Functional limit theorems for nonstationary marked Hawkes processes in the high intensity regime

BO LI AND GUODONG PANG

Abstract. We study marked Hawkes processes with an intensity process which has a non-stationary baseline intensity, a general self-exciting function of event "ages" at each time and marks. The marks are assumed to be conditionally independent given the event times, while the distribution of each mark depends on the event time, that is, time-varying. We first observe an immigration-birth (branching) representation of such a non-stationary marked Hawkes process, and then derive an equivalent representation of the process using the associated conditional inhomogeneous Poisson processes with stochastic intensities. We consider such a Hawkes process in the high intensity regime, where the baseline intensity gets large, while the self-exciting function and distributions of the marks are unscaled, and there is no time-scaling in the scaled Hawkes process. We prove functional law of large numbers and functional central limit theorems (FCLTs) for the scaled Hawkes processes in this asymptotic regime. The limits in the FCLTs are characterized by continuous Gaussian processes with covariance structures expressed with convolution functionals resulting from the branching representation. We also consider the special cases of multiplicative self-exciting functions, and an indicator type of non-decomposable self-exciting functions (including the cases of "ceasing" and "delayed" reproductions as well as their extensions with varying reproduction rates), and study the properties of the limiting Gaussian processes in these special cases.

1. Introduction

Hawkes processes were introduced in [24, 25, 27]. They have a self-exciting intensity process that depends on its entire history, and can capture positive auto-correlation, clustering effects and over-dispersion in the counting/arrival processes. They have been widely used in various applications, for example, finance (see the recent reviews in [2, 26]), seismology [39], neuron science [36, 37, 9], internet traffic and queueing [32, 13, 21, 8]. It is usually assumed that the baseline intensity function is constant, and the self-exciting function is a function of the "age" of an event (elapsed time of each event). In some studies, the self-exciting function can depend on some exogenous randomness, that is, random "marks" associated with each event, see for example [7, 5, 31, 28]. This is natural in many applications, since the "marks" carried with each event can affect the intensity process, besides the "age" of each event. They are similar to the "noises" in shot noise processes (see, e.g., [44, 45, 46, 35] and references therein). The marks are all assumed to be i.i.d. in the existing literature. However, the marks may be dependent across events as well as dependent on the event times. In this paper, we focus on a particular dependent structure where the marks are conditionally independent given the event times while their distributions depend on the associated event times, that is, a time-varying distribution.

Such non-stationarity may appear in many practical systems. For one example, in [39], a marked Hawkes process is used to model earthquake occurrences and residual analysis, where the marks are magnitude of each occurrence. The distributions of the magnitudes may depend on the epochs of each occurrence. For another example, the effect of random forces on a damped harmonic oscillator can be modeled by a marked Hawkes process (or shot noise process [48]), where the forces can have a conditional Gaussian distribution with mean zero and covariance matrix that depends on the time

Key words and phrases. Marked Hawkes process, non-stationarity, high intensity asymptotic regime, immigration-birth representation, functional limit theorems, continuous Gaussian processes.

of random force occurrence; see for example [44, Section 3.1.3] where such non-stationary shot noise processes are used. As commented in [7], the marks can be used to model the signals propagating along nervous fibers, where the signal activities could be stimulated externally, and their forces might have different distributions at their occurrences. In [34], a marked Hawkes process is studied where the marks are stochastic processes themselves and the impact of the marks is modeled to depend on the event times. Such marked Hawkes process has potential applications in neural activities [37]. In individual-based stochastic epidemic models [17, 19, 42, 43, 18], each individual is associated with a random force of infectivity, and the total infectivity force may be modeled as marked Hawkes process, where the marks represent the infectivity forces, which very likely depend on the time of infection.

Specifically, we consider a non-stationary marked Hawkes process $N = \{N(t) : t \geq 0\}$ with an intensity process

$$\lambda(t) = \lambda_0(t) + \sum_{j=1}^{N(t)} H(t - \tau_j, Z_j(\tau_j)), \quad t \geq 0, \tag{1.1}$$

where $\lambda_0 : \mathbb{R}_+ \to \mathbb{R}_+$ is a deterministic positive function representing the baseline intensity, $\{\tau_j : j \in \mathbb{N}\}$ are the event times of the process N, $H(t, z) : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+$, $d \geq 1$, is the exciting function, and $Z_j(\tau_j)$'s (or Z_j's for brevity) are the marks associated with the j^{th} event time τ_j. We assume that given the sequence of the event times $\{\tau_j : j \in \mathbb{N}\}$, the marks $\{Z_j\}$ are independent and the distribution of each mark Z_j depends on the associated event time τ_j (independent of the other event times), that is,

$$\mathbb{P}(Z_j(\tau_j) \leq z | \tau_j = u,\, \tau_{j'},\, \forall j' \leq j) = F_u(z), \quad z \in \mathbb{R}_+^d\,, \quad u \geq 0\,. \tag{1.2}$$

Note that the non-stationarity of the process N arises from three sources: the non-stationary initial intensity function $\lambda_0(\cdot)$, the non-stationary distribution of the marks $Z_j(\cdot)$, and the exciting function H. Such non-stationary distributions of variables $Z_j(\cdot)$ in (1.2) have been assumed for noises in shot noise processes [44, 45]. However, Hawkes processes with such marks are much more challenging to analyze.

The self-exciting function $H(t, z)$ can take any general form. We also discuss two types of special models in Section 3. The first type is a multiplicative function $H(t, z) = \tilde{H}(t)z$ for $z \in \mathbb{R}_+$ (or more generally, $H(t, z) = \tilde{H}(t)\tilde{G}(z)$ for $z \in \mathbb{R}^d$), which is the usual model studied in the literature, see Section 3.1. The special case where $\tilde{H}(t) \equiv 1$, self-exciting with marks only, is also discussed. The second type is an indicator-type non-decomposable self-exciting function: $H(t, z) = H_0\mathbf{1}(0 \leq t < z)$, or $H(t, z) = H_0\mathbf{1}(t \geq z)$ for some constant $H_0 > 0$ and $z \in \mathbb{R}_+$, which we refer to as the cases of "ceasing" or "delayed" reproductions (using the terminology in the migration-brith representation), respectively. These two cases have constant reproduction rate, so we also extend them to allow the rates to be time-varying, depending on the "ages" of the individuals. In particular, we have introduced non-decomposable self-exciting functions: $H(t, z) = \tilde{H}(t)\mathbf{1}(0 \leq t < z)$ and $H(t, z) = \tilde{H}(t)\mathbf{1}(t \geq z)$ for a measurable and locally integrable function $\tilde{H} : \mathbb{R}_+ \to \mathbb{R}_+$. In addition, a more general non-decomposable self-exciting function can take effect only over a random period of time, where the mark $z = (z_1, z_2) \in \mathbb{R}_+^2$ and $H(t, z) = H(t, (z_1, z_2)) = \tilde{H}(t)\mathbf{1}(z_1 \leq t < z_2)$, and where z_2 and z_1 represent the "ceasing" and the "delay" time in the reproduction, respectively (here $(z_2 - z_1)$ can be regarded as the "active" reproduction period). These models are studied in detail in Section 3.2. To the best of our knowledge, Hawkes processes with such self-exciting functions have not been studied in the literature.

In [27], Hawkes and Oakes gave an immigration-birth (branching) representation of linear Hawkes processes, which has become a very useful tool to study the properties of the Hawkes processes. Our first result is to observe that the immigration-birth representation also holds in our new model setup (Section 2.1). As a consequence, we provide an equivalent representation for the Hawkes process using the conditional (inhomogeneous) Poisson processes in the branching description. It gives a natural decomposition of the Hawkes process to facilitate the proofs of the functional limit theorems.

Since the exact analysis is prohibitive, we establish the functional law of large numbers (FLLN) and functional central limit theorems (FCLT) for the non-stationary marked Hawkes process N. We consider the high intensity asymptotic regime, in which the baseline intensity gets large ($\lambda_0^n(t)/n \to \bar{\lambda}_0(t)$ as the scaling parameter $n \to \infty$), while the self-exciting function H and the distributions of the "marks", $F_s(\cdot)$ are unscaled/fixed. The process N^n, indexed by the scaling parameter n, has the associated event time τ_j^n. We then consider the LLN and CLT scalings of the process N^n, with $\bar{N}^n = n^{-1}N^n$ and $\hat{N}^n = \sqrt{n}(\bar{N}^n - \mathbb{E}[\bar{N}^n])$. The asymptotic regime is clearly different from the conventional scaling regime where both time t and space are scaled simultaneously, because of the dependence of the self-exciting function on the "age" of each event. Most of the scaling limits for Hawkes processes are in this conventional asymptotic regime. For example, Barcry et al. [1] studied the stationary Hawkes process (no marks), and obtained Brownian motion limits in the FCLT, and Horst and Xu [28] studied the Hawkes processes with time-varying baseline intensity and i.i.d. marks, and proved Gaussian white noise limits. See Remarks 2.5 and 2.8 for further discussions and comparisons of the scaling limits in these two regimes. The only works concerning the large intensity regime are done in [21] and [20]. In [21], a linear stationary Hawkes process (with no marks) is studied, and FCLT is proved with a Gaussian limit process with a particular covariance structure. That is used to model the arrival process of infinite-server queues, which may have many practical applications. (A potential application of the large intensity regime in biology is to model chemical kinetics where the synthesis of molecules occurs in large numbers, see, e.g., [14] with a Cox process as the arrivals for an infinite-server tandem queueing model while Hawkes process could be potentially used.) In [20], Gao and Zhu recently proved limit theorems and large deviations for nonlinear Hawkes processes (with no marks) in the asymptotic regime with a large intensity function and a small exciting function (which is equivalent to the large intensity regime in the linear case) and the limiting process is a semimartingale Gaussian process. However, the approaches of proving weak convergence in [21] and [20] rely critically on the stationarity property of the process, using the stationary version of the Hawkes process, and cannot be extended to the non-stationary setting.

We focus on the proofs of the FCLT. We first show that the limiting Gaussian processes are well defined and have continuous sample paths. The covariance functions are expressed using convolution functionals, which is also natural from the branching representation of the Hawkes process. We also discuss the particular covariance properties of the limiting Gaussian processes in the special models of multiplicative and indicator-type non-decomposable self-exciting functions in Section 3. For models with multiplicative self-exciting functions, in the special case with a constant baseline intensity and no marks, we show that $\hat{N}(t+h) - \hat{N}(h)$ as $h \to \infty$ converges in distribution to a stationary Gaussian process $\hat{N}^\circ$ which can be characterized via a stochastic integral with respect to a two-sided Brownian motion. We prove that the covariance function of this limit $\hat{N}^\circ$ is in fact equivalent to that established in [21], although the expressions appear to take very different forms (see Proposition 3.1). We also study the corresponding properties and scaling limits of Hawkes processes with indicator-type non-decomposable self-exciting functions.

Non–stationarity brings substantial challenges in proving the weak convergence and analyzing the limit processes. The existing work on Hawkes processes relies heavily upon the convenient representations of the processes, see, e.g., the simple integral representation of the auxiliary processes in the decomposition of the centered Hawkes processes in Remark 2.4 and the associated martingales that can be constructed. We therefore must start from the immigration-birth branching representation and derive the representations of the Hawkes processes and decompositions using the conditional Poisson processes with stochastic intensities in the branching description. In particular, the sub-processes are represented via stochastic Volterra integrals with respect to martingales constructed from those conditional Poisson processes, and have complicated dependence structures. We then prove their weak convergence by checking the convergence of finite dimensional distributions and verifying the tightness criterion with the modulus of continuity directly. The proof of tightness for the model appears very challenging, for which requires to establish some nontrivial maximal inequalities and moment bounds for the oscillation of the subprocesses within small time-interval.

There are other relevant papers on scaling limits of (marked) Hawkes processes. In [31], CLT and large deviations results are obtained for Hawkes processes with i.i.d. marks. Gao and Zhu [23, 22] consider a linear Markovian Hawkes process with an exponential exciting function and with a constant baseline intensity and no marks, and prove CLTs and large deviations and studied large time asymptotic in the regime where the initial intensity is large. Zhu [51] also proves FCLT for nonlinear Hawkes processes using Poisson embedding, and obtains a Brownian limit process. Jaisson and Rosenbaum [29, 30] established FCLTs when the exciting functions have light and heavy tails. There have also been recent works on mean-field limits for Hawkes processes [15, 10, 9, 11]. These are all in the usual asymptotic regime with both time and space scalings.

1.1. Organization of the paper.

We give some notation used in the paper in the next subsection. We first provide the immigration-birth representation, and the resulting equivalent representations in Sections 2.1 and 2.2. We give the assumptions on the model, describe the high intensity regime and summarize the functional limit theorems in Section 2.3. We discuss the special case of multiplicative self-exciting functions in Section 3.1, and then introduce the special case of indicator-type non-decomposable self-exciting functions in Section 3.2, including the "ceasing" and "delayed" reproduction cases with constant or varying rates, as well as the more general model with both "delayed" and "ceasing" scenarios, i.e., a random active reproduction duration. The well-definedness of the Hawkes process is studied in Section 4. The proofs of the functional limit theorems are given in Section 5. The proofs about the limiting Gaussian processes in the special cases are given in Section 6.

1.2. Notation.

All random variables and processes are defined in a common complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. Throughout the paper, $\mathbb{N}$ denotes the set of natural numbers. $\mathbb{R}(\mathbb{R}_+)$ denotes the space of real (nonnegative) number. Let $\mathbb{D} = \mathbb{D}(\mathbb{R}_+, \mathbb{R})$ denote $\mathbb{R}$-valued function space of all càdlàg functions on $\mathbb{R}_+$. For $z \in \mathbb{R}^d$, $|z|$ denotes the Euclidean norm. $(\mathbb{D}, J_1)$ denotes space $\mathbb{D}$ equipped with Skorohod J_1 topology, see [4], which is complete and separable. Let $\mathbb{C}$ be subset of $\mathbb{D}$ for continuous functions. $\mathbb{D} \times \mathbb{C}$ denotes the product space endowed with the weak Skohorod topology [50]. $L^2(\mathbb{P})\big(L^4(\mathbb{P})\big)$ denotes the space of random variables with finite 2^{nd} (4^{th}) moment. For integrable function $f : \mathbb{R} \to \mathbb{R}$, its L^1 norm is denoted by $||f||_1$. Notations $\to$ and $\Rightarrow$ ($\to_h$ and $\Rightarrow_h$) mean convergence of real numbers and convergence in distribution(with respect to parameter h), respectively. Additional notation is introduced in the paper whenever necessary.

2. The Model and Results

2.1. Immigration-birth representation.

To facilitate the analysis, we present an immigration-birth representation for the non-stationary Hawkes process N described in (1.1) with chronological levels, which is the key to our results and proofs. The representation is a direct generalization of that presented first by Hawkes and Oakes in [27]. The existence and uniqueness of the process are discussed after introducing Assumption 1. As is expected, the immigration and birth processes will be conditional inhomogeneous Poisson processes.

Immigrants: Let $N_1(t) = \sup\{j \geq 1 : \tau_{1j} \leq t\}$ be the counting process of the first generation, which represents "immigrants" without extant parents in the process, and follows an inhomogeneous Poisson process with baseline intensity $\lambda_0(\cdot)$. The mark associated with the j^{th} immigrant at time τ_{1j} is denoted by Z_{1j}.

Birth property: For an individual (point) immigrated/born into the system at time τ, a mark Z is associated and its distribution depends on the occurrence time of τ, given by

$$\mathbb{P}\big(Z \in dz\big|\tau = s\big) = F_s(dz) \quad \forall z \in \mathbb{R}^d. \tag{2.1}$$

The individual produces children independently, which follows an inhomogeneous Poisson process with an intensity function $H(t-\tau, Z)$. Here, $(t-s)$ is the "age" of the individual at the current time t given the birth time $\tau = s \le t$.

Descendants: Given the l^{th} generation, the collection of children (points) produced by this generation is called the $(l+1)^{\text{th}}$ generation and denoted by N_{l+1}. The birth time and mark associated to the j^{th} newborn in the l^{th} generation are denoted by τ_{lj} and Z_{lj}. Let

$$\mathscr{G}_l(t) := \sigma\big\{N_l(s), Z_{lj} : 0 \le s \le t,\, j \le N_l(t)\big\} = \sigma\big\{\tau_{lj}, Z_{lj} : j \le N_l(t)\big\}$$

be the natural filtration generated by (N_l, Z_l), and $\mathscr{F}_l(t) := \bigvee_{l \ge k \ge 1} \mathscr{G}_k(t)$ represents the information produced by the generations 1 to l up to time t. Let $\mathscr{G}_l := \mathscr{G}_l(\infty)$ represent the information produced by the generation l up to time $+\infty$, and $\mathscr{F}_l := \mathscr{F}_l(\infty)$ represent the information produced by the generations 1 to ℓ up to time $+\infty$. By the birth property, given the filtration $\{\mathscr{F}_l(t)\}_{t\ge 0}$, N_{l+1} is a simple and conditional (inhomogeneous) Poisson process with the intensity process

$$\lambda_l(t) := \sum_{j=1}^{N_l(t)} H(t - \tau_{lj}, Z_{lj}) \in \mathscr{G}_l(t). \tag{2.2}$$

Clearly, N_{l+1} is independent of $\mathscr{F}_{l-1}$ conditioning on $\mathscr{G}_l$. In other words, conditioning on $\mathscr{F}_l$, $\sum_{j\ge 1} \delta_{(\tau_{(l+1)j}, Z_{(l+1)j})}$ is a Poisson random measure with intensity $\lambda_l(t)F_t(dz)dt$, where δ_x stands for the Dirac point mass at x (see, e.g., [3, Chapter O.5]).

Hawkes process: Let $N(t) = \sum_{l\ge 1} N_l(t)$ for $t \ge 0$. Then N is a point process with intensity process:

$$\begin{aligned}\lambda(t) = \sum_{l\ge 0} \lambda_l(t) &= \lambda_0(t) + \sum_{\tau_{lj} \le t} H(t - \tau_{lj}, Z_{lj}) \\ &= \lambda_0(t) + \sum_{\tau_j \le t} H(t - \tau_j, Z_j) \in \mathscr{F}_\infty(t),\end{aligned}$$

where $\{\tau_j\}$ are the resorting of $\{\tau_{lj}, l, j \ge 1\}$ representing the occurrence times of N and $\{Z_j\}$ are the associated marks. The distribution of Z_j depends on τ_j and is given by (2.1). Thus, this expression of the intensity process coincides with that defined in (1.1). Here $\mathscr{F}_\infty(t)$ represents the information produced by all the generations up to time t using the notation above.

Let $H(t,z)$ be the self-exciting function in (1.1). For $t, s > 0$ and $k \ge 1$, define

$$\begin{aligned} G(t,s) &:= \int_{\mathbb{R}^d} H(t-s, z) F_s(dz) \quad \text{and} \quad G^{k+1}(t,s) := \int_0^t G(t,u) G^k(u,s) du, \\ \psi(t,s) &:= \sum_{k \ge 1} G^k(t,s) = G(t,s) + \int_0^t G(t,u)\psi(u,s)du, \\ \phi_t(s,z) &:= \mathbf{1}_t(s) + \int_0^t \Big(1 + \int_0^t \psi(u,v)du\Big) H(v-s,z)dv, \end{aligned} \tag{2.3}$$

where $\mathbf{1}_t(s) := \mathbf{1}(s \le t)$. Since $H(t,s) = 0$ for $t < s$, $G^k(t,s) = 0$ for all $k \ge 1$. For the nontrivial case $t \ge s > 0$, there is an intuitive interpretation of the quantities above. For an individual (point) born to the system at time s, since $H(t-s,z)$ is the production rate at time t contributed by the individual (point) with mark z, $G(t,s)$ can be regarded as the expected rate of production contributed by the individual (point) to its 1^{st} latter generation, and $G^k(t,s)$ can be regarded as the expected contribution to its k^{th} latter generation, which is produced by its $(k-1)^{\text{th}}$ latter generation. Thus, $\psi(t,s)$ is the expected reproduction rate at time t of all the descendent of the individual (point) born at s. The integral representation of $\psi(t,s)$ is known as the Volterra equation with kernel G, and can be simply derived by summing on both sides of equation for G^k over $k = 1, 2, \cdots$. It can also be derived by conditioning on the information of the last generation. Finally, $\phi_t(s,z)$ is the expected number of members in the family of the point originated from the point born at s and associated with a mark value z, counting itself as well. Therefore by conditioning on the immigrant process N_1, we have

$$\mathbb{E}\big[\lambda(t)\big] = \lambda_0(t) + \int_0^t \psi(t,u)\lambda_0(u)du.$$

Noticing that infinite sums are used for the auxiliary functions, we first give a sufficient condition for their well–definednesses,

Assumption 1. *Assume that $H(t,z) = 0, \forall t < 0, z \in \mathbb{R}^d$. For any $T > 0$, the following hold:*

(i) For some locally integrable and measurable function $\varphi_T \ge 0$,

$$\int_{\mathbb{R}^d} H^2(t-s,z)F_s(dz) \le \varphi_T^2(t-s) \quad \forall s,t \in [0,T].$$

(ii) The function φ_T above is locally square integrable, that is, $\int_0^T \varphi_T^2(t)dt < \infty$.

Remark 2.1. *Under Assumption 1(i), we will have for every $T > 0$*

$$G(t,s) \le \varphi_T(t-s) \quad \text{for all } t,s \in [0,T]$$

which is locally integrable in $t > 0$. One can find from (2.3) that $t \to \psi(t,s)$ is the well-defined and locally integrable solution to the Volterra equation with kernel G and $\psi(t,s) = 0$ if $t < s$ by definition. Moreover, the representation above shows that (N, λ) is well-defined and finite with probability one. Therefore, by the Lemma in [27], N is the unique orderly point process on $\mathbb{R}_+$ with the conditional intensity process (1.1). Assumption 1(i) will be assumed throughout the paper.

Remark 2.2. *Assumption 1 is not strong. For the case of $H(t,z)$ being a bounded function, both the assumption always holds, for example, when $H(t,z) = \mathbf{1}(0 < t \le z), \forall t,z \ge 0$ considered in Section 3.2 and referred to as "ceasing" reproduction. In Section 3.1, we also consider the case of multiplicative function $H(t,z) = \tilde{H}(t)z$, $\forall t,z \ge 0$. In this case,*

$$\int_{\mathbb{R}_+} H^2(t-s,z)F_s(dz) = \tilde{H}^2(t-s)\int_{\mathbb{R}_+} z^2 F_s(dz).$$

It is sufficient to assume that $\tilde{H}$ is locally square integrable on $\mathbb{R}_+$ and $\sup_{s\in[0,T]}\int_{\mathbb{R}_+} z^2F_s(dz) < \infty$, that is, the second moment of the marks is locally bounded. If we further assume $F_s = F$ for some c.d.f. F on $[0,\infty)$, that is, in the i.i.d. case of one dimensional marks, letting $\mathbf{m}_1 = E[Z] = \int_0^\infty zF(dz)$, then we have

$$G(t,s) = \mathbf{m}_1\tilde{H}(t-s), \quad G^k(t,s) = \mathbf{m}_1^k\tilde{H}^{*k}(t-s), \quad \psi(t,s) = \sum_{k\ge1}\mathbf{m}_1^k\tilde{H}^{*k}(t-s) = \psi(t-s),$$

$$\phi_t(s,z) = \mathbf{1}_t(s) + \frac{z}{\mathbf{m}_1}\int_0^t \psi(u-s)du = \mathbf{1}_t(s) + \frac{z}{\mathbf{m}_1}\cdot\psi * 1(t-s)$$

where $\tilde{H}^{*k}$ denotes the k^{th} self-convolution of $\tilde{H}$.

Example 2.1. *We give an example of the non-stationary c.d.f. $F_t(\cdot)$ that may be applicable to model 'environmental' or 'seasonal' effects. Suppose there exists a sequence of deterministic times $0 = T_0 < T_1 < \cdots < T_k = T < \infty$ and the corresponding c.d.f.'s $F_{(i)}(\cdot)$, $i = 1, \ldots, k$, such that for*

$t \in [T_{i-1}, T_i)$ and for each $i = 1, \ldots, k$, $F_t(\cdot) = F_{(i)}(\cdot)$. It is possible that some of the $F_{(i)}(\cdot)$'s are common (for instance, in periodic settings), but not all.

Given the immigrant-birth representation, we further define for $k \geq l \geq 1$ and $t \geq 0$,

$$\begin{aligned} Y_{kl}(t) &:= \mathbb{E}\big[N_{k+1}(t)\big|\mathscr{F}_l\big] - \mathbb{E}\big[N_{k+1}(t)\big|\mathscr{F}_{l-1}\big], \\ X_l(t) &:= N_l(t) - \int_0^t \lambda_{l-1}(s)ds, \quad Y_l(t) := \sum_{k\geq l} Y_{kl}(t), \\ X(t) &:= \sum_{l\geq 1} X_l(t) \quad \text{and} \quad Y(t) := \sum_{l\geq 1} Y_l(t), \end{aligned} \tag{2.4}$$

and $M_l(t) := X_l(t) + Y_l(t)$, where we understand that $\mathbb{P}(\cdot|\mathscr{F}_0(t)) = \mathbb{P}(\cdot|\mathscr{F}_0) = \mathbb{P}(\cdot)$. It is clear that $Y_l(t) \in \mathscr{F}_l(t)$ and X_l is a $\{\mathscr{F}_l(t)\}_{t\geq 0}$-martingale under measure $\mathbb{P}(\cdot|\mathscr{F}_{l-1})$. From the immigration-birth representation, for the $(l-1)^{\text{th}}$ generation, X_l can be taken as its impact to its son generation, that is, the l^{th} generation, and Y_{kl} can be taken as its impact to the $(k+1)^{\text{th}}$ future generation with $(k+2-l) \geq 2$, and then, Y_l can be taken as its cumulated impact to the system with generation cap larger than 2. Thus, the centered Hawkes process N can be rewritten as

$$N(t) - \mathbb{E}[N(t)] = \sum_{l\geq 1} \big(X_l(t) + Y_l(t)\big) = \sum_{l\geq 1} M_l(t) = X(t) + Y(t), \quad t \geq 0. \tag{2.5}$$

One can check in (2.5) directly that

$$M_l(t) = \mathbb{E}[N(t)|\mathscr{F}_l] - \mathbb{E}[N(t)|\mathscr{F}_{l-1}], \quad t \geq 0. \tag{2.6}$$

These processes play an important role in the proof of weak convergence below.

Proposition 2.1. *Under Assumption 1(i), for every $t \geq 0$ and $l \in \mathbb{N}$, $X_l(t)$, $Y_l(t)$, $M_l(t)$, $X(t)$ and $Y(t)$ are well-defined variables in $L^2(\mathbb{P})$.*

Proof. We refer the proof to Proposition 4.2 where the stochastic processes $X, X_l \in \mathbb{D}$ and $Y, Y_l \in \mathbb{C}$ for all $l \in \mathbb{N}$, and Y_l satisfies a simplified equation. □

Remark 2.3. *Under Assumption 1(i), since $G(t,s)$ and $\varphi_T(t-s)$ are locally integrable in $t \in [0,T]$, $\Phi_T(t) := \sum_{k\geq 1} \varphi_T^{*k}(t)$ is well-defined and locally integrable on $[0,T]$. If, in addition, Assumption 1(ii) holds, then $\Phi_T \in L^2[0,T]$. These are shown in Lemma 4.1. This implies the well-posedness and finiteness of $\mathcal{U}f(s)$ in (2.7) for every f on a compact set of $[0,\infty)$ (see Corollary 4.1). Thus, the equivalent representations of Y_l and Y below in (2.14) and (2.15) are well-defined. We refer to Section 4 for further discussions on the well-definedness of the Hawkes process.*

We shall see from the proofs of the FCLTs that Assumption 1(ii) is only a technical condition for the weak convergence results. In addition, despite of the dependency among the chronological level l, condition (ii) also ensures that the correlated terms will not be too singular.

2.2. **Equivalent representations.** Before proceeding, we provide simplified expressions for Y_l and Y defined in (2.4), and which are well-defined under Assumption 1(i). Note that we focus on the weak convergence on finite interval. Let H be the self-exciting function in (1.1), and $\mathcal{B}_{b,c}$ be the collection of bounded measurable functions vanishing outside a compact set of $[0,\infty)$. For every

$f \in \mathcal{B}_{b,c}$, we define

$$\begin{aligned}
\mathcal{H}f(s,z) &:= \int_0^\infty f(u)H(u-s,z)du = \int_0^\infty f(s+u)H(u,z)du, \\
\mathcal{G}f(s) &:= \int_{\mathbb{R}^d} \mathcal{H}f(s,z)F_s(dz) = \int_0^\infty f(u)G(u,s)du \\
\mathcal{G}^{k+1}f(s) &:= \mathcal{G}(\mathcal{G}^k f)(s) = \int_0^\infty f(u)G^{k+1}(u,s)du \quad \text{for} \quad k \geq 1, \\
\mathcal{U}f(s) &:= \sum_{k\geq 0} \mathcal{G}^k f(s) = f(s) + \sum_{k\geq 1} \mathcal{G}^k f(s) = f(s) + \int_0^\infty f(u)\psi(u,s)du.
\end{aligned} \tag{2.7}$$

Since $f \in \mathcal{B}_{b,c}$ above is assumed to be 0 outside a compact set, all the integrals above are actually integrating on finite intervals, and $\mathcal{U}\mathbf{1}_t(\cdot), \mathcal{H}\mathcal{U}\mathbf{1}_t(\cdot,z) \in \mathcal{B}_{b,c}$ for all $t>0, z\in\mathbb{R}^d$.

