MATHEMATICS OF OPERATIONS RESEARCH
Vol. 40, No. 4, November 2015, pp. 797–819
ISSN 0364-765X (print) | ISSN 1526-5471 (online)

http://dx.doi.org/10.1287/moor.2014.0696

Affine Point Processes: Approximation and Efficient Simulation

Xiaowei Zhang
Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, xiaoweiz@ust.hk

Jose Blanchet
Department of Industrial Engineering and Operations Research, Columbia University, New York, New York 10027, jose.blanchet@columbia.edu

Kay Giesecke, Peter W. Glynn
Department of Management Science and Engineering, Stanford University, Stanford, California 94305, {giesecke@stanford.edu, glynn@stanford.edu}

We establish a central limit theorem and a large deviations principle for affine point processes, which are stochastic models of correlated event timing widely used in finance and economics. These limit results generate closed-form approximations to the distribution of an affine point process. They also facilitate the construction of an asymptotically optimal importance sampling estimator of tail probabilities. Numerical tests illustrate our results.

Keywords: affine point process; affine jump diffusion; central limit theorem; large deviations; rare-event simulation
MSC2000 subject classification: Primary: 60G55, 60F05, 60F10; secondary: 60J60, 60J75
OR/MS subject classification: Primary: stochastic model applications, diffusion, limit theorems; secondary: simulation
History: Received October 29, 2013; revised May 9, 2014; accepted October 8, 2014. Published online in *Articles in Advance* February 4, 2015.

1. Introduction. Point processes serve as stochastic models of event timing in many areas. In finance, point processes are used to describe credit defaults, arrivals of security orders, jumps in asset prices, and other economically significant events. Affine point processes constitute a particularly tractable class of models. These are specifications in which the arrival intensity is an affine function of an affine jump diffusion (AJD) (see Duffie et al. [24]). The transform of an affine point process is an exponentially affine function of the driving jump diffusion; the coefficients solve a system of ordinary differential equations (ODEs) (see Errais et al. [27]). The components of an affine point process are self- and cross-exciting and facilitate the description of complex event dependence structures. Due to their modeling flexibility and computational tractability, affine point processes are widely used in finance and economics (Aït-Sahalia et al. [2], Azizpour et al. [4], Bowsher [10], Embrechts et al. [26], and many others).

This paper analyzes the long-term asymptotics of affine point processes. We first establish a central limit theorem (CLT), which describes the typical behavior of the process in the long run, and which leads to a Gaussian approximation to the distribution of the process. The approximation can be evaluated quickly because the asymptotic mean and variance can be computed analytically. We then prove a large deviations (LD) principle, which characterizes the atypical behavior of the process, and which leads to an approximation of the tail of the distribution. The LD principle also facilitates the construction of an importance sampling (IS) scheme for estimating tail probabilities. We provide conditions guaranteeing the asymptotic optimality of this scheme. Numerical results illustrate the performance of the approximations and the simulation scheme.

Our results may be useful in the many cases where the ODEs governing the transform of an affine point process cannot be solved in closed form. To compute the distribution of the process, which is the key quantity required to address the eventual application, the numerical solution of a system of ODEs must be embedded within a numerical transform inversion algorithm. Such an algorithm typically uses thousands of evaluations of the transform, and each evaluation requires the numerical solution of an ODE system. This procedure is typically burdensome; the computational cost often renders empirical applications involving parameter estimation problems impractical. Our analytical and Monte Carlo approximations to the distribution of an affine point process provide a computationally efficient alternative to this procedure.[1]

Our work complements the literature on the LD analysis of Markov processes, which is a long-standing and extensive field (see Kontoyiannis and Meyn [41] and references therein). The compensator of an affine point process is a Markov additive functional of the form $I_f(t) = \int_0^t f(X(s))\,\mathrm{d}s$, where X is an AJD. The LD analysis

[1] Another approach is provided by Glasserman and Kim [38]. They develop the use of saddle point approximations as alternatives to numerical transform inversion, focusing on AJD.

for $I_f(t)$ with an unbounded functional f is challenging since its LD behavior is fundamentally different from that in the bounded case. In particular, if f is bounded, then $I_f(t)$ behaves like a random walk with light-tailed increments, namely, the tail probability of $t^{-1}I_f(t)$ decays exponentially as t increases. But if f is unbounded, the same tail probability may decay subexponentially fast (see Duffy and Meyn [25] and Blanchet et al. [8]). Our LD result provides an example in which the tail of $t^{-1}I_f(t)$ does decay exponentially fast for an unbounded functional f.

Prior work has studied the asymptotic behavior of point processes. Cvitanić et al. [15] and Giesecke et al. [34, 36] prove laws of large numbers, Dai Pra et al. [18], Dai Pra and Tolotti [17], Giesecke and Weber [32], and Spiliopoulos et al. [47] develop CLTs, whereas Dai Pra et al. [18] and Spiliopoulos and Sowers [46] examine LDs. These articles examine different, nonaffine systems of indicator point processes that represent default events in a pool of credit assets. They consider an asymptotic regime in which the number of system components (constituent assets) tends to infinity and the time horizon remains fixed. We, in contrast, focus on a system of nonterminating point processes with an affine structure and consider an asymptotic regime in which the time horizon tends to infinity, but the system size remains fixed. Moreover, the approximations we obtain may be appropriate for small systems with few components. Daley [19], Bordenave and Torrisi [9], Zhu [50, 51], and Bacry et al. [5] study the long-term asymptotic behavior of certain Hawkes processes, some of which are special cases of affine point processes. The intensity of a Hawkes process is a function of the path of the process only, while the intensity of an affine point process takes a more general form.

There is also prior work on rare-event simulation for systems of indicator point processes. Bassamboo and Jain [6] develop an asymptotically optimal IS scheme for a certain affine system with doubly stochastic structure. The key assumption is that events occur independently of one another, given the path of an AJD factor influencing all system components. The affine system we treat in this paper is richer: we do not require the narrowing doubly stochastic structure, and allow for the self- and cross-excitation effects that are relevant in many application contexts. Carmona and Crépey [12] develop an interacting particle scheme (IPS) (see Del Moral and Garnier [20]) for Markov chain systems. Further, Giesecke et al. [35] develop an IPS, Deng et al. [22] an asymptotically optimal sequential resampling scheme, and Giesecke and Shkolnik [31] an asymptotically optimal IS algorithm for general systems. These papers also consider a "large pool" rather than a "large horizon" regime.

The rest of the paper is organized as follows. Section 2 formulates the model and assumptions. Section 3 develops the CLT, while §4 analyzes LDs. Section 5 discusses extensions. Section 6 exploits the LD principle to develop an IS algorithm for estimating the tail of an affine point process, and proves the optimality of the scheme. Section 7 provides numerical results and §8 concludes. An appendix collects some proofs.

2. Problem formulation. Throughout the paper, we use the following notation:

- We take $\mathbb{R}^d_+ = \{v \in \mathbb{R}^d\colon v_i \geq 0, i = 1, \ldots, d\}$ and $\mathbb{R}^d_- = \{v \in \mathbb{R}^d\colon v_i \leq 0, i = 1, \ldots, d\}$.
- A vector $v \in \mathbb{R}^d$ is taken as a column vector, $v^\top$ denotes the transpose, $\|v\|$ denotes the Euclidean norm, and $\operatorname{diag}(v)$ denotes the diagonal matrix whose diagonal elements are v.
- For a matrix A, we write $A \succeq 0$ if A is symmetric positive semidefinite.
- $\mathbf{I}$ denotes the identity matrix, $\mathbf{0}$ denotes a zero matrix, and $\mathrm{Id}(i)$ denotes a matrix with all entries equal to 0 except the i-th diagonal entry, which is 1 (regardless of dimension).
- Let $I, J \subseteq \{1, \ldots, d\}$ be two index sets. For a vector $v \in \mathbb{R}^d$ and a matrix $A \in \mathbb{R}^{d\times d}$, we write $v_I = (v_i\colon i \in I)$ and $A_{I,J} = (A_{ij}\colon i \in I, j \in J)$.

We fix a complete probability space $(\Omega, \mathbb{P}, \mathcal{F})$ and a filtration $\{\mathcal{F}_t\colon t \geq 0\}$ satisfying the usual conditions of right continuity and completeness (see, for example, Karatzas and Shreve [40] for details). Let $W = (W(t)\colon t \geq 0)$ be a standard d-dimensional Brownian motion. Let $X = (X(t)\colon t \geq 0)$ be an affine jump-diffusion process in the sense of Duffie et al. [24]. In particular, X is a Markov process in a state space $\mathcal{S} \subseteq \mathbb{R}^d$ satisfying the jump-diffusion stochastic differential equation (SDE)

$$\mathrm{d}X(t) = \mu(X(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t) + \sum_{i=1}^n \gamma_i \int_{\mathbb{R}_+} z N_i(\mathrm{d}t, \mathrm{d}z) \tag{1}$$

with $X(0) = x_0$, where the drift and volatility functions are given by

$$\mu(x) = b - \beta x, \quad b \in \mathbb{R}^d, \ \beta \in \mathbb{R}^{d\times d}$$

$$\sigma(x)\sigma(x)^\top = a + \sum_{j=1}^d \alpha^j x_j, \quad a \in \mathbb{R}^{d\times d}, \ \alpha^j \in \mathbb{R}^{d\times d}, \ j = 1, \ldots, d.$$

Here, $\gamma_i \in \mathbb{R}^d$ and $N_i(\mathrm{d}t, \mathrm{d}z)$ is a random counting measure on $[0,\infty)\times\mathbb{R}_+$ with compensator measure $\Lambda_i(X(t))\,\mathrm{d}t\varphi_i(\mathrm{d}z)$, where φ_i is a probability measure on $\mathbb{R}_+$ and

$$\Lambda_i(x) = \lambda_i + \sum_{j=1}^{d} \kappa_{i,j} x_j, \quad \lambda \in \mathbb{R}^n, \quad \kappa \in \mathbb{R}^{n\times d}.$$

Moreover, we let Z^i denote a random variable drawn having distribution φ_i throughout the paper.

The SDE (1) has n jump components. The process defined by $N_i(t) = \int_0^t \int_0^\infty N_i(\mathrm{d}s, \mathrm{d}z)$ counts the number of jumps of the i-th component. The arrival intensity of N_i is $\Lambda_i(X)$. When N_i jumps, the process X exhibits a jump of size $\gamma_i Z^i$. Thus the parameter γ_i controls the sensitivity of X to the jumps of N_i.

An *affine point process* $L = (L_1, \ldots, L_n)$ is given by

$$L_i(t) \triangleq \int_0^t \int_{\mathbb{R}_+} z N_i(\mathrm{d}s, \mathrm{d}z).$$

We are interested in the long-term asymptotic behavior of

$$V(t) \triangleq \sum_{i=1}^{n} L_i(t).$$

Example 1. Suppose the parameters are specified as follows.

- $b = (b_1, \ldots, b_d)$ and $\beta = \mathrm{diag}(\beta_1, \ldots, \beta_d)$.
- The impact parameter $\gamma_i = (\delta_{i,1}, \ldots, \delta_{i,d})$ for $i = 1, \ldots, d$.
- The volatility function $\sigma(x) = \mathrm{diag}(\sigma_1\sqrt{x_1}, \ldots, \sigma_d\sqrt{x_d})$, so that $a = \mathbf{0}$ and $\alpha^i = \sigma_i \cdot \mathrm{Id}(i)$ for $i = 1, \ldots, d$.
- The intensity function $\Lambda_i(x) = \lambda_i + \kappa_i x_i$, so that $\lambda = (\lambda_1, \ldots, \lambda_d)$ and $\kappa = \mathrm{diag}(\kappa_1, \ldots, \kappa_d)$.

Then $X = (X_1, \ldots, X_d)$ satisfies

$$\mathrm{d}X_j(t) = (b_j - \beta_j X_j(t))\,\mathrm{d}t + \sigma_j\sqrt{X_j(t)}\,\mathrm{d}W_j(t) + \sum_{i=1}^{d} \delta_{i,j}\,\mathrm{d}L_i(t), \quad j = 1, \ldots, d,$$

where $b_j, \beta_j, \sigma_j, \delta_{i,j} > 0$. Moreover, the jump intensity of $L_i(t)$ is $\lambda_i + \kappa_i X_i(t)$ for some $\lambda_i, \kappa_i > 0$. The feedback term $\sum_{i=1}^{d} \delta_{i,j}\,\mathrm{d}L_i(t)$ introduces *self-* and *cross-excitation* into L. If $\delta_{i,j} = 0$ for all $i, j = 1, \ldots, d$, these effects are absent.

The following assumption will be imposed throughout the paper.

Assumption 1. (I) *There exist index sets $I = \{1, \ldots, m\}$ and $J = \{m+1, \ldots, d\}$ such that*
(1) $a \succeq 0$ *with* $a_{I,I} = \mathbf{0}$ (*hence* $a_{I,J} = \mathbf{0}$ *and* $a_{J,I} = \mathbf{0}$).
(2) $\alpha^i \succeq 0$ *and* $\alpha^i_{I,I} = \alpha^i_{i,i}\mathrm{Id}(i)$ *for* $i \in I$; $\alpha^i = \mathbf{0}$ *for* $i \in J$.
(3) $b \in \mathbb{R}^m_+ \times \mathbb{R}^{d-m}$.
(4) $\beta_{I,J} = \mathbf{0}$ *and* $\beta_{I,I}$ *is a Z-matrix, i.e.,* $\beta_{I,I}$ *has nonpositive off-diagonal elements.*
(5) $\lambda \in \mathbb{R}^n_+$, $\kappa \in \mathbb{R}^{n\times d}_+$ *with* $\kappa_{i,J} = \mathbf{0}$ *for* $i = 1, \ldots, n$.
(6) $\gamma_i \in \mathbb{R}^m_+ \times \mathbb{R}^{d-m}$ *for* $i = 1, \ldots, n$.
(II) $\alpha^i_{i,i} > 0$ *for each* $i = 1, \ldots, m$; $b_i > 0$ *for each* $i = 1, \ldots, m$; $\lambda_i + \sum_{j=1}^{m} \kappa_{i,j} > 0$ *for each* $i = 1, \ldots, n$.
(III) $\beta - \sum_{i=1}^{n} \mathbb{E}(Z^i)\gamma_i\kappa_i^\top$ *is positive stable, where* $\kappa_i^\top$ *is the i-th row of* κ, $i = 1, \ldots, n$.

Part (I) of Assumption 1 defines the *admissible* parameters of *canonical* affine processes, which include virtually all the affine processes used in practice. We refer the readers to Duffie et al. [23] for an extensive discussion on the topic and to Dai and Singleton [16] for more examples of canonical affine models. In particular, under such an assumption on the parameters $(a, \alpha, b, \beta, \lambda, \kappa, \gamma)$ of the SDE (1), the state space of a canonical AJD X is of the form $\mathcal{S} = \mathbb{R}^m_+ \times \mathbb{R}^{d-m}$. The first m components are of CIR type and they are the ones that truly govern the dynamics of the jump intensities and the volatilities, whereas the remaining $d-m$ components are of O-U type and their jump intensities and volatilities depend on the first m components. Moreover, it is easy to verify that the model in Example 1 indeed satisfies part (I) of Assumption 1.

Part (II) guarantees that the variance matrix of $X_1, \ldots, X_m$ is nondegenerate and that each L_i has a positive jump intensity. Moreover, that each b_i is positive is a very mild assumption, which is widely imposed in financial models (see, e.g., Filipović et al. [29]). Note that one can interpret $\mathbb{E}(Z^i)\gamma_i$ as the average impact of a jump of $L_i(t)$ on the intensity and κ_i as the jump frequency. Therefore part (III) states that the effect of jumps is dominated by that of mean reversion, which is represented by β. Namely, the jumps are neither too big nor too frequent so that X can be driven to the equilibrium by the force of mean reversion. Indeed, part (III) plays a crucial role in proving the ergodicity of $X(t)$, which is essential to derive the long-term behavior of $V(t)$ (see Proposition 9 in the appendix).

3. Typical behavior: CLT. Our first goal is to characterize the *typical* long-term behavior of $V = \sum_{i=1}^n L_i$. In particular, we will prove that

$$t^{-1/2}(V(t) - rt) \Rightarrow \eta \mathcal{N}(0, 1) \tag{2}$$

as $t \to \infty$, for some constants $r, \eta \in \mathbb{R}_+$ to be determined later, where $\Rightarrow$ denotes convergence in distribution and $\mathcal{N}(0, 1)$ is a Gaussian random variable with mean 0 and unit variance.

