ON THE HAWKES PROCESS WITH DIFFERENT EXCITING FUNCTIONS

BEHZAD MEHRDAD AND LINGJIONG ZHU

Abstract. The Hawkes process is a simple point process, whose intensity function depends on the entire past history and is self-exciting and has the clustering property. The Hawkes process is in general non-Markovian. The linear Hawkes process has immigration-birth representation. Based on that, Fierro et al. recently introduced a generalized linear Hawkes model with different exciting functions. In this paper, we study the convergence to equilibrium, large deviation principle, and moderate deviation principle for this generalized model. This model also has connections to the multivariate linear Hawkes process. Some applications to finance are also discussed.

Contents

1. Introduction

1.1. **Hawkes Process.** Let N be a simple point process on $\mathbb{R}$ and let $\mathcal{F}_t^{-\infty} := \sigma(N(C), C \in \mathcal{B}(\mathbb{R}), C \subset (-\infty, t])$ be an increasing family of σ-algebras. Any non-negative $\mathcal{F}_t^{-\infty}$-progressively measurable process λ_t with

$$\mathbb{E}\left[N(a,b]|\mathcal{F}_a^{-\infty}\right] = \mathbb{E}\left[\int_a^b \lambda_s ds \middle| \mathcal{F}_a^{-\infty}\right] \tag{1.1}$$

Date: 26 August 2014. *Revised:* 7 March 2015.
2000 *Mathematics Subject Classification.* 60G55, 60F10.
Key words and phrases. point process, Hawkes process, self and mutually exciting process, large deviations, moderate deviations, convergence to equilibrium, microstructure noise, ruin probabilities.

a.s. for all intervals $(a,b]$ is called an $\mathcal{F}_t^{-\infty}$-intensity of N. We use the notation $N_t := N(0,t]$ to denote the number of points in the interval $(0,t]$.

A Hawkes process is a simple point process N admitting an $\mathcal{F}_t^{-\infty}$-intensity

$$\lambda_t := \lambda\left(\int_{-\infty}^{t} h(t-s)N(ds)\right), \tag{1.2}$$

where $\lambda(\cdot):\mathbb{R}^+\to\mathbb{R}^+$ is locally integrable, left continuous, $h(\cdot):\mathbb{R}^+\to\mathbb{R}^+$ and we always assume that $\|h\|_{L^1}=\int_0^\infty h(t)dt<\infty$. In (1.2), $\int_{-\infty}^t h(t-s)N(ds)$ stands for $\int_{(-\infty,t)} h(t-s)N(ds)=\sum_{\tau<t} h(t-\tau)$, where τ are the occurrences of the points before time t.

In the literature, $h(\cdot)$ and $\lambda(\cdot)$ are usually referred to as exciting function and rate function respectively.

When $\lambda(\cdot)$ is linear, the Hawkes process is said to be linear and it is named after Hawkes [18]. The linear Hawkes process can be studied via immigration-birth representation, see e.g. Hawkes and Oakes [19]. When $\lambda(\cdot)$ is nonlinear, the Hawkes process is said to be nonlinear and the nonlinear Hawkes process was first introduced by Brémaud and Massoulié [6].

The law of large numbers and central limit theorems for linear Hawkes processes were studied in e.g. Bacry et al. [2], and the moderate deviations were studied in Zhu [29]. The central limit theorem for nonlinear Hawkes processes was obtained in Zhu [32]. Bordenave and Torrisi [5] obtained the large deviations for linear Hawkes processes and the large deviations for nonlinear Hawkes processes were studied in Zhu [33] and Zhu [34]. The limit theorems of some generalizations of the classical Hawkes processes have been studied in e.g. Karabash and Zhu [21] and Zhu [30].

The self-exciting and clutstering properties of the Hawkes process make it ideal to characterize the correlations in some complex systems, including finance. Bacry et al. [2], Bacry et al. [3] studied microstructure noise and Epps effect using the Hawkes models. Chavez-Demoulin et al. [7] studied value-at-risk. Errais et al. [14] used Hawkes process to model the credit risk. Embrechts et al. [12] fit the Hawkes process to financial data.

The Hawkes process has also been applied to many other fields, including seismology, see e.g. Hawkes and Adamopoulos [20], Ogata [23], sociology, see e.g. Crane and Sornette [9] and Blundell et al. [4], and neuroscience, see e.g. Chornoboy et al. [8], Pernice et al. [24], Pernice et al. [25]. For a survey of the Hawkes process and its applications, we refer to Liniger [22] and Zhu [28].

1.2. **Hawkes Process with Different Exciting Functions.** In this paper, we are interested to study an extension of the linear Hawkes process proposed by Fierro et al. [15]. It is based on the immigration-birth representation structure of the linear Hawkes process. The classical Hawkes process can be constructed from a homogeneous Poisson process (immigration) and using the same exciting function for different generations of offspring (birth). In some fields, e.g. seismology, where main shocks produce aftershocks with possibly different intensities, that naturally leads to the study of a Hawkes process with different exciting functions as proposed in Fierro et al. [15].

Let $(N^n)_{n\in\mathbb{N}}$ be a sequence of non-explosive simple point processes without common jumps so that

- N^0 is an inhomogeneous Poisson process with intensity $\gamma_0(t)$ at time t.

- For every $n \in \mathbb{N}$, N^n is a simple point process with intensity $\lambda_t^n = \int_0^t \gamma_n(t-s)N^{n-1}(ds)$, where the integral $\int_0^t \gamma_n(t-s)N^{n-1}(ds)$ denotes for $\int_{(0,t)} \gamma_n(t-s)N^{n-1}(ds) = \sum_{\tau\in N^{n-1}, 0<\tau<t} \gamma_n(t-\tau)$. Note that by definition, the intensity is $\mathcal{F}_t$-predictable.
- For every $n \in \mathbb{N}\cup\{0\}$, conditional on $N^0, \ldots, N^n$, N^{n+1} is a inhomogeneous Poisson process with intensity λ^{n+1}.

The existence of such a process was proved as Proposition 2.1. in Fierro et al. [15].

Using the notation of immigration-birth representation, N^0 is called the immigrant process and N^n the nth generation offspring process.

Let $N := \sum_{n=0}^{\infty} N^n$. N is said to be the Hawkes process with excitation functions $(\gamma_n)_{n\in\mathbb{N}\cup\{0\}}$. If $\gamma_0(t) \equiv \overline{\gamma}_0 > 0$ and $\gamma_n(t) = h(t)$ for any $n \in \mathbb{N}$, then the model reduces to the classical linear Hawkes process N with intensity at time t given by

$$\lambda_t = \overline{\gamma}_0 + \int_0^t h(t-s)N(ds). \tag{1.3}$$

Assumption 1. *(i)* $\overline{\gamma}_0 := \lim_{t\to\infty} \frac{1}{t}\int_0^t \gamma_0(s)ds$ *exists and is finite.*

(ii) $\rho := \sup_{n\in\mathbb{N}} \int_0^\infty \gamma_n(t)dt < 1$.

Under Assumption 1, Fierro et al. [15] showed that for any $t \geq 0$,

$$\mathbb{E}[N_t] = \int_0^t \sum_{n=0}^{\infty} (\gamma_0 * \cdots * \gamma_n)(s)ds < \infty. \tag{1.4}$$

Fierro et al. [15] proved the following law of large numbers result under Assumption 1,

$$\frac{N_t}{t} \to m, \qquad \text{almost surely, as } t \to \infty, \tag{1.5}$$

where $m := \sum_{n=0}^{\infty} m_n$, $m_0 := \overline{\gamma}_0$ and

$$m_n := \overline{\gamma}_0 \prod_{i=1}^{n} \int_0^\infty \gamma_i(u)du, \qquad n \in \mathbb{N}. \tag{1.6}$$

It is easy to check that in the case of classical linear Hawkes process (1.3),

$$m = \overline{\gamma}_0 \sum_{n=1}^{\infty} \|h\|_{L^1}^n = \frac{\overline{\gamma}_0}{1 - \|h\|_{L^1}}, \tag{1.7}$$

which is consistent with the results in Hawkes [18].

Assumption 2.

$$\lim_{t\to\infty} \sqrt{t}\left[\frac{1}{t}\int_0^t \gamma_0(s)ds - \overline{\gamma}_0\right] = 0, \tag{1.8}$$

and

$$\lim_{t\to\infty} \sqrt{t} \int_t^\infty \sum_{p=1}^{\infty} \gamma_p * \cdots * \gamma_1(s)ds = 0. \tag{1.9}$$

Further assume Assumption 2, Fierro et al. [15] also obtained the central limit theorem, which is the main result of their paper,

$$\frac{N_t - mt}{\sqrt{t}} \to N(0, \sigma^2), \tag{1.10}$$

in distribution as $t \to \infty$, where

$$\sigma^2 := \sum_{j=0}^{\infty} \left(1 + \sum_{p=1}^{\infty} \prod_{i=j+1}^{p+j} \int_0^{\infty} \gamma_i(u) du \right)^2 m_j. \tag{1.11}$$

It is easy to check that in the case of classical linear Hawkes process (1.3),

$$\sigma^2 = \sum_{j=0}^{\infty} \left(1 + \sum_{p=0}^{\infty} \|h\|_{L^1}^p \right)^2 \overline{\gamma}_0 \|h\|_{L^1}^j = \frac{\overline{\gamma}_0}{(1 - \|h\|_{L^1})^3}, \tag{1.12}$$

which is consistent with the results in Bacry et al. [2].

The paper is organized as the following. In Section 2, we show that there exists a stationary version of the Hawkes process with different exciting functions and we will show the convergence to the equilibrium. In Section 3, we will point out the connections of the Hawkes process with different exciting functions to the classical multivariate linear Hawkes process, which has been well studied in the literature. In Section 4, we obtain both the large deviations and the moderate deviations for the model. Finally, we discuss some applications to finance in Section 5.

2. Convergence to Equilibrium

Assume that $\gamma_0 \equiv \overline{\gamma}_0$ is a positive constant and Assumption 1 (ii) holds, then, there exists a stationary version of the Hawkes process $N^\dagger$ with exciting functions $(\gamma_n)_{n \in \mathbb{N}} \cup \{\overline{\gamma}_0\}$ constructed as follows.

Let $N^{\dagger,0}$ be a homogeneous Poisson process with intesntiy $\overline{\gamma}_0$ on $\mathbb{R}$ and for each $n \in \mathbb{N}$, $N^{\dagger,n}$ is an inhomogeneous Poisson process with intensity

$$\lambda_t^{\dagger,n} = \int_{-\infty}^{t} \gamma^n(t-s) N^{\dagger,n-1}(ds), \tag{2.1}$$

and $N^\dagger = \sum_{n=0}^{\infty} N^{\dagger,n}$.

The space of integer-valued measures is endowed with the vague topology, i.e. N^n converges to N if and only if for any continuous function ϕ with compact support, $\int \phi(x) N^n(dx) \to \int \phi(x) N(dx)$.

Given a simple point process N on $\mathbb{R}$, one can define $\theta_t N$ as the process shifted by time t, i.e. $\theta_t N(A) = N(A+t)$, where $A + t := \{s + t : s \in A\}$ for any Borel set A associated with the vague topology.

We say a sequence of simple point processes N^n converges to a simple point process N in distribution if for any Borel set A associated with the vague topology, $\lim_{n\to\infty} \mathbb{P}(N^n \in A) = \mathbb{P}(N \in A)$ and the convergences is in variation if

$$\lim_{n\to\infty} \sup_{A} |\mathbb{P}(N^n \in A) - \mathbb{P}(N \in A)| = 0. \tag{2.2}$$

This is the notation given in Brémaud and Massoulié [6].

In Daley and Vere-Jones [10]'s terminology, convergence in distribution (variation) is referred to as the weak (strong) convergence and the stationarity associated with the stationary limit is referred to as the weak (strong) stationarity.

For a given simple point process N on $\mathbb{R}$, let N^+ be its restriction to $\mathbb{R}^+$.

Theorem 1. *Let $N = \sum_{n=0}^{\infty} N^n$ be the Hawkes process with exciting functions $(\gamma_n)_{n \in \mathbb{N}} \cup \{\overline{\gamma}_0\}$ with empty history, i.e. $N(-\infty, 0] = 0$ and satisfies Assumption 1 (ii). Then, the following is true.*

(i) $\theta_s N$ converges to $N^\dagger$ weakly as $s \to \infty$, i.e. $(\theta_s N)^+$ converges in distribution to $(N^\dagger)^+$.