For a test function f, $\mathcal{H}f(s,z)$ can be taken as the cumulated effect of points generated directly by the individual (s,z) in the representation under f. Actually, given the individual (s,z), since it produces new points following a Poisson point process, say $\{\mathcal{E}(t), t\geq 0\}$, with intensity $H(t-s,z)dt$, we have from compensation formula that, c.f. [3, O.5],

$$\mathbb{E}\Big[\sum_{0\leq t<\infty} f(\mathcal{E}(t))\Big] = \int_0^\infty f(t)H(t-s,z)dt = \mathcal{H}f(s,z).$$

If $f = \mathbf{1}_t(\cdot)$ for some $t>0$, then $\mathcal{H}\mathbf{1}_t(s,z)$ is the expected number of points produced by (s,z) over $[0,t]$. And similar to the previous interpretations, $\mathcal{G}f(s), \mathcal{G}^k f(s)$ and $\mathcal{U}f(s)$ can be taken as the cumulated effect to the first generation, k^{th} later generation and all the family originated from the point born at s. With understanding that $\mathcal{G}^0 f(s) = f(s)$, we have

$$\mathcal{H}\mathcal{U}f(s,z) = \int_0^\infty \Big(f(v) + \int_0^\infty f(u)\psi(u,v)du\Big)H(v-s,z)dv,$$

and $\phi_t(s,z)$ defined in (2.3) can be rewritten as

$$\phi_t(s,z) = \mathbf{1}_t(s) + \mathcal{H}\mathcal{U}\mathbf{1}_t(s,z) \tag{2.8}$$

where $\mathbf{1}_t(s) = \mathbf{1}(s\leq t)$.

Let N_l and λ_l be the processes defined in (2.2). Define for every $f\in\mathcal{B}_{b,c}$,

$$\lambda_l f := \int_0^\infty f(t)\lambda_l(t)du = \sum_{j=1}^\infty \int_0^\infty f(t)H(t-\tau_{lj}, Z_{lj})dt = \sum_{j=1}^\infty \mathcal{H}f(\tau_{lj}, Z_{lj}) \in \mathscr{G}_l, \tag{2.9}$$

and $\lambda_0 f := \displaystyle\int_0^\infty \lambda_0(t)f(t)dt$. The second identity above in (2.9) follows from Fubini's theorem. The summation term can be regarded as the cumulated $\mathcal{H}f$-effect with respect to points generated by N_l. Noticing that the test function f vanishes outside a compact set, saying $[0,T]$, we have $\mathcal{H}f(\tau_{lj}, Z_{lj}) = 0$ for $j > N_l(T)$, which implies that the summation term is in fact a finite sum. We obtain the following results on $\lambda_l f$.

Lemma 2.1. *Under Assumption 1(i), for every $f,g\in\mathcal{B}_{b,c}$ and $l\geq k\geq 0$,*

$$\mathbb{E}\big[\lambda_l f \big| \mathscr{F}_k\big] = \lambda_k \mathcal{G}^{l-k} f, \tag{2.10}$$

and

$$\mathrm{Cov}(\lambda_l f, \lambda_l g) = \sum_{k=1}^{l} \int_0^\infty \mathcal{G}^{l-k}\Big(\int_{\mathbb{R}^d} \big(\mathcal{H}\mathcal{G}^{k-1}f(\cdot,z)\mathcal{H}\mathcal{G}^{k-1}g(\cdot,z)\big)F_\cdot(dz)\Big)(s)\lambda_0(s)ds, \tag{2.11}$$

where $\mathcal{G}^0 f(s) = f(s)$ and $\sum_{k=1}^0 = 0$ by convention.

Proof. We refer the proof to Proposition 4.1 and its discussion. The finiteness of identities under Assumption 1(i) are ensured from Lemma 4.1. □

Now, recall $\mathbf{1}_t(s)$ in (2.3), and observe that using (2.9), we can rewrite

$$\int_0^t \lambda_k(s)ds = \int_0^\infty \lambda_k(s)\mathbf{1}\big(s \in (0,t]\big)ds = \lambda_k \mathbf{1}_t.$$

Thus, by (2.10), we have for $k \geq l \geq 1$,

$$\mathbb{E}\big[N_{k+1}(t)\big|\mathscr{F}_l\big] = \mathbb{E}\big[\mathbb{E}\big[N_{k+1}(t)\big|\mathscr{F}_k\big]\big|\mathscr{F}_l\big] = \mathbb{E}\big[\lambda_k\mathbf{1}_t\big|\mathscr{F}_l\big] = \lambda_l\mathcal{G}^{k-l}\mathbf{1}_t.$$

Therefore, the expression of $Y_{kl}(t)$ in (2.4) can be rewritten as

$$Y_{kl}(t) = \lambda_l\mathcal{G}^{k-l}\mathbf{1}_t - \mathbb{E}\big[\lambda_l\mathcal{G}^{k-l}\mathbf{1}_t\big|\mathscr{G}_{l-1}\big], \tag{2.12}$$

and the process Y_l in (2.4) can be rewritten as

$$Y_l(t) = \sum_{k\geq l} Y_{kl}(t) = \lambda_l\mathcal{U}\mathbf{1}_t - \mathbb{E}\big[\lambda_l\mathcal{U}\mathbf{1}_t\big|\mathscr{G}_{l-1}\big]. \tag{2.13}$$

Moreover, they are clearly simpler and equivalent representations of our interested processes.

Since $\mathcal{U}\mathbf{1}_t(s)$ is the expected number over $[0,t]$ of points from the family originated from the point born at time s counting itself as well, the expected number of points in the system of the $(l+1)^{\text{th}}$ and the later generation will be

$$\mathbb{E}\left[\sum_{k\geq l} N_{k+1}(t)\Big|\mathscr{F}_l\right] = \mathbb{E}\left[\sum_{j=1}^{N_{l+1}(t)} \mathcal{U}\mathbf{1}_t(\tau_{(l+1)j})\Big|\mathscr{F}_l\right] = \int_0^\infty \mathcal{U}\mathbf{1}_t(s)\lambda_l(s)ds = \lambda_l\mathcal{U}\mathbf{1}_t.$$

Thus Y_l in (2.13) is the difference of the expectations under $\mathscr{F}_l$ and $\mathscr{F}_{l-1}$, and can be taken as the impact only contributed by $(l-1)^{\text{th}}$ generation as discussed in (2.4). By further looking at the information generated by λ_l, the last identity above can be written as the partial sum of $\mathcal{H}\mathcal{U}\mathbf{1}_t(\tau_{lj}, Z_{lj})$ as shown in (2.9), where $\mathcal{H}\mathcal{U}\mathbf{1}_t(s,z)$ is understood as the expected number of strictly later points originated from the point born at (s,z) over $[0,t]$. Thus,

$$Y_l(t) = \int_0^t\int_{\mathbb{R}^d} \mathcal{H}\mathcal{U}\mathbf{1}_t(s,z)\Big(N_l(ds,dz) - \lambda_{l-1}(s)F_s(dz)ds\Big), \tag{2.14}$$

which can also be checked from Fubini's theorem, where $N_l(ds,dz)$ is the conditional Poisson random measure with intensity $\lambda_{l-1}(s)F_s(dz)ds$ as introduced in the representation, and then

$$Y(t) = \int_0^t\int_{\mathbb{R}^d} \mathcal{H}\mathcal{U}\mathbf{1}_t(s,z)\Big(N(ds,dz) - \lambda(s)F_s(dz)ds\Big), \tag{2.15}$$

where $N(ds,dz) = \sum_{l\geq 1} N_l(ds,dz)$ is the marked Hawkes process.

Remark 2.4. *In the case of $H(t,z) = H(t)$, that is, a Hawkes process without marks, we have $G(t,s) = H(t-s)$ and $\psi(t,s) = \psi(t-s)$, abusing notation with the same ψ, where $\psi(t)$ satisfies*

$$\psi(t) = H(t) + \int_0^t H(t-u)\psi(u)du.$$

Thus, $\mathcal{H}\mathcal{U}\mathbf{1}_t(s,z) = \int_0^t \psi(u-s)du = \int_0^{t-s}\psi(u)du$ and by Fubini's theorem,

$$Y(t) = \int_0^t\int_0^t \psi(u-s)du\big(N(ds) - \lambda(s)ds\big) = \int_0^t \psi(t-s)X(s)ds$$

which appears in Lemma 4 in [1].

2.3. **Functional limit theorems.** We consider a sequence of the non-stationary marked Hawkes processes N^n indexed by n in the high intensity asymptotic regime, where the baseline intensity gets large, in the order $O(n)$, while the self-exciting function and the marks' distributions are fixed. The stochastic intensity for N^n in the n^{th} system is given by

$$\lambda^n(t) = \lambda_0^n(t) + \sum_{j=1}^{N^n(t)} H(t-\tau_j^n, Z_j^n(\tau_j^n)), \tag{2.16}$$

where the marks' distribution is given by

$$\mathbb{P}\big(Z_j^n(\tau_j^n) \in dz \big| \tau_j^n = u,\, \tau_{j'}^n,\, \forall j' \le j\big) = F_u(dz), \quad u > 0,\, z \in \mathbb{R}^d.$$

Using the representation (2.4) and (2.13), we define the following diffusion-scaled processes

$$\begin{aligned} \hat{X}_l^n(t) &:= \frac{1}{\sqrt{n}}\Big(N_l^n(t) - \mathbb{E}\big[N_l^n(t)\big|\mathscr{F}_{l-1}^n\big]\Big), \\ \hat{Y}_l^n(t) &:= \frac{1}{\sqrt{n}}\Big(\lambda_l^n \mathcal{U}\mathbf{1}_t - \mathbb{E}\big[\lambda_l^n \mathcal{U}\mathbf{1}_t\big|\mathscr{F}_{l-1}^n\big]\Big), \\ \hat{M}_l^n(t) := \hat{X}_l^n(t) + \hat{Y}_l^n(t) &= \frac{1}{\sqrt{n}}\left(\sum_{j=1}^{N_l^n(t)} \phi_t(\tau_{lj}^n, Z_{lj}^n) - \mathbb{E}\left[\sum_{j=1}^{N_l^n(t)} \phi_t(\tau_{lj}^n, Z_{lj}^n)\middle|\mathscr{F}_{l-1}^n\right]\right), \end{aligned} \tag{2.17}$$

for ϕ_t defined in (2.3) and $\lambda_l^n f$ defined in (2.9), where $\mathscr{F}_l^n = \sigma\big\{\tau_{lj}^n, Z_{lj}^n, j \ge 1\big\}$ and

$$\lambda_l^n(t) := \sum_{j=1}^{N^n(t)} H(t-\tau_{lj}^n, Z_{lj}^n(\tau_{lj}^n)).$$

Let

$$\hat{X}^n(t) := \sum_{l\ge 1} \hat{X}_l^n(t), \quad \hat{Y}^n(t) := \sum_{l \ge 1} \hat{Y}_l^n(t), \quad t \ge 0.$$

Thus, we have the diffusion-scaled process $\hat{N}^n$

$$\hat{N}^n(t) := \frac{1}{\sqrt{n}}\big(N^n(t) - \mathbb{E}[N^n(t)]\big) = \sum_{l\ge1} \hat{M}_l^n(t) = \sum_{l \ge 1}\big(\hat{X}_l^n(t) + \hat{Y}_l^n(t)\big) = \hat{X}^n(t) + \hat{Y}^n(t), \tag{2.18}$$

where applying Lemma 2.1,

$$\mathbb{E}\big[N^n(t)\big] = \sum_{l\ge1}\mathbb{E}\big[N_l^n(t)\big] = \lambda_0^n\mathbf{1}_t + \sum_{l\ge1}\mathbb{E}\big[\lambda_l^n\mathbf{1}_t\big] = \lambda_0^n\mathbf{1}_t + \sum_{l\ge1}\lambda_0^n\mathcal{G}^l\mathbf{1}_t = \lambda_0^n\mathcal{U}\mathbf{1}_t. \tag{2.19}$$

We make the following assumptions on the baseline intensity and the marks' distributions. Let $\bar{\lambda}_0^n(t) := n^{-1}\lambda_0^n(t)$.

Assumption 2. *Assume that for some locally integrable function $\bar{\lambda}_0$ on $[0,\infty)$,*

$$\sup_{t\in[0,T]}\big|\bar{\lambda}_0^n(t) - \bar{\lambda}_0(t)\big| \to 0 \quad as \quad n \to \infty.$$

Assumption 3. *For every $T > 0$,*

(i) the family $\big\{z \to \int_0^T H(t,z)dt\big\}$ is locally bounded on $\mathbb{R}^d$, that is, for all $K > 0$

$$\sup_{|z|\le K}\int_0^T H(t,z)dt < \infty;$$

(ii) $\Big(1+\int_0^T H(t-s,z)dt\Big)^2$ *is uniformly integrable with respect to* $\{F_s\}_{s\in[0,T]}$*, that is,*

$$\lim_{K\to\infty}\sup_{s\in[0,T]}\int_{|z|>K}\Big(1+\int_0^T H(t-s,z)dt\Big)^2 F_s(dz)=0.$$

Now, we are ready to present the FWLLN and FCLT for the non-stationary marked Hawkes process N^n with stochastic intensity λ^n defined in (2.16).

Theorem 2.1. *Under Assumptions 1 and 2, we have*

$$\bar N^n := \frac{1}{n}N^n \Rightarrow \bar N \quad in \quad (\mathbb{D},J_1) \quad as \quad n\to\infty,$$

where

$$\bar N(t) := \int_0^t \mathcal{U}\mathbf{1}_t(u)\bar\lambda_0(u)du = \int_0^t \Big(1+\int_0^t \psi(v,u)dv\Big)\bar\lambda_0(u)du, \quad t\ge 0,$$

with $\mathcal{U}\mathbf{1}_t(u)$ *given in* (2.7) *and* $\psi(u,v)$ *defined in* (2.3)*.*

Remark 2.5. *For the Hawkes process with i.i.d. marks, similar to Remark 2.2, we have reduced forms for our auxiliary functions, that is,*

$$G(t,s)=\int_{\mathbb{R}^d} H(t-s,z)F(dz)=G(t-s),$$

$$\psi(t,s)=\psi(t-s)=G(t-s)+\int_0^t G(t-u)\psi(u-s)du.$$

where ψ *is the renewal density with respect to* G*. (Here we abuse notation using the same* ψ*, see also Remark 2.4.) If* $\bar\lambda_0(\cdot)\equiv\bar\lambda_0$ *is a constant function, then*

$$\bar N(t)=\bar\lambda_0\int_0^t\Big(1+\int_0^t\psi(v-u)dv\Big)du=\bar\lambda_0\int_0^t\Big(1+1*\psi(u)\Big)du$$

by change of variable.

If the stability condition is satisfied, that is,

$$||G||_1=\int_0^\infty G(t)dt=\int_0^\infty\int_{\mathbb{R}^d}H(t,z)F(dz)dt<1, \tag{2.20}$$

we obtain the following result for the limit function $\bar N(t)$*:*

$$\frac{1}{n}\bar N(nt)\to\bar\lambda_0 t\big(1+||\psi||_1\big)=\frac{\bar\lambda_0 t}{1-||G||_1}, \tag{2.21}$$

uniformly in $t\in[0,T]$ $\mathbb{P}$*-a.s. as* $n\to\infty$*. Note that* $||G||_1$ *is the expected number of descendants a point can produce, and* $||G||_1<1$ *is also referred to as the subcritical condition from branching theory point of view, under which* $1+1*\psi(u)\to\frac{1}{1-||G||_1}$ *as* $u\to\infty$*.*

We recall the FLLN limit in [1], where Hawkes processes without marks are considered, that is, $H(t,z)=H(t)$*, and in addition,* $\lambda_0(\cdot)\equiv\bar\lambda_0$*. It is shown that under the stability condition* $||H||_1:=\int_0^\infty H(t)dt\in(0,1)$*,*

$$\sup_{t\in[0,T]}\left|\frac{1}{n}N(nt)-\frac{\bar\lambda_0\, t}{1-||H||_1}\right|\to 0 \quad as \quad n\to\infty, \tag{2.22}$$

almost-surely and in $L^2(\mathbb{P})$*. We observe that the limit coincides with that in* (2.21) *since without marks,* $||H||_1=||G||_1$*. It is also worth mentioning the LLN result in [31] that, for the Hawkes* *process with stationary marks of a common distribution $F(dz)$ and with $\lambda_0(\cdot) \equiv \bar{\lambda}_0$, under the stability condition in (2.20), almost-surely and in $L^2(\mathbb{P})$,*

$$\frac{1}{t}N(t) \to \frac{\bar{\lambda}_0}{1 - ||G||_1} \quad as \quad t \to \infty.$$

We first prove the following theorem for the convergence of $\{(\hat{X}^n_l, \hat{Y}^n_l), l \geq 1\}$ and $(\hat{X}^n, \hat{Y}^n)$ in the decomposition of $\hat{N}^n$.

Theorem 2.2. *Under Assumptions 1, 2 and 3, for each $l \geq 1$,*

$$\big(\hat{X}^n_l, \hat{Y}^n_l\big) \Rightarrow \big(\hat{X}_l, \hat{Y}_l\big) \quad in \quad \mathbb{D} \times \mathbb{C} \quad as \quad n \to \infty,$$

where the convergence is in the weak Skohorod topology on the product space $\mathbb{D} \times \mathbb{C}$, and $\big(\hat{X}_l, \hat{Y}_l\big)$ is a continuous Gaussian process of mean zero and covariance functions: for $t, s \geq 0$,

$$\begin{aligned}
\mathrm{Cov}(\hat{X}_l(t), \hat{X}_l(s)) &= \int_0^\infty \mathcal{G}^{l-1}\big(\mathbf{1}_t(\cdot)\mathbf{1}_s(\cdot)\big)(u)\bar{\lambda}_0(u)du \\
\mathrm{Cov}(\hat{X}_l(t), \hat{Y}_l(s)) &= \int_0^\infty \mathcal{G}^{l-1}\Big(\mathbf{1}_t(\cdot)\mathcal{G}\mathcal{U}\mathbf{1}_s(\cdot)\Big)(u)\bar{\lambda}_0(u)du \\
\mathrm{Cov}(\hat{Y}_l(t), \hat{Y}_l(s)) &= \int_0^\infty \mathcal{G}^{l-1}\Big(\int_{\mathbb{R}^d} \mathcal{H}\mathcal{U}\mathbf{1}_t(\cdot, z)\mathcal{H}\mathcal{U}\mathbf{1}_s(\cdot, z)F_\cdot(dz)\Big)(u)\bar{\lambda}_0(u)du.
\end{aligned}$$

In addition, the joint distribution of $(\hat{X}^n_l, \hat{Y}^n_l)_{l\geq 1}$ converges to that of $(\hat{X}_l, \hat{Y}_l)_{\geq 1}$, where $\{\big(\hat{X}_l, \hat{Y}_l\big), l \geq 1\}$ are independent over l. As a consequence, for each $l \geq 1$,

$$\hat{M}^n_l = \hat{X}^n_l + \hat{Y}^n_l \Rightarrow \hat{M}_l = \hat{X}_l + \hat{Y}_l \quad in \quad (\mathbb{D}, J_1) \quad as \quad n \to \infty,$$

and

$$\big(\hat{X}^n, \hat{Y}^n\big) \Rightarrow \big(\hat{X}, \hat{Y}\big) \quad in \quad \mathbb{D} \times \mathbb{C} \quad as \quad n \to \infty$$

where $\hat{X} = \sum_{l\geq 1} \hat{X}_l$ and $\hat{Y} = \sum_{l\geq 1} \hat{Y}_l$, and $(\hat{X}, \hat{Y})$ is a continuous Gaussian process of mean zero and covariance functions: for $t, s \geq 0$,

$$\begin{aligned}
\mathrm{Cov}(\hat{X}(t), \hat{X}(s)) &= \int_0^\infty \mathcal{U}\big(\mathbf{1}_t(\cdot)\mathbf{1}_s(\cdot)\big)(u)\bar{\lambda}_0(u)du, \\
\mathrm{Cov}(\hat{X}(t), \hat{Y}(s)) &= \int_0^\infty \mathcal{U}\Big(\mathbf{1}_t(\cdot)\mathcal{G}\mathcal{U}\mathbf{1}_s(\cdot)\Big)(u)\bar{\lambda}_0(u)du, \\
\mathrm{Cov}(\hat{Y}(t), \hat{Y}(s)) &= \int_0^\infty \mathcal{U}\Big(\int_{\mathbb{R}^d} \mathcal{H}\mathcal{U}\mathbf{1}_t(\cdot, z)\mathcal{H}\mathcal{U}\mathbf{1}_s(\cdot, z)F_\cdot(dz)\Big)(u)\bar{\lambda}_0(u)du.
\end{aligned} \tag{2.23}$$

For a test function $f \in \mathcal{B}_{b,c}$, we have by the definition of $\mathcal{U}f$ in (2.7) and ψ in (2.3) that

$$\begin{aligned}
\int_0^\infty \mathcal{U}f(u)\bar{\lambda}_0(u)du &= \int_0^\infty \Big(f(u) + \int_0^\infty f(v)\psi(v, u)dv\Big)\bar{\lambda}_0(u)du \\
&= \int_0^\infty f(u)\Big(\bar{\lambda}_0(u) + \int_0^u \psi(u, v)\bar{\lambda}_0(v)dv\Big)du
\end{aligned} \tag{2.24}$$

where Fubini's theorem and the fact $\psi(u, v) = 0$ for $v > u$ is applied. Therefore we obtain the following expressions for the covariance functions above.

Remark 2.6. *The integrals in Theorem 2.2 above are actually integrating on $u \in [0, t\wedge s]$. Applying (2.24), the covariance functions in (2.23) can be rewritten as*

$$\begin{aligned} \mathrm{Cov}(\hat{X}(t),\hat{X}(s)) &= \int_0^{t\wedge s}\Big(\bar{\lambda}_0(u)+\int_0^u \psi(u,v)\bar{\lambda}_0(v)dv\Big)du,\\ \mathrm{Cov}(\hat{X}(t),\hat{Y}(s)) &= \int_0^t \Big(\int_0^s \psi(w,u)dw\Big)\Big(\bar{\lambda}_0(u)+\int_0^u \psi(u,v)\bar{\lambda}_0(v)dv\Big)du,\\ \mathrm{Cov}(\hat{Y}(t),\hat{Y}(s)) &= \int_0^{t\wedge s}\Big(\int_{\mathbb{R}^d}\mathcal{HU}\mathbf{1}_t(u,z)\mathcal{HU}\mathbf{1}_s(u,z)F_u(dz)\Big)\Big(\bar{\lambda}_0(u)+\int_0^u \psi(u,v)\bar{\lambda}_0(v)dv\Big)du, \end{aligned}$$

where by definition

$$\mathcal{HU}\mathbf{1}_t(u,z) = \int_0^t \Big(1+\int_0^t \psi(w,v)dw\Big)H(v-u,z)dv.$$

Theorem 2.3. *Under Assumptions 1, 2 and 3,*

$$\hat{N}^n \Rightarrow \hat{N} \qquad in \quad (\mathbb{D},J_1) \quad as \quad n\to\infty, \tag{2.25}$$

where $\hat{N} := \{\hat{N}(t), t\geq 0\}$ is a centered continuous Gaussian process with covariance function $\hat{R}$:

$$\hat{R}(t,s) = \mathrm{Cov}\big(\hat{N}(t),\hat{N}(s)\big) = \int_0^\infty \mathcal{U}\bigg(\int_{\mathbb{R}^d}\phi_t(\cdot,z)\phi_s(\cdot,z)F_\cdot(dz)\bigg)(u)\bar{\lambda}_0(u)du, \tag{2.26}$$

for $t, s, \geq 0$, where $\phi_t(\cdot,\cdot)$ is defined in (2.3), see also (2.8). The limit $\hat{N}$ can be written as a sum of mutually independent continuous Gaussian processes $\hat{M}_l := \{\hat{M}_l(t), t\geq 0\}_{l\geq 1}$, that is, $\hat{N} = \sum_{l\geq 1}\hat{M}_l$, where $\hat{M}_l$ has mean zero and covariance function, for $t, s\geq 0$

$$\hat{R}_l(t,s) = \mathrm{Cov}\big(\hat{M}_l(t),\hat{M}_l(s)\big) = \int_0^\infty \mathcal{G}^{l-1}\bigg(\int_{\mathbb{R}^d}\phi_t(\cdot,z)\phi_s(\cdot,z)F_\cdot(dz)\bigg)(u)\bar{\lambda}_0(u)du. \tag{2.27}$$

From the immigration-birth representation presented in Section 2.1, the covariance function $\hat{R}$ in (2.26) can be understood as follows. The total Gaussian noise, the difference between N and its expectation, is the superposition of independent and small Gaussian noises caused directly by the point/individual of every generation originated from the immigration process, where each Gaussian noise is proportional to the size of the family that the point generates, and depends on the mark associated with the point, despite of their dependencies over generations. $\hat{M}_l$ is exactly the small Gaussian noises contributed by l^{th} generation as expected from the construction and discussion.

Remark 2.7. *Applying (2.24) the covariance functions above can also be written as*

$$\begin{aligned} \hat{R}_l(t,s) &= \int_0^{t\wedge s}\Big(\int_{\mathbb{R}^d}\phi_t(u,z)\phi_s(u,z)F_u(dz)\Big)\Big(\int_0^u G^{*(l-1)}(u,v)\bar{\lambda}_0(v)dv\Big)du\\ \hat{R}(t,s) &= \int_0^{t\wedge s}\Big(\int_{\mathbb{R}^d}\phi_t(u,z)\phi_s(u,z)F_u(dz)\Big)\Big(\bar{\lambda}_0(u)+\int_0^u \psi(u,v)\bar{\lambda}_0(v)dv\Big)du \end{aligned} \tag{2.28}$$

which are integrals on $[0, t\wedge s]$, and we understand $\int_0^u G^{*0}(u,v)\bar{\lambda}_0(v)dv = \bar{\lambda}_0(u)$.

Remark 2.8. *Following the notations in Remark 2.5, Theorem 1 in [1] shows that for the Hawkes process without marks, under the stability condition (similar to (2.20) without marks),*

$$\frac{1}{\sqrt{n}}\Big(N(nt)-\mathbb{E}\big[N(nt)\big]\Big) \Rightarrow \bigg(\frac{\bar{\lambda}_0}{(1-||H||_1)^3}\bigg)^{1/2} W(t) \tag{2.29}$$

in $\mathbb{D}$ with the Skorokhod J_1 topology as $n \to \infty$, where W is a standard Brownian motion. Theorem 2 in [31] shows that for the Hawkes process with stationary marks,

$$\frac{1}{\sqrt{t}}\Big(N(t) - \frac{\bar{\lambda}_0 t}{1-\|G\|_1}\Big) \Rightarrow N\Big(0, \frac{\bar{\lambda}_0\big(1+\sigma^2_{H,Z}\big)}{(1-\|G\|_1)^3}\Big), \tag{2.30}$$

where $\sigma^2_{H,Z} = \mathrm{Var}(\int_0^\infty H(t,Z)dt)$ and Z is a variable with c.d.f. F. It is evident that if there are no marks, then $\sigma^2_{H,Z} = 0$ and the variance formula $\frac{\bar{\lambda}_0(1+\sigma^2_{H,Z})}{(1-\|G\|_1)^3}$ coincides with $\frac{\bar{\lambda}_0}{(1-||H||_1)^3}$.

We note the main difference in the two scaling regimes lie in the role of the function H. In the conventional regime the scaling of the intensity process $\lambda(t)$ in (1.1) involves $\int_0^{nt} H(nt-s,z)ds$ and $\lambda_0(\cdot) \equiv \bar{\lambda}_0$, while in the large intensity regime, it scales $\lambda_0(t)$ by $\lambda_0^n(t)$ such that $\lambda_0^n(t)/n \to \bar{\lambda}_0(t)$, without any scaling on the integral $\int_0^t H(t-s,z)ds$. (This is similar to the scaling limits of shot noise processes, see [45].) The former appears to concern the stationarity behavior. This is confirmed by the following property: the variance function of the limit $\hat{N}(t)$ in (2.26) satisfies

$$\frac{1}{t}\hat{R}(t,t) \to \frac{\bar{\lambda}_0(1+\sigma^2_{H,Z})}{(1-||G||_1)^3}, \quad as \quad t \to \infty, \tag{2.31}$$

which coincides with the variance of the normal limit in (2.30). It should be an easy extension of the result in (2.29) in [1] that for the Hawkes process with stationary marks, under the stability condition in (2.20),

$$\frac{1}{\sqrt{n}}\Big(N(nt) - \mathbb{E}\big[N(nt)\big]\Big) \Rightarrow \Big(\frac{\bar{\lambda}_0(1+\sigma^2_{H,Z})}{(1-||H||_1)^3}\Big)^{1/2} W(t) \tag{2.32}$$

in $\mathbb{D}$ with the Skorokhod J_1 topology as $n \to \infty$. It remains open to show the FCLT for the non-stationary Hawkes process with time-varying marks in the conventional regime.

We also remark that in the limit theorems under the conventional scaling regime, as we see above, the stability condition in (2.20) plays a critical role, with or without marks. Without this condition, no FLLN or FCLT as in (2.22) and (2.29) can be proved. However, for the FLLN and FCLT in the high intensity regime, no such a stability condition is required.

