Conditional Importance Sampling for Event Counting Processes

Baeho Kim* and Alexander Shkolnik†

June 9, 2018

Abstract

This article develops a *conditional* important sampling (IS) for a wide range of stochastic point process models of event counting processes. We contribute to the event timing simulation literature by designing, implementing and testing a simple and practically efficient IS scheme for tail probabilities. Our approach to the design of the scheme is nontrivial and is of independent interest. Numerical results demonstrate its application to estimating rare-event probabilities and pricing interest rate derivatives with efficient performance.

Kewwords: Conditional importance sampling, Point process, Stochastic intensity, Rare-event probability

*Korea University Business School, Seoul, 02841, South Korea, Phone +82-2-3290-2626, Fax +82-2-922-7220, Email: `baehokim@korea.ac.kr`

†Consortium for Data Analytics in Risk (CDAR), University of California, Berkeley, CA 94720, USA, Email: `ads2@berkeley.edu`.

1 Introduction

The problem of estimating the likelihood that a given number of events is observed by some horizon arises naturally in a wide range of disciplines (e.g. finance, queuing, insurance, etc.). For instance, investors and fixed-income managers are concerned by the odd chance a portfolio of defaultable assets (e.g. loans, bonds, etc.) experiences a large number of credit events.[1] Insurance risk management centers around the probability of ruin due to some excessive number of claims. In reliability studies, the likelihood of systemic breakdown caused by the failure too many system components in too short a time is often emphasized. Unbiased and efficient algorithms for these problems are important not only for the numerical values they produce, but also for investigating how tail phenomena occur in a particular model via simulation.

This paper develops a novel, easy to simulate and fast Monte Carlo (MC) estimator of rare event probabilities described in the examples above. We focus event timing models whose dynamics may be prescribed by a (multivariate) point process. Such models and their simulation algorithms are widely used in the literature. The events arrive at some stochastic rate, or an "intensity" process which is typically modeled as a system of stochastic differential equations. Our algorithms accommodate any such intensity specification provided it may be simulated. However, plain MC is not adequate when the intensity values are too small to produce the sufficient number of events that hits the rare event threshold. For this reason, it is highly inaccurate at computing tail probabilities.

Importance sampling (IS) is a technique that facilitates efficient rare-event simulation.[2] In our setting, it entails the sampling under an "importance" measure (distinct from the reference distribution) for which the simulated event times occur more frequently. A classical measure change that yields such outcomes is known as exponential twisting (the approach originally dates back to Siegmund (1976)[3]). However, its application typically hinges on the knowledge of a problem related moment generating function (m.g.f.). Computing this function is usually not possible outside of simple event timing

[1] This tail probability is essential to the estimation of several popular (financial) risk measures such as value at risk (VaR) and expected shortfall (ES).

[2] It is well known that plain Monte Carlo is highly inefficient for estimating a rare-event probability, i.e., the number of simulation trials required to estimate this small probability scales in rough proportion to one over its square-root (Asmussen & Glynn 2007, Chapter VI).

[3] Siegmund's algorithm was originally designed for computing gambler's ruin probabilities.

models; for example, when the total intensity follows a (doubly-stochastic) Poisson process. In general, when the event time intensities are governed by correlated stochastic processes that exhibit interaction in the form of contagion or clustering,[4] the m.g.f. is intractable. This prohibits the application of IS schemes based on exponential twisting to many empirically motivated models of event timing.

Motivated by recent work[5], we utilize the Girsanov-Meyer, exponential martingale approach to construct our measure change. This is further coupled with a sequence of zero-variance distributions that converge (weakly) to our importance measure. The latter step provides further efficiency gains by ensuring each simulated path hits the rare event with probability one. This extends the scope of the algorithm to other parts of the distribution besides the (right) tail. Perhaps more importantly, the limiting measure possesses attractive properties for simulation. While the event counting process is no longer a true point process, its arrival times are uniformly distributed up to horizon. Furthermore, the remaining stochastic variables (those pertaining to the original intensity process) have the same dynamics as under the reference measure. Thus, remarkably, the simulation is simpler and faster than that of the plain MC estimator. Moreover, our approach facilitates a reduction in the sampling bias that often plagues event time simulation estimators. For a certain class of models, the simulation bias may be entirely eliminated (see Giesecke & Shkolnik (2018) for an in depth discussion).

Rare-event simulation literature often emphasizes asymptotic optimality[6] of IS estimators. However, a trade off usually arises between the restrictive conditions required for optimality and algorithm specifications that are too difficult to implement. We pursue weaker notions of optimality while preserving the speed and simplicity of our algorithms.[7] To this end, we prove that under mild conditions our estimator outperforms plain MC in terms of variance reduction. This ensures our estimator is almost always preferred to this standard option. Moreover, while most IS schemes are parametrized, we avoid the computation of any "optimal" parameter prior to, or during the simulation.

[4]These models are typically referred to as "self-exciting" point processes in the literature.

[5]e.g., see Giesecke & Shkolnik (2018), Blanchet & Ruf (2016), Giesecke & Shkolnik (2014) and Vanden-Eijnden & Weare (2012).

[6]This condition demands that the second moment of the estimator under the importance measure decay at twice the rate of decay of the rare event probability.

[7]It should also be noted that significant gains in speed can often offset the increased variance with a greater number of simulation trials (albeit at the slow, CLT square-root rate).

We test our algorithms on several credit portfolio applications using a popular reduced-form model of correlated name-by-name default timing. The model incorporates frailty and self-exciting features central to corporate debt modeling (see Duffie, Eckner, Horel & Saita (2009) and Azizpour, Giesecke & Schwenkler (2014) for example). Our numerical examples are motivated by risk management and bond pricing case studies, and illustrate the superior performance of our scheme over plain MC. In comparison, standard IS techniques (based on exponential twisting) cannot be applied to these case studies.

Previous research has studied tail estimation in the context of event timing simulation. A thorough review of this literature may be found in Giesecke & Shkolnik (2018) and Giesecke & Shkolnik (2014). We do not discuss here the literature on exponential twisting since its scope is restricted in our setting. Several authors have developed interacting particle schemed due to such difficulties (e.g., Del Moral & Garnier (2005), Carmona, Fouque & Vestal (2009), Carmona & Crépey (2010)). Our work is most closely related to Giesecke & Shkolnik (2010) which develops an asymptotically optimal IS scheme for Markov chain models under a strict set of assumptions.

The rest of this paper is structured as follows. Section 2 formulates the problem of our interest. Section 3 develops the proposed conditional IS scheme. Section 4 analyzes the asymptotic efficiency of the conditional IS estimator. Section 5 illustrates numerical examples. Section 6 concludes and proofs are provided as Appendices.

2 Problem Formulation

2.1 Preliminaries

Fix a measure space $(\Omega, \mathcal{F})$ equipped with a right continuous and complete information filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions. For some $n \in \mathbb{N}$, consider a set of distinct and totally inaccessible stopping times $\{\chi_i\}_{i=1}^n$ on $(\Omega, \mathcal{F})$. Each time is associated with an event indicator process $N_t^i = \mathbf{1}_{\{\chi_i \leq t\}}$ for $t \geq 0$, where $\mathbf{1}_{\mathscr{A}}$ is the indicator of $\mathscr{A} \in \mathcal{F}$. We define $N = (N^1, \ldots, N^n)$ and let $\{\tau_i\}_{i=1}^n$ be the *order* statistics of $\{\chi_i\}_{i=1}^n$. Thus, τ_k is the kth arrival of the aggregate counting process $\overline{N} = \sum_{i=1}^n N^i$.

Let $\mathbf{P}$ be the probability measure of interest on $(\Omega, \mathcal{F})$. In a statistical application such as a real-world risk management, one takes the physical probability measure as $\mathbf{P}$. For purposes of pricing derivatives, a risk-neutral measure should be taken as $\mathbf{P}$. Suppose that each N^i admits a stochastic $\mathbf{P}$-intensity p^i of the form $p^i = x^i(1 - N^i)$ for a positive, càdlàg process x^i. Each x^i is a conditional rate of default with the cumulant rate given by $H^i_\cdot = \int_0^\cdot x^i_s ds$. It holds that

$$N^i_\cdot - H^i_{\cdot \wedge \chi_i} \tag{1}$$

forms a $\mathbf{P}$-martingale for $i = 1, 2, \ldots, n$. We further assume each $H^i_\infty = \infty$, which is equivalent to the assumption that $\chi_i < \infty$ almost surely. If we define $\overline{p} = \sum_{i=1}^n p^i$, it follows that $\overline{N}$ admits $\overline{p}$ as its $\mathbf{P}$-intensity, where $\overline{p}$ is strictly positive almost everywhere on $[0, \tau_n)$ with $\tau_n < \infty$ almost surely under $\mathbf{P}$.

2.2 Objective

We wish to compute the right-tail probability of $\overline{N}$ given by

$$Y^\xi = \mathbf{P}\left(\overline{N}_T \geq \xi\right) \tag{2}$$

for a fixed horizon $T > 0$ and a given $\xi \in \{1, 2, \ldots, n\}$.[8] The true value of Y^ξ could be estimated via a *plain* Monte Carlo (pMC) simulation by generating a set of i.i.d. $\mathbf{P}$-samples $\left\{\hat{y}^\xi_m\right\}_{m=1}^M$ of $\mathbf{1}_{\{\overline{N}_T \geq \xi\}}$ for $M \in \mathbb{N}$, where the empirical mean given by

$$\hat{Y}^\xi_M = \frac{1}{M} \sum_{m=1}^{M} \hat{y}^\xi_m \tag{3}$$

is the pMC-estimator of Y^ξ.

However, the pMC method is computationally inefficient, as the central limit theorem provides that we need K times more replications to reduce the confidence interval of the pMC estimator by $\sqrt{K}$, regardless of the number of dimensions. A heavy computational burden is unavoidable under the pMC scheme, if one is interested in the rare event with a

[8] We are also interested in computing $\mathbf{P}\left(\overline{N}_T = \xi\right)$ for $\xi \in \{0, 1, 2, \ldots, n\}$.

small probability. In general, the pMC method would fail to induce a decent coverage of the confidence interval unless M is much larger than one over the probability of interest.

Moreover, the pMC scheme is subject to potential biases, if the implementation requires the discretization of the intensity process. For example, the conventional time-scaling algorithms based on time-change theorem of Meyer (1971) should be biased, unless the time-integrated intensity process can be exactly drawn. Giesecke & Shkolnik (2018) point out three potential sources of bias in the trivial time-scaling scheme: (i) intensity discretization, (ii) numerical integration, and (iii) event-time location.

The root mean square error (RMSE) is commonly used to evaluate the performance of different Monte Carlo methods. The RMSE is given by $\sqrt{\text{bias}^2 + \text{SE}^2}$, where the bias is defined as the difference between Y^ξ and $\hat{Y}^\xi_M$ with M as the number of simulation trials. The standard error (SE) is the estimated sample standard deviation of $\{\hat{y}^\xi_m\}_{m=1}^M$ divided by $\sqrt{M}$. Our goal is to achieve a relatively fast convergence rate of RMSE with a given computational budget by adopting an efficient simulation measure specific to the event of interest.

