\rho(\|h\|_1) < 1$ by Assumption 1 (e), we can apply Banach fixed-point theorem to the function

$$\theta\colon f(tT) \mapsto (\nu_t + h * f_{tT}).$$

Thus, we can get a fixed-point $\psi = f_\infty$ with recursion $f_k = \theta(f_{k-1})$. Then, we obtain recursively by the definitions of the function θ and the function f_k that

$$\begin{aligned} f_k(tT) &= \theta(f_{k-1}(tT)) = \nu_t + h * (f_{k-1})_{tT} = \nu_t + h * \theta(f_{k-2})_{tT} \\ &= \nu_t + h * (\nu_t + h * (f_{k-2})_{tT})_{tT} \\ &= \nu_t + \sum_{l=1}^{k-2} h^{*l} * \nu_t + h^{*(k-1)} * (f_1)_{tT}. \end{aligned}$$

For the initial value, we can choose $f_1 = 0$. Then, $f_k(tT)$ is nonnegative for any $k \in \mathbb{N}$ such that $k > 1$, any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$. This is due to the fact that the baseline is almost surely positive on the time interval $[0,1]$ almost everywhere by Assumption 1 (a) and the kernel h is nonnegative by Assumption 1 (d). Thus, we have shown Expression (42).

To show Equation (43), we first have for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ that

$$\begin{aligned} \widehat{h}(0) - \widehat{h}\bigg(\frac{1}{tT}\bigg) &= \int_0^\infty \big(1 - e^{-\frac{s}{tT}}\big)h(s)ds \\ &\geq \int_{tT}^\infty \big(1 - e^{-\frac{s}{tT}}\big)h(s)ds \\ &\geq (1 - e^{-1}) \int_{tT}^\infty \psi(s)ds. \end{aligned} \tag{45}$$

Here, we use the definition of the Laplace transform in the equality, Assumption 1 (d) in the first inequality, Definition (39) and Expression (42) in the last inequality.

From Assumption 2 (a), we obtain that the Laplace transform of the kernel $\widehat{h}$ is continuously differentiable at the time 0 and that its derivative is equal to

$$\widehat{h}'(0) = \int_0^\infty th(t)dt < \infty.$$

Then, we can apply the mean value theorem for the function $\widehat{\psi}(s) = \widehat{h}(s)/(1 - \widehat{h}(s))$ and we obtain that

$$\widehat{\psi}(s) = \widehat{\psi}(0) + \widehat{\psi}'(0)s + r(s)s. \tag{46}$$

Here, $r(s)$ is the remainder which satisfies $\lim_{s\to 0} r(s) = 0$.

Then, we can deduce for any time $t \in [0, 1]$ and any final time $T \in \mathbb{R}_+$ that

$$\begin{aligned} 0 &\leq \widehat{\psi}(0) - \widehat{\psi}\Big(\frac{1}{tT}\Big) \\ &= -\Big(\widehat{\psi}'(0) + r\Big(\frac{1}{tT}\Big)\Big)\frac{1}{tT} \\ &\leq \big(|\widehat{\psi}'(0)| + \big|r\big(\frac{1}{tT}\big)\big|\big)\frac{1}{tT}\mathbf{1}_{\{tT\geq 1\}} + \widehat{\psi}(0)\mathbf{1}_{\{tT<1\}}. \end{aligned}$$

Here, we use the fact that the function $\widehat{\psi}$ is decreasing in the first inequality, Equation (46) in the equality and the definition of $|\ |$ in the last inequality. As

$$\sup_{x\in[0,1]} |r(x)| < \infty,$$

we obtain

$$\Big|\widehat{\psi}(0) - \widehat{\psi}\Big(\frac{1}{tT}\Big)\Big| \leq C\Big(1 \wedge \frac{1}{tT}\Big). \tag{47}$$

Finally, we get for any time $t \in [0, 1]$ and any final time $T \in \mathbb{R}^+$ that

$$\begin{aligned} \Psi(tT) &= \int_0^{tT} \psi(s)\, ds \\ &= \widehat{\psi}(0) - \int_{tT}^\infty \psi(s)\, ds \\ &\leq \widehat{\psi}(0) + \frac{1}{1-e^{-1}}\Big(\widehat{\psi}(0) - \widehat{\psi}\Big(\frac{1}{tT}\Big)\Big) \\ &\leq \widehat{\psi}(0) + C\Big(1 \wedge \frac{1}{tT}\Big). \end{aligned}$$

Here, we use Definition (40) in the first equality, the definition of $\widehat{\psi}(0)$ in the second equality, Expression (45) in the first inequality and Expression (47) in the last inequality. Thus, we have proven Equation (43). With the same arguments, we can also show that Equation (44) holds. □

We next state the renewal equation for stochastic processes in the following lemma. This extends Lemma 3 from Bacry et al. (2013) in which the function from the renewal equation is nonrandom. This is required as the baseline ν is stochastic and time dependent.

