Insurance: Mathematics and Economics 101 (2021) 107–124

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Hawkes processes in insurance: Risk model, application to empirical data and optimal investment

Anatoliy Swishchuk [a], Rudi Zagst [b], Gabriela Zeller [b,*]

[a] *Department of Mathematics and Statistics, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada*
[b] *Chair of Mathematical Finance, Technical University of Munich, Parkring 11, 85748 Garching-Hochbrück, Germany*

ARTICLE INFO

Article history:
Available online 23 January 2021

JEL classification:
G22

Keywords:
Hawkes process
General compound Hawkes process
Risk model
FCLT
Diffusion approximation
Optimal investment for insurers
Incomplete market

ABSTRACT

In this paper we study a risk model with claim arrivals based on general compound Hawkes processes and show that it is suitable to model empirical insurance data. We review a law of large numbers and functional central limit theorem for this model and derive a pure diffusion approximation which allows analytical calculation of finite-time and infinite-time ruin probabilities. We use the approximation to study the influence of replacing the classical Poisson arrival process by a general compound Hawkes process on optimal investment strategies for an insurer in an incomplete market by applying results from asset–liability management.

1. Introduction

In risk theory, a central question is how to model the random process describing claim occurrences. In recent years, several extensions to the classical Cramér–Lundberg risk model introduced in Lundberg (1903) have been studied. This is relevant as in reality insurance claim arrivals cannot generally be assumed to follow a memoryless homogeneous Poisson process (e.g. Seal, 1983). One approach is to use a time-dependent intensity which is influenced by external factors such as environmental shocks. Albrecher and Asmussen (2006) study aggregate claims distributions and ruin probabilities for a risk process with claims according to a shot noise Cox process (a superposition of a homogeneous Poisson process and a Cox process with a Poisson shot noise intensity process) and Dassios and Jang (2003) use a Cox process to model claim arrivals for the pricing of catastrophe reinsurance. However, in financial applications it has been observed that time-dependence of event arrival rates and particularly temporal clustering of events can often not be entirely explained by exogenous factors (e.g. news, environmental changes), but might be caused by the arrival process itself (e.g. market reflexivity, Filimonov and Sornette, 2012; Hardiman et al., 2013). Thus self-exciting Hawkes processes, first introduced in Hawkes (1971), have gained attention due to their ability to reflect endogenously caused clustering. Bacry et al. (2015) give a good overview of the variety of recent applications of Hawkes processes in finance, such as modelling market activity (Da Fonseca and Zaatour, 2013; Lallouache and Challet, 2016; Embrechts et al., 2011; Chavez-Demoulin and McGill, 2012) and price modelling in high-frequency trading (Bacry et al., 2013; Zheng et al., 2014; Fauth and Tudor, 2012).

The first work to consider a risk model with Hawkes claims arrivals was Stabile and Torrisi (2010) who derive the asymptotic behaviour of infinite and finite horizon ruin probabilities and asymptotically efficient simulation laws assuming light-tailed claims. Their work was extended by Zhu (2013) who considered (subexponential) heavy tailed claims. Dassios and Zhao (2012) consider a risk process with a dynamic contagion process, generalizing the Hawkes process and the Cox process with shot noise intensity and thus including both self-excited and externally excited jumps. Jang and Dassios (2013) study a bivariate shot noise self-exciting process for insurance, including a constant rate of exponential decay that could be interpreted as the time value of money. Cheng and Seol (2018) derive diffusion approximations and expressions for the ruin probabilities of a risk model with Hawkes claims arrivals, providing numerical examples for exponential and Gamma-distributed jumps. They find that the diffusion limit is a Gaussian process which can be decomposed into a centred Gaussian process and an independent Brownian motion. Swishchuk (2017a) proposes a risk model with general compound Hawkes process (claim arrivals as a Hawkes process and claim sizes as an N-state Markov Chain) and derives a law of large numbers (LLN) and functional central limit theorem (FCLT)

* Corresponding author.
E-mail addresses: aswish@ucalgary.ca (A. Swishchuk), zagst@tum.de (R. Zagst), gabi.zeller@tum.de (G. Zeller).

https://doi.org/10.1016/j.insmatheco.2020.12.005

for this model. The FCLT allows to construct a diffusion approximation which leads to closed-form formulas for finite and infinite horizon ruin probabilities. A model with a general compound Hawkes process has been successfully used for Limit Order Book Modelling in Swishchuk (2017b), but has some drawbacks when applied to the insurance context.

In this paper, we use an approach that allows us to employ the convenient diffusion approximation results from Swishchuk (2017a) while being able to reasonably reflect claim sizes (which are usually assumed to follow a continuous distribution). We show that the model is suitable for an empirical insurance data set. To the best of our knowledge, this is the first work considering a risk model with Hawkes claims arrivals employing real empirical insurance data.

After testing the goodness of model fit and the diffusion approximations, we apply the results to calculate ruin probabilities. We then extend the work of Xie et al. (2008), who study the mean–variance efficient frontier and investment strategy for an insurer whose random risk process follows a Brownian motion with drift in an incomplete market, to the case of a risk model with Hawkes processes. We show how the higher risk of the Hawkes process (measured by the variance of the number of arrivals) influences the strategy and the obtainable return for a given risk limit in comparison to a classical Poisson process. In particular, we show that the analysis of the presence of clustering in the claim arrival process and the dependence between the evolution of the insurer's liability and the evolution of the tradeable risky assets are vital in order to choose an investment strategy in a way that ensures the allowed risk limit (in the form of a variance boundary on the terminal wealth) is adhered to. In the case of clustered claims arrivals, this will imply choosing a more conservative strategy and reducing the attainable expected return as compared to the case of non-clustered claims arrivals, even if the expected claim number and size do not differ. The paper is structured as follows: Section 2 reviews some definitions and results about Hawkes processes used in the sequel. In particular, we highlight the recent work of Swishchuk (2017a) on insurance risk models with general compound Hawkes processes (RH)[1]. Section 3 describes the empirical insurance data set and specifies the modelling of claim arrivals and claim sizes. We show that the model accurately reflects the characteristics of empirical data. Section 4 uses the results from Swishchuk (2017a) on diffusion approximations for the RH to calculate finite and infinite time ruin probabilities. Section 5 addresses the investment problem of an insurer in an incomplete market whose claims arrive according to a Hawkes process. Using the mean–variance based approach by Xie et al. (2008), we show how replacing the classical case of Poisson arrivals by Hawkes arrivals influences the risk (measured by the variance of terminal wealth) and thus alters the investment decision. Economically, this reflects that the insurer's claim arrival process might display a clustering effect and long-term memory, thus at each point in time information from the whole past of the process has to be considered. Section 6 concludes the paper by highlighting limitations and future research opportunities.

2. Background

In this section, we review some well-known definitions and past work on Hawkes processes and their application for insurance risk models.

2.1. Hawkes processes

2.1.1. Hawkes process: Definition and notation

The Hawkes process introduced by Hawkes (1971) is a simple point process with self-exciting property, clustering effect and long-term memory. It can be used to model a sequence of arrivals into a system over time, such as earthquake occurrences (Ogata, 1999), trade orders (Da Fonseca and Zaatour, 2013), credit defaults (Errais et al., 2010) or incoming insurance claims (Stabile and Torrisi, 2010; Cheng and Seol, 2018). The counting process, usually denoted $N(t)$, refers to the cumulative number of arrivals up until a time $t \geq 0$, and can be characterized by the corresponding point process $\mathbf{T} := (t_1, t_2, \ldots)$, the sequence of random arrival times at which the counting process $N(t)$ has jumped. We denote by $\mathcal{H}(t), t \geq 0$, the history of arrivals up to time t, i.e. $\{\mathcal{H}(t), t \geq 0\}$ is a filtration. For an extensive treatment of point processes we refer to Daley and Vere-Jones (2003).

Consider a counting process $N(t)$ with history $\mathcal{H}(t)$ for $t \geq 0$. If a non-negative, $\mathcal{H}(t)$-measurable function $\lambda^*(t)$ exists such that

$$\lambda^*(t) = \lim_{h \to 0} \frac{\mathbb{E}[N(t+h) - N(t) | \mathcal{H}(t)]}{h}, \quad (1)$$

then it is called conditional intensity function of $N(t)$. By Daley and Vere-Jones (2003), if the conditional intensity function exists, it uniquely determines the finite-dimensional distributions of the point process and can thus be used as a characterization.

The non-decreasing function

$$\Lambda(t) = \int_0^t \lambda^*(s) ds \quad (2)$$

is called the compensator of the counting process.

Definition 1 (*One-dimensional Hawkes Process*). Consider $N(t), t \geq 0$, a counting process with history $\{\mathcal{H}(t), t \geq 0\}$ that satisfies

$$\mathbb{P}(N(t+h) - N(t) = m | \mathcal{H}(t)) = \begin{cases} \lambda^*(t)h + o(h), & m = 1 \\ o(h) & m > 1 \\ 1 - \lambda^*(t)h + o(h), & m = 0 \end{cases} \quad (3)$$

Suppose the conditional intensity function is of the form

$$\lambda^*(t) = \lambda + \int_0^t \mu(t-s) dN(s) \quad (4)$$

where $\lambda > 0$ is called background intensity and $\mu : [0, \infty) \to (0, \infty)$ is called excitation function. Assume that $\mu(\cdot) \neq 0$ to avoid the trivial case of a homogeneous Poisson process. The process $N(\cdot)$ is called a Hawkes process.

Note that using the observed sequence of arrival times $(t_1, t_2, \ldots, t_k)$ up to time t, the conditional intensity can be written as

$$\lambda^*(t) = \lambda + \sum_{t_i < t} \mu(t - t_i) \quad (5)$$

Thus, the Hawkes process is a generally non-Markovian extension of the Poisson process. Note that the self-excitement of

[1] In Swishchuk (2017a) and Swishchuk (2017b), this model respectively the underlying process is abbreviated as *(RM)GCHP*. In particular, this highlights that it is only one rather general of numerous cases that are investigated in those works. However, as it is the only model employed in this present work, we shorten the abbreviation to *RH*, indicating a *Hawkes* risk model.

the process is reflected in the fact that a new arrival causes an increase in the intensity function and thus temporal clustering of **T**. Depending on the choice of $\lambda^*(t)$, this might even lead to an explosion of the process (infinitely many jumps would occur in a finite time interval), an event that should be avoided. In the one-dimensional case this is achieved by restricting our focus to so-called stationary Hawkes processes that fulfil the condition

$$\hat{\mu} := \int_0^\infty \mu(s)ds < 1. \tag{6}$$

2.1.2. Exponentially decaying Hawkes process: Definition and properties

A common choice for the excitation function $\mu(\cdot)$ is the one of exponential decay: $\mu(t) = \alpha e^{-\beta t}$ with parameters $\alpha, \beta > 0$. The conditional intensity function (4) of a Hawkes process with exponentially decaying intensity thus becomes

$$\lambda^*(t) = \lambda + \int_0^t \alpha e^{-\beta(t-s)}dN(s) = \lambda + \alpha \sum_{t_i<t} e^{-\beta(t-t_i)}. \tag{7}$$

In this case, the stationarity condition (6) corresponds to

$$\hat{\mu} = \int_0^\infty \alpha e^{-\beta s}ds = \frac{\alpha}{\beta} < 1 \iff \alpha < \beta. \tag{8}$$

Given an initial condition $\lambda^*(0) = \lambda_0$, (7) satisfies the SDE

$$d\lambda^*(t) = \beta(\lambda - \lambda^*(t))dt + \alpha dN(t), \qquad t \geq 0 \tag{9}$$

which can be solved as (Laub et al., 2015)

$$\lambda^*(t) = e^{-\beta t}(\lambda_0 - \lambda) + \lambda + \int_0^t \alpha e^{-\beta(t-s)}dN(s), \tag{10}$$

which is an extension of (7), e.g. for a process that started some time before the beginning of the observation period. Note that in the case of exponentially decaying intensity (7), the process $(\lambda^*(t), N(t))$ is a continuous-time Markov process (see Gao and Zhu, 2017) which is not the case for a general choice of excitation function in (4). In Da Fonseca and Zaatour (2013), this is used as the key property which allows to study the distributional properties of the process and compute explicitly the moments and autocorrelation function of the number of jumps of an exponential Hawkes process over a fixed interval. We briefly summarize their results in the following.

Proposition 1 (*Da Fonseca and Zaatour, 2013*)**.** *Given a Hawkes process $X(t) = (\lambda^*(t), N(t))$ with dynamic given by (9), the long-run expected value of the number of jumps during an interval of length τ is given by*

$$\mathbb{E}[N(\tau)] := \lim_{t\to\infty} E(t,\tau) := \lim_{t\to\infty} \mathbb{E}[N(t+\tau) - N(t)] = \frac{\lambda}{1-\alpha/\beta}\tau. \tag{11}$$

The variance is given by

$$\mathbb{V}ar[N(\tau)] := \lim_{t\to\infty} V(t,\tau) := \lim_{t\to\infty}\left(\mathbb{E}[(N(t+\tau)-N(t))^2] - E(t,\tau)^2\right) = \frac{\lambda}{1-\alpha/\beta}\left(\tau\left(\frac{1}{1-\alpha/\beta}\right)^2 + \left(1 - \left(\frac{1}{1-\alpha/\beta}\right)^2\right)\frac{1-e^{-\tau(\beta-\alpha)}}{\beta-\alpha}\right). \tag{12}$$

The covariance of the number of arrivals for two non-overlapping intervals of length τ with lag $\delta > 0$ is given by

$$\begin{aligned}\mathbb{C}ov(N(\tau),\delta) &:= \lim_{t\to\infty} C(t,\tau,\delta)\\ &:= \lim_{t\to\infty}\big(\mathbb{E}[(N(t+\tau)-N(t))(N(t+2\tau+\delta)-\\ &\quad - N(t+\tau+\delta))] - E(t,\tau)E(t+\tau+\delta,\tau)\big)\\ &= \frac{\lambda\beta\alpha(2\beta-\alpha)(e^{(\alpha-\beta)\tau}-1)^2}{2(\alpha-\beta)^4}e^{(\alpha-\beta)\delta}.\end{aligned} \tag{13}$$

Note that taking the limit for $t \to \infty$ (putting the process into its long-run stationary regime) to simplify dependence with respect to the initial value λ_0 requires again the stability of the process, implying $\alpha < \beta$.