3 Conditional Importance Sampling Measure

We propose a method of *conditional* IS (cIS), which is facilitated by adaptive measure changes conditional on the event of interest. For a given simulation horizon $T > 0$, define $\mathcal{E}^\xi_T = \left\{ \overline{N}_T \geq \xi \right\}$ and

$$\mathcal{L}^\xi_T = \exp\left(- \int_0^{\tau_\xi} \overline{p}_s ds \right) \prod_{i=1}^{\xi} \overline{p}_{\tau_i -} \frac{T^\xi}{\xi!} \tag{4}$$

for $\xi \in \mathbb{N}$. We construct the cIS simulation measure specific to the event $\mathcal{E}^\xi_T$ as

$$\mathbf{Q}^\xi_T(\mathscr{A}) = \mathbf{E}^{\mathbf{P}} \left(\frac{\mathbf{1}_{\mathcal{E}^\xi_T \cap \mathscr{A}}}{\mathcal{L}^\xi_T} \right) \tag{5}$$

for all $\mathscr{A} \in \mathcal{G}_\xi$, where $\mathcal{G}_\xi = \mathcal{F}_{\tau_\xi}$.[9]

[9] By design $\mathbf{Q}^\xi_T$ is absolutely continuous with respect to $\mathbf{P}$.

Algorithm 1 Conditional Monte Carlo to estimate $\mathbf{P}\left(\overline{N}_T \geq \xi\right)$

procedure CONDITIONALIS(ξ, T, M) $\triangleright$ M is the number of cIS trials
  Initialize $\hat{Y} \leftarrow 0$
  for $m \in \{1, \ldots, M\}$ **do**
    Draw an ordered sample of $\{\tau_1, \ldots, \tau_\xi\}$ from $U(0, T)$
    Compute $\mathcal{L}_T^\xi$ from a conditional sample of $\overline{p}$ under $\mathbf{Q}_T^\xi$
    Update $\hat{Y} \leftarrow \hat{Y} + \mathcal{L}_T^\xi$
  end for
  return $\hat{Y}/M$
end procedure

Theorem 3.1. *$\mathbf{Q}_T^\xi$ is a probability measure on $\left(\Omega, \mathcal{G}_\xi\right)$ and $\mathbf{P}\left(\mathcal{E}_T^\xi\right) = \mathbf{E}^{\mathbf{Q}_T^\xi}\left(\mathcal{L}_T^\xi\right)$; i.e., the estimator of $\mathbf{P}\left(\mathcal{E}_T^\xi\right)$ under $\mathbf{Q}_T^\xi$ is $\mathcal{L}_T^\xi$. Moreover,*

(i) If $\{u_i\}_{i=1}^\xi$ denotes the order statistics of ξ i.i.d. uniform random variables on $[0, T]$, we have $\tau_i \overset{\mathcal{D}}{=} u_i$ under $\mathbf{Q}_T^\xi$ for $i = 1, \ldots, \xi$.

(ii) $\mathbf{Q}_T^\xi\left(I_k = i \middle| \mathcal{F}_{\tau_k -}\right) = \mathbf{P}\left(I_k = i \middle| \mathcal{F}_{\tau_k -}\right) = p^i \big/ \overline{p}$ holds for all $k = 1, \ldots, \xi$, where I_k is the component of N at which the k^{th} event of $\overline{N}$ occurs.

(iii) Any $\mathbf{P}$-local martingale that has no jumps in common with $\overline{N}$ is a $\mathbf{Q}_T^\xi$-local martingale.

Proof. See Appendix A.1. □

The proposed cIS algorithm, summarized in Algorithm 1, does not require *ad-hoc* level selection or tuning procedures to compute $\mathbf{P}\left(\overline{N}_T \geq \xi\right)$. The cIS algorithm generates an unbiased estimator of $Y = \mathbf{P}\left(\overline{N}_T \geq \xi\right)$, if one can evaluate $\mathcal{L}_T^\xi$ exactly under $\mathbf{Q}_T^\xi$. Specifically, the cIS estimator is unbiased, if one can exactly sample the skeleton of $\overline{p}$ on $\{\tau_1, \ldots, \tau_\xi\}$ uniformly distributed on $[0, T]$, and evaluate the time-integrated transform $\exp\left(-\int_0^{\tau_\xi} \overline{p}_s ds\right)$ without biases. Giesecke & Shkolnik (2018) find that the cIS estimator $\mathcal{L}_T^\xi$ can be sampled both exactly and efficiently in a Markovian setting under some conditions, when exact samples of $\overline{p}$ can be obtained and the bridge transform $\mathbf{E}^{\mathbf{Q}_Y^\xi}\left(e^{-\int_t^s \overline{p}_u du} \middle| \overline{p}_t, \overline{p}_s\right)$ can be evaluated in closed-form for $0 \leq t < s$.

Corollary 3.2. *A similar cIS algorithm is applicable in that* $\mathbf{P}\left(\overline{N}_T = \xi\right) = \mathbf{E}^{\widehat{\mathbf{Q}}_T^{\xi}}\left(\widehat{\mathcal{L}}_T^{\xi}\right)$, *where the limit conditional measure* $\widehat{\mathbf{Q}}_T^{\xi}$ *is redefined for the modified set* $\widehat{\mathcal{E}}_T^{\xi} = \left\{\overline{N}_T = \xi\right\}$ *and* $\widehat{\mathcal{L}}_T^{\xi}$ *is given by*

$$\widehat{\mathcal{L}}_T^{\xi} = \exp\left(-\int_0^T \overline{p}_s ds\right) \prod_{j=1}^{\xi} \overline{p}_{\tau_j -} \frac{T^{\xi}}{\xi!} \tag{6}$$

for a fixed simulation horizon $T > 0$. *Parts (i)–(iii) of Theorem 3.1 hold.*

Proof. See Appendix A.2. □

4 Asymptotic Optimality

This type of analysis is usually carried out in some appropriate asymptotic regime. We let $\xi = \lceil \mu n \rceil$ for $\mu \in (0, 1)$ and adopt the rare event regime $y_n = \mathbf{P}(\overline{N}_T \geq \lceil \mu n \rceil) \to 0$ as $n \uparrow \infty$. Observe that

$$\frac{1}{n}\mathbf{E}^{\mathbf{P}}(\overline{N}_T) = \int_0^T \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}^{\mathbf{P}}(p_s^i)\, ds. \tag{7}$$

Thus, $\frac{1}{n}\sum_{i=1}^{n} \mathbf{E}^{\mathbf{P}}(p_t^i) < \frac{\mu}{T}$ for all $0 \leq t \leq T$ and all n large enough is a sufficient condition for the rare event regime to be satisfied. It is not difficult to show that variance of the plain MC estimator $\mathbf{Var}^{\mathbf{P}}(\mathbf{1}_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}) = y_n(1 - y_n)$. We compare this to the $\mathbf{Q}_T^{\xi}$-variance of the estimator $\mathcal{L}_T^{\xi}$ to obtain the following result.

Theorem 4.1. *Consider a* σ*-algebra* $\mathcal{H} = \sigma\left(\mathbf{1}_{\{\tau_{\lceil \mu n \rceil} \leq T\}}\right)$ *and define* $\mathcal{H}_k = \mathcal{F}_{\tau_k} \vee \mathcal{H}$ *for* $k = 0, 1, \ldots n$ *with* $\tau_0 = 0$. *Suppose that* $\frac{1}{n}\sum_{i=1}^{n} \mathbf{E}^{\mathbf{P}}\left(p_{\tau_k -}^i \middle| \mathcal{H}_{k-1}\right) < \frac{2\mu}{eT}$ *holds for* $k = 1, \ldots, \lceil \mu n \rceil$ *and n sufficiently large under* $\mathbf{P}$ *almost surely. Then,*

$$\frac{\mathbf{Var}^{\mathbf{Q}_T^{\xi}}(\mathcal{L}_T^{\xi})}{\mathbf{Var}^{\mathbf{P}}(\mathbf{1}_{\{\overline{N}_T \geq \lceil \mu n \rceil\}})} \to 0 \qquad as \quad n \uparrow \infty. \tag{8}$$

Remark 4.2. *The assumption on the* $\{p^i\}$ *may be further relaxed in terms of convergence rates as* $n \uparrow \infty$. *Note that the condition is a stronger version of the rare event* *condition above and intuitively requires the average event count intensity to be small.*

Proof. Under our hypothesis we have $y_n \to 0$ under $n \uparrow \infty$. Expanding (8), we obtain

$$\lim_{n\uparrow\infty} \frac{\mathbf{E}^{\mathbf{P}}(\mathcal{L}_T^{\xi}\mathbf{1}_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}) - y_n^2}{y_n(1-y_n)} = \lim_{n\uparrow\infty} \frac{\mathbf{E}^{\mathbf{P}}(\mathcal{L}_T^{\xi}\mathbf{1}_{\{\overline{N}_T \geq \lceil \mu n \rceil\}})}{y_n}. \tag{9}$$

By Jensen's inequality $\mathbf{E}^{\mathbf{P}}(\mathcal{L}_T^{\xi}\mathbf{1}_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}) = \mathbf{E}^{\mathbf{Q}_T^{\xi}}\left((\mathcal{L}_T^{\xi})^2\right) \geq y_n^2$. Also, since

$$\mathbf{E}^{\mathbf{P}}\left(\prod_{i=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_i -}\mathbf{1}_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}\right) = \mathbf{E}^{\mathbf{P}}\left(\mathbf{1}_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}\mathbf{E}^{\mathbf{P}}\left(\overline{p}_{\tau_1 -} \cdots \mathbf{E}^{\mathbf{P}}\left(\overline{p}_{\tau_{\lceil \mu n \rceil}-} \middle| \mathcal{H}_{\lceil \mu n \rceil - 1}\right) \cdots \middle| \mathcal{H}_0\right)\right)$$

applying our assumptions,

$$\mathbf{E}^{\mathbf{P}}\left(\mathcal{L}_T^{\xi}\mathbf{1}_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}\right) < \frac{(2\lceil \mu n \rceil / e)^{\lceil \mu n \rceil}}{\lceil \mu n \rceil !}\mathbf{E}^{\mathbf{P}}\left(\exp\left(-\int_0^{\tau_{\lceil \mu n \rceil}} \overline{p}_s\, ds\right)\mathbf{1}_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}\right) \tag{10}$$

Since the two random variables under the expectation are negatively associated and that $\int_0^{\tau_{\lceil \mu n \rceil}} \overline{p}_s\, ds$ is equal in $\mathbf{P}$-law to $\lceil \mu n \rceil$ standard i.i.d. exponential (Meyer 1971), we obtain

$$\mathbf{E}^{\mathbf{P}}(\mathcal{L}_T^{\xi}\mathbf{1}_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}) < \frac{e^{-\lceil \mu n \rceil}\lceil \mu n \rceil^{\lceil \mu n \rceil}}{\lceil \mu n \rceil !} y_n, \tag{11}$$

which provides that the right side of (9) tends to zero for large n as required. □

5 Numerical Examples

This section illustrates the performance of the cIS scheme in different settings. The simulations were performed on a PC with an Intel® Core™ i7-4790 3.60 GHz processor and 8.0 GB RAM. The code was written in julia Version 0.6.0.