(ii) If we further assume that $\eta := \sup_{n\in\mathbb{N}} \int_0^\infty t\gamma_n(t)dt < \infty$, then, $\theta_s N$ converges to $N^\dagger$ strongly as $s \to \infty$, i.e. $(\theta_s N)^+$ converges in variation to $(N^\dagger)^+$.

Proof. (i) For both $N^\dagger$ and N, let $\theta_s N^\dagger$ and $\theta_s N$ be the shifted version obtained by setting time s as the origin and shift the process backwards in times by s to bring the origin back to 0, that is, $\theta_s N(A) = N(A+s)$, where $A+s := \{t+s : t \in A\}$ for any Borel set A. We can decompose $\theta_s N^\dagger$ into two components, one component has the same dynamics as $\theta_s N$, being built from the points generated by the homogeneous Poisson process $\overline{\gamma}_0$ and its offspring $(N^n)_{n\geq 1}$ after time $-s$, the other component $N^\dagger_{-s}$ that consists of the offspring of the points generated by homogeneous Poisson process $\overline{\gamma}_0$ before time $-s$. Hence, we have

$$\lambda^{\dagger,1}_{-s}(t) = \int_{-\infty}^{-s} \gamma_1(t-u)N^{\dagger,0}_{-s}(du), \qquad t \geq -s, \tag{2.3}$$

and

$$\lambda^{\dagger,n}_{-s}(t) = \int_{-\infty}^{t} \gamma_n(t-u)N^{\dagger,n-1}_{-s}(du), \qquad t \geq -s, n \geq 2. \tag{2.4}$$

Let us define

$$H_n(t) := \int_t^\infty \gamma_n(s)ds, \qquad n \geq 1. \tag{2.5}$$

It is easy to compute that

$$\mathbb{E}[\lambda^{\dagger,1}_{-s}(t)] = \int_{-\infty}^{-s} \gamma_1(t-u)\overline{\gamma}_0 du = \overline{\gamma}_0 \int_{t+s}^\infty \gamma_1(u)du = \overline{\gamma}_0 H_1(t+s), \tag{2.6}$$

and

$$\begin{aligned}\mathbb{E}[\lambda^{\dagger,2}_{-s}(t)] &= \overline{\gamma}_0 \|\gamma_1\|_{L^1} \int_{t+s}^\infty \gamma_2(u)du + \overline{\gamma}_0 \int_0^{t+s} \gamma_2(t+s-u)H_1(u)du \\ &= \overline{\gamma}_0 \|\gamma_1\|_{L^1} H_2(t+s) + \overline{\gamma}_0(\gamma_2 * H_1)(t+s).\end{aligned} \tag{2.7}$$

Similarly, we have

$$\mathbb{E}[\lambda^{\dagger,3}_{-s}(t)] = \overline{\gamma}_0\|\gamma_1\|_{L^1}\|\gamma_2\|_{L^1}H_3(t+s)+\overline{\gamma}_0\|\gamma_1\|_{L^1}(\gamma_3*H_2)(t+s)+\overline{\gamma}_0(\gamma_3*\gamma_2*H_1)(t+s). \tag{2.8}$$

Iteratively, we get

$$\mathbb{E}[\lambda^{\dagger,n}_{-s}(t)] = \overline{\gamma}_0 \sum_{j=0}^{n-1} \left(\sum_{i=1}^{j} \|\gamma_i\|_{L^1} \right) \left(\gamma_n * \cdots * \gamma_{j+2} * H_{j+1}\right)(t+s), \tag{2.9}$$

where $\prod_{i=1}^0 \|\gamma_i\|_{L^1} := 1$.

Therefore, for any $T > 0$,

$$\begin{aligned}
\mathbb{P}(N^{\dagger}_{-s}(0,T) > 0) &= 1 - \mathbb{E}\left[\exp\left(-\int_0^T \lambda^{\dagger}_s(t)dt\right)\right] \\
&\leq \mathbb{E}\left[\int_0^T \lambda^{\dagger}_{-s}(t)dt\right] \\
&= \int_0^T \sum_{n=1}^{\infty} \mathbb{E}[\lambda^{\dagger,n}_{-s}(t)]dt,
\end{aligned} \tag{2.10}$$

where $\lambda^{\dagger}_s(t) := \sum_{n=1}^{\infty} \lambda^{\dagger,n}_{-s}(t)$ is the intensity of $N^{\dagger}_{-s}$ at time t.

Since $\gamma_n(t)$ is integrable for any n, $H_n(t) \to 0$ as $t \to \infty$. Thus, $\mathbb{E}[\lambda^{\dagger,n}_{-s}(t)] \to 0$ as $s \to \infty$ for any t. Moreover,

$$\mathbb{E}[\lambda^{\dagger,n}_{-s}(t)] \leq \overline{\gamma}_0 \sum_{j=0}^{n-1} \left(\sum_{i=1}^{j} \|\gamma_i\|_{L^1}\right) H_{j+1}(0) \left(\gamma_n * \cdots * \gamma_{j+2} * 1\right)(t+s) \tag{2.11}$$

Thus, for any t,

$$\limsup_{s\to\infty} \mathbb{E}[\lambda^{\dagger,n}_{-s}(t)] = \overline{\gamma}_0 n \prod_{i=1}^{n} \|\gamma_i\|_{L^1} \leq \overline{\gamma}_0 n \rho^n, \tag{2.12}$$

by Assumption 1 (ii), which is summable in n. Therefore, for any $T > 0$,

$$\mathbb{P}(N^{\dagger}_{-s}(0,T) > 0) \leq \int_0^T \sum_{n=1}^{\infty} \mathbb{E}[\lambda^{\dagger,n}_{-s}(t)]dt \to 0, \tag{2.13}$$

as $s \to \infty$. Hence, we proved the weak asymptotic stationarity of N.

(ii) Since $\int_0^{\infty} t\gamma_n(t)dt < \infty$ for any $n \in \mathbb{N}$,

$$\begin{aligned}
\int_0^{\infty} (\gamma_n * \gamma_{n-1} * \cdots \gamma_{j+2} * H_{j+1})(t+s)dt &\leq \int_0^{\infty} (\gamma_n * \gamma_{n-1} * \cdots \gamma_{j+2} * H_{j+1})(t)dt \\
&= \|\gamma_n\|_{L^1} \cdots \|\gamma_{j+2}\|_{L^1} \|H_{j+1}\|_{L^1} < \infty,
\end{aligned} \tag{2.14}$$

since $\|H_{j+1}\|_{L^1} = \int_0^{\infty}\int_t^{\infty} \gamma_{j+1}(s)dsdt = \int_0^{\infty} t\gamma_{j+1}(t)dt < \infty$. Together with Assumption 1 (ii) and the proofs in part (i), we get

$$\int_0^{\infty} \sum_{n=1}^{\infty} \mathbb{E}[\lambda^{\dagger,n}_{-s}(t)]dt \leq \overline{\gamma}_0 \sum_{n=1}^{\infty} n\rho^{n-1}\eta < \infty, \tag{2.15}$$

and therefore

$$\mathbb{P}(N^{\dagger}_{-s}(0,\infty) > 0) \leq \int_0^{\infty} \sum_{n=1}^{\infty} \mathbb{E}[\lambda^{\dagger,n}_{-s}(t)]dt \to 0, \tag{2.16}$$

as $s \to \infty$. Hence, we proved the strong asymptotic stationarity of N. □

Remark 2. *In Theorem 1, we assumed that $\gamma_0(t) \equiv \overline{\gamma}_0$ being a constant. It will be interesting to extend the convergence to equilibrium results in Theorem 1 under a weaker assumption.*

3. Connections to Multivariate Hawkes Processes

In this section, we will show that the Hawkes process with different exciting functions is related to the multivariate Hawkes process, see e.g. Hawkes [18], Liniger [22], Bacry et al. [2]. A multivariate Hawkes process is multidimensional point process $(N_1(t),\dots,N_d(t))$ such that for any $1\leq i\leq d$, $N_i(t)$ is a simple point process with intensity

$$\lambda_i(t):=\nu_i+\sum_{j=1}^{d}\int_0^t\phi_{ij}(t-s)N_j(ds), \tag{3.1}$$

where ν_i are non-negative constants and $\phi_{ij}(t)$ are non-negative real-valued functions, and $\|\phi_{ij}\|_{L^1}<\infty$. If the spectral radius of the matrix $(\|\phi_{ij}\|_{L^1})_{1\leq i,j\leq d}$ is less than 1, then, we have the law of large numbers, see e.g. Bacry et al. [2]

$$\frac{1}{t}(N_1(t),\dots,N_d(t))^t\rightarrow(I-\Phi)^{-1}\nu, \tag{3.2}$$

as $t\rightarrow\infty$ where $\nu=(\nu_1,\dots,\nu_d)^t$ and $\Phi=(\|\phi_{ij}\|_{L^1})_{1\leq i,j\leq d}$.

Let us consider a special case of the Hawkes process with exciting functions $(\gamma_n)_{n\in\mathbb{N}\cup\{0\}}$ by letting $\gamma_0(t)\equiv\overline{\gamma}_0$, $\gamma_n(t)=h(t)$ if $n\in\mathbb{N}$ is odd and $\gamma_n(t)=g(t)$ if $n\in\mathbb{N}$ is even. We can consider two mutually exciting processes N^{even} and N^{odd} defined as

$$N^{\text{even}}:=\sum_{n=0}^{\infty}N^{2n}\qquad\text{and}\qquad N^{\text{odd}}:=\sum_{n=0}^{\infty}N^{2n+1}. \tag{3.3}$$

N^{even} and N^{odd} are mutually exciting since N^n is generated based on N^{n-1} and a jump in N^{2n} will lead to more jumps for N^{2n+1} and a jump in N^{2n+1} will on the other hand contribute to more jumps for N^{2n+2}. By the law of large numbers result due to Fierro et al. [15],

$$\frac{N_t}{t}=\frac{N_t^{\text{even}}}{t}+\frac{N_t^{\text{odd}}}{t}\rightarrow m \tag{3.4}$$

a.s. as $t\rightarrow\infty$, where

$$\begin{aligned}m&=\overline{\gamma}_0\sum_{n=1}^{\infty}\prod_{i=1}^{n}\|\gamma_i\|_{L^1}\\&=\overline{\gamma}_0\left(\|h\|_{L^1}+\|h\|_{L^1}\|g\|_{L^1}+\|h\|_{L^1}\|g\|_{L^1}+\cdots\right)\\&=\frac{1+\|h\|_{L^1}}{1-\|h\|_{L^1}\|g\|_{L^1}}.\end{aligned} \tag{3.5}$$

Now, let us point out the connections to the multivariate Hawkes process. The intensity of N^{even} is given by

$$\begin{aligned}\lambda_t^{\text{even}}&=\sum_{n=0}^{\infty}\lambda_t^{2n}\\&=\overline{\gamma}_0+\sum_{n=1}^{\infty}\int_0^t h(t-s)N^{2n-1}(ds)\\&=\overline{\gamma}_0+\int_0^t h(t-s)N^{\text{odd}}(ds).\end{aligned} \tag{3.6}$$

Similarly, the intensity of N^{odd} is given by

$$\lambda_t^{\text{odd}} = \int_0^t g(t-s)N^{\text{even}}(ds). \tag{3.7}$$

Therefore, $(N_t^{\text{even}}, N_t^{\text{odd}})$ is a bivariate Hawkes process with

$$\nu = \begin{pmatrix} \overline{\gamma}_0 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \Phi = \begin{pmatrix} 0 & \|h\|_{L^1} \\ \|g\|_{L^1} & 0 \end{pmatrix}. \tag{3.8}$$

Thus, by (3.2),

$$\frac{1}{t}\begin{pmatrix} N_t^{\text{even}} \\ N_t^{\text{odd}} \end{pmatrix} \to \begin{pmatrix} 1 & -\|h\|_{L^1} \\ -\|g\|_{L^1} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \overline{\gamma}_0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\overline{\gamma}_0}{1-\|h\|_{L^1}\|g\|_{L^1}} \\ \frac{\overline{\gamma}_0\|h\|_{L^1}}{1-\|h\|_{L^1}\|g\|_{L^1}} \end{pmatrix}, \tag{3.9}$$

as $t \to \infty$, which is consistent with (3.5).