We next provide a brief proof of the claim in (2.31). For the diffusion-scaled limit $\hat{N}$ after large intensity limit in Theorem 2.3, if $F_s = F$ and $||G||_1 < 1$, that is, Hawkes process with stationary marks and stability condition in (2.20), we already have in Remark 2.5 that

$$G(t,s) = G(t-s) \quad and \quad \psi(t,s) = \psi(t-s).$$

Thus, $\phi_t(s,z)$ defined in (2.3) is also a function of $(t-s)$, and abusing notation, we denote it as $\phi(t-s,z)$. Then, for $t \geq s > 0$,

$$\begin{aligned}
\phi_t(s,z) = \phi(t-s,z) &= 1 + \int_0^t H(u-s,z)du + \int_0^t du \int_0^u \psi(u-v)H(v-s,z)dv \\
&= 1 + \int_0^{t-s} H(u,z)du + \int_0^{t-s} du \int_0^u \psi(u-v)H(v,z)dv \\
&\xrightarrow{t\to\infty} 1 + \int_0^\infty H(u,z)du + \int_0^\infty du \int_0^u \psi(u-v)H(v,z)dv \\
&= 1 + \big(1+||\psi||_1\big)\int_0^\infty H(v,z)dv = 1 + \frac{\int_0^\infty H(u,z)du}{1-||G||_1}
\end{aligned} \tag{2.33}$$

where Fubini's theorem is applied and $1 + ||\psi||_1 = \frac{1}{1-||G||_1}$ *under the stability condition* (2.20)*. Plugging into* (2.28)*, the covariance function* $\hat{R}$ *in* (2.26) *reads*

$$\begin{aligned}\hat{R}(t,t) &= \int_0^t \Big(\int_{\mathbb{R}^d} \big(\phi_t(v,z)\big)^2 F_v(dz)\Big)\Big(\bar{\lambda}_0(v) + \int_0^v \psi(v,u)\bar{\lambda}_0(u)du\Big)dv \\ &= \bar{\lambda}_0 \int_0^t \mathbb{E}\big[\phi^2(t-v,Z)\big]\big(1 + 1 * \psi(v)\big)dv \\ &= t\cdot\bar{\lambda}_0 \int_0^1 \mathbb{E}\big[\phi^2(tv,Z)\big]\big(1 + 1 * \psi(t(1-v))\big)dv\end{aligned}$$

where the identities $F_v(dz) = F(dz), \phi_t(s,z) = \phi(t-s,z)$ *and* $\psi(u,v) = \psi(u-v)$ *are applied in the second line, and change of variable is applied in the last line. Letting* $t \to \infty$*, we have from* (2.33) *and the fact* $1 + ||\psi||_1 = \frac{1}{1-||G||_1}$ *that* (2.31) *holds.*

3. Special Models

In this section, we illustrate with various examples the covariance structures of the Gaussian limit process. We consider the special cases of the multiplicative self-exciting function and the indicator type of non-decomposable self-exciting function. We will discuss the corresponding sufficient conditions for Assumptions 1 and 3 in these cases. We also study some asymptotic properties of the limiting Gaussian processes.

3.1. **Multiplicative self-exciting function.** We consider $H(t,z) = \tilde{H}(t)\tilde{G}(z), \forall t > 0, z \in \mathbb{R}^d$ being a multiplicative function, where $\tilde{H} : \mathbb{R}_+ \to \mathbb{R}_+$ and $\tilde{G} : \mathbb{R}^d \to \mathbb{R}_+$ are measurable functions. By change of variable, the problem with $\mathbb{R}^d$-valued marks in this setting is equivalent to that with $\mathbb{R}_+$-valued marks. Therefore, w.l.o.g., we assume in this subsection that $H(t,z) = \tilde{H}(t)z$ for $z \in \mathbb{R}_+$, $\tilde{H} : \mathbb{R}_+ \to \mathbb{R}_+$ and $\{F_s, s \geq 0\}$ is a sequence of distributions on $\mathbb{R}_+$. We assume that $\forall T > 0$,

$$\int_0^T \tilde{H}^2(u)du < \infty \quad \text{and} \quad \lim_{K\to\infty} \sup_{s\in[0,T]} \int_{z>K} z^2 F_s(dz) = 0.$$

Denote by $\mathbf{m}_{1,s} = \int_0^\infty zF_s(dz)$ and $\mathbf{m}_{2,s} = \int_0^\infty z^2 F_s(dz)$. Assumptions 1 and 3 hold by taking $\varphi_T(t-s) = \tilde{H}(t-s)\Big(\sup_{u\in[0,T]} \mathbf{m}_{2,u}\Big)^{1/2}$, where the second locally uniformly integrable condition above implies $\sup_{u\in[0,T]} \mathbf{m}_{2,u} < \infty$. Recalling G^k and ψ defined in (2.3), and noticing that $\tilde{H}(u-v) = 0 = G^k(u,v) = \psi(u,v)$ for $u < v$, we have $G(t,s) = \mathbf{m}_{1,s}\tilde{H}(t-s)$,

$$\begin{aligned}G^{k+1}(t,s) &= \mathbf{m}_{1,s}\int_s^t G^k(t,u)\tilde{H}(u-s)du, \\ \psi(t,s) &= \mathbf{m}_{1,s}\Big(\tilde{H}(t-s) + \int_s^t \psi(t,u)\tilde{H}(u-s)du\Big),\end{aligned} \tag{3.1}$$

which means given a new point born at s, the reproduction rate of the point depends on the "age" of the point $(t-s)$ and is proportional to the value of the mark.

Corollary 3.1. $(\hat{X},\hat{Y})$ *in Theorem 2.2 has the following representations*

$$\hat{X}(t) = W_1(t) \quad \text{and} \quad \hat{Y}(t) = \int_0^t \Big(\int_0^v \frac{\psi(v,u)}{\mathbf{m}_{1,u}} W_2(du)\Big)dv, \tag{3.2}$$

and the limit $\hat{N}$ *is given by*

$$\hat{N}(t) = W_1(t) + \int_0^t \Big(\int_0^v \frac{\psi(v,u)}{\mathbf{m}_{1,u}} W_2(du)\Big)dv.$$

Here (W_1, W_2) is a two-dimensional Gaussian process such that for all $a, b \in \mathbb{R}$ and $t, s \geq 0$

$$\begin{aligned}&\mathbb{E}\big[(aW_1(t)+bW_2(t))(aW_1(s)+bW_2(s)\big] \\ &= \int_0^{t\wedge s} (a^2+2ab\mathbf{m}_{1,u}+b^2\mathbf{m}_{2,u})\Big(\bar\lambda_0(u)+\int_0^u \psi(u,v)\bar\lambda_0(v)dv\Big)du.\end{aligned} \tag{3.3}$$

Proof. We have from Remark 2.6 and (3.1) that

$$\mathcal{H}\mathcal{U}\mathbf{1}_t(s,z) = z\int_0^t \tilde H(u-s)du + z\int_0^t \psi(u,v)du\int_0^u \tilde H(v-s)dv = \frac{z}{\mathbf{m}_{1,s}}\int_0^t \psi(u,s)du.$$

Thus, the covariance functions for $\big(\hat X, \hat Y\big)$ in Theorem 2.2 are given by, for $t, s > 0$

$$\begin{aligned}\mathrm{Cov}(\hat X(t),\hat X(s)) &= \int_0^\infty \mathbf{1}_t(u)\mathbf{1}_s(u)\Big(\bar\lambda_0(u)+\int_0^u \psi(u,v)\bar\lambda_0(v)dv\Big)du,\\ \mathrm{Cov}(\hat X(t),\hat Y(s)) &= \int_0^\infty \mathbf{1}_t(u)\Big(\frac{\int_0^s \psi(v,u)dv}{\mathbf{m}_{1,u}}\Big)\mathbf{m}_{1,u}\Big(\bar\lambda_0(u)+\int_0^u \psi(u,v)\bar\lambda_0(v)dv\Big)du,\\ \mathrm{Cov}(\hat Y(t),\hat Y(s)) &= \int_0^\infty \Big(\frac{\int_0^t \psi(v,u)dv}{\mathbf{m}_{1,u}}\Big)\Big(\frac{\int_0^s \psi(v,u)dv}{\mathbf{m}_{1,u}}\Big)\mathbf{m}_{2,u}\Big(\bar\lambda_0(u)+\int_0^u \psi(u,v)\bar\lambda_0(v)dv\Big)du,\end{aligned}$$

applying Remark 2.6. The representation follows from the integrals

$$\hat X(t) = \int_0^\infty \mathbf{1}_t(u)W_1(du) \quad \text{and} \quad \hat Y(t) = \int_0^\infty \Big(\frac{1}{\mathbf{m}_{1,u}}\int_0^t \psi(v,u)dv\Big)W_2(du), \tag{3.4}$$

and the fact that $\psi(u,v) = 0$ for $u < v$. To see why the Itô integrals in the representations are meaningful, by definition, we have for all $a, b \in \mathbb{R}$ and $u \geq 0$,

$$(a^2+2ab\mathbf{m}_{1,u}+b^2\mathbf{m}_{2,u}) = \int_0^\infty (a+bz)^2 F_u(dz) \geq 0,$$

and thus, (W_1, W_2) in (3.3) is a well-defined Gaussian process and has independent increments. □

If the marks has an identical distribution, that is, $F_s = F$ as in Remarks 2.2 and 2.5, the covariance functions for $(\hat X, \hat Y)$ in Theorem 2.2 can be simplified to

$$\begin{aligned}\mathrm{Cov}\big(\hat X(t),\hat X(s)\big) &= \bar\lambda_0 * 1(t\wedge s) + \bar\lambda_0*\psi*1(t\wedge s),\\ \mathrm{Cov}\big(\hat X(t),\hat Y(s)\big) &= \int_0^t \psi*1(s-u)\big(\bar\lambda_0(u)+\psi*\bar\lambda_0(u)\big)du,\\ \mathrm{Cov}\big(\hat Y(t),\hat Y(s)\big) &= \frac{\mathbf{m}_2}{\mathbf{m}_1^2}\int_0^{t\wedge s} \psi*1(t-u)\psi*1(s-u)\big(\bar\lambda_0(u)+\psi*\bar\lambda_0(u)\big)du,\end{aligned} \tag{3.5}$$

where $\psi(u) = \sum_{k\geq 1}\mathbf{m}_1^k \tilde H^{*k}(u)$.

Remark 3.1. *In general, Hawkes process N does not have the Markov property, and neither has its Gaussian limit $\hat N$. Even in the exponential case, the Hawkes process $N(t)$ itself is not Markovian. But intensity process $\lambda(t)$ is Markovian, so is the pair $(N(t), \lambda(t))$. If $\tilde H(t) = e^{\beta t}$ for some $\beta \in \mathbb{R}$, it can be checked that*

$$G^k(t,s) = e^{\beta(t-s)}\frac{\mathbf{m}_{1,s}}{(k-1)!}\Big(\int_s^t \mathbf{m}_{1,u}du\Big)^{k-1} \quad \text{and} \quad \psi(t,s) = \mathbf{m}_{1,s}\exp\Big(\int_s^t (\beta+\mathbf{m}_{1,u})du\Big).$$

The desired limit process can be written as

$$\hat N(t) = W_1(t) + \int_0^t \Big(e^{\int_0^u(\beta+\mathbf{m}_{1,r})dr}\int_0^u e^{-\int_0^v(\beta+\mathbf{m}_{1,r})dr}W_2(dv)\Big)du,$$

where (W_1, W_2) *is defined in Corollary 3.1. By the independent increments of* (W_1, W_2)*, the derivative of* $\hat{Y}$ *in* (3.2) *satisfies*

$$d\dot{\hat{Y}}(t) = (\beta + \mathbf{m}_{1,t})\dot{\hat{Y}}(t)dt + W_2(dt)$$

which is an Ornstein–Uhlenbeck (OU) process driven by the continuous Gaussian noise W_2*. Therefore, it is necessarily that* $\dot{\hat{Y}}$ *is a Markov process but not time homogenous, and so will be* $(\hat{N}, \dot{\hat{Y}})$*.*

If we further assume $\bar{\lambda}_0(t) = e^{\beta t}$ *for the same* β *and i.i.d. marks, that is* $F_s = F$*, then*

$$\bar{\lambda}_0(t) + \int_0^t \psi(t,s)\bar{\lambda}_0(s)ds = e^{(\beta+\mathbf{m}_1)t}.$$

Let (B_1, B_2) *be two–dimensional Brownian motion with*

$$\mathbb{E}\big[(aB_1 + bB_2(t))(aB_1(s) + bB_2(s))\big] = (a^2 + 2ab\mathbf{m}_1 + b^2\mathbf{m}_2)\big(t \wedge s\big).$$

We have from the covariance function (3.3)*,*

$$W_1(t) = \int_0^t e^{\alpha u} B_1(du) \quad \text{and} \quad W_2(t) = \int_0^t e^{\alpha u} B_2(du),$$

for (W_1, W_2) *defined in Corollary 3.1 where we write* $\alpha := \dfrac{\beta + \mathbf{m}_1}{2}$ *for simplicity. Therefore,*

$$\hat{N}(t) = \int_0^t e^{\alpha u} B_1(du) + \int_0^t e^{2\alpha u}\Big(\int_0^u e^{-\alpha v} B_2(dv)\Big)du.$$

One can check directly that $e^{-\alpha t}\big(\hat{N}(t), \dot{\hat{Y}}(t)\big)$ *satisfies*

$$d\begin{pmatrix} e^{-\alpha t}\hat{N}(t) \\ e^{-\alpha t}\dot{\hat{Y}}(t)\end{pmatrix} = \begin{pmatrix} -\alpha & 1 \\ 0 & \alpha \end{pmatrix}\begin{pmatrix} e^{-\alpha t}\hat{N}(t) \\ e^{-\alpha t}\dot{\hat{Y}}(t)\end{pmatrix} dt + \begin{pmatrix} B_1(dt) \\ B_2(dt)\end{pmatrix},$$

which is a time-homogenous Markov process of OU type. We also refer the reader to the linear Markovian Hawkes process studied in the literature (see, e.g., [38, 22, 23] *and* [12, Exercise 7.2.5]*).*

3.1.1. *Stationary limit associated with the limiting Gaussian process.* Recall the limiting Gaussian process $\hat{N}$ in Theorem 2.3. In this subsection, we are interested in the stationary limit

$$\hat{N}^\circ(t) := \lim_{h\to\infty}\big(\hat{N}(t+h) - \hat{N}(h)\big), \quad t \ge 0, \tag{3.6}$$

as well as the other limit processes $\hat{X}$ and $\hat{Y}$. Here we assume that

(i) $F_s(dz) = \delta_1(dz)$ is a degenerate distribution at 1;
(ii) $\bar{\lambda}_0(s) \equiv \bar{\lambda}_0$ is a constant function on $[0,\infty)$ for some constant $\bar{\lambda}_0 > 0$; and
(iii) $||\tilde{H}||_1 = \displaystyle\int_0^\infty \tilde{H}(u)du < 1$.

The last condition (iii) is referred to as the stability condition in the literature, see for example, [1, 2, 6]. Under condition (iii), there is a unique stationary version of the Hawkes process, whose realizations of point sets are invariant under simultaneous shifts of their time arguments. Recalling the intuitive interpretation of $\mathcal{U}\mathbf{1}_t$ in (2.3), we have in this special case,

$$\lim_{t\to\infty}\mathcal{U}\mathbf{1}_t(s) = \frac{1}{1 - ||\tilde{H}||_1}$$

for every $s > 0$, that is, every family is finitely numbered.

Since $\hat{N}$ is a Gaussian process, the limit $\hat{N}^\circ$, if it exists, must also be a Gaussian process, and hence, we only need to check the limits of the covariance functions in Theorem 2.2. We show the following convergence result and characterize the limit processes as an Itô integral with respect to a two-sided Brownian motion. It is worth noting that the covariance function of $\hat{N}^\circ(t)$ in

(3.9) coincides with the result in Theorem 2 of [21] (in particular, our expression of the covariance function of $\hat{N}^\circ$ (3.9) is equivalent to the second expression in (3.10) given in [21]). It is no surprising since the Hawkes process tends to stationarity as the reference point goes to infinity. The model setup in [21] starts from stationarity at time zero and the proof technique relies heavily on the process in stationarity. The proof of the following proposition is given in Section 6.1. In the proof of the equivalence of the covariance function of the limit process between our expression (3.9) and the expression (3.10) given in [21], an expression for $\hat{\phi}$ in (3.11) in terms of the renewal density ψ in (2.3) is found, and where ψ in this case (with $\mathbf{m}_1 = 1$ under the above condition (i)) is given by

$$\psi(s+u,s) = \psi(u) = \tilde{H}(u) + \psi * \tilde{H}(u), \quad \text{for } s,u>0. \tag{3.7}$$

Proposition 3.1. *Let $\hat{N}$ be the scaled limit process in Theorem 2.3. We have for all $t \geq 0$,*

$$\big(\hat{N}(t+h) - \hat{N}(h)\big) \Rightarrow_h \hat{N}^\circ(t) \stackrel{d}{=} W(t) + \int_{-\infty}^{\infty} \Big(\int_0^t \psi(u-v)du\Big) W(dv), \tag{3.8}$$

as $h \to \infty$, where $W = \{W(t), t \in \mathbb{R}\}$ is a two-sided Brownian motion with $W(0) = 0$ and variance $\frac{\bar{\lambda}_0}{1-||\tilde{H}||_1}$. The covariance function of the process $\hat{N}^\circ$ is given by

$$\mathrm{Cov}\big(\hat{N}^\circ(t), \hat{N}^\circ(s)\big) = \frac{\bar{\lambda}_0}{1-||\tilde{H}||_1} \int_{\mathbb{R}} \Big(\mathbf{1}_t(v) + \int_0^t \psi(u-v)du\Big)\Big(\mathbf{1}_s(v) + \int_0^s \psi(u-v)du\Big) dv \tag{3.9}$$

$$= \bar{\lambda}_0 \Big(\int_{t\wedge s}^{t \vee s} du \int_0^{t\wedge s} \hat{\phi}(u-v)dv + K(t\wedge s)\Big), \tag{3.10}$$

where $\hat{\phi} : [0,\infty) \to [0,\infty)$ is defined as a function satisfying the integral equation:

$$\hat{\phi}(t) = \frac{\tilde{H}(t)}{1-||\tilde{H}||_1} + \int_0^\infty \tilde{H}(t+u)\hat{\phi}(u)du + \int_0^t \tilde{H}(t-u)\hat{\phi}(u)du, \tag{3.11}$$

and $K(t)$ is given by

$$K(t) = \frac{t}{1-||\tilde{H}||_1} + 2\int_0^t \int_0^u \hat{\phi}(u-v)dvdu.$$

We also have for $t \geq 0$, as $h \to \infty$,

$$\begin{aligned} \big(\hat{X}(t+h) - \hat{X}(h)\big) &\Rightarrow_h \hat{X}^\circ(t) \stackrel{d}{=} W(t), \\ \big(\hat{Y}(t+h) - \hat{Y}(h)\big) &\Rightarrow_h \hat{Y}^\circ(t) \stackrel{d}{=} \int_{-\infty}^{\infty} \Big(\int_0^t \psi(u-v)du\Big) W(dv). \end{aligned} \tag{3.12}$$

3.1.2. *The case where $\tilde{H} \equiv 1$, self-exciting with marks only.* The condition (iii), $||H||_1 < 1$, in the previous discussion is crucial in deriving stability, under which the family originated from a newborn is finite. In this subsection, we consider the case that $H(t,z) = z$ and $\bar{\lambda}_0(\cdot) \equiv \bar{\lambda}_0$. In this case the stability condition fails to hold, the point (s,z) in the system will produce new points in every exponential distributed time with parameter $\bar{\lambda}_0 z$, which makes the family infinite. That may be the reason why this special case has never been considered in the literature (to the best of our knowledge). It can be checked that neither the FLLN and FCLT holds in the conventional scaling regime as in (2.22) and (2.29). However, in the large intensity regime, the FLLN and FCLT we have established apply to this setting. We next discuss what the limits simplify to in this special case.

First, we have from Remark 3.1 for all $t \geq s \geq 0$,

$$G(t,s) = \mathbf{m}_{1,s} \quad \text{and} \quad \psi(t,s) = \mathbf{m}_{1,s} \exp\Big(\int_s^t \mathbf{m}_{1,u} du\Big), \tag{3.13}$$

which gives

$$1+\int_s^t \psi(t,u)du = \exp\Big(\int_s^t \mathbf{m}_{1,u}du\Big). \tag{3.14}$$

Therefore, applying Theorem 2.1,

$$\bar{N}(t) = \bar{\lambda}_0 \int_0^t \exp\Big(\int_0^u \mathbf{m}_{1,v}dv\Big)du. \tag{3.15}$$

In the case of stationary marks, we obtain $\bar{N}(t) = \frac{\bar{\lambda}_0}{\mathbf{m}_1}(e^{\mathbf{m}_1 t}-1)$. It is clear that unlike (2.21), we have

$$\frac{1}{n}\bar{N}(nt) \to \bar{\lambda}_0 \exp\Big(\int_0^\infty \mathbf{m}_{1,u}du\Big)$$

as $n\to\infty$, which is finite if and only if $\int_0^\infty \mathbf{m}_{1,u}du<\infty$.

We can also characterize $\hat{N}$ as follows whose proof is given in subsection 6.1.

Corollary 3.2. *Let $\eta(t) := \displaystyle\int_0^t \mathbf{m}_{1,u}du$ for $t>0$. Then the covariance for $\hat{N}$ is given by*

$$\begin{aligned}\hat{R}(t,s) = \bar{\lambda}_0 \int_0^{t\wedge s} e^{\eta(u)}du + \bar{\lambda}_0 \int_0^t e^{\eta(v)}dv \int_0^s e^{\eta(v')}dv' \int_0^{v\wedge v'} \mathbf{m}_{2,u}e^{-\eta(u)}du,\\ + \bar{\lambda}_0\int_0^s e^{\eta(u)}\eta(t\wedge u)du + \bar{\lambda}_0\int_0^t e^{\eta(u)}\eta(s\wedge u)du.\end{aligned} \tag{3.16}$$

If further the mark is stationary and we have $\mathbf{m}_{1,u}=\mathbf{m}_1$ and $\mathbf{m}_{2,u}=\mathbf{m}_2$, then

$$\hat{R}(t,s) = \bar{\lambda}_0\frac{\mathbf{m}_2-\mathbf{m}_1^2}{\mathbf{m}_1^2}\Big(\frac{e^{\mathbf{m}_1(t\wedge s)}-1}{\mathbf{m}_1} - \big(e^{\mathbf{m}_1 t}+e^{\mathbf{m}_1 s}\big)(t\wedge s)\Big) + \bar{\lambda}_0\frac{\mathbf{m}_2}{\mathbf{m}_1^3}\Big(e^{\mathbf{m}_1(t+s)} - e^{\mathbf{m}_1(t\vee s)}\Big).$$

Motivated by the previous stationary limit in Proposition 3.1, we are also interested in the limit of (3.6) for the example in this subsection. The proof is also given in subsection 6.1.

Proposition 3.2. *(i) If $\eta(\infty)<\infty$ and $\int_0^\infty \mathbf{m}_{2,v}dv<\infty$, we have*

$$\big(\hat{N}(t+h)-\hat{N}(h)\big) \Rightarrow_h B\big(\bar{\lambda}_0 e^{\eta(\infty)}\cdot t\big) + t\cdot\Big(\bar{\lambda}_0 e^{2\eta(\infty)}\int_0^\infty \mathbf{m}_{2,v}e^{-\eta(v)}dv\Big)^{1/2}\cdot\xi\,, \tag{3.17}$$

where ξ is a standard normal variable, independent of the standard Brownian motion B.

(ii) If $\eta(\infty)=\infty$ and $\int_0^\infty \mathbf{m}_{2,v}e^{-\eta(v)}dv<\infty$, letting $k(t)$ be the unique solution to $\int_0^{k(t)} e^{\eta(u)}du = t$ for all $t>0$, we have

$$\big(\hat{N}(k(t+h))-\hat{N}(k(h))\big) \Rightarrow_h B(\bar{\lambda}_0 t) + t\cdot\Big(\bar{\lambda}_0\int_0^\infty \mathbf{m}_{2,v}e^{-\eta(v)}dv\Big)^{1/2}\cdot\xi, \tag{3.18}$$

where ξ is a standard normal variable, independent of the standard Brownian motion B.

In particular, if $F_u=F$ for some common distribution function F on $\mathbb{R}_+$, then we can take $k(t)=\frac{\ln(t\mathbf{m}_1)}{\mathbf{m}_1}$ where $\mathbf{m}_1=\int_0^\infty zF(dz)$ so that the limit above reads

$$\big(\hat{N}(k(t+h))-\hat{N}(k(h))\big) \Rightarrow_h B(\bar{\lambda}_0 t) + t\cdot\Big(\frac{\bar{\lambda}_0\mathbf{m}_2}{\mathbf{m}_1}\Big)^{1/2}\cdot\xi. \tag{3.19}$$

3.2. **Indicator-type non-decomposable self-exciting functions.** In this subsection, we consider an indicator-type non-decomposable self-exciting function H, where the associated functions are denoted similarly with subscripts.

More specifically, we firstly take $H(t,z) = H_-(t,z) := H_0\mathbf{1}(t \in [0,z))$ for $t, z \in \mathbb{R}_+$. By scaling the functions and the process, without loss of generality, we take $H_0 = 1$. Recall the immigration-birth representation of the Hawkes process N in Section 2.1. The process $N_{l+1}(t)$ is the number of children (points) produced in the $(l+1)^{\text{th}}$ generation with the intensity process $\lambda_l(t)$ given in (2.2), that is,

$$\lambda_l(t) = \sum_{j=1}^{N_l(t)} \mathbf{1}(0 \le t - \tau_{lj} < Z_{lj}). \tag{3.20}$$

This may be interpreted that the birth rate of individuals in the l^{th} generation is only positive if their "age" $t - \tau_{lj}$ is less than Z_{lj}. Thus, the marks Z_{lj} can be regarded as a random threshold of the age of individuals at which they will stop reproducing. Thus, the model can be regarded as an immigration-birth model with "ceasing" reproduction. It can be found that Assumption 1 are always satisfied. We only need to assume the locally tightness of $\{F_u(\cdot), u \ge 0\}$, that is

$$\lim_{K\to\infty} \sup_{u\in[0,T]} \int_{|z|>K} F_u(dz) = 0 \quad \text{for every } T > 0,$$

to satisfy Assumption 3.

Then we take $H(t,z) = H_+(t,z) := \mathbf{1}(t \ge z)$ for $t, z \in \mathbb{R}_+$ under the same condition above, thus

$$\lambda_l(t) = \sum_{j=1}^{N_l(t)} \mathbf{1}(t - \tau_{lj} \ge Z_{lj}). \tag{3.21}$$

This may be interpreted that the birth rate of individuals in the l^{th} generation is only positive if their "age" $t - \tau_{lj}$ is greater than or equal to Z_{lj}. Thus, the marks Z_{lj} can be regarded as a random threshold of the age of individuals at which they will start reproducing. Thus, the model can be regarded an immigration-birth model with "delayed" reproduction.

For the last, we further take $H(t,z) = H(t,(z_1,z_2)) = \tilde{H}(t)\mathbf{1}(z_1 \le t < z_2)$ for $t \ge 0$ and $z = (z_1, z_2) \in \mathbb{R}_+^2$ being a two-dimensional mark, which is referred to as the case of varying reproduction rates with the presence of "ceasing" and "delayed" reproduction, that is,

$$\lambda(t) = \lambda_0(t) + \sum_{k=1}^{N(t)} \tilde{H}(t - \tau_k)\mathbf{1}(Z_k^{(1)} \le t - \tau_k < Z_k^{(2)}). \tag{3.22}$$

In additional to the locally tightness of $\{F_u(\cdot), u \ge 0\}$, we further assume the reproduction function $\tilde{H}$ being locally squared integrable to satisfy Assumptions 1 and 3. Since z_1 and z_2 representing the lower and the upper bound of age for reproduction, we always understand $\mathbb{P}\big(Z^2 > Z^1 \ge 0 \big| \tau = s\big) = 1$ for all $s > 0$ for the mark's distribution in (2.1).

The phenomenon of delayed and/or ceasing reproduction is common in biology, see, e.g., [41, 33, 16, 40]. The nonstationarity of the thresholds representing the delay or ceasing can be because of seasonal birth or environmental effects [47, 49]. There might be potential applications of Hawkes processes with such immigration-birth representations of delayed and/or ceased reproductions in biological or population dynamics.