5.1 Networked default clustering

Multivariate counting process models are widely used to measure and manage the correlated default risk in the system. For a fixed $n \in \mathbb{N}$ defaultable names, the stopping times $\{\chi_i\}_{i=1}^n$ models the default times of the individual constituent of the system, and a central (systemic) quantity of interest is the total default count by a fixed time $T > 0$, is given by $\overline{N}_T$.

Our objective is to estimate the tail probability $\mathbf{P}\left(\overline{N}_T \geq \xi\right)$ for various ξ, where $\mathbf{P}$ refers to the physical probability measure. In a bottom-up formulation, consider a systematic risk factor $\eta^0 \geq 0$ and a set of idiosyncratic factor processes $\{\eta^i\}_{i=1}^n$ so that the state process $X = (x^1, \dots, x^n)$ is given by

$$x^i = \omega_i \eta^0 + \eta^i, \tag{12}$$

where each default indicator process N^i admits $p^i = x^i\left(1 - N^i\right)$ as its $\mathbf{P}$-intensity (see Section 2.1). Here, $\omega_i > 0$ is the systematic factor loading of the i^{th} name in the system. We further assume that η^0 is the strong solution of the SDE given by

$$d\eta_t^0 = \kappa_0(\theta_0 - \eta_t^0)dt + \sigma_0\sqrt{\eta_t^0}dW_t^0, \tag{13}$$

where W^0 is a $\mathbf{P}$-Brownian motion with some $\kappa_0 > 0$, $\theta_0 > 0$ and $\sigma_0 > 0$. Furthermore, we assume that η^i is governed by the SDE under the statistical probability measure $\mathbf{P}$

$$d\eta_t^i = \kappa_i(\theta_i - \eta_t^i)dt + \sigma_i\sqrt{\eta_t^i}dW_t^i + \delta_i \cdot dN_t, \tag{14}$$

where $(W^1, \dots, W^n)$ is a vector of mutually independent $\mathbf{P}$-Brownian motions, $\kappa_i > 0$ is the mean-reversion rate, $\theta_i > 0$ is the long-run mean level, $\sigma_i > 0$ is the diffusive volatility, and the vector $\delta_i = (\delta_{i1}, \dots, \delta_{in})$ represents name i's sensitivity to defaults in the system for $i = 1, \dots, n$. Similar models are specified and analyzed by Giesecke, Kim & Zhu (2011) among others.

For each η^i, we introduce the *interarrival* process h^{ij} satisfying

$$\eta_t^i = \sum_{j=1}^{n} h_{t-\tau_{j-1}}^{ij} \mathbf{1}_{\left\{\overline{N}_t = j-1\right\}}, \tag{15}$$

where $\tau_0 = 0$ and $h_0^{ij} = \eta_{\tau_{j-1}}^i$ for $j = 1, \dots, n$. As the process h^{ij} follows the Feller diffusion $dh_t^{ij} = \kappa_i(\theta_i - h_t^{ij})dt + \sigma_i\sqrt{h_t^{ij}}dW_t^i$, the distribution of h_s^{ij} for $s > 0$ given h_0^{ij} is non-central chi-squared upto a scale factor. As shown by Cox, Ingersoll & Ross

Algorithm 2 The cIS algorithm to estimate $\mathbf{P}\left(\overline{N}_T \geq \xi\right)$ with $T > 0$ under the default intensity model described in equations (12)–(14)

procedure cIS_TailProbability(ξ, T, M) $\quad\triangleright$ M is the number of cIS trials
  Initialize $\hat{Y} \leftarrow 0$
  for $m \in \{1, \ldots, M\}$ **do**
    Initialize $\mathcal{K} \leftarrow \{1, \ldots, n\}$ and $x_0^i \leftarrow \omega^i \eta_0^0 + \eta_0^i \;\; \forall i \in \mathcal{K}$
    Set $t \leftarrow 0$ and $\mathcal{L}_T^{\xi} \leftarrow T^{\xi}/\xi!$
    Draw an ordered sample of $\{\tau_1, \ldots, \tau_\xi\}$ from $U(0, T)$
    for $k \in \{1, \ldots, \xi\}$ **do**
      Sample $\eta_{\tau_k}^0$ given η_t^0 from the transition density of h^{0k}
      Update $\mathcal{L}_T^{\xi} \leftarrow \mathcal{L}_T^{\xi} \times \varphi^{0k}\left(\sum_{i\in\mathcal{K}} \omega^i,\; \tau_k - t;\; h_0^{0k},\; h_{\tau_k - t}^{0k}\right)$
      for $i \in \mathcal{K}$ **do**
        Sample $\eta_{\tau_k}^i$ given η_t^i from the transition density of h^{ik}
        Update $\mathcal{L}_T^{\xi} \leftarrow \mathcal{L}_T^{\xi} \times \varphi^{ik}\left(1,\; \tau_k - t;\; h_0^{ik},\; h_{\tau_k - t}^{ik}\right)$
        Set $x_{\tau_k}^i \leftarrow \omega^i \eta_{\tau_k}^0 + \eta_{\tau_k}^i$
      end for
      Update $\mathcal{L}_T^{\xi} \leftarrow \mathcal{L}_T^{\xi} \times \sum_{i\in\mathcal{K}} x_{\tau_k}^i$
      Draw index $I_k = j$ from the distribution $\left\{x^j / \sum_{i\in\mathcal{K}} x^i\right\}_{j\in\mathcal{K}}$ and set $t \leftarrow \tau_k$
      Update $\mathcal{K} \leftarrow \mathcal{K} - \{j\}$ and $\eta_t^i \leftarrow \eta_t^i + \delta_{ij} \; \forall i \in \mathcal{K}$
    end for
    Update $\hat{Y} \leftarrow \hat{Y} + \mathcal{L}_T^{\xi}$
  end for
  return $\hat{Y}/M$
end procedure

(1985), h_s^{ij} can be drawn from the non-central chi-squared transition density from h_0^{ij} as

$$\mathbf{P}\left(h_s^{ij} \leq z \middle| h_0^{ij}\right) = F_{\chi_d'^2(\nu)}\left(\frac{4\kappa z}{\sigma^2\left(1 - e^{-\kappa s}\right)}\right), \tag{16}$$

where $d = \frac{4\kappa\theta}{\sigma^2}$ is the degree of freedom and $\nu = \frac{4\kappa e^{-\kappa s} h_0^{ij}}{\sigma^2\left(1 - e^{-\kappa s}\right)}$ is the non-centrality parameter. Moreover, the bridge transform

$$\varphi^{ij}(a, s; h_s^{ij}, h_0^{ij}) = \mathbf{E}^{\mathbf{P}}\left(e^{-a\int_0^s h_t^{ij} dt} \middle| h_s^{ij}, h_0^{ij}\right) \tag{17}$$

can be computed analytically for $a > 0$; see Broadie & Kaya (2006) for a formula. Algorithm 2 summarizes the cIS algorithm to estimate $\mathbf{P}\left(\overline{N}_T \geq \xi\right)$ with $T > 0$.

Table 1: Estimation results of $\mathbf{P}(\overline{N}_T \geq \xi)$ for $T = 1$ year under the default intensity model described in equations (12)–(14).

ξ	cIS est.	99% C.I.	**JAM** est.	cIS error	cIS var.	Var. ratio
5	4.1304E-01	$\pm$ 1.6602E-03	4.1170E-01	1.10336	2.0769E-01	1.1021E+00
6	2.7401E-01	$\pm$ 1.0668E-03	2.7246E-01	1.06868	8.5751E-02	2.1626E+00
7	1.7050E-01	$\pm$ 6.4858E-04	1.7019E-01	1.04420	3.1696E-02	4.1051E+00
8	1.0006E-01	$\pm$ 3.7045E-04	9.9831E-02	1.01630	1.0340E-02	7.8809E+00
9	5.5701E-02	$\pm$ 2.0704E-04	5.5506E-02	1.02030	3.2299E-03	1.4503E+01
10	2.9500E-02	$\pm$ 1.1127E-04	2.9830E-02	1.03533	9.3281E-04	2.7122E+01
11	1.5061E-02	$\pm$ 5.8152E-05	1.5335E-02	1.05983	2.5481E-04	5.0921E+01
12	7.3892E-03	$\pm$ 2.9299E-05	7.5448E-03	1.08842	6.4683E-05	9.7959E+01
13	3.4710E-03	$\pm$ 1.3591E-05	3.5546E-03	1.07480	1.3917E-05	2.1992E+02
14	1.5909E-03	$\pm$ 6.6923E-06	1.5886E-03	1.15469	3.3747E-06	4.1492E+02
15	7.0593E-04	$\pm$ 3.1368E-06	6.9904E-04	1.21972	7.4138E-07	7.9935E+02
16	3.0338E-04	$\pm$ 1.3589E-06	2.9016E-04	1.22958	1.3915E-07	1.9433E+03
17	1.2631E-04	$\pm$ 5.7287E-07	1.4004E-04	1.24495	2.4728E-08	4.9734E+03
18	5.1644E-05	$\pm$ 2.4117E-07	6.1286E-05	1.28187	4.3825E-09	1.2052E+04
19	2.0557E-05	$\pm$ 1.0167E-07	2.5432E-05	1.35761	7.7891E-10	3.3811E+04
20	7.9875E-06	$\pm$ 4.2342E-08	9.1633E-06	1.45512	1.3509E-10	9.9570E+04
21	3.0265E-06	$\pm$ 1.6349E-08	3.0732E-06	1.48279	2.0139E-11	2.6282E+05
22	1.1232E-06	$\pm$ 6.1108E-09	1.5413E-06	1.49348	2.8137E-12	9.4529E+05
23	4.0873E-07	$\pm$ 2.5559E-09	N/A	1.71654	4.9223E-13	N/A
24	1.4514E-07	$\pm$ 8.4581E-10	N/A	1.59963	5.3905E-14	N/A

Note. This table reports cIS ($M = 5 \times 10^5$) and **JAM** ($M = 5 \times 10^5$) estimates of $Y = \mathbf{P}\left(\overline{N}_T \geq \xi\right)$ for 20 values of ξ. The **JAM** estimates are reported in the fourth column. The fifth column reports the relative error $\hat{\sigma}_Y / Y$ of the cIS estimator. The sixth column reports the $\mathbf{Q}_T^{\xi}$-variance of the cIS estimator. The last column reports the variance ratio, the sample variance of the exact **JAM** estimator over the sample variance of the cIS estimator.

We present estimates of the tail probability $\mathbf{P}(\overline{N}_T \geq \xi)$ in Table 1 for the portfolio of size $n = 100$ with time horizon $T = 1$ year. For each name of $i = 1, \ldots, n$, we uniformly draw ω_i from $[0, 1]$, κ_i from $[0.5, 1.5]$, θ_i from $[0.001, 0.051]$, and set $\sigma_i = \min(\sqrt{2\kappa_i\theta_i}, \bar{\sigma}_i)$, where $\bar{\sigma}_i$ is drawn from $[0, 0.2]$ uniformly. That is, the parameters of η^0 and η^i satisfy the Feller condition to ensure that $x^i > 0$ holds under $\mathbf{P}$ almost surely for $i = 1, \ldots, n$. We set the initial value $\eta_0^i = \theta_i$. For the systematic factors, we set $\theta_0 = 0.02$, $\kappa_0 = 1.0$, $\sigma_0 = 0.1$, and $\eta_0^0 = \theta_0$. The jump sensitivies are constructed by drawing each δ_{ij} from $[0, 1/n]$ uniformly. The selected parameters model a system with $\mathbf{P}(\overline{N}_T = 0) = 0.0281186615220496$.