To guarantee the finiteness of the asymptotic variance η^2, we will impose the following assumption in this section.

Assumption 2. *There exists $\epsilon > 0$ for which $\mathbb{E}(Z^i)^{2+\epsilon} < \infty$ for all $i = 1, \ldots, n$.*

To prove the CLT (2), we first construct a local martingale U of the form

$$U(t) \triangleq V(t) - rt + A^\top (X(t) - X(0))$$

for some appropriately chosen $r \in \mathbb{R}$ and $A \in \mathbb{R}^d$, then derive a CLT for $U(t)$, and finally, show that the term $A^\top (X(t) - X(0))$ is asymptotically negligible.

3.1. Construction of local martingale. We have

$$U(t) = \sum_{i=1}^n \int_0^t \int_{\mathbb{R}_+} z N_i(\mathrm{d}s, \mathrm{d}z) - rt + \int_0^t A^\top (b - \beta X(s))\, \mathrm{d}s + \int_0^t A^\top \sigma(X(s))\, \mathrm{d}W(s) + \sum_{i=1}^n \int_0^t \int_{\mathbb{R}_+} A^\top \gamma_i z\, N_i(\mathrm{d}s, \mathrm{d}z).$$

Define the compensated random measure

$$\tilde{N}_i(\mathrm{d}s, \mathrm{d}z) \triangleq N_i(\mathrm{d}s, \mathrm{d}z) - \Lambda_i(X(s))\, \mathrm{d}s \varphi_i(\mathrm{d}z) = N_i(\mathrm{d}s, \mathrm{d}z) - (\lambda_i + \kappa_i^\top X(s))\, \mathrm{d}s \varphi_i(\mathrm{d}z).$$

It then follows that

$$\begin{aligned} U(t) &= \sum_{i=1}^n \int_0^t \int_{\mathbb{R}_+} (1 + A^\top \gamma_i) z \tilde{N}_i(\mathrm{d}s, \mathrm{d}z) + \sum_{i=1}^n \int_0^t (\lambda_i + \kappa_i^\top X(s))\, \mathrm{d}s \int_{\mathbb{R}_+} (1 + A^\top \gamma_i) z \varphi_i(\mathrm{d}z) - rt \\ &\quad + \int_0^t A^\top (b - \beta X(s))\, \mathrm{d}s + \int_0^t A^\top \sigma(X(s))\, \mathrm{d}W(s) \\ &= I_1(t) + I_2(t) + \int_0^t \left[\sum_{i=1}^n (1 + A^\top \gamma_i) \mathbb{E}(Z^i) \kappa_i^\top - A^\top \beta \right] X(s)\, \mathrm{d}s + \left[\sum_{i=1}^n \lambda_i (1 + A^\top \gamma_i) \mathbb{E}(Z^i) + A^\top b - r \right] t, \end{aligned} \tag{3}$$

where $I_1(t) \triangleq \sum_{i=1}^n \int_0^t \int_{\mathbb{R}_+} (1 + A^\top \gamma_i) z \tilde{N}_i(\mathrm{d}s, \mathrm{d}z)$ and $I_2(t) \triangleq \int_0^t A^\top \sigma(X(s))\, \mathrm{d}W(s)$. Note that I_1 and I_2 are both local martingales. Hence, if we choose r and A such that

$$\sum_{i=1}^n (1 + A^\top \gamma_i) \mathbb{E}(Z^i) \kappa_i^\top - A^\top \beta = 0$$

$$\sum_{i=1}^n \lambda_i (1 + A^\top \gamma_i) \mathbb{E}(Z^i) + A^\top b - r = 0,$$

then U is a local martingale in light of (3). Part (III) of Assumption 1 implies that the matrix $\beta - \sum_{i=1}^n \mathbb{E}(Z^i) \gamma_i \kappa_i^\top$ is nonsingular, so we can solve the above equations explicitly as follows:

$$\begin{aligned} A^\top &= \left(\sum_{i=1}^n \mathbb{E}(Z^i) \kappa_i^\top \right) \left(\beta - \sum_{i=1}^n \mathbb{E}(Z^i) \gamma_i \kappa_i^\top \right)^{-1} \\ r &= A^\top b + \sum_{i=1}^n \lambda_i \mathbb{E}(Z^i)(1 + A^\top \gamma_i). \end{aligned} \tag{4}$$

From now on, we will fix the values of r and A as given in (4). We have established the following result.

Proposition 1. *Under Assumptions 1 and 2, U is a local martingale.*

3.2. CLT for U. We will apply the local martingale CLT to U. To that end, we need to calculate the *predictable quadratic variation* $\langle U\rangle$ so as to compute the asymptotic variance η^2. See Protter [45] or Andersen et al. [3] for the definition and calculation of predictable quadratic variations.

Taking A and r as in (4), it follows from (3) that

$$U(t)=\sum_{i=1}^n\int_0^t\int_{\mathbb{R}_+}(1+A^\top\gamma_i)z\,\tilde{N}_i(\mathrm{d}s,\mathrm{d}z)+\int_0^t A^\top\sigma(X(s))\,\mathrm{d}W(s).$$

Therefore

$$\begin{aligned}\langle U\rangle(t)&=\sum_{i=1}^n\int_0^t\int_{\mathbb{R}_+}(1+A^\top\gamma_i)^2z^2\varphi_i(\mathrm{d}z)\Lambda_i(X(s))\,\mathrm{d}s+\int_0^t A^\top\sigma(X(s))\sigma(X(s))^\top A\,\mathrm{d}s\\&=\sum_{i=1}^n(1+A^\top\gamma_i)^2\,\mathbb{E}(Z^i)^2\int_0^t(\lambda_i+\kappa_i^\top X(s))\,\mathrm{d}s+\int_0^t A^\top\Big(a+\sum_{j=1}^d\alpha^jX_j(s)\Big)A\,\mathrm{d}s\\&=\Big(A^\top aA+\sum_{i=1}^n\lambda_iC_i\Big)t+\sum_{j=1}^d\Big(A^\top\alpha^jA+\sum_{i=1}^n\kappa_{i,j}C_i\Big)\int_0^t X_j(s)\,\mathrm{d}s,\end{aligned}\tag{5}$$

where $C_i\triangleq(1+A^\top\gamma_i)^2\,\mathbb{E}(Z^i)^2$.

Proposition 2. *Under Assumptions 1 and 2,*

$$\lim_{t\to\infty}\frac{\langle U\rangle(t)}{t}=A^\top aA+C^\top\lambda+(A^\top\alpha A+C^\top\kappa)\,\mathbb{E}_\pi X(0)\triangleq\eta^2\quad a.s.,\tag{6}$$

where $C\in\mathbb{R}^n$ with elements $C_i=(1+A^\top\gamma_i)^2\,\mathbb{E}(Z^i)^2$ and $A^\top\alpha A=(A^\top\alpha^1A,\ldots,A^\top\alpha^dA)$. Moreover, $\mathbb{E}_\pi X(0)$, where π is the stationary distribution of X, is given by (27).

Proof. This follows immediately from the strong law of large numbers for X (see Proposition 9) and (5). □

We also need the following technical result whose proof is deferred to §A.1.

Proposition 3. *Under Assumptions 1 and 2, for any $T>0$,*

$$\lim_{j\to\infty}\mathbb{E}\Big[\sup_{0\le t\le jT}j^{-1}|U(t)-U(t-)|^2\Big]=0.$$

Proposition 4. *Under Assumptions 1 and 2,*

$$t^{-1/2}U(t)\Rightarrow\mathcal{N}(0,\eta^2)$$

as $t\to\infty$, where η^2 is given by (6).

Proof. This follows from Propositions 2 and 3, and the local martingale CLT (see pp. 338–340 of Ethier and Kurtz [28]). □

3.3. CLT for V. Now, we are in a position to state our first main result. Note that the asymptotic mean and asymptotic variance of V can be analytically calculated.

Theorem 1. *Under Assumptions 1 and 2,*

$$t^{-1/2}(V(t)-rt)\Rightarrow\mathcal{N}(0,\eta^2)$$

as $t\to\infty$, where

$$\begin{aligned}r&=A^\top b+\sum_{i=1}^n\lambda_i\,\mathbb{E}(Z^i)(1+A^\top\gamma_i)\\\eta^2&=A^\top aA+C^\top\lambda+(A^\top\alpha A+C^\top\kappa)B,\\A^\top&=\Big(\sum_{i=1}^n\mathbb{E}(Z^i)\kappa_i^\top\Big)\Big(\beta-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_i\kappa_i^\top\Big)^{-1},\\B&=\Big(\beta-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_i\kappa_i^\top\Big)^{-1}\Big(b+\sum_{i=1}^n\lambda_i\,\mathbb{E}(Z^i)\gamma_i\Big),\\C_i&=(1+A^\top\gamma_i)^2\,\mathbb{E}(Z^i)^2,\quad i=1,\ldots,n.\end{aligned}$$

Proof. Proposition 9 asserts that $X(t) \Rightarrow X(\infty)$ as $t \to \infty$, where $X(\infty)$ has distribution π. Hence $t^{-1/2}X(t) \to 0$ in probability as $t \to \infty$. Note that $V(t) - rt = U(t) - A^\top(X(t) - X(0))$. It then follows immediately from Proposition 4 that

$$t^{-1/2}(V(t) - rt) \Rightarrow \mathcal{N}(0, \eta^2)$$

as $t \to \infty$. $\square$

4. Atypical behavior: LDs principle. As will be illustrated in §7, the Gaussian approximation implied by the CLT (2) is often not accurate enough for the tail of the distribution of $V(t)$. To obtain more accurate tail estimates, we will characterize the *atypical* behavior of $V(t)$ through a LD principle.

Note that Theorem 1 indicates that $t^{-1}V(t) \to r$ in probability as $t \to \infty$. Consequently, $\mathbb{P}(V(t) \geq Rt) \to 0$ as $t \to \infty$ for $R > r$. We will prove that under mild conditions, $V(t)$ satisfies the following LD principle:

$$\lim_{t\to\infty} t^{-1} \log \mathbb{P}(V(t) \geq Rt) = -\mathcal{I}(R),$$

where the *rate function* $\mathcal{I}(\cdot)$ will be defined later. The Gärtner-Ellis theorem provides a mechanism for establishing such an asymptotic result. A key role is played by the limiting cumulant generating function (CGF) of $V(t)$, which is given by

$$\lim_{t\to\infty} t^{-1} \log \mathbb{E} \exp(\theta V(t)). \tag{7}$$

It turns out that we need a moment condition on the jump size distribution φ_i that is stronger than Assumption 2 to guarantee the existence of (7). More specifically, for the rest of the paper, we will assume that the jump size distribution is light tailed, i.e., it has a finite exponential moment.

Assumption 3. $\sup\{\theta \in \mathbb{R}\colon \mathbb{E}e^{\theta Z^i} < \infty\} > 0$ *for each* $i = 1, \ldots, n$.

4.1. Road map. Because the analysis is technically involved, we outline here the proof of our second main result of this paper, i.e., the LD principle for $V(t)$.

The two challenges in applying the Gärtner-Ellis theorem in our context is: (i) to compute the limit CGF (7) and (ii) to establish its *steepness* (see, for example, Dembo and Zeitouni [21]).

To address the first challenge, we construct a martingale of the following exponential form:

$$M(t) = M_\theta(t) \triangleq \exp[\theta V(t) - \phi t + u^\top (X(t) - X(0))]$$

for some appropriately chosen $\phi \in \mathbb{R}$ and $u \in \mathbb{R}^d$. (Note that both ϕ and u clearly depend on the choice of θ, but we suppress this dependence when no ambiguity can arise.) In §4.2, we apply Itô's formula, similarly as §3.1, to derive sufficient conditions for ϕ and u, so that $M(t)$ be a local martingale. It turns out that ϕ can be expressed explicitly in terms of u, whereas u satisfies a system of nonlinear equations (13). We then follow the idea developed in Cheridito et al. [13] to prove that $M(t)$ is indeed a martingale with such chosen ϕ and u. In particular, we define a sequence of stopping times $\{\tau_l\colon l = 1, 2, \ldots\}$ such that $X(t)$ is bounded for $t < \tau_l$. The "stopped" version of $M(t)$ can be shown to be a martingale, and thus induces an equivalent probability measure Q^l. It is easy to see $\mathbb{E}M(T)\mathbb{I}(\tau_l \geq T) = Q^l(\tau_l \geq T)$, so in order that $M(t)$ be a martingale, it suffices to prove that $X(t)$ is nonexplosive under both $\mathbb{P}$ and Q^l, which can be deduced by virtue of Girsanov's theorem and the admissibility of the parameters.

Being a martingale, $M(t)$ induces a probability measure Q, in which case

$$\mathbb{E} \exp[\theta V(t) - \phi t] = \mathbb{E}^Q \exp[-u^\top (X(t) - X(0))].$$

Nevertheless, the system of nonlinear equations (13) may have multiple solutions. The subtlety is to identify the probabilistically meaningful solution u that makes $\mathbb{E}^Q \exp[-u^\top(X(t) - X(0))]$ bounded, so that ϕ is indeed the limiting CGF, namely,

$$\lim_{t\to\infty} t^{-1} \log \mathbb{E} \exp(\theta V(t)) = \phi.$$

To that end, we carefully characterize the nonlinear system (13) in §4.3. A key observation is that $(\theta, u) = (0, \mathbf{0})$ satisfies (13) and it entails $\phi(0) = 0$, which ought to be true if ϕ is the limiting CGF. It is then conceivable that the desirable solution u should satisfy $u(0) = \mathbf{0}$. We treat u as an implicit function and establish its existence in a neighborhood of the origin by analyzing the nonsingularity of an associated Jacobian matrix. The maximal interval of existence of the desired solution u is determined by the nonsingularity of the Jacobian matrix. The relevant matrix analysis is fairly tractable thanks to the structure of the parameters (i.e., Assumption 1).

In §4.4, we further show that $\mathbb{E}^Q \exp[-u^\top(X(t)-X(0))] = O(1)$ as $t\to\infty$ for the properly chosen solution u by analyzing the stochastic stability of $X(t)$ via the Foster-Lyapunov method (see, for example, Meyn and Tweedie [43]). By then, we will prove that ϕ is indeed the limiting CGF for the properly chosen solution u.

The second challenge in applying the Gärtner-Ellis theorem is to show that the range of ϕ is $(0,\infty)$. Indeed, we can show that ϕ is monotonically increasing so it suffices to show that $\phi(\theta)\to 0$ as $\theta\downarrow 0$, whereas $\phi(\theta)\to\infty$ as $\theta\uparrow\infty$. The limits turn out to be essentially determined by the behavior of the aforementioned Jacobian matrix at its nonsingularity boundary.

With the two challenges addressed, we can safely apply the Gärtner-Ellis theorem to establish the LD principle for $V(t)$ in §4.5.