Lemma 3. *We assume that Assumption 1 holds. We introduce a stochastic process $g : \mathbb{R}_+ \times \Omega \to \mathbb{R}^d$ adapted to the filtration $\overline{\mathcal{F}}$ and which is locally bounded almost surely. Then, there exists a stochastic process $f : \mathbb{R}^+ \times \Omega \to \mathbb{R}^d$ which is locally bounded almost surely and solution to the renewal equation for any time $t \in \mathbb{R}^+$ and almost surely*

$$f(t) = g(t) + h * f_t. \tag{48}$$

The solution is given for any time $t \in \mathbb{R}_+$ and almost surely by

$$f_g(t) = g(t) + \psi * g_t. \tag{49}$$

Moreover, the solution f_g is unique in the almost surely and almost everywhere sense. Namely, we have

$$\mathbb{P}\big(f_g(t) = f(t) \text{ for any time } t \in \mathbb{R}_+ \text{ and almost everywhere}\big) = 1 \tag{50}$$

for any stochastic process f satisfying the renewal equation (48).

Proof of Lemma 3. First, we have that the resolvent kernel of the kernel ψ is integrable by Assumption 1 (e) and resolvent kernel properties. We also have that the stochastic process g is locally bounded almost surely by assumption of the lemma. Thus, we can deduce that the stochastic process f_g defined in Equation (49) is locally bounded almost surely.

Moreover, we show in what follows that f_g satisfies the renewal equation (48). It is sufficient to prove for any time $t \in \mathbb{R}_+$ and almost surely that

$$h * (f_g)_t = \psi * g_t \tag{51}$$

First, we get by Definition (49) for any time $t \in \mathbb{R}_+$ and almost surely that

$$h * (f_g)_t = h * (g(t) + \psi * g_t)_t. \tag{52}$$

Moreover, we obtain by the definition of the resolvent kernel (39) for any time $t \in \mathbb{R}_+$ that

$$h(t) = \psi(t) - h * \psi_t. \tag{53}$$

Finally, Equations (52) and (53) yield Equation (51).

We show now that the solution f_g is unique in the almost surely and almost everywhere sense. Namely, we show Equation (50) for any stochastic process f satisfying the renewal equation (48). First, we get as both processes f_g and f satisfy the renewal equation (48) for any time $t \in \mathbb{R}_+$ and almost surely that

$$f(t) - f_g(t) = h * (f - f_g)_t. \tag{54}$$

We introduce the stochastic process v of dimension d such that its ith component for any index $i = 1, \dots, d$ is equal to

$$v^{(i)}(t) = |f^{(i)}(t) - f_g^{(i)}(t)|.$$

Then, we can deduce from Equation (54) for any time $t \in \mathbb{R}_+$ and almost surely that

$$v(t) = h * v_t. \tag{55}$$

This yields for any time $t \in \mathbb{R}_+$ and almost surely that

$$\int_0^\infty v(t)dt \leq ||h||_1 \int_0^\infty v(t)dt. \tag{56}$$

Finally, we get by Assumption 1 (d) for any time $t \in \mathbb{R}_+$ and almost surely that

$$\int_0^\infty v(t)dt < \int_0^\infty v(t)dt. \tag{57}$$

Thus, we have shown Equation (50) for any stochastic process f satisfying the renewal equation (48).

□

Before introducing the next lemma, we give some definition. First, we denote the sum of the baseline and the convolution of the resolvent kernel and the baseline for any time $t \in [0, 1]$ by

$$\mu_t = \nu_t + (\psi_{Tt}) * \nu_t. \tag{58}$$

We have that the stochastic process μ is adapted to the filtration $(\mathcal{F}^\nu)_{t\in[0,1]}$ by Definition (38). Moreover, we define the limit of the stochastic process μ_t for any time $t \in [0, 1]$ as

$$\mu_{t,L} = \big((I - \|h\|_1)^{-1}\mathbf{1}_{\{t\in(0,1]\}} + \mathbf{1}_{\{t=0\}}\big)\nu_t \tag{59}$$

Finally, we define the notation small tau in probability as $Y_T = o_{\mathbb{P}}(Z_T)$, which means that $\frac{Y_T}{Z_T}\mathbf{1}_{\{Z_T\neq 0\}} \xrightarrow{\mathbb{P}} 0$ as the final time $T \to \infty$ for Y_T and Z_T which are random variables.

The following lemma exhibits a martingale representation of the intensity $(\lambda_{tT})_{t\in[0,1]}$. It is based on the convolution of the resolvent kernel and the martingale $(M_{tT})_{t\in[0,1]}$ introduced in Definition (7). The proof is based on the application of the renewal equation obtained in Lemma 3. It extends Lemma 4 in Bacry et al. (2013) and Proposition 2.1 (p. 606) in Jaisson and Rosenbaum (2015), who consider a constant and nonrandom baseline.