In Table 1, we compare the cIS estimator with the exact *jump approximation method*

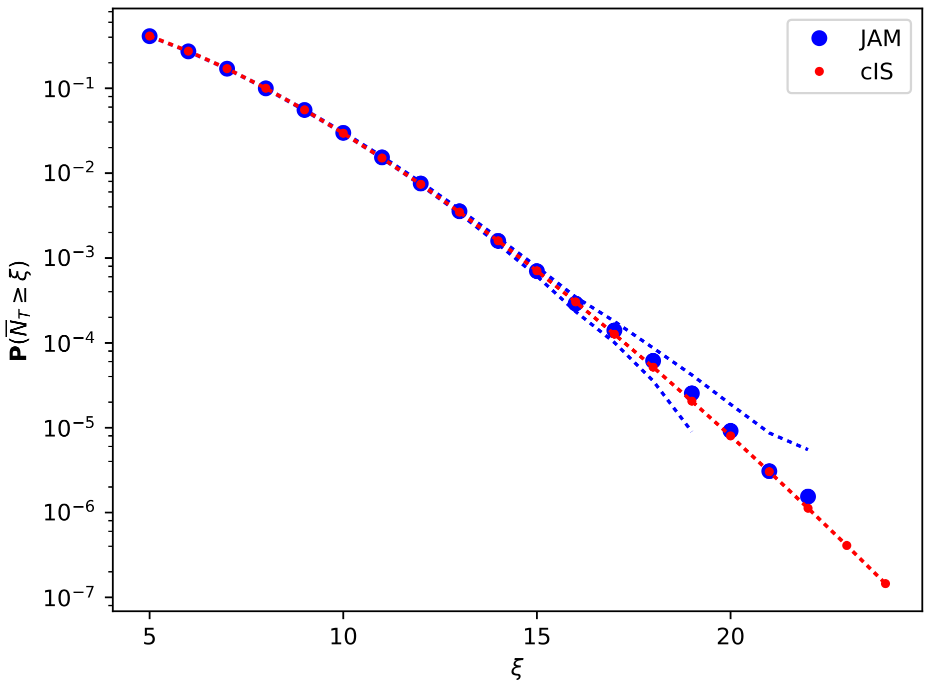


Figure 1: The cIS and **JAM** estimates (markers) and 99% confidence intervals (dashed lines) for $\mathbf{P}(\overline{N}_T \geq \xi)$ for $\xi \in \{5, 6, \ldots, 24\}$ with $M = 5 \times 10^5$. Confidence intervals extending below zero are omitted.

(**JAM**) estimator, which is shown to be superior to the naïve pMC estimator; see Giesecke & Shkolnik (2018) for reference. Figure 1 shows the cIS and **JAM** estimates along with their 99% confidence intervals by running $M = 5 \times 10^5$ trials of cIS and exact **JAM** trials, respectively.[10] Note that the exact **JAM** method fails to generate the rare events and return $[0, 0]$ as the *empirical* confidence interval of $\mathbf{P}(\overline{N}_T \geq \xi)$ for $\xi \geq 23$. However, the event $\{\overline{N}_T \geq \xi\}$ always occurs under the cIS simulation measure $\mathbf{Q}_T^{\xi}$ for any choice of ξ. As shown, a significant variance reduction that can be achieved by the cIS scheme for the rare-event probability estimation.

The full distribution of $\overline{N}_T$ can be estimated via the cIS scheme in the context of Corollary 3.2. The distribution of $\overline{N}_T$ under $\mathbf{P}$ can be represented by $\{\phi_\xi\}_{\xi=0}^{n}$, where $\phi_\xi = \mathbf{P}\left(\overline{N}_T = \xi\right)$. Table 2 and Figure 2 illustrate that the cIS scheme can substantially reduce the variance of $\mathbf{P}\left(\overline{N}_T = \xi\right)$ for large ξ, if compared to the performance of the

[10] Separate cIS experiments need to be performed for each of the values of ξ.

Table 2: Estimation results of $\mathbf{P}(\overline{N}_T = \xi)$ for $T = 1$ year under the default intensity model described in equations (12)–(14).

ξ	cIS est.	99% C.I.	**JAM** est.	cIS error	cIS var.	Var. ratio
5	1.3866E-01	± 1.1543E-04	1.3924E-01	0.22850	1.0039E-03	1.1929E+02
6	1.0357E-01	± 1.0546E-04	1.0227E-01	0.27953	8.3809E-04	1.0674E+02
7	7.0413E-02	± 8.5618E-05	7.0358E-02	0.33377	5.5234E-04	1.1318E+02
8	4.4282E-02	± 6.2882E-05	4.4325E-02	0.38980	2.9794E-04	1.3329E+02
9	2.6088E-02	± 4.2267E-05	2.5676E-02	0.44474	1.3461E-04	1.7135E+02
10	1.4518E-02	± 2.6450E-05	1.4495E-02	0.50010	5.2715E-05	2.4588E+02
11	7.6823E-03	± 1.5649E-05	7.7905E-03	0.55914	1.8451E-05	3.7449E+02
12	3.8926E-03	± 8.7986E-06	3.9901E-03	0.62046	5.8332E-06	6.0045E+02
13	1.8894E-03	± 4.6787E-06	1.9660E-03	0.67972	1.6494E-06	1.0360E+03
14	8.8565E-04	± 2.4311E-06	8.8958E-04	0.75351	4.4534E-07	1.7275E+03
15	4.0176E-04	± 1.1895E-06	4.0888E-04	0.81271	1.0661E-07	3.2547E+03
16	1.7610E-04	± 5.6182E-07	1.5011E-04	0.87577	2.3783E-08	5.3459E+03
17	7.5104E-05	± 2.5167E-07	7.8757E-05	0.91982	4.7723E-09	1.3668E+04
18	3.1105E-05	± 1.0989E-07	3.5853E-05	0.96980	9.0995E-10	3.2286E+04
19	1.2557E-05	± 4.8029E-08	1.6269E-05	1.04993	1.7381E-10	7.6550E+04
20	4.9464E-06	± 2.0352E-08	6.0902E-06	1.12942	3.1210E-11	1.5060E+05
21	1.8998E-06	± 8.2876E-09	1.5319E-06	1.19743	5.1753E-12	2.2671E+05
22	7.1593E-07	± 3.1627E-09	1.5413E-06	1.21262	7.5369E-13	1.5760E+06
23	2.6348E-07	± 1.3013E-09	N/A	1.35571	1.2759E-13	N/A
24	9.4678E-08	± 4.6784E-10	N/A	1.35639	1.6492E-14	N/A

Note. This table reports cIS ($M = 5 \times 10^5$) and **JAM** ($M = 5 \times 10^5$) estimates of $Y = \mathbf{P}\left(\overline{N}_T = \xi\right)$ for 20 values of ξ. The **JAM** estimates are reported in the fourth column. The fifth column reports the relative error $\hat{\sigma}_Y/Y$ of the cIS estimator. The sixth column reports the $\widehat{\mathbf{Q}}^{\xi}_T$-variance of the cIS estimator. The last column reports the variance ratio, the sample variance of the exact **JAM** estimator over the sample variance of the cIS estimator.

exact **JAM** method. Figure 3 illustrates the cIS estimates and their 99.999% confidence levels of ϕ_ξ for $\xi \in \{0, 1, \ldots, n\}$.

A risk manager and/or policymaker should be concerned about failure of an abnormally large fraction of the total population in the system. The default rate in the system is given by $D_T = \overline{N}_T/n$, where the distribution of $D_T \in [0, 1]$ represents the likelihood of failure by time $T > 0$ of any fraction of the population in the system; see Giesecke & Kim (2011) for similar failure-based measures of systemic risk. To measure and quantify the downside credit risk in the system, quantile-based tail-risk measures are often used. For example, the Value-at-Risk (VaR) at level $\alpha \in (0, 1)$ is defined based

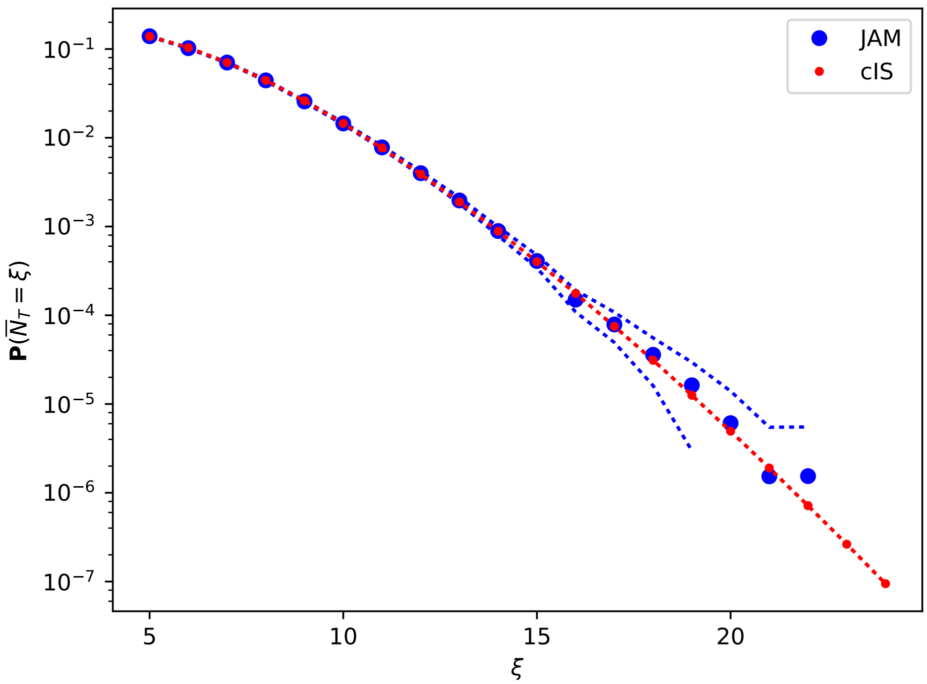


Figure 2: The cIS and **JAM** estimates (markers) and 99% confidence intervals (dashed lines) for $\mathbf{P}(\overline{N}_T = \xi)$ for $\xi \in \{5, 6, \dots, 24\}$ with $M = 5 \times 10^5$. Confidence intervals extending below zero are omitted.

on a descrete distribution as

$$\mathrm{VaR}_\alpha\left(\overline{N}_T\right) = \min\left\{ x \geq 0 : \sum_{\xi=x}^{n} \phi_\xi \leq 1 - \alpha \right\}. \tag{18}$$

Subsequently, the definition of expected shortfall (ES) is given by

$$\mathrm{ES}_\alpha\left(\overline{N}_T\right) = \frac{1}{\alpha} \int_0^\alpha \mathrm{VaR}_\beta\left(\overline{N}_T\right) d\beta, \tag{19}$$

which can be computed based on a discrete distribution as

$$\mathrm{ES}_\alpha\left(\overline{N}_T\right) = \omega_\alpha \mathcal{V} + (1 - \omega_\alpha) \sum_{\xi > \mathcal{V}} \xi \frac{\phi_\xi}{\sum_{\xi > \mathcal{V}} \phi_\xi}, \tag{20}$$