4.2. Construction of exponential martingale M. Let $Y(t)=\theta V(t)-\phi t+u^\top(X(t)-X(0))$. Itô's formula implies that

$$M(t) = 1+\int_0^t M(s-)\,\mathrm{d}Y^c(s)+\tfrac{1}{2}\int_0^t M(s-)\,\mathrm{d}[Y]^c(s)+\sum_{0<s\le t}(M(s)-M(s-)), \tag{8}$$

where Y^c is the path-by-path continuous part of Y and $[Y]^c$ is the path-by-path continuous part of the quadratic variation process $[Y]$. Note that

$$\begin{aligned} Y^c(t) &= -\phi t+\int_0^t u^\top(b-\beta X(s))\,\mathrm{d}s+\int_0^t u^\top\sigma(X(s))\,\mathrm{d}W(s)\\ &= (u^\top b-\phi)t-\int_0^t u^\top\beta X(s)\,\mathrm{d}s+\int_0^t u^\top\sigma(X(s))\,\mathrm{d}W(s), \end{aligned} \tag{9}$$

and

$$[Y]^c(t) = \int_0^t u^\top\sigma(X(s))\sigma(X(s))^\top u\,\mathrm{d}s = (u^\top a u)t+\sum_{j=1}^n u^\top\alpha^j u\int_0^t X_j(s)\,\mathrm{d}s, \tag{10}$$

and letting $G(t)=\sum_{i=1}^n\int_0^t\int_{\mathbb{R}_+}\gamma_i z N_i(\mathrm{d}s,\mathrm{d}z)$,

$$\begin{aligned} \sum_{0<s\le t}(M(s)-M(s-)) &= \sum_{0<s\le t} M(s-)(e^{\theta(V(s)-V(s-))+u^\top(G(s)-G(s-))}-1)\\ &= \int_0^t\int_{\mathbb{R}_+} M(s-)\sum_{i=1}^n(e^{\theta z+u^\top\gamma_i z}-1)N_i(\mathrm{d}s,\mathrm{d}z)\\ &= \int_0^t\int_{\mathbb{R}_+} M(s-)\sum_{i=1}^n(e^{(\theta+u^\top\gamma_i)z}-1)\tilde{N}_i(\mathrm{d}s,\mathrm{d}z)\\ &\quad+\sum_{i=1}^n(\mathbb{E}\,e^{(\theta+u^\top\gamma_i)Z^i}-1)\int_0^t M(s-)(\lambda_i+\kappa_i^\top X(s))\,\mathrm{d}s. \end{aligned} \tag{11}$$

Plugging (9), (10), and (11) into (8) yields that

$$\begin{aligned} M(t) &= 1+\int_0^t M(s-)u^\top\sigma(X(s))\,\mathrm{d}W(s)+\int_0^t\int_{\mathbb{R}^d} M(s-)\sum_{i=1}^n(e^{(\theta+u^\top\gamma_i)z}-1)\tilde{N}_i(\mathrm{d}s,\mathrm{d}z)\\ &\quad+\int_0^t M(s-)\Big[u^\top b-\phi+\tfrac{1}{2}u^\top a u+\sum_{i=1}^n\lambda_i(\mathbb{E}\,e^{(\theta+u^\top\gamma_i)Z^i}-1)\Big]\,\mathrm{d}s\\ &\quad+\tfrac{1}{2}\sum_{j=1}^n u^\top\alpha^j u\int_0^t M(s-)X_j(s)\,\mathrm{d}s+\int_0^t M(s-)\Big[\sum_{i=1}^n(\mathbb{E}\,e^{(\theta+u^\top\gamma_i)Z^i}-1)\kappa_i^\top-u^\top\beta\Big]X(s)\,\mathrm{d}s. \end{aligned}$$

Therefore M is a local martingale if $\phi\in\mathbb{R}$ and $u\in\mathbb{R}^d$ satisfy

$$u^\top b-\phi+\tfrac{1}{2}u^\top a u+\sum_{i=1}^n\lambda_i(\mathbb{E}\,e^{(\theta+u^\top\gamma_i)Z^i}-1)=0, \tag{12}$$

and

$$\sum_{i=1}^d u_i\beta_{i,j}-\tfrac{1}{2}u^\top\alpha^j u-\sum_{i=1}^n(\mathbb{E}\,e^{(\theta+u^\top\gamma_i)Z^i}-1)\kappa_{i,j}=0,\quad j=1,\ldots,d. \tag{13}$$

As a matter of fact, if θ, u, and ϕ satisfy the last two equations, then M is indeed a martingale. Yet, to show this fact is by no means trivial. (Note that Novikov's condition is difficult to verify in our setting.) We proceed similarly as in Cheridito et al. [14], in which the authors study generic jump-diffusion processes with possible explosions.

PROPOSITION 5. *Suppose u and ϕ satisfy (12) and (13). Under Assumptions 1 and 3, $(M(t)\colon t\in[0,T])$ is a martingale for each $T>0$.*

PROOF. See §A.2 in the appendix. □

REMARK 1. Note that (13) may have multiple solutions u for a given θ. For instance, consider the simple case, where $\kappa=\mathbf{0}$ and β is diagonal. Then, for each $j=1,\dots,m$, $u_j=0$ or $u_j=2\beta_{j,j}/\alpha^j_{j,j}$, thereby yielding 2^m multiple solutions for u_I in total! See also Zhang et al. [49] for the discussion on multiple solutions (θ,u) for an affine point process when the underlying AJD is one dimensional. The challenge here is not only to address the existence of a solution to (13), but also to identify the probabilistically meaningful solution branch that serves our purpose.

4.3. Characterization of nonlinear system (13). Note that by part (I) of Assumption 1, $\alpha^j=\mathbf{0}$, $\kappa_{i,j}=0$ for $i=1,\dots,n$, $j=m+1,\dots,d$ and that $\beta_{i,j}=0$ for $i=1,\dots,m$ and $j=m+1,\dots,d$. Therefore it follows from (13) that

$$\sum_{i=m+1}^{d} u_i\beta_{i,j}=0,\quad j=m+1,\dots,d,$$

which, written in matrix form, is equivalent to

$$u_J^\top\beta_{J,J}=\mathbf{0},$$

where $J=\{m+1,\dots,d\}$. We will fix the two index sets in the rest of the paper: $I=\{1,\dots,m\}$ and $J=\{m+1,\dots,d\}$.

It then follows immediately from the block lower triangular form (Assumption 1) of β and Lemma 3 that $\beta_{J,J}$ is nonsingular. Hence $u_J=\mathbf{0}$, i.e., $u_i=0$ for $i=m+1,\dots,d$.

REMARK 2. We offer a heuristic interpretation for the fact that $u_J=u_J(\theta)\equiv\mathbf{0}$ for all θ. Note that $V(t)$ behaves "similarly" as its compensator

$$\sum_{i=1}^{n}\int_0^t \Lambda_i(X(s))\,\mathrm{d}s\int_{\mathbb{R}_+} z\,\varphi_i(\mathrm{d}z),$$

in the sense that they have the same expected value. The key observation is that the intensity functions $\Lambda_i(x)$ are independent of $X_J(t)$. Hence, only $X_i(t)$, $i=1,\dots,m$ are necessary to "offset" the randomness of $V(t)$, which heuristically explains why $u_i\equiv 0$ for $i=m+1,\dots,d$.

Now that we know $u_J\equiv\mathbf{0}$, we can focus on the first m components of u, i.e., u_I, and further simplify (12) and (13). In particular, by the assumptions on the structure of α and a, (12), and (13) can be simplified to

$$u_I^\top b_I-\phi+\sum_{i=1}^{n}\lambda_i(\mathbb{E}\,e^{(\theta+u_I^\top\gamma_{i,I})Z^i}-1)=0, \tag{14}$$

and

$$\sum_{i=1}^{m}u_i\beta_{i,j}-\tfrac{1}{2}\alpha^j_{j,j}u_j^2-\sum_{i=1}^{n}(\mathbb{E}\,e^{(\theta+u_I^\top\gamma_{i,I})Z^i}-1)\kappa_{i,j}=0,\quad j=1,\dots,m, \tag{15}$$

where $\gamma_{i,j}$ denotes the j-th component of γ_i and $\gamma_{i,I}=(\gamma_{i,1},\dots,\gamma_{i,m})$. Obviously, $\phi=\phi(\theta)$ is directly computable from $u=u(\theta)$ by (14). As a result, we will focus on the system of equations (15).

We need a solution to (15) that will make ϕ, computed from (14), is, in fact, the limiting CGF of $V(t)$. Hence we expect that $\phi(0)=0$. Note that $(\theta,u_I)=(0,\mathbf{0})$ satisfies (15), and that $\phi(0)=0$ if $u_I(0)=\mathbf{0}$. Therefore it is plausible that the appropriate solution branch $u_I(\theta)$ to (15) ought to satisfy $u_I(0)=\mathbf{0}$. To facilitate the analysis of the Equations (15), define $F_j(\theta,v)\colon\mathbb{R}\times\mathbb{R}^m\to\mathbb{R}$ as follows:

$$F_j(\theta,v)=\sum_{i=1}^{m}v_i\beta_{i,j}-\tfrac{1}{2}\alpha^j_{j,j}v_j^2-\sum_{i=1}^{n}(\mathbb{E}\,e^{(\theta+v^\top\gamma_{i,I})Z^i}-1)\kappa_{i,j}, \tag{16}$$

so that $F_j(\theta, u_I)$ equals the left-hand side (LHS) of the j-th equation of (15). Set $F(\theta, v) = (F_1(\theta, v), \ldots, F_m(\theta, v))\colon \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$. Then,

$$\frac{\partial F_j}{\partial v_l} = \beta_{l,j} - \sum_{i=1}^{n} \kappa_{i,j} \gamma_{i,l} \mathbb{E}(Z^i e^{(\theta + v^\top \gamma_{i,I}) Z^i}), \quad 1 \le l \ne j \le m$$

$$\frac{\partial F_j}{\partial v_j} = \beta_{j,j} - \alpha^j_{j,j} v_j - \sum_{i=1}^{n} \kappa_{i,j} \gamma_{i,j} \mathbb{E}(Z^i e^{(\theta + v^\top \gamma_{i,I}) Z^i}).$$

Let $\mathcal{J}(\theta, v) \triangleq (\partial F_j / \partial v_l)_{1 \le j, l \le m}$ denote the Jacobian matrix of F regarding v. Then

$$\mathcal{J}(\theta, v)^\top = \beta_{I,I} - \operatorname{diag}((\alpha^1_{1,1} v_1, \ldots, \alpha^m_{m,m} v_m)) - \sum_{i=1}^{n} \mathbb{E}(Z^i e^{(\theta + v^\top \gamma_{i,I}) Z^i}) \gamma_{i,I} \kappa_{i,I}^\top. \tag{17}$$

Therefore $\mathcal{J}(\theta, v)$ is a Z-matrix by part (I) of Assumption 1. Further, it follows from part (III) of Assumption 1 that

$$\mathcal{J}(0, \mathbf{0})^\top = \beta_{I,I} - \sum_{i=1}^{n} \mathbb{E}(Z^i) \gamma_{i,I} \kappa_{i,I}^\top \tag{18}$$

is an *M-matrix*, and thus is nonsingular (see, for example, Berman and Plemmons [7]) for the definition of M-matrices. Since $F(0, \mathbf{0}) = \mathbf{0}$, it then follows from the Implicit Function Theorem that, for any $\theta \in \mathbb{R}$ in a neighborhood of 0, there exists a unique $u_I^* = u_I^*(\theta)$ in the neighborhood of the origin in $\mathbb{R}^m$ such that $u_I^*(0) = \mathbf{0}$ and $F(\theta, u_I^*) = \mathbf{0}$. Moreover, letting $\mathcal{D}_{u_I^*}$ denote the maximal interval of existence of $u_I^*(\theta)$, including 0, the Implicit Function Theorem implies that $\mathcal{D}^o_{u_I^*} = (\underline{\theta}, \bar{\theta})$, where

$$\bar{\theta} \triangleq \min\{\theta \in \mathcal{D}_{u_I^*} \cap \mathbb{R}_+\colon \mathcal{J}(\theta, u_I^*(\theta)) \text{ is singular}\} \tag{19}$$

$$\underline{\theta} \triangleq \max\{\theta \in \mathcal{D}_{u_I^*} \cap \mathbb{R}_-\colon \mathcal{J}(\theta, u_I^*(\theta)) \text{ is singular}\} \tag{20}$$

with the convention that $\min\{\varnothing\} = \infty$ and $\max\{\varnothing\} = -\infty$. We have the following characterization of $\bar{\theta}$ and $\underline{\theta}$.

Proposition 6. *Suppose Assumptions 1 and 3 hold. Then $\underline{\theta} = -\infty$; moreover, $\bar{\theta} = \infty$ if $\kappa = \mathbf{0}$, and $\bar{\theta} < \infty$ otherwise.*

Proof. See §A.3 in the appendix. □

4.4. Limiting CGF of V. By Proposition 5, $M(t) = \exp[\theta V(t) - \phi t + u^\top (X(t) - X(0))]$ is a martingale if u and ϕ solve the Equations (14) and (15). It follows that

$$\mathbb{E} \exp(\theta V(t) - \phi t) = \mathbb{E}^Q \exp[-u(X(t) - X(0))],$$

where Q is the equivalent probability measure induced by $M(t)$, i.e., $(\mathrm{d}Q / \mathrm{d}\mathbb{P})|_{\mathcal{F}_t} = M(t)$. Hence, to show that ϕ is the limiting CGF of $V(t)$, it suffices to prove that

$$\mathbb{E}^Q \exp[-u^\top (X(t) - X(0))] = O(1). \tag{21}$$

As discussed in Remark 1, the subtlety lies in that there may exist multiple solutions $u_I(\theta)$ to the Equations (13) for a given θ. We will show that u^* as defined in §4.3 makes (21) valid, so that ϕ^*, solved from (14), is indeed the limiting CGF of $V(t)$.

To prove (21), it suffices to study the stochastic stability of $X(t)$ under the probability measure Q^*_θ, where Q^*_θ denotes the probability measure induced by $M^*_\theta(t) = \exp[\theta V(t) - \phi^*(\theta) t + u^*(\theta)^\top (X(t) - X(0))]$. It turns out that depending on whether θ is positive, we need different levels of stochastic stability of $X(t)$ under Q^*_θ. Note that, by Lemma 5, $u_I^*(\theta) \in \mathbb{R}^m_+$ for $\theta \ge 0$ and $u_I^*(\theta) \in \mathbb{R}^m_-$ for $\theta < 0$. Further, note that $X_I(t) \in \mathbb{R}^m_+$. Hence

$$\exp[-u^*(\theta)^\top (X(t) - X(0))] = \exp[-u_I^*(\theta)^\top (X_I(t) - X_I(0))]$$

is bounded for all t if $\theta \ge 0$, and unbounded if $\theta < 0$ unless $u^* \equiv \mathbf{0}$. Consequently, $X(t)$ being ergodic under Q^*_θ is sufficient for (21) if $\theta \ge 0$, while exponential ergodicity is required if $\theta < 0$ (see, for example, Meyn and Tweedie [43]). More detailed discussions will be provided in the appendix.

Let $\mathcal{D}_{\phi^*}$ denote the domain of

$$\phi^*(\theta) = u_I^*(\theta)^\top b_I + \sum_{i=1}^{n} \lambda_i (\mathbb{E}\, e^{(\theta + u_I^*(\theta)^\top \gamma_{i,I}) Z^i} - 1). \tag{22}$$

Proposition 7. *Under Assumptions 1 and 3,*

$$\mathbb{E}^{Q^*_\theta}\exp[-u^*(\theta)^\top(X(t)-X(0))]=O(1)$$

as $t\to\infty$ *for* $\theta\in\mathcal{D}_{\phi^*}$.

Proof. See §A.4. $\square$

Corollary 1. *Under Assumptions 1 and 3,*

$$\phi^*(\theta)=\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}\exp(\theta V(t))$$

for $\theta\in\mathcal{D}_{\phi^*}$.

Proof. Since $\mathbb{E}^{Q^*_\theta}\exp[-u^*(\theta)^\top(X(t)-X(0))]=O(1)$ for $\theta\in\mathcal{D}_{\phi^*}$ by Proposition 7, it follows that

$$e^{-\phi^*(\theta)t}\cdot\mathbb{E}\exp(\theta V(t))=\mathbb{E}^{Q^*_\theta}\exp[-u^*(\theta)^\top(X(t)-X(0))]=O(1),$$

yielding that

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}\exp(\theta V(t))=\phi^*(\theta)$$

for $\theta\in\mathcal{D}_{\phi^*}$. $\square$

4.5. LD for V**.** With the limiting CGF of $V(t)$ available, we can apply the Gärtner-Ellis theorem to establish the LD for $V(t)$. The key step in the derivation is to show that, for any $R>0$, there exists a unique θ_R such that $\phi^{*\prime}(\theta_R)=R$, or equivalently that ϕ^* is *steep* (see, for example, Dembo and Zeitouni [21]). The details are provided in the appendix.

Theorem 2. *Let r be the equilibrium mean of $V(t)$ given in Theorem 1. Under Assumptions 1 and 3,*

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{P}(V(t)\ge Rt)=-\mathcal{I}(R)$$

for $R>r$, whereas

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{P}(V(t)\le Rt)=-\mathcal{I}(R)$$

for $0<R<r$, where $\mathcal{I}(R)=\theta^*R-\phi^*(\theta^*)$, and θ^* uniquely solves $\phi^{*\prime}(\theta^*)=R$.