Indeed, we can work in a more generating setting. Let $(A_i)_{i=1}^d$ be a partition of $\mathbb{N} \cup \{0\}$, i.e. $A_i \cap A_j = \emptyset$ for any $i \neq j$ and $\cup_{i=1}^d A_i = \mathbb{N} \cup \{0\}$. Assume that $\gamma_0 \equiv \overline{\gamma}_0$ and $(\gamma_n)_{n\in\mathbb{N}}$ may not be homogeneous. Define

$$N_1 = \sum_{n\in A_1} N^n, \qquad N_2 = \sum_{n\in A_2} N^n, \qquad \cdots\cdots \qquad N_d = \sum_{n\in A_d} N^n. \tag{3.10}$$

Then, the d-dimensional process $(N_1, \ldots, N_d)$ has the mutually exciting property and it is more general than the classical multivariate Hawkes process. Since we proved convergence to equilibrium in Theorem 1, by ergodic theorem,

$$\frac{1}{t}(N_1, N_2, \ldots, N_d) \to \left(\sum_{n\in A_1} m_n, \sum_{n\in A_2} m_n, \ldots, \sum_{n\in A_d} m_n \right), \tag{3.11}$$

a.s. as $t \to \infty$, where m_n is defined in (1.6).

4. Moderate and Large Deviations

In this section, we are interested to study the moderate and large deviations for $\mathbb{P}(\frac{N_t}{t} \in \cdot)$. The large deviations for classical Hawkes processes have been well studied in the literature for both linear and nonlinear cases, see e.g. Bordenave and Torrisi [5], Zhu [33] and Zhu [34]. The moderate deviations for classical Hawkes processes have been studied for the linear case, see e.g. Zhu [29].

In the linear case, let us assume that

$$\lambda_t = \nu + \int_0^t h(t-s)N(ds), \tag{4.1}$$

where $\|h\|_{L^1} < 1$ and $\int_0^\infty th(t)dt < \infty$. Bordenave and Torrisi [5] proved a large deviation principle for $(\frac{N_t}{t} \in \cdot)$ with the rate function

$$I(x) = \begin{cases} x\log\left(\frac{x}{\nu + x\|h\|_{L^1}}\right) - x + x\|h\|_{L^1} + \nu & \text{if } x \in [0,\infty) \\ +\infty & \text{otherwise} \end{cases}. \tag{4.2}$$

Moreover, Karabash and Zhu [21] obtained a large deviation principle for the linear Hawkes process with random marks.

For nonlinear Hawkes processes, i.e. when $\lambda(\cdot)$ is nonlinear, Zhu [33] first considered the case that $h(\cdot)$ is exponential, i.e. when the Hawkes process is Markovian and obtained a large deviation principle for $(N_t/t \in \cdot)$ Then, Zhu [33] also proved the large deviation principle for the case when $h(\cdot)$ is a sum of exponentials and used that as an approximation to recover the result for the linear case proved in Bordenave and Torrisi [5] and also for a special class of general nonlinear Hawkes processes. For the most general $h(\cdot)$ and $\lambda(\cdot)$, Zhu [34] proved a process-level, i.e. level-3 large deviation principle for the Hawkes process and used contraction principle to obtain a large deviation principle for $(N_t/t \in \cdot)$.

The large deviations result for $(N_t/t \in \cdot)$ is helpful to study the ruin probabilities of a risk process when the claims arrivals follow a Hawkes process. Stabile and Torrisi [26] considered risk processes with non-stationary Hawkes claims arrivals and studied the asymptotic behavior of infinite and finite horizon ruin probabilities under light-tailed conditions on the claims. The corresponding result for heavy-tailed claims was obtained by Zhu [31].

Before we proceed, let us recall that a sequence of probability measures $(P_n)_{n\in\mathbb{N}}$ on a topological space X satisfies a large deviation principle with speed n and rate function $I : X \to \mathbb{R}$ if I is non-negative, lower semicontinuous and for any measurable set A,

$$-\inf_{x\in A^o} I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log P_n(A) \leq \limsup_{n\to\infty} \frac{1}{n} \log P_n(A) \leq -\inf_{x\in\overline{A}} I(x). \tag{4.3}$$

Here, A^o is the interior of A and $\overline{A}$ is its closure. We refer to Dembo and Zeitouni [11] or Varadhan [27] for general background of large deviations and the applications.

Theorem 3. *Under Assumption 1 and $N(-\infty, 0] = 0$, for any $\theta \in \mathbb{R}$, $\Gamma(\theta) := \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]$ exists and*

$$\Gamma(\theta) = \overline{\gamma}_0(e^{f_\infty(\theta)} - 1), \tag{4.4}$$

where $f_\infty(\theta) := \lim_{M\to\infty} f_M(\theta)$ exists on the extended real line and f_M is defined recursively as $f_n(\theta) = \theta + \|\gamma_{M+1-n}\|_{L^1}(e^{f_{n-1}(\theta)} - 1)$, $1 \leq n \leq M$ and $f_0(\theta) = \theta$.

Remark 4. *It is easy to compute that*

$$\frac{\partial f_M(\theta)}{\partial\theta} = 1 + \|\gamma_1\|_{L^1} e^{f_{M-1}(\theta)} \frac{\partial f_{M-1}(\theta)}{\partial\theta}. \tag{4.5}$$

By iterating and setting $\theta = 0$, we can show that

$$\left.\frac{\partial\Gamma(\theta)}{\partial\theta}\right|_{\theta=0} = \sum_{p=0}^{\infty} m_p = m. \tag{4.6}$$

This verifies the limit in the law of large numbers (1.5).

Remark 5. *In the case of classical linear Hawkes process, say $\lambda_t = \nu + \int_0^t h(t-s)N(ds)$, it is easy to see that $\overline{\gamma}_0 = \nu$ and $\gamma_n = h$ for any $n \in \mathbb{N}$. Thus $\Gamma(\theta) = \nu(f(\theta) - 1)$, where $f(\theta)$ satisfies $f(\theta) = e^{\theta + \|h\|_{L^1}(f(\theta)-1)}$. Then, it is easy to check that $I(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \Gamma(\theta)\}$ gives* (4.2). *More generally, for example, if we assume that $\gamma_n = h$ for odd $n \in \mathbb{N}$ and $\gamma_n = g$ for even $n \in \mathbb{N}$, then, $\Gamma(\theta) = \overline{\gamma}_0(f(\theta) - 1)$, where $f(\theta)$ satisfies*

$$f(\theta) = e^{\theta + \|h\|_{L^1}(e^{\theta + \|g\|_{L^1}(f(\theta)-1)} - 1)}. \tag{4.7}$$

Proof of Theorem 3. For any $M \in \mathbb{N}$, $\theta \in \mathbb{R}$, and continuous deterministic function $G(s)$, $0 \leq s \leq t$,

$$
\begin{aligned}
(4.8)\qquad &\mathbb{E}\left[e^{\int_0^t G(t-s)N^M(ds)+\theta\sum_{n=0}^{M-1}N_t^n}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{\int_0^t G(t-s)N^M(ds)}\big|N^0, N^1, \ldots, N^{M-1}\right]e^{\theta\sum_{n=0}^{M-1}N_t^n}\right] \\
&= \mathbb{E}\left[e^{\int_0^t (e^{G(t-s)}-1)\lambda_s^M ds}e^{\theta\sum_{n=0}^{M-1}N_t^n}\right] \\
&= \mathbb{E}\left[e^{\int_0^t (e^{G(t-s)}-1)\int_0^s \gamma_M(s-u)N^{M-1}(du)ds}e^{\theta\sum_{n=0}^{M-1}N_t^n}\right] \\
&= \mathbb{E}\left[e^{\int_0^t [\int_u^t (e^{G(t-s)}-1)\gamma_M(s-u)ds]N^{M-1}(du)}e^{\theta\sum_{n=0}^{M-1}N_t^n}\right] \\
&= \mathbb{E}\left[e^{\int_0^t [\int_0^{t-u} (e^{G(t-u-s)}-1)\gamma_M(s)ds]N^{M-1}(du)}e^{\theta\sum_{n=0}^{M-1}N_t^n}\right]
\end{aligned}
$$

Therefore, we have for any $M \in \mathbb{N}$ and $\theta \in \mathbb{R}$,

$$
\mathbb{E}\left[e^{\theta\sum_{n=0}^{M}N_t^n}\right] = e^{\int_0^t (e^{F_M(t-s)}-1)\gamma_0(s)ds}, \tag{4.9}
$$

where $F_M(t) = G_0(t, \theta, M)$ and

$$
G_n(t, \theta, M) = \theta + \int_0^t (e^{G_{n+1}(t-s,\theta,M)} - 1)\gamma_n(s)ds, \qquad 0 \leq n \leq M-1, \tag{4.10}
$$

and $G_M(t, \theta, M) = \theta$.

It is easy to see that for any given $M \in \mathbb{N}$,

$$
\lim_{t\to\infty} F_M(t) = f_M(\theta), \tag{4.11}
$$

where $f_n(\theta) = \theta + \|\gamma_{M+1-n}\|_{L^1}(e^{f_{n-1}(\theta)} - 1)$, $1 \leq n \leq M$.

Since for $\theta \geq 0$, $e^{\theta\sum_{n=0}^{M}N_t^n}$ is increasing in M and for $\theta < 0$, it is decreasing in M, by monotone convergence theorem,

$$
\mathbb{E}[e^{\theta N_t}] = \lim_{M\to\infty}\mathbb{E}\left[e^{\theta\sum_{n=0}^{M}N_t^n}\right] = e^{\int_0^t (e^{\lim_{M\to\infty}F_M(t-s)}-1)\gamma_0(s)ds}. \tag{4.12}
$$

Since for $\theta \geq 0$, $F_M(t)$ is increasing in both M and t and for $\theta < 0$, $F_M(t)$ is decreasing in both M and t, we have

$$
\lim_{t\to\infty}\lim_{M\to\infty} F_M(t) = \lim_{M\to\infty}\lim_{t\to\infty} F_M(t) = \lim_{M\to\infty} f_M, \tag{4.13}
$$

and for any $\theta < 0$, $f_M(\theta)$ is decreasing in M and $f_\infty(\theta) := \lim_{M\to\infty} f_M(\theta)$ exists. For any $\theta \geq 0$, $f_M(\theta)$ is increasing in M and the limit $f_\infty(\theta) := \lim_{M\to\infty} f_M(\theta)$ exists on extended positive real line $[0, \infty]$. Hence, we conclude that

$$
\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta N_t}] = \overline{\gamma}_0(e^{f_\infty(\theta)} - 1) \tag{4.14}
$$

exists on the extended real line. □

Theorem 6. *Under Assumption 1 and $N(-\infty, 0] = 0$, $\mathbb{P}(N_t/t \in \cdot)$ satisfies a large deviation principle with rate function*

$$
I(x) := \sup_{\theta\in\mathbb{R}}\{\theta x - \Gamma(\theta)\}. \tag{4.15}
$$

Proof. Because we already had Theorem 3, we can apply Gärtner-Ellis theorem to obtain the large deviation principle if we can check the essential smoothness condition.

Let us defined the set

$$\mathcal{D}_\Gamma := \{\theta : \Gamma(\theta) < \infty\}. \tag{4.16}$$

Recall that we assumed $\rho := \sup_{n\in\mathbb{N}} \int_0^\infty \gamma_n(t)dt < 1$. For a classical linear Hawkes process with immigration rate ν and exciting function $h(t)$ and $\|h\|_{L^1} < 1$. The limit $\lim_{t\to\infty} \frac{1}{t}\log\mathbb{E}[e^{\theta N_t}]$ exists and is finite for any $\theta \leq \|h\|_{L^1} - 1 - \log\|h\|_{L^1}$. By comparing with the classical Hawkes process, there exists some constant $\theta_c \geq \rho - 1 - \log\rho > 0$ so that for any $\theta \leq \theta_c$, $\Gamma(\theta) < \infty$. More precisely, let us define $\theta_c := \sup\{\theta : \Gamma(\theta) < \infty\}$. Hence, we showed that the interior of $\mathcal{D}_\Gamma$ contains a nonempty neighborhood of the origin. Next, we need to show that for any $\theta < \theta_c$, $\Gamma(\theta)$ is differentiable at θ. By the definition of $f_M(\theta)$, for each $M \in \mathbb{N}$, $f_M(\theta)$ is differentiable on $(-\infty, \theta_c)$. Let $B_\delta(\theta)$ be the closure of an open ball of radius $\delta > 0$ centered at θ so that the closure of the ball is inside $(-\infty, \theta_c)$. We know that $f_M(\theta)$ converges to $f_\infty(\theta)$ at least pointwisely. For fixed M, by the definition $f_M(\theta)$ is smooth and hence differentiable. On the closure of $B_\delta(\theta)$, which is a compact set, $f'_M(\theta)$ is increasing in θ for any $\theta < \theta_c$ since we can show that $\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta\sum_{n=0}^M N_t^n}]$ is convex in θ. For a sequence of increasing functions that converge pointwisely, we have the uniform convergence. Therefore, f'_M converges uniformly. Together with the fact that $f_M \to f_\infty$ pointwisely, we conclude that $f_\infty(\theta)$ is differentiable for $\theta < \theta_c$ and $f'_\infty(\theta) = \lim_{M\to\infty} f'_M(\theta)$. For any $0 < \theta < \theta_c$, $f_M(\theta)$ is increasing in θ for any $M \in \mathbb{N}$. Thus

$$\frac{\partial f_M(\theta)}{\partial\theta} = 1 + \|\gamma_1\|_{L^1} e^{f_{M-1}(\theta)} \frac{\partial f_{M-1}(\theta)}{\partial\theta} \geq 1, \tag{4.17}$$

and thus

$$\frac{\partial f_M(\theta)}{\partial\theta} \geq 1 + \|\gamma_1\|_{L^1} e^{f_{M-1}(\theta)} \to +\infty, \tag{4.18}$$

as $\theta \uparrow \theta_c$. Thus we proved steepness. The proof is complete. $\square$

Let $C_1, C_2, \ldots$ be a sequence of real-valued i.i.d. random variables with finite mean $\mathbb{E}[C_1]$ and variance $\mathrm{Var}[C_1]$, independent of the point process N_t. Fierro et al. [15] showed that

$$\frac{\sum_{i=1}^{N_t} C_i - \mathbb{E}[C_1]\mathbb{E}[N_t]}{\sqrt{t}} \to N\left(0, m\mathrm{Var}[C_1] + \mathbb{E}[C_1]\sigma^2\right), \tag{4.19}$$

in distribution as $t \to \infty$.