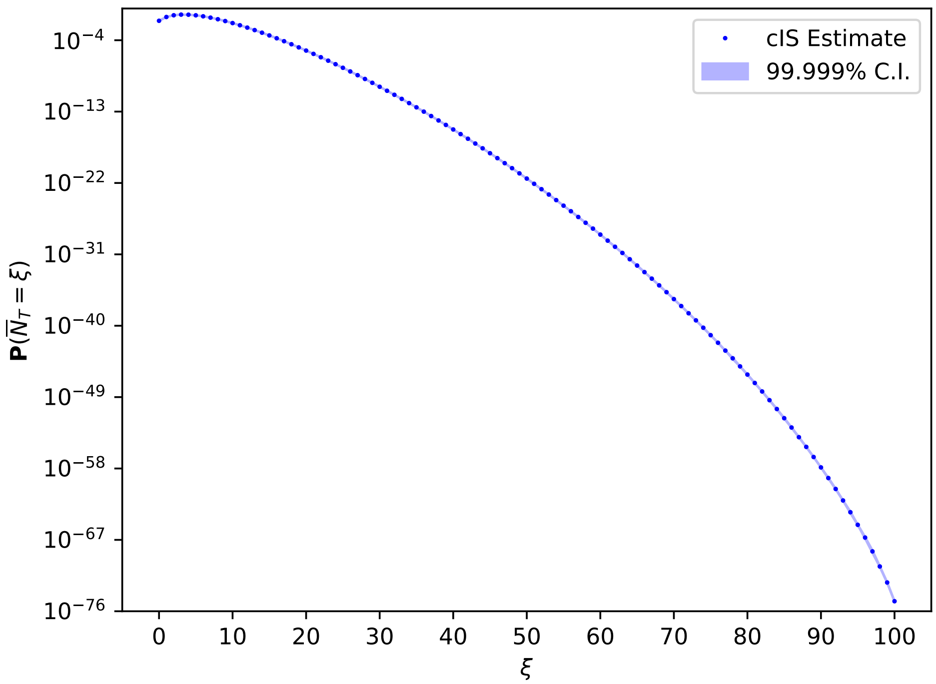


Figure 3: The cIS estimates (markers) and 99.999% confidence intervals (shaded area) of $\mathbf{P}(\overline{N}_T = \xi)$ for $\xi \in \{0, 1, \ldots, 100\}$ with $M = 5 \times 10^5$.

where $\mathcal{V} = \mathrm{VaR}_\alpha(\overline{N}_T)$ and $\omega_\alpha = \left(\sum_{\xi \leq \mathcal{V}} \phi_\xi - \alpha\right) / (1-\alpha)$.[11] Figure 4 depicts both VaR and ES of $\overline{N}_T$ under $\mathbf{P}$ across various confindence levels α between 80% and 99.999% estimated from the cIS scheme with $M = 5 \times 10^5$. It is noteworthy that the *shortfall* measure (s_α) is proposed by Bertsimas, Lauprete & Samarov (2004) as

$$s_\alpha\left(\overline{N}_T\right) = \mathrm{ES}_\alpha\left(\overline{N}_T\right) - \mathbf{E}^{\mathbf{P}}\left(\overline{N}_T\right), \tag{21}$$

where the subtraction from the mean leads to the mean-shortfall systemic risk analysis more akin to the classical mean-variance optimization problem.

[11] As $n \uparrow \infty$, note that $\omega_\alpha \downarrow 0$ holds and $\mathrm{ES}_\alpha\left(\overline{N}_T\right)$ converges to $\mathbf{E}^{\mathbf{P}}\left(\overline{N}_T \middle| \overline{N}_T > \mathrm{VaR}_\alpha\left(\overline{N}_T\right)\right)$.

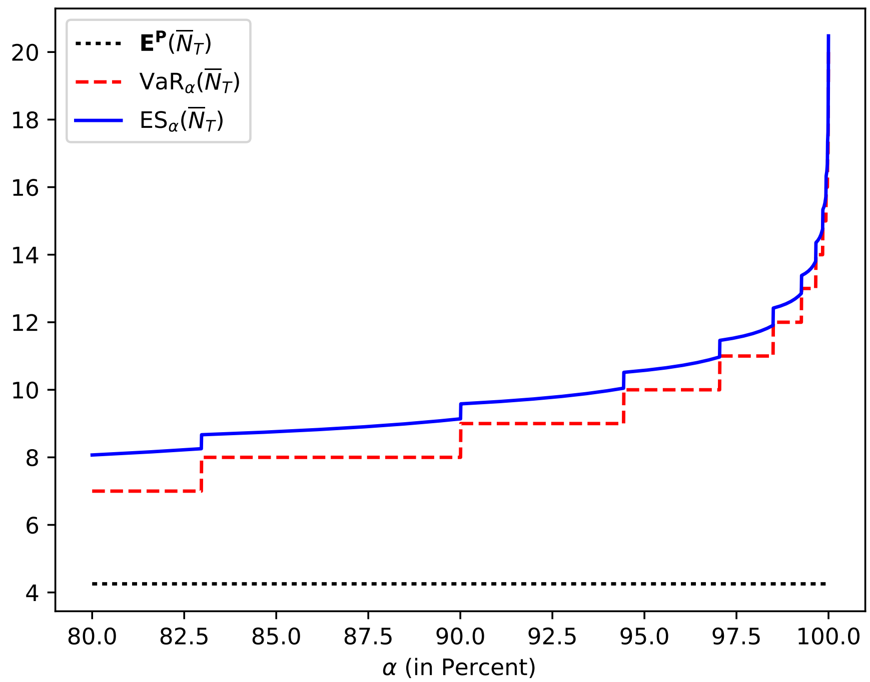


Figure 4: Mean value, Value at Risk (VaR) and Expected shortfall (ES) of $\overline{N}_T$ under $\mathbf{P}$ across various confindence levels α between 80% and 99.999% estimated from the cIS scheme with $M = 5 \times 10^5$.

5.2 Zero-coupon bond pricing

It is well-documented by literature that jumps play an important role in the dynamics of interest rates. For instance, Johannes (2004) emphasizes the role of jumps in continuous-time interest rate models both statistically and economically. However, most of the short-rate models with jump diffusion do not allow closed-form expressions of their integral transforms. This makes it difficult to apply the models to obtain the unbiased pricing estimators of interest rate derivates analytically.

Suppose that the short-rate process follows the SDE under the risk-neutral pricing measure $\mathbf{P}$ given by

$$dr_t = \kappa\left(\theta - r_t\right)dt + \sigma\sqrt{r_t}dW_t + \delta_t d\pi_t, \tag{22}$$

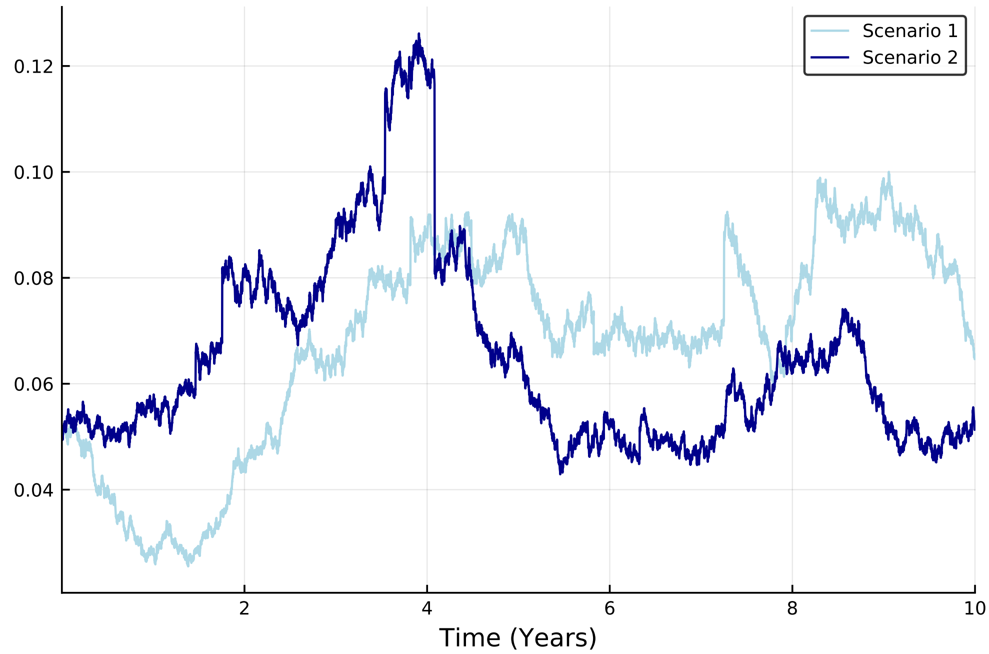


Figure 5: Sample scenarios under the shor-rate model specified in (22) with jump sensitivities given by $\delta_t = r_t \left(\exp\left(\eta_t\right) - 1\right)$, where $\eta_t \sim \text{Normal}\left((\theta - r_t)u, v^2\right)$. The selected parameters are $(\kappa, \theta, \sigma, \lambda_0, \lambda_1, u, v, r_0) = (0.1, 0.08, 0.05, 1.0, 5.0, 1.0, 1.0, 0.05)$. Both positive and negative jumps occur, while r_t keeps its nonnegativity.

where π is the jump arrival counting process with the intensity $\Lambda(r)$, and the jump size is represented by $\delta = \delta(r)$. The jump intensity is assumed to take the form $\Lambda(r_t) = \lambda_0 + \lambda_1 r_t$ for $\lambda_0, \lambda_1 > 0$. We consider a special case when the state-dependent jump sensitivities given by $\delta_t = \delta(r_t) = r_t \left(\exp\left(\eta_t\right) - 1\right)$, where $\eta_t \sim \text{Normal}\left((\theta - r_t)u, v^2\right)$ for $u, v > 0$. In this setting, the log-normal jump component allows both positive and negative jumps, while keeping r_t nonnegative.[12] The sign-indefinite jumps provide additional source of mean reversion; i.e. there is a greater chance of a positive jump at lower interest rate levels than θ, and a higher chance of a negative jump at high levels of interest rate. Figure 5 illustrates the sample scenarios for $T = 10$ years.

Estimating the zero-coupon bond price via cIS is a special case with $n = 1$ and the event indicator process $N^1 = \overline{N}$ admits r as its $\mathbf{P}$-intensity on $[0, \tau_1)$, where $\mathbf{P}$ is the risk-neutral pricing measure. Specifically, the price of a zero-coupon bond with unit

[12] A generalized version of the log-normal jump size model has been introduced by Johannes (2004). Giesecke & Smelov (2013) examine the case of positive jump sizes, which are uniformly distributed between two strictly positive values.

face value and time-to-maturity $T > 0$ takes the form of the time-integrated transform of r given by

$$B_0(T) = \mathbf{E}^{\mathbf{P}}\left(e^{-\int_0^T r_t dt}\right) = 1 - \mathbf{P}\left(\overline{N}_T \geq 1\right) = 1 - \mathbf{E}^{\mathbf{Q}_T^{\xi=1}}\left(\mathcal{L}_T^{\xi=1}\right), \tag{23}$$

where $\mathbf{Q}_T^{\xi=1}$ is the cIS simulation measure conditional on the event $\{\overline{N}_T \geq 1\}$. In other words, the zero-coupon bond price can be interpreted as the $\mathbf{P}$-probability that no events occur during the interval $[0, T]$ in a doubly stochastic counting process with its $\mathbf{P}$-intensity $r \geq 0$.