Proof. See §A.5. $\square$

5. Extensions. Theorems 1 and 2 can be extended to a more general setting. In particular, letting $w\in\mathbb{R}^n$ be a *nonzero* vector, define $J=\sum_{i=1}^n w_iL_i$. We then have the following CLT and LD principle for J.

Theorem 3. *Under Assumptions 1 and 2,*

$$t^{-1/2}(J(t)-rt)\Rightarrow\mathcal{N}(0,\eta^2)$$

as $t\to\infty$, where

$$r=A^\top b+\sum_{i=1}^n\lambda_i\mathbb{E}(Z^i)(w_i+A^\top\gamma_i)$$

$$\eta^2=A^\top aA+C^\top\lambda+(A^\top\alpha A+C^\top\kappa)B,$$

$$A^\top=\left(\sum_{i=1}^n\mathbb{E}(Z^i)w_i\kappa_i^\top\right)\left(\beta-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_i\kappa_i^\top\right)^{-1},$$

$$B=\left(\beta-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_i\kappa_i^\top\right)^{-1}\left(b+\sum_{i=1}^n\lambda_i\mathbb{E}(Z^i)\gamma_i\right),$$

$$C_i=(w_i+A^\top\gamma_i)^2\,\mathbb{E}(Z^i)^2,\quad i=1,\ldots,n.$$

THEOREM 4. *Let $u^*(\theta)\colon \mathbb{R}\to\mathbb{R}^d$ be the implicit function defined as the unique solution branch with $u^*(0)=\mathbf{0}$ of the system of nonlinear equations*

$$\sum_{i=1}^{d} u_i\beta_{i,j} - \tfrac{1}{2}u^\top \alpha^j u - \sum_{i=1}^{n}(\mathbb{E}\, e^{(\theta w_i + u^\top \gamma_i)Z^i} - 1)\kappa_{i,j} = 0, \quad j = 1,\ldots,d.$$

Let $\mathcal{I}(R)=\theta^* R-\phi^*(\theta^*)$, and θ^* uniquely solves $\phi^{*\prime}(\theta^*)=R$, where

$$\phi^*(\theta) = u^*(\theta)^\top b + \sum_{i=1}^{n}\lambda_i(\mathbb{E}\, e^{(\theta w_i + u^*(\theta)^\top \gamma_i)Z^i} - 1).$$

Under Assumptions 1 and 3,

$$\lim_{t\to\infty}\frac{1}{t}\log \mathbb{P}(J(t)\ge Rt) = -\mathcal{I}(R), \quad \textit{for} \begin{cases} r<R<0, & \textit{if } w\in\mathbb{R}^n_-,\\ R>r, & \textit{otherwise}\end{cases}$$

and

$$\lim_{t\to\infty}\frac{1}{t}\log \mathbb{P}(J(t)\le Rt) = -\mathcal{I}(R), \quad \textit{for} \begin{cases} 0<R<r, & \textit{if } w\in\mathbb{R}^n_+,\\ R<r, & \textit{otherwise}.\end{cases}$$

The proofs of these results are very similar to those of Theorems 1 and 2. The only noteworthy difference is that, in proving the steepness of the function ϕ^*, which is essential for the LD principle, one needs to characterize the domain of the function u^*, whose form depends on the sign of w. For instance, provided that $\kappa\ne\mathbf{0}$, $\mathcal{D}_{u^*}$ is unbounded below and bounded above if $w\in\mathbb{R}^n_+$, consistent with Proposition 6, whereas it is bounded from both sides if w has mixed signs, i.e., there exist $w_i>0$ and $w_j<0$ for some i and j. We omit the details.

6. Efficient simulation: Importance sampling. In some applications such as the computation of risk measures for security portfolios, one requires accurate estimates of rare-event probabilities. Monte Carlo simulation can be used to estimate these probabilities. However, it is well known that the number of simulation trials required to achieve a prescribed relative error is roughly inversely proportional to the probability of interest. Hence, plain Monte Carlo (pMC) simulation is highly inefficient for estimating rare-event probabilities, essentially because the variance of the estimator is too large relative to the probability of interest. We develop a provably efficient IS scheme to address this issue when estimating the tail of $J(t)$. The LD analysis of §4 guides the design of an appropriate change of measure.

Suppose we are interested in computing $\mathbb{P}(J(t)>Rt)$ for $R\in(r,0)$ if $w\in\mathbb{R}^n_-$ and $R\in(r,\infty)$ otherwise. (The left-tail $\mathbb{P}(J(t)<Rt)$ can be treated in the same fashion.) The LD Theorem 4 implies that $\mathbb{P}(J(t)>Rt)$ decays to 0 exponentially fast as $t\to\infty$. Hence the number of pMC trials required to achieve a given relative precision grows exponentially in t. We design an IS scheme in which the number of simulation trials grows subexponentially in t.

Given the key role θ^* plays in the logarithmic asymptotics of Theorem 4, it is natural to consider an IS estimator associated with the equivalent measure $Q^*_{\theta^*}$ induced by the martingale $M^*_{\theta^*}$. More specifically, consider the IS estimator

$$\begin{aligned} H(t) &\triangleq M^*_{\theta^*}(t)^{-1}\mathbb{1}(J(t)\ge Rt)\\ &= \exp\bigl[-\theta^* J(t) + \phi^*(\theta^*)t - u^*(\theta^*)^\top (X(t)-X(0))\bigr]\mathbb{1}(J(t)\ge Rt). \end{aligned} \tag{23}$$

Note that by Girsanov's theorem, under Q^*_θ, the process X satisfies the SDE (1) with parameters $(a,\alpha,b,\beta^*,\lambda^*,\kappa^*)$ and measure φ^*_i, where

$$\begin{aligned} \lambda^*_i &= \lambda_i\int_{\mathbb{R}_+} e^{(\theta^* w_i + u^*(\theta^*)^\top\gamma_i)z}\varphi_i(\mathrm{d}z),\\ \kappa^*_i &= \kappa_i\int_{\mathbb{R}_+} e^{(\theta^* w_i + u^*(\theta^*)^\top\gamma_i)z}\varphi_i(\mathrm{d}z),\\ \beta^* &= \begin{pmatrix} \beta_{I,I}-\operatorname{diag}(\alpha^1_{1,1}u^*_1(\theta^*),\ldots,\alpha^m_{m,m}u^*_m(\theta^*)), & \mathbf{0},\\ \beta_{J,I}, & \beta_{J,J}, \end{pmatrix},\\ \varphi^*_i(\mathrm{d}z) &= \frac{e^{(\theta^* w_i + u^*(\theta^*)^\top\gamma_i)z}\varphi_i(\mathrm{d}z)}{\int_{\mathbb{R}_+} e^{(\theta^* w_i + u^*(\theta^*)^\top\gamma_i)y}\varphi_i(\mathrm{d}y)}. \end{aligned} \tag{24}$$

THEOREM 5. *Under Assumptions 1 and 3, the IS estimator (23) is asymptotically optimal, i.e.,*

$$\lim_{t\to\infty}\frac{\log \mathbb{E}^{Q^*_{\theta^*}} H(t)^2}{2\log \mathbb{E}^{Q^*_{\theta^*}} H(t)} = 1. \tag{25}$$

PROOF. Note that

$$\begin{aligned}\mathbb{E}^{Q^*_{\theta^*}} H(t)^2 &= \mathbb{E}^{Q^*_{\theta^*}} \exp\{-2[\theta^* J(t) - \phi^*(\theta^*)t + u^*(\theta^*)^\top (X(t)-X(0))]\}\mathbb{1}(J(t)\ge Rt)\\ &\le \mathbb{E}^{Q^*_{\theta^*}} \exp\{-2[\theta^* Rt - \phi^*(\theta^*)t + u^*(\theta^*)^\top (X(t)-X(0))]\}\\ &= e^{-2\mathcal{I}(R)t}\cdot \mathbb{E}^{Q^*_{\theta^*}} \exp[-2u^*(\theta^*)^\top (X(t)-X(0))],\end{aligned}$$

where $\mathcal{I}(R) = \theta^*\cdot R - \phi^*(\theta^*)$. It follows that

$$\frac{\log \mathbb{E}^{Q^*_{\theta^*}} H(t)^2}{\log \mathbb{E}^{Q^*_{\theta^*}} H(t)} \ge \frac{-2\mathcal{I}(R)t + \log \mathbb{E}\exp[-2u^*(\theta^*)^\top (X(t)-X(0))]}{\log \mathbb{P}(L(t)\ge Rt)}.$$

An argument similar to the one used in the proof of Proposition 7 shows that $\mathbb{E}^{Q^*_{\theta^*}} \exp[-2u^*(\theta^*)^\top (X(t) - X(0))] = O(1)$ as $t\to\infty$. Hence

$$\underline{\lim}_{t\to\infty}\frac{\log \mathbb{E}^{Q^*_{\theta^*}} H(t)^2}{2\log \mathbb{E}^{Q^*_{\theta^*}} H(t)} \ge 1$$

by Theorem 4. On the other hand, note that $\mathbb{E}^{Q^*_{\theta^*}} H(t)^2 \ge (\mathbb{E}^{Q^*_{\theta^*}} H(t))^2$ by Jensen's inequality, from which it follows that

$$\overline{\lim}_{t\to\infty}\frac{\log \mathbb{E}^{Q^*_{\theta^*}} H(t)^2}{2\log \mathbb{E}^{Q^*_{\theta^*}} H(t)} \le 1,$$

completing the proof. □

7. Numerical experiments.

This section provides numerical results for the model specification of Example 1. The components of $X = (X_1, \ldots, X_d)$ satisfy the AJD

$$\mathrm{d}X_j(t) = (b_j - \beta_j X_j(t))\,\mathrm{d}t + \sigma_j\sqrt{X_j(t)}\,\mathrm{d}W_j(t) + \sum_{i=1}^n \delta_i\,\mathrm{d}L_i(t), \quad j = 1,\ldots,d, \tag{26}$$

where $b_j, \beta_j, \sigma_j, \delta_i > 0$. The jump intensity is $\Lambda_i(X(t)) = \lambda_i + \kappa_i X_i(t)$ for some $\lambda_i, \kappa_i > 0$. We take $d = 3$, $\beta = (2.0, 2.1, 2.2)$, $b = (6.0, 6.1, 6.2)$, $\sigma = (0.5, 0.6, 0.7)$, $\delta = (0.2, 0.3, 0.4)$, $\lambda = (0, 0, 0)$, $\kappa = (1.0, 1.1, 1.2)$, and set φ_i as the exponential distribution with mean 1 for $i = 1, 2, 3$. We consider the two choices $w = (1, 1, 1)$ and $w = (1, -1, 1)$.

7.1. Gaussian approximation.

The CLT 3 implies the following Gaussian approximation:

$$J(t) \overset{\mathcal{D}}{\approx} rt + \eta\sqrt{t}\cdot \mathcal{N}(0,1)$$

for large t, where $\overset{\mathcal{D}}{\approx}$ denotes approximate equality in distribution. To illustrate the quality of the approximation, we compare the distribution of $(J(t) - rt)/(\eta\sqrt{t})$ with a standard normal distribution for each of several values $t > 0$. The inverse Fourier transform is used to compute the distribution of $(J(t) - rt)/(\eta\sqrt{t})$ (see Errais et al. [27] for details on computing the Fourier transform and Abate and Whitt [1] for the numerical inversion). Tables 1 and 2 report the results. Figure 1 shows the corresponding density functions. While the Gaussian approximation performs quite well in the center of the distribution, there is significant error in the tail.

7.2. Efficient simulation.

We now show the asymptotic optimality of the IS estimator (23). Its implementation is briefly discussed below. We first compute the tilting parameter θ^* by solving the set of nonlinear equations in Theorem 4 and compute the new set of parameters (24). Then we generate samples of $(X(t), J(t))$ under the distribution $Q^*_{\theta^*}$, i.e., the SDE (26) with parameters (24). To that end, we iteratively simulate the sequence of jump times $\tau_1 < \tau_2 < \cdots < \tau_K$ with $\tau_{K-1} < t \le \tau_K$. For each iteration, given τ_{k-1} and $X(\tau_{k-1})$, we simulate the next jump time τ_k and identify the source of the jump, say, L_{i_k}. Given τ_k and $X(\tau_{k-1})$, it is easy to simulate $X(\tau_k-)$ since X_j behaves as a CIR process between jump times, i.e.,

$$\mathrm{d}X_j(t) = (b_j - \beta_j X_j(t))\,\mathrm{d}t + \sigma_j\sqrt{X_j(t)}\,\mathrm{d}W_j(t), \quad t\in[\tau_{k-1}, \tau_k),$$

Table 1. Gaussian approximation.

l	$t=1$	$t=5$	$t=10$	$t=50$	$t=100$	Std. norm.
			$\mathbb{P}((J(t)-rt)/(\eta\sqrt{t})<l)$			
−3.0	2.036E−03	2.153E−03	2.257E−03	2.859E−03	3.101E−03	1.350E−03
−2.5	1.638E−03	1.916E−03	3.094E−03	6.069E−03	6.780E−03	6.210E−03
−2.0	1.2417E−03	8.639E−03	1.537E−02	2.218E−02	2.313E−02	2.275E−02
−1.5	1.493E−02	6.275E−02	7.081E−02	7.243E−02	7.158E−02	6.681E−02
−1.0	1.598E−01	2.155E−01	2.036E−01	1.810E−01	1.749E−01	1.587E−01
−0.5	5.263E−01	4.434E−01	4.038E−01	3.510E−01	3.387E−01	3.085E−01
			$\mathbb{P}((J(t)-rt)/(\eta\sqrt{t})>l)$			
0.0	8.000E−01	6.655E−01	6.157E−01	5.511E−01	5.361E−01	5.000E−01
0.5	7.112E−02	1.739E−01	2.143E−01	2.670E−01	2.793E−01	3.085E−01
1.0	2.330E−02	8.047E−02	1.051E−01	1.361E−01	1.431E−01	1.587E−01
1.5	8.015E−03	3.436E−02	4.643E−02	5.999E−02	6.264E−02	6.681E−02
2.0	3.737E−03	1.440E−02	1.936E−02	2.360E−02	2.409E−02	2.275E−02
2.5	2.821E−03	6.710E−03	8.448E−03	9.166E−03	9.024E−03	6.210E−03
3.0	2.864E−03	4.135E−03	4.636E−03	4.471E−03	4.298E−03	1.350E−03

Note. Distribution function of $(J(t)-rt)/(\eta\sqrt{t})$ with $w=(1,1,1)$.

and the marginal distribution of the CIR process is noncentral chi-squared after proper scaling (see Glasserman [37]). We then simulate $X_j(\tau_k)$ via $X_j(\tau_k-)+\delta_{i_k}Z^{i_k}$ for all $j=1,\ldots,d$ and simulate $J(\tau_k)$ via $J(\tau_k)=J(\tau_{k-1})+w_{i_k}Z^{i_k}$, where Z^{i_k} is drawn from the distribution $\varphi^*_{i_k}$. In the last iteration, we simulate $X(t)$ given $X(\tau_{K-1})$. We refer the reader to §5.2 of Giesecke et al. [33] for details of the above iterative simulation approach.