Lemma 4. *We assume that Assumptions 1 and 2 hold. Then, we have for any time $t \in [0, 1]$ and any final time $T \in \mathbb{R}_+$ that the intensity λ_{tT} has the martingale representation*

$$\lambda_{tT} = \mu_t + \psi * dM_{tT}. \tag{60}$$

We also have for any time $t \in [0,1]$ when the final time $T \to \infty$ that

$$\sup_{T^{-\frac{1}{2}} \leq t \leq 1} \sqrt{T}\big(\mu_t - \mu_{t,L}\big) \xrightarrow{\mathbb{P}} 0. \tag{61}$$

Proof of Lemma 4. We first reexpress the intensity process as the renewal equation. We have for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ that

$$\begin{aligned} \lambda_{tT} &= \nu_t + h * dN_{tT} \\ &= \nu_t + h * (\lambda_{tT} + dM_{tT}) \\ &= \big(\nu_t + h * dM_{tT}\big) + h * \lambda_{tT}. \end{aligned} \tag{62}$$

Here, we use Definition (9) and Definition (38) in the first equality, Definition (7) in the second equality and algebraic manipulation in the third equality.

Thus, the intensity process $(\lambda_{tT})_{t\in[0,1]}$ is solution to the renewal equation (48) almost surely by Equation (62). More specifically, we consider the stochastic processes $g(tT)$ defined as

$$g(tT) = \nu_t + h * dM_{tT}$$

for any time $t \in [0,1]$ and a fixed $T \in \mathbb{R}_+$. To obtain the form of the intensity process $(\lambda_{tT})_{t\in[0,1]}$, we apply Lemma 3. First, we have that $g(tT)$ are stochastic processes adapted to the filtration $(\mathcal{F}_{tT})_{t\in[0,1]}$ which are locally bounded almost surely by Definition 2, Assumptions 2 (b) and 2 (c). Then, we have for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ that

$$\begin{aligned} \lambda_{tT} &= g(tT) + \psi * g_{tT} \\ &= \nu_t + h * dM_{tT} + (\psi_{Tt}) * (\nu_t + h * dM_{tT}) \\ &= (\nu_t + (\psi_{Tt}) * \nu_t) + (h + \psi * h_{tT}) * dM_{tT}. \end{aligned}$$

Here, we use Equation (49) from Lemma 3 in the first equality, the definition of the stochastic process $(g(tT))_{t\in[0,1]}$ in the second equality, algebraic manipulation in the third equality. Then, we have for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ that

$$\begin{aligned} \lambda_{tT} &= (\nu_t + (\psi_{Tt}) * \nu_t) + (h + \psi * h_{tT}) * dM_{tT} \\ &= (\nu_t + (\psi_{Tt}) * \nu_t) + \psi * dM_{tT}. \\ &= \mu_t + \psi * dM_{tT}. \end{aligned}$$

Here, we use Definition (39) in the second equality and Definition (58) in the third equality. Thus, we can obtain Equation (60).

We show now Equation (61). First, we get for any time $t \in (0,1]$ and any final time $T \in \mathbb{R}_+$ that

$$\begin{aligned} \mu_t - \mu_{t,L} &= \nu_t + (\psi_{Tt}) * \nu_t - (I - \|h\|_1)^{-1} \nu_t \\ &= \nu_t + (\psi_{Tt}) * \nu_t - (1 + \widehat{\psi}(0)) \nu_t \\ &= (\psi_{Tt}) * \nu_t - \widehat{\psi}(0) \nu_t \\ &= \int_0^{tT} \psi(s) \nu_{t - \frac{s}{T}} ds - \widehat{\psi}(0) \nu_t. \end{aligned}$$

Here, we use Definition (58) and Definition (59) in the first equality, Definition (39) in the second equality, algebraic manipulation in the third equality and Definition (37) in the fourth equality. Then, we have for any time $t \in (0, 1]$ and any final time $T \in \mathbb{R}_+$ that

$$\begin{aligned} \mu_t - \mu_{t,L} &= \int_0^{tT} \psi(s)\nu_{t-\frac{s}{T}} ds - \widehat{\psi}(0)\nu_t \\ &= \int_0^{tT} \psi(s)\nu_{t-\frac{s}{T}} ds - \int_0^{\infty} \psi(s) ds \nu_t \\ &= \int_0^{tT} \psi(s)\big(\nu_{t-\frac{s}{T}} - \nu_t\big) ds - \int_{tT}^{\infty} \psi(s) ds. \end{aligned} \tag{63}$$

Here, we use Definition (36) in the second equality and algebraic manipulation in the third equality. Moreover, we obtain that

$$\begin{aligned} \sup_{T^{-\frac{1}{2}} \leq t \leq 1} \sqrt{T}\big|\mu_t - \mu_{t,L}\big| &= \sup_{T^{-\frac{1}{2}} \leq t \leq 1} \sqrt{T}\Big| \int_0^{tT} \psi(s)(\nu_{t-\frac{s}{T}} - \nu_t) ds - \int_{tT}^{\infty} \psi(s) ds\Big| \\ &\leq \sup_{T^{-\frac{1}{2}} \leq t \leq 1} \sqrt{T}\Big(\Big| \int_0^{tT} \psi(s)(\nu_{t-\frac{s}{T}} - \nu_t) ds\Big| + \Big| \int_{tT}^{\infty} \psi(s) ds\Big|\Big) \\ &\leq \sup_{T^{-\frac{1}{2}} \leq t \leq 1} \sqrt{T}\Big| \int_0^{tT} \psi(s)(\nu_{t-\frac{s}{T}} - \nu_t) ds\Big| \\ &\quad + \sup_{T^{-\frac{1}{2}} \leq t \leq 1} \sqrt{T}\Big| \int_{tT}^{\infty} \psi(s) ds\Big|. \end{aligned} \tag{64}$$

Here, we use Equation (63) in the first equality, the triangular inequality in the first inequality and supremum properties in the second inequality.