Example 3.1. *In these models, the marks are associated with the 'age' thresholds of delayed and/or creasing reproductions. The piece-wise type of non-stationary distributions $F_t(\cdot)$ described in Example 2.1 can be clearly used in this setting to capture seasonal birth or environmental effects. We describe another type of non-stationarity: assume that $Z_k(\tau_k) = \int_{\tau_k}^{\tau_k + Z_k^*} \zeta(u)du$ where $\zeta(\cdot)$ is a deterministic function. Suppose that Z_k^*'s are iid with c.d.f. F. Then the distribution of $Z_k(\tau_k)$ given*

$\tau_k = t$ *is* $F_t(z) = F\big(\int_t^{t+z}\zeta(u)du\big)$. *One may interpret that the aging process is at a time-varying speed* $\zeta(\cdot)$*, so that the realized 'delay' or 'ceasing' thresholds have a non-stationary distribution.*

3.2.1. *The case of "ceasing" reproduction.* In this case, $H_-(v-u,z) = \mathbf{1}\big(v \in [u, u+z)\big)$ for all $v,u,z>0$. Recalling G^k and ψ defined in (2.3), we have $G_-(t,s) = F^c_s(t-s)$ for $t>s>0$, and

$$\psi_-(t,s) := \sum_{k\ge 1} G^k_-(t,s) = F^c_s(t-s) + \int_s^t \psi_-(t,u)F^c_s(u-s)du.$$

It follows from Remark 2.6 that for $t>u>0$ and $z>0$,

$$\mathcal{HU}\mathbf{1}_t(u,z) = \int_0^t \Big(1 + \int_0^t \psi_-(w,v)dw\Big)\mathbf{1}(z > v-u)\mathbf{1}(v\ge u)dv.$$

Corollary 3.3. *The covariance functions of* $(\hat X, \hat Y)$ *are given by*

$$\mathrm{Cov}\big(\hat X(t), \hat X(s)\big) = \int_0^{t\wedge s}\Big(\bar\lambda_0(u) + \int_0^u \psi_-(u,v)\bar\lambda_0(v)dv\Big)du,$$
$$\mathrm{Cov}(\hat X(t), \hat Y(s)) = \int_0^t \Big(\int_0^s \psi_-(w,u)dw\Big)\Big(\bar\lambda_0(u) + \int_0^u \psi_-(u,v)\bar\lambda_0(v)dv\Big)du,$$

and

$$\begin{aligned}\mathrm{Cov}(\hat Y(t), \hat Y(s)) = &\int_0^t dv\int_0^s dv'\Big(1+\int_0^t \psi_-(w,v)dw\Big)\Big(1+\int_0^s\psi_-(w,v')dw\Big)\\ &\times\int_0^{v\wedge v'} F^c_u\big((v\vee v')-u\big)\Big(\bar\lambda_0(u) + \int_0^u\psi_-(u,w)\bar\lambda_0(w)dw\Big)du.\end{aligned}$$

In the case of i.i.d. marks, that is $F_s = F$ for all $s>0$, we have $G_-(t,s) = F^c(t-s)$,

$$\psi_-(t,s) = \psi_-(t-s) = F^c(t-s) + \psi_- * F^c(t-s) \quad \text{for } t>s>0,$$

and ψ_- is the renewal density corresponding to F^c (which can be regarded as an improper probability density function F^c). Then

$$\begin{aligned}\mathrm{Cov}\big(\hat X(t), \hat X(s)\big) &= \bar\lambda_0 * 1(t\wedge s) + \bar\lambda_0 * \psi_- * 1(t\wedge s),\\ \mathrm{Cov}(\hat X(t), \hat Y(s)) &= \int_0^t \big(\psi_- * 1(s-u)\big)\big(\bar\lambda_0(u) + \psi_- * \bar\lambda_0(u)\big)du,\\ \mathrm{Cov}(\hat Y(t), \hat Y(s)) &= \int_0^t dv\int_0^s dv'\big(1+\psi_- * 1(t-v)\big)\big(1+\psi_- * 1(s-v')\big)\\ &\quad\times\int_0^{v\wedge v'} F^c\big((v\vee v')-u\big)\big(\bar\lambda_0(u) + \psi_- * \bar\lambda_0(u)\big)du.\end{aligned}$$

In the following, similar to Proposition 3.1, we present the asymptotic behavior of $\hat N$ at infinity, with time scaled down by the speed of $\bar N$ in Theorem 2.1.

Proposition 3.3. *Assume that* $\bar\lambda_0(t) \equiv \bar\lambda_0$ *for some constant* $\bar\lambda_0 > 0$. *Let* $\mathbf{m}_1 = \int_0^\infty y dF(y)$ *and* $\mathbf{m}_2 = \int_0^\infty y^2 dF(y)$. *The following hold.*

(i) If $\mathbf{m}_1 \in (0,1)$*, then we have*

$$\bar N(t+h) - \bar N(h) \to_h \frac{\bar\lambda_0}{1-\mathbf{m}_1}t,$$
$$\hat N(t+h) - \hat N(h) \Rightarrow_h W_1(t) + W_2(t),$$

where $\big(W_1, W_2\big)$ *is a two-dimensional Gaussian process such that*

$$\begin{aligned}\mathrm{Cov}\big(W_1(t), W_1(s)\big) &= \frac{\bar\lambda_0}{1-\mathbf{m}_1} t\wedge s,\\ \mathrm{Cov}\big(W_1(t), W_2(s)\big) &= \frac{\bar\lambda_0}{1-\mathbf{m}_1}\int_0^t du\int_0^s \psi_-(v-u)dv,\\ \mathrm{Cov}\big(W_2(t), W_2(s)\big) &= \frac{\bar\lambda_0}{1-\mathbf{m}_1}\int_{-\infty}^t dv\int_{-\infty}^s dv'\int_{|v-v'|}^\infty F^c(u)du\\ &\quad\times\Big(\mathbf{1}(v\ge 0)+\int_0^t \psi_-(w-v)dw\Big)\Big(\mathbf{1}(v'\ge 0)+\int_0^s \psi_-(w-v')dw\Big).\end{aligned}$$

(ii) If $\mathbf{m}_1 = 1$ *and* $\mathbf{m}_2<\infty$, *then we have*

$$\bar N\big(\sqrt{t+h}\big)-\bar N(\sqrt{h}) \to_h \frac{\bar\lambda_0}{\mathbf{m}_2}t,$$

$$\hat N\big(\sqrt{t+h}\big)-\hat N(\sqrt{h}) \Rightarrow_h \sqrt{\frac{\bar\lambda_0}{\mathbf{m}_2}}B(t)+\frac{\sqrt{\bar\lambda_0}}{\mathbf{m}_2}t\times\xi,$$

where ξ *is a standard normal variable, independent of the standard Brownian motion* B.

(iii) If $\mathbf{m}_1>1$, *letting* $\rho_->0$ *be the constant such that* $\int_0^\infty e^{-\rho_- y}F^c(y)dy = 1$, *and defining* $\mathfrak{f}(\rho_-) := \rho_-^2\int_0^\infty y e^{-\rho_- y}F^c(y)dy$, *then we have*

$$\bar N\Big(\frac{\ln(t+h)}{\rho_-}\Big)-\bar N\Big(\frac{\ln h}{\rho_-}\Big) \to_h \frac{\bar\lambda_0}{\mathfrak{f}(\rho_-)}t,$$

$$\hat N\Big(\frac{\ln(t+h)}{\rho_-}\Big)-\hat N\Big(\frac{\ln h}{\rho_-}\Big) \Rightarrow_h \sqrt{\frac{\bar\lambda_0}{\mathfrak{f}(\rho_-)}}B(t)+\frac{\sqrt{\bar\lambda_0}}{\mathfrak{f}(\rho_-)}t\times\xi,$$

where ξ *is a standard normal variable, independent of the standard Brownian motion* B.

3.2.2. *The case of "delayed" reproduction.* In this case, $H_+(v-u,z) = \mathbf{1}\big(v\ge z+u\big)$ for all $u,v,z>0$. We have $G_+(t,s) = F_s(t-s)$, and

$$\psi_+(t,s) := \sum_{k\ge 1} G_+^k(t,s) = F_s(t-s)+\int_s^t \psi_+(t,u)F_s(u-s)du,$$

for $t>s>0$. It follows from Remark 2.6 that for $t>u>0$ and $z>0$,

$$\mathcal{H}\mathcal{U}\mathbf{1}_t(u,z) = \int_0^t\Big(1+\int_0^t\psi_+(w,v)dw\Big)\mathbf{1}(z\le v-u)dv.$$

Corollary 3.4. *The covariance functions of* $(\hat X,\hat Y)$ *are given by*

$$\begin{aligned}\mathrm{Cov}\big(\hat X(t),\hat X(s)\big) &= \int_0^{t\wedge s}\Big(\bar\lambda_0(u)+\int_0^u\psi_+(u,v)\bar\lambda_0(v)dv\Big)du,\\ \mathrm{Cov}(\hat X(t),\hat Y(s)) &= \int_0^t\Big(\int_0^s\psi_+(w,u)dw\Big)\Big(\bar\lambda_0(u)+\int_0^u\psi_+(u,v)\bar\lambda_0(v)dv\Big)du,\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}(\hat Y(t),\hat Y(s)) &= \int_0^t dv\int_0^s dv'\Big(1+\int_0^t\psi_+(w,v)dw\Big)\Big(1+\int_0^s\psi_+(w,v')dw\Big)\\ &\quad\times\int_0^{v\wedge v'}F_u\big((v\wedge v')-u\big)\Big(\bar\lambda_0(u)+\int_0^u\psi_+(u,w)\bar\lambda_0(w)dw\Big)du.\end{aligned}$$

In the case of i.i.d. marks, that is $F_s = F$ for all $s > 0$, we have $G_+(t,s) = F(t-s)$,

$$\psi_+(t,s) = \psi_+(t-s) = F(t-s) + \psi_+ * F(t-s) \quad \text{for } t > s > 0,$$

and ψ_+ is the renewal density corresponding to F (which can be regarded as an improper probability density function). Then

$$\begin{aligned}
\mathrm{Cov}\big(\hat{X}(t), \hat{X}(s)\big) &= \bar{\lambda}_0 * 1(t \wedge s) + \bar{\lambda}_0 * \psi_+ * 1(t \wedge s), \\
\mathrm{Cov}(\hat{X}(t), \hat{Y}(s)) &= \int_0^t \big(\psi_+ * 1(s-u)\big)\big(\bar{\lambda}_0(u) + \psi_+ * \bar{\lambda}_0(u)\big)du, \\
\mathrm{Cov}(\hat{Y}(t), \hat{Y}(s)) &= \int_0^t dv \int_0^s dv' \big(1 + \psi_+ * 1(t-v)\big)\big(1 + \psi_+ * 1(s-v')\big) \\
&\qquad \times \int_0^{v \wedge v'} F\big((v \wedge v') - u\big)\big(\bar{\lambda}_0(u) + \psi_+ * \bar{\lambda}_0(u)\big)du.
\end{aligned} \tag{3.23}$$

We also have the following asymptotic behavior of $\hat{N}$ along with $\bar{N}$.

Proposition 3.4. *Assume that $\bar{\lambda}_0(t) \equiv \bar{\lambda}_0$ for some constant $\bar{\lambda}_0 > 0$. Let ρ_+ be the constant such that $\displaystyle\int_0^\infty e^{-\rho_+ y} F(y)dy = 1$ and define $\mathfrak{f}(\rho_+) := \rho_+^2 \int_0^\infty y e^{-\rho_+ y} F(y)dy$. Then we have*

$$\begin{aligned}
\bar{N}\Big(\frac{\ln(t+h)}{\rho_+}\Big) - \bar{N}\Big(\frac{\ln h}{\rho_+}\Big) &\to_h \frac{\bar{\lambda}_0}{\mathfrak{f}(\rho_+)}\, t, \\
\hat{N}\Big(\frac{\ln(t+h)}{\rho_+}\Big) - \hat{N}\Big(\frac{\ln h}{\rho_+}\Big) &\to_h \sqrt{\frac{\bar{\lambda}_0}{\mathfrak{f}(\rho_+)}} B(t) + \frac{\sqrt{\bar{\lambda}_0}}{\mathfrak{f}(\rho_+)} \sqrt{\frac{\int_0^\infty e^{-2\rho_+ y} F(y)dy}{1 - \int_0^\infty e^{-2\rho_+ y} F(y)dy} t} \times \xi,
\end{aligned}$$

where ξ is a standard normal variable, independent of the standard Brownian motion B.

3.2.3. *Varying reproduction rates in the cases of "ceasing" and "delayed" reproductions.* In both cases of "ceasing" and "delayed" reproductions above, the reproduction rates are assumed to be constant H_0 during the active reproduction periods. However, the reproduction rates may depend on the "ages" of individuals during the active reproduction periods. Specifically, the self-exciting function H is given by $H(t,z) = \tilde{H}(t)\mathbf{1}(t \in [0,z))$ and $H(t,z) = \tilde{H}(t)\mathbf{1}(t \ge z)$ for a nonnegative measurable function $\tilde{H} : \mathbb{R}_+ \to \mathbb{R}_+$, respectively, in the cases of "ceasing" and "delayed" reproductions. Thus, the intensity processes $\lambda_l(t)$ in (3.20) and (3.20) in the immigration-birth representations are, respectively, given by

$$\lambda_l(t) = \sum_{j=1}^{N_l(t)} \tilde{H}(t - \tau_{lj})\mathbf{1}(0 \le t - \tau_{lj} < Z_{lj}), \tag{3.24}$$

and

$$\lambda_l(t) = \sum_{j=1}^{N_l(t)} \tilde{H}(t - \tau_{lj})\mathbf{1}(t - \tau_{lj} \ge Z_{lj}). \tag{3.25}$$

There may be many possibilities for the function $\tilde{H}$, representing the varying nature of the reproduction rate as the "age" increases. The limit processes $(\hat{X}, \hat{Y})$ in Theorem 2.2 have covariance functions as given in Remark 2.6. In the case of "ceasing" reproduction with $H(t,z) = \tilde{H}(t)\mathbf{1}(t \in [0,z))$, we have $G(t,s) = \tilde{H}(t-s)F_s^c(t-s)$ for $t > s$,

$$\psi(t,s) = \tilde{H}(t-s)F_s^c(t-s) + \int_s^t \psi(t,u)\tilde{H}(u-s)F_s^c(u-s)du,$$

and

$$\mathcal{HU}\mathbf{1}_t(u,z) = \int_0^t \Big(1 + \int_0^t \psi(w,v)dw\Big)\tilde{H}(v-u)\mathbf{1}(v \in [u, u+z))dv.$$

Thus, the covariance functions $\mathrm{Cov}(\hat{X}(t), \hat{X}(s))$ and $\mathrm{Cov}(\hat{X}(t), \hat{Y}(s))$ are as given in Remark 2.6 using the function $\psi(t,s)$ above, while the covariance function

$$\begin{aligned}\mathrm{Cov}\big(\hat{Y}(t), \hat{Y}(s)\big) = &\int_0^t dv \int_0^s dv' \Big(1 + \int_0^t \psi(w,v)dw\Big)\Big(1 + \int_0^t \psi(w,v')dw\Big) \\ &\times \int_0^{v\wedge v'} \tilde{H}(v-u)\tilde{H}(v'-u)F_u^c\big((v\vee v') - u\big)\Big(\bar{\lambda}_0(u) + \int_0^u \psi(u,v)\bar{\lambda}_0(v)dv\Big)du.\end{aligned}$$

Similarly for the case of "delayed" reproduction.

A more general model is a Hawkes process $N(t)$ with the following intensity process:

$$\lambda(t) = \lambda_0(t) + \sum_{k=1}^{N(t)} \tilde{H}(t-\tau_k)\mathbf{1}(Z_k^{(1)} \le t - \tau_k < Z_k^{(2)}), \quad t \ge 0. \tag{3.26}$$

Here we assume that the pairs $\big\{\big(Z_k^{(1)}, Z_k^{(2)}\big), k \in \mathbb{N}\big\}$ are i.i.d. random vectors on $\mathbb{R}_+^2$ with $\mathbb{P}\big(Z_k^{(2)} > Z_k^{(1)}\big) = 1$. We denote by $F(dz_1, dz_2)$ the joint c.d.f. of $\big(Z_k^{(1)}, Z_k^{(2)}\big)$ and F_1 and F_2 the associated marginal c.d.f. In this model, the indicator function $\mathbf{1}(Z_k^{(1)} \le t - \tau_k < Z_k^{(2)})$ means that the self-exciting function $\tilde{H}$ takes effects only when the "age" $t - \tau_k$ of the event k is bigger than or equal to $Z_k^{(1)}$ and less than $Z_k^{(2)}$. The variable $Z_k^{(1)}$ can be regarded as the "delay" for the event k to take effect while $Z_k^{(2)}$ can be regarded as the threshold of the "age" of the event k to stop exciting. Then, it is evident that $Z_k^{(2)} - Z_k^{(1)}$ can be regarded as the active exciting duration for the event k.

For this general model, we have for $t \ge 0$,

$$\begin{aligned}H\big(t, (z_1, z_2)\big) &= \tilde{H}(t)\mathbf{1}(z_1 \le t < z_2) = \tilde{H}(t)\big(\mathbf{1}(z_1 \le t) - \mathbf{1}(z_2 \le t)\big), \\ \psi(t+s, s) = \psi(t) &= \tilde{H}(t)\big(F_1(t) - F_2(t)\big) + \int_0^t \psi(t-u)\tilde{H}(u)\big(F_1(u) - F_2(u)\big)du.\end{aligned}$$

Proposition 3.5. *Let $(\hat{X}, \hat{Y})$ be associated the limit Gaussian process of Hawkes with intensity in (3.26). Then the covariance functions of $(\hat{X}, \hat{Y})$ are given by the following: for $t, s \ge 0$,*

$$\mathrm{Cov}(\hat{X}(t), \hat{X}(s)) = \int_0^{t\wedge s} \big(\bar{\lambda}_0(u) + \psi * \bar{\lambda}_0(u)\big)du\,,$$

$$\mathrm{Cov}(\hat{X}(t), \hat{Y}(s)) = \int_0^t 1 * \psi(s-u)\big(\bar{\lambda}_0(u) + \psi * \bar{\lambda}_0(u)\big)du\,,$$

$$\begin{aligned}\mathrm{Cov}\big(\hat{Y}(t), \hat{Y}(s)\big) = &\int_0^t dv \int_0^s dv' \big(1 + \psi * 1(t-v)\big)\big(1 + \psi * 1(s-v')\big) \int_0^{v\wedge v'} \Big[\big(\bar{\lambda}_0(u) + \psi * \bar{\lambda}_0(u)\big) \\ &\times \tilde{H}(v-u)\tilde{H}(v'-u)\big(F_1(v\wedge v' - u) - F(v\wedge v' - u, v\vee v' - u)\big)\Big]du.\end{aligned}$$

4. Well-definedness of the processes X, Y and N

Since the sub-processes are defined related to inhomogeneous Poisson process with independent marks, we first give the following characterization, which is a direct result of the exponential formula for Poisson random measure, see for example [3, Chapter O.5]. As a consequence, the well-posedness and finiteness statements in the previous sections are ensured under Assumption 1(i), which is assumed throughout the paper.

Proposition 4.1. *Let (A, Z) be a marked Poisson process on the product space $\mathbb{R}_+ \times \mathbb{R}^d$ with characteristic measure $\mu(s)dsF_s(dz)$. Let $\{\tau_j, Z_j;\, j \geq 1\}$ be the occurrence times and the associated marks. For any bounded measurable function $f(s,z)$ on $\mathbb{R}_+ \times \mathbb{R}^d$ such that $f(s,z) = 0$ for all large $s > 0$ and $z \in \mathbb{R}^d$, we have for all $\theta \in \mathbb{R}$,*

$$\begin{aligned}
\mathbb{E}\Big[\sum_{j=1}^{\infty} f(\tau_j, Z_j)\Big] &= \int_0^\infty \int_{\mathbb{R}^d} f(s,z)\mu(s)dsF_s(dz), \\
\mathrm{Var}\Big(\sum_{j=1}^{\infty} f(\tau_j, Z_j)\Big) &= \int_0^\infty \int_{\mathbb{R}^d} f^2(s,z)\mu(s)dsF_s(dz), \qquad (4.1)\\
\mathbb{E}\Big[\exp\Big(i\theta\sum_{j=1}^{\infty} f(\tau_j, Z_j)\Big)\Big] &= \exp\Big(\int_0^\infty \int_{\mathbb{R}^d} \big(e^{i\theta f(s,z)} - 1\big)\mu(s)dsF_s(dz)\Big).
\end{aligned}$$

Let $g(s,z)$ be another bounded measurable function, we have

$$\mathrm{Cov}\Big(\sum_{j=1}^{\infty} f(\tau_j, Z_j), \sum_{j=1}^{\infty} g(\tau_j, Z_j)\Big) = \int_0^\infty \int_{\mathbb{R}^d} f(s,z)g(s,z)\mu(s)dsF_s(dz). \qquad (4.2)$$

For any $f \in \mathcal{B}_{b,c}$, recalling $\lambda_l f$ in (2.9) and $\mathcal{H}f$ in (2.7), since $\{N_l, Z_{lj}, j \geq 1\}$ fulfills the condition of Proposition 4.1 under $\mathbb{P}(\cdot|\mathscr{F}_{l-1})$, we obtain

$$\mathbb{E}[\lambda_l f|\mathscr{F}_{l-1}] = \int_0^\infty \int_{\mathbb{R}^d} \mathcal{H}f(s,z)\lambda_{l-1}(s)dsF_s(dz) = \int_0^\infty \lambda_{l-1}(s)\mathcal{G}f(s)ds = \lambda_{l-1}\mathcal{G}f \in \mathscr{G}_{l-1},$$

which gives (2.10) in Lemma 2.1. Moreover, (2.11) follows from the following

$$\begin{aligned}
\mathrm{Var}(\lambda_l f) &= \sum_{k=1}^{l} \mathbb{E}\Big[\Big(\mathbb{E}\big[\lambda_l f|\mathscr{F}_k\big] - \mathbb{E}\big[\lambda_l f|\mathscr{F}_{k-1}\big]\Big)^2\Big] \\
&= \sum_{k=1}^{l} \mathbb{E}\Big[\int_0^\infty \lambda_{l-k}(s)ds \int_{\mathbb{R}^d} F_s(dz)\Big(\mathcal{H}\mathcal{G}^{k-1}f(s,z)\Big)^2\Big] \\
&= \sum_{k=1}^{l} \int_0^\infty \mathcal{G}^{l-k}\Big(\int_{\mathbb{R}^d} \big(\mathcal{H}\mathcal{G}^{k-1}f\big)^2(\cdot,z)F_\cdot(dz)\Big)(s)\lambda_0(s)ds.
\end{aligned}$$

Lemma 4.1. *Let $\mathcal{G}$ and $\mathcal{H}$ be defined in (2.7) and φ_T be the function in Assumption 1(i).*

(i) Under Assumption 1(i), for every nonnegative $f, g \in \mathcal{B}_b[0,T]$, we have

$$\mathcal{G}^k f(s) \leq \int_0^T f(u)\varphi_T^{*k}(u-s)du,$$

$$\int_{\mathbb{R}^d} \mathcal{H}f(s,z)\mathcal{H}g(s,z)F_s(dz) \leq \Big(\int_0^T f(u)\varphi_T(u-s)du\Big)\Big(\int_0^T g(u)\varphi_T(u-s)du\Big)$$

for all $s \in [0,T]$ and $k \geq 1$, where φ_T^{*k} denotes the k^{th} convolution of φ_T. The function $\Phi_T(t) := \sum_{k\geq 1}\varphi_T^{*k}(t)$ is a well-defined integrable function on $[0,T]$.

(ii) If, in addition, Assumption 1(ii) holds, then $\Phi_T \in L^2[0,T]$.

Proof. Since $f \in \mathcal{B}_b[0,T]$, we have $\mathcal{G}f(s) = \int_0^T f(u)G(u,s)du = \int_0^T f(u)\int_{\mathbb{R}^d} H(u-s,z)F_s(dz)$. Under Assumption 1(i), applying the Cauchy–Schwarz inequality, we have

$$\mathcal{G}f(s) \leq \int_0^T f(u)du\Big(\int_{\mathbb{R}^d} H^2(u-s,z)F_s(dz)\Big)^{1/2} \leq \int_0^T f(u)\varphi_T(u-s)du.$$

The inequality for $\mathcal{G}^k f$ is proved by induction and the fact that $H(u-v,z) = 0$ for $u < v$.

Similarly, we have from Fubini's theorem and the Cauchy–Schwarz inequality that

$$\begin{aligned}\int_{\mathbb{R}^d}\mathcal{H}f(s,z)\mathcal{H}g(s,z)F_s(dz) &= \int_0^T\int_0^T f(u)g(v)dudv\Big(\int_{\mathbb{R}^d}H(u-s,z)H(v-s,z)F_s(dz)\Big)\\ &\le \int_0^T\int_0^T f(u)g(v)dudv\varphi_T(u-s)\varphi_T(v-s)\\ &= \Big(\int_0^T f(u)\varphi_T(u-s)du\Big)\Big(\int_0^T g(u)\varphi_T(u-s)du\Big).\end{aligned}$$

Denoting by $\varphi_{\delta,T}(t)=e^{-\delta t}\varphi_T(t)$ for $\delta>0$, such that $\int_0^T e^{-\delta u}\varphi_T(u)du=p<1$. It is true that $\int_0^T\varphi_T(t)dt<\infty$ if and only if $\int_0^T\varphi_{\delta,T}(t)dt<\infty$, and $\varphi_T^{*k}(t)=e^{\delta t}\varphi_{\delta,T}^{*k}(t)$ for $k\ge 1$. Thus,

$$\int_0^T\varphi_T^{*k}(t)dt\le\ e^{\delta T}\int_0^T\varphi_{\delta,T}^{*k}(t)dt\le e^{\delta T}p^k,$$

and

$$\int_0^T\Phi_T(t)dt=\ \int_0^T\sum_{k\ge1}\varphi_T^{*k}(t)dt\le e^{\delta T}\frac{p}{1-p}<\infty.$$

Therefore, Φ_T is well defined and integrable on $[0,T]$.

If, in addition, Assumption 1(ii) holds, that is, $\varphi_T\in L^2[0,T]$, then for every $k,l\ge1$, we have

$$\begin{aligned}\int_0^T\varphi_T^{*(k+1)}(t)\varphi_T^{*(l+1)}(t)dt &= \int_0^T\int_0^T\varphi_T^{*k}(s)ds\varphi_T^{*l}(r)dr\int_0^T\varphi_T(t-s)\varphi_T(t-r)dt\\ &\le\Big(\int_0^T\varphi_T^{*k}(s)ds\Big)\Big(\int_0^T\varphi_T^{*l}(r)dr\Big)\Big(\int_0^T\varphi_T^2(t)dt\Big)\end{aligned}$$

where $\varphi_T(u)=0$ for $u<0$ is used in the first identity, and the Cauchy–Schwarz inequality is used in the second line. Here the inequality also holds for $k,l\ge0$ with the understanding that $\int_0^T\varphi_T^{*0}(u)du=1$. The square integrability of Φ_T is thus proved by Fubini's theorem. □

Applying Lemma 4.1(i) and using Proposition 4.1, we have the following estimation for test functions in $\mathcal{B}_b[0,T]$ under Assumption 1(i), which will be used to prove the well-posedness in Proposition 4.2.

Corollary 4.1. *Under Assumption 1(i), for every nonnegative $f,g\in\mathcal{B}_b[0,T]$, we have*

$$\mathcal{U}f(s)=\sum_{l\ge0}\mathcal{G}^lf(s)\le f(s)+\int_0^T f(u)\Phi_T(u-s)du.$$

Thus, $\mathcal{U}f\in\mathcal{B}_b[0,T]$ is well defined. Applying Lemma 4.1(i) to (2.10) and (2.11), we obtain

$$\begin{aligned}\mathbb{E}\big[\lambda_lf\big] &\le\int_0^T\lambda_0(u)du\Big(\int_0^T f(v)\varphi_T^{*l}(v-u)dv\Big)=\int_0^T f(u)\varphi_T^{*l}*\lambda_0(u)du,\\ \mathrm{Cov}(\lambda_lf,\lambda_lg) &\le\sum_{k=1}^l\int_0^T\Big(\int_0^T f(v)\varphi_T^{*k}(v-u)dv\int_0^T g(v)\varphi_T^{*k}(v-u)dv\Big)\varphi_T^{*(l-k)}*\lambda_0(u)du,\end{aligned}\tag{4.3}$$

where we understand that $\varphi_T^{*0}*\lambda_0(u)=\lambda_0(u)$ and the inequalities above hold for all $l\ge0$.

With Lemmas 2.1 and 4.1, we can then prove Proposition 2.1, but we like to give a more detailed version as follows.

Proposition 4.2. *Under Assumption 1(i), Y_l,M_l,X,Y in (2.4) are well-defined stochastic processes, $X\in\mathbb{D}$ and $Y,Y_l\in\mathbb{C}$ for all $l\in\mathbb{N}$, and Y_l can be written as (2.13).*

Proof. Applying Corollary 4.1 to (2.13) directly, we see that $\lambda_l \mathcal{U}\mathbf{1}_t$ is well-defined for every $l \geq 0$. Moreover, for $0 \leq s < t \leq T$, let $\mathbf{1}_{s,t}(u) = \mathbf{1}_t(u) - \mathbf{1}_s(u) = \mathbf{1}(u \in (s,t])$, we have

$$0 \leq \mathcal{G}\mathcal{U}\mathbf{1}_t(u) - \mathcal{G}\mathcal{U}\mathbf{1}_s(u) = \mathcal{G}\mathcal{U}\mathbf{1}_{s,t}(u) \leq \int_0^\infty \mathbf{1}_{s,t}(v)\Phi_T(v-u)dv = \int_s^t \Phi_T(v-u)dv.$$

Thus, $t \to \mathcal{G}\mathcal{U}\mathbf{1}_t(u)$ is uniformly equicontinuous on $[0,T]$, which shows

$$Y_l(t) = \int_0^\infty \lambda_l(u)\big(\mathcal{G}\mathcal{U}\mathbf{1}_t(u) + \mathbf{1}_t(u)\big)du - \int_0^\infty \lambda_{l-1}(u)\mathcal{G}\mathcal{U}\mathbf{1}_t(u)du$$

is actually the difference of two increasing and continuous functions.