We estimate $\mathbb{P}(J(t)<Rt)$ for $w=(1,1,1)$ and $\mathbb{P}(J(t)>Rt)$ for $w=(1,-1,1)$ for different values of $t>0$. When comparing the computational costs of the pMC and the IS, we assume the confidence interval (CI) is constructed at the 95% level, and the target relative precision is 10%, namely, the half-length of the CI should be within 10% of the estimated value. More specifically, let p denote the probability to be estimated, v denote the variance of the estimator, and m denote the number of samples to be generated. Then the (approximate) 95% CI is $p\pm 1.96\sqrt{v/n}$, and hence we require $1.96\sqrt{v/n}\le 0.1p$, which yields

$$n\ge \frac{19.6^2 v}{p^2}.$$

We first use a relatively large sample size to estimate p and v, then estimate the necessary sample sizes to achieve the target relative precision for both the pMC and IS estimators, and finally, estimate the CPU time

Table 2. Gaussian approximation.

l	$t=1$	$t=5$	$t=10$	$t=50$	$t=100$	Std. norm.
			$\mathbb{P}((J(t)-rt)/(\eta\sqrt{t})<l)$			
−3.0	2.400E−03	2.509E−03	2.664E−03	3.129E−03	3.292E−03	1.350E−03
−2.5	3.025E−03	4.058E−03	4.941E−03	6.709E−03	7.167E−03	6.210E−03
−2.0	8.179E−03	1.544E−02	1.862E−02	2.249E−02	2.318E−02	2.275E−02
−1.5	3.46E−02	6.226E−02	6.703E−02	6.944E−02	6.932E−02	6.681E−02
−1.0	1.390E−01	1.819E−01	1.796E−01	1.706E−01	1.677E−01	1.587E−01
−0.5	3.923E−01	3.779E−01	3.600E−01	3.329E−01	3.260E−01	3.085E−01
			$\mathbb{P}((J(t)-rt)/(\eta\sqrt{t})>l)$			
0.0	6.690E−01	5.969E−01	5.684E−01	5.304E−01	5.215E−01	5.000E−01
0.5	1.511E−01	2.237E−01	2.500E−01	2.835E−01	2.911E−01	3.085E−01
1.0	6.184E−02	1.085E−01	1.254E−01	1.456E−01	1.499E−01	1.587E−01
1.5	2.410E−02	4.734E−02	5.552E−02	6.385E−02	6.533E−02	6.681E−02
2.0	9.804E−03	1.956E−02	2.262E−02	2.461E−02	2.470E−02	2.275E−02
2.5	4.903E−03	8.487E−03	9.345E−03	9.220E−03	8.993E−03	6.210E−03
3.0	3.478E−03	4.641E−03	4.775E−03	4.361E−03	4.192E−03	1.350E−03

Note. Distribution function of $(J(t)-rt)/(\eta\sqrt{t})$ with $w=(1,-1,1)$.

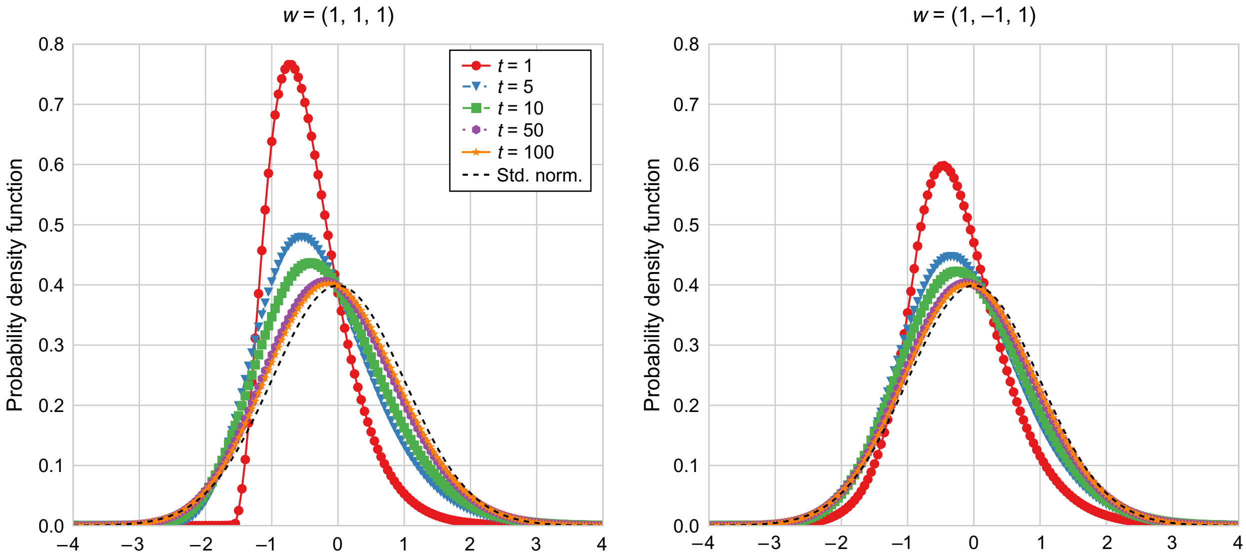

FIGURE 1. (Color online) Central limit convergence.

For different values of $t > 0$, the density function of $(J(t) - rt)/(\eta\sqrt{t})$ is computed via the inverse Fourier transform. The asymptotic mean r and the asymptotic variance η^2 can be calculated analytically.

used to complete the necessary sample sizes. The simulation algorithm is written in C with the random number generator from GNU Scientific Library (GSL-1.16). It is run on a Mac computer with OS X 10.8.4, processor 3.4 GHz Intel Core i7, and memory 32 GB 1333 MHz DDR3. The numerical results are reported in Table 3 for the case $w = (1, 1, 1)$ and $\mathbb{P}(J(t) < Rt)$, and Table 4 for $w = (1, -1, 1)$ and $\mathbb{P}(J(t) > Rt)$.

8. Conclusions. Affine point processes have broad applications in finance, economics, and many other areas due to their model flexibility and analytical tractability. In this paper, we have studied the long-term asymptotic behaviors of this type of processes. In particular, we have established a CLT and an LD principle to respectively characterize their typical and atypical behaviors. The tractable affine structure permits us to calculate the key quantities such as the asymptotic mean, asymptotic variance, and the LDs rate function explicitly. Furthermore, applying the LD result, we have developed an asymptotically optimal IS algorithm for simulating certain rare events associated with the affine point process. Numerical experiments illustrated the Gaussian approximation induced by the CLT and the efficiency of the IS estimator.

TABLE 3. Asymptotic optimality of the IS estimator.

t	p (pMC)	T (pMC)	p (IS)	v (IS)	T (IS)	VR	Log ratio
1	4.340E−01	0.03993	4.372E−01	9.365E−02	0.1119	2.627E+00	0.7590
2	2.780E−01	0.12670	2.820E−01	6.834E−02	0.05929	2.963E+00	0.7550
5	8.825E−02	0.54910	9.078E−02	1.207E−02	0.19180	6.837E+00	0.8120
10	1.805E−02	3.22100	1.685E−02	5.772E−04	0.35800	2.870E+01	0.8641
20	9.289E−04	136.40000	9.097E−04	2.330E−06	0.53690	3.900E+02	0.9044
30	4.624E−05	3,892	4.565E−05	8.042E−09	0.8159	5.676E+03	0.9209
50	n/a	n/a	1.651E−07	1.327E−13	1.53200	1.244E+06	0.9433
100	n/a	n/a	1.469E−13	1.690E−25	3.90900	8.692E+11	0.9631
200	n/a	n/a	1.738E−25	3.149E−49	9.79600	5.521E+23	0.9786
500	n/a	n/a	6.968E−61	8.095E−120	43.73000	8.607E+58	0.9896
1,000	n/a	n/a	7.744E−120	1.529E−237	121.90000	5.066E+117	0.9940

Notes. Estimation of $\mathbb{P}(J(t) < Rt)$ with $w = (1, 1, 1)$ and $R = 0.6r = 0.6 \times 18.28 = 10.97$. p denotes the probability estimated via pMC or IS; T denotes the elapsed CPU time (seconds); v denotes the estimated variance; "VR" denotes the variance reduction ratio; "log ratio" denotes the ratio (25). Since the pMC estimator is $\mathbb{I}(J(t) < Rt)$, its variance is simply $p(1-p)$, so we do not report it here. The CPU time is estimated using the estimated sample size required to achieve the 10% relative precision in constructing a 95% CI. Due to the prohibitively long CPU time, we do not run the pMC for large values of t.

TABLE 4. Asymptotic optimality of the IS estimator.

t	p (pMC)	T (pMC)	p (IS)	v (IS)	T (IS)	VR	Log ratio
1	1.80E−01	0.1501	1.806E−01	8.775E−02	0.1003	1.686E+00	0.6185
2	1.640E−01	0.3137	1.645E−01	6.414E−02	0.1794	2.143E+00	0.6634
5	1.193E−01	0.4655	1.187E−01	3.184E−02	0.3346	3.285E+00	0.7228
10	7.000E−02	0.9870	7.045E−02	1.175E−02	0.4123	5.573E+00	0.7712
20	2.427E−02	4.8390	2.583E−02	1.878E−03	0.5763	1.340E+01	0.8169
50	1.684E−03	197.5000	1.731E−03	1.171E−05	1.5090	1.476E+02	0.8749
75	2.065E−04	2,368	1.894E−04	1.647E−07	2.7780	1.150E+03	0.8996
100	n/a	n/a	2.401E−05	2.989E−09	3.6590	8.033E+03	0.9144
200	n/a	n/a	6.178E−09	2.723E−16	10.3600	2.269E+07	0.9446
500	n/a	n/a	1.415E−19	2.289E−37	39.3400	6.182E+17	0.9710
1,000	n/a	n/a	3.615E−37	2.204E−72	112.1000	1.640E+35	0.9828
2,000	n/a	n/a	4.116E−72	4.120E−142	351.9000	9.990E+69	0.9902

Note. Estimation of $\mathbb{P}(J(t) > Rt)$ with $w = (1, -1, 1)$ and $R = 1.5r = 1.5 \times 6.081 = 9.152$.

Acknowledgments. The authors would like to thank the anonymous referee for careful reading of our manuscript and providing us with constructive comments. The research is partially supported by the Hong Kong Research Grants Council [Direct Allocated Grant DAG12EG01].

Appendix A. Additional technical results and proofs.

A.1. Proof of Proposition 3. We first need the following two results regarding the stochastic stability of the AJD process X.

PROPOSITION 8. *Suppose Assumption 1 holds. Suppose also that either $\mathbb{E} Z^i < \infty$ or $\kappa_i = \mathbf{0}$ for all $i = 1, \ldots, n$. Then X is a nonexplosive process.*

PROOF. See Lemma 9.2 of Duffie et al. [23]. □

PROPOSITION 9. *Under Assumptions 1 and 2, X has a unique stationary distribution π. Moreover, $\mathbb{P}(X(t) \in \cdot) \to \pi$ in total variation as $t \to \infty$. Also,*

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t X(s)\,\mathrm{d}s = \mathbb{E}_\pi X(0) = \left(\beta - \sum_{i=1}^n \mathbb{E}(Z^i)\gamma_i \kappa_i^\top\right)^{-1} \left(b + \sum_{i=1}^n \lambda_i \mathbb{E}(Z^i)\gamma_i\right) \quad a.s. \tag{27}$$

PROOF. See Chapter 2 of Zhang [48]. □

We also need the following lemma, which states that the number of jumps is roughly proportional to the length of the time interval.

LEMMA 1. *Let $N_i(t) = \int_0^t \int_0^\infty N_i(\mathrm{d}s, \mathrm{d}z)$. Under Assumptions 1 and 2,*

$$\lim_{j\to\infty} j^{-1}\, \mathbb{E} N_i(jT) = T\, \mathbb{E}_\pi \Lambda_i(X(0))$$

for any $T > 0$ and $i = 1, \ldots, n$, where π is stationary distribution of $X(t)$.

PROOF. Fix $T > 0$ and $i = 1, \ldots, n$. It follows from Proposition 9 that $\mathbb{E} X(t) \to \mathbb{E}_\pi X(0)$ as $t \to \infty$. Hence, for any $\epsilon > 0$, there exists $j_0 > 0$ such that

$$\mathbb{E}\, \Lambda_i(X(jT)) < \mathbb{E}_\pi \Lambda_i(X(0)) + \epsilon$$

for all $j > j_0$ since $\Lambda_i(x)$ is affine in x; in addition, $\mathbb{E} \int_0^t \Lambda_i(X(s))\,\mathrm{d}s < \infty$ for all $t > 0$. Moreover, $X(t)$ is nonexplosive by Proposition 8, and thus $N_i(t)$ is nonexplosive, from which we conclude that $N_i(t) - \int_0^t \Lambda_i(X(s))\,\mathrm{d}s$ is a martingale (see Theorems T8 and T9 of Brémaud [11]). Therefore

$$\begin{aligned} j^{-1}\, \mathbb{E} N_i(jT) &= j^{-1}\, \mathbb{E} \int_0^{jT} \Lambda_i(X(s))\,\mathrm{d}s \\ &= j^{-1} \int_0^{jT} \mathbb{E}\, \Lambda_i(X(s))\,\mathrm{d}s \quad \text{(by Fubini's theorem)} \\ &\le j^{-1} \int_0^{j_0 T} \mathbb{E}\, \Lambda_i(X(s))\,\mathrm{d}s + j^{-1}(\mathbb{E}_\pi \Lambda_i(X(0)) + \epsilon)(jT - j_0 T). \end{aligned}$$

It follows that

$$\overline{\lim_{j\to\infty}}\, j^{-1}\, \mathbb{E} N_i(jT) \le T(\mathbb{E}_\pi \Lambda_i(X(0)) + \epsilon).$$

Likewise, we can show

$$\varliminf_{j\to\infty} j^{-1}\,\mathbb{E}\, N_i(jT) \geq T(\mathbb{E}_\pi\, \Lambda_i(X(0)) - \epsilon).$$

Sending $\epsilon \downarrow 0$ completes the proof. $\square$

Proof of Proposition 3. It follows from (3) and (4) that

$$U(t) = \sum_{i=1}^n \int_0^t \int_{\mathbb{R}_+} (1 + A^\top \gamma_i) z \tilde{N}_i(\mathrm{d}s, \mathrm{d}z) + \int_0^t A^\top \sigma(X(s))\, \mathrm{d}W(s).$$

Therefore the pure jump part of U is

$$\sum_{i=1}^n \int_0^t \int_{\mathbb{R}_+} (1 + A^\top \gamma_i) z N_i(\mathrm{d}s, \mathrm{d}z),$$

where $g_i(z) \triangleq (1 + A^\top \gamma_i) z$.

Let $(Z_l^i\colon l \geq 1)$ be a sequence of iid random variable's with common distribution φ_i. Note that

$$\sup_{0\leq t\leq jT} j^{-1} |U(t) - U(t-)|^2 = \sup_{1\leq i\leq n} \sup_{1\leq l\leq N_i(jT)} j^{-1} g_i(Z_l^i)^2.$$

Hence

$$\begin{aligned}
\mathbb{P}\Big(\sup_{0\leq t\leq jT} j^{-1}|U(t) - (t-)|^2 > x\Big) &= \mathbb{E}\Big[\mathbb{P}\Big(\sup_{1\leq i\leq n} \sup_{1\leq l\leq N_i(jT)} j^{-1} g_i(Z_l^i)^2 > x \,\Big|\, N_i(jT),\ i = 1, \ldots, n\Big)\Big] \\
&\leq \mathbb{E}\Big[\sum_{i=1}^n \sum_{l=1}^{N_i(jT)} \mathbb{P}(j^{-1} g_i(Z_l^i)^2 > x \,\Big|\, N_i(jT),\ i = 1, \ldots, n)\Big] \\
&= \mathbb{E}\Big[\sum_{i=1}^n \sum_{l=1}^{N_i(jT)} \mathbb{P}(j^{-1} g_i(Z_l^i)^2 > x)\Big] \\
&= \sum_{i=1}^n \mathbb{E}\, N_i(jT)\, \mathbb{P}(j^{-1} g_i(Z_1^i)^2 > x).
\end{aligned}$$

It follows that, for any $\delta > 0$,

$$\begin{aligned}
\mathbb{E} \sup_{0\leq t\leq jT} j^{-1}|U(t) - U(t-)|^2 &\leq \delta + \int_\delta^\infty \mathbb{P}\Big(\sup_{0\leq t\leq jT} j^{-1}|U(t) - U(t-)|^2 > x\Big)\, \mathrm{d}x \\
&\leq \delta + \int_\delta^\infty \sum_{i=1}^n \mathbb{E}\, N_i(jT)\, \mathbb{P}(j^{-1} g_i(Z_1^i)^2 > x)\, \mathrm{d}x \\
&= \delta + \sum_{i=1}^n \mathbb{E}\, N_i(jT) \int_\delta^\infty \mathbb{P}(g_i(Z_1^i)^2 > jx)\, \mathrm{d}x \\
&\leq \delta + \sum_{i=1}^n \mathbb{E}\, N_i(jT) \int_\delta^\infty (jx)^{-(1+\epsilon/2)}\, \mathbb{E}\, g_i(Z_1^i)^{2+\epsilon}\, \mathrm{d}x \\
&= \delta + 2\epsilon^{-1}\delta^{-(\epsilon/2)} \sum_{i=1}^n j^{-(1+\epsilon/2)}\, \mathbb{E}\, N_i(jT)\, \mathbb{E}\, g_i(Z_1^i)^{2+\epsilon},
\end{aligned}$$

where the Markov inequality is applied in the penultimate step. Note that $\mathbb{E}\, g_i(Z_1^i)^{2+\epsilon} < \infty$ for $i = 1, \ldots, n$ by Assumption 2. It then follows immediately from Proposition 1 that

$$\varlimsup_{j\to\infty} \mathbb{E} \sup_{0\leq t\leq jT} j^{-1}|U(t) - U(t-)|^2 \leq \delta.$$

Now, sending $\delta \downarrow 0$ concludes the proof. $\square$

A.2. Proof of Proposition 5. Let (θ, u) be a solution to (13). Define $\hat{\lambda} \in \mathbb{R}_+^n$ and $\hat{\kappa} \in \mathbb{R}_+^{n\times d}$ such that

$$\hat{\lambda}_i = \lambda_i \int_{\mathbb{R}_+} e^{(\theta + u^\top \gamma_i) z} \varphi_i(\mathrm{d}z) \tag{28}$$

$$\hat{\kappa}_i = \kappa_i \int_{\mathbb{R}_+} e^{(\theta + u^\top \gamma_i) z} \varphi_i(\mathrm{d}z), \tag{29}$$

where $\kappa_i^\top$ and $\hat{\kappa}_i^\top$ is the i-th row of κ and $\hat{\kappa}$, respectively, for $i = 1, \ldots, n$. Moreover, define $\hat{\beta} \in \mathbb{R}^{d\times d}$ such that

$$\hat{\beta}_j = \beta_j - \alpha^j u, \quad j = 1, \ldots, n$$

where β_j and $\hat{\beta}_j$ are the j-th column of β and $\hat{\beta}$, respectively, for $j = 1, \ldots, d$. Note that $\alpha^j = \mathbf{0}$ and $u_j = 0$ for $j = m+1, \ldots, n$ from which it follows immediately that

$$\hat{\beta} = \begin{pmatrix} \beta_{I,I} - \operatorname{diag}(\alpha^1_{1,1}u_1, \ldots, \alpha^m_{m,m}u_m) & \mathbf{0} \\ \beta_{J,I} & \beta_{J,J} \end{pmatrix}. \tag{30}$$

Hence $\hat{\beta}_{I,I} = \beta_{I,I} - \operatorname{diag}(\alpha^1_{1,1}u_1, \ldots, \alpha^m_{m,m}u_m)$ has nonpositive off-diagonal elements since β has nonpositive off-diagonal elements. Therefore we have the following proposition.