We introduce

$$I = \sup_{T^{-\frac{1}{2}} \leq t \leq 1} \sqrt{T}\Big| \int_0^{tT} \psi(s)(\nu_{t-\frac{s}{T}} - \nu_t) ds\Big|.$$

First, we can deduce from integral and norm properties that

$$I \leq \sup_{T^{-\frac{1}{2}} \leq t \leq 1} \sqrt{T} \int_0^{tT} \psi(s)\big|\nu_{t-\frac{s}{T}} - \nu_t\big| ds.$$

Then, we get from Assumption 2 (c) almost surely that

$$I \leq \sup_{T^{-\frac{1}{2}} \leq t \leq 1} \sqrt{T} \int_0^{tT} \psi(s) \frac{Cs}{T} ds.$$

We also obtain from algebraic manipulation almost surely that

$$I \leq \sup_{T^{-\frac{1}{2}} \leq t \leq 1} \frac{C}{\sqrt{T}} \int_0^{tT} \psi(s) s ds.$$

Moreover, supremum properties yield almost surely that

$$I \leq \frac{C}{\sqrt{T}} \int_0^\infty \psi(s) s ds.$$

Finally, we obtain by Assumption 2 (a) and Definition (39) when the final time $T \to \infty$ that

$$I \xrightarrow{\mathbb{P}} 0. \tag{65}$$

We introduce now

$$II = \sup_{T^{-\frac{1}{2}} \leq t \leq 1} \sqrt{T} \Big| \int_{tT}^\infty \psi(s) ds \Big|.$$

First, we can deduce from supremum properties and Expression (42) from Lemma 2 that

$$II = \sqrt{T} \Big| \int_{\sqrt{T}}^\infty \psi(s) ds \Big|.$$

Then, we obtain by algebraic manipulation that

$$II = \Big| \int_{\sqrt{T}}^\infty \sqrt{T} \psi(s) ds \Big|.$$

In addition, we get by supremum properties that

$$II \leq \Big| \int_{\sqrt{T}}^\infty s \psi(s) ds \Big|.$$

Finally, we obtain by Assumption 2 (a) and Definition (39) when the final time $T \to \infty$ that

$$II \xrightarrow{\mathbb{P}} 0. \tag{66}$$

Thus, Expressions (64), (65) and (66) yield Equation (61).

□

We introduce some notation prior to the next lemma. First, we denote by X_i the process X evaluated at the end of the ith interval for any interval index $i = 1, \ldots, M$, namely $X_i = X_{i\Delta}$. We also define $\bar{X}_i$ as the average of the stochastic process X on the ith interval, namely

$$\bar{X}_i = \Delta^{-1} \int_{(i-1)\Delta}^{i\Delta} X_t \, dt. \tag{67}$$

In addition, we define the increment of the martingale M on the ith interval as

$$\varepsilon_i = \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} dM_t. \tag{68}$$

Moreover, we denote the increment related to the Hawkes component on the ith interval by

$$\epsilon_i = \frac{1}{\Delta}\Big(\int_0^{(i-1)\Delta} \Delta_i \Psi(-t)\, dM_t + \int_{(i-1)\Delta}^{i\Delta} \Psi(i\Delta - t)\, dM_t \Big). \tag{69}$$

Finally, we define the sum of ε_i and ϵ_i as u_i, namely

$$u_i = \varepsilon_i + \epsilon_i. \tag{70}$$

The following lemma is a decomposition of the estimation error u_i as the sum of the error originating from the martingale ε_i and another related to the Hawkes component ϵ_i for any interval index $i = 1, \dots, M$.

Lemma 5. *We assume that Assumptions 1 and 2 hold. Then, we have for any final time $T \in \mathbb{R}_+$ and any interval index $i = 1, \dots, M$ the decomposition*

$$\widehat{\lambda}_i = \bar{\mu}_i + u_i. \tag{71}$$

We also have when the final time $T \to \infty$ that

$$\sup_{i \in \mathbb{N} \text{ such that } \lfloor \sqrt{T} \rfloor < i \leq M} \sqrt{T}\big(\widehat{\lambda}_i - \bar{\mu}_{i,L} - u_i\big) \xrightarrow{\mathbb{P}} 0. \tag{72}$$

Proof of Lemma 5. First, we have by Definition (35) for any index $i = 1, \dots, M$ that

$$\widehat{\lambda}_i = \frac{1}{\Delta}(N_{i\Delta-} - N_{(i-1)\Delta}).$$

Then, we obtain by Definition (7) for any index $i = 1, \dots, M$ that

$$\widehat{\lambda}_i = \frac{1}{\Delta}(M_{i\Delta-} - M_{(i-1)\Delta}) + \bar{\lambda}_i.$$