Proposition 2 (*Da Fonseca and Zaatour, 2013*)**.** *A direct consequence from the last result is the autocorrelation function of the number of jumps over intervals of length τ separated by a time lag of δ:*

$$Acf(\tau,\delta) = \frac{e^{-2\beta\tau}(e^{\alpha\tau}-e^{\beta\tau})^2\alpha(\alpha-2\beta)}{2(\alpha(\alpha-2\beta)(e^{(\alpha-\beta)\tau}-1)+\beta^2\tau(\alpha-\beta))}e^{(\alpha-\beta)\delta}. \tag{14}$$

Note that this expression is always positive for $\alpha < \beta$ (stationarity condition) and exponentially decaying with the lag δ.

2.1.3. Simulation, parameter inference and goodness of fit testing

In this section, we briefly explain the methods we use for Hawkes process simulation, parameter inference and goodness of fit testing of a Hawkes model. All these techniques are well-known and a good summary can be found in e.g. Laub et al. (2015). For Hawkes process simulation we rely on the modified thinning algorithm first introduced in Ogata (1981) which is an adaption of the thinning algorithm used to simulate an inhomogeneous Poisson process by Lewis and Shedler (1979).

In order to fit an exponentially decaying Hawkes process to our empirical dataset, we use maximum likelihood parameter inference. The log-likelihood function for a realization of $(N(t))$, $t \geq 0$ over $[0, T]$ as derived in Laub et al. (2015) in this case is

$$l = \sum_{i=1}^{N(T)} log(\lambda + \alpha A(i)) - \lambda T + \frac{\alpha}{\beta}\sum_{i=1}^{N(T)}\left(e^{-\beta(T-t_i)} - 1\right) \tag{15}$$

where

$$A(i) = \begin{cases} 0 & i = 1,\\ \sum_{j=1}^{i-1} e^{-\beta(t_i-t_j)} = e^{-\beta(t_i-t_{i-1})}(1 + A(i-1)) & i \in \{2,\ldots,k\}.\end{cases} \tag{16}$$

Some numerical disadvantages of using the maximum likelihood estimation method, such as bias for small sample sizes, getting stuck in local optima and performance issues for large samples (especially for high-frequency trading applications) have been addressed by e.g. Da Fonseca and Zaatour (2013), Filimonov and Sornette (2015). This motivated the derivation of a generalized method of moments for parameter estimation in Da Fonseca and Zaatour (2013) based on their results stated in Propositions 1 and 2. However, for our purposes, we found the maximum likelihood method to be instantaneous and reliable. There are many methods of checking the goodness of fit of a fitted Hawkes model to point data, an overview can be found in Laub et al. (2015). We will use the basic test relying on the random time change theorem here.

Theorem 1 (*Random Time Change Theorem Brown et al., 2002*). *Let $\{t_1, t_2, \dots, t_k\}$ be a realization over time $[0, T]$ from a point process with conditional intensity function $\lambda^*(\cdot)$. If $\lambda^*(\cdot)$ is positive over $[0, T]$ and $\Lambda(T) < \infty$ a.s. then the transformed points $\{t_1^*, \dots, t_k^*\} = \{\Lambda(t_1), \dots, \Lambda(t_k)\}$ form a Poisson process with unit rate. $\Lambda(\cdot)$ denotes the compensator of the point process.*

As the closed form of the compensator for an exponential Hawkes process is known from (2), one can test the quality of the parameter estimation by transforming the original timepoints and performing standard fitness tests for a unit rate Poisson process on the transformed datapoints. Many generalizations of Hawkes processes have been studied and used in financial applications, e.g. multi-dimensional (self- and mutually-exciting) Hawkes processes (Embrechts et al., 2011; Aït-Sahalia et al., 2015) or marked Hawkes processes (Bacry et al., 2013; Fauth and Tudor, 2012; Karabash and Zhu, 2015). In this work, we restrict ourselves to one-dimensional Hawkes processes with exponential decay. This seems reasonable for the dataset at hand and considering that this is the first application of a risk model with Hawkes processes to real insurance data. Promising ideas for further studies with more general Hawkes processes are summarized in Section 6.

2.2. Risk model with Hawkes processes

In general, a risk model intends to describe the available capital of an insurance company (or part of it) over time and is of the form

$$R(t) = u + ct - \sum_{i=1}^{N_t} Y_i, \tag{17}$$

where u denotes the initial capital and c denotes the (continuous) premium rate. $N(t)$ is a counting process describing the number of claims occurring in the interval $(0, t]$ and $\{Y_i\}$ is a sequence of non-negative random variables describing the claim sizes. Usually, the $\{Y_i\}$ are assumed i.i.d. with distribution function G and finite first two moments $\mathbb{E}[Y_1] = m_1$ and $\mathbb{E}[Y_1^2] = m_2$. $N(t)$ and $\{Y_i\}$ are assumed to be independent. In the classical case $N(t)$ is a homogeneous Poisson process, but from now on we assume $N(t)$ to be a stationary Hawkes process with exponentially decaying intensity.

An important event to study in risk theory is the occurrence of ruin, i.e. the event that the capital falls below 0 for the first time (or equivalently another fixed lower bound, e.g. given by regulatory requirements).

The ruin time given the initial capital u is thus defined as

$$\tau(u) = \inf\{t > 0 : R(t) < 0 \mid R(0) = u\} \quad \text{where} \quad \inf\emptyset = \infty. \tag{18}$$

The probability that ruin occurs until a fixed finite time-horizon t is then

$$\begin{aligned}\Psi(u, t) &= \mathbb{P}(\tau(u) \le t \mid R(0) = u)\\ &= \mathbb{P}(\inf_{0<s\le t} R(s) < 0 \mid R(0) = u),\end{aligned} \tag{19}$$

and the infinite-horizon ruin probability is accordingly

$$\Psi(u) = \lim_{t\to\infty} \Psi(u, t) = \mathbb{P}(\inf_{t>0} R(t) < 0 \mid R(0) = u). \tag{20}$$

Even for the classical model, closed-form solutions for these ruin probabilities are only available for special choices of the claim size distribution (see e.g. Asmussen and Albrecher, 2010). This has motivated seeking bounds for the ruin probability, such as the Lundberg inequality for the classical case (see Asmussen and Albrecher, 2010) and an adapted version for the Cox case by Embrechts et al. (1993). Another approach is to approximate the risk process by a diffusion approximation, which for the classical model was first done by Iglehart (1969) and Grandell (1977). Schmidli (1994) considered the case where borrowing money and investing surpluses is allowed and the recent work of Basu (2016) studies a renewal-process based risk-reserve process with dividend payments. As mentioned, many authors have recently suggested risk models with Hawkes claims arrivals. For our application, we will focus on the work of Swishchuk (2017a) who introduces the following risk model.

Definition 2 (*Risk Model with General Compound Hawkes Process (RH) Swishchuk, 2017a*). Let $N(t)$ be any one-dimensional Hawkes process as defined above. Let (X_i) be an ergodic continuous-time finite (or countably infinite) Markov Chain, independent of $N(t)$, with state space X, and let $a(x)$ be any bounded function on X. Then a general compound Hawkes process is defined as

$$H(t) = H(0) + \sum_{i=1}^{N(t)} a(X_i). \tag{21}$$

Define the risk process $R(t)$ based on a general compound Hawkes process as

$$R(t) = u + ct - \sum_{i=1}^{N(t)} a(X_i), \tag{22}$$

where u is the initial capital, c is the premium rate, (X_i) is a continuous-time Markov Chain on the state space $X = \{1, \dots, n\}$, $N(t)$ is a Hawkes process, $a(x)$ is a bounded function on X. $N(t)$ and (X_i) are independent.

A related model has been successfully used for modelling Limit Order Book dynamics in Swishchuk (2017b) where the $a(X_i)$ are interpreted as a conditionally dependent sequence of price changes which are not fixed to one-tick movements. Likewise, in many cases in the insurance context, the usual assumption of i.i.d. claim sizes might be too restrictive such that the above model offers a possible generalization. In our particular context, we will adapt the interpretation of the Markov Chain as described in Section 3.3 in order to appropriately reflect incoming claim sizes. Swishchuk (2017a) proves a law of large numbers for this model and uses it to derive a net profit condition and the premium principle based on the expected value principle.

Theorem 2 (*Law of Large Numbers for RH Swishchuk, 2017a*). *Let $R(t)$ be the risk model defined in Definition 2, and let (X_i) be a Markov Chain with state space X and stationary probabilities π_i^*. We suppose that $0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1$. Then*

$$\lim_{t\to\infty} \frac{R(t)}{t} = c - a^* \frac{\lambda}{1-\hat{\mu}}, \tag{23}$$

where $a^* = \sum_{i\in X} a(i)\pi_i^*$.

Corollary 1 (*Net Profit Condition and Premium Principle Swishchuk, 2017a*). *The net profit condition for RH is given as*

$$c > a^* \frac{\lambda}{1-\hat{\mu}}. \tag{24}$$

The premium principle for RH, based on the expected value principle, is given as

$$c = (1+\theta)a^* \frac{\lambda}{1-\hat{\mu}}, \tag{25}$$

where θ denotes the safety loading.

3. Implementation of risk model results with empirical data

In this section, we first explain in detail the empirical insurance data set used in the sequel. We show that the use of a Poisson process for claim arrivals would not be suitable and fit an exponential Hawkes process to the claim arrival process. We model claim sizes according to an adapted version of the *RH* and then show that this model is indeed suitable to reflect the data set under consideration.

3.1. Empirical data

The data set was provided by a large German Insurance group and comprises claim occurrences from the class of legal expenses insurance, which refers to insurance protection covering the costs of a legal dispute (e.g. lawyer expenses or fees). For this class of insurance, we observe that often once a legal dispute occurs and is reported to the insurance company, multiple payments from (or triggered by) this case have to be expected in the subsequent time period. This might be due to multiple receivables from lawyers and consultants, an appeal of the court case being lodged or more legal matters being uncovered and reported resulting from the initial reporting. Particularly, it might be expected that clients who once start a legal dispute usually show a higher willingness to continue to pursue it or start another one. Therefore we suspect a Hawkes process might be suitable to model claim arrivals for this class of insurance claims.

The dataset used in the following comes from the subclass of *legal insurance against damage compensation in consequence of incidents related to traffic*. We have conducted the same procedure for other datasets from different subclasses and an outlook is given in Section 6.

3.2. Claim arrivals

First, we need to clarify the kind of arrival times we consider. The empirical data set is divided according to the *reporting year* of a claim. This is not necessarily equal to the year of the original *claim occurrence*, indeed there is extensive academic work dedicated to understanding the dynamics of delayed (IBNR and RBNS) claims, for instance Dassios and Zhao (2013), Yuen et al. (2005), Boumezoued and Devineau (2017). For the same *claim occurrence* we often observe multiple *payment dates* corresponding to cash outflows for the insurance company. It is interesting for the company to understand the characteristics of these cash outflows, as for a case which is not typically closed after a single payment, it is beneficial to already estimate an amount which should be reserved for future *liabilities* from this known case. Modelling the endogenous structure of initial payments triggering future payments by use of a Hawkes process gives rise to interesting insights about the structure of the claim payment process.

The dataset we consider consists of claims which occurred in the years 2007 to 2011, were reported with a delay of three years during 2010 to 2014 and have corresponding payment dates during the time period from 01 January 2010 to 28 July 2016. For a detailed explanation of this data set choice, see Appendix A.1. For each claim payment, only the day of the payment is recorded as any finer granularity is not of particular interest to the insurance company. As the process is aggregated over multiple clients and claim occurrences, on some days there are multiple arrivals with the same timestamp (day). As the Hawkes process is a simple point process (see Definition 1), multiple arrivals with the same timestamp are theoretically not possible. In these cases, we distribute the indistinguishable arrivals uniformly over their arrival day in order to generate distinct timestamps. Note that this modification is only necessary for the parameter estimation step, as in all further applications we use increments of at least one day (thus the slightly shifted arrivals will be counted "indistinguishably" again). First, we test whether the claim arrivals could be described by a memoryless Poisson distribution as assumed by the classical model, where we use the same criteria as in Da Fonseca and Zaatour (2013) for trade clustering in stock and futures data. In Fig. 1, the interarrival times of claim payments are plotted against an exponential distribution. The figure indicates clearly that a Poisson model would not be suitable for the data. Fig. 2 displays the number of payments per week (7 days) over the whole time period of 2400 days, where clustering of payment occurrences over time can be observed.

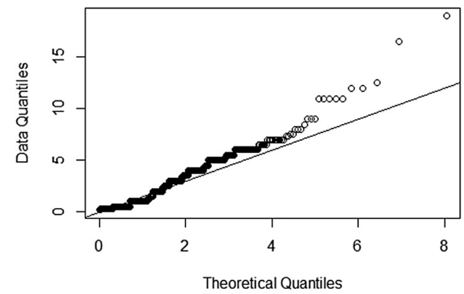


Fig. 1. The plot of empirical interarrival times against an exponential distribution indicates that a Poisson model would not be a suitable fit.