We have the following result.

Theorem 7. *Assume $N(-\infty, 0] = 0$ and Assumption 1. Further assume that $\mathbb{E}[e^{\theta C_1}] < \infty$ for $\theta \in (-\gamma, \gamma)$ for some $\gamma > 0$. Then, the limit $\Gamma_C(\theta) := \lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta\sum_{i=1}^{N_t} C_i}]$ exists and indeed $\Gamma_C(\theta) = \Gamma(\log\mathbb{E}[e^{\theta C_1}])$, where $\Gamma(\cdot)$ is defined in Theorem 3. Moreover, $\mathbb{P}(\frac{1}{t}\sum_{i=1}^{N_t} C_i \in \cdot)$ satisfies a large deviation principle with rate function*

$$I_C = \sup_{\theta\in\mathbb{R}}\{\theta x - \Gamma_C(\theta)\}. \tag{4.20}$$

Proof. For any $\theta\in\mathbb{R}$ so that $\mathbb{E}\left[e^{\theta\sum_{i=1}^{N_t}C_i}\right]<\infty$, we have

$$\begin{aligned}\mathbb{E}\left[e^{\theta\sum_{i=1}^{N_t}C_i}\right]&=\mathbb{E}\left[\mathbb{E}\left[e^{\theta\sum_{i=1}^{N_t}C_i}|N_t\right]\right]\\&=\sum_{k=0}^{\infty}\mathbb{E}[e^{\theta\sum_{i=1}^{k}C_i}]\mathbb{P}(N_t=k)\\&=\sum_{k=0}^{\infty}e^{k\log\mathbb{E}[e^{\theta C_1}]}\mathbb{P}(N_t=k)\\&=\mathbb{E}\left[e^{\log\mathbb{E}[e^{\theta C_1}]N_t}\right].\end{aligned}\tag{4.21}$$

Hence, by Theorem 3, we get $\Gamma_C(\theta):=\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta\sum_{i=1}^{N_t}C_i}]=\Gamma(\log\mathbb{E}[e^{\theta C_1}])$. Following the proof of Theorem 6, we conclude that $\mathbb{P}(\frac{1}{t}\sum_{i=1}^{N_t}C_i\in\cdot)$ satisfies a large deviation principle with rate function $I_C(x)$ given by (4.20). $\square$

Let $X_1,\ldots,X_n$ be a sequence of real-valued i.i.d. random variables with mean 0 and variance σ^2. Assume that $\mathbb{E}[e^{\theta X_1}]<\infty$ for θ in some ball around the origin. For any $\sqrt{n}\ll a_n\ll n$, a moderate deviation principle says that for any Borel set A,

$$\begin{aligned}-\inf_{x\in A^o}\frac{x^2}{2\sigma^2}&\leq\liminf_{n\to\infty}\frac{n}{a_n^2}\log\mathbb{P}\left(\frac{1}{a_n}\sum_{i=1}^{n}X_i\in A\right)\\&\leq\limsup_{n\to\infty}\frac{n}{a_n^2}\log\mathbb{P}\left(\frac{1}{a_n}\sum_{i=1}^{n}X_i\in A\right)\leq-\inf_{x\in A^o}\frac{x^2}{2\sigma^2}.\end{aligned}\tag{4.22}$$

In other words, $\mathbb{P}(\frac{1}{a_n}\sum_{i=1}^{n}X_i\in\cdot)$ satisfies a large deviation principle with the speed $\frac{a_n^2}{n}$. The above classical result can be found for example in Dembo and Zeitouni [11]. Moderate deviation principle fills in the gap between central limit theorem and large deviation principle.

The moderate deviation principle for classical linear Hawkes process has been studied in Zhu [29]. For the remaining of this section, let us prove the moderate deviation principle for the Hawkes process with different exciting functions.

Theorem 8. *Assume that Assumptions 1 and 2 hold. For any Borel set A and time sequence $a(t)$ such that $\sqrt{t}\ll a(t)\ll t$, we have the following moderate deviation principle.*

$$\begin{aligned}-\inf_{x\in A^o}J(x)&\leq\liminf_{t\to\infty}\frac{t}{a(t)^2}\log\mathbb{P}\left(\frac{N_t-mt}{a(t)}\in A\right)\\&\leq\limsup_{t\to\infty}\frac{t}{a(t)^2}\log\mathbb{P}\left(\frac{N_t-mt}{a(t)}\in A\right)\leq-\inf_{x\in\overline{A}}J(x),\end{aligned}\tag{4.23}$$

where $J(x)=\frac{x^2}{2\sigma^2}$ and $\sigma^2=\sum_{j=0}^{\infty}\left(1+\sum_{p=1}^{\infty}\prod_{i=j+1}^{p+j}\int_0^\infty\gamma_i(u)du\right)^2m_j$.

Before we proceed to the proof of Theorem 8, let us prove a series of lemmas.

Lemma 9. *Consider the equation*

$$x=\theta+(e^x-1)\rho.\tag{4.24}$$

The equation has two distinct solutions if $\theta < \rho - 1 - \log\rho$ and has one solution if $\theta = \rho - 1 - \log\rho$. Let $x(\theta)$ be the minimal solution of (4.24) if the solutions exist. Then, $x(\theta) \le 0$ if $\theta \le 0$ and $x(\theta) \ge 0$ if $\theta \ge 0$.

Proof. Let $F(x) := x - \theta - (e^x - 1)\rho$. It is easy to compute that $F'(x) = 1 - \rho e^x$ and $F''(x) = -\rho e^x$. Thus, $F(x)$ is strictly concave and its maximum is achieved at $x = \log(1/\rho)$. Therefore, the equation (4.24) has no solutions if $F(\log(1/\rho)) < 0$, has one solution if $F(\log(1/\rho)) = 0$ and has two solutions if $F(\log(1/\rho)) > 0$. It is easy to check that $F(\log(1/\rho)) = -\theta - \log\rho + \rho - 1$. Now assume that $\theta \le \rho - 1 - \log\rho$ so that (4.24) has solutions. Observe that if $\theta \ge 0$, then $F(0) = -\theta < 0$ and $F(\log(1/\rho)) \ge 0$, where $\log(1/\rho) > 0$, thus $x(\theta) \ge 0$. Similarly $x(\theta) \le 0$ when $\theta \le 0$. $\square$

Lemma 10. *Given $0 \le \theta \le \rho - 1 - \log\rho$, if $f(t,\theta) \le x(\theta)$ for any $t \ge 0$, then $F_n(t,\theta) \le x(\theta)$ uniformly in $t \ge 0$ and $n \in \mathbb{N}$, where*

$$F_n(t,\theta) := e^{\theta + \int_0^t (e^{f(t-s,\theta)} - 1)\gamma_n(s)ds}, \qquad t \ge 0. \tag{4.25}$$

Similarly, given $\theta \le 0$, if $f(t,\theta) \ge x(\theta)$ for any $t \ge 0$, then $F_n(t,\theta) \ge x(\theta)$.

Proof. Let us first assume that $0 \le \theta \le \rho - 1 - \log\rho$. By the definition of $x(\theta)$ and the assumption $f(t,\theta) \le x(\theta)$ for any $t \ge 0$, it is easy to see that

$$\begin{aligned} F_n(t,\theta) &\le e^{\theta + \int_0^t (e^{x(\theta)} - 1)\gamma_n(s)ds} \\ &\le e^{\theta + (e^{x(\theta)} - 1)\int_0^t \gamma_n(s)ds} \\ &\le e^{\theta + (e^{x(\theta)} - 1)\|\gamma_n\|_{L^1}} \\ &\le e^{\theta + (e^{x(\theta)} - 1)\rho} \\ &= x(\theta), \end{aligned} \tag{4.26}$$

where we used the fact that $x(\theta) \ge 0$ for $\theta \ge 0$ in Lemma 9. Similarly, one can show that, given $\theta \le 0$, if $f(t,\theta) \ge x(\theta)$ for any $t \ge 0$, then $F_n(t,\theta) \ge x(\theta)$. $\square$

Lemma 11. *For any fixed θ, there is some $k_1 \ge \frac{1}{1-\rho}$ so that for any sufficiently large t,*

$$\left| G_n\left(s, \frac{a(t)}{t}\theta, M\right) \right| \le k_1 \frac{a(t)}{t}|\theta|, \tag{4.27}$$

uniformly for $1 \le n \le M$, $M \in \mathbb{N}$ and $s \ge 0$, where G_n was defined in (4.10).

Proof. Given $\theta \ge 0$, by Lemma 10 and (4.10), we have $G_n(t,\theta,M) \le x(\theta)$. Notice that in Lemma 9, $x(0) = 0$ and $x'(0) = \frac{1}{1-\rho} > 1$ since $\rho < 1$. Therefore, for $0 \le \theta \ll 1$, there exists some $k_1 \ge \frac{1}{1-\rho}$ so that $0 \le x(\theta) \le k_1\theta$. Therefore, $x\left(\frac{a(t)}{t}\theta\right) \le k_1 \frac{a(t)}{t}\theta$ for any sufficiently large t. Hence, for $\theta \ge 0$, for sufficiently large t, $G_n\left(s, \frac{a(t)}{t}\theta, M\right) \le k_1\frac{a(t)}{t}|\theta|$ uniformly for $1 \le n \le M$, $M \in \mathbb{N}$ and $s \ge 0$. Similarly, given $\theta \le 0$, by Lemma 10 and the discussions above, $G_n(t,\theta,M) \ge x(\theta) \ge k_1\theta$ for $\theta \le 0$ and $|\theta| \ll 1$. Hence, we proved the desired result. $\square$

Lemma 12. *Let us define*

$$C_n^1(s,M) := 1 + \int_0^s C_{n+1}^1(s-r,M)\gamma_n(r)dr, \tag{4.28}$$

and

$$C_n^2(s,M) := \int_0^s \left(C_{n+1}^2(s-r,M) + \frac{1}{2}[C_{n+1}^1(s-r,M)]^2 \right) \gamma_n(r)dr, \tag{4.29}$$

where $C_M^1(s,M) = 1$ *and* $C_M^2(s,M) = 0$, $s \geq 0$, $n \leq M$, *and* $M \in \mathbb{N}$. *Then, we have*

$$C_n^1(s,M) \leq \frac{1}{1-\rho} \qquad \text{and} \qquad C_n^2(s,M) \leq \frac{1}{(1-\rho)^3}. \tag{4.30}$$

Proof. Let us use induction on n. For $n = M$, $C_M^1(s,M) = 1 \leq \frac{1}{1-\rho}$ since $\rho < 1$. Now assume $C_{n+1}^1(s,M) \leq \frac{1}{1-\rho}$, we get

$$\begin{aligned} C_n^1(s,M) &\leq 1 + \int_0^s \frac{1}{1-\rho}\gamma_n(r)dr \\ &\leq 1 + \frac{\|\gamma_n\|_{L^1}}{1-\rho} \\ &\leq \frac{1}{1-\rho}. \end{aligned} \tag{4.31}$$