In addition, we get by Definition (68) for any index $i = 1, \dots, M$ that

$$\widehat{\lambda}_i = \varepsilon_i + \bar{\lambda}_i.$$

Moreover, we get by Equation (60) from Lemma 4 for any index $i = 1, \dots, M$ that

$$\widehat{\lambda}_i = \varepsilon_i + \bar{\mu}_i + \frac{1}{\Delta}\psi * (M_{i\Delta-} - M_{(i-1)\Delta}).$$

Furthermore, we can deduce by Definitions (38) , (40) and (41) for any index $i = 1, \dots, M$ that

$$\widehat{\lambda}_i = \varepsilon_i + \bar{\mu}_i + \frac{1}{\Delta}\Big(\int_0^{(i-1)\Delta} \Delta_i \Psi(-t)\, dM_t + \int_{(i-1)\Delta}^{i\Delta} \Psi(i\Delta - t)\, dM_t \Big).$$

This yields Equation (71) with the use of Definition (69). Finally, we obtain Equation (72) by Equation (61) from Lemma 4. □

We introduce the rescaled error process X which for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ satisfies

$$X_t = \sqrt{T}\big(\widehat{\Lambda}_t - \Lambda_t\big). \tag{73}$$

We also introduce ξ_i which is a random vector of dimension d defined for any interval index $i = 1, \dots, M$ by

$$\xi_i = \sqrt{T} u_i. \tag{74}$$

Moreover, we define the time discretized filtration for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}^+$ as

$$\mathbf{F}_t = \mathcal{F}_{\Delta \lfloor tT \rfloor}.$$

Furthermore, we use $\mathbb{E}_{i-1}$, Var_{i-1} and Cov_{i-1} instead of using $\mathbb{E}[\cdot|\mathcal{F}_{(i-1)\Delta}]$, $\mathrm{Var}[\cdot|\mathcal{F}_{(i-1)\Delta}]$ and $\mathrm{Cov}[\cdot|\mathcal{F}_{(i-1)\Delta}]$ for any interval index $i = 1, \dots, M$. Finally, we define the notation small tau in probability uniformly for any time $t \in [0,1]$ as $Y_T(t) = o^u_{\mathbb{P}}(Z_T(t))$, which means that

$$\sup_{0 \le t \le 1} \left| \frac{Y_T(t)}{Z_T(t)} \mathbf{1}_{\{Z_T(t) \neq 0\}} \right| \xrightarrow{\mathbb{P}} 0$$

when the final time $T \to \infty$ for Y_T and Z_T which are a family of stochastic processes.

The following lemma gives a discretization in time of the rescaled error process X based on the random vectors ξ. It also reexpresses the rescaled error process X as the sum of a martingale with respect to the filtration $\mathbf{F}$ and another random variable. To get the martingale with respect to the filtration $\mathbf{F}$, we compensate the random variables ξ_i by their conditional expectations $\mathbb{E}_{i-1}[\xi_i]$ for any interval index $i = 1, \dots, M$.

Lemma 6. *We assume that Assumptions 1 and 2 hold. Then, we can discretize the rescaled error process X_t for any time $t \in [0,1]$ when the final time $T \to \infty$ as*

$$X_t = \sum_{i=1}^{\lfloor tT \rfloor} \xi_i + o^u_{\mathbb{P}}(1). \tag{75}$$

We can also reexpress the rescaled error process X_t for any time $t \in [0,1]$ when the final time $T \to \infty$ as

$$X_t = \sum_{i=1}^{\lfloor tT \rfloor} \big(\xi_i - \mathbb{E}_{i-1}[\xi_i]\big) + \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}[\xi_i] + o^u_{\mathbb{P}}(1). \tag{76}$$

Proof of Lemma 6. First, we get by Definition (73) for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ that

$$X_t = \sqrt{T}\big(\widehat{\Lambda}_t - \Lambda_t\big).$$

We also can deduce by Definition (12) for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ that

$$X_t = \sqrt{T}\bigg(\frac{N_{tT}}{T} - \Lambda_t\bigg).$$

Then, this can be rewritten by an algebraic manipulation for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ as

$$X_t = \sqrt{T}\bigg(\sum_{i=1}^{\lfloor tT \rfloor} \frac{(N_{i\Delta-} - N_{(i-1)\Delta})}{T} - \Lambda_t\bigg).$$

In addition, we obtain by Definition (35) for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ that

$$X_t = \sqrt{T}\bigg(\sum_{i=1}^{\lfloor tT \rfloor} \widehat{\lambda}_i - \Lambda_t\bigg).$$

Moreover, we get by an algebraic manipulation for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ that

$$X_t = \sqrt{T}\Bigg(\sum_{i=1}^{\lfloor \sqrt{T} \rfloor} \widehat{\lambda}_i + \sum_{i=\lfloor \sqrt{T} \rfloor + 1}^{\lfloor tT \rfloor} \widehat{\lambda}_i - \Lambda_t\Bigg).$$

Finally, we get by another algebraic manipulation for any time $t \in [0,1]$ and any final time $T \in \mathbb{R}_+$ that

$$X_t = \sqrt{T}\bigg(\sum_{i=1}^{\lfloor \sqrt{T} \rfloor} \widehat{\lambda}_i - \Lambda_{\sqrt{T}}\bigg) + \sqrt{T}\Bigg(\sum_{i=\lfloor \sqrt{T} \rfloor + 1}^{\lfloor tT \rfloor} \widehat{\lambda}_i - \Lambda_t + \Lambda_{\sqrt{T}}\Bigg).$$