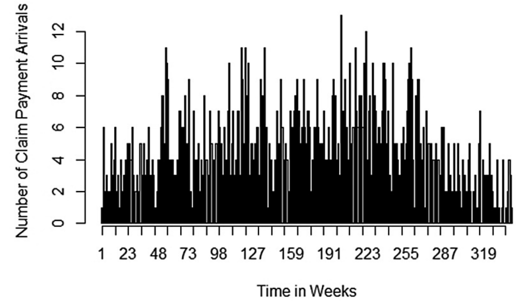


Fig. 2. The number of claim payments per week during the time period 01 Jan 2010 to 28 July 2016 gives further indication as to the presence of clustering.

Next, again following Da Fonseca and Zaatour (2013), we compute the empirical autocorrelation of the number of payments during intervals of fixed length τ separated by a lag of length δ. Thus we compute (see (12) and (13)):

$$AC(t, \tau, \delta) = \frac{C(t, \tau, \delta)}{\sqrt{V(t, \tau)V(t + \tau + \delta, \tau)}}. \tag{26}$$

We choose the interval length as (7, 14, 21, 28) days and let the lag range from 0 to 240 days by steps of 1. We plot the resulting autocorrelation values as a function of the time lag δ in Fig. 3.

We note that the autocorrelation between the number of payments in two intervals is a decreasing function of the time lag for all chosen interval lengths. This corroborates the use of a Hawkes process as it indicates that incoming arrivals in one period influence closely subsequent periods and this memory effect decays as time moves on. Note that a Poisson process would assume independence between the number of arrivals in subsequent intervals and thus a constant autocorrelation of 0 which is clearly not the case for the data.

Given these insights into the nature of payment arrival times, we proceed to fit a Hawkes process with exponentially decaying intensity to the arrival process. This process is chosen as

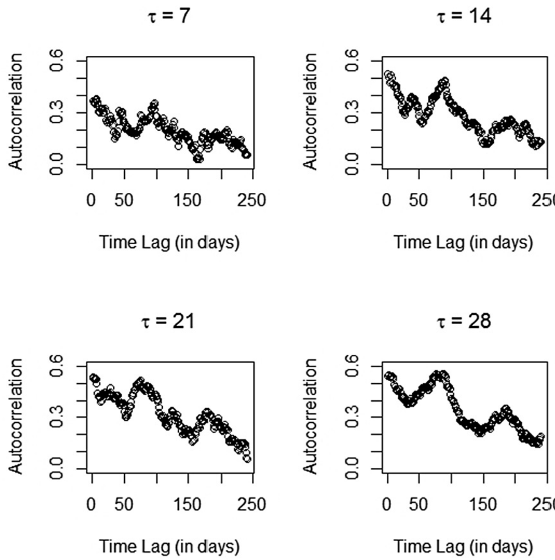


Fig. 3. Empirical autocorrelation function, as in (26), of the number of claim payments on intervals of length τ as a function of the time lag δ between the intervals. The overall trend is decreasing for all interval lengths. Clearly, the number of claim payment arrivals on consecutive intervals cannot be assumed independent.

Table 1
Parameter estimates $\hat{\lambda}, \hat{\alpha}, \hat{\beta}$ from MLE as well as the estimated number of arrivals on a unit interval $\mathbb{E}[N(1)] = \frac{\hat{\lambda}}{1-\hat{\alpha}/\hat{\beta}}$ (see (11)) compared to its empirical counterpart $\widehat{\mathbb{E}[N(1)]}$.

Parameter	ML Estimate
λ	0.1467
α	0.0260
β	0.0334
$\mathbb{E}[N(1)]$	0.6621
$\widehat{\mathbb{E}[N(1)]}$	0.6483

it is able to capture the features of the data such as clustering and autocorrelation decay (see Figs. 2 and 3) and is analytically tractable in the sense of Section 2.1.2. To this end, we estimate parameters λ, α and β from (7) using the maximum likelihood estimation described in Section 2.1.3. We use the *Nelder–Mead* optimization routine from the R package *lme4* to minimize the negative log-likelihood function. In order to avoid getting stuck in a local optimum, we repeat the optimization with 100 random starting values $(\lambda_0, \alpha_0, \beta_0)$ drawn uniformly from the interval $(0, 500)$ (such that $\alpha_0 < \beta_0$) and proceed with the estimations which yield the smallest value of the objective. The results are summarized in Table 1.

In order to test the goodness of fit of the Hawkes model, we first use the approach described in Section 2.1.3 and plot the transformed interarrival times against a unit Exponential distribution as shown in Fig. 4. We observe an improvement over the fit of a classical Poisson model in Fig. 1.

Furthermore, using (11), (12) and (14), we compare the theoretical expected value (on a unit interval), variance (on intervals of different length) and autocorrelation (on intervals of different lengths and time lags) of the number of jumps of an exponential Hawkes process with parameters from Table 1 with the corresponding values of the empirical arrivals in Table 1, Table 2 and Fig. 5 respectively.

Overall, these tests lead us to conclude that a risk model with claim arrivals according to a stationary Hawkes process

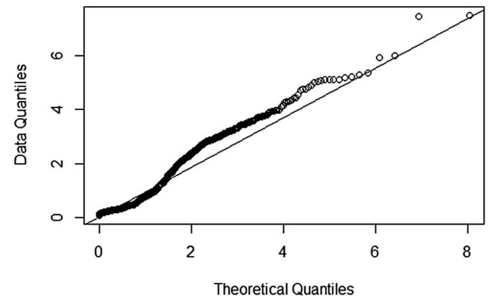


Fig. 4. The plot of transformed interarrival times against a unit exponential distribution indicates that an exponential Hawkes model with parameters $\hat{\lambda}, \hat{\alpha}, \hat{\beta}$ from Table 1 would be an acceptable fit.

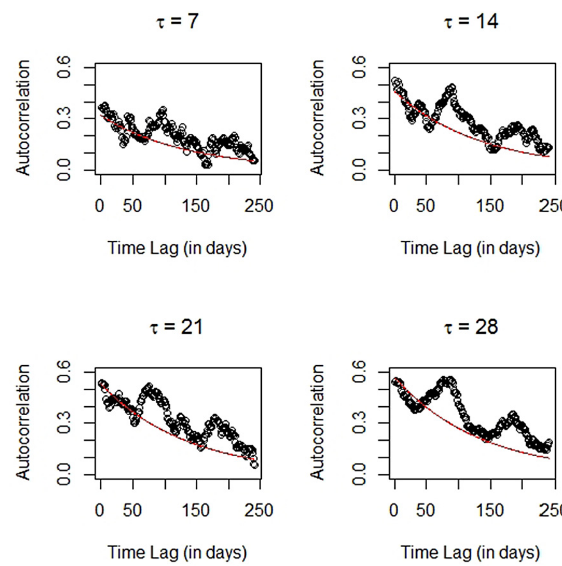


Fig. 5. Empirical autocorrelation function, as in (26), of the number of claim payments on intervals of length τ as a function of the time lag δ between the intervals compared to the corresponding theoretical values from Proposition 2 for an exponential Hawkes process with parameters $\hat{\lambda} = 0.1467, \hat{\alpha} = 0.0260, \hat{\beta} = 0.0334$.

Table 2
Comparison of the theoretical variance of the number of arrivals of a Hawkes process with parameters $\hat{\lambda} = 0.1467, \hat{\alpha} = 0.0260, \hat{\beta} = 0.0334$ according to (12) to the corresponding empirical variance.

Interval length (days)	Variance (theo.)	Variance (emp.)
7	6.9208	6.9995
14	18.2587	19.0570
21	33.7908	36.1109
28	53.3054	58.3725
35	76.6013	81.7779
42	103.4878	112.5545
49	133.7836	152.2974
56	167.3165	191.2224
63	203.9232	207.0156
70	243.4485	265.7647

with exponentially decaying intensity is suitable for our dataset.[2] We turn our attention to describing the sizes of outgoing claim payments in the next section.

[2] As we are interested in studying endogenous clustering and the chosen Hawkes process already fits the data well, we refrain from a comprehensive comparison with other classes of arrival processes as this would transcend the scope of this work.

3.3. Claim sizes

In insurance literature the claim sizes $\{Y_i\}$ are usually supposed to be i.i.d. with distribution G having finite first two moments $\mathbb{E}[Y_1] = m_1$ and $\mathbb{E}[Y_1^2] = m_2$. A common choice for G is an exponential distribution, say $Exp(\gamma)$, as for this choice one can e.g. obtain closed-form solutions for the ruin probabilities in the classical case (see e.g. Asmussen and Albrecher, 2010). Risk models with Hawkes process arrivals and i.i.d. claim sizes are studied e.g. in Stabile and Torrisi (2010) and Cheng and Seol (2018). However, for those models, the results are not easily applicable to empirical data and ruin probabilities can only be obtained numerically. Thus, we would like to make use of the theoretical results for *RH* which assumes claim sizes to follow a finite number of fixed jump sizes governed by the evolution of a Markov chain.[3] However, as we work with an aggregated portfolio of insurance claims from different payment streams, it would not be reasonable to assume a dependence of directly subsequent claim sizes in the overall portfolio. Thus, we "reinterpret" the Markov Chain (X_i) and the function $a(x)$ from Definition 2 in order for the modelled claim sizes to approximate an i.i.d. sequence following the empirical distribution of observed claim sizes. The approximation can be made arbitrarily well by increasing the number N of states of the Markov Chain.

Let $\hat{G}$ be the empirical distribution function of the claim sizes and let B be the maximum observed claim size, thus $\hat{G}(B) = 1$. We set equidistant boundaries $(b_1, b_2, \ldots, b_N = B)$ and define $\pi^* = (\pi_1^*, \ldots, \pi_N^*)$ as

$$\begin{aligned}
\pi_1^* &= \hat{G}(b_1)\\
\pi_2^* &= \hat{G}(b_2) - \pi_1^*\\
&\cdots\\
\pi_N^* &= \hat{G}(b_N) - \sum_{i=1}^{N-1}\pi_i^* = 1 - \sum_{i=1}^{N-1}\pi_i^*
\end{aligned} \tag{27}$$

Note that by definition $\sum_{i=1}^{N}\pi_i^* = 1$.

Let (X_i) be a Markov Chain on $X = \{1, \ldots, N\}$ with transition matrix

$$P = \begin{pmatrix}\pi_1^* & \pi_2^* & \cdots & \pi_N^*\\ \cdots & \cdots & \cdots & \cdots\\ \pi_1^* & \pi_2^* & \cdots & \pi_N^*\end{pmatrix}.$$

We know e.g. by Norris (2009) that, as (X_i) is an irreducible Markov Chain on a finite state space, it has a unique stationary distribution. Indeed, we can easily verify that the stationary distribution is again given by π^*:

$$\begin{aligned}
\pi^* P &= (\pi_1^*, \ldots, \pi_N^*)\begin{pmatrix}\pi_1^* & \cdots & \pi_N^*\\ \cdots & \cdots & \cdots\\ \pi_1^* & \cdots & \pi_N^*\end{pmatrix}\\
&= (\pi_1^*\sum_{i=1}^{N}\pi_i^*, \ldots, \pi_N^*\sum_{i=1}^{N}\pi_i^*)\\
&= (\pi_1^*, \ldots, \pi_N^*) = \pi^*.
\end{aligned} \tag{28}$$

Furthermore, as the columns of P are constant, for each state $k \in X$ it holds

$$\mathbb{P}(X_{i+1} = k \mid X_i = j) = \mathbb{P}(X_{i+1} = k \mid X_i = l) = \pi_k^* \quad \forall j, l \in X, i \in \mathbb{N} \tag{29}$$

and by the Markov property and the law of total probability

$$\begin{aligned}
\mathbb{P}(X_{i+1} = k) &= \sum_{j\in X}\mathbb{P}(X_{i+1} = k \mid X_i = j)\mathbb{P}(X_i = j)\\
&= \pi_k^*\sum_{j\in X}\mathbb{P}(X_i = j) = \pi_k^* \qquad \forall k \in X, i \in \mathbb{N}.
\end{aligned} \tag{30}$$

Thus, the probability of realizing one state is independent of the previous state and (X_i) essentially describes an i.i.d. sequence.

Now, let Y be a random variable with c.d.f. $\hat{G}$, $A_i := \{\omega : Y(\omega) \in (b_{i-1}, b_i]\}$, and set

$$\begin{aligned}
a(i) &= \mathbb{E}[Y \mid b_{i-1} < Y \le b_i] = \mathbb{E}[Y \mid A_i]\\
&= \frac{\mathbb{E}[Y\mathbf{1}_{A_i}]}{\mathbb{P}(A_i)} = \frac{\mathbb{E}[Y\mathbf{1}_{A_i}]}{\pi_i^*}
\end{aligned} \tag{31}$$

Then

$$a^* = \sum_{i=1}^{N}\pi_i^* a(i) = \sum_{i=1}^{N}\mathbb{E}[Y\mathbf{1}_{A_i}] = \mathbb{E}[Y]$$

and $a(X_i)$ describes an i.i.d. sequence that approximates the distribution $\hat{G}$ arbitrarily close as the number of states $N \to \infty$.

For the empirical example, Table 3 gives the values of the equidistant boundaries $b = (b_1, \ldots, b_N = B)$, state values $a_i := a(i)$ and the distribution π_i^* along with the value of a^* (which coincides with $\mathbb{E}[Y_1]$) for the case of a 5-state Markov Chain. Note that to replicate empirical claim sizes, usually significantly more states would be used, for Table 3 the size $N = 5$ is chosen for the sake of presentation. Note that the number of states should be chosen such that there is no segment without observations, as this would not lead to an irreducible Markov chain.

Table 3
Boundaries $b = (b_1, \ldots b_N)$, stationary distribution π^*, state values $a(i)$ and expected value a^* under the stationary distribution for a 5-state Markov Chain and empirical claim sizes.