Applying Fubini's theorem, we have from (4.3),

$$\mathbb{E}\Big[\sum_{l\geq 1}\int_0^T \lambda_l(u)du\Big] = \sum_{l\geq 1}\mathbb{E}\big[\lambda_l \mathbf{1}_T\big] \leq \int_0^T \Phi_T * \lambda_0(u)du < \infty,$$

which gives $\mathbb{P}$-a.s. $\sum_{l\geq 1}\int_0^T \lambda_l(u)du$ is finitely valued. Thus, $\sum_{l\geq 1}\lambda_l\mathcal{U}\mathbf{1}_t$ is well-defined as well as $Y(t)$ for every $t \in [0,T]$, and $Y \in \mathbb{C}$ by the same reasoning above.

On the other hand, since $\mathbb{E}\big[X_l(t)\big|\mathscr{F}_{l-1}\big] = 0$, we have for every $t \in [0,T]$,

$$\mathbb{E}\Big[\Big(\sum_{l\geq 1} X_l(t)\Big)^2\Big] = \sum_{l\geq 1}\mathbb{E}\big[X_l^2(t)\big] = \sum_{l\geq 1}\mathbb{E}\big[\lambda_{l-1}\mathbf{1}_t\big] \leq \int_0^t \big(\lambda_0(u) + \Phi_T * \lambda_0(u)\big)du,$$

which shows that $X(t) = \sum_{l\geq 1} X_l(t)$ in $L^2(\mathbb{P})$. Since X_l^2 is an $\{\mathscr{F}_l(t)\}_{t\geq 0}$-submartingale under $\mathbb{P}(\cdot|\mathscr{F}_{l-1})$, by applying Doob's maximal inequality, we have

$$\begin{aligned}\varepsilon^2\mathbb{P}\Big(\max_{t\in[0,T]} X_l^2(t) > \varepsilon^2\Big) &\leq \mathbb{E}\big[X_l^2(T)\big] = \mathbb{E}\big[\lambda_{l-1}\mathbf{1}_T\big] \\ &\leq \int_0^T \varphi_T^{*(l-1)} * \lambda_0(u)du \leq e^{\delta T}p^{l-1}\int_0^T \lambda_0(u)du\end{aligned}$$

for $\delta > 0$ and $p \in (0,1)$ defined in the proof of Lemma 4.1. Let $\theta \in (p^{1/2}, 1)$. We then obtain

$$\begin{aligned}\mathbb{P}\Bigg(\sup_{t\in[0,T]}\Big|\sum_{l\geq m} X_l(t)\Big| \geq \varepsilon\Bigg) &\leq \mathbb{P}\Bigg(\sum_{k\geq 0}\sup_{t\in[0,T]}\big|X_{m+k}(t)\big| \geq \sum_{k\geq 0}\frac{\varepsilon\theta^k}{1-\theta}\Bigg) \\ &\leq \sum_{k\geq 0}\mathbb{P}\Bigg(\sup_{t\in[0,T]}\big|X_{m+k}(t)\big| \geq \frac{\varepsilon\theta^k}{1-\theta}\Bigg) \\ &\leq p^{m-1}\Bigg(\Big(\frac{1-\theta}{\varepsilon}\Big)^2 e^{\delta T}\int_0^T \lambda_0(u)du \cdot \sum_{k\geq 0}(p\theta^{-2})^k\Bigg). \end{aligned} \tag{4.4}$$

Applying the Borel–Cantelli lemma, we conclude that X is the limit of $\sum_{l=1}^m X_l$ a.s.-$\mathbb{P}$ under the uniform topology and thus $X \in \mathbb{D}$. □

5. Proofs of the functional limit theorems

This section is dedicated to the proofs of our main results. Throughout this section, $T > 0$ is a fixed constant, we focus on the interval $[0,T]$ and always assume $t, s \leq T$. We write $\varphi(t) = \varphi_T(t)$ and $\Phi_T(t) = \Phi(t)$ in Assumption 1 for this T, and further define $\varphi(t) = 0$ for $t < 0$ w.l.o.g., that is

$$\int_{\mathbb{R}^d} H^2(t-s,z)F_s(dz) \leq \varphi^2(t-s) \quad \text{for all } t, s \in [0,T].$$

We mainly focus on proving the FCLT for $\hat{N}^n$, Theorem 2.2 and 2.3, and the FWLLN for $\bar{N}$, Theorem 2.1 is proved thereafter. Recalling the identities for the pre-limit process $\hat{N}^n$ in (2.18)

$$\hat{N}^n(t) = \sum_{l\geq 1} \hat{M}_l^n(t) = \sum_{l\geq 1} \big(\hat{X}_l^n(t) + \hat{Y}_l^n(t)\big) = \hat{X}^n(t) + \hat{Y}^n(t)$$

and their definitions in (2.17), as well as $\mathcal{H}, \mathcal{G}, \mathcal{U}, \mathbf{1}_t$ and ϕ_t defined in (2.7) and (2.3), it holds that $\mathbf{1}_t(\cdot), \mathcal{G}^k\mathbf{1}_s(\cdot), \mathcal{U}\mathbf{1}_t(\cdot) \in \mathcal{B}_b[0,T]$. Instead of working on $\hat{N}$ directly, our proof starts from the convergence of subprocess $\hat{M}_l^n = \hat{X}_l^n + \hat{Y}_l^n$ in Section 5.1, and then the convergence of $\hat{N}^n = \hat{X}^n + \hat{Y}^n$ in Section 5.2, which is defined as the infinite sum of the subprocesses. Since $\hat{X}_l^n$ and $\hat{Y}_l^n$ are highly dependent but have nice path property as shown in Proposition 4.2, it is more convenient to work on the tightness of $\big(\hat{X}_l^n, \hat{Y}_l^n\big)$, and so is the joint discussion over $l \in \mathbb{N}$. Specifically, the proof of the convergence of the processes $\hat{N}^n$ proceeds in the following steps:

Step 1: The existence of the limit Gaussian processes $\hat{M}_l, \hat{N}$ and $\hat{X}_l, \hat{Y}_l$ with sample paths in $\mathbb{C}$ (Lemmas 5.1 and 5.2). Given their covariance functions in (2.26) and (2.27), it is sufficient to check their joint continuities.

Step 2: The convergence of finite dimensional distributions of $\hat{M}_l^n$ to $\hat{M}_l$, and $\hat{N}^n$ to $\hat{N}$, respectively (Lemmas 5.3 and 5.6). To overcome the dependency of $\hat{M}_l^n$ among $l \in \mathbb{N}$, the convergence is proved under conditional probability.

Step 3: Verifying the tightness criterion with the modulus of continuity as in [4, Theorem13.3] and completing the proofs (Lemmas 5.5 and 5.9, respectively). With the presence of non-stationary distribution marks, the noises captured by Y can only be expressed as an integral with respect to a martingale as stated in Section 2.2. This is in contrast with the stationary case seen in Remark 2.4, where the process Y is expressed as a simple integral functional of X. To tackle this challenge, we must investigate the moments of the increments of the associated processes directly.

5.1. **Convergence of the bracket processes.** Notice that under Assumption 2,

$$\sup_n \int_0^T \bar{\lambda}_0^n(u)du < \infty \quad \text{and} \quad \sup_{t\in[0,T]} \left| \int_0^t \bar{\lambda}_0^n(u)du - \int_0^t \bar{\lambda}_0(u)du \right| \to 0,$$

and $\bar{\lambda}_0^n$ and $\bar{\lambda}_0$ are deterministic functions. In the following, we understand that $\bar{\lambda}_0(u) = 0$ for $u < 0$.

Lemma 5.1. *Let $\hat{R}, \hat{R}_l$ be defined in (2.26) and (2.27). Under Assumption 1(i), for any $\delta > 0$,*

$$\sum_{l\geq 1} \sup_{\substack{|t-t'|\leq\delta, |s-s'|\leq\delta \\ t,t',s,s'\in[0,T]}} \left|\hat{R}_l(t',s') - \hat{R}_l(t,s)\right| \leq 2\left(1 + \int_0^T \Phi(u)du\right)^3 \sup_{t\in[0,T]} \int_{t-\delta}^t \bar{\lambda}_0(u)du. \tag{5.1}$$

If, in addition, Assumption 2 holds, there exist continuous modifications of the centered Gaussian processes $\{\hat{M}_l, l \in \mathbb{N}\}$ and $\hat{N}$ with covariance functions $\{\hat{R}_l, l \in \mathbb{N}\}$ and $\hat{R}$, respectively.

Proof. Recalling the covariance functions $\hat{R}_l, \hat{R}$ of the limit processes $\hat{M}_l, \hat{N}$, their existence as Gaussian processes follows from the consistency condition for the Gaussian distributional property. To prove $\hat{M}_l, \hat{N} \in \mathbb{C}$, it is sufficient to check the inequalities in the lemma.

For every $0 \leq t < t' \leq T$ and $s \in [0,T]$, we have by definition (2.3) that

$$\big(\phi_{t'}(v,z) - \phi_t(v,z)\big)\phi_s(v,z) = \big(\mathbf{1}_{t,t'}(v) + \mathcal{H}\mathcal{U}\mathbf{1}_{t,t'}(v,z)\big)\big(\mathbf{1}_s(v) + \mathcal{H}\mathcal{U}\mathbf{1}_s(v,z)\big)$$

where $\mathbf{1}_{t,t'}(v) := \mathbf{1}_{t'}(v) - \mathbf{1}_t(v) = \mathbf{1}(v \in (t,t'])$. Applying Lemma 4.1(i), we have

$$0 \leq \int_{\mathbb{R}^d} \big(\phi_{t'}(v,z) - \phi_t(v,z)\big)\phi_s(v,z)F_v(dz)$$

$$\leq \Big(\mathbf{1}_{t,t'}(v) + \int_0^T \mathbf{1}_{t,t'}(u)\Phi(u-v)du\Big)\Big(\mathbf{1}_s(v) + \int_0^T \mathbf{1}_s(u)\Phi(u-v)du\Big)$$
$$\leq \Big(\mathbf{1}_{t,t'}(v) + \int_0^T \mathbf{1}_{t,t'}(u)\Phi(u-v)du\Big)\Big(1 + \int_0^T \Phi(u)du\Big).$$

Therefore, for $t, t', s, s' \in [0,T]$ with $t' > t$ and $s' > s$, we have from the inequality above that

$$\begin{aligned}
&\hat{R}_l(t',s') - \hat{R}_l(t,s) = \hat{R}_l(t',s') - \hat{R}_l(t,s') + \hat{R}_l(t,s') - \hat{R}_l(t,s)\\
&\leq \Big(1 + \int_0^T \Phi(u)du\Big)\int_0^T \mathbf{1}_{t,t'}(u)\Big(\varphi^{*(l-1)} * \bar{\lambda}_0(u) + \Phi * \varphi^{*(l-1)} * \bar{\lambda}_0(u)\Big)du\\
&\quad + \Big(1 + \int_0^T \Phi(u)du\Big)\int_0^T \mathbf{1}_{s,s'}(u)\Big(\varphi^{*(l-1)} * \bar{\lambda}_0(u) + \Phi * \varphi^{*(l-1)} * \bar{\lambda}_0(u)\Big)du\\
&\leq \Big(1 + \int_0^T \Phi(u)du\Big)^2\Big(\int_0^T \varphi^{*(l-1)}(u)du\Big)\Big(\sup_{v\in[0,T]}\int_0^T \big(\mathbf{1}_{s,s'}(u) + \mathbf{1}_{t,t'}(u)\big)\bar{\lambda}_0(u-v)du\Big)
\end{aligned}$$

for every $l \geq 2$, where we make use of the fact in the last inequality that for $f, g, h \geq 0$,

$$\int_0^T h(u) f * g(u)du \leq \int_0^T f(u)du \sup_{v\in[0,T]}\int_0^T h(u)g(u-v)du,$$

and the inequality also holds for the case $l = 1$ with $\big(\int_0^T \varphi^{*(0)}(u)du\big)$ understood as 1. Similarly for the other cases of (t, t', s, s'). Summing over $l \in \mathbb{N}$ proves the inequality (5.1). $\square$

Lemma 5.2. *For the covariance functions for $\hat{X}_l, \hat{Y}_l$ in Theorem 2.2, under Assumption 1(i), for every $\delta > 0$ and $s, s', t, t' \in [0,T]$ with $|s'-s| \vee |t'-t| \leq \delta$,*

$$\Big|\mathbb{E}\big[\hat{X}_l(t')\hat{X}_l(s')\big] - \mathbb{E}\big[\hat{X}_l(t)\hat{X}_l(s)\big]\Big| \leq 2\Big(\int_0^T \varphi^{*(l-1)}(u)du\Big) \times \sup_{v\in[0,T]}\int_{v-\delta}^v \bar{\lambda}_0(u)du\,,$$
$$\Big|\mathbb{E}\big[\hat{Y}_l(t')\hat{Y}_l(s')\big] - \mathbb{E}\big[\hat{Y}_l(t)\hat{Y}_l(s)\big]\Big| \leq 2\Big(\int_0^T \Phi(u)du\Big)^2\Big(\int_0^T \varphi^{*(l-1)}(u)du\Big) \times \sup_{v\in[0,T]}\int_{v-\delta}^v \bar{\lambda}_0(u)du.$$

Thus we conclude the existence of continuous sample paths of $(\hat{X}_l, \hat{Y}_l)$, summing over $l = 1, 2, \cdots$ also proves that for the processes $(\hat{X}, \hat{Y})$.

Proof. The inequality for the covariance functions for $\hat{X}_l$ and $\hat{Y}_l$ can be derived following the same procedure as the proof of Lemma 5.1. Under Assumption 1(i), we have from Lemma 4.1 that $\sum_{l\geq 1}\int_0^T \varphi^{*(l-1)}(u)du < \infty$ which gives the assertion for $(\hat{X}, \hat{Y})$. $\square$

To prove the convergence of the finite-dimensional distributions of $\hat{M}_l^n$, the continuity theorem for characteristic functions is used. Noticing that applying Proposition 4.1 to (2.17), the covariance function of $\hat{M}_l^n(t)$ is given by, for $s, t \geq 0$,

$$\begin{aligned}
\hat{R}_l^n(t,s) &:= \mathbb{E}\big[\hat{M}_l^n(t)\hat{M}_l^n(s)\big] = \mathbb{E}\big[\mathbb{E}\big[\hat{M}_l^n(t)\hat{M}_l^n(s)\big|\mathscr{F}_{l-1}^n\big]\big]\\
&= \int_0^\infty \mathcal{G}^{l-1}\Big(\int_{\mathbb{R}^d}\phi_t(\cdot,z)\phi_s(\cdot,z)F_\cdot(dz)\Big)(u)\bar{\lambda}_0^n(u)du \to \hat{R}_l(t,s)
\end{aligned} \tag{5.2}$$

where $\hat{R}_l$ is defined in (2.27), as $n \to \infty$ under Assumption 2.

Lemma 5.3. *Under Assumptions 1(i), 2 and 3, the finite-dimensional distributions of the processes $\hat{M}_l^n$ converge to those of $\hat{M}_l$, and the limit processes $\{\hat{M}_l, l \geq 1\}$ are mutually independent.*

Proof. We consider first the limit distribution of $\hat{M}_l^n(t_0)$ at fixed $t_0 \in [0,T]$, $l \geq 1$. We need the following inequalities:

$$\big|e^u - e^v\big| \leq |u-v| \quad \text{for all complex numbers } u, v \text{ with } \Re(u), \Re(v) \leq 0, \tag{5.3}$$

$$|e^{iu} - 1 - iu| \leq \frac{1}{2}|u|^2 \quad \text{and} \quad \Big|e^{iu} - 1 - iu + \frac{u^2}{2}\Big| \leq \frac{1}{6}|u|^3, \quad \text{for all } u \in \mathbb{R}. \tag{5.4}$$

Applying Proposition 4.1 to $\hat{M}_l^n$ in (2.17), we have for all $\theta \in \mathbb{R}$,

$$\mathbb{E}\Big[\exp(i\theta \hat{M}_l^n(t_0))\Big|\mathscr{F}_{l-1}^n\Big] = \exp\bigg(\int_0^T\int_{\mathbb{R}^d}\Big(e^{\frac{i\theta}{\sqrt{n}}\phi_{t_0}(s,z)} - 1 - \frac{i\theta}{\sqrt{n}}\phi_{t_0}(s,z)\Big)F_s(dz)\lambda_{l-1}^n(s)ds\bigg).$$

Making use of (5.4), for every $\tilde{K}_1 > 0$, we obtain that there are complex numbers θ_1, θ_2 with $|\theta_1|, |\theta_2| \leq 1$, such that

$$\begin{aligned}
&e^{\frac{i\theta}{\sqrt{n}}\phi_{t_0}(s,z)} - 1 - \frac{i\theta}{\sqrt{n}}\phi_{t_0}(s,z)\\
&\quad = -\frac{\theta^2}{2n}\phi_{t_0}^2(s,z) + \theta_1\frac{\theta^2}{n}\phi_{t_0}^2(s,z)\mathbf{1}(|z| > \tilde{K}_1) + \frac{\theta_2}{6}\frac{\theta^3}{n^{3/2}}\phi_{t_0}^3(s,z)\mathbf{1}(|z| \leq \tilde{K}_1).
\end{aligned}$$

Further applying (5.3) and (5.2) gives $\mathbb{P}\big(\cdot\,\big|\mathscr{F}_{l-1}^n\big)$-a.s.

$$\begin{aligned}
&\bigg|\mathbb{E}\big[\exp(i\theta\hat{M}_l^n(t_0))\big|\mathscr{F}_{l-1}^n\big] - \exp\Big(-\frac{\theta^2}{2}\hat{R}_l^n(t_0,t_0)\Big)\bigg|\\
&\leq \bigg|\int_0^T\int_{\mathbb{R}^d}\Big(e^{\frac{i\theta}{\sqrt{n}}\phi_{t_0}(s,z)} - 1 - \frac{i\theta}{\sqrt{n}}\phi_{t_0}(s,z)\Big)F_s(dz)\lambda_{l-1}^n(s)ds + \frac{\theta^2}{2}\hat{R}_l^n(t_0,t_0)\bigg|\\
&\leq \frac{\theta^2}{2}\bigg|\int_0^T\Big(\int_{\mathbb{R}^d}\phi_{t_0}^2(s,z)F_s(dz)\Big)\bar{\lambda}_{l-1}^n(s)ds - \hat{R}_l^n(t_0,t_0)\bigg|\\
&\qquad + \theta^2\int_0^T\int_{|z|>\tilde{K}_1}\phi_{t_0}^2(s,z)F_s(dz)\bar{\lambda}_{l-1}^n(s)ds\\
&\qquad + \frac{|\theta|^3}{6}\frac{1}{\sqrt{n}}\int_0^T\int_{|z|\leq\tilde{K}_1}\phi_{t_0}^3(s,z)F_s(dz)\bar{\lambda}_{l-1}^n(s)ds\\
&=: \frac{\theta^2}{2}I_1^n + \theta^2 I_2^n + \frac{|\theta|^3}{6}I_3^n.
\end{aligned}$$

Therefore, given the fact (5.2) under Assumption 2, we only need to show that $I_k^n (k = 1,2,3)$ goes to zero as $n \to \infty$. Notice that under Assumption 1(i), for some $\tilde{K}_2 > 0$,

$$\phi_{t_0}(s,z) \leq \tilde{K}_2\Big(1 + \int_0^T H(t-s,z)dt\Big) \quad \text{for all } s \in [0,T],\ z \in \mathbb{R}^d. \tag{5.5}$$

For $l = 1$, $I_1^n = 0$ by definition. For $l \geq 2$, we have from Lemma 2.1 that

$$\mathbb{E}\big[(I_1^n)^2\big] = \frac{1}{n^2}\mathrm{Var}\bigg(\lambda_{l-1}^n\Big(\int_{\mathbb{R}^d}\phi_{t_0}^2(\cdot,z)F_\cdot(dz)\Big)\bigg).$$

Further applying (4.3), (5.5) and Lemma 4.1(i) gives

$$\begin{aligned}
\mathbb{E}\big[(I_1^n)^2\big] &\leq \frac{1}{n^2}\sum_{k=1}^{l-1}\int_0^T\bigg(\int_0^T\Big(\int_{\mathbb{R}^d}\phi_{t_0}^2(t,z)F_t(dz)\Big)\varphi^{*k}(t-s)dt\bigg)^2\varphi^{*(l-1-k)} * \lambda_0^n(s)ds\\
&\leq \frac{\tilde{K}_2^4}{n^2}\Big(1 + \int_0^T\varphi(t)dt\Big)^4\sum_{k=1}^{l-1}\int_0^T\Big(\int_0^T\varphi^{*k}(t-s)dt\Big)^2\varphi^{*(l-1-k)} * \lambda_0^n(s)ds
\end{aligned}$$

$$= O\bigg(\frac{1}{n^2}\int_0^T \lambda_0^n(s)ds\bigg).$$

For I_2^n, we have from (5.5),

$$\int_{|z|>\tilde{K}_1} \phi_{t_0}^2(s,z)F_s(dz) \le \tilde{K}_2^2 \int_{|z|>\tilde{K}_1} \bigg(1+\int_0^T H(t-s,z)dt\bigg)^2 F_s(dz),$$

which goes to zero as $\tilde{K}_1 \to \infty$ by Assumption 3(ii), and uniformly in $s \in [0,T]$. This means that for all $\varepsilon > 0$, we can take $\tilde{K}_1 > 0$ such that

$$\mathbb{E}\big[I_2^n\big] \le \varepsilon \mathbb{E}\big[\bar{\lambda}_{l-1}^n \mathbf{1}_T\big] \le \varepsilon \int_0^T \varphi^{*(l-1)} * \bar{\lambda}_0^n(t)dt.$$

For I_3^n, we have from (5.5) that under Assumption 3(i),

$$\sup_{t\in[0,T],|z|\le\tilde{K}_1} |\phi_{t_0}(t,z)| < \infty.$$

Thus, we have

$$\begin{aligned}\mathbb{E}\big[I_3^n\big] &\le \sup_{t\in[0,T],|z|\le\tilde{K}_1} \frac{|\phi_{t_0}^3(t,z)|}{\sqrt{n}}\mathbb{E}\big[\bar{\lambda}_{l-1}^n\mathbf{1}_T\big] \\ &\le \sup_{t\in[0,T],|z|\le\tilde{K}_1} \frac{|\phi_{t_0}^3(t,z)|}{\sqrt{n}}\bigg(\int_0^T \varphi^{*(l-1)} * \bar{\lambda}_0^n(u)du\bigg).\end{aligned}$$

Therefore, we have shown that as $n\to\infty$, in probability, for $l \ge 1$,

$$\mathbb{E}\Big[\exp(i\theta \hat{M}_l^n(t_0))\Big|\mathscr{F}_{l-1}^n\Big] \Rightarrow \exp\bigg(\frac{-\theta^2}{2}\hat{R}_l(t_0,t_0)\bigg).$$

For the convergence of the finite dimensional distributions of $\hat{M}_l^n$, it is sufficient to consider the characteristic function of the linear combinations of $\{\hat{M}_l^n(t), t\in[0,T]\}$ over a finite number of t. The proof above stays the same but replacing $\mathbf{1}_{t_0}$ by the corresponding linear combinations of $\{\mathbf{1}_t(s),\, t\in[0,T]\}$, which is always a bounded function in s and the corresponding $I_k^n, k=1,2,3$ always converge to zero as $n\to\infty$. Moreover, noticing that $\hat{M}_{l-1}^n \in \mathscr{F}_{l-1}^n$ and the limit is proved by conditioning on $\mathscr{F}_{l-1}^n$, we can also conclude the independency of the limit processes $\hat{M}_l$ over l. This finishes the proof. □

Following the same procedure, we can also have the convergence of finite-dimensional distribution for $(\hat{X}_l^n, \hat{Y}_l^n)$ with $\hat{M}_l^n$ replaced by the linear sum of $\hat{X}_l^n$ and $\hat{Y}_l^n$ in the proof above.

Lemma 5.4. *Under Assumptions 1(i), 2 and 3, the finite-dimensional distributions of the processes $(\hat{X}_l^n, \hat{Y}_l^n)$ to those of $(\hat{X}_l, \hat{Y}_l)$, which is the 2-dimensional Gaussian process with the desired covariance functions in Theorem 2.2, and the limit processes $\{(\hat{X}_l, \hat{Y}_l), l\ge 1\}$ are mutually independent.*

For the tightness of the auxiliary processes in Theorem 2.2, we mainly focus on that of $(\hat{X}_l^n, \hat{Y}_l^n)$ on the product space with weak topology, for which it is sufficient to check the tightness of each component. Then, the tightness of $\hat{M}_l^n$ in $(\mathbb{D}, J_1)$ follows from the identity $\hat{M}_l^n = \hat{X}_l^n + \hat{Y}_l^n$ in (2.17) and the fact $\hat{Y}_l^n \in \mathbb{C}$ shown in Proposition 4.2. Recall the modules of continuity

$$\omega''(X,\delta) := \sup_{\substack{s<r<t\le s+\delta \\ s,r,t\in[0,T]}} \big|X(t)-X(r)\big| \wedge \big|X(r)-X(s)\big| \quad \text{and} \quad \omega(Y,\delta) := \sup_{\substack{s<t\le s+\delta \\ s,t\in[0,T]}} \big|Y(t)-Y(s)\big|,$$

for $X\in\mathbb{D}, Y\in\mathbb{C}$. In the proof, we always assume $0\le s<r<t\le T$.

Lemma 5.5. *Under Assumption 1, for every $l \in \mathbb{N}$, there exists some constant $K_l > 0$, independent of $\bar\lambda_0^n$, so that for all $\varepsilon > 0$ and $\delta > 0$,*

$$\begin{aligned}
\mathbb{P}\big(\omega''(\hat X_1^n,\delta)\ge\varepsilon\big) &\le \frac{K_1}{\varepsilon^4}\bigg(\int_0^T \bar\lambda_0^n(u)du\bigg)\sup_{t\in[0,T]}\bigg(\int_{t-2\delta}^t \bar\lambda_0^n(u)du\bigg),\\
\mathbb{P}\big(\omega''(\hat X_{l+1}^n,\delta)\ge\varepsilon\big) &\le \frac{K_{l+1}\delta}{\varepsilon^4}\bigg(1+\int_0^T \bar\lambda_0^n(u)du\bigg)^2,\\
\mathbb{P}\Big(\omega(\hat Y_l^n,\delta)\ge\varepsilon\Big) &\le \frac{K_l\delta}{\varepsilon^2}\bigg(\int_0^T \bar\lambda_0^n(u)du\bigg).
\end{aligned} \tag{5.6}$$

Proof. We follow the idea from the maximal inequality in [4, Theorem 10.3].

For the increment of $\hat Y_l^n$ in (5.6), we considering an arbitrary $[a,b]\subset[0,T]$, and let D_k be the set of dyadic rationals $t_{kj} = a+(b-a)\frac{j}{2^k}, 0\le j\le 2^k$ on $[a,b]$, and $D=\cup_{k\ge1}D_k$ be a dense subset of $[a,b]$. By the continuity of $\hat Y_l^n$, we have

$$\sup_{\substack{s<t\\ s,t\in[a,b]}}\big|\hat Y_l^n(t)-\hat Y_l^n(s)\big| = \sup_{\substack{s<t\\ s,t\in D}}\big|\hat Y_l^n(t)-\hat Y_l^n(s)\big| \le 2\sum_{k\ge1}\sup_{\substack{2^k(t-s)=b-a\\ s,t\in D_k}}\big|\hat Y_l^n(t)-\hat Y_l^n(s)\big|, \tag{5.7}$$

similar to the last inequality at [4, page 109]. This can also be understood from the fact that every $s,t\in[a,b]$ can be uniquely approximated by $\sum_{k\ge1}s_k$ and $\sum_{k\ge1}t_k$ with $s_k,t_k\in\big\{0,\frac{(b-a)}{2^k}\big\}$, and the inequality above follows from the triangle inequality.