It is clear that $C_M^2(s,M) = 0 \leq \frac{1}{(1-\rho)^3}$. Now assume that $C_{n+1}^2(s,M) \leq \frac{1}{(1-\rho)^3}$ and apply the inequality $C_{n+1}^1(s,M) \leq \frac{1}{1-\rho}$ that we have just proved,

$$\begin{aligned} C_n^2(s,M) &\leq \int_0^s \left[\frac{1}{(1-\rho)^3} + \frac{1}{2}\frac{1}{(1-\rho)^2} \right] \gamma_n(r)dr \\ &\leq \rho \left[\frac{1}{(1-\rho)^3} + \frac{1}{2}\frac{1}{(1-\rho)^2} \right] \\ &\leq \frac{1}{(1-\rho)^3}. \end{aligned} \tag{4.32}$$

□

Lemma 13. *Given any fixed* $\theta \in \mathbb{R}$, *let* t *be sufficiently large so that* $k_1 \frac{a(t)}{t}|\theta| \leq \frac{1-\rho}{4}$. *Then, we have*

$$\left| G_n\left(s, \frac{a(t)}{t}\theta, M\right) - C_n^1(s,M)\frac{a(t)}{t}\theta - C_n^2(s,M)\left(\frac{a(t)}{t}\theta\right)^2 \right| \leq k_2 \left[\frac{a(t)}{t}|\theta|\right]^3, \tag{4.33}$$

where $C_n^1(s,M)$ *and* $C_n^2(s,M)$ *are defined in Lemma 12 and*

$$k_2 := \frac{4\left[\frac{\rho}{2(1-\rho)^3}\left[k_1 + \frac{1}{1-\rho}\right] + \rho k_1^3\right]}{(4-\rho)(1-\rho)}. \tag{4.34}$$

Proof. Let us prove by induction. For $n = M$, $G_M(s, \frac{a(t)}{t}\theta, M) = \frac{a(t)}{t}\theta$, $C_M^1(s,M) = 1$ and $C_M^2(s,M) = 0$, thus (4.33) holds. Assume (4.33) is true for $n+1$. Notice that

$$G_n\left(s, \frac{a(t)}{t}\theta, M\right) = \frac{a(t)}{t}\theta + \int_0^s \left[e^{G_{n+1}(s-r, \frac{a(t)}{t}\theta, M)} - 1\right]\gamma_n(r)dr. \tag{4.35}$$

Since $|e^x - 1 - x - \frac{x^2}{2}| \leq |x|^3$ for $|x| < 1$ and

$$\left| G_{n+1}\left(s, \frac{a(t)}{t}\theta, M\right) \right| \leq k_1 \frac{a(t)}{t}|\theta| \leq \frac{1-\rho}{4} < 1, \tag{4.36}$$

we have

$$\begin{aligned}(4.37)\qquad &\left|G_n\left(s,\frac{a(t)}{t}\theta,M\right)-\left[\frac{a(t)}{t}\theta+\int_0^s\left[G_{n+1}\left(s-r,\frac{a(t)}{t}\theta,M\right)\right.\right.\right.\\&\qquad\left.\left.\left.+\frac{1}{2}\left(G_{n+1}\left(s-r,\frac{a(t)}{t}\theta,M\right)\right)^2\right]\gamma_n(r)dr\right]\right|\\&\leq\int_0^s\gamma_n(r)\left[k_1\frac{a(t)}{t}|\theta|\right]^3dr\\&\leq\rho k_1^3\left(\frac{a(t)|\theta|}{t}\right)^3.\end{aligned}$$

In addition, by using the induction,

$$\begin{aligned}(4.38)\qquad L:=&\left|\left[G_{n+1}\left(s-r,\frac{a(t)}{t}\theta,M\right)\right]^2-C_n^1(s-r,M)^2\left(\frac{a(t)}{t}\theta\right)^2\right|\\=&\left|G_{n+1}\left(s-r,\frac{a(t)}{t}\theta,M\right)-C_n^1(s-r,M)\left(\frac{a(t)}{t}\theta\right)\right|\\&\cdot\left|G_{n+1}\left(s-r,\frac{a(t)}{t}\theta,M\right)+C_n^1(s-r,M)\left(\frac{a(t)}{t}\theta\right)\right|\\\leq&\left[C_n^2(s-r,M)\left(\frac{a(t)}{t}\theta\right)^2+k_2\left(\frac{a(t)}{t}|\theta|\right)^3\right]\\&\cdot\left[\left|G_{n+1}\left(s-r,\frac{a(t)}{t}\theta,M\right)\right|+C_n^1(s-r,M)\frac{a(t)}{t}|\theta|\right].\end{aligned}$$

Using the bounds in Lemma 12, we obtain

$$(4.39)\qquad L\leq\left[\frac{1}{(1-\rho)^3}\left(\frac{a(t)}{t}\theta\right)^2+k_2\left(\frac{a(t)}{t}|\theta|\right)^3\right]\left[k_1\frac{a(t)}{t}|\theta|+\frac{1}{1-\rho}\frac{a(t)}{t}|\theta|\right].$$

Since $k_1\frac{a(t)}{t}|\theta|\leq\frac{1-\rho}{4}$ and $k_1\geq\frac{1}{1-\rho}$, we get

(4.40)

$$\begin{aligned}L=&\left(\frac{a(t)}{t}|\theta|\right)^3\frac{1}{(1-\rho)^3}\left[k_1+\frac{1}{1-\rho}\right]+k_2\left(\frac{a(t)}{t}|\theta|\right)^3\left[k_1\frac{a(t)}{t}|\theta|+\frac{1}{1-\rho}\frac{a(t)}{t}|\theta|\right]\\\leq&\left(\frac{a(t)}{t}|\theta|\right)^3\frac{1}{(1-\rho)^3}\left[k_1+\frac{1}{1-\rho}\right]+k_2\left(\frac{a(t)}{t}|\theta|\right)^3\left[\frac{1-\rho}{4}+\frac{1}{1-\rho}\frac{1-\rho}{4k_1}\right]\\\leq&\left(\frac{a(t)}{t}|\theta|\right)^3\left[\frac{1}{(1-\rho)^3}\left[k_1+\frac{1}{1-\rho}\right]+k_2\frac{1-\rho}{2}\right].\end{aligned}$$

Let us put (4.37) and (4.40) together and define

$$\begin{aligned}(4.41)\quad I:=&G_n\left(s,\frac{a(t)}{t}\theta,M\right)-\frac{a(t)}{t}\theta\left[1+\int_0^sC_{n+1}^1\left(s-r,\frac{a(t)}{t}\theta,M\right)\gamma_n(r)dr\right]\\&-\left[\frac{a(t)}{t}\theta\right]^2\left[\int_0^s\left[C_{n+1}^2\left(s-r,\frac{a(t)}{t}\theta,M\right)\right.\right.\\&\qquad\left.\left.+\frac{1}{2}\left[C_{n+1}^1\left(s-r,\frac{a(t)}{t}\theta,M\right)\right]^2\right]\gamma_n(r)dr\right].\end{aligned}$$

Therefore,

$$
\begin{aligned}
|I| &\leq \int_0^s k_2 \left[\frac{a(t)}{t}|\theta|\right]^3 \gamma_n(r)dr \\
&\quad + \frac{1}{2}\int_0^s \left[\frac{1}{(1-\rho)^3}\left[k_1 + \frac{1}{1-\rho}\right] + k_2\frac{1-\rho}{2}\right]\left(\frac{a(t)}{t}|\theta|\right)^3 \gamma_n(r)dr \\
&\qquad + \rho k_1^3 \left(\frac{a(t)}{t}|\theta|\right)^3 \\
&\leq \left(\frac{a(t)}{t}|\theta|\right)^3 \left[\rho k_2 + \frac{\rho}{2}\left[\frac{1}{(1-\rho)^3}\left[k_1 + \frac{1}{1-\rho}\right] + k_2\frac{1-\rho}{2}\right] + \rho k_1^3\right] \\
&\leq \left(\frac{a(t)}{t}|\theta|\right)^3 \left[\left[\frac{\rho(1-\rho)}{4} + \rho\right] k_2 + \frac{\rho}{2(1-\rho)^3}\left[k_1 + \frac{1}{1-\rho}\right] + \rho k_1^3\right] \\
&= k_2 \left(\frac{a(t)}{t}|\theta|\right)^3.
\end{aligned}
\tag{4.42}
$$

This proves the desired result. $\square$

We observe that k_1 and k_2 only depend on ρ, so our bound in Lemma 13 is uniform in $n \leq M$, $M \in \mathbb{N}$ and $s \geq 0$. Now, let us go back to the proof of Theorem 8.

Proof of Theorem 8. We are interested to prove that the limit

$$
\lim_{t\rightarrow\infty} \frac{t}{a(t)^2} \log \mathbb{E}\left[e^{\frac{a(t)}{t}\theta(N_t - mt)}\right]
\tag{4.43}
$$

exists and it can be computed explicitly.

Notice that

$$
\begin{aligned}
&\lim_{t\rightarrow\infty} \frac{t}{a(t)^2} \log \mathbb{E}\left[e^{\frac{a(t)}{t}\theta(N_t - mt)}\right] \\
&= \lim_{t\rightarrow\infty}\lim_{M\rightarrow\infty} \frac{t}{a(t)^2}\left[\int_0^t (e^{F_M(t-s)} - 1)\gamma_0(s)ds - ma(t)\theta\right] \\
&= \lim_{t\rightarrow\infty}\lim_{M\rightarrow\infty} \frac{t}{a(t)^2}\left[\int_0^t \left(e^{C_1^1(t-r,M)\frac{a(t)}{t}\theta + C_1^2(t-r,M)(\frac{a(t)}{t}\theta)^2 + O(\frac{a(t)}{t}|\theta|)^3} - 1\right)\gamma_0(r)dr \right. \\
&\qquad\qquad \left. - ma(t)\theta\right] \\
&= \lim_{t\rightarrow\infty}\lim_{M\rightarrow\infty} \frac{t}{a(t)^2}\left[\int_0^t C_1^1(t-r,M)\gamma_0(r)dr - mt\right]\frac{a(t)}{t}\theta \\
&\quad + \lim_{t\rightarrow\infty}\lim_{M\rightarrow\infty} \frac{t}{a(t)^2}\left(\int_0^t \left[C_1^2(t-r,M) + \frac{1}{2}C_1^1(t-r,M)^2\right]\gamma_0(r)dr\right)\left(\frac{a(t)}{t}\theta\right)^2 \\
&\quad + \lim_{t\rightarrow\infty} O\left(\left(\frac{a(t)}{t}|\theta|\right)^3\right)\int_0^t \gamma_0(r)dr \frac{t}{a(t)^2} \\
&=: I_1 + I_2 + I_3,
\end{aligned}
\tag{4.44}
$$

where we used the Taylor expansion of e^x for $x = o(1)$.

The next step is to carry out careful analysis on I_1, I_2, I_3, the last three terms in (4.44).

For the first term I_1 in (4.44), it is easy to see that

$$I_1 = \lim_{t\to\infty}\lim_{M\to\infty}\frac{t}{a(t)}\theta\left[\frac{1}{t}\int_0^t C_1^1(t-r,M)\gamma_0(r)dr - m\right] \tag{4.45}$$
$$= \lim_{t\to\infty}\frac{t}{a(t)}\theta\left[\frac{1}{t}\int_0^t C_1^1(t-r,\infty)\gamma_0(r)dr - m\right],$$

where $C_1^1(t,\infty)$ can be computed via iteration in (4.28) as

$$C_1^1(t,\infty) = 1 + \sum_{n=1}^{\infty}(\gamma_n * \cdots * \gamma_1 * 1)(t). \tag{4.46}$$

Let us recall that $m = \sum_{n=0}^{\infty} m_n$, where $m_0 = \overline{\gamma}_0$ and $m_n = \overline{\gamma}_0\int_0^\infty(\gamma_n * \cdots * \gamma_1)(t)dt$. Therefore, we have

$$|I_1| \leq |\theta|\limsup_{t\to\infty}\frac{t}{a(t)}\left|\frac{1}{t}\int_0^t\gamma_0(s)ds - \overline{\gamma}_0\right|\int_0^\infty h(s)ds \tag{4.47}$$
$$+ |\theta|\limsup_{t\to\infty}\overline{\gamma}_0\frac{t}{a(t)}\int_t^\infty h(s)ds$$
$$+ |\theta|\limsup_{t\to\infty}\int_0^\infty h(s)\left|\frac{1}{a(t)}\int_{t-s}^t\gamma_0(u)du\right|ds,$$

where $h(t) := \sum_{n=1}^{\infty}(\gamma_n * \cdots * \gamma_1)(t)$ and thus $\int_0^\infty h(t)dt = \frac{\rho}{1-\rho} < \infty$. The first two terms in (4.47) are zero due to Assumption 2 and the third term in (4.47) is zero by dominated convergence theorem.