Parameter	Value
$(b_1, \ldots, b_5 = B)$	(2014.2, 4028.4, 6042.6, 8056.8, 10071)
$(\pi_1^*, \ldots, \pi_5^*)$	(0.9017, 0.0720, 0.0206, 0.0032, 0.0026)
$(a_1, \ldots, a_5)$	(499.5056, 2821.8888, 4743.6872, 7049.5920, 9199.8750)
a^*	797.3672
$E[X_1]$	797.3672

To corroborate that the generated claim sizes indeed describe claims with distribution function $\hat{G}$, we compare in Fig. 6 the distribution function of the empirical claims with its counterpart from claims generated by the Markov Chain approach with 50 states and draw the corresponding QQ plot. Overall, we conclude that empirical claim sizes are replicated reasonably well within the framework of *RH*.

3.4. Risk process

In order to simulate the risk model from Definition 2, we need to estimate values for the initial capital u and the premium rate c. As this information is not inferable from the empirical data set, we calculate the premium rate using the expected value principle and Corollary 1 with a safety loading of $\theta = 0.2$ to obtain

$$c = (1+\theta)a^*\frac{\lambda}{1-\alpha/\beta} = 633.5552. \tag{32}$$

This seems a reasonable number considering the number of policies in the portfolio and the mean yearly premium for such contracts. It has to be kept in mind that in practice, the majority of policyholders are likely to never occur a claim, providing additional premium income for the company which is not considered in our data set.

[3] Let us reiterate that while in our particular context, we do not assume dependence between subsequent claim sizes, in general the Markov Chain approach of *RH* offers a generalization of the usual i.i.d. assumption into contexts where dependence between claim sizes should be assumed, such as modelling of claims from natural catastrophes (see e.g. Albrecher and Teugels, 2006, Boudreault et al., 2006, Albrecher and Boxma, 2004).

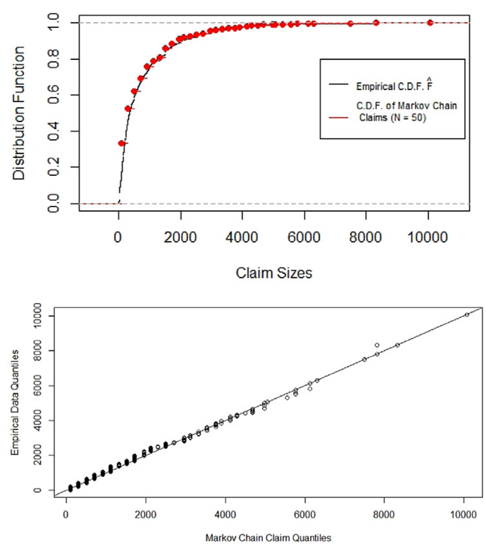


Fig. 6. The distribution function of claim sizes generated by the Markov Chain approach for a 50-state Markov Chain against the empirical distribution function of claim sizes indicates that the approach replicates claim sizes as assumed. This is corroborated by the corresponding QQ plot.

We set the initial capital as $u = 8000$ which seems reasonable given a mean claim size of around 800 and an expected number of 0.6483 claims per day. In fact, u is best thought of as a variable - i.e. how much initial capital has to be provided in order for the ruin probability over a certain period to be below a given bound. We generate $L = 1000$ simulations of a risk process with u and c given above, the arrival process being an exponential Hawkes process with parameters $\hat{\lambda} = 0.1467, \hat{\alpha} = 0.0260, \hat{\beta} = 0.0334$ and claim sizes being generated by a Markov chain with 50 states according to the procedure described above. Fig. 7 compares the underlying empirical risk process to the first 50 simulations. In order to assess whether our simulated paths accurately depict the empirical one, we compare the fluctuations over time and the final value at time $T = 2400$ using the metrics

$$\hat{S}(L) = \frac{1}{L}\sum_{i=1}^{L}\frac{\max(\hat{R}_i(t)) - \min(\hat{R}_i(t))}{\max(R(t)) - \min(R(t))}, \tag{33}$$

$$\hat{F}(L) = \frac{1}{L}\sum_{i=1}^{L}\frac{\hat{R}_i(T)}{R(T)}, \tag{34}$$

where L denotes the number of simulated paths, $\hat{R}_i(t)$ refer to the simulated risk processes and $R(t)$ to the benchmark (empirical process). Note that the metric $\hat{S}$ was suggested by Zhang (2016) in the context of comparing the fit of Hawkes models with exponential and power law kernels to empirical data. Table 4 gives an overview of the results. Overall, we conclude that RH is able to describe the empirical data reasonably well.

4. Diffusion approximation and ruin probabilities

After reassuring that RH is suitable for empirical data, we review that the risk process can be approximated by a jump–diffusion process following Swishchuk (2017a). From this jump–diffusion approximation, we construct an approximation by a pure diffusion process which allows analytical calculation of estimates for finite-time and infinite-horizon ruin probabilities.

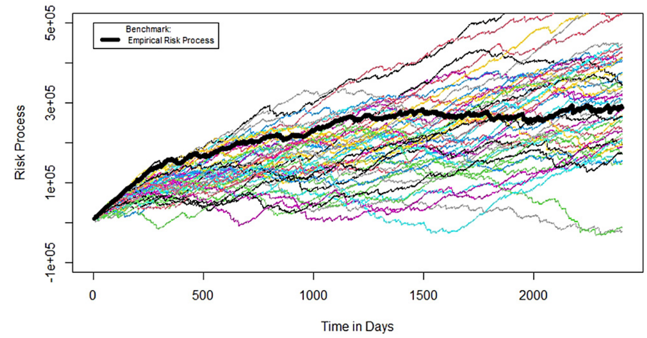


Fig. 7. Plotting the empirical risk process with parameters $u = 8000$ and $c = 633.5552$ against 50 simulated paths of RH with arrivals following an exponential Hawkes process with $\hat{\lambda} = 0.1467, \hat{\alpha} = 0.0260, \hat{\beta} = 0.0334$ and claims following a Markov Chain as described in Section 3.3. Note that in the first part of the graph, the empirical process seems to have an extraordinarily high upward drift and is on the "upper bound" of the simulations. This is reasonable as for the first period, our data set is naturally missing payments from claims which were reported before the start of the observation period and continue to induce payments within it. This dynamic disappears as we pass on further in time, the high drift vanishes and the empirical process is well covered by the simulations.

Table 4
We assess how well simulations using a Hawkes arrival process and claim sizes generated by the Markov Chain approach replicate the empirical risk process using fluctuations and final capital values as metrics.

Parameter	Value
$\hat{S}$	1.2202
$\hat{F}$	1.1321
$R(T)$ (Empirical)	287829.1651
$\mathbb{E}[R(T)]$ (Simulation)	325865.5270
$\sqrt{\mathbb{V}ar[R(T)]}$ (Simulation)	140836.3261

4.1. Approximation by jump–diffusion process

Theorem 3 (*FCLT, Approximation by Jump–Diffusion Process Swishchuk, 2017a*). *Let $R(t)$ be the risk model defined in Definition 2, and (X_i) be an ergodic Markov Chain with stationary distribution π^*. We suppose that*
$0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1$ and $\int_0^\infty s\mu(s)ds < \infty$. Then:

$$\lim_{t\to\infty}\frac{R(t) - (ct - a^*N(t))}{\sqrt{t}} \overset{D}{=} \hat{\sigma}\Phi(0,1) \tag{35}$$

(or in Skorokhod topology (see Skhorokhod, 2014))

$$\lim_{n\to\infty}\frac{R(nt) - (cnt - a^*N(nt))}{\sqrt{n}} \overset{D}{=} \hat{\sigma}W(t) \tag{36}$$

where $\Phi(\cdot,\cdot)$ is the standard Normal c.d.f. and $W(t)$ is a standard Wiener process.

$$\begin{aligned}
\hat{\sigma} &:= \sigma^*\sqrt{\lambda/(1-\hat{\mu})}, \quad (\sigma^*)^2 := \sum_{i\in X}\pi_i^*\nu(i)\\
a^* &:= \sum_{i\in X}\pi_i^* a(i), \quad d(i) := a^* - a(i)\\
\nu(i) &:= d(i)^2 + \sum_{j\in X}(g(j)-g(i))^2P(i,j)\\
&\quad - 2d(i)\sum_{j\in X}(g(j)-g(i))P(i,j)\\
g &:= (P + \Pi^* - I)^{-1}(d(1),\dots,d(n))'
\end{aligned} \tag{37}$$

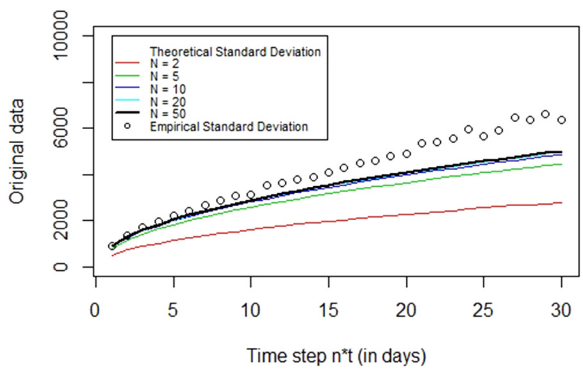


Fig. 8. Error estimation of jump–diffusion approximation by comparison of standard deviations of empirical process in (39) and corresponding theoretical values from Theorem 3. Theoretical values are calculated for a Markov chain with $N = (2, 5, 10, 20, 50)$ states, where the standard deviation naturally approaches the empirical value as N increases.

where P is the transition matrix for (X_i) and Π^* is the matrix of stationary probabilities of P, meaning that the rows of Π^* coincide with the stationary distribution.

Proof. See Swishchuk (2017a).

Note that the theorem holds analogously for a model with Poisson process arrivals with rate λ_P and i.i.d. claim sizes $\{Y_i\}$ with $\hat{\sigma} := \sqrt{\lambda_P}\sqrt{\mathbb{V}ar[Y_1]}$. The theorem implies that $R(t)$ can be approximated by the jump–diffusion process

$$R(t) \approx u + ct - a^* N(t) + \hat{\sigma} W(t) \tag{38}$$

where a^* and $\hat{\sigma}$ are defined above, $N(t)$ is a Hawkes process and $W(t)$ is a standard Wiener process. To assess the accuracy of the approximation, we proceed as suggested in Swishchuk et al. (2019) and compare the standard deviation on the right-hand side of (36) multiplied by $\sqrt{n}$ to its empirical counterpart on the left-hand side, that is the standard deviation of

$$R(nt) - (cnt - a^* N(nt)) = u - \sum_{k=1}^{N(nt)} (a(X_k) - a^*) \tag{39}$$

To this end, we choose t as the original time scale of one day and let n run from 1 day to 30 days by steps of 1 day. At each step int, we compute the value of the process

$$(R(int) - (cint - a^* N(int)))- \\ - (R((i-1)nt) - (c(i-1)nt - a^* N((i-1)nt))).$$

We compare the standard deviation of these values to the standard deviation theoretically obtained on the right-hand side of (36) multiplied by $\sqrt{n}$. Note that this approximation should be naturally only accurate for large n, however due to our relatively short time frame of $T = 2400$ days, for large n the left-hand side of (36) is only based on few observations. To ensure statistical significance, we thus chose the sequence of n such that each observation for the empirical standard deviation value is based on at least 80 data points. The results are summarized in Fig. 8 and we can see that for small n ($n \leq 15$) they look quite accurate.

4.2. Approximation by pure diffusion process

Theorem 4 (*FCLT 2, Approximation by Pure Diffusion Process*). *Let $R(t)$ be the risk model from Definition 2, and let (X_i) be an ergodic Markov chain with stationary distribution π^*. We suppose that $0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1$ and $\int_0^\infty s\mu(s)ds < \infty$. Then:*

$$\lim_{t\to\infty} \frac{R(t) - (ct - a^* \frac{\lambda}{1-\hat{\mu}} t)}{\sqrt{t}} \stackrel{D}{=} \bar{\sigma}\Phi(0,1) \tag{40}$$

(or in Skorokhod topology (see Skhorokhod, 2014))

$$\lim_{n\to\infty} \frac{R(nt) - (cnt - a^* \frac{\lambda}{1-\hat{\mu}} nt)}{\sqrt{n}} \stackrel{D}{=} \bar{\sigma} W(t) \tag{41}$$

where $\Phi(\cdot,\cdot)$ is the standard Normal c.d.f. and $W(t)$ is a standard Wiener process and $\bar{\sigma} = \sqrt{\hat{\sigma}^2 + \left(a^* \sqrt{\frac{\lambda}{(1-\hat{\mu})^3}}\right)^2}$ where a^* and $\hat{\sigma}$ are defined in Theorem 3.

Note that for the classical case of a Poisson process with rate λ_P and i.i.d. claim sizes $\{Y_i\}$ we obtain

$$\bar{\sigma}_P = \sqrt{\lambda_P \mathbb{V}ar[X_1] + \mathbb{E}[X_1]^2 \lambda_P} = \sqrt{\lambda_P \mathbb{E}[X_1^2]} \tag{42}$$

which is consistent with classical results.

Proof. See Appendix A.2.