Note that the joint process of $(r, \overline{N})$ is self-affecting due to the entangled dependence structure between the state of r, the jump intensity $\Lambda(r)$, and the jump size $\delta(r)$. This implies that the jump times of r cannot be generated independently in the simulation of r. However, a notable computational advantage can be attained by taking a sequence of measure changes under the state-dependent specification of the $\mathbf{P}$-intensities of the jump term π. That is, the jump-counting process π and the event-counting process $\overline{N}$ satisfy a so-called doubly stochastic property in that they are conditionally independent given the state of r.

We sample the jump arrivals in the intensity dynamics by the *exact* **JAM** method as illustrated in Giesecke & Shkolnik (2018). Its basic idea is to construct a simulation measure under which the jump arrival intensity is a pure jump process with jumps occuring at the jump arrival times. Specifically, let $0 = \mathcal{S}_0 < \mathcal{S}_1 < \mathcal{S}_2 < \dots$ be the ordered jump arrival times of π. We then construct a jump-simulation measure $\mathbf{Q}^J$ under which the jump arrival times are sampled based on the twisted jump arrival intensity whose paths are piecewise constant between two consecutive jumps; i.e., the measure $\mathbf{Q}^J$ is constructed with the twisted jump intensities ψ defined as $\psi_k = \Lambda(r_{\mathcal{S}_k})$ for $k \geq 0$. The Radon-Nikodym derivative of $\mathbf{Q}^J$ with respect to $\mathbf{P}$ is given by $\mathcal{Z}_\infty^J$, which takes the form of

$$\mathcal{Z}_\infty^J = \exp\left(\int_0^\infty \left(\Lambda(r_s) - \psi_{\pi_s}\right) ds\right) \prod_{i=1}^\infty \frac{\psi_{i-1}}{\Lambda\left(r_{\mathcal{S}_{i-}}\right)}. \tag{24}$$

The Doob martingale of $\mathcal{Z}_\infty^J$ is given by $\mathcal{Z}_{\tau_1}^J = \mathbf{E}^{\mathbf{P}}\left(\mathcal{Z}_\infty^J \middle| \mathcal{F}_{\tau_1}\right) = \left(\mathcal{L}_{\tau_1}^J\right)^{-1}$ for $\tau_1 > 0$. It follows that $\mathcal{L}^J_{\tau_1}$ is given by

$$\mathcal{L}^J_{\tau_1} = \exp\left(\int_0^{\tau_1}\left(-\Lambda(r_s)+\psi_{\pi_s}\right)ds\right)\prod_{i=1}^{\pi_{\tau_1}}\frac{\Lambda\left(r_{S_{i-}}\right)}{\psi_{i-1}} \tag{25}$$

so that we can compute $\mathbf{P}\left(\overline{N}_T \geq 1\right) = \mathbf{E}^{\mathbf{Q}_T^{\xi=1}}\left(\mathbf{E}^{\mathbf{Q}^J}\left(\underbrace{\mathcal{L}_T^{\xi=1}\,\mathcal{L}^J_{\tau_1}}_{:=\mathcal{L}^\star_T}\right)\right)$.

We introduce the *interjump* short-rate process ρ^j satisfying

$$r_t = \sum_{j\geq 1}\rho^j_{t-S_{j-1}}\mathbf{1}_{\{\pi_t=j-1\}}, \tag{26}$$

where $\rho^j_0 = r_{S_{j-1}}$ for $j \geq 1$. As the process ρ^j follows the Feller diffusion without jump arrivals, the exact **JAM** scheme is available so that one can avoid sampling the time-integrated transform of the jump intensity $\Lambda(\rho^j) = \lambda_0 + \lambda_1\rho^j$ from its Markov property along with the existence of the closed-form expression of the bridge transform $\varphi\left(a, s; \rho^j_s, \rho^j_0\right) = \mathbf{E}^{\mathbf{P}}\left(e^{-a\int_0^s \rho^j_t dt}\middle|\, \rho^j_s,\ \rho^j_0\right)$ for $a > 0$ and $s > 0$. Algorithm 3 summarizes the cIS algorithm to estimate a zero-coupon bond price with maturity $T > 0$.

Beliaeva & Nawalkha (2012) propose a multinomial tree model to obtain an approximated bond price. Their proposed methodology illustrates how to superimpose mixed jump-diffusion trees by recombining multinomial jump trees on the diffusion tree for the stochastic short-rate model extended with various types of jumps. Although the multinomial tree does not generate simulation errors, its discretized nature along with linear approximation should produce non-negligible biases; see Beliaeva & Nawalkha (2012) for details.

We adopt the *trapezoidal discretization* scheme as the benchmark pMC method. Specifically, we divide the time inverval $[0, T]$ into $\left\lceil\sqrt{M}\right\rceil$ equal time steps of length. [13] Assuming that at most one jump can occur at each discretized time point, we sequentially generate the value of $r_j \triangleq r_{t_j}$ conditional on r_{j-1} as $r_j \approx r_{j-} + \delta_j\Delta\pi_j$, where r_{j-} can be drawn from the noncentral chi-squared transition density from r_{j-1} given by equation (16). Furthermore, we approximate the probability of jump arrival at time t_j as

[13] Specifically, we set the number of discretization time steps equal to the square-root of the number of simulation trials, as suggested by Duffie & Glynn (1995).

Algorithm 3 The cIS algorithm to estimate a zero-coupon bond price with $T > 0$

procedure CIS_BONDPRICING(T, M) ▷ M is the number of cIS trials
  Initialize $\hat{Y} \leftarrow 0$
  for $m \in \{1, \dots, M\}$ **do**
    Set $\mathcal{L}_T^\star \leftarrow T,\ r_t \leftarrow r_0$ and draw τ from $U(0, T)$
    while $\tau > 0$ **do**
      Draw Δ from the exponential distribution with rate $\Lambda(r_t) = \lambda_0 + \lambda_1 r_t$
      if $\Delta \geq \tau$ **then**
        Sample $r_{t+\tau}$ given r_t from the transition density of r without jump arrivals
        Compute $A = e^{-\lambda_0 \tau} \varphi\left(1 + \lambda_1,\ \tau;\ r_t,\ r_{t+\tau}\right)$
        Update $\mathcal{L}_T^\star \leftarrow \mathcal{L}_T^\star \times r_{t+\tau} \times A e^{\Lambda(r_t)\tau}$ and $\tau \leftarrow 0$
      else
        Sample $r_{t+\Delta}$ given r_t from the transition density of r without jump arrivals
        Compute $A = e^{-\lambda_0 \Delta} \varphi\left(1 + \lambda_1,\ \Delta;\ r_t,\ r_{t+\Delta}\right)$
        Set $\mathcal{L}_T^\star \leftarrow \mathcal{L}_T^\star \times A e^{\Lambda(r_t)\Delta} \Lambda(r_{t+\Delta}) / \Lambda(r_t)$
        Draw $\eta_{t+\Delta}$ from $\mathcal{N}\left((\theta - r_{t+\Delta})u, v^2\right)$ and set $\delta_{t+\Delta} \leftarrow r_{t+\Delta}\left(\exp(\eta_{t+\Delta}) - 1\right)$
        Update $r_t \leftarrow r_{t+\Delta} + \delta_{t+\Delta}$ and $\tau \leftarrow \tau - \Delta$
      end if
    end while
    Update $\hat{Y} \leftarrow \hat{Y} + \mathcal{L}_T^\star$
  end for
  return 1 - $\hat{Y}/M$
end procedure

$\mathbf{P}\left(\Delta\pi_j = 1\right) \approx \left(\lambda_0 + \lambda_1 r_{j-}\right) h$ for small $h > 0$. We obtain the pMC estimator of zero-coupon bond price by the trapezoidal rule as $\frac{1}{M}\sum_{i=1}^{M} \exp\left(-h \sum_{j=1}^{\lceil \sqrt{M} \rceil} \frac{r_{j-} + r_{j-1}}{2}\right)$. The bias of the the pMC estimator with a given number of time discretization steps is estimated using 10^9 trials to estimate the expectation of the estimator, and then taking the difference with the true value, which is estimated by cIS algorithm with 10^{10} iterations.

Table 3 and Figure 6 report the numerical results for the cIS, multinomial tree, and trapezoidal discretization schemes on the zero-coupon bond pricing. The RMSE convergence rate of our proposed cIS scheme is substantially faster than those of the multinomial tree and the trapezoidal discretization schemes.

Table 3: Estimation results under the short-rate model in (22) for a bond price with maturity $T = 3$ years.

Method	Trials (K)	Steps	Bias	SE	RMSE	Time (sec)
Conditional Importance Sampling	50	N/A	0	1.5477E-04	1.5477E-04	0.2164
	500			4.9092E-05	4.9092E-05	1.4017
	5,000			1.5513E-05	1.5513E-05	14.1515
	50,000			4.8123E-06	4.8123E-06	142.4679
Multinomial Tree	N/A	1,000	-3.7366E-04	0	3.7366E-04	0.2403
		2,500	-1.5910E-04		1.5910E-04	1.5892
		7,000	-8.7096E-05		8.7096E-05	12.0744
		20,000	-5.3870E-05		5.3870E-05	115.0961
Trapezoidal Discretization	10	100	5.2978E-05	2.6705E-04	2.7225E-04	0.2594
	50	225	2.1852E-05	1.1908E-04	1.2107E-04	0.9502
	250	500	8.5263E-06	5.3402E-05	5.4078E-05	10.6361
	1,250	1,120	5.3556E-06	2.3793E-05	2.4389E-05	119.5019

Note. The jump sensitivities are specified as $\delta_t = r_t \left(\exp\left(\eta_t\right) - 1\right)$, where $\eta_t \sim \text{Normal}\left((\theta - r_t)u, v^2\right)$. The selected parameters are $(\kappa, \theta, \sigma, \lambda_0, \lambda_1, u, v, r_0) = (0.1, 0.08, 0.05, 1.0, 5.0, 1.0, 1.0, 0.05)$. The true price of the bond as estimated by cIS algorithm with 10^{10} iterations is 0.8447251.

6 Conclusion

This paper proposes a simple but efficient importance sampling method for a wide range of counting processes. The proposed algorithm facilitates the conditional Monte Carlo simulation based on the limit of conditional probability measures specific to the event of interest. Numerical results illustrate the performance of the methodology that generates unbiased simulation estimators of rare-event probabilities of clustered defaults in a network and exact fixed-income security prices under the jump-diffusion interest rate models.