Now, this yields as $1/\sqrt{T}$ is negligible when the final time $T \to \infty$ for any time $t \in [0,1]$ that

$$X_t = \sqrt{T}\Bigg(\sum_{i=\lfloor \sqrt{T} \rfloor + 1}^{\lfloor tT \rfloor} \widehat{\lambda}_i - \Lambda_t + \Lambda_{\sqrt{T}}\Bigg) + o^u_{\mathbb{P}}(1).$$

This can be rewritten by Definitions (11) and (59) when the final time $T \to \infty$ for any time $t \in [0,1]$ as

$$X_t = \sqrt{T} \sum_{i=\lfloor \sqrt{T} \rfloor + 1}^{\lfloor tT \rfloor} \big(\widehat{\lambda}_i - \bar{\mu}_{i,L}\big) + o^u_{\mathbb{P}}(1).$$

Then, we can deduce by Expression (72) from Lemma 5 when the final time $T \to \infty$ for any time $t \in [0,1]$ that

$$X_t = \sqrt{T} \sum_{i=\lfloor \sqrt{T} \rfloor + 1}^{\lfloor tT \rfloor} u_i + o_{\mathbb{P}}^u(1).$$

Moreover, this yields as $1/\sqrt{T}$ is negligible when the final time $T \to \infty$ for any time $t \in [0,1]$ that

$$X_t = \sqrt{T} \sum_{i=1}^{\lfloor tT \rfloor} u_i + o_{\mathbb{P}}^u(1).$$

Finally, we get Equation (75) from Definition (74). Then, we can deduce Equation (76) by algebraic manipulation and since u_i for any interval index $i = 1, \dots, M$ is integrable from Lemma 2, Assumptions 2 (b) and (c). □

As the rescaled error process X takes a martingale form in Equation (76) from Lemma 6, we can use the toolkit from central limit theorems relying on triangular array of martingale increments. More specifically, the proof of Theorem 1 is based on an application of Theorem IX.7.28 (pp. 590-591) in Jacod and Shiryaev (2003). We first give its statement in the theorem which follows.

Theorem 2 (Theorem IX.7.28 in Jacod and Shiryaev (2003))**.** *We assume that the stochastic process Z is a continuous d-dimensional local martingale such that $\mathbb{E}(|Z_t|^2) < \infty$ for any time $t \in [0,1]$ and that each ξ_i for any index $i = 1, \dots, M$ is square-integrable. Let $\widetilde{C}$, G and B be continuous stochastic processes adapted to the filtration $(\mathcal{F}_{tT})_{t\in[0,1]}$, null at 0 and taking values in the space of d'-dimensional semimartingales, $\mathbb{R}^q \otimes \mathbb{R}^{d'}$ and $\mathbb{R}^q$ respectively. If for any time $t > 0$, any $u > 0$, any $1 \le j, k \le q$, any $1 \le l \le d'$ and any uniformly integrable bounded martingale N which are orthogonal to all components of Z we have*

$$\sup_{0 \le t \le 1} \left| \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}[\xi_i] - B_t \right| \xrightarrow{\mathbb{P}} 0, \tag{77}$$

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathrm{Cov}_{i-1}[\xi_i] \xrightarrow{\mathbb{P}} \widetilde{C}_t = \int_0^t c_u c_u^T \, du, \tag{78}$$

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}[\xi_i^T \Delta_i M'] \xrightarrow{\mathbb{P}} G_t, \tag{79}$$

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\left[|\xi_i|^2 \mathbf{1}_{\{|\xi_i| > u\}}\right] \xrightarrow{\mathbb{P}} 0. \tag{80}$$

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}[\xi_i^T \Delta_i M'] \xrightarrow{\mathbb{P}} 0. \tag{81}$$

Then, there is a canonical d-dimensional standard Brownian extension of the stochastic basis $\mathcal{B}$. This extension includes the canonical standard Brownian motion W_t which satisfies as $T \to \infty$ that

$$\sqrt{T}\big(\widehat{\Lambda}_t - \Lambda_t\big) \xrightarrow{\mathcal{D}-s} \int_0^t c_s dW_s.$$

We first show that Condition (77) (Condition (7.27) in Theorem IX.7.28 from Jacod and Shiryaev (2003)) holds with $B_t = 0$ in the following lemma. This proves that the sum of the conditional expectations $\mathbb{E}_{i-1}[\xi_i]$ converges to 0 in probability and uniformly in time.