The theorem implies that $R(t)$ can be approximated by the pure diffusion process

$$R(t) \approx u + ct - a^* \frac{\lambda}{1-\hat{\mu}} t + \bar{\sigma} W(t) \tag{43}$$

where a^* and $\bar{\sigma}$ are defined above and $W(t)$ is a standard Wiener process. We use the same approach as above with $\hat{\mu} := \frac{\alpha}{\beta}$ for the exponential Hawkes process. This time we compare the standard deviation of

$$R(nt) - \left(cnt - a^* \frac{\lambda}{1-\alpha/\beta} nt\right) = u - \left(\sum_{k=1}^{N(nt)} a(X_k) - a^* \frac{\lambda}{1-\alpha/\beta} nt\right) \tag{44}$$

with its counterpart on the right-hand side in (41), that is $\sqrt{n}\sqrt{t}\bar{\sigma}$. The results can be seen in Fig. 9. In this case the approximation is not very accurate. This most likely originates in the approximation via the CLT which always entails a hardly measurable approximation error depending on the model parameters. Indeed, if we go back to the (more accurate) jump–diffusion approximation (38)

$$R(t) \approx u + ct - a^* N(t) + \hat{\sigma} W(t),$$

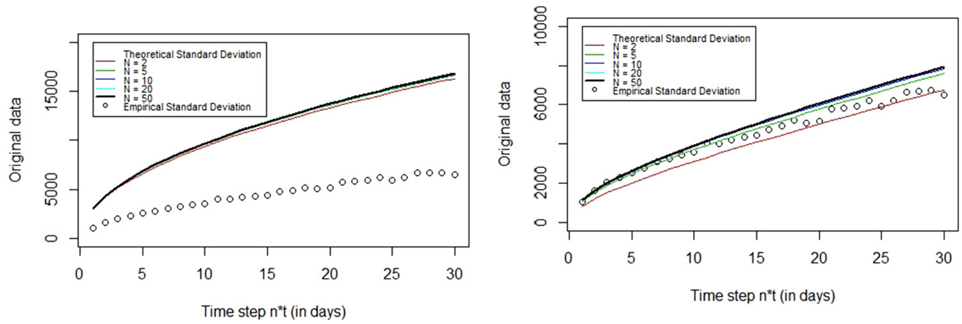
where $\hat{\sigma} = \sqrt{\frac{\lambda}{1-\alpha/\beta}}\sigma^*$, $N(t)$ is a Hawkes process and $W(t)$ is a standard Wiener process, we obtain the variance of the risk process as

$$\begin{aligned}
\mathbb{V}ar[R(t)] &= (a^*)^2 \mathbb{V}ar[N(t)] + \hat{\sigma}^2 t \\
&= (a^*)^2 \left(\frac{\lambda}{1-\alpha/\beta} t \left(\frac{1}{1-\alpha/\beta} \right)^2 + (*) \right) + (\sigma^*)^2 \frac{\lambda}{1-\alpha/\beta} t \\
&= t \left((\sigma^*)^2 \frac{\lambda}{1-\alpha/\beta} + (a^*)^2 \frac{\lambda}{(1-\alpha/\beta)^3} \right) - (**) \\
&= t(\bar{\sigma})^2 - (**)
\end{aligned} \tag{45}$$

where

$$(*) = \frac{\lambda}{1-\alpha/\beta} \left(1 - \left(\frac{1}{1-\alpha/\beta} \right)^2 \right) \left(\frac{1-e^{-t(\beta-\alpha)}}{\beta-\alpha} \right) \tag{46}$$
$$(**) = -(a^*)^2 (*) > 0$$

where we have used the variance of the number of jumps $N(t)$ of a Hawkes process from (12). We note that the pure diffusion approximation naturally does not capture the (negative) non-linear



(a) Original Data. (b) Original Data, with correction $(**)$ from (46).

Fig. 9. (a) Error Estimation of pure diffusion approximation by comparison of standard deviations of empirical process in (44) and corresponding theoretical values from Theorem 4. Theoretical values are calculated for a Markov Chain with $N = (2, 5, 10, 20, 50)$ states, where the standard deviation increases as N increases. We observe that the theoretical values largely overestimate the empirical ones for this data set. (b) Plotting the theoretical standard deviation values from Theorem 4, corrected by $(**)$ from (46), shows a very close match of theoretical and empirical values.

influence $(**)$ on the variance. Indeed, if we plot the standard deviation of the process

$$R(nt) - \left(cnt - a^* \frac{\lambda}{1-\alpha/\beta} nt \right) = u - \left(\sum_{k=1}^{N(nt)} a(X_k) - a^* \frac{\lambda}{1-\alpha/\beta} nt \right)$$

against $\sqrt{\bar{\sigma}^2 nt - (**)(nt)}$ in Fig. 9, we see a very close match. However, the absolute value of the term $(**)$ decreases as the difference $\beta - \alpha$ increases, thus for suitable parameters the additional error of the pure diffusion approximation becomes almost negligible. Note that for the classical case of a Poisson process with rate λ_P and i.i.d. claim sizes $\{Y_i\}$, it holds

$$\begin{aligned} \mathbb{V}ar[R(t)] &= \mathbb{E}[Y_1]^2 \mathbb{V}ar[N(t)] + (\sqrt{\lambda_P}\sqrt{\mathbb{V}ar[Y_1]})^2 t \\ &= \mathbb{E}[Y_1]^2 \lambda_P t + \lambda_P \mathbb{V}ar[Y_1] t \\ &= t\lambda_P \mathbb{E}[Y_1^2] = t(\bar{\sigma}_P)^2 \end{aligned}$$

which is consistent with the approximation by the pure diffusion process and classical results.

4.3. Ruin probabilities

We now turn our attention to estimating ruin probabilities for *RH* using the approximations derived in the last section. In the case of the (more accurate) jump–diffusion approximation, we compare numerically the ruin probabilities from simulations of *RH* and simulations of the diffusion process with Hawkes process jumps from (38) and observe in Fig. 10 that they are quite similar in all cases. Using the pure diffusion approximation, we can apply well-known formulas for ruin probabilities of diffusion processes (see Iglehart, 1969; Whitt, 1970; Asmussen and Albrecher, 2010) to give closed-form expressions for the ruin probabilities for the case of *RH*.

Theorem 5 (*Ruin Probabilities for RH*)**.** *The ruin probability in the interval* $(0, \tau]$ *for a risk model as in Theorem 4 is given by*

$$\Psi(u,\tau) = \Phi\Big(- \frac{u + (c - a^*\lambda/(1-\hat{\mu}))\tau}{\bar{\sigma}\sqrt{\tau}} \Big) + e^{-\frac{2(c-a^*\lambda/(1-\hat{\mu}))}{\bar{\sigma}^2}u} \Phi\Big(- \frac{u - (c - a^*\lambda/(1-\hat{\mu}))\tau}{\bar{\sigma}\sqrt{\tau}} \Big), \quad (47)$$

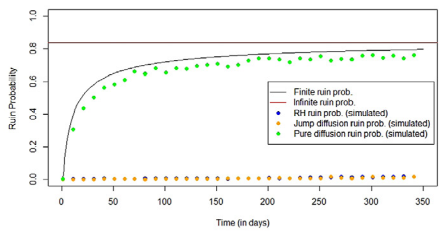
and the ultimate ruin probability for a risk model as in Theorem 4 is given by

$$\Psi(u) = e^{-\frac{2(c-a^*\lambda/(1-\hat{\mu}))}{\bar{\sigma}^2}u}. \quad (48)$$

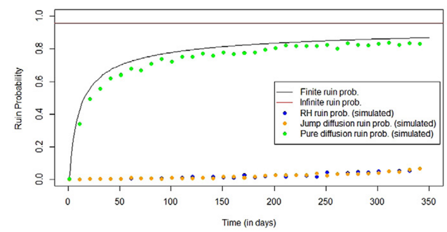
In Fig. 10, we compute and plot the finite-horizon ruin probability for increasing t from 1 until 350 days as well as the infinite-horizon ruin probability. We compare them with values obtained from 1000 simulations of the corresponding *RH* and of the two approximation processes. For the parameters obtained from the original data set, the numerical results from the jump–diffusion approximation are quite accurate, while clearly the theoretical and simulated ruin probabilities for the pure diffusion approximation overestimate the simulated ones for *RH*, an effect that originates from the large estimation error in (43) (see Figs. 10(a) and 10(b)). As we have seen above, the diffusion coefficient tends to overestimate the standard deviation of the risk process and therefore an exaggerated ruin probability is assigned. Assuming the same claim size parameters and a different set of parameters for the Hawkes arrival process such that $\beta - \alpha$ is no longer very close to 0 leads to a significant increase in the accuracy of the ruin probability estimation via Theorem 5 as displayed in Fig. 10(c). However, for similar parameters but with a branching ratio close to 1 (indicating a very strong presence of clustering), we again observe a large estimation error in Fig. 10(d).[4]

We conclude that the pure diffusion approximation of *RH* is generally suitable for calculating ruin probabilities for a risk model with Hawkes process claims, however, for any application it has to be kept in mind that the approximation error can become very large depending on the model parameters. In any case, having such a closed-form formula for the ruin probability can give helpful indications in practice, where an insurer often faces the challenge of estimating how much capital has to be reserved at one point in time in order to limit the probability of the event that the value of (a part of) his insurance portfolio falls below a certain lower bound during a future time period. Of course, the relevant threshold here would usually not be 0 (indicating a positive probability of insolvency), but at least a certain amount acting as an *emergency risk buffer* according to regulatory requirements. For some claim classes it is thus essential to understand how dependencies between claim arrivals can cause temporal clustering and how this feature affects the risk of “ruin”. The ruin probability estimates from our model with self-exciting Hawkes processes might serve as a helpful tool here.

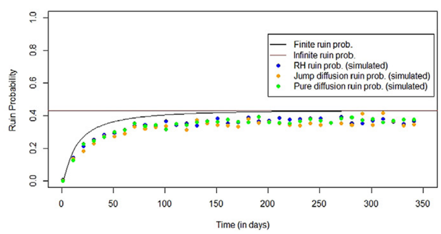
[4] Note that general results for the rate of convergence for the diffusion approximation of risk processes are available, see Swishchuk (2000). In this case, they are of order $\frac{C(T)}{\sqrt{n}}$, $0 \le t \le T$, where T is a horizon time, $C(T)$ is a constant and n refers to the scaling of the time axis as above.



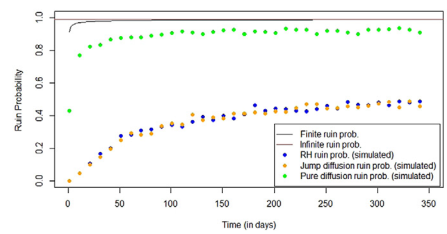
(a) $(\lambda, \alpha, \beta) = (0.1467, 0.0260, 0.0334), \theta = 0.2$



(b) $(\lambda, \alpha, \beta) = (0.1467, 0.0260, 0.0334), \theta = 0.05$



(c) $(\lambda, \alpha, \beta) = (1, 2, 5), \theta = 0.2$



(d) $(\lambda, \alpha, \beta) = (1, 2, 2.1), \theta = 0.2$

Fig. 10. We calculate finite-time ruin probabilities for intervals $[0, t)$, where t ranges from 1 to 350 days by steps of 1, and infinite-time ruin probability according to Theorem 5. We compare their values to finite-time ruin probabilities for t from 1 to 350 days by steps of 10 obtained from 1000 simulations of *RH* and the pure and jump–diffusion approximation process. We observe that for any parameter set, the simulated probabilities for *RH* and the jump–diffusion process are quite similar. In the case of the pure diffusion process and the theoretical values, the accuracy of the approximation deteriorates whenever the difference between β and α is relatively small (as should be expected from (46)).

5. Optimal investment

As an application of the Hawkes risk model, we would like to study how replacing the classical Poisson process for claim arrivals by a self-exciting Hawkes process influences the risk of the insurance company and their optimal investment decisions. To this end, we approximate the risk process by the pure diffusion process in (43) and regard it as the company's *liability* evolving according to a Brownian motion with drift. We then use the results by Xie et al. (2008) on mean–variance portfolio selection with multiple risky assets and one liability in an incomplete market where the risk from the liability cannot be completely hedged by trading the available assets and therefore the event of ruin cannot be excluded. Xie et al. (2008) work in a mean–variance framework, thus the goal is to maximize the expected terminal wealth while limiting the variance. They derive the mean–variance efficient frontier and the optimal strategy by applying the general stochastic LQ control technique.