Next, let us consider the second term I_2 in (4.44). In Lemma 12, we obtained a uniform bound on $C_n^1(s,M)$ and $C_n^2(s,M)$. Hence, we have

$$I_2 = \lim_{t\to\infty}\lim_{M\to\infty}\frac{\theta^2}{t}\int_0^t\left[C_1^2(t-r,M) + \frac{1}{2}C_1^1(t-r,M)^2\right]\gamma_0(r)dr \tag{4.48}$$
$$= \theta^2\lim_{t\to\infty}\frac{1}{t}\int_0^\infty\left[C_1^2(t-r,\infty) + \frac{1}{2}C_1^1(t-r,\infty)^2\right]\gamma_0(r)dr$$
$$= \theta^2\lim_{t\to\infty}\left[C_1^2(t,\infty) + \frac{1}{2}C_1^1(t,\infty)^2\right]\overline{\gamma}_0$$
$$= \frac{1}{2}\sigma^2\theta^2,$$

since

$$\overline{\gamma}_0\left[C_1^2(\infty,\infty) + \frac{1}{2}C_1^1(\infty,\infty)^2\right] = \frac{1}{2}\sum_{j=0}^{\infty}\left(1 + \sum_{p=1}^{\infty}\prod_{i=j+1}^{p+j}\int_0^\infty\gamma_i(u)du\right)^2 m_j = \frac{1}{2}\sigma^2. \tag{4.49}$$

Let us verify (4.49). First, fix M. We will let M go to infinity later. We can do that since all of our estimates for convergence in M are uniform in t, so we can interchange the two limits. Let us define $\Gamma(i,j,t) = (\gamma_i * \gamma_{i+1} * \cdots * \gamma_j)(t)$ for $i \leq j$. Another look at (4.28) reveals that

$$C_n^1(s,M) = 1 + [C_{n+1}^1(\cdot,M) * \gamma_n](s). \tag{4.50}$$

Since $C^1_M(s,M)=1$, we get

$$C^1_n(s,M)=1+\sum_{k=n}^{M-1}(\Gamma(n,k)*1)(s), \tag{4.51}$$

for $n\leq M$. Similarly,

$$C^2_n(s,M)=[C^2_{n+1}(\cdot,M)*\gamma_n](s)+\frac{1}{2}[[C^1_{n+1}(\cdot,M)]^2*\gamma_n](s). \tag{4.52}$$

Therefore,

$$C^2_n(s,M)=\frac{1}{2}\sum_{k=n+1}^{M-1}[(C^1_k)^2*\Gamma(n,k-1)](s). \tag{4.53}$$

Let us define

$$m(i,j):=\prod_{k=i+1}^{j}\int_0^\infty\gamma_k(u)du, \tag{4.54}$$

for $j>i$ and $m(i,i):=1$. Hence, $m(0,j)=m_j$. Moreover, let us define

$$\begin{aligned} I_1(M)&:=\left[C^2_1(\infty,M)+\frac{1}{2}[C^1_1(\infty,M)]^2\right]\overline{\gamma}_0 \\ &=\frac{\overline{\gamma}_0}{2}\sum_{n=1}^{M-1}\left[1+\sum_{k=n}^{M-1}\int_0^\infty\Gamma(n,k,s)ds\right]^2 m(0,n-1), \end{aligned} \tag{4.55}$$

and

$$\begin{aligned} I_2(M,t)&:=\frac{1}{t}\left[\left(C^2_1(\cdot,M)+\frac{1}{2}\left(C^1_1(\cdot,M)\right)^2\right)*\gamma_0\right](t) \\ &=\frac{1}{2t}\left[\sum_{n=2}^{M-1}(C^1_n)^2*\Gamma(1,n-1)*\gamma_0+(C^1_1)^2*\gamma_0\right](t) \end{aligned} \tag{4.56}$$

Now,

$$\begin{aligned} &I_1(M)-I_2(M,t) \\ &=\frac{1}{2}\sum_{n=1}^{M}\left[1+\sum_{k=n}^{M-1}(\Gamma(n,k,\cdot)*1)(\infty)\right]^2\left[\overline{\gamma}_0 m(0,n-1)-\frac{(\Gamma(1,n-1)*\gamma_0)(t)}{t}\right] \\ &\quad+\frac{1}{2t}\sum_{n=1}^{M-1}\left[\left[1+\sum_{k=n}^{M-1}(\Gamma(n,k,\cdot)*1)(\infty)\right]^2-\left[1+\sum_{k=n}^{M-1}(\Gamma(n,k,\cdot)*1)(\cdot)\right]^2\right] \\ &\qquad\qquad *\Gamma(1,n-1)*\gamma_0(t). \end{aligned} \tag{4.57}$$

We denote the first sum by L_1 and the second sum by L_2. Since $C_n^1(t,M) \leq \frac{1}{1-\rho}$ for all n and t by Lemma 12,

$$\begin{aligned}
L_1 &\leq \frac{1}{2(1-\rho)^2} \sum \overline{\gamma}_0 m(0,n-1) - \frac{(\Gamma(1,n-1)*\gamma_0)(t)}{t} \\
&\leq \frac{1}{2(1-\rho)^2} \left| \frac{1}{t}\int_0^t \gamma_0(s)ds - \overline{\gamma}_0 \right| \int_0^\infty h(s)ds \\
&\qquad + \overline{\gamma}_0 \int_t^\infty h(s)ds + \int_0^\infty h(s)\frac{1}{t}\int_{t-s}^t \gamma_0(u)duds.
\end{aligned} \tag{4.58}$$

The right hand side of the above equation goes to zero as $t \to \infty$ since $h(\cdot)$ is integrable. Next, let us bound L_2.

$$\begin{aligned}
L_2 &= \frac{1}{2t}\sum_{n=1}^{M-1} \left[\left[1 + \sum_{k=n}^{M-1} (\Gamma(n,k,\cdot)*1)(\infty) \right] - \left[1 + \sum_{k=n}^{M-1} (\Gamma(n,k,\cdot)*1)(\cdot) \right] \right] \\
&\quad \cdot \left[\left[1 + \sum_{k=n}^{M-1} (\Gamma(n,k,\cdot)*1)(\infty) \right] + \left[1 + \sum_{k=n}^{M-1} (\Gamma(n,k,\cdot)*1)(\cdot) \right] \right] \\
&\qquad\qquad * \Gamma(1,n-1) * \gamma_0(t) \\
&\leq \frac{1}{2t}\sum_{n=1}^{M-1} \frac{2}{1-\rho} \left[\left[1 + \sum_{k=n}^{M-1} (\Gamma(n,k,\cdot)*1)(\infty) \right] - \left[1 + \sum_{k=n}^{M-1} (\Gamma(n,k,\cdot)*1)(\cdot) \right] \right] \\
&\qquad\qquad * \gamma_0(t) \cdot \rho^n \\
&\leq \frac{1}{2t}\sum_{n=1}^{M-1} \frac{2\rho^n}{1-\rho} \int_0^t \left[\int_{t-r}^\infty \Gamma(1,k,s)ds \right] \gamma_0(r)dr \\
&\leq \frac{1}{2t}\sum_{n=1}^{M-1} \frac{2\rho^n}{1-\rho} \left[\int_0^t \Gamma(1,k,s) \int_{t-r}^t \gamma_0(r)drds + \int_t^\infty \Gamma(1,k,s) \int_0^t \gamma_0(r)drds \right] \\
&\leq \sum_{n=1}^{M-1} \frac{2\rho^n}{1-\rho} \left[\int_t^\infty h(s)ds(\overline{\gamma}_0+1) + \int_0^\infty h(s)ds\frac{1}{t}\int_{t-r}^t \gamma_0(r)dr \right],
\end{aligned} \tag{4.59}$$

which goes to zero by dominated convergence theorem. The difference in M is given by

$$\begin{aligned}
&I_1(\infty) - I_1(M) \\
&= \frac{\overline{\gamma}_0}{2}\sum_{n=M}^{\infty} \left[1 + \sum_{k=n}^{M-1} (\Gamma(n,k,\cdot)*1)(\infty) \right]^2 m(0,n-1) \\
&\leq \frac{\overline{\gamma}_0}{2}\frac{1}{(1-\rho)^2}\sum_{n=M}^{\infty} \rho^n,
\end{aligned} \tag{4.60}$$

and for sufficiently large t (uniformly in M),

$$\begin{aligned}
(4.61)\qquad & I_2(\infty,t)-I_2(M,t)\\
&=\frac{1}{2t}\sum_{n=M}^{\infty}\left[1+\sum_{k=n}^{M-1}\int_0^{\cdot}\Gamma(n,k,s)ds\right]^2*\Gamma(1,n-1)*\gamma_0(t)\\
&\leq\frac{1}{2(1-\rho)^2}\sum_{n=M}^{\infty}\rho^n\frac{1}{t}\int_0^t\gamma_0(r)dr\\
&\leq\frac{\overline{\gamma}_0+1}{2(1-\rho)^2}\sum_{n=M}^{\infty}\rho^n.
\end{aligned}$$

Hence, we proved (4.49).

Finally, let us show that the third term I_3 in (4.44) is zero in the limit. For some universal constant $K>0$,

$$\begin{aligned}
(4.62)\qquad |I_3|&\leq\limsup_{t\rightarrow\infty}K\left(\frac{a(t)}{t}|\theta|\right)^3\left(\int_0^t\gamma_0(r)dr\right)\frac{t}{a(t)^2}\\
&=\limsup_{t\rightarrow\infty}K\frac{a(t)}{t}|\theta|^3\frac{\int_0^t\gamma_0(r)dr}{t}\\
&\leq\limsup_{t\rightarrow\infty}K\frac{a(t)}{t}|\theta|^3\overline{\gamma}_0\\
&=0.
\end{aligned}$$

Hence, we proved that

$$\lim_{t\rightarrow\infty}\frac{t}{a(t)^2}\log\mathbb{E}\left[e^{\frac{a(t)}{t}\theta(N_t-mt)}\right]=\frac{1}{2}\theta^2\sigma^2. \tag{4.63}$$

By Gätner-Ellis theorem, the proof is complete. □

Remark 14. *Following the same proof of Theorem 8, we can show that for any* $\theta\in\mathbb{R}$, $\lim_{t\rightarrow\infty}\mathbb{E}[e^{\frac{i\theta}{\sqrt{t}}(N_t-mt)}]=e^{-\frac{\theta^2}{2}\sigma^2}$. *In other words, our method gives an altenative proof to the central limit theorem that was obtained in Fierro et al.* [15].

Remark 15. *Indeed, one can also consider the moderate deviations in the presence of random marks, i.e. for a sequence of i.i.d. real-valued random variables* $C_1,C_2,\ldots$ *with mean* $\mathbb{E}[C_1]$ *and variance* $Var[C_1]$ *independent of* N_t, *we expect that for a sequence* $a(t)$ *so that* $\sqrt{t}\ll a(t)\ll t$, $\mathbb{P}(\frac{\sum_{i=1}^{N_t}Ci-\mathbb{E}[C_1]\mathbb{E}[N_t]}{a(t)}\in\cdot)$ *follows a large deviation principle with rate function* $J_C(x):=\frac{x^2}{2\sigma_C^2}$, *where* $\sigma_C^2:=m\,Var[C_1]+\mathbb{E}[C_1]\sigma^2$ *The proofs are similar to the proofs of moderate deviations for the unmarked case and we will not go into the details in this paper.*

5. Applications to Finance

5.1. **Microstructure Noise.** Let X_t stand for some asset price at time t. The signature plot can be defined for X_t over a time period $[0,T]$ at the time scale τ as

$$\hat{C}(\tau):=\frac{1}{T}\sum_{n=0}^{\lfloor T/\tau\rfloor}(X_{(n+1)\tau}-X_{n\tau})^2. \tag{5.1}$$

This is also known as the realized volatility. The microstructure noise effect is described by an increase of the realized volatility when the time scale τ decreases. This behavior is different from what one would expect if X_t is a Brownian motion, for which $\hat{C}(\tau)$ will be constant in τ as $T \to \infty$.