Proposition 10. *Under Assumptions 1 and 3, the parameters $(a, \alpha, b, \hat{\beta}, \hat{\lambda}, \hat{\kappa}, \gamma)$ are admissible, where $\hat{\lambda}$, $\hat{\kappa}$, and $\hat{\beta}$ are respectively defined by (28), (29), and (30).*

Note that, with (θ, u) at hand, we can rewrite $M(t)$ as

$$M(t) = 1 + \int_0^t M(s-)u^\top \sigma(X(s))\,\mathrm{d}W(s) + \int_0^t \int_{\mathbb{R}^d_+} M(s-)\sum_{i=1}^n (e^{(\theta+u^\top\gamma_i)z} - 1)\tilde{N}_i(\mathrm{d}s, \mathrm{d}z),$$

or equivalently,

$$\mathrm{d}M(t) = M(t-)\left[u^\top \sigma(X(t))\,\mathrm{d}W(t) + \sum_{i=1}^n \int_{\mathbb{R}_+} (e^{(\theta+u^\top\gamma_i)z} - 1)\tilde{N}_i(\mathrm{d}t, \mathrm{d}z)\right]. \tag{31}$$

Hence

$$M(t) = \exp\biggl(\int_0^t u^\top \sigma(X(s))\,\mathrm{d}W(s) - \tfrac{1}{2}\int_0^t u^\top \sigma(X(s))\sigma^\top(X(s))u\,\mathrm{d}s + \sum_{i=1}^n \int_0^t \int_{\mathbb{R}_+} (\theta + u^\top\gamma_i)zN_i(\mathrm{d}s, \mathrm{d}z) - \sum_{i=1}^n \int_0^t \int_{\mathbb{R}_+} (e^{(\theta+u^\top\gamma_i)z} - 1)\varphi_i(\mathrm{d}z)\Lambda_i(X(s))\,\mathrm{d}s\biggr). \tag{32}$$

Proof of Proposition 5. It follows from (31) and (32) that $M(t)$ is a positive local martingale, and thus a supermartingale. Consequently, it suffices to show that $\mathbb{E}\,M(T) = 1$. Consider a new set of parameters $(a, \alpha, b, \hat{\beta}, \hat{\lambda}, \hat{\kappa}, \gamma)$, defined via (28), (29), and (30). Moreover, let

$$\hat{\varphi}_i(\mathrm{d}z) = \frac{e^{(\theta+u^\top\gamma_i)z}\varphi_i(\mathrm{d}z)}{\int_{\mathbb{R}_+} e^{(\theta+u^\top\gamma_i)y}\varphi_i(\mathrm{d}y)} \tag{33}$$

for $i = 1, \ldots, n$. Set

$$\hat{\mu}(x) = b - \beta x + \sigma(x)\sigma(x)^\top u.$$

Since $u_J = \mathbf{0}$, $a_{I,I} = \mathbf{0}$, $a_{I,J} = \mathbf{0}$, and $a_{J,I} = \mathbf{0}$, it follows that

$$\sigma(x)\sigma^\top(x)u = \left(a + \sum_{j=1}^n x_j\alpha^j\right)u = \sum_{j=1}^n x_j\alpha^j u,$$

yielding that $\hat{\mu}(x) = b - \hat{\beta}x$.

Proposition 10 indicates that the parameters $(a, \alpha, b, \hat{\beta}, \hat{\lambda}, \hat{\kappa}, \gamma)$ are admissible. Hence we may consider an AJD $\hat{X}(t) \in \mathbb{R}^m_+ \times \mathbb{R}^{d-m}$ satisfying

$$\mathrm{d}\hat{X}(t) = \hat{\mu}(\hat{X}(t))\,\mathrm{d}t + \sigma(\hat{X}(t))\,\mathrm{d}W(t) + \sum_{i=1}^n \int_{\mathbb{R}_+} \gamma_i z\hat{N}_i(\mathrm{d}t, \mathrm{d}z), \tag{34}$$

where $\hat{N}_i(\mathrm{d}t, \mathrm{d}z)$ is a counting random measure on $[0, \infty) \times \mathbb{R}^d$ with compensator $\hat{\Lambda}_i(\hat{X}(t))\,\mathrm{d}t\,\hat{\varphi}_i(\mathrm{d}z)$, where $\hat{\Lambda}_i(x) = \hat{\lambda}_i + \hat{\kappa}_i^\top x$, for $i = 1, \ldots, n$.

For each $l \geq 1$, we define the stopping times

$$\tau_l = \inf\{t > 0\colon \|X(t)\| \geq l\} \qquad \text{and} \qquad \hat{\tau}_l = \inf\{t > 0\colon \|\hat{X}(t)\| \geq l\}.$$

Both $X(t)$ and $\hat{X}(t)$ are nonexplosive by Proposition 8. Therefore these stopping times satisfy

$$\lim_{l\to\infty} \mathbb{P}(\tau_l \geq T) = \lim_{l\to\infty} \mathbb{P}(\hat{\tau}_l \geq T) = 1. \tag{35}$$

For each l, let $X^{\tau_l}(t) = X(t \wedge \tau_l)$ be the stopped processes associated with $(\tau_l\colon l \geq 1)$ and similarly let $M^{\tau_l}(t) = M(t \wedge \tau_l)$. Note that

$$\begin{aligned}
&\mathbb{E}\exp\left(\frac{1}{2}\int_0^{t\wedge\tau_l} u^\top \sigma(X(s))\sigma^\top(X(s))u\,\mathrm{d}s + \sum_{i=1}^n \int_0^{t\wedge\tau_l} \int_{\mathbb{R}_+} f_i(z)\varphi_i(\mathrm{d}z)\Lambda_i(X(s))\,\mathrm{d}s\right) \\
&\quad = \mathbb{E}\exp\left(\frac{1}{2}\sum_{j=1}^m \alpha^j_{j,j}u_j^2 \int_0^{t\wedge\tau_l} X_j(s)\,\mathrm{d}s + \sum_{i=1}^n \mathbb{E}\,f_i(Z^i)\int_0^{t\wedge\tau_l} (\lambda_i + \kappa_i^\top X(s))\,\mathrm{d}s\right) \\
&\quad < \infty
\end{aligned}$$

where $f_i(z) = e^{(\theta+u^\top\gamma_i)z}((\theta + u^\top\gamma_i)z - 1) + 1$ for $i = 1, \ldots, n$ since $X^{\tau_l}(s)$ is bounded for $s \in [0, \tau_l)$.

It follows from (32) and Théorème IV.3 of Lépingle and Mémin [42] that $(M^{\tau_l}(t)\colon t\in[0,T])$ is a martingale. Hence, for each $l\geq 1$, $M^{\tau_l}(t)$ induces a probability measure Q^l equivalent to P defined by $(\mathrm{d}Q^l/\mathrm{d}\mathbb{P})|_{\mathcal{F}_t}=M^{\tau_l}(t)$ for $t\in[0,T]$. It follows from Girsanov's theorem that, for each $l\geq 1$,

$$W^l(t)=W(t)-\int_0^t\sigma^\top(X^{\tau_l}(s))u\,\mathrm{d}s$$

is a standard d-dimensional Brownian motion under Q^l. In addition, $N_i(\mathrm{d}t,\mathrm{d}z)$ has compensator $\hat{\Lambda}_i(X^{\tau_l}(t))\,\mathrm{d}t\,\hat{\varphi}_i(\mathrm{d}z)$ under Q^l for each $i=1,\dots,n$.

Note that we can rewrite the SDE (1) for $t\in[0,\tau^l)$ as

$$\mathrm{d}X(t)=\hat{\mu}(X(t))\,\mathrm{d}t+\sigma(X(t))\,\mathrm{d}W^l(t)+\sum_{i=1}^n\int_{\mathbb{R}_+}\gamma_i z\hat{N}_i(\mathrm{d}t,\mathrm{d}z). \tag{36}$$

By comparing (34) with (36), we conclude that $(X(t)\colon t\in[0,\tau_l))$ under Q^l has the same distribution as $(\hat{X}(t)\colon t\in[0,\hat{\tau}_l))$ under $\mathbb{P}$. Therefore, by (35),

$$\mathbb{E}\,M^{\tau_l}(T)\mathbb{1}_{\{\tau_l\geq T\}}=Q^l(\tau_l\geq T)=\mathbb{P}(\hat{\tau}_l\geq T)\to 1$$

as $l\to\infty$. Moreover, note that

$$\mathbb{E}\,M^{\tau_l}(T)\mathbb{1}_{\{\tau_l\geq T\}}=\mathbb{E}\,M(T)\mathbb{1}_{\{\tau_l\geq T\}}\to\mathbb{E}\,M(T)\mathbb{1}_{\{\tau_\infty\geq T\}}$$

as $l\to\infty$ by the Monotone Convergence theorem, where $\tau_\infty\triangleq\inf\{t>0\colon \|X(t)\|=\infty\}$. The nonexplosiveness of X implies that $\tau_\infty=\infty$ $\mathbb{P}$-a.s. Therefore we conclude that $\mathbb{E}\,M(T)=1$. □

A.3. Proof of Proposition 6

Lemma 2. *Let A be a Z-matrix, so that there exists $s\in\mathbb{R}$ and a nonnegative matrix B for which $A=s\mathbf{I}-B$. Then the following three statements are equivalent*:
(1) *A is an M-matrix.*
(2) *$s>\rho(B)$, where $\rho(B)$ denotes the spectral radius of B.*
(3) *For any vector $v\neq\mathbf{0}$, there exists a nonnegative diagonal matrix D such that $v^\top ADv>0$.*

Proof. See Berman and Plemmons [7, pp. 132–134]. □

We then have the following immediate result.

Corollary 2. *Let A be an M-matrix and D be a nonnegative diagonal matrix with the same dimension. Then $A+D$ is an M-matrix.*

Lemma 3. *Under Assumption 1, $\beta_{I,I}$ is an M-matrix and β is positive stable.*

Proof. Note that by part (I) of Assumption 1, $\beta_{I,I}$ is a Z-matrix and $\mathbb{E}(Z^i)\gamma_{i,I}\kappa_{i,I}^\top$ is a nonnegative matrix. Therefore $\beta_{I,I}-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_{i,I}\kappa_{i,I}^\top$ is a Z-matrix as well. Moreover, note that $\beta_{I,J}=\mathbf{0}$ and $\kappa_{i,J}=0$ for all $i=1,\dots,n$, where $I=\{1,\dots,m\}$ and $J=\{m+1,\dots,d\}$,

$$\beta-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_i\kappa_i^\top=\begin{pmatrix}\beta_{I,I}-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_{i,I}\kappa_{i,I}^\top & \mathbf{0}\\ \beta_{J,I}-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_{i,J}\kappa_{i,I}^\top & \beta_{J,J}\end{pmatrix}.$$

It then follows from part (III) of Assumption 1 that both $\beta_{I,I}-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_{i,I}\kappa_{i,I}^\top$ and $\beta_{J,J}$ are postive stable. Therefore $\beta_{I,I}-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_{i,I}\kappa_{i,I}^\top$ is an M-matrix.

On the other hand, note that there exists $s>0$ and a nonnegative matrix B for which $\beta_{I,I}=s\mathbf{I}-B$ since $\beta_{I,I}$ is a Z-matrix. Hence we can write

$$\beta_{I,I}-\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_{i,I}\kappa_{i,I}^\top=s\mathbf{I}-\left(B+\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_{i,I}\kappa_{i,I}^\top\right).$$

It then follows from Lemma 2 that

$$s>\rho\left(B+\sum_{i=1}^n\mathbb{E}(Z^i)\gamma_{i,I}\kappa_{i,I}^\top\right) \tag{37}$$

and thus $s>\rho(B)$ since $\mathbb{E}(Z^i)\gamma_{i,I}\kappa_{i,I}^\top\in\mathbb{R}_+^{m\times m}$ for all $1\leq i\leq n$ (see, for example, p. 491 of Horn and Johnson [39]). Consequently, $\beta_{I,I}$ is an M-matrix by Lemma 2. Therefore

$$\beta=\begin{pmatrix}\beta_{I,I} & \mathbf{0}\\ \beta_{J,I} & \beta_{J,J}\end{pmatrix}$$

is positive stable because $\beta_{I,I}$ and $\beta_{J,J}$ are positive stable. □

Lemma 4. *Under Assumptions 1 and 3, $\mathcal{J}(\theta, u_I^*(\theta))$ is an M-matrix if and only if $\theta \in (\underline{\theta}, \bar{\theta})$, where $\bar{\theta}$ and $\underline{\theta}$ are respectively defined in (19) and (20).*

Proof. Let $s(\theta) = \max\{0, \mathcal{J}(\theta, u_I^*(\theta))_{i,i}, 1 \le i \le m\}$ and $B(\theta) = s(\theta)\mathbf{I} - \mathcal{J}(\theta, u_I^*(\theta))$. Since $\mathcal{J}(\theta, u_I^*(\theta))$ is a Z-matrix by the representation (17), it follows that $B(\theta) \in \mathbb{R}_+^{m\times m}$. By the Perron-Frobenius theorem, $\rho(B(\theta))$ is an eigenvalue of $B(\theta)$, where $\rho(\cdot)$ denotes the spectral radius. Therefore $\mathcal{J}(\theta, u_I^*(\theta)) = s(\theta)\mathbf{I} - B(\theta)$ is singular if $s(\theta) = \rho(B(\theta))$. Note that $\mathcal{J}(0, \mathbf{0})$ is an M-matrix by (18). It then follows from Lemma 2 that $s(0) > \rho(B(0))$. Furthermore, since $s(\theta)$ and elements of $B(\theta)$ are continuous in θ, we conclude that $\theta \in (\underline{\theta}, \bar{\theta})$ if and only if $s(\theta) > \rho(B(\theta))$, which is true if and only if $\mathcal{J}(\theta, u_I^*(\theta))$ is an M-matrix. $\square$

Lemma 5. *Under Assumptions 1 and 3, $\nabla u_I^*(\theta) \in \mathbb{R}_+^m$ for $\theta \in (\underline{\theta}, \bar{\theta})$.*