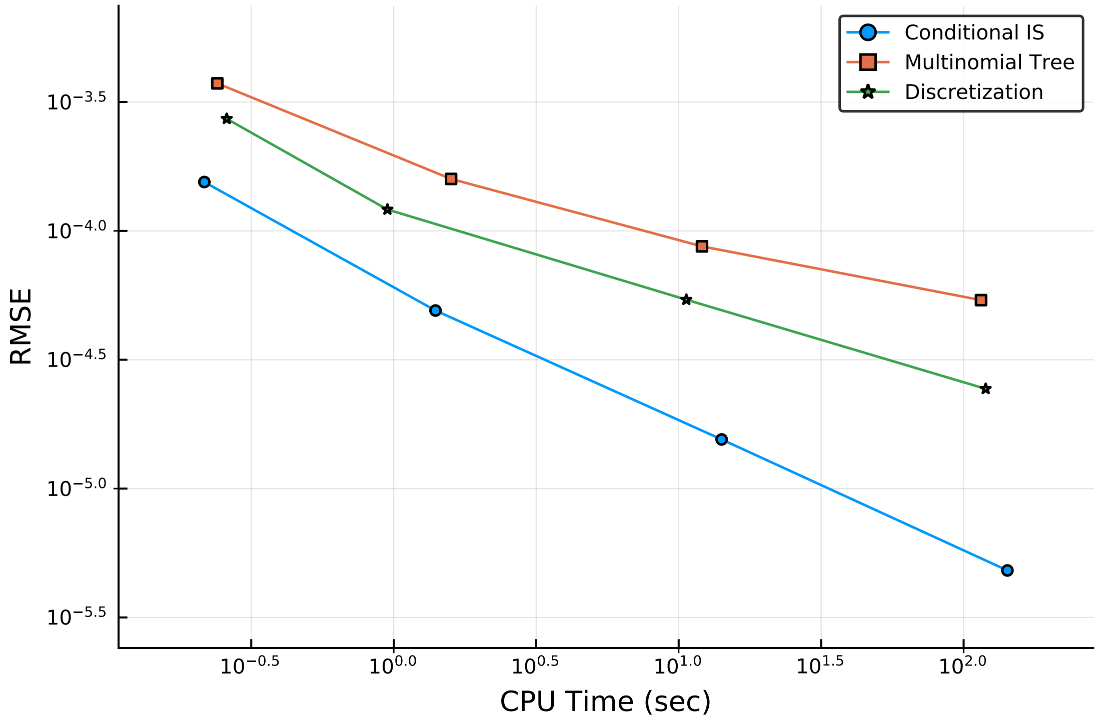
References

Asmussen, S. & P Glynn (2007), *Stochastic Simulation – Algorithms and Analysis*, Springer, New York.

Azizpour, Shahriar, Kay Giesecke & Gustavo Schwenkler (2014), Exploring the sources of default correlation. Working Paper, Stanford University.

Beliaeva, Natalia & Sanjay Nawalkha (2012), 'Pricing american interest rate options un-

Figure 6: Convergence of the RMSEs for a bond price under the short-rate model in (22) with maturity $T = 3$ years.



Note. The jump sensitivities are specified as $\delta_t = r_t \left(\exp\left(\eta_t\right) - 1\right)$, where $\eta_t \sim \text{Normal}\left((\theta - r_t)u, v^2\right)$. The selected parameters are $(\kappa, \theta, \sigma, \lambda_0, \lambda_1, u, v, r_0) = (0.1, 0.08, 0.05, 1.0, 5.0, 1.0, 1.0, 0.05)$.

der the jump-extended constant-elasticity-of-variance short rate models', *Journal of Banking and Finance* **36**, 151–163.

Bertsimas, Dimitris, Geoffrey J. Lauprete & Alexander Samarov (2004), 'Shortfall as a risk measure: properties, optimization and applications', *Journal of Economic Dynamics and Control* **28**, 1353–1381.

Blanchet, J. & J. Ruf (2016), 'A weak convergence criterion for constructing changes of measure', *Stochastic Models* **32**(2), 233–252.

Broadie, M. & O. Kaya (2006), 'Exact simulation of stochastic volatility and other affine jump diffusion processes', *Operations Research* **54**(2), 217–231.

Carmona, René, Jean-Pierre Fouque & Douglas Vestal (2009), 'Interacting particle systems for the computation of rare credit portfolio losses', *Finance and Stochastics* **13**(4), 613–633.

Carmona, René & Stéphane Crépey (2010), 'Particle methods for the estimation of Markovian credit portfolio loss distributions', *International Journal of Theoretical and Applied Finance* **13**(4), 577–602.

Cox, J., J. Ingersoll & S. Ross (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**, 385–408.

Del Moral, Pierre & Josselin Garnier (2005), 'Genealogical particle analysis of rare events', *Annals of Applied Probability* **15**, 2496–2534.

Duffie, Darrell, Andreas Eckner, Guillaume Horel & Leandro Saita (2009), 'Frailty correlated default', *Journal of Finance* **64**, 2089–2123.

Duffie, Darrell & Peter Glynn (1995), 'Efficient monte carlo estimation of security prices", *Annals of Applied Probability* **4**(5), 897–905.

Giesecke, Kay & Alexander Shkolnik (2010), Importance sampling for indicator Markov chains, *in* B.Johansson, S.Jain, J.Montoya-Torres, J.Hugan & E.Yücesan, eds, 'Proceedings of the 2010 Winter Simulation Conference', IEEE Press, pp. 967–975.

Giesecke, Kay & Alexander Shkolnik (2014), Optimal importance sampling of default losses. Working Paper, Stanford University.

Giesecke, Kay & Alexander Shkolnik (2018), Reducing bias in event time simulation via measure changes. Mathematics of Operations Research, Forthcoming.

Giesecke, Kay & Baeho Kim (2011), 'Systemic risk: What defaults are telling us', *Management Science* **57**(8), 1387–1405.

Giesecke, Kay, Baeho Kim & Shilin Zhu (2011), 'Monte carlo algorithms for default timing problems', *Management Science* **57**(12), 2115–2129.

Giesecke, Kay & Dmitry Smelov (2013), 'Exact sampling of jump diffusions', *Operations Research* **61**(4), 894–907.

Johannes, Michael (2004), 'The statistical and economic role of jumps in continuous-time interest rate models', *Journal of Finance* **59**(1), 227–260.

Meyer, P.-A. (1971), 'Démonstration simplifée d'un théorème de knight', *Séminaire de Probabilités V. Lecture Notes in Mathematics 191* pp. 191–195.

Protter, Philip (2005), *Stochastic Integration and Differential Equations*, Springer-Verlag, New York.

Siegmund, D (1976), 'Importance sampling in the Monte Carlo study of sequential tests', *The Annals of Statistics* **4**(4), 673–684.

Vanden-Eijnden, E. & J. Weare (2012), 'Rare event simulation of small noise diffusions', *Communications on Pure and Applied Mathematics* **65**(12), 1770–1803.

A Proofs

A.1 Proof of Theorem 3.1

Proof. Let $u^\gamma = \left(\gamma \overline{p}\right)^{-1}$ be a nonnegative càdlàg process on $[0, \tau_n)$ for $\gamma \in \mathbb{N}$.[14] Define

$$\mathcal{Z}^\gamma_\infty = \exp\left(\sum_{i=1}^n \int_0^{\tau_n} p^i_s \left(1 - u^\gamma_s\right) ds + N^i_{\tau_i} \log\left(u^\gamma_{\tau_i -}\right)\right) \tag{27}$$

$$= \exp\left(\int_0^{\tau_n} \overline{p}_s ds - \frac{\tau_n}{\gamma}\right) \prod_{i=1}^n \left(\gamma \overline{p}_{\tau_i -}\right)^{-1}. \tag{28}$$

Our assumption $\tau_n < \infty$ $\mathbf{P}$-almost surely and that $\gamma < \infty$ implies that, for all $T > 0$, $\mathcal{Z}^\gamma_T = \mathbf{E}^\mathbf{P}\left(\mathcal{Z}^\gamma_\infty \middle| \mathcal{F}_T\right)$ is a (uniformly integrable) Doob martingale with $\mathbf{E}^\mathbf{P}\left(\mathcal{Z}^\gamma_T\right) = 1$; see Theorem 3.1 in Giesecke & Shkolnik (2018). Furthermore, $\mathbf{Q}^\gamma_T = Z^\gamma_T \mathbf{P}$ is an absolutely continuous probability measure in the sense that $\mathbf{Q}^\gamma_T\left(\mathscr{A}\right) = \mathbf{E}^\mathbf{P}\left(\mathcal{Z}^\gamma_T \mathbf{1}_\mathscr{A}\right)$ holds for all $\mathscr{A} \in \mathcal{F}_T$. Since $\mathcal{E}^\xi_T \in \mathcal{G}_\xi$, we have

$$\mathbf{E}^\mathbf{P}\left(\mathcal{Z}^\gamma_{\tau_\xi} \mathbf{1}_{\mathcal{E}^\xi_T \cap \mathscr{A}}\right) = \mathbf{E}^\mathbf{P}\left(\mathbf{E}^\mathbf{P}\left(\mathcal{Z}^\gamma_{\tau_\xi} \mathbf{1}_{\mathcal{E}^\xi_T \cap \mathscr{A}} \middle| \mathcal{G}_\xi\right)\right) = \mathbf{Q}^\gamma_T\left(\mathcal{E}^\xi_T \bigcap \mathscr{A}\right) \tag{29}$$

for all $\mathscr{A} \in \mathcal{G}_\xi$. Note that $\overline{N}$ is a $\mathbf{Q}^\gamma_T$-Poisson process with rate γ^{-1} on $[0, T \wedge \tau_n)$. This implies that $\mathbf{Q}^\gamma_T\left(\mathcal{E}^\xi_T\right) = \frac{e^{-T/\gamma}}{\gamma^\xi}\left(\frac{T^\xi}{\xi!} + \alpha(\gamma)\right)$ holds, where $\alpha(\gamma) \downarrow 0$ as $\gamma \uparrow \infty$. It follows

[14] We follow the convention $1/0 = \infty$.

that $\mathcal{Z}^{\gamma}_{\tau_\xi}\mathbf{1}_{\mathcal{E}^\xi_T\bigcap\mathscr{A}} = \frac{e^{-\tau_\xi/\gamma}}{\mathcal{L}^\xi_T\gamma^\xi}$ holds and $\gamma^\xi\mathbf{Q}^\gamma_T\left(\mathcal{E}^\xi_T\right)$ can be expressed as

$$\mathbf{E}^{\mathbf{P}}\left(\frac{\mathbf{1}_{\mathcal{E}^\xi_T}}{\mathcal{L}^\xi_T}e^{-\tau_\xi/\gamma}\frac{T^\xi}{\xi!}\right) = e^{-T/\gamma}\left(\frac{T^\xi}{\xi!}+\alpha(\gamma)\right) \tag{30}$$

by taking $\mathscr{A} = \Omega$ in equation (29). Since $\frac{\mathbf{1}_{\mathcal{E}^\xi_T}}{\mathcal{L}^\xi_T}e^{-\tau_\xi/\gamma}\frac{T^\xi}{\xi!}$ is increasing as $\gamma\uparrow\infty$ and non-negative, by monotone convergence we have

$$\mathbf{Q}^\xi_T\left(\Omega\right) = \mathbf{E}^{\mathbf{P}}\left(\frac{\mathbf{1}_{\mathcal{E}^\xi_T}}{\mathcal{L}^\xi_T}\right) = 1\,, \tag{31}$$

which proves that $\mathbf{Q}^\xi_T \ll \mathbf{P}$ is a well-defined probability measure on $\left(\Omega,\mathcal{G}_\xi\right)$.