Lemma 7. *We assume that Assumptions 1 and 2 hold. Then, we have when the final time $T \to \infty$ that*

$$\sup_{0\leq t\leq 1} \left| \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}[\xi_i] \right| \xrightarrow{\mathbb{P}} 0.$$

Proof of Lemma 7. We first obtain by Definition (74) that

$$\sup_{0\leq t\leq 1} \left| \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}[\xi_i] \right| = \sup_{0\leq t\leq 1} \left| \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\Big[\sqrt{T}u_i\Big] \right|. \tag{82}$$

Then, we can deduce by Definition (70) that

$$\sup_{0\leq t\leq 1} \left| \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\Big[\sqrt{T}u_i\Big] \right| = \sup_{0\leq t\leq 1} \left| \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\Big[\sqrt{T}(\varepsilon_i + \epsilon_i)\Big] \right|. \tag{83}$$

We introduce

$$I = \sup_{0\leq t\leq 1} \left| \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\Big[\sqrt{T}\varepsilon_i\Big] \right|$$

and

$$II = \sup_{0\leq t\leq 1} \left| \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\Big[\sqrt{T}\epsilon_i\Big] \right|$$

We get by supremum properties that

$$\sup_{0\leq t\leq 1} \left| \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\Big[\sqrt{T}u_i\Big] \right| \leq I + II. \tag{84}$$

For the first term I, we can deduce by Definition (68) that

$$I = \sup_{0\leq t\leq 1} \left| \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\Big[\frac{\sqrt{T}}{\Delta} \int_{(i-1)\Delta}^{i\Delta} dM_t\Big] \right|.$$

This leads as the stochastic process M is a martingale with respect to the filtration $\overline{\mathcal{F}}$ to

$$I = 0. \tag{85}$$

For the second term II, we can deduce by Definition (69) that

$$II = \sup_{0\leq t\leq 1}\Big|\sum_{i=1}^{\lfloor tT\rfloor}\mathbb{E}_{i-1}\Big[\frac{\sqrt{T}}{\Delta}\Big(\int_0^{(i-1)\Delta}\Delta_i\Psi(-t)\,dM_t+\int_{(i-1)\Delta}^{i\Delta}\Psi(i\Delta-t)\,dM_t\Big)\Big]\Big|.$$

This leads as the stochastic process M is a martingale with respect to the filtration $\overline{\mathcal{F}}$ and by Lemma 2 when the final time $T\to\infty$ to

$$II \xrightarrow{\mathbb{P}} 0. \tag{86}$$

Moreover, we get from Expressions (84), (85) and (86) when the final time $T\to\infty$ that

$$\sup_{0\leq t\leq 1}\Big|\sum_{i=1}^{\lfloor tT\rfloor}\mathbb{E}_{i-1}\Big[\sqrt{T}u_i\Big]\Big| \xrightarrow{\mathbb{P}} 0. \tag{87}$$

Finally, Expressions (82), (83) and (87) yield the lemma. □

We show that Condition (78) (Condition (7.28) in in Theorem IX.7.28 from Jacod and Shiryaev (2003)) holds with $\widetilde{C}_t = \int_0^t c_u c_u^T$ in the following lemma. This proves that the sum of the covariances converges to the asymptotic covariance in probability.

Lemma 8. *We assume that Assumptions 1 and 2 hold. Then, we have for any time $t\in[0,1]$ when the final time $T\to\infty$ that*

$$\sum_{i=1}^{\lfloor tT\rfloor}\mathrm{Cov}_{i-1}[\xi_i] \xrightarrow{\mathbb{P}} \int_0^t c_u c_u^T\,du.$$

Proof of Lemma 8. First, the random vectors ξ_i^2 are integrable from Lemma 2, Assumptions 2 (b) and (c). Then, covariance properties yield for any time $t\in[0,1]$ and any final time $T\in\mathbb{R}_+$ that

$$\sum_{i=1}^{\lfloor tT\rfloor}\mathrm{Cov}_{i-1}[\xi_i] = \sum_{i=1}^{\lfloor tT\rfloor}\mathbb{E}_{i-1}[\xi_i\xi_i^T] - \sum_{i=1}^{\lfloor tT\rfloor}\mathbb{E}_{i-1}[\xi_i]\mathbb{E}_{i-1}[\xi_i^T]. \tag{88}$$

In addition, we obtain for the first term in the right side of Equation (88) by an extension of the arguments from the proof of Corollary 1 in Bacry et al. (2013) with Assumptions 1 (e) and 2 for any time $t\in[0,1]$ when the final time $T\to\infty$ that

$$\sum_{i=1}^{\lfloor tT\rfloor}\mathbb{E}_{i-1}[\xi_i\xi_i^T] \xrightarrow{\mathbb{P}} \int_0^t c_u c_u^T\,du. \tag{89}$$

To deal with the second term in the right side of Equation (88), we use Burkholder-Davis-Gundy inequalites (see Expression (2.1.32) in Jacod and Protter (2012) (p. 39)) with Assumptions 2 (b) and (c). This yields for any time $t \in [0,1]$ when the final time $T \to \infty$ that

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}[\xi_i]\mathbb{E}_{i-1}[\xi_i^T] \xrightarrow{\mathbb{P}} 0. \tag{90}$$

Finally, Expressions (88), (89) and (90) lead to the proposition. □

We show now that Condition (80) (Condition (7.30) in Theorem IX.7.28 from Jacod and Shiryaev (2003)) holds in the proposition that follows. This proves the Lindeberg condition for the central limit theorem.