If X_t^1 and X_t^2 are the prices of two assets, we can define

$$\hat{\rho}(\tau) := \frac{\hat{C}_{12}(\tau)}{\sqrt{\hat{C}_1(\tau)\hat{C}_2(\tau)}}, \tag{5.2}$$

where

$$\hat{C}_{12}(\tau) := \frac{1}{T}\sum_{n=0}^{\lfloor T/\tau \rfloor}(X^1_{(n+1)\tau} - X^1_{n\tau})(X^2_{(n+1)\tau} - X^2_{n\tau}), \tag{5.3}$$

and $\hat{C}_1(\tau)$ and $\hat{C}_2(\tau)$ are defined similarly as in (5.1).

The Epps effect, named after Epps [13] describes the pheonomenon that the correlation coefficient $\hat{\rho}(\tau)$ increases in τ and it tends to zero as $\tau \to 0$.

Bacry et al. [3] studied the signature plot $\hat{C}(\tau)$ as in (5.1) for the price model, $X_t = N_1(t) - N_2(t)$, where (N_1, N_2) is a bivariate Hawkes process and they also studied correlation coefficient $\hat{\rho}(\tau)$ as in (5.2) for $X_t^1 = N_1(t) - N_2(t)$, $X_t^2 = N_3(t) - N_4(t)$, where (N_1, N_2, N_3, N_4) is a multivariate Hawkes process. They considered the case of long horizon, i.e. the large T limit and hence studied the macroscopic properties of a multivariate Hawkes process, see e.g. [2], [3]. The large T limit can correspond to a trading day realization of the price model. In [3], they considered for instance a realization of 20 hours (Figure 2 in [3]).

Following the ideas in [2], [3], one can do the same analysis for the Hawkes process with differnt exciting functions. For example, we can fix a partition (A_1, A_2) for $\mathbb{N} \cup \{0\}$ and let $N_1 = \sum_{n \in A_1} N^n$ and $N_2 = \sum_{n \in A_2} N^n$. Then, we can study the signature plot $\hat{C}(\tau)$ for $X_t = N_1(t) - N_2(t)$. One can also fix a partition (A_1, A_2, A_3, A_4) for $\mathbb{N} \cup \{0\}$ and let $N_i = \sum_{n \in A_i} N^n$, $1 \leq i \leq 4$. Then, we can study the correlation coefficient $\hat{\rho}(\tau)$ for $X_t^1 = N_1(t) - N_2(t)$, $X_t^2 = N_3(t) - N_4(t)$.

In the context of the Hawkes process with different exciting functions, since we already proved ergodicity in Theorem 1, by considering large T, i.e. letting $T \to \infty$, by ergodic theorem,

$$\hat{C}(\tau) \to C(\tau) := \frac{1}{\tau}\mathbb{E}[(X_\tau)^2], \tag{5.4}$$

and

$$\hat{\rho}(\tau) \to \rho(\tau) := \frac{\mathbb{E}[X_\tau^1 X_\tau^2]}{\sqrt{\mathbb{E}[(X_\tau^1)^2]\mathbb{E}[(X_\tau^2)^2]}}, \tag{5.5}$$

as $T \to \infty$, where the expectations are taken over the stationary version of the processes. Heuristically, as $\tau \to 0$, $\mathbb{E}[X_\tau^1 X_\tau^2] = O(\tau^2)$, $\mathbb{E}[(X_\tau^1)^2] = O(\tau)$ and $\mathbb{E}[(X_\tau^2)^2] = O(\tau)$. Thus, as $\tau \to 0$, $\rho(\tau) = O(\tau)$ and this explains the vanishing correlation coefficient as $\tau \to 0$ in the Epps effect.

Our main result is that $C(\tau)$ and $\rho(\tau)$ can be computed by evaluating $\mathbb{E}[(X_\tau^1)^2]$, $\mathbb{E}[(X_\tau^2)^2$, and $\mathbb{E}[X_\tau^1 X_\tau^2]$:

Proposition 16.

$$
\begin{aligned}
\mathbb{E}[(X_\tau^1)^2] &= \sum_{i\in A_1}\overline{\gamma}_0 m_i\tau + \sum_{i,j\in A_1}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu \\
&\quad + \sum_{i\in A_2}\overline{\gamma}_0 m_i\tau + \sum_{i,j\in A_2}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu - 2\sum_{i\in A_1,j\in A_2}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu
\end{aligned}
\tag{5.6}
$$

$$
\begin{aligned}
\mathbb{E}[(X_\tau^2)^2] &= \sum_{i\in A_3}\overline{\gamma}_0 m_i\tau + \sum_{i,j\in A_3}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu \\
&\quad + \sum_{n\in A_4}\overline{\gamma}_0 m_n\tau + \sum_{i,j\in A_4}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu - 2\sum_{i\in A_3,j\in A_4}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu
\end{aligned}
\tag{5.7}
$$

$$
\begin{aligned}
\mathbb{E}[X_\tau^1 X_\tau^2] &= \sum_{i\in A_1,j\in A_3}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu + \sum_{i\in A_2,j\in A_4}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu \\
&\quad - \sum_{i\in A_2,j\in A_3}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu - \sum_{i\in A_1,j\in A_4}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu,
\end{aligned}
\tag{5.8}
$$

where $\rho(\cdot,\cdot,\cdot)$ are defined iteratively as $\rho(\cdot,\cdot,t)=\rho(\cdot,\cdot-t)$, $t>0$, and for $t>s$, $i\geq 1$,

$$
\begin{aligned}
\rho(i,i,t-s) &= \int_{-\infty}^t\int_{-\infty}^s \gamma_i(t-u)\gamma_i(s-v)\rho(i-1,i-1,|u-v|)dudv \\
&\quad + \int_{-\infty}^s \gamma_i(t-u)\gamma_i(s-u)\overline{\gamma}_0 m_{i-1}du,
\end{aligned}
\tag{5.9}
$$

and $\rho(0,0,t-s)=(\overline{\gamma}_0)^2$, $t>s$ and for $j\geq i+1$, $t>s$,

$$
\rho(i,j,t-s) = \int_{-\infty}^s \gamma_j(s-u)\rho(i,j-1,t-u)du,
\tag{5.10}
$$

and finally,

$$
\rho(i,i+1,t-s) = \begin{cases} \int_{-\infty}^s \gamma_i(s-u)\rho(i,i,t-u)du & \text{if } t>s \\ \int_{-\infty}^s \gamma_i(s-u)\rho(i,i,t-u)du + \gamma_i(s-t)\overline{\gamma}_0 m_i & \text{if } t<s \end{cases}.
\tag{5.11}
$$

Proof. Let $N^i(dt) := N^i_{t+d\delta} - N^i_t$.

First, for any $i\in\mathbb{N}\cup\{0\}$,

$$
\frac{1}{d\delta}\mathbb{E}[N^i(dt)] = \overline{\gamma}_0 m_i,
\tag{5.12}
$$

where m_i is defined in (1.6) for $i\in\mathbb{N}$ and $m_0:=1$.

Second, since N^i is a simple point process,

$$
\frac{1}{d\delta}\mathbb{E}[N^i(dt)N^i(dt)] = \frac{1}{d\delta}\mathbb{E}[N^i(dt)] = \overline{\gamma}_0 m_i.
\tag{5.13}
$$

Third, for any $t\neq s$, by stationarity, we can define

$$
\rho(i,j,t-s) := \frac{1}{(d\delta)^2}\mathbb{E}[N^i(dt)N^j(ds)].
\tag{5.14}
$$

Therefore, we can compute that

(5.15)

$$
\begin{aligned}
\mathbb{E}[(X_\tau^1)^2] &= \mathbb{E}\left[\left(\sum_{n\in A_1}\int_0^t N^n(ds) - \sum_{n\in A_2}\int_0^t N^n(ds)\right)^2\right] \\
&= \sum_{i\in A_1}\overline{\gamma}_0 m_i \tau + \sum_{i,j\in A_1}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu \\
&\quad + \sum_{i\in A_2}\overline{\gamma}_0 m_i\tau + \sum_{i,j\in A_2}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu - 2\sum_{i\in A_1, j\in A_2}\int_0^\tau\int_0^\tau \rho(i,j,s-u)dsdu
\end{aligned}
$$

Similarly, we can show (5.7) and (5.8).

What remains is to compute $\rho(\cdot,\cdot,\cdot)$. By symmetry,

$$\rho(i,i,t) = \rho(i,i,-t), \qquad -\infty < t < \infty. \tag{5.16}$$

Therefore, for $t > s$, and $i \geq 1$,

$$
\begin{aligned}
\rho(i,i,t-s) &= \mathbb{E}[\lambda_t^i \lambda_s^i] \\
&= \mathbb{E}\left[\int_{-\infty}^t \gamma_i(t-u)N^{i-1}(du)\int_{-\infty}^s \gamma_i(s-v)N^{i-1}(dv)\right] \\
&= \int_{-\infty}^t\int_{-\infty}^s \gamma_i(t-u)\gamma_i(s-v)\rho(i-1,i-1,|u-v|)dudv \\
&\qquad + \int_{-\infty}^s \gamma_i(t-u)\gamma_i(s-u)\overline{\gamma}_0 m_{i-1}du.
\end{aligned} \tag{5.17}
$$

It is clear that $\rho(0,0,t-s) = (\overline{\gamma}_0)^2$ for any $t > s$.

Fourth, for $j \geq i+1$,

$$
\begin{aligned}
\rho(i,j,t-s) &= \frac{1}{(d\delta)^2}\mathbb{E}[N^i(dt)N^j(ds)] \\
&= \frac{1}{d\delta}\mathbb{E}[N^i(dt)\lambda_s^j] \\
&= \frac{1}{d\delta}\mathbb{E}\left[N^i(dt)\int_{-\infty}^s \gamma_j(s-u)N^{j-1}(du)\right] \\
&= \int_{-\infty}^s \gamma_j(s-u)\rho(i,j-1,t-u)du.
\end{aligned} \tag{5.18}
$$

Fifth and finally,

(5.19)

$$
\begin{aligned}
\rho(i,i+1,t-s) &= \frac{1}{(d\delta)^2}\mathbb{E}[N^i(dt)N^{i+1}(ds)] \\
&= \frac{1}{d\delta}\mathbb{E}[N^i(dt)\lambda_s^{i+1}] \\
&= \frac{1}{d\delta}\mathbb{E}\left[N^i(dt)\int_{-\infty}^s \gamma_i(s-u)N^i(du)\right] \\
&= \begin{cases} \int_{-\infty}^s \gamma_i(s-u)\rho(i,i,t-u)du & \text{if } t > s \\ \int_{-\infty}^s \gamma_i(s-u)\rho(i,i,t-u)du + \gamma_i(s-t)\overline{\gamma}_0 m_i & \text{if } t < s \end{cases}.
\end{aligned}
$$

□

5.2. **Asymptotic Ruin Probabilities for a Risk Process with Hawkes Arrivals with Different Exciting Functions.** In this section, we study the applications to ruin probabilities. The applications of the Hawkes processes to ruin probabilities in insurnace have been studied in Stabile and Torrisi [26], Zhu [31] for instance. The advantage of using a Hawkes processes than a standard Poisson process is that the arrivals of the claims will have a contagion and clustering effect. We consider the following risk model for the surplus process R_t of an insurance portfolio,

$$R_t = u + pt - \sum_{i=1}^{N_t} C_i, \tag{5.20}$$

where $u > 0$ is the initial reserve, $p > 0$ is the constant premium and the C_i's are i.i.d. positive random variables with $\mathbb{E}[e^{\theta C_1}] < \infty$ for any $\theta \in \mathbb{R}$. C_i represents the claim size at the ith arrival time, these being independent of N_t, the Hawkes process with exciting functions $(\gamma_n)_{n\in\mathbb{N}\cup\{0\}}$.

For $u > 0$, let

$$\tau_u = \inf\{t > 0 : R_t \leq 0\}, \tag{5.21}$$

and denote the infinite and finite horizon ruin probabilities by

$$\psi(u) = \mathbb{P}(\tau_u < \infty), \quad \psi(u, uz) = \mathbb{P}(\tau_u \leq uz), \quad u, z > 0. \tag{5.22}$$

We first consider the case when the claim sizes have light-tails, i.e. there exists some $\theta > 0$ so that $\mathbb{E}[e^{\theta C_1}] < \infty$.

By the law of large numbers,

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{N_t} C_i = m\mathbb{E}[C_1]. \tag{5.23}$$

By Theorem 7, $\Gamma_C(\theta) := \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta \sum_{i=1}^{N_t} C_i}]$ exists.