5.1. The market model

Let $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ be a complete probability space equipped with the filtration $\mathbb{F} = (\mathcal{F}(t))_{t\in[0,T]}$ generated by an $(n+1)$-dimensional Brownian motion $\{(W_0(t), W_1(t), \dots, W_n(t))' : t \in [0, T]\}$ for $n \in \mathbb{N}$, where $0 < T < \infty$ is a fixed time horizon, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_T = \mathcal{F}$ and the superscript "′" indicates the transpose of a vector or matrix. Denote by $\mathcal{C}([0, T]; \mathbb{R}^{n\times k})$ the class of continuous bounded deterministic functions on $[0, T]$ with values in $\mathbb{R}^{n\times k}$. Assume the company at time $t = 0$ is equipped with an initial endowment of $w > 0$ and an initial liability l, such that its net initial wealth is $x = w - l > 0$. Consider a financial market with $(m + 1)$ assets being traded continuously, where $m \le n$ and the assets are labelled $i = 0, 1, \dots, m$, where $i = 0$ refers to the riskfree asset. Assume the company can dynamically adjust its investment portfolio during the time period $[0, T]$ without incurring transaction fees or short-selling restrictions. Trading takes place self-financing without taking consumption into account. The price of the risk-free asset $S_0(t)$ evolves according to the ODE

$$dS_0(t) = r(t)S_0(t)dt, \qquad S_0^0 = 1, \tag{49}$$

where $r(t) \in \mathcal{C}([0, T]; \mathbb{R}^+)$ denotes the risk-free interest rate. The price processes $S_1(t), \dots, S_m(t)$ of the risky assets evolve according to the SDEs

$$dS_i(t) = b_i(t)S_i(t)dt + \sigma_i(t)S_i(t)dW(t), \quad S_0^i = s_i \in \mathbb{R}, \tag{50}$$

where $i = 1, \dots, m$, $W(t) := (W_1(t), \dots, W_n(t))'$ and $\sigma_i(t) := (\sigma_{i1}(t), \dots, \sigma_{in}(t)) \in \mathcal{C}([0, T]; \mathbb{R}^{1\times n})$ denotes the *volatility* vector of the ith risky asset, thus define the matrix

$$\sigma(t) := (\sigma_1'(t), \dots, \sigma_m'(t))' \in \mathcal{C}([0, T]; \mathbb{R}^{m\times n}) \tag{51}$$

and let

$$b(t) := (b_1(t), \dots, b_m(t))' \in \mathcal{C}([0, T]; \mathbb{R}^{m\times 1}) \tag{52}$$

denote the *rate of return* of the risky assets. Denote the company's cumulative liability at time t by $L(t)$ and assume $L(t)$ evolves according to the SDE

$$dL(t) = g(t)dt + \upsilon(t)dB(t), \qquad L(0) = l, \tag{53}$$

where $\{B(t) : t \in [0, T]\}$ is a standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$. Denote by $\rho_j(t)$ the correlation coefficient between $B(t)$ and $W_j(t)$ for $j = 1, \dots, n$ and let

$$\rho(t) := (\rho_1(t), \dots, \rho_n(t))' \in \mathcal{C}([0, T]; \mathbb{R}^{n\times 1}) \tag{54}$$

be the *correlation coefficient vector*. Thus $B(t)$ can be expressed as

$$B(t) = \rho(t)'W(t) + \sqrt{1 - \rho(t)'\rho(t)}W_0(t) \tag{55}$$

where $B(t), W_0(t), W_1(t), \dots, W_n(t)$ are standard Brownian motions and $\rho(t)'\rho(t) \le 1$ for all $t \in [0, T]$. Combining (53) and (55), $L(t)$ evolves according to

$$dL(t) = g(t)dt + \upsilon(t)\rho(t)'dW(t) + \upsilon(t)\sqrt{1 - \rho(t)'\rho(t)}dW_0(t), \quad L_0 = l. \tag{56}$$

Thus, the evolution of the liability is generally assumed to be dependent of the risky assets' prices and in particular for $\rho(t)'\rho(t) < 1$ the risk arising from the liability can never be completely eliminated by trading the assets. For $\rho(t)'\rho(t) = 1$ the assets and the liability are driven by the same source of randomness but as long as $n > m$, the market is incomplete. Only for the case $n = m$ and $\rho(t)'\rho(t) = 1$, the risk from the liability can be completely hedged by trading the m available assets. Assume that for $i = 1, \dots, m$ and $j = 1, \dots, n$, $r(t), b_i(t), \sigma_{ij}(t), g(t), \upsilon(t), \rho_j(t)$ are deterministic functions of t, $b_i(t) > r(t)$, and there exists $\epsilon > 0$ such that $\sigma(t)\sigma(t)' \geq \epsilon \mathbb{I}_m$ for any $t \in [0, T]$, where $\mathbb{I}_m$ is the $m \times m$ identity matrix.

Denote by $\eta_i(t)$ the number of units of asset i held by the company at time t. Then $\varphi_i(t) := \eta_i(t)S_i(t)$ denotes the amount of money invested in asset i at time t. Let $\varphi(t) := (\varphi_1(t), \dots, \varphi_m(t))'$, then we call the process $\varphi := \{\varphi(t) : t \in [0, T]\}$ a *trading strategy*.

Let $X(t)$ be the *net wealth* of the company at time t. We only consider self-financing trading strategies, i.e. assume $X(t)$ evolves according to

$$dX(t) = \sum_{i=0}^{m} \eta_i(t) dS_i(t) - dL(t), \qquad X_0 = x > 0. \tag{57}$$

Inserting (49), (50) and (56), this is equivalent to

$$\begin{aligned} dX(t) &= (r(t)X(t) + (b(t) - r(t)\mathbb{1})'\varphi(t) - g(t))dt + (\varphi(t)'\sigma(t) \\ &\quad - (\upsilon(t)\rho(t))')dW(t) - \upsilon(t)\sqrt{1 - \rho(t)'\rho(t)}dW_0(t), \\ X(0) &= x, \end{aligned} \tag{58}$$

where $\mathbb{1}$ denotes the unit vector of length m.

The *set of admissible trading strategies* for initial wealth x is defined as

$$\mathcal{A}(x) := \left\{\varphi : \varphi(t) \in \mathcal{L}^2_{\mathcal{F}}([0, T]; \mathbb{R}^m), (X(t), \varphi(t)) \text{ satisfies } (58)\right\} \tag{59}$$

where $\mathcal{L}^2_{\mathcal{F}}([0, T]; \mathbb{R}^m)$ denotes all $\mathbb{R}^m$-valued, progressively measurable and square integrable random variables on $[0, T]$ under $\mathbb{P}$ with $\left(E\left[\int_0^T |\varphi(t)|^2 dt\right]\right)^{\frac{1}{2}} < \infty$.

A strategy $\varphi \in \mathcal{A}(x)$ is considered *optimal* if it solves the optimization problem

$$P(\chi) \quad \min_{\varphi \in \mathcal{A}(x)} \left(-\mathbb{E}[X(T)] + \chi \mathbb{V}ar[X(T)]\right) \tag{60}$$

which is equivalent to the classical mean–variance model where $\chi \in [0, \infty)$ expresses the weight (or importance) assigned to the objective $\mathbb{V}ar[X(T)]$ by the company.

5.2. Optimal strategy and mean–variance efficient frontier

Xie et al. (2008) find the solution of problem $P(\chi)$ by introducing the auxiliary problem

$$A(\chi, \omega) \quad \min_{\varphi \in \mathcal{A}(x)} \mathbb{E}[\chi X^2(T) - \omega X(T)] \tag{61}$$

whose relation to the original problem is derived in Zhou (2000). They use the general stochastic LQ optimal control theory from Yong and Zhou (1999) to obtain the optimal feedback control and optimal cost functional of the auxiliary problem $A(\chi, \omega)$. Substituting the obtained solution into the original setting and inserting the value of the auxiliary variable ω^*, they derive the solution of the original problem as stated in the following theorem.

Theorem 6 (*Optimal Feedback Control for $P(\chi)$ Xie et al., 2008*). *The optimal feedback control (i.e. the optimal strategy) of the problem $P(\chi)$ is given by*

$$\varphi^*(X(t)) = -\tau(t)\big(X(t) + \vartheta(t) - \gamma\big) - \zeta(t), \qquad t \in [0, T] \tag{62}$$

where

$$\tau(t) := \big(\sigma(t)\sigma(t)'\big)^{-1}\big(b(t) - r(t)\mathbb{1}\big)' \in \mathcal{C}\big([0, T]; \mathbb{R}^{m\times 1}\big)$$

$$\zeta(t) := \big(\sigma(t)\sigma(t)'\big)^{-1}\sigma(t)(-\upsilon(t)\rho(t)) \in \mathcal{C}\big([0, T]; \mathbb{R}^{m\times 1}\big) \tag{63}$$

$$\varsigma(t) := (b(t) - r(t)\mathbb{1})\tau(t) \in \mathcal{C}\big([0, T]; \mathbb{R}^{+}\big) \tag{64}$$

$$\kappa(t) := -(b(t) - r(t)\mathbb{1})\zeta(t) - g(t) \in \mathcal{C}([0, T]; \mathbb{R}) \tag{65}$$

$$\vartheta(t) := \int_t^T \kappa(s)e^{\int_s^t r(z)dz} ds + \gamma\left(1 - e^{-\int_t^T r(z)dz}\right) \tag{66}$$

$$a_0 := \int_0^T \kappa(t)e^{\int_t^T r(z)dz - \int_0^T \varsigma(z)dz} dt + xe^{\int_0^T (r(z) - \varsigma(z))dz}$$

$$a_1 := 1 - e^{-\int_0^T \varsigma(z)dz}, \ \gamma := \gamma^* = \frac{1}{2\chi(1 - a_1)} + \frac{a_0}{1 - a_1}.$$

Note that the expected terminal wealth under an optimal solution of $P(\chi)$ is then given by

$$\mathbb{E}[X^*(T)] = \frac{2\chi a_0 + a_1}{2\chi(1 - a_1)}.$$

Theorem 7 (*Mean–variance Efficient Frontier Xie et al., 2008*). *The efficient frontier of the mean–variance portfolio selection problem $P(\chi)$, if it ever exists, is given by*

$$\mathbb{V}ar[X^*(T)] = \frac{e^{-\int_0^T \varsigma(z)dz}}{1 - e^{-\int_0^T \varsigma(z)dz}}\big[\mathbb{E}[X^*(T)] - \mathcal{D}_1\big]^2 + \mathcal{D}_2 \tag{67}$$

with

$$\mathcal{D}_1 := xe^{\int_0^T r(z)dz} + \int_0^T \kappa(t)e^{\int_t^T r(z)dz} dt$$

and

$$\begin{aligned} \mathcal{D}_2 := \int_0^T \Big[& \upsilon^2(t)(1 - \rho(t)'\rho(t)) + \delta(t)'\delta(t) - \\ & \delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t)\Big] e^{\int_t^T (2r(z) - \varsigma(z))dz} dt \end{aligned}$$

for $\mathbb{E}[X^*(T)] \geq \mathcal{D}_1$, where $\varsigma(t)$ and $\kappa(t)$ are defined in (64) and (65) *respectively and $\delta(t) := -\upsilon(t)\rho(t) \in \mathcal{C}([0, T]; \mathbb{R}^{n\times 1})$.*

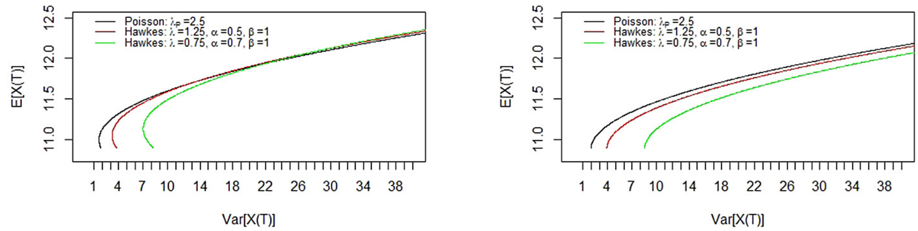
Note that $\mathcal{D}_2 \geq 0$ as shown in Xie et al. (2008). In particular, the only case where $\mathcal{D}_2 = 0$ holds, i.e. a risk-free portfolio on the efficient frontier can be reached, occurs for $n = m$ and $\rho(t)'\rho(t) = 1$. In this case $\sigma(t)'\big(\sigma(t)\sigma(t)'\big)^{-1}\sigma(t) = \mathbb{I}_m$ and thus

$$\begin{aligned} &[(-\upsilon(t)\sqrt{1 - \rho(t)'\rho(t)})^2 + \delta(t)'\delta(t) \\ &\quad - \delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t)] \\ &= \big[-\upsilon(t)0 + \delta(t)'\delta(t) - \delta(t)'\delta(t)\big] = 0. \end{aligned}$$

This is the case when the financial market is complete and the risk from the liability can be completely hedged by trading the assets. Xie et al. (2008) give numerical examples and show how the incompleteness of the market and the introduction of a liability influence the efficient frontier. They also show the influence of varying parameters $T, x, g(t), \upsilon(t), \rho(t)$ on the efficient frontier and the optimal strategy. Here, particularly varying the diffusion $\upsilon(t)$ of the liability is relevant.

5.3. Application to Hawkes risk model

We apply the results of Xie et al. (2008) to see how the mean–variance efficient frontier and the optimal strategy of an insurance company change when claims arrive according to a



(a) $\rho = (0.65, 0.1)$. (b) $\rho = (0, 0)$.

Fig. 11. Comparison of mean–variance efficient frontiers for parameters given in Tables 5 and 7. This is consistent with the observation in Figure 1e of Xie et al. (2008) that increasing the volatility of the liability will shift the frontier away from the vertical axis. However, notice that in the case $\rho \neq (0, 0)$ the steepness of the frontier increases such that the frontiers intersect at a certain variance level. This implies that for an investor that is willing to accept a very high level of risk, the Hawkes process case implies higher attainable expected returns.

Table 5
Asset Parameters, taken from (Xie et al., 2008). The initial wealth x was adapted to fit this example.

Parameter	n	m	T	x	r	b	$\sigma = (\sigma^1, \sigma^2)$	$\rho = (\rho^1, \rho^2)'$
Value	2	1	1	10	0.06	0.12	(0.15, 0.25)	(0.65, 0.10)′

Table 6
Comparison of the expected number and variance of arrivals on an interval of length t for a Poisson and an exponential Hawkes process. Formulas for the Hawkes process case are taken from (Da Fonseca and Zaatour, 2013). Note that for the Poisson case $\mathbb{E}[N(t)] = \mathbb{V}ar[N(t)]$ whereas for the Hawkes case it can be easily shown that $\mathbb{E}[N(t)] < \mathbb{V}ar[N(t)]$ for any $t > 0$.