Lemma 9. *We assume that Assumptions 1 and 2 hold. Then, we have for any time $t \in [0,1]$ and any $u \in \mathbb{R}_+$ satisfying $u > 0$ when the final time $T \to \infty$ that*

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\left[|\xi_i|^2 \mathbf{1}_{\{|\xi_i|>u\}}\right] \xrightarrow{\mathbb{P}} 0.$$

Proof of Lemma 9. Since convergence in L^1 implies convergence in probability, it is sufficient to show the convergence in L^1. More specifically, we prove for any time $t \in [0,1]$ and any $u \in \mathbb{R}_+$ satisfying $u > 0$ when the final time $T \to \infty$ that

$$\mathbb{E}\Big[\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\left[|\xi_i|^2 \mathbf{1}_{\{|\xi_i|>u\}}\right]\Big] \to 0. \tag{91}$$

First, we have by linearity of the expectation for any time $t \in [0,1]$ and any $u \in \mathbb{R}_+$ satisfying $u > 0$ that

$$\mathbb{E}\Big[\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}\left[|\xi_i|^2 \mathbf{1}_{\{|\xi_i|>u\}}\right]\Big] = \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}\Big[\mathbb{E}_{i-1}\left[|\xi_i|^2 \mathbf{1}_{\{|\xi_i|>u\}}\right]\Big]. \tag{92}$$

Then, we get by conditional expectation properties for any time $t \in [0,1]$ and any $u \in \mathbb{R}_+$ satisfying $u > 0$ that

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}\Big[\mathbb{E}_{i-1}\left[|\xi_i|^2 \mathbf{1}_{\{|\xi_i|>u\}}\right]\Big] = \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}\left[|\xi_i|^2 \mathbf{1}_{\{|\xi_i|>u\}}\right]. \tag{93}$$

Moreover, we can deduce by Hölder's inequality for any time $t \in [0,1]$ and any $u \in \mathbb{R}_+$ satisfying $u > 0$ that

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}\left[|\xi_i|^2 \mathbf{1}_{\{|\xi_i|>u\}}\right] \leq \sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}\left[|\xi_i|^{2+\eta}\right]^{\frac{1}{2+\eta}} \mathbb{E}\left[\mathbf{1}_{\{|\xi_i|>u\}}\right]^{\frac{1}{q}}. \tag{94}$$

Here, we have that $q \in \mathbb{R}$ which satisfies

$$\frac{1}{2+\eta} + \frac{1}{q} = 1.$$

Also, the random variables $|\xi_i|^{2+\eta}$ are integrable from Lemma 2, Assumptions 2 (b) and (c). Finally, we can conclude by an application of Burkholder-Davis-Gundy inequalites with Assumptions 2 (b) and (c). □

Moreover, we show that Condition (81) (Condition (7.31) in Theorem IX.7.28 from Jacod and Shiryaev (2003)) holds in the following lemma.

Lemma 10. *We assume that Assumptions 1 and 2 hold. Then, we have for any time $t \in [0,1]$ and for any bounded stochastic process M' which is a martingale with respect to the filtration $\overline{\mathcal{F}}$ of dimension d when the final time $T \to \infty$ that*

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}[\xi_i^T \Delta_i M'] \xrightarrow{\mathbb{P}} 0.$$

Proof of Lemma 10. We first have that the rescaled error process X is a purely discontinuous martingale with respect to the filtration $\overline{\mathcal{F}}$ almost surely in the sense of Definition I.4.11 (b) (p. 40) from Jacod and Shiryaev (2003). This is obtained by Definition (73) and Definition 2. Thus, we can deduce that the product of the martingales XM' is a martingale with respect to the filtration $\overline{\mathcal{F}}$ almost surely for any bounded $\overline{\mathcal{F}}$-martingale M' of dimension d by Definition I.4.11 (a) (p. 40) from Jacod and Shiryaev (2003).

Thus, we can deduce for any time $t \in [0,1]$ and for any bounded $\overline{\mathcal{F}}$-martingale M' of dimension d almost surely that

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}[\Delta_i X^T \Delta_i M'] = 0.$$

Finally, this yields by Equation (75) from Lemma 6 for any time $t \in [0,1]$ and for any bounded $\overline{\mathcal{F}}$-martingale M' of dimension d when the final time $T \to \infty$ that

$$\sum_{i=1}^{\lfloor tT \rfloor} \mathbb{E}_{i-1}[\xi_i^T \Delta_i M'] \xrightarrow{\mathbb{P}} 0.$$

Thus, the lemma is shown. □

In what follows, we deliver the proof of Theorem 1, which is based on an application of Theorem 2, namely Theorem IX.7.28 (pp. 590-591) in Jacod and Shiryaev (2003).

Proof of Theorem 1. This is based on an application of Theorem 2, namely Theorem IX.7.28 (pp. 590-591) in Jacod and Shiryaev (2003). We now verify that Conditions (77) to (81) from Theorem 2 are satisfied. First, we set the reference martingale $Z_t = 0$ which is a square-integrable martingale with respect to the filtration $\overline{\mathcal{F}}$. Thus, Condition (79) is directly satisfied. In addition, we have that Condition (77) holds by Lemma 7. Moreover, we can deduce that Condition (78) is satisfied with the use of Lemma 8. We also get that Condition (80) holds by Lemma 9. Finally, we obtain that Condition (81) is satisfied by an application of Lemma 10. □