Proof. Note that
$$\nabla u_I^*(\theta) = -\mathcal{J}(\theta, u_I^*(\theta))^{-1} \nabla_\theta F(\theta, u_I^*(\theta)), \tag{38}$$
and that
$$\nabla_\theta F(\theta, v)^\top = -\sum_{i=1}^n \mathbb{E}(Z^i e^{(\theta + v^\top \gamma_{i,I})Z^i}) \kappa_{i,I}^\top \in \mathbb{R}_-^m. \tag{39}$$
It follows from Lemma 4 that, for any $\theta \in (\underline{\theta}, \bar{\theta})$, $\mathcal{J}(\theta, u_I^*(\theta))$ is an M-matrix, and thus $J(\theta, u_I(\theta))^{-1} \in \mathbb{R}_+^{m\times m}$ (see p. 137 of Berman and Plemmons [7]). Hence $\nabla u_I^*(\theta) \in \mathbb{R}_+^m$ for any $\theta \in (\underline{\theta}, \bar{\theta})$ by (38) and (39). $\square$

Lemma 6. *Under Assumptions 1 and 3, $\mathcal{J}(\theta, v)$ is an M-matrix for all $(\theta, v) \in \mathbb{R}_- \times \mathbb{R}_-^m$.*

Proof. Following the notations in the proof of Lemma 3, we have $\beta_{I,I} = s\mathbf{I} - B$ for some $s > 0$ and some $B \in \mathbb{R}_+^{m\times m}$. Then by (17),
$$\mathcal{J}(\theta, v)^\top = s\mathbf{I} - \left(B + \sum_{i=1}^n \mathbb{E}(Z^i e^{(\theta + v^\top \gamma_{i,I})Z^i}) \gamma_{i,I} \kappa_{i,I}^\top\right) + \operatorname{diag}(-\alpha_{1,1}^1 v_1, \ldots, -\alpha_{m,m}^m v_m). \tag{40}$$
It follows from (37) that, for all $(\theta, v) \in \mathbb{R}_- \times \mathbb{R}_-^m$,
$$s > \rho\left(B + \sum_{i=1}^n \mathbb{E}(Z^i e^{(\theta + v^\top \gamma_{i,I})Z^i}) \gamma_{i,I} \kappa_{i,I}^\top\right)$$
since $\gamma_{i,I}(\cdot) \in \mathbb{R}_+^m$ for all $1 \le i \le n$, yielding that
$$s\mathbf{I} - \left(B + \sum_{i=1}^n \mathbb{E}(Z^i e^{(\theta + v^\top \gamma_{i,I})Z^i}) \gamma_{i,I} \kappa_{i,I}^\top\right)$$
is an M-matrix by Lemma 2. It further implies that $\mathcal{J}(\theta, v)$ is an M-matrix for all $(\theta, v) \in \mathbb{R}_- \times \mathbb{R}_-^m$ by (40) and Corollary 2. $\square$

Lemma 7. *Under Assumptions 1 and 3, $\underline{\theta} = -\infty$, where $\underline{\theta}$ is defined in (20).*

Proof. Lemma 5 together with $u_I^*(0) = \mathbf{0}$ implies that $u_I^*(\theta) \in \mathbb{R}_-^m$ for all $\theta \in (\underline{\theta}, 0)$. Hence, by Lemma 6, there does not exist $\theta < 0$ such that $\mathcal{J}(\theta, u_I^*(\theta))$ is singular, yielding that $\underline{\theta} = -\infty$. $\square$

Lemma 8. *Under Assumptions 1 and 3, $\bar{\theta} = \infty$ if $\kappa = \mathbf{0}$, and $\bar{\theta} < \infty$ otherwise, where $\bar{\theta}$ is defined in (19).*

Proof. If $\kappa = \mathbf{0}$, then $\nabla u_I(\theta) \equiv \mathbf{0}$ by (38) and (39). In particular, Equations (15) can be solved trivially by $u_I^*(\theta) = 0$ for all $\theta \in \mathbb{R}$. Hence $\mathcal{D}_{u_I^*} = \mathbb{R}$, where $\mathcal{D}_{u_I^*}$ is the domain of $u_I^*(\theta)$, implying that $\bar{\theta} = \infty$.

If $\kappa \ne \mathbf{0}$, then by part (I) of Assumption 1, there exists $j \in \{1, \ldots, m\}$ such that $\kappa_{i,j} > 0$. Note that $u_I^*(\theta) \in \mathbb{R}_+^m$ for all $\theta \in (0, \bar{\theta})$ by Lemma 5. Therefore it follows from (15) that
$$\sum_{k=1}^m u_k^*(\theta)\beta_{k,j} - \tfrac{1}{2}\alpha_{j,j}^j u_j^*(\theta)^2 = (\mathbb{E}\, e^{(\theta + u_I^*(\theta)^\top \gamma_{i,I})Z^i} - 1)\kappa_{i,j} \ge (\mathbb{E}\, e^{\theta Z^i} - 1)\kappa_{i,j}.$$
By part (I) of Assumption 1, $\beta_{I,I}$ has nonpositive off-diagonal elements. Therefore
$$u_j^*(\theta)\beta_{j,j} - \tfrac{1}{2}\alpha_{j,j}^j u_j^*(\theta)^2 \ge (\mathbb{E}\, e^{\theta Z^i} - 1)\kappa_{i,j}$$
for $\theta \in \mathcal{D}_{u_I^*} \cap \mathbb{R}_+$. Since $\alpha_{j,j}^j > 0$ by part (II) of Assumption 1, the LHS of the last equality is upper bounded, and thus θ must be bounded so that Equations (15) have a solution. Therefore $\mathcal{D}_{u_I^*}$ is upper bounded, thereby $\bar{\theta} < \infty$. $\square$

Proof of Proposition 6. It is an immediate result from Lemmas 7 and 8. $\square$

A.4. Proof of Proposition 7. We first discuss the case where $\theta \geq 0$. In this case, it suffices to show that X is ergodic under Q_θ^*.

PROPOSITION 11. *Suppose that Assumptions 1 and 3 hold. Then, for any $\theta \in \mathcal{D}^o_{u_I^*} = (-\infty, \bar{\theta})$, X has a unique stationary distribution $\pi^{Q_\theta^*}$ under the probability measure Q_θ^*. Moreover, $Q_\theta^*(X(t) \in \cdot) \to \pi^{Q_\theta^*}(\cdot)$ in total variation as $t \to \infty$.*

PROOF. Fix θ. Let $\hat{\lambda}$, $\hat{\kappa}$, $\hat{\beta}_{I,I}$, and $\hat{\varphi}_i$ be defined as in (28), (29), (30), and (33) with $u = u^*$. Let $\hat{Z}^i$ be a random variable with distribution $\hat{\varphi}_i$. Then, we can rewrite (17) as

$$\mathcal{J}(\theta, u_I^*(\theta))^\top = \hat{\beta}_{I,I} - \sum_{i=1}^n \mathbb{E}(\hat{Z}^i)\gamma_{i,I}\hat{\kappa}_{i,I}^\top,$$

which is positive stable for $\theta \in (-\infty, \bar{\theta})$. Hence

$$\hat{\beta} - \sum_{i=1}^n \mathbb{E}(\hat{Z}^i)\gamma_i\hat{\kappa}_i^\top = \begin{pmatrix} \hat{\beta}_{I,I} - \sum_{i=1}^n \mathbb{E}(\hat{Z}^i)\gamma_{i,I}\hat{\kappa}_{i,I}^\top & \mathbf{0} \\ \beta_{J,I} - \sum_{i=1}^n \mathbb{E}(\hat{Z}^i)\gamma_{i,J}\hat{\kappa}_{i,I}^\top & \beta_{J,J} \end{pmatrix}$$

is positive stable.

Note that Proposition 10 asserts that the parameters $(a, \alpha, b, \hat{\beta}, \hat{\lambda}, \hat{\kappa}, \gamma)$ are admissible. Moreover, it is easy to see $\hat{\varphi}_i(\cdot)$ satisfies Assumption 3. Consequently, we can apply Proposition 9 to the AJD that satisfies the SDE with parameters $(a, \alpha, b, \hat{\beta}, \hat{\lambda}, \hat{\kappa}, \gamma)$. Note that such an SDE is exactly the one that X satisfies under Q_θ^*. Hence we conclude that X has a unique stationary distribution $\pi^{Q_\theta^*}$ under Q_θ^* and $Q_\theta^*(X(t) \in \cdot) \to \pi^{Q_\theta^*}(\cdot)$ in total variation as $t \to \infty$.

Proposition 11 asserts that X is ergodic under probability measure Q_θ^*, thereby $\mathbb{E}^{Q_\theta^*} f(X(t)) \to \int_{\mathcal{S}} f(x)\pi^{Q_\theta^*}(\mathrm{d}x)$ as $t \to \infty$ for any bounded function f. Hence we have the following corollary.

COROLLARY 3. *Under Assumptions 1 and 3, for $\theta \in \mathcal{D}_{\phi^*} \cap \mathbb{R}_+$,*

$$\mathbb{E}^{Q_\theta^*} \exp[-u^*(\theta)^\top (X(t) - X(0))] = O(1)$$

as $t \to \infty$.

PROOF. It follows from Lemma 5 that $u_I^*(\theta) \in \mathbb{R}_+^m$ for $\theta \in \mathcal{D}_{\phi^*} \cap \mathbb{R}_+$. Therefore

$$\exp(-u^*(\theta)^\top X(t)) = \exp\left(-\sum_{i=1}^m u_i^*(\theta)X_i(t)\right) \leq 1$$

since $X_i(\cdot) \geq 0$ for $i = 1, \ldots, m$. Consequently, by Proposition 11, we have

$$\mathbb{E}^{Q_\theta^*} \exp[-u^*(\theta)^\top (X(t) - X(0))] \to \int_{\mathcal{S}} e^{-u^*(\theta)^\top (x - x_0)} \pi^{Q_\theta^*}(\mathrm{d}x) < \infty$$

as $t \to \infty$ for $\theta \in \mathcal{D}_{\phi^*} \cap \mathbb{R}_+$. $\square$

We now discuss the case when $\theta < 0$. In this case, it suffices to show that X is exponentially ergodic under Q_θ^*. To that end, we will apply the Foster-Lyapunov method (see Meyn and Tweedie [43]) for an extensive exposition of this approach. The key to the Foster-Lyapunov approach is to find an appropriate test function that satisfies the so-called Lyapunov inequality, more specifically (41) in our setting.

Let $\mathcal{A}_\theta^*$ denote the infinitesimal generator of X under Q_θ^*, i.e.,

$$\mathcal{A}_\theta^* f(x) = \frac{1}{2}\sum_{i,j=1}^d \left(a_{i,j} + \sum_{k=1}^m \alpha_{i,j}^k x_k\right) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + (b - \hat{\beta}x) \cdot \nabla f(x) + \sum_{i=1}^n (\hat{\lambda}_i + \hat{\kappa}_i \cdot x) \int_{\mathbb{R}_+} (f(x + \gamma_i z) - f(x))\hat{\varphi}_i(\mathrm{d}z)$$

for any twice continuously differentiable function $f: \mathbb{R}^d \to \mathbb{R}$.

LEMMA 9. *Fix $\theta < 0$. Under Assumptions 1 and 3, there exist $w > 0$ and $c \in \mathbb{R}_+^d$ with $c_J = \mathbf{0}$ such that*

$$\mathcal{A}_\theta^* g(x) \leq -g(x) + w \tag{41}$$

for all $x \in \mathcal{S} = \mathbb{R}_+^m \times \mathbb{R}^{d-m}$, where $g(x) = e^{(c - u^*)^\top x}$.

Proof. Noting that $c_J = u_J^* = \mathbf{0}$, $g(x)$ is independent of x_J. Then, a direct calculation yields that

$$\begin{aligned}\mathscr{A}_\theta^* g(x) &= g(x)\bigg[(c-u^*)^\top(b-\hat{\beta}x)+\tfrac{1}{2}\sum_{k=1}^m \alpha_{k,k}^k (c_k-u_k^*)^2 x_k + \sum_{i=1}^n (\hat{\lambda}_i+\hat{\kappa}_i^\top x)\int_{\mathbb{R}_+}(e^{\gamma_i^\top (c-u^*)z}-1)\hat{\varphi}_i(\mathrm{d}z)\bigg]\\ &= g(x)\bigg[u^{*\top}\beta x - \tfrac{1}{2}\sum_{k=1}^m \alpha_{k,k}^k u_k^{*2}x_k - \sum_{i=1}^n \kappa_i^\top x\cdot \mathbb{E}\, e^{(\theta+\gamma_i^\top u^*)Z^i} - c^\top \beta x + \tfrac{1}{2}\sum_{k=1}^m \alpha_{k,k}^k c_k^2 x_k + \sum_{i=1}^n \kappa_i^\top x \cdot \mathbb{E}\, e^{(\theta+\gamma_i^\top c)Z^i} + D_1\bigg]\\ &= g(x)\big[(F(\theta,u_I^*)-F(\theta,c_I))^\top x_I + D_1\big]\\ &= g(x)\big[-F(\theta,c_I)^\top x_I + D_1\big],\end{aligned}$$

where $F\colon \mathbb{R}\times\mathbb{R}^m\to\mathbb{R}^m$ is defined by (16) and

$$D_1 = (c-u^*)^\top b + \sum_{i=1}^n \lambda_i \int_{\mathbb{R}_+}(e^{(\theta+\gamma_i^\top c)z} - e^{(\theta+\gamma_i^\top u^*)z})\varphi_i(\mathrm{d}z).$$

The Taylor expansion indicates that

$$F(\theta,c_I)-F(\theta,\mathbf{0}) = \mathcal{J}(\theta,\mathbf{0})^\top c_I + o(\|c_I\|)$$

as $\|c_I\|\downarrow 0$. Since $\mathcal{J}(\theta,\mathbf{0})$ is an M-matrix by Lemma 6, it follows that there exists $q\in\mathbb{R}_{++}^m$ such that $\mathcal{J}(\theta,\mathbf{0})^\top q\in\mathbb{R}_{++}^m$ (see Berman and Plemmons [7, p. 136]). Hence we can set $c_I=\epsilon q$ for a sufficient small $\epsilon>0$, so that $\mathcal{J}(\theta,\mathbf{0})^\top c_I + o(\|c_I\|)\in\mathbb{R}_{++}^m$. Moreover, noting that

$$F_j(\theta,\mathbf{0}) = \sum_{i=1}^n \kappa_{i,j}(1-\mathbb{E}\, e^{\theta Z^i})\geq 0,\quad j=1,\ldots,m,$$

we conclude that $F(\theta,c_I)\in\mathbb{R}_{++}^m$ for $\epsilon>0$ small enough. It follows that there exists $D_2>0$ large enough such that $F(\theta,c_I)^\top x_I - D_1 > 1$ for all $x\in\mathcal{S}\backslash K$, where $K=\{x\in\mathcal{S}\colon 0\leq x_i\leq D_2, 1\leq i\leq m\}$. On the other hand, obviously $w\triangleq \sup_{x\in K} g(x)|F(\theta,c_I)^\top x_I - D_1 - 1|<\infty$. Therefore $\mathscr{A}_\theta^* g(x)\leq -g(x)+w$ for all $x\in\mathcal{S}$. □

Corollary 4. *Under Assumptions 1 and 3, for* $\theta<0$,

$$\mathbb{E}^{Q_\theta^*}\exp[-u^*(\theta)^\top (X(t)-X(0))] = O(1)$$

as $t\to\infty$.

Proof. We need the concept of the f*-norm*, denoted by $\|\psi\|_f$ for a positive measurable function $f\geq 1$ and a signed measure ψ as follows:

$$\|\psi\|_f \triangleq \sup_{|h|\leq f}\left|\int f(x)\psi(\mathrm{d}x)\right|.$$

Proposition 11 guarantees that X has a unique stationary distribution $\pi^{Q_\theta^*}$ under Q_θ^*. Let $g(x)=\exp[(c-u^*)^\top x]$ for $x\in\mathcal{S}$, where c is specificed as in Lemma 9. It then follows from Lemma 9 and Theorem 6.1 of Meyn and Tweedie [43] that

$$\left\|\mathbb{P}^{Q_\theta^*}(X(t)\in\cdot\mid X(0)) - \pi^{Q_\theta^*}(\cdot)\right\|_{f_1}\to 0 \tag{42}$$

as $t\to\infty$, where $f_1\triangleq g+1$.