Applying Proposition 4.1 and Lemma 4.1(i) to (2.17), we have

$$\begin{aligned}
\mathbb{E}\big[\big(\hat Y_l^n(t)-\hat Y_l^n(s)\big)^2\big|\mathscr{F}_{l-1}^n\big] &= \frac1n\int_0^T\int_{\mathbb{R}^d}\big(\mathcal{H}\mathcal{U}\mathbf{1}_{s,t}(u,z)\big)^2F_u(dz)\lambda_{l-1}^n(u)du\\
&\le \int_0^T\bigg(\int_0^T \mathbf{1}_{s,t}(v)\Phi(v-u)dv\bigg)^2\bar\lambda_{l-1}^n(u)du\\
&\le (t-s)\int_0^T\int_0^T \mathbf{1}_{s,t}(v)\Phi^2(v-u)\bar\lambda_{l-1}^n(u)dudv
\end{aligned}$$

where the Cauchy–Schwarz inequality is applied in the last inequality. Further applying Corollary 4.1, we have

$$\mathbb{E}\big[\big(\hat Y_l^n(t)-\hat Y_l^n(s)\big)^2\big] \le (t-s)\int_0^T \mathbf{1}_{s,t}(u)\Phi^2*\varphi^{*(l-1)}*\bar\lambda_0^n(u)du. \tag{5.8}$$

Summing over all choices of $(s,t)=(t_{k(j-1)},t_{kj})$, we have for $k\ge1$,

$$\mathbb{P}\Bigg(\sup_{\substack{2^k(t-s)=b-a\\ s,t\in D_k}}\big|\hat Y_l^n(t)-\hat Y_l^n(s)\big|\ge\varepsilon\Bigg) \le \frac{b-a}{2^k\varepsilon^2}\int_0^T \mathbf{1}_{a,b}(u)\Phi^2*\varphi^{*(l-1)}*\bar\lambda_0^n(u)du. \tag{5.9}$$

Finally, for $\theta\in(1/\sqrt2,1)$, we have from (5.7) that

$$\begin{aligned}
&\mathbb{P}\Bigg(\sup_{\substack{s<t\\ s,t\in[a,b]}}\big|\hat Y_l^n(t)-\hat Y_l^n(s)\big|\ge\varepsilon\Bigg)\\
&\le \mathbb{P}\Bigg(\sum_{k\ge1}\sup_{\substack{2^k(t-s)=b-a\\ s,t\in D_k}}\big|\hat Y_l^n(t)-\hat Y_l^n(s)\big| \ge \frac{\varepsilon(1-\theta)}{2\theta}\sum_{k\ge1}\theta^k\Bigg)\\
&\le \Bigg(\sum_{k\ge1}\frac{(b-a)}{2^k\theta^{2k}}\frac{4\theta^2}{\varepsilon^2(1-\theta)^2}\Bigg)\int_0^T \mathbf{1}_{a,b}(u)\Phi^2*\varphi^{*(l-1)}*\bar\lambda_0^n(u)du
\end{aligned}$$

$$= \frac{4(b-a)\theta^2}{\varepsilon^2(2\theta^2-1)(1-\theta)^2}\int_0^T \mathbf{1}_{a,b}(u)\Phi^2 * \varphi^{*(l-1)} * \bar{\lambda}_0^n(u)du. \tag{5.10}$$

Then, (5.6) for $\hat{Y}_l^n$ is proved by applying [4, Theorem 7.4].

For the increment of $\hat{X}_l^n$ in (5.6), noticing that N_l^n is an inhomogeneous Poisson process with intensity λ_{l-1}^n under $\mathbb{P}(\cdot|\mathscr{F}_{l-1}^n)$, we have

$$\begin{aligned}&\mathbb{E}\big[(\hat{X}_l^n(t)-\hat{X}_l^n(r))^2(\hat{X}_l^n(r)-\hat{X}_l^n(s))^2\big|\mathscr{F}_{l-1}^n\big]\\ =&\frac{1}{n^2}\int_r^t \lambda_{l-1}^n(u)du\int_s^r \lambda_{l-1}^n(u)du \le \frac{1}{n^2}\Big(\int_s^t \lambda_{l-1}^n(u)du\Big)^2 = \big(\bar{\lambda}_{l-1}^n\mathbf{1}_{s,t}\big)^2.\end{aligned} \tag{5.11}$$

The inequality (5.6) for $\hat{X}_1^n$ is classical by applying [4, Theorem 10.4].

For $l \ge 2$, applying the Cauchy–Schwarz inequality and (4.3), we have

$$\begin{aligned}\big(\lambda_0^n\mathcal{G}^{l-1}\mathbf{1}_{s,t}\big)^2 &\le (t-s)\Big(\int_0^T \lambda_0^n(v)dv\Big)\Big(\int_0^T \lambda_0^n(v)dv\int_0^T \mathbf{1}_{s,t}(u)\big(\varphi^{*(l-1)}(u-v)\big)^2du\Big),\\ \mathrm{Var}(\lambda_{l-1}\mathbf{1}_{s,t}) &\le (t-s)\sum_{k=1}^{l-1}\int_0^T\int_0^T \mathbf{1}_{s,t}(u)\big(\varphi^{*k}(u-v)\big)^2\varphi^{*(l-1-k)} * \lambda_0(v)dvdu.\end{aligned}$$

This implies that

$$\begin{aligned}\mathbb{E}\big[\big(\bar{\lambda}_{l-1}^n\mathbf{1}_{s,t}\big)^2\big] \le (t-s)&\Big(\int_0^T \bar{\lambda}_0^n(v)dv\Big)\Big(\int_0^T \mathbf{1}_{s,t}(u)\big(\varphi^{*(l-1)}\big)^2 * \bar{\lambda}_0^n(u)dv\Big)\\ &+\frac{(t-s)}{n}\Big(\int_0^T \mathbf{1}_{s,t}(u)\sum_{k=1}^{l-1}\big(\varphi^{*k}\big)^2 * \varphi^{*(l-1-k)} * \bar{\lambda}_0^n(u)du\Big).\end{aligned}$$

Then, the probability bound for the maximal value of $\big|\hat{X}_l^n(t)-\hat{X}_l^n(r)\big| \wedge \big|\hat{X}_l^n(r)-\hat{X}_l^n(s)\big|$ over all choose of adjacent triples $(s,r,t) \in D_k \subset [a,b]$ (that is, $t-r=r-s=\frac{(b-a)}{2^k}$) can thus be found, similar to (5.9). And the maximal inequalities for $\hat{X}_l^n$ in (5.6) is derived similar to the proof of [4, Theorem 10.3]. This completes the proof. □

5.2. **Convergence of the Hawkes processes.** In this subsection, we focus on the infinite sum processes $\hat{N}^n = \sum_{l\ge1}\hat{M}_l^n, \hat{X}^n = \sum_{l\ge1}\hat{X}_l^n$ and $\hat{Y}^n = \sum_{l\ge1}\hat{Y}_l^n$ in (2.18). Notice that $\hat{X}_l^n, \hat{Y}_l^n, \hat{M}_l^n$ are not independent over l, and $\hat{X}^n \in \mathbb{D}, \hat{Y}^n \in \mathbb{C}$ proved in Proposition 4.2. In the proof of the convergence of finite–dimensional distribution, Lemmas 5.3 and 5.4 are applied. However, in the proof of tightness, Lemma 5.5 cannot be applied directly, since the modules of continuity ω'' is not sub-additive in $\mathbb{D}$. We have to calculate and estimate the fourth moment of $\hat{X}^n$.

Lemma 5.6. *Under Assumption 1(i), 2 and 3, the finite–dimensional distribution of the process $\hat{N}^n$ converges to that of $\hat{N}$.*

Proof. Consider first the limit distribution of $\sum_{l\ge1}\hat{M}_l^n(t_0)$ for fixed $t_0 \in [0,T]$, where we make use of [4, Theorem 3.2] by showing that

$$\sum_{l\ge1}\hat{M}_l^n(t_0) = \lim_{m\to\infty}\sum_{l=1}^m \hat{M}_l^n(t_0) \Rightarrow_n \sum_{l\ge1}\hat{M}_l(t_0) = \lim_{m\to\infty}\sum_{l=1}^m \hat{M}_l(t_0).$$

To this end, we only need to check the following facts:

$$\sum_{l=1}^m \hat{M}_l^n(t_0) \Rightarrow_n \sum_{l=1}^m \hat{M}_l(t_0) \Rightarrow_m \hat{N}(t_0), \tag{5.12}$$

$$\lim_{m\to\infty}\limsup_n \mathbb{P}\left(\Big|\sum_{l=m}^{\infty}\hat{M}_l^n(t_0)\Big|\geq\varepsilon\right)=0 \quad \text{for each } \varepsilon>0. \tag{5.13}$$

For every $m\in\mathbb{N}$ and $\theta\in\mathbb{R}$, by conditioning we have

$$\begin{aligned}
&\mathbb{E}\bigg[\exp\bigg(\sum_{l=1}^{m} i\theta\hat{M}_l^n(t_0)\bigg)\bigg]-\exp\bigg(-\frac{\theta^2}{2}\sum_{l=1}^{m}\hat{R}_l^n(t_0,t_0)\bigg)\\
&=\sum_{k=1}^{m}\mathbb{E}\bigg[\exp\bigg(i\theta\sum_{l=1}^{k-1}\hat{M}_l^n(t_0)-\frac{\theta^2}{2}\sum_{l=k+1}^{m}\hat{R}_l^n(t_0,t_0)\bigg)\bigg(e^{-i\theta\hat{M}_k^n(t_0)}-e^{-\frac{\theta^2}{2}\hat{R}_k^n(t_0,t_0)}\bigg)\bigg]\\
&=\sum_{k=1}^{m}\mathbb{E}\bigg[e^{i\theta\sum_{l=1}^{k-1}\hat{M}_l^n(t_0)}\Big(\mathbb{E}\big[e^{i\theta\hat{M}_k^n(t_0)}\big|\mathscr{F}_{k-1}^n\big]-e^{-\frac{\theta^2}{2}\hat{R}_k^n(t_0,t_0)}\Big)e^{-\frac{\theta^2}{2}\sum_{l=k+1}^{m}\hat{R}_l^n(t_0,t_0)}\bigg]
\end{aligned}$$

where we understand that $\sum_{l=m+1}^{m}=\sum_{l=1}^{0}=0$. This also gives

$$\Big|\mathbb{E}\Big[e^{i\theta\sum_{l=1}^{m}\hat{M}_l^n(t_0)}\Big]-e^{-\frac{\theta^2}{2}\sum_{l=1}^{m}\hat{R}_l^n(t_0,t_0)}\Big|\leq\sum_{k=1}^{m}\mathbb{E}\Big[\Big|\mathbb{E}\big[e^{i\theta\hat{M}_k^n(t_0)}\big|\mathscr{F}_{k-1}^n\big]-e^{-\frac{\theta^2}{2}\hat{R}_k^n(t_0,t_0)}\Big|\Big].$$

Notice that we have actually shown in the proof of Lemma 5.3 that, as $n\to\infty$,

$$\mathbb{E}\Big[\Big|\mathbb{E}\big[\exp\big(i\theta\hat{M}_l^n(t_0)\big)\big|\mathscr{F}_{l-1}^n\big]-\exp\big(-\frac{\theta^2}{2}\hat{R}_l^n(t_0,t_0)\big)\Big|\Big]\to 0.$$

The first condition (5.12) is thus proved by applying Assumption 2 to the deterministic function

$$\begin{aligned}
\sum_{l=1}^{m}\hat{R}_l^n(t_0,t_0)&=\int_0^T\sum_{l=1}^{m}\mathcal{G}^{l-1}\bigg(\int_{\mathbb{R}^d}\phi_{t_0}^2(\cdot,z)F_\cdot(dz)\bigg)(s)\bar{\lambda}_0^n(s)ds\\
&\to_n\int_0^T\sum_{l=1}^{m}\mathcal{G}^{l-1}\bigg(\int_{\mathbb{R}^d}\phi_{t_0}^2(\cdot,z)F_\cdot(dz)\bigg)(s)\bar{\lambda}_0(s)ds\\
&\to_m\int_0^T\mathcal{U}\bigg(\int_{\mathbb{R}^d}\phi_{t_0}^2(\cdot,z)F_\cdot(dz)\bigg)(s)\bar{\lambda}_0(s)ds=\hat{R}(t_0,t_0).
\end{aligned}$$

On the other hand, since $\mathbb{E}\big[\hat{M}_l^n(t_0)\big|\mathscr{F}_{k-1}^n\big]=0$ for $l\geq k\geq 1$, it is straightforward that

$$\begin{aligned}
\mathbb{E}\bigg[\bigg(\sum_{l=m}^{\infty}\hat{M}_l^n(t_0)\bigg)^2\bigg]&=\sum_{l=m}^{\infty}\mathbb{E}\Big[\big(\hat{M}_l^n(t_0)\big)^2\Big]=\sum_{l\geq m}\mathbb{E}\bigg[\bar{\lambda}_{l-1}^n\bigg(\int_{\mathbb{R}^d}\phi_{t_0}^2(\cdot,z)F_\cdot(dz)\bigg)^2\bigg]\\
&\leq\tilde{K}_2^4\bigg(1+\int_0^T\varphi(u)du\bigg)^4\bigg(\int_0^T\sum_{l\geq m}\varphi^{*(l-1)}(u)du\bigg)\bigg(\int_0^T\bar{\lambda}_0^n(u)du\bigg)
\end{aligned}$$

where $\tilde{K}_2$ is the constant defined in (5.5). This proves (5.13).

Similar to the end of the proof of Lemma 5.3, for the convergence of finite dimensional distributions of $\hat{N}^n$, it is sufficient to consider the linear span of $\{\hat{N}^n(t),t\in[0,T]\}$, and the proof above stays the same with $\mathbf{1}_{t_0}$ replaced by the associated linear combination of $\{\mathbf{1}_t(s),s\in[0,T]\}$. This completes the proof. $\square$

Lemma 5.7. *Under Assumption 1(i), 2 and 3, the finite–dimensional distribution of the processes $(\hat{X}^n,\hat{Y}^n)$ converges to that of $(\hat{X},\hat{Y})$.*

To prove the tightness of $(\hat{X}^n,\hat{Y}^n)$ and $\hat{N}^n$, we need the moments for $\hat{X}^n$ in (2.17).

Lemma 5.8. *Under Assumption 1(i), for $0 \le s < r < t \le T$, we have*

$$
\begin{aligned}
&\mathbb{E}\Big[\big(\hat{X}^n(t) - \hat{X}^n(r)\big)^2\big(\hat{X}^n(r) - \hat{X}^n(s)\big)^2\Big] \\
&= \sum_{k\ge 0}\mathbb{E}\Big[\bar{\lambda}^n_k\mathbf{1}_{s,r}\cdot\bar{\lambda}^n_k\mathbf{1}_{r,t} + \bar{\lambda}^n_k\mathcal{G}\mathcal{U}\mathbf{1}_{s,r}\cdot\bar{\lambda}^n_k\mathbf{1}_{r,t} + \bar{\lambda}^n_k\mathbf{1}_{s,r}\cdot\bar{\lambda}^n_k\mathcal{G}\mathcal{U}\mathbf{1}_{r,t}\Big] \\
&\quad + n^{-1}\sum_{k\ge 0}\mathbb{E}\Big[\bar{\lambda}^n_k\big(\mathbf{1}_{s,r}\cdot\mathcal{G}\mathcal{U}\mathbf{1}_{r,t}\big) + 2\bar{\lambda}^n_k\big(\mathbf{1}_{s,r}\cdot\mathcal{G}\mathcal{U}\big(\mathbf{1}_{s,r}\cdot\mathcal{G}\mathcal{U}\mathbf{1}_{r,t}\big)\big)\Big] = J^n_1 + J^n_2.
\end{aligned}
\tag{5.14}
$$

Proof. We start from showing that $\hat{X}^n(t) \in L^4(\mathbb{P})$ for all $t \in [0,T]$.

Let $\varphi_\delta(u) = e^{-\delta u}\varphi(u)$ for $\delta > 0$ such that $\int_0^T \varphi_\delta(u)du = p < 1$ as in the proof of Lemma 4.1. Then $\int_0^t \varphi^{*l}(u)du = \int_0^t e^{\delta u}\varphi^{*l}_\delta(u)du \le p^l e^{\delta t}$ for all $l \ge 1$ and $t \in [0,T]$. Therefore, applying (4.3), we obtain

$$
\begin{aligned}
\mathbb{E}\big[\lambda^n_l\mathbf{1}_t\big] &\le \int_0^t\bigg(\int_v^t \varphi^{*l}(u-v)du\bigg)\lambda^n_0(v)dv \le p^l\int_0^t e^{\delta(t-v)}\lambda^n_0(v)dv, \\
\mathrm{Var}\big(\lambda^n_l\mathbf{1}_t\big) &\le \sum_{k=1}^{l}\int_0^t\bigg(\int_w^t\bigg(\int_v^t\varphi^{*k}(u-v)du\bigg)^2\varphi^{*(l-k)}(v-w)dv\bigg)\lambda^n_0(w)dw \\
&\le \sum_{k=1}^{l}p^{l+k}\bigg(\int_0^t e^{2\delta(t-u)}\lambda^n_0(u)du\bigg) \le p^l\frac{p}{1-p}\bigg(\int_0^t e^{2\delta(t-u)}\lambda^n_0(u)du\bigg),
\end{aligned}
$$

where the two inequalities also hold for the trivial case $l = 0$. On the other hand, since $N^n_l(t)$ is Poisson distributed with parameter $\lambda^n_{l-1}\mathbf{1}_t$ under $\mathbb{P}(\cdot|\mathscr{F}^n_{l-1})$ for every $l \ge 1$,

$$
n^2\mathbb{E}\big[(\hat{X}^n_l(t))^4\big] = \mathbb{E}\Big[\lambda^n_{l-1}\mathbf{1}_t + 3\big(\lambda^n_{l-1}\mathbf{1}_t\big)^2\Big] = \mathbb{E}\big[\lambda^n_{l-1}\mathbf{1}_t\big] + 3\mathbb{E}^2\big[\lambda^n_{l-1}\mathbf{1}_t\big] + 3\mathrm{Var}\big(\lambda^n_{l-1}\mathbf{1}_t\big).
$$

Plugging the inequalities into the identity above gives

$$
\mathbb{E}\big[(\hat{X}^n_l(t))^4\big] \le \frac{p^{l-1}}{n^2}\Bigg(\int_0^t e^{\delta(t-v)}\lambda^n_0(v)dv + 3\bigg(\int_0^t e^{\delta(t-v)}\lambda^n_0(v)dv\bigg)^2 + \frac{3p}{1-p}\bigg(\int_0^t e^{2\delta(t-v)}\lambda^n_0(v)dv\bigg)\Bigg).
$$

Further applying Hölder's inequality, we have

$$
\mathbb{E}\big[(\hat{X}^n(t))^4\big] \le \mathbb{E}\Bigg[\sum_{l\ge 1}\big(q^{-l/4}\hat{X}^n_l(t)\big)^4\Bigg]\bigg(\sum_{l\ge 1}q^{l/3}\bigg)^3 < \infty
$$

for $q \in (p,1)$ by making use of the inequalities above, which proves $\hat{X}^n(t) \in L^4(\mathbb{P})$.

Now, we are ready to prove (5.14) from the identity

$$
\begin{aligned}
&\mathbb{E}\Big[\big(\hat{X}^n(t) - \hat{X}^n(r)\big)^2\big(\hat{X}^n(r) - \hat{X}^n(s)\big)^2\Big] \\
&= \lim_{m\to\infty}\mathbb{E}\Bigg[\bigg(\sum_{l=1}^m \hat{X}^n_l(t) - \hat{X}^n_l(r)\bigg)^2\bigg(\sum_{l=1}^m \hat{X}^n_l(r) - \hat{X}^n_l(s)\bigg)^2\Bigg] \\
&= \lim_{m\to\infty}\sum_{1\le i,i',j,j'\le m} K^n(i,i',j,j')
\end{aligned}
$$

where for $i, i', j, j' \in \mathbb{N}$,

$$
K^n(i,i',j,j') := \mathbb{E}\Big[(\hat{X}^n_i(t) - \hat{X}^n_i(r))(\hat{X}^n_{i'}(t) - \hat{X}^n_{i'}(r))(\hat{X}^n_j(r) - \hat{X}^n_j(s))(\hat{X}^n_{j'}(r) - \hat{X}^n_{j'}(s))\Big].
$$

Since N^n_l is a conditional Poisson process under $\mathbb{P}(\cdot|\mathscr{F}^n_{l-1})$, it can be checked that

$$
K^n(i,i',j,j')
$$

$$= \begin{cases} \mathbb{E}\big[(\hat{X}_i^n(t) - \hat{X}_i^n(r))^2(\hat{X}_j^n(r) - \hat{X}_j^n(s))^2\big] & \text{if } i = i' \text{ and } j = j \\ \mathbb{E}\big[(\hat{X}_i^n(t) - \hat{X}_i^n(r))^2(\hat{X}_j^n(r) - \hat{X}_j^n(s))(\hat{X}_{j'}^n(r) - \hat{X}_{j'}^n(s))\big] & \text{if } i = i' > j \vee j' \text{ and } j \neq j' \\ 0 & \text{otherwise} \end{cases}$$

For the non-trivial cases, the following results hold from Proposition 4.1 and Lemma 2.1.

(i) For $i = j \geq 1$, we have already shown in (5.11) that

$$K^n(i,i,i,i) = \mathbb{E}\big[\bar{\lambda}_{i-1}^n \mathbf{1}_{s,r} \cdot \bar{\lambda}_{i-1}^n \mathbf{1}_{r,t}\big].$$

(ii) For $i > j \geq 1$, by conditioning on $\mathscr{F}_j^n$,

$$K^n(i,i,j,j) = \mathbb{E}\Big[\bar{\lambda}_{i-1}^n \mathbf{1}_{r,t} \cdot \big(\hat{X}_j^n(r) - \hat{X}_j^n(s)\big)^2\Big] = \mathbb{E}\Big[\bar{\lambda}_j^n \mathcal{G}^{i-j-1}\mathbf{1}_{r,t} \cdot \big(\hat{X}_j^n(r) - \hat{X}_j^n(s)\big)^2\Big].$$

On the other hand, applying Proposition 4.1, we have for $\theta_1, \theta_2 \in \mathbb{R}$,

$$\begin{aligned} &\mathbb{E}\Big[\exp\Big(i\theta_1\big(N_l^n(r) - N_l^n(s) - \lambda_{l-1}^n\mathbf{1}_{s,r}\big) + i\theta_2\lambda_l^n f\Big)\Big|\mathscr{F}_{l-1}^n\Big] \\ &= \mathbb{E}\left[\exp\left(\sum_{k=1}^{N_l^n(T)}\big(i\theta_1\mathbf{1}_{s,r}(\tau_{lk}^n) + i\theta_2\mathcal{H}f(\tau_{lk}^n, Z_{lk}^n)\big) - i\theta_1\lambda_{l-1}^n\mathbf{1}_{s,r}\right)\middle|\mathscr{F}_{l-1}^n\right] \\ &= \exp\left(\int_0^T\int_{\mathbb{R}^d}\Big(e^{i\theta_1\mathbf{1}_{s,r}(u) + i\theta_2\mathcal{H}f(u,z)} - 1 - i\theta_1\mathbf{1}_{s,r}(u)\Big)F_u(dz)\lambda_{l-1}^n(u)du\right). \end{aligned}$$

Taking derivatives $\frac{\partial^2}{\partial\theta_1\partial\theta_2}, \frac{\partial^3}{\partial^2\theta_1\partial\theta_2}$ and setting $\theta_1 = \theta_2 = 0$, we have

$$\begin{aligned} \mathbb{E}\big[(X_l^n(r) - X_l^n(s)) \cdot \lambda_l^n f\big|\mathscr{F}_{l-1}^n\big] &= \int_0^T \lambda_{l-1}^n(u)\mathbf{1}_{r,s}(u)\mathcal{G}f(u)du = \lambda_{l-1}^n\big(\mathbf{1}_{r,s}\cdot\mathcal{G}f\big), \\ \mathbb{E}\big[(X_l^n(r) - X_l^n(s))^2 \cdot \lambda_l^n f\big|\mathscr{F}_{l-1}^n\big] &= \lambda_{l-1}^n\big(\mathbf{1}_{r,s}\cdot\mathcal{G}f\big) + \big(\lambda_{l-1}^n\mathbf{1}_{r,s}\big)\cdot\big(\lambda_{l-1}^n\mathcal{G}f\big), \end{aligned} \tag{5.15}$$

Therefore, taking $f(u) = \mathcal{G}^{i-j-1}\mathbf{1}_{r,t}(u)$ in the second identity in (5.15) gives

$$K^n(i,i,j,j) = \mathbb{E}\Big[\bar{\lambda}_{j-1}^n\mathcal{G}^{i-j}\mathbf{1}_{r,t}\cdot\bar{\lambda}_{j-1}^n\mathbf{1}_{s,r} + \frac{1}{n}\bar{\lambda}_{j-1}^n\big(\mathbf{1}_{s,r}\cdot\mathcal{G}^{i-j}\mathbf{1}_{r,t}\big)\Big].$$

(iii) For $j > i \geq 1$, by conditioning on $\mathscr{F}_i^n$, we have

$$K^n(i,i,j,j) = \mathbb{E}\big[(\hat{X}_i^n(t) - \hat{X}_i^n(r))^2\cdot\bar{\lambda}_i^n\mathcal{G}^{j-i-1}\mathbf{1}_{s,r}\big] = \mathbb{E}\big[\bar{\lambda}_{i-1}^n\mathbf{1}_{r,t}\cdot\bar{\lambda}_{i-1}^n\mathcal{G}^{j-i}\mathbf{1}_{s,r}\big],$$

where the conditionally independence between $(\hat{X}_i^n(t) - \hat{X}_i^n(r))$ and $\bar{\lambda}_i^n\mathcal{G}^{j-i-1}\mathbf{1}_{s,r} \in \mathscr{F}_i^n(r)$ is used in the second identity.

(iv) For $i > j > j'$, by conditioning on $\mathscr{F}_j^n$ and $\mathscr{F}_{j'}^n$ successively, we have

$$\begin{aligned} &\mathbb{E}\big[(\hat{X}_i^n(t) - \hat{X}_i^n(r))^2(\hat{X}_j^n(r) - \hat{X}_j^n(s))(\hat{X}_{j'}^n(r) - \hat{X}_{j'}^n(s))\big] \\ &= \mathbb{E}\big[\bar{\lambda}_j^n\mathcal{G}^{i-j-1}\mathbf{1}_{r,t}\cdot(\hat{X}_j^n(r) - \hat{X}_j^n(s))(\hat{X}_{j'}^n(r) - \hat{X}_{j'}^n(s))\big] \\ &= n^{-1/2}\mathbb{E}\big[\bar{\lambda}_{j-1}^n\big(\mathbf{1}_{s,r}\cdot\mathcal{G}^{i-j}\mathbf{1}_{r,t}\big)\cdot(\hat{X}_{j'}^n(r) - \hat{X}_{j'}^n(s))\big] \\ &= n^{-1}\mathbb{E}\big[\bar{\lambda}_{j'-1}^n\big(\mathbf{1}_{s,r}\cdot\mathcal{G}^{j-j'}\big(\mathbf{1}_{s,r}\cdot\mathcal{G}^{i-j}\mathbf{1}_{r,t}\big)\big],\end{aligned}$$

where we make use of the first identity in (5.15) in the last two identities above.

Since $K^n \geq 0$ for all choice of i, i', j, j', the formula in (5.14) is thus proved. □

Now, we are ready to prove the tightness of $\hat{X}^n$ and $\hat{Y}^n$, respectively, following the same idea used in the proof of Lemma 5.5, and the tightness of $\hat{N}^n = \hat{X}^n + \hat{Y}^n$ follows from the fact $\hat{Y}^n \in \mathbb{C}$. Notice the correlated terms appeared in (5.14), comparing to (5.11).

Lemma 5.9. *Under Assumption 1, there is $K_\infty > 0$ for every $\varepsilon, \delta > 0$,*

$$\mathbb{P}\big(\omega''(\hat{X}^n, \delta) \geq \varepsilon\big) \leq \frac{K_\infty}{\varepsilon^4}\bigg(\sup_{t\in[0,T]}\int_{t-2\delta}^{t}\big(1+\bar{\lambda}_0^n(u)\big)du\bigg)\bigg(\int_0^T\big(1+\bar{\lambda}_0^n(u)\big)du\bigg), \tag{5.16}$$

$$\mathbb{P}\big(\omega(\hat{Y}^n, \delta) \geq \varepsilon\big) \leq \frac{K_\infty \delta}{\varepsilon^2}\Big(\int_0^T \bar{\lambda}_0^n(u)du\Big). \tag{5.17}$$

Proof. Let J_1^n and J_2^n be defined in (5.14).