We subsequently have

$$\mathbf{P}\left(\mathcal{E}^\xi_T\right) = \mathbf{E}^{\mathbf{P}}\left(\mathbf{1}_{\mathcal{E}^\xi_T}\right) = \mathbf{E}^{\mathbf{P}}\left(\mathbf{1}_{\mathcal{E}^\xi_T}\frac{\mathcal{L}^\xi_T}{\mathcal{L}^\xi_T}\right) = \mathbf{E}^{\mathbf{Q}^\xi_T}\left(\mathcal{L}^\xi_T\right), \tag{32}$$

as $\mathcal{L}^\xi_T > 0$ holds under $\mathbf{P}$ almost surely by our assumption.

(i) We construct a conditional measure $\mathbf{Q}^{\gamma|\xi}_T = \frac{\mathbf{1}_{\mathcal{E}^\xi_T}\mathcal{Z}^\gamma_T}{\mathbf{Q}^\gamma_T\left(\mathcal{E}^\xi_T\right)}\mathbf{P}$ on $\left(\Omega,\mathcal{G}_\xi\right)$ in the sense that $\mathbf{Q}^{\gamma|\xi}_T\left(\mathscr{A}\right) = \mathbf{Q}^\gamma_T\left(\mathscr{A}\,\middle|\,\mathcal{E}^\xi_T\right)$ for all $\mathscr{A}\in\mathcal{G}_\xi$. By Fatou's lemma with $\mathcal{Z}^\gamma_T \geq 0$, we have

$$\liminf_{\gamma\uparrow\infty}\mathbf{Q}^{\gamma|\xi}_T\left(\mathscr{A}\right) = \liminf_{\gamma\uparrow\infty}\frac{\mathbf{E}^{\mathbf{P}}\left(\mathbf{1}_{\mathscr{A}\bigcap\mathcal{E}^\xi_T}\mathcal{Z}^\gamma_T\right)}{\mathbf{Q}^\gamma_T\left(\mathcal{E}^\xi_T\right)} = \mathbf{E}^{\mathbf{P}}\left(\frac{\mathbf{1}_{\mathscr{A}\bigcap\mathcal{E}^\xi_T}}{\mathcal{L}^\xi_T}\right) = \mathbf{Q}^\xi_T\left(\mathscr{A}\right). \tag{33}$$

Therefore, we have by Portmanteau's theorem that $\mathbf{Q}^{\gamma|\xi}_T \Rightarrow \mathbf{Q}^\xi_T$ as $\gamma\uparrow\infty$. We also have

$$\mathbf{Q}^{\gamma|\xi}_T = \frac{\mathbf{1}_{\mathcal{E}^\xi_T}\mathbf{Q}^\gamma_T}{\mathbf{Q}^\gamma_T\left(\mathcal{E}^\xi_T\right)} = \frac{\mathbf{1}_{\mathcal{E}^\xi_T}\left(\mathcal{Z}^\gamma_{\tau_\xi}\big/\mathcal{Z}^1_{\tau_\xi}\right)\mathbf{Q}^1_T}{\mathbf{Q}^\gamma_T\left(\mathcal{E}^\xi_T\right)} = \frac{\mathbf{1}_{\mathcal{E}^\xi_T}e^{(1-\gamma^{-1})\tau_\xi}\mathbf{Q}^1_T}{e^{-T/\gamma}\left(\frac{T^\xi}{\xi!}+\alpha(\gamma)\right)}. \tag{34}$$

Now consider a $\mathbf{Q}^\xi_T$-continuity set $A_{(t_1,\dots,t_\xi)} = \left\{\tau_1\leq t_1, \tau_2\leq t_2, \dots, \tau_\xi\leq t_\xi\right\}$ for any $0 \leq t_1 \leq t_2 \leq \cdots \leq t_\xi \leq T$. As $\tau_\xi \leq T$ and $e^{\tau_\xi - T/\gamma} \leq e^{(1-\gamma^{-1})\tau_\xi} \leq e^{\tau_\xi}$ must hold on $\mathcal{E}_T^\xi \subseteq A_{(t_1,\ldots,t_\xi)}$, we have

$$\frac{\mathbf{E}^{\mathbf{Q}_T^1}\left(\mathbf{1}_{A_{(t_1,\ldots,t_\xi)}} e^{\tau_\xi}\right)}{\frac{T^\xi}{\xi!} + \alpha(\gamma)} \leq \mathbf{Q}_T^{\gamma|\xi}\left(A_{(t_1,\ldots,t_\xi)}\right) \leq \frac{\mathbf{E}^{\mathbf{Q}_T^1}\left(\mathbf{1}_{A_{(t_1,\ldots,t_\xi)}} e^{\tau_\xi}\right)}{e^{-T/\gamma}\left(\frac{T^\xi}{\xi!} + \alpha(\gamma)\right)}, \tag{35}$$

where Meyer (1971)'s time-change theorem implies that

$$\mathbf{E}^{\mathbf{Q}_T^1}\left(\mathbf{1}_{A_{(t_1,\ldots,t_\xi)}} e^{\tau_\xi}\right) = \int_0^{t_1}\int_{x_1}^{t_2}\cdots\int_{x_{\xi-1}}^{t_\xi} dx_\xi \cdots dx_2 dx_1 = \frac{T^\xi}{\xi!} F\left(t_1,\ldots,t_\xi\right), \tag{36}$$

where $F\left(t_1,\ldots,t_\xi\right)$ is the distribution of the $\{u_i\}_{i=1}^{\xi}$. It follows that

$$\frac{\frac{T^\xi}{\xi!} F\left(t_1,\ldots,t_\xi\right)}{\frac{T^\xi}{\xi!} + \alpha(\gamma)} \leq \mathbf{Q}_T^{\gamma|\xi}\left(A_{(t_1,\ldots,t_\xi)}\right) \leq \frac{\frac{T^\xi}{\xi!} F\left(t_1,\ldots,t_\xi\right)}{e^{-T/\gamma}\left(\frac{T^\xi}{\xi!} + \alpha(\gamma)\right)}. \tag{37}$$

Taking $\gamma \uparrow \infty$, we conclude that

$$\mathbf{Q}_T^{\gamma|\xi} \to F\left(t_1,\ldots,t_\xi\right), \tag{38}$$

which proves the uniform arrivals of the first ξ event times on $[0,T]$ under $\mathbf{Q}_T^\xi$.

(ii) Let $k \in \{1,\ldots,\xi\}$ and consider $I_k \in \{1,\ldots,n\}$. By the conditional change of measure formula, we have

$$\frac{\mathbf{E}^{\mathbf{Q}_T^\gamma}\left(\mathbf{1}_{\mathcal{E}_T^\xi}\middle|\mathcal{F}_{\tau_k-}\right)}{\mathbf{Q}_T^\gamma\left(\mathcal{E}_T^\xi\right)} \mathbf{Q}_T^{\gamma|\xi}\left(I_k = i\middle|\mathcal{F}_{\tau_k-}\right) = \frac{\mathbf{E}^{\mathbf{Q}_T^\gamma}\left(\mathbf{1}_{\mathcal{E}_T^\xi \bigcap \{I_k=i\}}\middle|\mathcal{F}_{\tau_k-}\right)}{\mathbf{Q}_T^\gamma\left(\mathcal{E}_T^\xi\right)}. \tag{39}$$

Since $\mathbf{1}_{\mathcal{E}_T^\xi} = 0$ if $\tau_k > T$ and τ_k is $\mathcal{F}_{\tau_k-}$-measurable, we obatin

$$\begin{aligned}\mathbf{Q}_T^{\gamma|\xi}\left(I_k = i\middle|\mathcal{F}_{\tau_k-}\right) &= \mathbf{1}_{\{\tau_k \leq T\}} \frac{\mathbf{Q}_T^\gamma\left(\mathcal{E}_T^\xi \bigcap \{I_k = i\}\middle|\mathcal{F}_{\tau_k-}\right)}{\mathbf{Q}_T^\gamma\left(\mathcal{E}_T^\xi\middle|\mathcal{F}_{\tau_k-}\right)} &(40)\\ &= \mathbf{1}_{\{\tau_k \leq T\}} \mathbf{Q}_T^\gamma\left(I_k = i\middle|\mathcal{F}_{\tau_k-}\right) &(41)\end{aligned}$$

$$= \mathbf{1}_{\{\tau_k \le T\}} \frac{p^i_{\tau_k -} u_{\tau_k -}}{\overline{p}_{\tau_k -} u_{\tau_k -}} = \mathbf{1}_{\{\tau_k \le T\}} \frac{p^i_{\tau_k -}}{\overline{p}_{\tau_k -}}. \tag{42}$$

Since $\mathbf{Q}^{\gamma|\xi}_T \left(\tau_k \le T \right) = 1$ holds for all $k \in \{1, \ldots, \xi\}$, we conclude that

$$\mathbf{E}^{\mathbf{Q}^{\gamma|\xi}_T} \left(\mathbf{1}_{\mathscr{A}} \mathbf{1}_{\{ I_k = i \}} \right) = \mathbf{E}^{\mathbf{Q}^{\gamma|\xi}_T} \left(\mathbf{1}_{\mathscr{A}} \frac{p^i_{\tau_k -}}{\overline{p}_{\tau_k -}} \right) \tag{43}$$

for all $\mathscr{A} \in \mathcal{F}_{\tau_k -}$. Taking $\gamma \uparrow \infty$ by Portmanteau's theorem we obtain

$$\mathbf{Q}^{\xi}_T \left(I_k = i \middle| \mathcal{F}_{\tau_k -} \right) = \frac{p^i_{\tau_k -}}{\overline{p}_{\tau_k -}}. \tag{44}$$

(iii) This is a direct consequence of the Girsanov-Meyer theorem (see Protter (2005, Theorem III.41).

□

A.2 Proof of Corollary 3.2

Proof. This follows directly by the arguments in the proof of Theorem 3.1 with some slight adjustments that lead to significant simplification. Again we take the sequence of conditional measures $\widehat{\mathbf{Q}}^{\gamma}_T(\cdot \,|\, \mathcal{E}^{\xi}_T)$ but for the modified set $\widehat{\mathcal{E}}^{\xi}_T = \left\{ \overline{N}_T = \xi \right\}$. Observe that on $\widehat{\mathcal{E}}^{\xi}_T$,

$$\mathcal{Z}^{\gamma}_T / \widehat{\mathbf{Q}}^{\gamma}_T (\widehat{\mathcal{E}}^{\xi}_T) = \frac{e^{T/\gamma}}{\widehat{\mathcal{L}}^{\xi}_T}. \tag{45}$$

We follow the same arguments that take $\gamma \uparrow \infty$ as in the proof of Theorem 3.1. Statement (i) is simple to prove due to the new definition of $\widehat{\mathcal{E}}^{\xi}_T$. Statements (ii)–(iii) are proved in the same way as in Theorem 3.1. □