To exclude the trivial case, we assume that

$$m\mathbb{E}[C_1] < p < \frac{\Gamma_C(\theta_c)}{\theta_c}, \tag{5.24}$$

where the critical value θ_c is defined as

$$\theta_c := \sup\{\theta : \Gamma_C(\theta) < \infty\}. \tag{5.25}$$

The first inequality in (5.24) is the usual net profit condition in ruin theory and the second inequality in (5.24) guarantees that the equation $\Gamma_C(\theta) = p\theta$ has a unique positive solution $\theta^\dagger < \theta_c$.

To see this, let $G(\theta) = \Gamma_C(\theta) - p\theta$. Notice that $G(0) = 0$, $G(\infty) = \infty$, and that G is convex. We also have $G'(0) = m\mathbb{E}[C_1] - p < 0$ and $\Gamma_C(\theta_c) - \rho\theta_c > 0$ by (5.24). Therefore, there exists only one solution $\theta^\dagger \in (0, \theta_c)$ of $\Gamma_C(\theta^\dagger) = p\theta^\dagger$.

Theorem 17 (Infinite Horizon)**.** *Assume* (5.24)*, we have* $\lim_{u\to\infty} \frac{1}{u} \log \psi(u) = -\theta^\dagger$*, where* $\theta^\dagger \in (0, \theta_c)$ *is the unique positive solution of* $\Gamma_C(\theta) = p\theta$*.*

Proof. Let us first quote a result from Glynn and Whitt [16]. Let S_n be random variables and $\tau_u = \inf\{n : S_n > u\}$ and $\psi(u) = \mathbb{P}(\tau_u < \infty)$. Assume that there exist some $\gamma, \epsilon > 0$ so that

(i) $\kappa_n(\theta) = \log \mathbb{E}[e^{\theta S_n}]$ is well defined and finite for $\gamma - \epsilon < \theta < \gamma + \epsilon$.

(ii) $\limsup_{n\to\infty} \mathbb{E}[e^{\theta(S_n - S_{n-1})}] < \infty$ for $-\epsilon < \theta < \epsilon$.
(iii) $\kappa(\theta) = \lim_{n\to\infty} \frac{1}{n}\kappa_n(\theta)$ exists and is finite for $\gamma - \epsilon < \theta < \gamma + \epsilon$.
(iv) $\kappa(\gamma) = 0$ and κ is differentiable at γ with $0 < \kappa'(\gamma) < \infty$.
Then, Glynn and Whitt [16] showed that $\lim_{u\to\infty} \frac{1}{u}\log\psi(u) = -\gamma$.

Take $S_t = \sum_{i=1}^{N_t} C_i - pt$ and $\kappa_t(\theta) = \log\mathbb{E}[e^{\theta S_t}]$. By Theorem 7, we have $\lim_{t\to\infty}\frac{1}{t}\kappa_t(\theta) = \Gamma_C(\theta) - p\theta$. Consider $\{S_{nh}\}_{n\in\mathbb{N}}$. We have $\lim_{n\to\infty}\frac{1}{n}\kappa_{nh}(\theta) = h\Gamma_C(\theta) - hp\theta$. By checking the conditions (i)-(iv), we get

$$\lim_{u\to\infty}\frac{1}{u}\log\mathbb{P}\left(\sup_{n\in\mathbb{N}} S_{nh} > u\right) = -\theta^{\dagger}. \tag{5.26}$$

Finally, notice that

$$\sup_{t\in\mathbb{R}^+} S_t \geq \sup_{n\in\mathbb{N}} S_{nh} \geq \sup_{t\in\mathbb{R}^+} S_t - ph. \tag{5.27}$$

Hence, $\lim_{u\to\infty}\frac{1}{u}\log\psi(u) = -\theta^{\dagger}$. □

Theorem 18 (Finite Horizon)**.** *Under the same assumptions as in Theorem 17, we have*

$$\lim_{u\to\infty}\frac{1}{u}\log\psi(u, uz) = -w(z), \quad \textit{for any } z > 0, \tag{5.28}$$

where

$$w(z) = \begin{cases} zI_C\left(\frac{1}{z} + p\right) & \textit{if } 0 < z < \frac{1}{\Gamma_C'(\theta^{\dagger}) - p} \\ \theta^{\dagger} & \textit{if } z \geq \frac{1}{\Gamma_C'(\theta^{\dagger}) - p} \end{cases}. \tag{5.29}$$

Proof. The proof is similar to that in Stabile and Torrisi [26] and we omit it here. □

Next, we are interested to study the case when the claim sizes have heavy tails, i.e. $\mathbb{E}[e^{\theta C_1}] = +\infty$ for any $\theta > 0$.

A distribution function B is subexponential, i.e. $B \in \mathcal{S}$ if

$$\lim_{x\to\infty}\frac{\mathbb{P}(C_1 + C_2 > x)}{\mathbb{P}(C_1 > x)} = 2, \tag{5.30}$$

where C_1, C_2 are i.i.d. random variables with distribution function B. Let us denote $B(x) := \mathbb{P}(C_1 \geq x)$ and let us assume that $\mathbb{E}[C_1] < \infty$ and define $B_0(x) := \frac{1}{\mathbb{E}[C]}\int_0^x \overline{B}(y)dy$, where $\overline{F}(x) = 1 - F(x)$ is the complement of any distribution function $F(x)$. The examples and properties of subexponential distributions can be found in the book by Asmussen and Albrecher [1].

Goldie and Resnick [17] showed that if $B \in \mathcal{S}$ and satisfies some smoothness conditions, then B belongs to the maximum domain of attraction of either the Frechet distribution or the Gumbel distribution. In the former case, $\overline{B}$ is regularly varying, i.e. $\overline{B}(x) = L(x)/x^{\alpha+1}$, for some $\alpha > 0$ and we write it as $\overline{B} \in \mathcal{R}(-\alpha - 1)$, $\alpha > 0$.

We assume that $B_0 \in \mathcal{S}$ and either $\overline{B} \in \mathcal{R}(-\alpha - 1)$ or $B \in \mathcal{G}$, i.e. the maximum domain of attraction of Gumbel distribution. $\mathcal{G}$ includes Weibull and lognormal distributions.

When the arrival process N_t satisfies a large deviation result, the probability that it deviates away from its mean is exponentially small, which is dominated by subexonential distributions. The results in Zhu [31] for the asymptotics of ruin probabilities for risk processes with non-stationary, non-renewal arrivals and subexponential claims can be applied in the context of Hawkes arrivals with different exciting functions. We have the following infinite-horizon and finite-horizon ruin probability estimates when the claim sizes are subexponential.

Theorem 19. *Assume the net profit condition $p > m\mathbb{E}[C_1]$.*

(i) (Infinite-Horizon)

$$\lim_{u\to\infty}\frac{\psi(u)}{\overline{B}_0(u)}=\frac{m\mathbb{E}[C_1]}{p-m\mathbb{E}[C_1]}. \tag{5.31}$$

(ii) (Finite-Horizon) For any $T>0$,

(5.32)

$$\lim_{u\to\infty}\frac{\psi(u,uz)}{\overline{B}_0(u)}=\begin{cases}\frac{m\mathbb{E}[C_1]}{p-m\mathbb{E}[C_1]}\left[1-\left(1+\left(1-\frac{m\mathbb{E}[C_1]}{p}\right)\frac{T}{\alpha}\right)^{-\alpha}\right] & \text{if } \overline{B}\in\mathcal{R}(-\alpha-1)\\ \frac{m\mathbb{E}[C_1]}{p-m\mathbb{E}[C_1]}\left[1-e^{-(1-\frac{m\mathbb{E}[C_1]}{p})T}\right] & \text{if } B\in\mathcal{G}\end{cases}.$$

Acknowledgements

The authors are extremely grateful to the editor and the referees for a very careful reading of the manuscript and also for the very helpful suggestions.

References

[1] Asmussen, S. and H. Albrecher. *Ruin Probabilities.* Second Edition, World Scientific, Singapore, 2010.
[2] Bacry, E., Delattre, S., Hoffmann, M. and J. F. Muzy. (2013). Scaling limits for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications* **123**, 2475-2499.
[3] Bacry, E., Delattre, S., Hoffmann, M. and J. F. Muzy. (2013). Modeling microstructure noise with mutually exciting point processes. *Quantitative Finance.* **13**, 65-77.
[4] Blundell, C., Heller, K. A. and J. M. Beck. Modelling reciprocating relationships with Hawkes processes. (2012). *Advances in Neural Information Processing Systems.* **25**.
[5] Bordenave, C. and Torrisi, G. L. (2007). Large deviations of Poisson cluster processes. *Stochastic Models*, **23**, 593-625.
[6] Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.*, **24**, 1563-1588.
[7] Chavez-Demoulin, V., Davison, A. C. and A. J. McNeil. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance.* **5**, 227-234.
[8] Chornoboy, E. S., Schramm, L. P. and A. F. Karr. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybern.* **59**, 265-275.
[9] Crane, R. and D. Sornette. (2008) Robust dynamic classes revealed by measuring the response function of a social system. *Proc. Nat. Acad. Sci. USA* **105**, 15649.
[10] Daley, D. J. and Vere-Jones, D. *An Introduction to the Theory of Point Processes*, Volume I and II, 2nd edition. Springer-Verlag, New York, 2003.
[11] Dembo, A. and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd Edition, Springer, 1998
[12] Embrechts, P., Liniger, T. and L. Lin. (2011). Multivariate Hawkes processes: an application to financial data. *J. Appl. Prob. Spec. Vol.* **48A**, 367-378.
[13] Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association.* **74**, 291-298.
[14] Errais, E., Giesecke, K. and Goldberg, L. (2010). Affine point processes and portfolio credit risk. *SIAM J. Financial Math.* **1**, 642-665.
[15] Fierro, R., Leiva, V. and J. Møller (2015). The Hawkes process with different exciting functions and its asymptotic behavior. *To appear in J. Appl. Probab.*
[16] Glynn, P. W. and W. Whitt. (1994). Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *J. Appl. Probab.* **31A**, 131-156.
[17] Goldie, C. M. and S. Resnick. (1988). Distributions that are both subexponential and in the domain of attraction of an extreme value distribution. *Adv. Appl. Probab.* **20**, 706-718.

[18] Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83-90.
[19] Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.* **11**, 493-503.
[20] Hawkes, A. G. and L. Adamopoulos. (1973). Cluster models for earthquakes-regional comparisons. *Bull. Int. Statist. Inst.* **45**, 454-461.
[21] Karabash, D. and L. Zhu. (2015). Limit theorems for marked Hawkes processes with application to a risk model. *To appear in Stochastic Models*.
[22] Liniger, T. (2009). *Multivariate Hawkes Processes*. PhD thesis, ETH.
[23] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83**, 9-27.
[24] Pernice, V., Staude B., Carndanobile, S. and S. Rotter. (2012). How structure determines correlations in neuronal networks. *PLoS Computational Biology.* **85**:031916.
[25] Pernice, V., Staude B., Carndanobile, S. and S. Rotter. (2011). Recurrent interactions in spiking networks with arbitrary topology. *Physical Review E.* **7**:e1002059.
[26] Stabile, G. and Torrisi, G. L. (2010). Risk processes with non-stationary Hawkes arrivals. *Methodol. Comput. Appl. Prob.* **12** 415-429.
[27] Varadhan, S. R. S. *Large Deviations and Applications*. SIAM, Philadelphia, 1984.
[28] Zhu, L. (2013). *Nonlinear Hawkes Processes*. PhD thesis, New York University.
[29] Zhu, L. (2013). Moderate deviations for Hawkes processes. *Statistics & Probability Letters.* **83**, 885-890.
[30] Zhu, L. (2014). Limit theorems for a Cox-Ingersoll-Ross process with Hawkes jumps. *Journal of Applied Probability.* **51**, 699-712.
[31] Zhu, L. (2013). Ruin probabilities for risk processes with non-stationary arrivals and subexponential claims. *Insurance: Mathematics and Economics.* **53** 544-550.
[32] Zhu, L. (2013). Central limit theorem for nonlinear Hawkes processes. *Journal of Applied Probability.* **50** 760-771.
[33] Zhu, L. (2015). Large deviations for Markovian nonlinear Hawkes Processes. *Annals of Applied Probability.* **25**, 548-581.
[34] Zhu, L. (2014). Process-level large deviations for nonlinear Hawkes point processes. *Annales de l'Institut Henri Poincaré.* **50**, 845-871.

Courant Institute of Mathematical Sciences
New York University
251 Mercer Street
New York, NY-10012
United States of America
E-mail address: `mehrdad@cims.nyu.edu`

School of Mathematics
University of Minnesota-Twin Cities
206 Church Street S.E.
Minneapolis, MN-55455
United States of America
E-mail address: `zhul@umn.edu`