Since $\mathbf{1}_{s,r} + \mathbf{1}_{r,t} = \mathbf{1}_{s,t}$, we have

$$\begin{aligned} J_1^n &\leq \sum_{k\geq 0}\mathbb{E}\Big[\bar{\lambda}_k^n\mathbf{1}_{s,t}\cdot\bar{\lambda}_k^n\mathbf{1}_{s,t} + \bar{\lambda}_k^n\mathcal{G}\mathcal{U}\mathbf{1}_{s,r}\cdot\bar{\lambda}_k^n\mathbf{1}_{s,t} + \bar{\lambda}_k^n\mathbf{1}_{s,t}\cdot\bar{\lambda}_k^n\mathcal{G}\mathcal{U}\mathbf{1}_{r,t}\Big] \\ &= \sum_{k\geq 0}\mathbb{E}\big[\bar{\lambda}_k^n\mathbf{1}_{s,t}\cdot\bar{\lambda}_k^n\mathcal{U}\mathbf{1}_{s,t}\big] = \sum_{k\geq 0}\Big(\mathbb{E}\big[\bar{\lambda}_k^n\mathbf{1}_{s,t}\big]\mathbb{E}\big[\bar{\lambda}_k^n\mathcal{U}\mathbf{1}_{s,t}\big] + \mathrm{Cov}\big(\bar{\lambda}_k^n\mathbf{1}_{s,t}\,,\,\bar{\lambda}_k^n\mathcal{U}\mathbf{1}_{s,t}\big)\Big). \end{aligned}$$

Applying Lemma 4.1(i) and the fact that $\mathcal{G}^k\mathcal{U}\mathbf{1}_{s,t}(u) \leq \mathcal{U}\mathbf{1}_{s,t}(u)$ for $k \geq 0$, we have

$$\begin{aligned} \sum_{k\geq 0}\mathbb{E}\big[\bar{\lambda}_k^n\mathbf{1}_{s,t}\big]\mathbb{E}\big[\bar{\lambda}_k^n\mathcal{U}\mathbf{1}_{s,t}\big] &= \sum_{k\geq 0}\big(\bar{\lambda}_0^n\mathcal{G}^k\mathbf{1}_{s,t}\cdot\bar{\lambda}_0^n\mathcal{G}^k\mathcal{U}\mathbf{1}_{s,t}\big) \\ &\leq \big(\bar{\lambda}_0^n\mathcal{U}\mathbf{1}_{s,t}\big)^2 \leq \bigg(\int_0^T\mathbf{1}_{s,t}(u)\big(\bar{\lambda}_0^n(u)+\Phi*\bar{\lambda}_0^n(u)\big)du\bigg)^2. \end{aligned} \tag{5.18}$$

Applying (4.3) and a similar fact that $\varphi^{*k}(u) + \varphi^{*k}*\Phi(u) \leq \Phi(u)$ for $k \geq 1$, we have

$$\begin{aligned} &\sum_{k\geq 0}\mathrm{Cov}\big(\bar{\lambda}_k^n\mathbf{1}_{s,t}\,,\,\bar{\lambda}_k^n\mathcal{U}\mathbf{1}_{s,t}\big) = \sum_{k\geq 1}\mathrm{Cov}\big(\bar{\lambda}_k^n\mathbf{1}_{s,t}\,,\,\bar{\lambda}_k^n\mathcal{U}\mathbf{1}_{s,t}\big) \\ &\leq \frac{1}{n}\sum_{k\geq 1}\int_0^T\bigg(\int_0^T\mathbf{1}_{s,t}(u)\varphi^{*k}(u-v)du\bigg)\bigg(\int_0^T\mathbf{1}_{s,t}(u)\Phi(u-v)du\bigg)\big(\bar{\lambda}_0^n(v)+\Phi*\bar{\lambda}_0^n(v)\big)dv \\ &\leq \frac{(t-s)}{n}\int_0^T\bigg(\int_0^T\mathbf{1}_{s,t}(u)\Phi^2(u-v)du\bigg)\big(\bar{\lambda}_0^n(v)+\Phi*\bar{\lambda}_0^n(v)\big)dv, \end{aligned} \tag{5.19}$$

where the Cauchy-Schwarz inequality is applied in the last inequality under Assumption 1(ii).

On the other hand, applying Cauchy's inequality,

$$\int_0^u\Phi(v)dv \leq u^{1/2}\Big(\int_0^u\Phi^2(v)dv\Big)^{1/2} \leq u^{1/2}\tilde{K}_3 \leq \frac{1}{2} \quad \text{for all } u \leq \delta_0,$$

where $\delta_0 > 0$ is a constant such that $\delta_0^{1/2}\tilde{K}_3 \leq \frac{1}{2}$ and $\tilde{K}_3 = (\int_0^{\delta_0}\Phi^2(v)dv)^{1/2}$.

Applying Lemma 4.1(i), we obtain

$$\mathbf{1}_{s,t}(u)\mathcal{G}\mathcal{U}\mathbf{1}_{s,t}(u) \leq \mathbf{1}_{s,t}(u)\int_0^T\mathbf{1}_{s,t}(v)\Phi(v-u)dv \leq \mathbf{1}_{s,t}(u)\bigg(\int_0^{t-s}\Phi(v)dv\bigg)$$

which shows that for $0 < t-s \leq \delta_0$,

$$\begin{aligned} J_2^n &\leq \frac{1}{n}\bigg(\int_0^{t-s}\Phi(v)dv\bigg)\bigg(1+2\bigg(\int_0^{t-s}\Phi(v)dv\bigg)\bigg)\sum_{k\geq 0}\mathbb{E}\big[\bar{\lambda}_k^n\mathbf{1}_{s,t}\big] \\ &\leq \frac{2\tilde{K}_3}{n}(t-s)^{1/2}\bigg(\int_0^T\mathbf{1}_{s,t}(u)\big(\bar{\lambda}_0^n(u)+\Phi*\bar{\lambda}_0^n(u)\big)du\bigg). \end{aligned} \tag{5.20}$$

Thus, under Assumption 1, for all $0 \leq s < r < t \leq T$ with small $(t-s)$, we have

$$\mathbb{E}\big[\big(\hat{X}^n(t)-\hat{X}^n(r)\big)^2\big(\hat{X}^n(r)-\hat{X}^n(s)\big)^2\big]$$

$$\leq \Big(\int_0^T \mathbf{1}_{s,t}(u)\big(\bar\lambda_0^n(u) + \Phi * \bar\lambda_0^n(u)\big)du\Big)^2$$
$$+ 2\tilde K_3 (t-s)^{1/2}\Big(\int_0^T \mathbf{1}_{s,t}(u)\big(\bar\lambda_0^n(u) + \Phi * \bar\lambda_0^n(u)\big)du\Big)$$
$$+ (t-s)\int_0^T \mathbf{1}_{s,t}(u)\big(\Phi^2 * \bar\lambda_0^n(u) + \Phi^2 * \Phi * \bar\lambda_0^n(u)\big)du. \tag{5.21}$$

Now, for every small interval $[a,b]$, we take $D_k = \{t_{kj}, j = 0, 1, \cdots, 2^k\}$ such that

$$\Big(\int_a^t \big(1 + \bar\lambda_0^n(u) + \Phi * \bar\lambda_0^n(u)\big)du\Big)\Big|_{t=t_{kj}} = \frac{j}{2^k}\int_a^b \big(1 + \bar\lambda_0^n(u) + \Phi * \bar\lambda_0^n(u)\big)du.$$

Then $D_k \subset D_{k+1}$ and $D = \cup_{k\geq 1} D_k$ is a dense subset of $[a,b]$. Taking $(t,r,s) = (t_{kj}, t_{k(j-1)}, t_{k(j-2)})$ in (5.21) gives, for every $\varepsilon > 0$,

$$\varepsilon^4 \mathbb{P}\big(\big|\hat X^n(t) - \hat X^n(r)\big| \wedge \big|\hat X^n(r) - \hat X^n(s)\big| \geq \varepsilon\big)\Big|_{(t,r,s)=(t_{kj},t_{k(j-1)},t_{k(j-2)})}$$
$$\leq \left(\frac{\int_a^b (1 + \bar\lambda_0^n(u) + \Phi * \bar\lambda_0^n(u))du}{2^{k-1}}\right)^2 + 2\tilde K_3 \left(\frac{\int_a^b (1 + \bar\lambda_0^n(u) + \Phi * \bar\lambda_0^n(u))du}{2^{k-1}}\right)^{3/2}$$
$$+ \frac{\int_a^b \big(1 + \bar\lambda_0^n(u) + \Phi * \bar\lambda_0^n(u)\big)du}{2^{k-1}} \int_0^T \mathbf{1}_{s,t}(u)\Big(\Phi^2 * \bar\lambda_0^n(u) + \Phi^2 * \Phi * \bar\lambda_0^n(u)\Big)du.$$

Summing over all choices of $j = 2, 3, \cdots, 2^k$, we have

$$\varepsilon^4 \mathbb{P}\left(\sup_{(s,r,t)=(t_{k(j-2)},t_{k(j-1)},t_{kj})} \big|\hat X^n(t) - \hat X^n(r)\big| \wedge \big|\hat X^n(r) - \hat X^n(s)\big| \geq \varepsilon\right)$$
$$\leq 2\frac{\big(\int_a^b \big(1 + \bar\lambda_0^n(u) + \Phi * \bar\lambda_0^n(u)\big)du\big)^2}{2^{k-1}} + 4\tilde K_3 \frac{\big(\int_a^b (1 + \bar\lambda_0^n(u) + \Phi * \bar\lambda_0^n(u))du\big)^{3/2}}{2^{\frac{k-1}{2}}} \tag{5.22}$$
$$+ 2\frac{\int_a^b \big(1 + \bar\lambda_0^n(u) + \Phi * \bar\lambda_0^n(u)\big)du}{2^{k-1}} \times \int_0^T \mathbf{1}_{a,b}(u)\Big(\Phi^2 * \bar\lambda_0^n(u) + \Phi^2 * \Phi * \bar\lambda_0^n(u)\Big)du.$$

Then similar to the derivation of (5.10), by taking $\theta \in (2^{-1/8}, 1)$, we have for all small $[a,b] \subset [0,T]$,

$$\mathbb{P}\left(\sup_{a\leq s<r<t\leq b} \big|\hat X^n(t) - \hat X^n(r)\big| \wedge \big|\hat X^n(r) - \hat X^n(s)\big| \geq \varepsilon\right) \leq \frac{K_\infty}{2\varepsilon^4}\left(\int_a^b \big(1 + \bar\lambda^n(u)\big)du\right)^2$$

for some K_∞ depending on Φ but independent of λ_0^n and ε. This further gives (5.16) for $\hat X^n$.

For the increment of $\hat Y_l^n$, since $\mathbb{E}\big[\hat Y_l^n(t)\big|\mathscr{F}_{k-1}^n\big] = 0$ for $l \geq k \geq 1$, it follows from (5.8) that

$$\mathbb{E}\big[(\hat Y^n(t) - \hat Y^n(s))^2\big] = \sum_{l\geq 1} \mathbb{E}\big[(\hat Y_l^n(t) - \hat Y_l^n(s))^2\big]$$
$$\leq (t-s)\int_0^T \mathbf{1}_{s,t}(u)\big(\Phi^2 * \bar\lambda_0^n(u) + \Phi^2 * \Phi * \bar\lambda_0^n(u)\big)du,$$

which further gives (5.17) for $\hat Y^n$ by the same reasoning for $\hat Y_l^n$. This finishes the proof. $\square$

5.3. **Proof of Theorem 2.1.** Assume Assumption 1 and 2 holds. Recalling the representation in (2.18) and the expectation of N^n in (2.19), we have

$$\big(\bar N^n(t) - \bar\lambda_0 \mathcal{U}\mathbf{1}_t\big) = \frac{1}{\sqrt n}\hat X^n(t) + \frac{1}{\sqrt n}\hat Y^n(t) + \big(\bar\lambda_0^n \mathcal{U}\mathbf{1}_t - \bar\lambda_0\mathcal{U}\mathbf{1}_t\big).$$

On the other hand, from the inequality (4.4), we already have for every $\varepsilon > 0$

$$\mathbf{P}\Big(\sup_{t\in[0,T]}\Big|\frac{1}{\sqrt{n}}\hat{X}^n(t)\Big|\geq\varepsilon\Big)=O\Big(\frac{1}{n^2\varepsilon^2}\int_0^T\lambda_0^n(u)du\Big),$$

under Assumption 1(i). Since $\hat{Y}^n(0)=0$ by assumption and $\hat{Y}^n\in\mathbb{C}$, we have

$$\lim_{n\to\infty}\mathbf{P}\Big(\sup_{t\in[0,T]}\big|\hat{Y}^n(t)\big|<\infty\Big)=1.$$

from (5.17) under Assumption 1. And under Assumption 2,

$$\bar{\lambda}_0^n\mathcal{U}\mathbf{1}_t-\bar{\lambda}_0\mathcal{U}\mathbf{1}_t=\int_0^T\mathcal{U}\mathbf{1}_t(u)\big(\bar{\lambda}_0^n(u)-\bar{\lambda}_0(u)\big)du\Rightarrow 0$$

under uniform topology on $[0,T]$, as $n\to\infty$. The FWLLN for $\bar{N}^n$ in Theorem 2.1 is thus proved.

6. Proofs for the special models

Since the Gaussian space is closed under weak convergence, the propositions on the Gaussian limits of the special models in Section 3 are proved by following the same idea of checking the limits of the covariance functions of the increments of $\big(\hat{X},\hat{Y}\big)$ under suitable time scales. Firstly, we prove that under the stability condition, the large intensity limit of a stationary Hawkes process coincides with the stationary limit of the large intensity Hawkes process shown in [21]. Motivated by that, we also prove the stationary limit of Gaussian process for the indicator-type non-decomposable Hawkes process in Section 3.2.

6.1. **Proof of Proposition 3.1, 3.2 and Corollary 3.2.**

Proof of Proposition 3.1. By assumption, ψ in this model is defined from the renewal equation (3.7). Applying Theorem 2.2 and Remark 2.6, it can be found that for $t,s>0$

$$\begin{aligned}&\mathbb{E}\big[\big(\hat{X}(t+h)-\hat{X}(h)\big)\big(\hat{X}(s+h)-\hat{X}(h)\big)\big]\\ =&\int_0^T\mathcal{U}\big(\mathbf{1}_{h,t+h}(\cdot)\mathbf{1}_{h,s+h}(\cdot)\big)(u)\bar{\lambda}_0(u)du=\int_h^{(t\wedge s)+h}\big(\bar{\lambda}_0(u)+\psi*\bar{\lambda}_0(u)\big)du\\ =&\bar{\lambda}_0\int_0^{t\wedge s}\big(1+\psi*1(u+h)\big)du\to_h=\frac{\bar{\lambda}_0}{1-||\tilde{H}||_1}(t\wedge s)\end{aligned}$$

where we have used $\mathbf{1}_{s,t}(u)=\mathbf{1}_t(u)-\mathbf{1}_s(u)=\mathbf{1}(u\in(s,t])$, and the fact that $||f*g||_1=||f||_1\cdot||g||_1$ in the last identity, Similarly, passing $h\to\infty$, we obtain

$$\begin{aligned}&\mathbb{E}\big[\big(\hat{X}(t+h)-\hat{X}(h)\big)\big(\hat{Y}(s+h)-\hat{Y}(h)\big)\big]\\ =&\int_0^T\mathcal{U}\big(\mathbf{1}_{h,t+h}(\cdot)\mathcal{G}\mathcal{U}\mathbf{1}_{h,s+h}(\cdot)\big)(u)\bar{\lambda}_0(u)du\\ =&\int_0^T\mathbf{1}_{h,t+h}(u)\int_0^T\mathbf{1}_{h,s+h}(v)\psi(v-u)dv\big(\bar{\lambda}_0(u)+\psi*\bar{\lambda}_0(u)\big)du\\ =&\bar{\lambda}_0\int_0^t\big(1+\psi*1(u+h)\big)du\int_0^s\psi(v-u)dv\to_h\frac{\bar{\lambda}_0}{1-||\tilde{H}||_1}\int_0^t dv\int_0^s du\psi(u-v).\end{aligned}$$

Since $F_s(dz)=\delta_1(dz)$, we have $\mathcal{H}f(s,z)=\mathcal{G}\mathcal{U}f(s)=\int_0^T f(u)\psi(u-s)du$, thus

$$\begin{aligned}&\mathbb{E}\big[\big(\hat{Y}(t+h)-\hat{Y}(h)\big)\big(\hat{Y}(s+h)-\hat{Y}(h)\big)\big]\\ =&\int_0^T\mathcal{U}\Big(\int_{\mathbb{R}}\mathcal{H}\mathcal{U}\mathbf{1}_{h,t+h}(\cdot,z)\mathcal{H}\mathcal{U}\mathbf{1}_{h,s+h}(\cdot,z)F_\cdot(dz)\Big)(u)\bar{\lambda}_0(u)du\end{aligned}$$

$$
\begin{aligned}
&= \int_h^{t+h} dv \int_h^{s+h} dv' \int_0^{v\wedge v'} \psi(v-u)\psi(v'-u)\big(\bar{\lambda}_0(u) + \psi * \bar{\lambda}_0(u)\big)du \\
&= \bar{\lambda}_0 \int_0^t \psi(v-u)dv \int_0^s \psi(v'-u)dv' \int_{-h}^{v\wedge v'} \big(1 + \psi * 1(u+h)\big)du \\
&\to_h \frac{\bar{\lambda}_0}{1-||\tilde{H}||_1} \int_0^t dv \int_0^s dv' \Big(\int_0^{v\wedge v'} \psi(v-u)\psi(v'-u)du\Big) \\
&\qquad + \frac{\bar{\lambda}_0}{1-||\tilde{H}||_1} \int_0^t dv \int_0^s dv' \Big(\int_0^{\infty} \psi(v+u)\psi(v'+u)du\Big).
\end{aligned}
$$

Given the covariance functions above, we can actually have an equivalent Itô integral formula for the Gaussian process $\big(\hat{X}^\circ, \hat{Y}^\circ\big)$. Let W be a two-sided Brownian motions with $W(0) = 0$ and variance $\frac{\bar{\lambda}_0}{1-||\tilde{H}||_1}$. Letting $h \to \infty$, we have for $t \geq 0$

$$
\begin{aligned}
\big(\hat{X}(t+h) - \hat{X}(h)\big) &\Rightarrow_h \hat{X}^\circ(t) \stackrel{d}{=} W(t), \\
\big(\hat{Y}(t+h) - \hat{Y}(h)\big) \Rightarrow_h \hat{Y}^\circ(t) &\stackrel{d}{=} \int_0^t \int_0^t \psi(u-v)du W(dv) + \int_0^\infty \int_0^t \psi(u+v)du W(-dv) \\
&= \int_{-\infty}^{\infty} \Big(\int_0^t \psi(u-v)du\Big) W(dv)
\end{aligned}
$$

Here the finiteness of $\displaystyle\int_0^\infty \Big(\int_0^t \psi(u+v)du\Big) W(-dv)$ follows from the completeness of Gaussian distribution in $L^2(\mathbb{P})$ space that

$$
\int_0^\infty \Big(\int_0^t \psi(u+v)du\Big)^2 dv \leq ||\psi||_1 \int_0^\infty \int_0^t \psi(u+v)du dv \leq t \times ||\psi||_1^2 < \infty.
$$

This gives the results in (3.12) and (3.8), and it is straightforward to derive the associated covariance function (3.9). Actually, one can also have Itô expressions for $\big(\hat{X}_l, \hat{Y}_l\big)$.

We next prove the equivalent between (3.9) and (3.10). We first show that (3.9) is equal to $K(t)$ for the case $t = s$. For the integrals on the right hand side, we have

$$
\int_{\mathbb{R}} \Big(\mathbf{1}_t(v) + \int_0^t \psi(u-v)du\Big)^2 dv = t + 2\int_0^t dv \int_0^t \psi(u-v)du + \int_{\mathbb{R}} \Big(\int_0^t \psi(u-v)du\Big)^2 dv.
$$

Applying the change of variables, we have

$$
\begin{aligned}
\int_{\mathbb{R}} \Big(\int_0^t \psi(u-v)du\Big)^2 dv &= 2\int_0^t du \int_0^u du' \int_{\mathbb{R}} \psi(u-v)\psi(u'-v)dv \\
&= 2\int_0^t du \int_0^u dv \int_0^\infty \psi(u-v+w)\psi(w)dw.
\end{aligned}
$$

Plugging the identities above into (3.9) gives

$$
\begin{aligned}
\mathbb{E}\big[\hat{N}^\circ(t)\hat{N}^\circ(t)\big] &= \frac{\bar{\lambda}_0}{1-||\tilde{H}||_1}\Big(t + 2\int_0^t du \int_0^u \psi(v)dv + 2\int_0^t du \int_0^u dv \int_0^\infty \psi(v+w)\psi(w)dw\Big) \\
&= \frac{\bar{\lambda}_0}{1-||\tilde{H}||_1}\Big(t + 2\int_0^t du \int_0^u \Big(\psi(v) + \int_0^\infty \psi(v+w)\psi(w)dw\Big)dv\Big). \qquad (6.1)
\end{aligned}
$$

Now, we take for $t > 0$,

$$
\kappa(t) := \psi(t) + \int_0^\infty \psi(t+u)\psi(u)du. \qquad (6.2)
$$

It can be checked directly that

$$\begin{aligned}
&\int_0^\infty \tilde{H}(t+u)\Big(\int_0^\infty \psi(u+v)\psi(v)dv\Big)du + \int_0^t \tilde{H}(t-u)\Big(\int_0^\infty \psi(u+v)\psi(v)dv\Big)du \\
= &\int_0^\infty \psi(u)dv\Big(\int_0^u \tilde{H}(t+u-v)\psi(v)dv\Big) + \int_0^\infty \psi(v)dv\Big(\int_0^t \tilde{H}(t-u)\psi(u+v)du\Big) \\
= &\int_0^\infty \psi(u)\big(\tilde{H} * \psi(t+u)\big)du = \int_0^\infty \psi(u)\psi(t+u)du - \int_0^\infty \psi(u)\tilde{H}(t+u)du,
\end{aligned}$$

where we make use of $\psi = \tilde{H} + \tilde{H} * \psi$ in the last identity. Similarly,

$$\int_0^\infty \tilde{H}(t+u)\psi(u)du + \int_0^t \tilde{H}(t-u)\psi(u)du = \int_0^\infty \tilde{H}(t+u)\psi(u)du + \big(\psi(t) - \tilde{H}(t)\big).$$

Therefore, κ satisfies the equation

$$\kappa(t) = \tilde{H}(t) + \int_0^\infty \tilde{H}(t+u)\kappa(u)du + \int_0^t \tilde{H}(t-u)\kappa(u)du,$$

and $\hat{\phi}(u) = \dfrac{\bar{\lambda}_0}{1 - ||\tilde{H}||_1}\kappa(u)$ in (3.11), and (6.1) equals to $K(t)$. The case $t = s$ is proved.

For the case $t > s$ in (3.9), it is sufficient to check that

$$\begin{aligned}
\int_s^t du \int_0^s \kappa(u-v)dvdu &= \int_{\mathbb{R}} \Big(\mathbf{1}_{s,t}(v) + \int_s^t \psi(u-v)du\Big)\Big(\mathbf{1}_s(v) + \int_0^s \psi(u-v)du\Big)dv \\
&= \int_s^t du \int_0^s dv\psi(u-v) + \int_0^\infty \Big(\int_s^t \psi(u+v)du\Big)\Big(\int_0^s \psi(u+v)du\Big)dv \\
&\quad + \int_0^\infty \Big(\int_s^t \psi(u-v)du\Big)\Big(\int_0^s \psi(u-v)du\Big)dv. \qquad (6.3)
\end{aligned}$$

By the change of variables,

$$\begin{aligned}
&\int_0^\infty \Big(\int_s^t \psi(u+v)du\Big)\Big(\int_0^s \psi(u+v)du\Big)dv \\
= &\int_s^t du \int_0^s dv \int_0^\infty \psi(u+w)\psi(v+w)dw = \int_s^t du \int_0^s dv \int_v^\infty \psi(u-v+w)\psi(w)dw
\end{aligned}$$

and

$$\begin{aligned}
&\int_0^\infty \Big(\int_s^t \psi(u-v)du\Big)\Big(\int_v^s \psi(u-v)du\Big)dv \\
= &\int_s^t du \int_0^s dv \int_0^\infty \psi(u-w)\psi(v-w)dw = \int_s^t du \int_0^s dv \int_0^v \psi(u-v+w)\psi(w)dw
\end{aligned}$$

where we make use of the fact that $\psi(v-w) = 0$ for $v < w$ and $u > v$ in the last identity. Recalling κ in (6.2), the identity (6.3) is thus proved. This completes the proof. □

Proof of Corollary 3.2. Plugging (3.13) and (3.14) into Remark 2.6, we have

$$\begin{aligned}
\mathrm{Cov}\big(\hat{X}(t), \hat{X}(s)\big) &= \bar{\lambda}_0 \int_0^{t\wedge s} e^{\eta(u)}du, \\
\mathrm{Cov}\big(\hat{X}(t), \hat{Y}(s)\big) &= \bar{\lambda}_0 \int_0^t e^{\eta(u)}du \int_0^s dv\mathbf{1}(v > u)\mathbf{m}_{1,u}e^{\eta(v)-\eta(u)} = \bar{\lambda}_0 \int_0^s e^{\eta(v)}\eta(t \wedge v)dv,
\end{aligned} \qquad (6.4)$$

noticing that $\psi(v,u)=0$ for $u>v$. By the definition of $\mathcal{H}\mathcal{U}f$ in (2.7), for $t>u>0$,

$$\begin{aligned}\mathcal{H}\mathcal{U}\mathbf{1}_t(u,z) &= \int_0^t H(v-u,z)dv + \int_0^t \Big(\int_0^t \psi(w,v)dw\Big) H(v-u,z)dv \\ &= z\Big(\int_u^t dw + \int_u^t dw \int_u^w \psi(w,v)dv\Big) = z\int_u^t e^{\eta(v)-\eta(u)}dv \end{aligned} \tag{6.5}$$

by making use of (3.14). Therefore, the covariance functions of $\hat{Y}$ in Remark 2.6 reads

$$\begin{aligned}\mathrm{Cov}\big(\hat{Y}(t),\hat{Y}(s)\big) &= \bar{\lambda}_0\int_0^{t\wedge s} \Big(\int_u^t e^{\eta(v)-\eta(u)}dv\Big)\Big(\int_u^s e^{\eta(v')-\eta(u)}dv'\Big)\mathbf{m}_{2,u}e^{\eta(u)}du \\ &= \bar{\lambda}_0\int_0^t e^{\eta(v)}dv\int_0^s e^{\eta(v')}dv'\int_0^{v\wedge v'}\mathbf{m}_{2,u}e^{-\eta(u)}du. \end{aligned} \tag{6.6}$$

The covariance for $\hat{N}$ follows from the identity $\hat{N}=\hat{X}+\hat{Y}$. □

Proof of Proposition 3.2. If $\int_0^\infty \mathbf{m}_{1,u}du<\infty$ and $\int_0^\infty \mathbf{m}_{2,v}dv<\infty$, for every $t,s,h>0$, we have from (6.4) that as $h\to\infty$,

$$\mathrm{Cov}\Big(\big(\hat{X}(t+h)-\hat{X}(h)\big),\big(\hat{X}(s+h)-\hat{X}(h)\big)\Big) = \bar{\lambda}_0\int_h^{(t\wedge s)+h} e^{\eta(u)}du \to \bar{\lambda}_0 e^{\eta(\infty)}(t\wedge s)$$

and

$$\begin{aligned}&\mathrm{Cov}\Big(\big(\hat{X}(t+h)-\hat{X}(h)\big),\big(\hat{Y}(s+h)-\hat{Y}(h)\big)\Big)\\ &= \bar{\lambda}_0\int_h^{s+h} e^{\eta(v)}\big(\eta((t+h)\wedge v)-\eta(v)\big)dv \le \bar{\lambda}_0 e^{\eta(\infty)}\big(\eta(\infty)-\eta(h)\big)s\to 0.\end{aligned}$$

For the covariance of $\hat{Y}$, we have from (6.6) that as $h\to\infty$,

$$\begin{aligned}&\mathrm{Cov}\Big(\big(\hat{Y}(t+h)-\hat{Y}(h)\big),\big(\hat{Y}(s+h)-\hat{Y}(h)\big)\Big)\\ &= \bar{\lambda}_0\int_h^{t+h} e^{\eta(v)}dv\int_h^{s+h} e^{\eta(v')}dv'\int_0^{v\wedge v'}\mathbf{m}_{2,u}e^{-\eta(u)}du \to \Big(\bar{\lambda}_0 e^{2\eta(\infty)}\int_0^\infty \mathbf{m}_{2,u}e^{-\eta(u)}du\Big)\cdot t\cdot s.\end{aligned}$$

The limit above gives the limit in the first case.

On the other hand, suppose that $\displaystyle\int_0^\infty \mathbf{m}_{1,v}dv=\infty$. Let k be the function defined in Proposition 3.2, we have

$$k(t)\wedge k(s) = k(t\wedge s) \quad \text{and} \quad e^{\eta(k(t))}k'(t)=1 \quad t>0.$$

Similar to the calculations above, we have by change of variable

$$\begin{aligned}&\mathrm{Cov}\big(\hat{X}(k(t+h))-\hat{X}(k(h)),\hat{X}(k(s+h))-\hat{X}(h)\big)\\ &= \bar{\lambda}_0\int_{k(h)}^{k((t\wedge s)+h)} e^{\eta(u)}du = \bar{\lambda}_0\int_h^{(t\wedge s)+h} e^{\eta(k(u))}k'(u)du = \bar{\lambda}_0(t\wedge s)\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Cov}\Big(\big(\hat{X}(t+h)-\hat{X}(h)\big),\big(\hat{Y}(s+h)-\hat{Y}(h)\big)\Big)\\ &= \bar{\lambda}_0\int_{k(h)}^{k(s+h)} e^{\eta(v)}\big(\eta(k(t+h)\wedge v)-\eta(v)\big)dv\\ &\le \bar{\lambda}_0\big(k(s+h)-k(h)\big)e^{\eta(\infty)}\big(\eta(\infty)-\eta(k(h))\big)\to 0\end{aligned}$$

as $h \to \infty$, and for the covariance of $\hat{Y}$,

$$
\begin{aligned}
&\mathrm{Cov}\Big(\big(\hat{Y}(t+h)-\hat{Y}(h)\big), \big(\hat{Y}(s+h)-\hat{Y}(h)\big)\Big) \\
&= \bar{\lambda}_0 \int_{k(h)}^{k(t+h)} e^{\eta(v)} dv \int_{k(h)}^{k(s+h)} e^{\eta(v')} dv' \int_0^{v\wedge v'} \mathbf{m}_{2,u} e^{-\eta(u)} du \\
&= \bar{\lambda}_0 \int_{h}^{t+h} e^{\eta(k(v))} k'(v) dv \int_{h}^{s+h} e^{\eta(k(v'))} k'(v') dv' \int_0^{k(v\wedge v')} \mathbf{m}_{2,u} e^{-\eta(u)} du \\
&\to \Big(\bar{\lambda}_0 \int_0^\infty \mathbf{m}_{2,u} e^{-\eta(u)} du\Big) \cdot t \cdot s,
\end{aligned}
$$

as $h \to \infty$, which gives the desired limit.