Process	Parameters	$\mathbb{E}[N(t)]$	$\mathbb{V}ar[N(t)]$
Poisson	λ_P	$\lambda_P t$	$\lambda_P t$
Hawkes	λ, α, β	$\frac{\lambda}{1-\alpha/\beta} t$	$\frac{\lambda}{1-\alpha/\beta}(t(\frac{1}{1-\alpha/\beta})^2 + (1 - (\frac{1}{1-\alpha/\beta})^2)\frac{1-e^{-t(\beta-\alpha)}}{\beta-\alpha})$

Hawkes process[5] instead of a homogeneous Poisson process. For the assets, we use the parameters from the numerical example in Xie et al. (2008) which are summarized in Table 5. We use the pure diffusion approximation from (43) and treat the negative risk process of the company as the liability, i.e. we set

$$\begin{aligned} L(t) &:= -R(t) \approx -u + \left(\frac{\lambda}{1-\alpha/\beta} a^* - c\right) t + \bar{\sigma} W(t) \\ &= -u - \theta a^* \frac{\lambda}{1-\alpha/\beta} t + \bar{\sigma} W(t) \end{aligned} \tag{68}$$

where we use that $W(t)$ is a standard Brownian motion and therefore has the same distribution as $-W(t)$. Keeping the previous notation, this corresponds to $l = -u$, $g(t) \equiv -\theta a^* \frac{\lambda}{1-\alpha/\beta}$ and $\upsilon(t) \equiv \bar{\sigma}$. For the sake of presentation and comparability, we will now assume that incoming claim sizes are i.i.d. with $a^* = \mathbb{E}[X_1] =: m_1 = 0.5$, $\mathbb{V}ar[X_1] = 0.5$ and $\mathbb{E}[X_1^2] =: m_2 = 1$. We compare claim arrival processes with the same expected number of arrivals on any interval in order to analyse the difference in risk arising from the variance of the number of arrivals. Table 6 gives an overview of these quantities for Poisson and Hawkes processes. Table 7 gives the parameter values chosen for our numerical example where we compare a Poisson and two Hawkes processes.

In Fig. 11 we plot the mean–variance efficient frontiers for all three cases using (67). We recognize that changing the Poisson process to a Hawkes process with increasing presence of clustering will shift the frontier away from the vertical axis and thus lead to a higher risk for the same expected return level. Table 8 shows the values of $\mathcal{D}_1$ (expected terminal wealth of minimum-variance portfolio) and $\mathcal{D}_2$ (minimum attainable variance), indicating the higher risk introduced by the Hawkes process. However, the frontier becomes "steeper" and, above an intersection point at a certain level of risk, higher returns can be attained (this is due to the correlation of the liability and the asset processes — if we choose $\rho = (0, 0)$, the shift is strictly to the lower right as can be seen in Fig. 11). As effectively we increase the diffusion of the liability while leaving other parameters untouched, these results are consistent with Xie et al. (2008), Figure 1e.

Optimal strategy for given expected return level

In the following, we change the notation from $\varphi(X(t))$ (used in the previous section to emphasize dependence on the state variable $X(t)$) to $\varphi(t)$, as we like to think of the strategy as dynamically evolving over time. As in Xie et al. (2008), first assume the company would like to obtain a certain expected return level, say 15%, i.e. $\mathbb{E}[X^*(T)] = x \cdot 1.15 = 11.5$. Table 9 shows the minimum attainable variance of terminal wealth and the optimal strategy at the initial time 0 for all three cases. As could be expected from Fig. 11, in the case of the Hawkes processes, a higher variance must be accepted in order to attain the desired level of expected return, which entails the need for a higher initial investment in the risky asset. For the same expected return level, we plot a realization of the optimal investment in the risky asset $\varphi_1^*(t)$ over time in Fig. 12. To this end, standard Brownian motions $W_0(t)$ and $W(t) = (W_1(t), W_2(t))'$ are simulated over $[0, T] = [0, 1]$ on a grid of step size 0.001, and the evolution of wealth $X(t)$ and the optimal strategy $\varphi^*(t)$ are calculated iteratively using Eqs. (58) and (62) respectively.

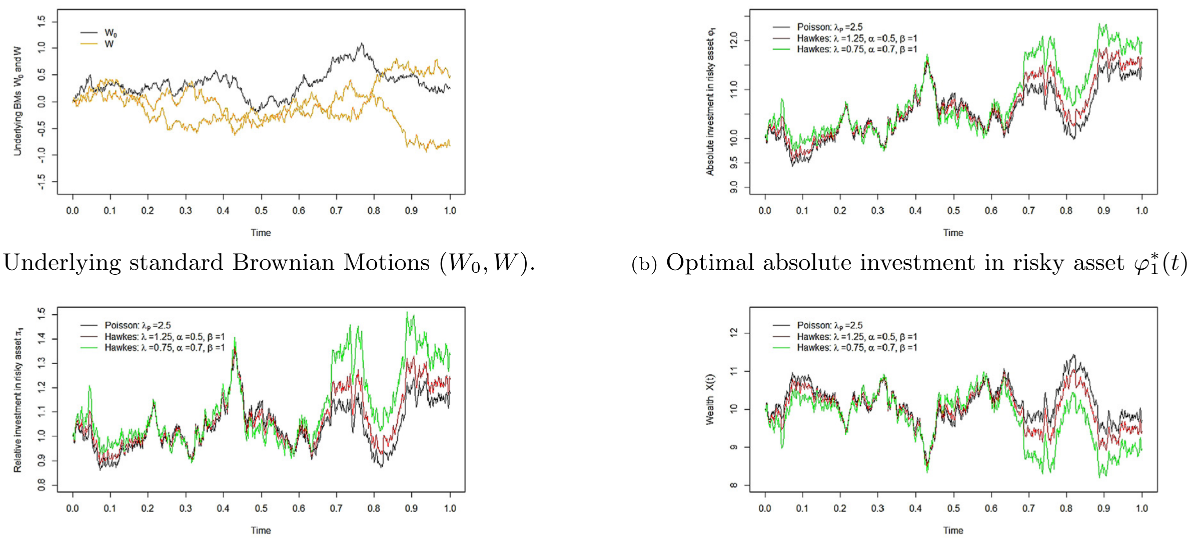
Optimal strategy for given maximum level of risk

Usually an insurance company's main interest and responsibility is to limit the risk it is subjected to. So now we want to study attainable return levels and corresponding optimal strategies for a given level of accepted risk. In the first step, we set the maximum level of risk (measured by variance of terminal wealth) to $\bar{v} = 8$ and list in Table 10 the corresponding attainable return levels and corresponding optimal strategies for all three cases. We use the same pure diffusion approximation as above with parameters given in Table 7.

It has to be kept in mind that as the computation of the mean–variance efficient frontier and the optimal strategies rely on the pure diffusion approximation of the risk process, the approximation error might be large. Indeed, as we have seen in Section 4, the approximation tends to *overestimate* the standard deviation of the risk process and thus the risk assigned in the Hawkes case might tend to be exaggerated. Thus in order to get to a more accurate estimation, we use our knowledge about the jump–diffusion approximation. By (38) it is given by

$$\begin{aligned} \hat{L}(t) &:= -R(t) \approx -u - ct + \hat{\sigma} W(t) + a^* N(t) \\ &= -u - (1+\theta) m_1 \frac{\lambda}{1-\alpha/\beta} t + \hat{\sigma} W(t) + m_1 N(t) \\ &=: -u + \hat{g} t + \hat{\sigma} W(t) + m_1 N(t) \end{aligned} \tag{69}$$

[5] Note again that the analysis is restricted to Hawkes processes with exponentially decaying intensity due to their unique analytical tractability.



(a) Underlying standard Brownian Motions (W_0, W).

(b) Optimal absolute investment in risky asset $\varphi_1^*(t)$.

(c) Optimal relative investment in risky asset $\pi_1^*(t)$.

(d) Wealth $X^*(t)$ given optimal strategy.

Fig. 12. One realization of the optimal investment in the risky asset until the final time $T = 1$, for $\rho = (0.65, 0.1)$. Standard Brownian motions $W_0(t)$ and $W(t) = (W_1(t), W_2(t))'$ were simulated over $[0, T] = [0, 1]$ on a grid of step size 0.001, and the evolution of wealth $X^*(t)$ and the optimal strategy $\varphi_1^*(t)$ calculated iteratively using Eqs. (58) and (62). One can observe that in Figs. 12(b), 12(c), 12(d), the fluctuations in the absolute/relative investment in the risky asset and the wealth under the optimal strategy are the larger, the *riskier* (by our measure the larger the variance of the number of jumps) the chosen point process is.

Table 7
Liability parameters, where the safety loading for the premium calculation is $\theta = 0.2$. Parameters are chosen such that $\mathbb{E}[N(1)]$ and $g(t)$ are equal for all cases as we would only like to study the difference arising from the change in variance. The second Hawkes process has a higher share of endogenous events α/β, which corresponds to stronger influence of self-excitement and clustering. The background rate λ is adjusted accordingly to keep the expected number of arrivals constant.

Process	(λ, α, β)	$\mathbb{E}[N(1)]$	$\mathbb{V}ar[N(1)]$	$g(t) \equiv -\theta m_1 \mathbb{E}[N(1)]$	$\upsilon(t) \equiv \bar{\sigma}$	$\mathbb{V}ar[L(T)]$
Poisson	$(2.5, -, -)$	2.5	2.5000	−0.25	1.3693	1.8750
Hawkes 1	$(1.25, 0.5, 1)$	2.5	4.0980	−0.25	1.9365	3.4799
Hawkes 2	$(0.75, 0.7, 1)$	2.5	5.9393	−0.25	2.8626	8.1944

Table 8
Values of $\mathcal{D}_1$ and $\mathcal{D}_2$ for the three processes and two different choices of ρ. Note that $\mathcal{D}_1$ corresponds to the expected terminal wealth of the minimum-variance portfolio and $\mathcal{D}_2$ to its variance. As shown in Xie et al. (2008), $\mathcal{D}_2 > 0$ unless the market is complete and the liability is completely hedgeable by the tradeable assets. Naturally, $\mathcal{D}_2$ increases for the Hawkes process cases.

Process	$\mathcal{D}_1$	$\mathcal{D}_2$	$\mathcal{D}_1$	$\mathcal{D}_2$
	$\rho = (0.65, 0.1)$		$\rho = (0, 0)$	
Poisson	10.9980	1.6055	10.8760	1.9497
Hawkes 1	11.0486	3.2110	10.8760	3.8994
Hawkes 2	11.1311	7.0167	10.8760	8.5210

Table 9
Optimal strategies $\varphi^*(t)$ at $t = 0, X(t) = X_0 = x$ from (62) for a given expected return level of 15%, thus $\mathbb{E}[X^*(T)] = 11.5$. Note that in the Hawkes cases, more is initially invested in the risky asset in the case $\rho = (0.65, 0.1)$. This is in accordance with the findings of Xie et al. (2008), Fig. 3b, that the initial investment in the risky asset must increase when the volatility of the liability increases (c.p.) to attain the same expected return level. For the case $\rho = (0, 0)$, the strategy does not depend on the choice of jump process (as can be seen from Eqs. (63) and (62) respectively, ζ and therefore φ^* do not depend on υ), but the attainable variance increases in the Hawkes cases. This corresponds to the observation in Fig. 11 that for attaining this level of $\mathbb{E}[X^*(T)]$, a higher risk $\mathbb{V}ar[X^*(T)]$ must be accepted.

Process	$\mathbb{V}ar[X^*(T)]$	$(\varphi_0^*(0), \varphi_1^*(0))$	$\mathbb{V}ar[X^*(T)]$	$(\varphi_0^*(0), \varphi_1^*(0))$
	$\rho = (0.65, 0.1)$		$\rho = (0, 0)$	
Poisson	7.4294	$(-0.0201, 10.0201)$	10.9495	$(-0.0029, 10.0029)$
Hawkes 1	7.9210	$(-0.0272, 10.0272)$	12.8992	$(-0.0029, 10.0029)$
Hawkes 2	10.1618	$(-0.0389, 10.0389)$	17.5208	$(-0.0029, 10.0029)$

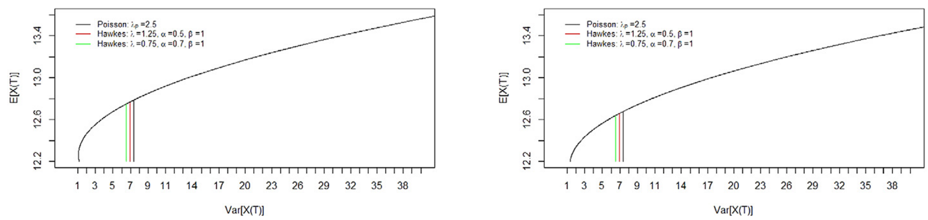
where $N(t)$ is the number of jumps of a Hawkes process with parameters (λ, α, β) on the interval $(0, t]$. This corresponds to

Table 10
Optimal strategies $\varphi^*(t)$ at $t = 0, X(t) = X_0 = x$ from (62) for a given maximum variance level of $\bar{v} = 8$. In the Hawkes case, when restricted to the same level of variance, the initial investment in the risky asset must be lowered and a lower expected return level $\mathbb{E}[X^*(T)]$ can be attained. For the second Hawkes case, in the case $\rho = (0.65, 0.1)$ part of the initial wealth would be invested in the riskfree asset instead of short-selling it to allow a higher initial investment in the risky asset. In the case $\rho = (0, 0)$, the variance restriction of $\bar{v} = 8$ would even not be attainable for the more volatile Hawkes process.

Process	$\mathbb{E}[X^*(T)]$	$(\varphi_0^*(0), \varphi_1^*(0))$	$\mathbb{E}[X^*(T)]$	$(\varphi_0^*(0), \varphi_1^*(0))$
	$\rho = (0.65, 0.1)$		$\rho = (0, 0)$	
Poisson	11.5240	$(-0.4051, 10.4051)$	11.3876	$(1.7984, 8.2016)$
Hawkes 1	11.5038	$(-0.0877, 10.0877)$	11.2972	$(3.2480, 6.7520)$
Hawkes 2	11.3374	$(2.5680, 7.4320)$	–	–

$\hat{l} = -u$, $\hat{g}(t) \equiv -(1+\theta)m_1 \frac{\lambda}{1-\alpha/\beta}$ and $\hat{\upsilon}(t) \equiv \hat{\sigma}$, where the values in this case are given in Table 11. Representing the liability process by $\hat{L}(t)$ instead of $L(t)$, i.e. splitting the random part into a continuous part (governed by the evolution of the Brownian motion $W(t)$) and a jump part (governed by the evolution of a Hawkes process $N(t)$), allows to explicitly distinguish between two sources of randomness: one part of the liability stemming from a source of financial risk correlated with the asset processes (represented by a Brownian motion), and the other part coming from a source of pure insurance risk (namely a jump process representing incoming insurance claims).