Note that $c\in\mathbb{R}_+^d$ with $c_J=\mathbf{0}$ so $g(x)\geq \exp(-u^{*\top}x)\triangleq f_2(x)$ for all $x\in\mathcal{S}$. Hence, by the definition of the f-norm,

$$\left\|\mathbb{P}^{Q_\theta^*}(X(t)\in\cdot\mid X(0)) - \pi^{Q_\theta^*}(\cdot)\right\|_{f_2}\leq \left\|\mathbb{P}^{Q_\theta^*}(X(t)\in\cdot\mid X(0)) - \pi^{Q_\theta^*}(\cdot)\right\|_{f_1}. \tag{43}$$

It follows from (42) and (43) that

$$\left\|\mathbb{P}^{Q_\theta^*}(X(t)\in\cdot\mid X(0)) - \pi^{Q_\theta^*}(\cdot)\right\|_{f_2}\to 0$$

as $t\to\infty$, and thus

$$\mathbb{E}^{Q_\theta^*}\exp\big[-u^*(\theta)^\top(X(t)-X(0))\big]\to\int_{\mathcal{S}} e^{-u^*(\theta)^\top (x-x_0)}\pi^{Q_\theta^*}(\mathrm{d}x)<\infty$$

as $t\to\infty$. □

Proof of Proposition 7. It is an immediate result from Corollaries 3 and 4. □

A.5. Proof of Theorem 2. By the expression (22), the domain of ϕ^* can be written as

$$\mathscr{D}_{\phi^*} = \mathscr{D}_{u_I^*} \cap \left(\bigcap_{\{i:\, \lambda_i > 0\}} \{\theta\colon\ \mathbb{E}\, e^{(\theta + u_I^*(\theta)^\top \gamma_{i,I}) Z^i} < \infty\} \right).$$

Note that $\mathscr{D}^o_{u^*} = (-\infty, \bar{\theta})$ by Proposition 6, where $\bar{\theta}$ is defined by (19), and that $u^*(\theta) \in \mathbb{R}^d_-$ for $\theta \le 0$ by Lemma 5. Then $\mathscr{D}^o_{\phi^*} = (-\infty, \tilde{\theta})$, where

$$\tilde{\theta} = \bar{\theta} \wedge \min_{\{i:\, \lambda_i > 0\}} \hat{\theta}_i \tag{44}$$

and $\hat{\theta}_i > 0$ is such that

$$\hat{\theta}_i + u_I^*(\hat{\theta}_i)^\top \gamma_{i,I} = \sup\{\theta\colon\ \mathbb{E}\, e^{\theta Z^i} < \infty\}. \tag{45}$$

Moroever, note that

$$\phi^{*\prime}(\theta) = \nabla u_I^*(\theta)^\top b_I + \sum_{i=1}^n \lambda_i (1 + \nabla u_I^*(\theta)^\top \gamma_{i,I}) \mathbb{E}\, Z^i e^{(\theta + u_I^*(\theta)^\top \gamma_{i,I}) Z^i} \ge 0 \tag{46}$$

since $\nabla u^*(\theta) \in \mathbb{R}^d_+$. Consequently, to show that ϕ^* is steep, it suffices to show that the range of $\phi^{*\prime}(\theta)$ is $(0, \infty)$ for $\theta \in (-\infty, \tilde{\theta})$.

Lemma 10. *Under Assumptions 1 and 3, $\phi^{*\prime}(\theta) \to 0$ as $\theta \downarrow -\infty$.*

Proof. Note that by sending $\theta \downarrow -\infty$, Equations (15) are reduced to

$$\sum_{i=1}^m u_i \beta_{i,j} - \tfrac{1}{2} \alpha^j_{j,j} u_j^2 + \sum_{i=1}^n \kappa_{i,j} = 0, \quad j = 1, \ldots, m.$$

It can be easily shown by the Miranda existence test (see Miranda [44] or Frommer et al. [30]) that the above nonlinear equations about u have a unique solution $\underline{u} = (\underline{u}_1, \ldots, \underline{u}_m)$ that lies in $\mathbb{R}^m_-$. Hence $u_I^*(\theta) \to \underline{u}$ as $\theta \downarrow -\infty$. Then by (17),

$$\lim_{\theta \downarrow -\infty} \mathscr{J}(\theta, u_I^*(\theta))^\top \to \beta_{I,I} - \operatorname{diag}(\alpha^1_{1,1} \underline{u}_1, \ldots, \alpha^m_{m,m} \underline{u}_m)),$$

which is nonsingular by Lemma 6. Further, note that $\nabla_\theta F(\theta, u_I^*(\theta)) \to 0$ as $\theta \downarrow -\infty$ by (39). It then follows from (38) that

$$\lim_{\theta \downarrow -\infty} \nabla u_I^*(\theta) = \lim_{\theta \downarrow -\infty} \mathscr{J}(\theta, u_I^*(\theta))^{-1} \nabla_\theta F(\theta, u_I^*(\theta)) = \mathbf{0}.$$

Therefore, by (46), we conclude that $\phi^{*\prime}(\theta) \to 0$ as $\theta \downarrow -\infty$. □

Lemma 11. *Under Assumptions 1 and 3, $\phi^{*\prime}(\theta) \to \infty$ as $\theta \uparrow \tilde{\theta}$.*

Proof. By (44), we will discuss two cases, i.e., $\tilde{\theta} = \bar{\theta}$ and $\tilde{\theta} = \hat{\theta}_i$ for some i such that $\lambda_i > 0$.

Case 1. Assume $\tilde{\theta} = \bar{\theta}$. First note that if $\kappa = \mathbf{0}$, then $\lambda_i > 0$ by part (II) Assumption 1, and $\bar{\theta} = \infty$ by Lemma 8, implying that $\hat{\theta}_i = \infty$ for all $i = 1, \ldots, n$. Hence we can incorporate this scenario, i.e., $\kappa = \mathbf{0}$, into the discussion of Case 2 later.

If $\kappa \ne \mathbf{0}$, then $\bar{\theta} < \infty$ by Lemma 8. It then follows from (19), the definition of $\bar{\theta}$ that $\mathscr{J}(\bar{\theta}, u_I^*(\bar{\theta}))$ is singular. Hence

$$\lim_{\theta \uparrow \bar{\theta}} \| \mathscr{J}(\theta, u_I^*(\theta))^{-1} \| = \infty.$$

Apparently, for all i such that $\lambda_i > 0$, we have $\bar{\theta} < \hat{\theta}_i$ so $\mathbb{E}\, e^{(\bar{\theta} + u_I^*(\bar{\theta})^\top \gamma_{i,I}) Z^i} < \infty$. Hence

$$\lim_{\theta \uparrow \bar{\theta}} \nabla_\theta F(\theta, u_I^*(\theta)) = -\sum_{i=1}^n \mathbb{E}(Z^i e^{(\bar{\theta} + u_I^*(\bar{\theta})^\top \gamma_{i,I}) Z^i}) \kappa_{i,I}^\top,$$

which is finite. Therefore $\|\nabla u_I^*(\theta)\| \to \infty$ as $\theta \uparrow \bar{\theta}$ by (38). Since $b_i > 0$ for $i = 1, \ldots, m$ by part (II) Assumption 1, it then immediately follows from (46) that $\phi^{*\prime}(\theta) \to \infty$ as $\theta \uparrow \tilde{\theta}$.

Case 2. If $\tilde{\theta} = \hat{\theta}_i$ for some i such that $\lambda_i > 0$, then by (46)

$$\phi^{*\prime}(\theta) \ge \lambda_i \mathbb{E}\, Z^i e^{(\theta + u_I^*(\theta)^\top \gamma_{i,I}) Z^i} \ge \lambda_i \mathbb{E}\, e^{(\theta + u_I^*(\theta)^\top \gamma_{i,I}) Z^i} \mathbb{I}(Z^i \ge 1) \to \infty$$

as $\theta \uparrow \hat{\theta}_i$ by (45). Hence $\phi^{*\prime}(\theta) \to \infty$ as $\theta \uparrow \tilde{\theta}$. □

Proof of Theorem 2. Since $\phi(\theta)$ is increasing in θ, it follows from Lemmas 10 and 11 that, for each $R > 0$, there exists a unique $\theta^* \in (-\infty, \tilde{\theta})$ such that $\phi^{*\prime}(\theta^*) = R$. We then can apply the Gärtner-Ellis theorem to prove Theorem 2. □

References

[1] Abate J, Whitt W (1992) The Fourier-series method for inverting transforms of probability distributions. *QUESTA* 10(1):5–88.
[2] Aït-Sahalia Y, Cacho-Diaz J, Laeven RJA (2014) Modeling financial contagion using mutually exciting jump processes. *J. Financial Econom.* Forthcoming.

[3] Andersen PK, Borgan Ø, Gill RD, Keiding N (1993) *Statistical Models Based on Counting Processes* (Springer, New York).
[4] Azizpour S, Giesecke K, Kim B (2011) Premia for correlated default risk. *J. Econom. Dynam. Control* 35(8):1340–1357.
[5] Bacry E, Delattre S, Hoffmann M, Muzy JF (2013) Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes Their Appl.* 123:2475–2499.
[6] Bassamboo A, Jain S (2006) Efficient importance sampling for reduced form models in credit risk. Wieland FP, Perrone LF, Liu J, Lawson BG, Nicol DM, Fujimoto RM, eds. *Proc.* 2006 *Winter Simulation Conf.* (IEEE, Piscataway, NJ), 741–749.
[7] Berman A, Plemmons RJ (1987) *Nonnegative Matrices in the Mathematical Sciences* (SIAM, Philadelphia).
[8] Blanchet J, Glynn PW, Meyn SP (2013) Large deviations for the empirical mean of an M/M/1 queue. *QUESTA* 73:425–446.
[9] Bordenave C, Torrisi GL (2007) Large deviations of Poisson cluster processes. *Stochastic Models* 23:593–625.
[10] Bowsher CG (2007) Modelling security market events in continuous time: Intensity based, multivariate point process models. *J. Econometrics* 141:876–912.
[11] Brémaud P (1981) *Point Processes and Queues: Martingale Dynamics* (Springer, New York).
[12] Carmona R, Crépey S (2010) Particle methods for the estimation of Markovian credit portfolio loss distributions. *Internat. J. Theoret. Appl. Finance* 13(4):577–602.
[13] Cheridito P, Filipović D, Kimmel RL (2007) Market price of risk specification for affine models: Theory and evidence. *J. Financial Econom.* 83:123–170.
[14] Cheridito P, Filipović D, Yor M (2005) Equivalent and absolutely continuous measure changes for jump-diffusion processes. *Ann. Appl. Probab.* 15:1713–1732.
[15] Cvitanić J, Ma J, Zhang J (2012) Law of large numbers for self-exciting correlated defaults. *Stochastic Processes Their Appl.* 122(8):2781–2810.
[16] Dai Q, Singleton KJ (2000) Specification analysis of affine term structure models. *J. Finance* 55:1943–1978.
[17] Dai Pra P, Tolotti M (2009) Heterogeneous credit portfolios and the dynamics of the aggregate losses. *Stochastic Processes Their Appl.* 119:2913–2944.
[18] Dai Pra P, Runggaldier WJ, Sartori E, Tolotti M (2009) Large portfolio losses: A dynamic contagion model. *Ann. Appl. Probab.* 19:347–394.
[19] Daley DJ (1972) Asymptotic properties of stationary point processes with generalized clusters. *Z. Wahrsch. Verw. Gebiete.* 21:65–76.
[20] Del Moral P, Garnier J (2005) Genealogical particle analysis of rare events. *Ann. Appl. Probab.* 15:2496–2534.
[21] Dembo A, Zeitouni O (1998) *Large Deviations Techniques and Applications*, 2nd ed. (Springer, New York).
[22] Deng S, Giesecke K, Lai TL (2012) Sequential importance sampling and resampling for dynamic portfolio credit risk. *Oper. Res.* 60(1):78–91.
[23] Duffie D, Filipović D, Schachermayer W (2003) Affine processes and applications in finance. *Ann. Appl. Probab.* 13(3):984–1053.
[24] Duffie D, Pan J, Singleton KJ (2000) Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68(6):1343–1376.
[25] Duffy KR, Meyn SP (2010) Most likely paths to error when estimating the mean of a reflected random walk. *Performance Evaluation* 67(12):1290–1303.
[26] Embrechts P, Liniger T, Lin L (2011) Multivariate Hawkes processes: An application to financial data. *J. Appl. Prob.* 48:367–378.
[27] Errais E, Giesecke K, Goldberg LR (2010) Affine point processes and portfolio credit risk. *SIAM J. Finan. Math.* 1:642–665.
[28] Ethier SN, Kurtz TG (1986) *Markov Processes: Characterization and Convergence* (John Wiley & Sons, New York).
[29] Filipović D, Mayerhofer E, Schneider P (2013) Density approximations for multivariate affine jump-diffusion processes. *J. Econometrics* 176:93–111.
[30] Frommer A, Lang B, Schnurr M (2004) A comparison of the Moore and Miranda existence tests. *Computing* 72:349–354.
[31] Giesecke K, Shkolnik A (2014) Optimal importance sampling of default losses. Working paper.
[32] Giesecke K, Weber S (2006) Credit contagion and aggregate losses. *J. Econom. Dynam. Control* 30:741–767.
[33] Giesecke K, Kim B, Zhu S (2011) Monte Carlo algorithms for default timing problems. *Management Sci.* 57(12):2115–2129.
[34] Giesecke K, Spiliopoulos K, Sowers RB (2013) Default clustering in large portfolios: Typical events. *Ann. Appl. Probab.* 23(1): 348–385.
[35] Giesecke K, Kakavand H, Mousavi M, Takada H (2010) Exact and efficient simulation of correlated defaults. *SIAM J. Finan. Math.* 1:868–896.
[36] Giesecke K, Spiliopoulos K, Sowers RB, Sirignano JA (2015) Large portfolio asymptotics for loss from default. *Math. Finance* 25(1):77–114.
[37] Glasserman P (2003) *Monte Carlo Methods in Financial Engineering* (Springer, New York).
[38] Glasserman P, Kim K-K (2009) Saddlepoint approximation for affine jump-diffusion models. *J. Econom. Dynam. Control* 33:37–52.
[39] Horn RA, Johnson CR (1985) *Matrix Analysis* (Cambridge University Press, Cambridge, UK).
[40] Karatzas I, Shreve SE (1991) *Brownian Motion and Stochastic Calculus*, 2nd ed. (Springer, New York).
[41] Kontoyiannis I, Meyn SP (2003) Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.* 13(1):304–362.
[42] Lépingle D, Mémin J (1978) Sur l'integrabilité uniforme des martingales exponentielles. *Z. Wahrsch. Verw. Gebiete* 42(3):175–203.
[43] Meyn SP, Tweedie RL (1993) Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.* 25:518–548.
[44] Miranda C (1940) Un'osservazione su un teorema di Brouwer. *Bolletino Unione Mathematica Italiana* 3:5–7.
[45] Protter PE (2003) *Stochastic Integration and Differential Equations*, 2nd ed. (Springer, Berlin).
[46] Spiliopoulos K, Sowers RB (2014) Default clustering in large pools: Large deviations. *SIAM J. Finan. Math.* Forthcoming.
[47] Spiliopoulos K, Sirignano JA, Giesecke K (2014) Fluctuation analysis for the loss from default. *Stoch. Proc. Appl.* 124(7):2322–2362.
[48] Zhang X (2011) Computing rare-event probabilities for affine models and general state space Markov processes. Doctoral dissertation, Stanford University, Stanford, CA.
[49] Zhang X, Glynn PW, Giesecke K, Blanchet J (2009) Rare event simulation for a generalized Hawkes process. Rossetti MD, Hill RR, Johansson B, Dunkin A, Ingalls RG, eds. *Proc.* 2009 *Winter Simulation Conf.* (IEEE, Piscataway, NJ),1291–1298.
[50] Zhu L (2013) Central limit theorem for nonlinear Hawkes processes. *J. Appl. Probab.* 50(3):760–771.
[51] Zhu L (2014) Process-level large deviations for nonlinear Hawkes point processes. *Annales de l'Institut Henri Poincaré* 50(3):845–871.