In the case of stationary mark $F_s = F$, we have $\mathbf{m}_{1,s} = \mathbf{m}_1$ and $\eta(t) = \mathbf{m}_1 t$. Thus, all the covariance functions above stays the same with k replaced by $\frac{\ln \mathbf{m}_1 t}{\mathbf{m}_1}$, in which case $e^{\eta(k(v))} k'(v) = 1$ for all $v > 0$ and $\int_0^\infty \mathbf{m}_{2,u} e^{-\eta(u)} du = \frac{\mathbf{m}_2}{\mathbf{m}_1}$. This finishes the proof. □

6.2. Proofs of Propositions 3.3 and 3.4.

Proof of Proposition 3.3. It can be found from Theorem 2.1 that $\bar{N}$ equals to the variance function of $\hat{X}$ in Theorem 2.2. Therefore, we only need to focus on the limits of the covariance functions of the increments of $\big(\hat{X}, \hat{Y}\big)$. Applying Theorem 2.2, for every $t > r > 0$ and $s > r > 0$,

$$
\begin{aligned}
\mathrm{Cov}\big(\hat{X}(t)-\hat{X}(r), \hat{X}(s)-\hat{X}(r)\big) &= \int_r^{t\wedge s} \big(\bar{\lambda}_0(u) + \psi_- * \bar{\lambda}_0(u)\big) du, \\
\mathrm{Cov}\big(\hat{X}(t)-\hat{X}(r), \hat{Y}(s)-\hat{Y}(r)\big) &= \int_r^t du \int_r^s \psi_-(v-u) dv \big(\bar{\lambda}_0(u) + \psi_- * \bar{\lambda}_0(u)\big) du.
\end{aligned}
\tag{6.7}
$$

and making use of the fact $\psi_-(w-v) = 0$ for $w < v$, we obtain

$$
\begin{aligned}
&\mathrm{Cov}\big(\hat{Y}(t)-\hat{Y}(r), \hat{Y}(s)-\hat{Y}(r)\big) \\
&= \int_0^{t\wedge s} \big(\bar{\lambda}_0(u) + \psi_- * \bar{\lambda}_0(u)\big) du \Big(\int_0^\infty \mathcal{HU}\mathbf{1}_{r,t}(u,z) \mathcal{HU}\mathbf{1}_{r,t}(u,z) F(dz)\Big) \\
&= \int_0^t dv \int_0^s dv' \Big(\mathbf{1}(v \ge r) + \int_r^t \psi_-(w-v) dw\Big) \Big(\mathbf{1}(v' \ge r) + \int_r^s \psi_-(w-v') dw\Big) \\
&\quad \times \int_0^{v\wedge v'} F^c\big((v\vee v') - u\big) \big(\bar{\lambda}_0(u) + \psi_- * \bar{\lambda}_0(u)\big) du.
\end{aligned}
\tag{6.8}
$$

Recalling that ψ_- is given by the renewal equation

$$
\psi_-(t) = F^c(t) + \int_0^t \psi_-(u) F^c(t-u) du,
$$

our proof relies on the dominated convergence theorem and the renewal theorems.

Case (i). Assume that $\mathbf{m}_1 \in (0,1)$. Since $F^c * 1(t) \to \mathbf{m}_1$ as $t \to \infty$, we have

$$
\bar{\lambda}_0(t) + \psi_- * \bar{\lambda}_0(t) = \bar{\lambda}_0\big(1 + \psi_- * 1(t)\big) \to \frac{\bar{\lambda}_0}{1-\mathbf{m}_1} \quad \text{as} \quad t \to \infty.
$$

Applying the change of variables to (6.7), we have for $s, t > 0$,

$$
\begin{aligned}
&\mathrm{Cov}\big(\hat{X}(t+h)-\hat{X}(h), \hat{X}(s+h)-\hat{X}(h)\big) \\
&= \int_0^{t\wedge s} \big(\bar{\lambda}_0(u+h) + \psi_- * \bar{\lambda}_0(u+h)\big) du \to_h \frac{\bar{\lambda}_0}{1-\mathbf{m}_1}\big(t \wedge s\big)
\end{aligned}
$$

and

$$\begin{aligned}
&\mathrm{Cov}\big(\hat{X}(t+h)-\hat{X}(h),\hat{Y}(s+h)-\hat{Y}(h)\big)\\
&=\int_0^t du\int_0^s dv\psi_-(v-u)\big(\bar{\lambda}_0(u+h)+\psi_-*\bar{\lambda}_0(u+h)\big)\to_h \frac{\bar{\lambda}_0}{1-\mathbf{m}_1}\int_0^t\int_0^s\psi_-(v-u)dvdu.
\end{aligned}$$

Similarly, applying the change of variables to the second identity in (6.8), we obtain

$$\begin{aligned}
&\mathrm{Cov}\big(\hat{Y}(t+h)-\hat{Y}(h),\hat{Y}(s+h)-\hat{Y}(h)\big)\\
&=\int_{-h}^t dv\int_{-h}^s\Big(\mathbf{1}(v\ge 0)+\int_0^t\psi_-(w-v)dw\Big)\Big(\mathbf{1}(v'\ge 0)+\int_0^s\psi_-(w-v')dw\Big)\\
&\quad\times\int_{-h}^{v\wedge v'}F^c\big((v\vee v')-u\big)\big(\bar{\lambda}_0(u+h)+\psi_-*\bar{\lambda}_0(u+h)\big)du\\
&\to_h \frac{\bar{\lambda}_0}{1-\mathbf{m}_1}\int_{-\infty}^t dv\int_{-\infty}^s dv'\int_{|v-v'|}^\infty F^c(u)du\\
&\quad\times\Big(\mathbf{1}(v\ge 0)+\int_0^t\psi_-(w-v)dw\Big)\Big(\mathbf{1}(v'\ge 0)+\int_0^s\psi_-(w-v')dw\Big)
\end{aligned}$$

which is bounded by $\frac{\mathbf{m}_1\bar{\lambda}_0\cdot t\cdot s}{(1-\mathbf{m}_1)^3}$, and where we use the fact $v\vee v'-v\wedge v'=|v-v'|$.

Case (ii). Assume that $\mathbf{m}_1=1$. Then ψ_- is the renewal density of proper p.d.f. F^c, we have from Smith's key renewal theorem and the elementary renewal theorem that

$$\begin{aligned}
\psi_-(t)=F^c(t)+\psi_-*F^c(t)&\to_t\int_0^\infty F^c(y)dy\Big/\Big(\int_0^\infty yF^c(y)dy\Big)=\frac{2\mathbf{m}_1}{\mathbf{m}_2}=\frac{2}{\mathbf{m}_2},\\
\frac{\bar{\lambda}_0(t)+\psi_-*\bar{\lambda}_0(t)}{t}=\bar{\lambda}_0\frac{1+\psi_-*1(t)}{t}&\to_t\bar{\lambda}_0\Big(\int_0^\infty yF^c(y)dy\Big)^{-1}=\frac{2\bar{\lambda}_0}{\mathbf{m}_2}.
\end{aligned}$$

Applying the change of variables to (6.7) and the dominated convergence theorem, we obtain that as $h\to\infty$,

$$\begin{aligned}
&\mathrm{Cov}\big(\hat{X}(\sqrt{t+h})-\hat{X}(\sqrt{h}),\hat{X}(\sqrt{s+h})-\hat{X}(\sqrt{h})\big)\\
&=\int_0^{t\wedge s}\frac{\bar{\lambda}_0(\sqrt{u+h})+\psi_-*\bar{\lambda}_0(\sqrt{u+h})}{2\sqrt{u+h}}du\to\frac{\bar{\lambda}_0}{\mathbf{m}_2}(t\wedge s)
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Cov}\big(\hat{X}(\sqrt{t+h})-\hat{X}(\sqrt{h}),\hat{Y}(\sqrt{s+h})-\hat{Y}(\sqrt{h})\big)\\
&=\int_0^t\frac{\bar{\lambda}_0(\sqrt{u+h})+\psi_-*\bar{\lambda}_0(\sqrt{u+h})}{2\sqrt{u+h}}du\Big(\int_{\sqrt{h}}^{\sqrt{s+h}}\psi_-(v-\sqrt{u+h})dv\Big)\to 0,
\end{aligned}$$

where we need the fact that, for $u>0$,

$$\int_{\sqrt{h}}^{\sqrt{s+h}}\psi_-(v-\sqrt{u+h})dv=\int_0^{(\sqrt{s+h}-\sqrt{u+h})^+}\psi_-(v)dv\to_h 0.$$

For the covariance of the increments for $\hat{Y}$ in (6.8), the dominated convergence theorem is applied, where we first show the boundedness of the function. For every $u\le\sqrt{t+h}$,

$$\begin{aligned}
\mathcal{HU}\mathbf{1}_{\sqrt{h},\sqrt{t+h}}(u,z)&=\int_{\sqrt{h}}^{\sqrt{t+h}}\mathbf{1}(v\in[u,u+z))dv+\int_{\sqrt{h}}^{\sqrt{t+h}}\psi_-(w-v)dw\int_0^{\sqrt{t+h}}\mathbf{1}(v\in[u,u+z))dv\\
&\le\big(\sqrt{t+h}-\sqrt{h}\big)\big(1+Mz\big)\le\frac{\sqrt{t}}{\sqrt{h}}\big(1+Mz\big),
\end{aligned}$$

where M denote an upper bound for ψ_- given (6.2). Moreover, for fixed $u \in (0,1), z > 0$, and every h large enough so that $\sqrt{uh} + z < \sqrt{h}$, we have from (6.2),

$$\sqrt{h} \cdot \mathcal{HU}\mathbf{1}_{\sqrt{h},\sqrt{t+h}}(\sqrt{uh}, z) = \sqrt{h} \int_{\sqrt{h}}^{\sqrt{t+h}} dw \int_{\sqrt{uh}}^{\sqrt{uh}+z} \psi_-(w-v)dv \to \frac{t}{2} z \frac{2}{\mathbf{m}_2} = \frac{tz}{\mathbf{m}_2}.$$

With the boundedness and the limits above, we are ready to obtain

$$\begin{aligned}
&\mathrm{Cov}\big(\hat{Y}(\sqrt{t+h}) - \hat{Y}(\sqrt{h}), \hat{Y}(\sqrt{s+h}) - \hat{Y}(\sqrt{h})\big) \\
&= \bar{\lambda}_0 \int_0^{1+\frac{t\wedge s}{h}} \frac{1 + \psi_- * 1(\sqrt{uh})}{\sqrt{h}} \frac{du}{2\sqrt{u}} \Big(\int_0^\infty F(dz) \sqrt{h} \cdot \mathcal{HU}\mathbf{1}_{\sqrt{h},\sqrt{t+h}}(\sqrt{uh}, z) \\
&\qquad \times \sqrt{h} \cdot \mathcal{HU}\mathbf{1}_{\sqrt{h},\sqrt{s+h}}(\sqrt{uh}, z) \Big) \\
&\to_h \frac{ts\bar{\lambda}_0}{\mathbf{m}_2} \int_0^1 \frac{du}{\mathbf{m}_2} = \frac{\bar{\lambda}_0 \cdot t \cdot s}{\mathbf{m}_2^2}.
\end{aligned}$$

Case (iii). Assume that $\mathbf{m}_1 > 1$. Recall that $\rho_- > 0$ is the solution to $\int_0^\infty e^{-\rho_- y} F^c(y) dy = 1$. Then $e^{-\rho_- t}\psi_-(t)$ is the renewal density of the proper p.d.f. $e^{-\rho_- y}F^c(y)$. Similar to (6.2) in Case (ii), we have

$$\begin{aligned}
e^{-\rho_- t}\psi_-(t) &\to_t \Big(\int_0^\infty y e^{-\rho_- y} F^c(y) dy \Big)^{-1}, \\
\frac{\bar{\lambda}_0(t) + \psi_- * \bar{\lambda}_0(t)}{e^{\rho_- t}} &\to_t \Big(\rho_- \int_0^\infty y e^{-\rho_- y} F^c(y) dy \Big)^{-1}.
\end{aligned}$$

To simplify notation, we denote by $\frac{\ln(t+1)}{\rho_-} = k(t)$ for $t > 0$. Then

$$\int_0^{k(t)} \rho_- e^{\rho_- u} du = t \quad \text{and} \quad k'(t) = \big(\rho_- e^{\rho_- u}\big)^{-1}\Big|_{u=k(t)} \quad \forall t > 0.$$

Applying the change of variables to (6.7), as $h \to \infty$, we obtain

$$\begin{aligned}
&\mathrm{Cov}\big(\hat{X}(k(t+h)) - \hat{X}(k(h)), \hat{X}(k(s+h)) - \hat{X}(k(h))\big) \\
&= \int_0^{t\wedge s} \frac{\bar{\lambda}_0(u) + \psi_- * \bar{\lambda}_0(u)}{\rho_- e^{\rho_- u}}\Big|_{u=k(v+h)} dv \to_h \frac{\bar{\lambda}_0 \cdot t \wedge s}{\rho_-^2 \int_0^\infty y e^{-\rho_- y} F^c(y) dy},
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Cov}\big(\hat{X}(k(t+h)) - \hat{X}(k(h)), \hat{Y}(k(s+h)) - \hat{Y}(k(h))\big) \\
&= \int_0^t \frac{\bar{\lambda}_0(u) + \psi_- * \bar{\lambda}_0(u)}{\rho_- e^{\rho_- u}}\Big|_{u=k(v+h)} dv \int_{k(h)}^{k(s+h)} \psi_-(w - k(v+h)) dw \to_h 0,
\end{aligned}$$

where for $v \in (0,t)$ we have

$$\int_{k(h)}^{k(s+h)} \psi_-(w - k(v+h)) dw = \int_0^{(k(s+h)-k(v+h))^+} \psi_-(w) dw \to_h 0$$

since $k(s+h) - k(h) \to_h 0$.

For the covariance of $\hat{Y}$ in (6.8), similar to the previous case, we first have

$$\begin{aligned}
&\mathcal{HU}\mathbf{1}_{k(h),k(t+h)}(u,z) \\
&= \int_{k(h)}^{k(t+h)} \mathbf{1}(v \in [u, u+z)) dv + \int_{k(h)}^{k(t+h)} \psi_-(w-v) dw \int_0^{k(t+h)} \mathbf{1}(v \in [u, u+z)) dv
\end{aligned}$$

$$\leq \big(k(t+h)-k(h)\big) + M\int_{k(h)}^{k(t+h)} e^{\rho_-(w-v)}dw \int_u^{u+z} dv$$
$$= (k(t+h)-k(h)) + \frac{M}{\rho_-^2} t e^{-\rho_- u}\big(1-e^{-\rho_- z}\big)$$

where M denotes an upper bound for $e^{-\rho_- t}\psi_-(t)$. Moreover, for every $u,z>0$ and h large enough,

$$\mathcal{HU}\mathbf{1}_{k(h),k(t+h)}(u,z) = \; o(1) + \int_0^t \frac{\psi_-(w-v)}{\rho_- e^{\rho_-(w-v)}}\Big|_{w=k(s+h)} ds \int_u^{u+z} dv e^{-\rho_- v}$$
$$\to_h \frac{t\cdot e^{-\rho_- u}\cdot(1-e^{-\rho_- z})}{\rho_-^2 \int_0^\infty y e^{-\rho_- y} F^c(y)dy}.$$

Therefore, we have from the change of variables and the dominated convergence theorem that

$$\mathrm{Cov}\big(\hat{Y}(k(t+h)) - \hat{Y}(k(h)), \hat{Y}(k(s+h)) - \hat{Y}(k(h))\big)$$
$$= \int_0^{k((t\wedge s)+h)} \big(\bar{\lambda}_0(u) + \psi_- * \bar{\lambda}_0(u)\big) \int_0^\infty F(dz)\Big(\mathcal{HU}\mathbf{1}_{k(h),k(t+h)}(u,z)\mathcal{HU}\mathbf{1}_{k(h),k(s+h)}(u,z)\Big)$$
$$\to_h \frac{t\times s}{\big(\rho_-^2\int_0^\infty y e^{-\rho_- y}F^c(y)dy\big)^2}\Big(\int_0^\infty \big(\bar{\lambda}_0(u)+\psi_- * \bar{\lambda}_0(u)\big)e^{-2\rho_- u}du\Big)\int_0^\infty \big(1-e^{-\rho_- z}\big)^2 F(dz)$$
$$= \frac{\bar{\lambda}_0}{\rho_-^4\big(\int_0^\infty y e^{-\rho_- y}F^c(y)dy\big)^2}\times t\times s$$

where we make use of the fact that

$$\int_0^\infty e^{-sz}F(dz) = s\int_0^\infty e^{-sy}F(y)dy = 1 - s\int_0^\infty e^{-sy}F^c(y)dy.$$

This finishes the proof. □

Proof of Proposition 3.4. The proof follows the idea similar to Case (iii) in Proposition 3.3. Let ρ_+ be the constant such that $\int_0^\infty e^{-\rho_+ y}F(y)dy = 1$. Then we have

$$e^{-\rho_+ t}\psi_+(t) \to_t \left(\int_0^\infty y e^{-\rho_+ y}F(y)dy\right)^{-1}$$
$$\frac{\bar{\lambda}_0(t) + \psi_+ * \bar{\lambda}_0(t)}{e^{\rho_+ t}} \to_t \left(\rho_+\int_0^\infty y e^{-\rho_+ y}F(y)dy\right)^{-1},$$

from the renewal theorem. Let $k(t) = \frac{\ln(t+1)}{\rho_+}$ for $t>0$. Similar to (6.7), we obtain

$$\mathrm{Cov}\big(\hat{X}(k(t+h)) - \hat{X}(k(h)), \hat{X}(k(s+h)) - \hat{X}(k(h))\big)$$
$$= \int_0^{t\wedge s} \frac{\bar{\lambda}_0(u) + \psi_+ * \bar{\lambda}_0(u)}{\rho_+ e^{\rho_+ u}}\Big|_{u=k(v+h)} dv \to_h \frac{\bar{\lambda}_0\cdot t\wedge s}{\rho_+^2\int_0^\infty y e^{-\rho_+ y}F^c(y)dy},$$

and

$$\mathrm{Cov}\big(\hat{X}(k(t+h)) - \hat{X}(k(h)), \hat{Y}(k(s+h)) - \hat{Y}(k(h))\big)$$
$$= \int_0^t \frac{\bar{\lambda}_0(u) + \psi_+ * \bar{\lambda}_0(u)}{\rho_+ e^{\rho_- u}}\Big|_{u=k(v+h)} dv \int_{k(h)}^{k(s+h)} \psi_+(w-k(v+h))dw \to_h 0.$$

For the covariance of $\hat{Y}$, similar to (6.8), we have

$$\mathcal{HU}\mathbf{1}_{k(h),k(t+h)}(u,z) = \int_{k(h)}^{k(t+h)} \mathbf{1}(z\leq v-u)dv + \int_{k(h)}^{k(t+h)} \psi_+(w-v)dw \int_0^{k(t+h)} \mathbf{1}(z\leq v-u)dv$$

$$\leq \big(k(t+h)-k(h)\big) + M\int_{k(h)}^{k(t+h)} e^{\rho_+(w-v)}dw \int_0^{k(t+h)} dv$$
$$\leq \big(k(t+h)-k(h)\big) + \frac{Mt}{\rho_+^2}.$$

For $u<k(t+h)$, making use of the fact $\psi(w-v)=0$ for $w<v$, we have

$$\begin{aligned}\mathcal{HU}\mathbf{1}_{k(h),k(t+h)}(u,z) =&\ o(1) + \int_{k(h)}^{k(t+h)} dw \int_{u+z}^{w} \psi_+(w-v)dv \\ =&\ o(1) + e^{-\rho_+(u+z)}\int_0^t \frac{\psi * 1(v-u-z)}{\rho_+ e^{\rho_+(v-u-z)}}\bigg|_{v=k(s+h)} ds \\ \to_h&\ \frac{t\cdot e^{-\rho_+(u+z)}}{\rho_+^2\int_0^\infty ye^{-\rho_+ y}F(y)dy}.\end{aligned}$$

Therefore, applying the dominated convergence theorem, we obtain

$$\begin{aligned}&\operatorname{Cov}\big(\hat{Y}(k(t+h))-\hat{Y}(k(h)),\hat{Y}(k(s+h))-\hat{Y}(k(h))\big)\\ \to_h&\ \frac{t\times s}{\big(\rho_+^2\int_0^\infty ye^{-\rho_+ y}F(y)dy\big)^2}\bigg(\int_0^\infty\big(\bar{\lambda}_0(u)+\psi_+ * \bar{\lambda}_0(u)e^{-2\rho_+ u}\big)du\bigg)\bigg(\int_0^\infty e^{-2\rho_+ z}F(dz)\bigg)\\ =&\ \frac{\bar{\lambda}_0}{\rho_+^4\big(\int_0^\infty ye^{-\rho y}F(y)dy\big)^2}\frac{\int_0^\infty e^{-2\rho_+ y}F(y)dy}{1-\int_0^\infty e^{-2\rho_+ y}F(y)dy}\times t\times s.\end{aligned}$$

This finishes the proof. □

Acknowledgements

The authors are thankful to the anonymous referees for their helpful comments that have greatly improved the exposition of the paper. G. Pang was supported in part by NSF grants DMS-1715875 and DMS-2108683.

References

[1] Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-Francois Muzy. Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499, 2013.

[2] Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.

[3] Jean Bertoin. *Lévy Processes*, volume 121. Cambridge university press Cambridge, 1996.

[4] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 1999.

[5] P Brémaud, G Nappo, and GL Torrisi. Rate of convergence to equilibrium of marked Hawkes processes. *Journal of Applied Probability*, 39(1):123–136, 2002.

[6] Pierre Brémaud and Laurent Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3):1563–1588, 1996.

[7] Pierre Brémaud and Laurent Massoulié. Power spectra of general shot noises and Hawkes point processes with a random excitation. *Advances in Applied Probability*, 34(1):205–222, 2002.

[8] Xinyun Chen. Perfect sampling of Hawkes processes and queues with Hawkes arrivals. *arXiv preprint arXiv:2002.06369*, 2020.

[9] Julien Chevallier. Fluctuations for mean-field interacting age-dependent Hawkes processes. *Electronic Journal of Probability*, 22, 2017.

[10] Julien Chevallier. Mean-field limit of generalized Hawkes processes. *Stochastic Processes and their Applications*, 127(12):3870–3912, 2017.

[11] Julien Chevallier, A Duarte, E Löcherbach, and G Ost. Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. *Stochastic Processes and their Applications*, 129(1):1–27, 2019.

[12] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes Volume I: Elementary Theory and Methods*. Springer, New York, 2003.

[13] Andrew Daw and Jamol Pender. Queues driven by Hawkes processes. *Stochastic Systems*, 8(3):192–229, 2018.

[14] Justin Dean, Ayalvadi Ganesh, and Edward Crane. Functional large deviations for Cox processes and $Cox/g/\infty$ queues, with a biological application. *The Annals of Applied Probability*, 30(5):2465–2490, 2020.

[15] Sylvain Delattre, Nicolas Fournier, and Marc Hoffmann. Hawkes processes on large networks. *The Annals of Applied Probability*, 26(1):216–261, 2016.

[16] Peter T Ellison and Mary Ann Ottinger. A comparative perspective on reproductive aging, reproductive cessation, post-reproductive life, and social behavior. *in Sociality, Hierarchy, Health: Comparative Biodemography: A Collection of Papers*, pages 315–338, 2014.

[17] Raphaël Forien, Guodong Pang, and Étienne Pardoux. Epidemic models with varying infectivity. *SIAM Journal on Applied Mathematics*, forthcoming, 2021.

[18] Raphaël Forien, Guodong Pang, and Étienne Pardoux. Recent advances in epidemic modeling: Non-Markov stochastic models and their scaling limits. page arXiv:2106.08466, 2021.

[19] Raphaël Forien, Guodong Pang, Étienne Pardoux, and Arsene Brice Zotsa-Ngoufack. Epidemic models with varying infectivity and immunity/susceptibility. *in preparation*, 2021.

[20] Fuqing Gao and Lingjiong Zhu. Some asymptotic results for nonlinear Hawkes processes. *Stochastic Processes and their Applications*, 128(12):4051–4077, 2018.

[21] Xuefeng Gao and Lingjiong Zhu. Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues. *Queueing Systems*, 90(1-2):161–206, 2018.

[22] Xuefeng Gao and Lingjiong Zhu. Large deviations and applications for Markovian Hawkes processes with a large initial intensity. *Bernoulli*, 24(4A):2875–2905, 2018.

[23] Xuefeng Gao and Lingjiong Zhu. Limit theorems for Markovian Hawkes processes with a large initial intensity. *Stochastic Processes and their Applications*, 128(11):3807–3839, 2018.

[24] Alan G Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3):438–443, 1971.

[25] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[26] Alan G Hawkes. Hawkes processes and their applications to finance: a review. *Quantitative Finance*, 18(2):193–198, 2018.

[27] Alan G Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[28] Ulrich Horst and Wei Xu. Functional limit theorems for marked Hawkes point measures. *arXiv preprint arXiv:1908.06703*, 2019.

[29] Thibault Jaisson and Mathieu Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *The annals of applied probability*, 25(2):600–631, 2015.

[30] Thibault Jaisson and Mathieu Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *The Annals of Applied Probability*, 26(5):2860–2882, 2016.

[31] Dmytro Karabash and Lingjiong Zhu. Limit theorems for marked Hawkes processes with application to a risk model. *Stochastic Models*, 31(3):433–451, 2015.

[32] David T Koops, Mayank Saxena, Onno J Boxma, and Michel Mandjes. Infinite-server queues with Hawkes input. *Journal of Applied Probability*, 55(3):920–943, 2018.

[33] Mirkka Lahdenperä, Khyne U Mar, and Virpi Lummaa. Reproductive cessation and post-reproductive lifespan in asian elephants and pre-industrial humans. *Frontiers in zoology*, 11(1):1–14, 2014.

[34] Young Lee, Kar Wai Lim, and Cheng Soon Ong. Hawkes processes with stochastic excitations. In *International Conference on Machine Learning*, pages 79–88. PMLR, 2016.

[35] Bo Li and Guodong Pang. Shot noise processes with randomly delayed cluster arrivals and dependent noises in the large intensity regime. *Advances in Applied Probability*, forthcoming, 2021.

[36] E Löcherbach. Spiking neurons: Interacting Hawkes processes, mean field limits and oscillations. *ESAIM: Proceedings and Surveys*, 60:90–103, 2017.

[37] Hongyuan Mei and Jason M Eisner. The neural Hawkes process: A neurally self-modulating multivariate point process. In *Advances in Neural Information Processing Systems*, pages 6754–6764, 2017.

[38] David Oakes. The Markovian self-exciting process. *Journal of Applied Probability*, 12(1):69–77, 1975.

[39] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401):9–27, 1988.

[40] Teri J Orr and Marlene Zuk. Reproductive delays in mammals: an unexplored avenue for post-copulatory sexual selection. *Biological Reviews*, 89(4):889–912, 2014.

[41] Craig Packer, Marc Tatar, and Anthony Collins. Reproductive cessation in female mammals. *Nature*, 392(6678):807–811, 1998.

[42] Guodong Pang and Étienne Pardoux. Functional central limit theorems for epidemic models with varying infectivity. *arXiv:2009.11861*, 2021.

[43] Guodong Pang and Étienne Pardoux. Functional law of large numbers and PDEs for epidemic models with infection-age dependent infectivity. *arXiv:2106.03758*, 2021.

[44] Guodong Pang and Murad S Taqqu. Nonstationary self-similar Gaussian processes as scaling limits of power-law shot noise processes and generalizations of fractional Brownian motion. *High Frequency*, 2(2):95–112, 2019.

[45] Guodong Pang and Yuhang Zhou. Functional limit theorems for a new class of non-stationary shot noise processes. *Stochastic Processes and their Applications*, 128(2):505–544, 2018.

[46] Guodong Pang and Yuhang Zhou. Functional limit theorems for shot noise processes with weakly dependent noises. *Stochastic Systems*, 10(2):99–191, 2020.

[47] Mikael Sandell. The evolution of seasonal delayed implantation. *The Quarterly Review of Biology*, 65(1):23–42, 1990.

[48] Enrico Scalas and Noèlia Viles. A functional limit theorem for stochastic integrals driven by a time-changed symmetric α-stable lévy process. *Stochastic processes and their applications*, 124(1):385–410, 2014.

[49] Shripad Tuljapurkar. Delayed reproduction and fitness in variable environments. *Proceedings of the National Academy of Sciences*, 87(3):1139–1143, 1990.

[50] Ward Whitt. *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and their Application to Queues*. Springer Science & Business Media, 2002.

[51] Lingjiong Zhu. Central limit theorem for nonlinear Hawkes processes. *Journal of Applied Probability*, 50(3):760–771, 2013.

School of Mathematics, Nankai University, Tianjin, 300071 China
Email address: `libo@nankai.edu.cn`

The Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, College of Engineering, Pennsylvania State University, University Park, PA 16802
Email address: `gup3@psu.edu`