As naturally Theorem 6 is only applicable for a liability following a Brownian motion with drift (without jumps), we need an



(a) $\rho = (0.65, 0.1)$. (b) $\rho = (0, 0)$.

Fig. 13. Mean–variance efficient frontier for parameters given in Tables 5 and 11. Note that there is only one frontier in each case as the difference lies in the jump part which is not depicted in the frontier, but affects the variance boundary. Due to the higher variance of the number of jumps of the Hawkes processes, the boundary has to be shifted further to the left which implies a lower attainable expected terminal wealth.

Table 11
Liability parameters for the jump–diffusion approximation, where the safety loading for the premium calculation is $\theta = 0.2$. Note that $\hat{\upsilon}$ does not differ for the three cases, instead the difference in variance of the liability process stems from the added jump process $N(t)$. Furthermore, the last column indicates a smaller difference in variance of the liability process than assumed by the pure diffusion approximation (last column of Table 7).

Process	(λ, α, β)	$\mathbb{E}[N(1)]$	$\mathbb{V}ar[N(1)]$	$\hat{g}(t) \equiv -(1+\theta)m_1\mathbb{E}[N(1)]$	$\hat{\upsilon}(t) \equiv \hat{\sigma}$	$\mathbb{V}ar[\hat{L}(T)]$
Poisson	$(2.5, -, -)$	2.5	2.5	−1.5	1.1180	1.8750
Hawkes 1	$(1.25, 0.5, 1)$	2.5	4.0980	−1.5	1.1180	2.2744
Hawkes 2	$(0.75, 0.7, 1)$	2.5	5.9393	−1.5	1.1180	2.7347

approximative approach in order to use $\hat{L}(t)$ instead of $L(t)$ for the liability process. In order to find an *approximative optimal strategy* for a given level of accepted risk $\bar{\upsilon}$, we initially omit the jump part of $\hat{L}(t)$ and use Theorem 6 with the liability process given by

$$\widetilde{\hat{L}}(t) = -u + \hat{g}t + \hat{\upsilon}\rho' W(t) + \hat{\upsilon}\sqrt{1-\rho'\rho}W_0(t)$$

where $\hat{g}$ and $\hat{\upsilon}$ are given in Table 11. This implies that the jumps are considered to be a part of the liability which is not taken into consideration when calculating the optimal strategy. This is an assumption that could be interpreted in practice as a part of the liability that is not hedgeable by trading the assets in the market and therefore has to be taken into account through separate capital acting as a *risk buffer*. This means that in order to take into account the additional variance introduced by the jumps, the original variance boundary has to be adjusted according to

$$\begin{aligned}\mathbb{V}ar[X^*(T)] &= \mathbb{V}ar[A^*(T) - \hat{L}(T)] \\ &= \mathbb{V}ar[A^*(T) - (\widetilde{\hat{L}}(T) + m_1 N(T))] \\ &\overset{\text{jumps indep.}}{=} \mathbb{V}ar[A^*(T) - \widetilde{\hat{L}}(T)] + m_1^2\mathbb{V}ar[N(T)] \overset{!}{=} \bar{\upsilon} \\ &\iff \mathbb{V}ar[A^*(T) - \widetilde{\hat{L}}(T)] \overset{!}{=} \bar{\upsilon} - m_1^2\mathbb{V}ar[N(T)] =: \hat{\bar{\upsilon}}\end{aligned}$$

where $A^*(T)$ denotes the terminal value of the *asset part under the optimal strategy*. The efficient frontier in this case (note that it is the same for all three processes as the difference lies in the jump part that is not depicted in the frontier) for both choices of ρ is shown in Fig. 13 together with the shifted variance boundaries.

When comparing the result, in particular the attainable expected terminal wealth, to the one from the pure diffusion approximation in Table 10, the expectation of the *jump part*, that is $m_1\mathbb{E}[N(T)] = 1.25$ has to be subtracted from the estimate given by the mean–variance frontier obtained using $\widetilde{\hat{L}}(t)$. The results for parameters in Table 11 are given in Table 12. In order to corroborate that this approach is feasible, we set the number of simulation runs to $K = 10000$ and set the variance boundary for each case to $\hat{\bar{\upsilon}}$. We simulate the continuous part of $\hat{L}(t)$ on a grid with step size $\delta = 0.001$ over $[0, T] = [0, 1]$ according to

$$\widetilde{\hat{L}}(t) = -u + \hat{g}t + \hat{\upsilon}\rho' W(t) + \hat{\upsilon}\sqrt{1-\rho'\rho}W_0(t),$$

Table 12
For two choices of ρ we give the attainable expected terminal wealth (corrected for jumps) and the corresponding optimal strategy at time $t = 0$ for an original variance boundary of $\bar{\upsilon} = 8$, where we correct for the variance introduced by the jump part of the liability which is not included in the calculation of the optimal strategy. We observe that the approach still indicates a lower attainable return and a lower initial investment in the risky asset in the Hawkes cases. However, the differences are not as extreme as in Table 10, in particular the variance restriction is now attainable by the second Hawkes process in case $\rho = (0, 0)$. This comparison between the processes is more realistic as it eliminates the large overestimation of risk of the pure diffusion approximation in the Hawkes cases.

Process	$m_1^2\mathbb{V}ar[N(T)]$	$\hat{\bar{\upsilon}}$	$\mathbb{E}[X^*(T)](-1.25)$	$\varphi^*(0)$
Poisson	0.625	7.375	12.7862 (11.5362)	(0.0164, 9.9836)
Hawkes 1	1.0245	6.9755	12.7694 (11.5194)	(0.2860, 9.7140)
Hawkes 2	1.4848	6.5152	12.7493 (11.4993)	(0.6083, 9.3917)
(a) $\rho = (0.65, 0.1)$				
Process	$m_1^2\mathbb{V}ar[N(T)]$	$\hat{\bar{\upsilon}}$	$\mathbb{E}[X^*(T)](-1.25)$	$\varphi^*(0)$
Poisson	0.625	7.375	12.6770 (11.4270)	(1.7815, 8.2185)
Hawkes 1	1.0245	6.9755	12.6598 (11.4098)	(2.0564, 7.9436)
Hawkes 2	1.4848	6.5152	12.6393 (11.3893)	(2.3853, 7.6147)
(b) $\rho = (0, 0)$				

then calculate $X^*(t)$ and $\varphi_1^*(t)$ iteratively using (58) and Theorem 6 respectively and set $\varphi_0^*(t) = X^*(t) - \varphi_1^*(t)$. Over all simulation runs, we calculate empirical mean and variance of $X^*(T)$ and compare them with the theoretical values in Table 12. We then simulate $\hat{L}(t)$ including jumps as

$$\hat{L}(t) = -u + \hat{g}t + \hat{\upsilon}\rho' W(t) + \hat{\upsilon}\sqrt{1-\rho'\rho}W_0(t) + m_1 N(t)$$

where for each run, the same realizations of $W(t)$ and $W_0(t)$ as above are used and $N(t)$ is the number of jumps of a Hawkes (or Poisson) process on $(0, t]$.[6] For the simulation of the jumps it is worth noting that instead of simulating a point process on $[0, T]$

[6] All simulations are conducted using the statistical software R. For the Brownian motions, the increments on a grid as described are drawn as i.i.d. Normal r.v.; for the Hawkes processes, the simulation routine based on Ogata's modified thinning algorithm (Ogata, 1981) from the *hawkes* package is used.

for each of the K simulations, we simulate one process on $[0, KT]$ and map the realization of jumps on the interval $((k-1)T, kT]$ to the interval $(0, T]$ to be used in the kth simulation. This ensures that the theoretical mean and variance of the number of jumps are met by the simulations and not negatively distorted over the relatively short time span [0, 1]. This could otherwise happen as a Hawkes process with an expected value of 2.5 jumps per time unit might simply not have enough time to develop any clustering. The optimal strategy $\varphi_1^*(t)$, i.e. the optimal amount invested in the risky asset calculated from Theorem 6, is kept fixed as calculated under $\widetilde{\hat{L}}(t)$ (thus $\hat{\varphi}_1^*(t) = \varphi_1^*(t)$), but the wealth process $\hat{X}^*(t)$ is calculated anew according to (58) with $\hat{L}(t)$ in place of $\widetilde{\hat{L}}(t)$ and the investment in the riskless asset is adjusted accordingly as $\hat{\varphi}_0^*(t) = \hat{X}^*(t) - \hat{\varphi}_1^*(t)$. The results of the simulation are given in Table 13, where we observe that the optimal strategy calculated by using the modified jump–diffusion approximation with a shifted variance boundary $\hat{\bar{v}}$ adheres to the original boundary $\bar{v}$ when jumps are included, i.e. the theoretical approach described above is feasible.

In this chapter we have dealt with the case of an insurance company that can invest in assets traded on the market, but is burdened by a liability process that continuously affects the available capital and cannot be completely hedged by trading the assets. This is relevant in practice as managing asset investments in a way that maximizes returns while assuring sustainable and responsible liability management is a crucial challenge for any insurer. The liability process represents the risk process, i.e. the obligations arising from having to pay claims which arrive randomly (in size and number) over time. By approximating the risk process as studied in the previous section by either a jump–diffusion or a pure diffusion process, the application of results by Xie et al. (2008) to study the mean–variance efficient frontier and the optimal dynamic investment strategy (i.e. a scheme to decide which portion of the available wealth to invest in a riskless asset (typically thought of as a bank account) and a risky asset (e.g. a stock index) respectively) under constraints on either the expected terminal wealth or – more relevant in practice in this case – the variance of the terminal wealth (in this framework used to measure overall risk). We have highlighted the economic implications of operating under a claim process where claims tend to display clustering as opposed to claims occurring independently over time by substituting a Poisson claims process by two examples of self-exciting Hawkes processes. For a given level of accepted risk, although the expected claim number and size over the observed time horizon is identical, the clustering characteristics of the Hawkes process entails a higher risk which implies a lower *risk allowance* in the asset investment strategy and thus lower attainable returns. Therefore, if an insurer observes or suspects the risk process of his portfolio (or a sub-portfolio for a certain line of business) to display clustering (or generally overdispersion), he needs to be aware that in general, this affects the volatility of the liability process in such a way that for a given risk boundary, his investment strategy needs to be chosen more conservatively, meaning that a higher portion of the overall wealth must be invested in the riskless asset and thus a lower expected return is attainable. In the case where asset and liability processes are assumed uncorrelated (e.g. the liability process is assumed to represent pure insurance risk), this observation holds regardless of the chosen level of risk (recall the shift of the mean–variance efficient frontier to the lower right in the right panel of Fig. 11). The set-up used in this work furthermore allows for the asset and liability processes to be correlated. In this case, the frontier is shifted such that the frontiers intersect (recall the left panel of Fig. 11), however, the above results still hold for all but unreasonably high levels of allowed risk. We have furthermore studied the case where the risk process is approximated by a jump–diffusion process. As we have seen in Section 4, this is often the more accurate approximation, and it further allows to distinguish the random part of the liability process into a Brownian motion part correlated with the asset evolution and a pure jump part independent from it. The jump part can be interpreted as a (pure insurance) risk that is not hedgeable by trading the risky asset and therefore needs an a priori risk buffer assigned. The riskier the jump process (in our case, the stronger the clustering it displays), the more the variance boundary needs to be adjusted downwards before solving the optimal investment problem under the remaining continuous liability process. When applying the resulting optimal investment strategy under the original liability process (including jumps), the original variance boundary is adhered to. Our analysis emphasizes the insurer's need to not only estimate the expected future number and size of claims accurately, but also their temporal distribution over the observed period, i.e. the presence of clustering, in order to make sure given *risk allowances* are not breached undeliberately. To this end, the analysis of the presence of clustering in the claims process as well as the potential correlation of (part of) the insurer's liability with the tradeable assets need to be taken into account when deciding on how conservatively an optimal investment strategy must be chosen.

6. Conclusion

In this paper, we have introduced a risk model with claim arrivals based on a general compound Hawkes process (RH) and shown that it is suitable to model empirical data from the class of legal expenses insurance. We have studied its theoretical properties (LLN and FCLT) and derived a pure diffusion approximation which allows the calculation of ruin probabilities and application of results from asset–liability management to study the influence of a Hawkes claim arrival process on optimal investment strategies for an insurer in an incomplete market.

Of course, the assumption of a one-dimensional Hawkes process with exponential intensity is only a first step. Thus future work could be devoted to applying Hawkes processes with different intensity functions (e.g. power law). In this work, we restricted our focus to modelling portfolios with claims from one subclass only. A promising generalization would be to use a multi-dimensional marked Hawkes process to study a portfolio of claims from different subclasses, their mutual influence and development over time. This is of particular interest to an insurance company for reservation purposes, as it could help classify claims according to their initial characteristics and thus better estimate the amount of capital to be reserved. With respect to the optimal investment problem, the mean–variance approach is of course only one possibility. As another objective for an insurer could be to find an optimal investment strategy given a limit on the ruin probability, e.g. the results of Browne (1995) could be used as a starting point to apply to a Hawkes model case next.

Acknowledgements

G. Zeller thanks the Department of Mathematics and Statistics, University of Calgary, for their hospitality during her stay in Calgary. A. Swishchuk wishes to acknowledge NSERC for partial support of this project. The authors would like to thank ERGO Insurance Group for providing the empirical data sets, their co-operation and support. The authors furthermore thank the two anonymous referees whose valuable feedback greatly helped to improve the presentation of this paper.

Table 13
For $K = 10000$ simulations, we give the original and shifted variance boundaries and the theoretically attainable expected terminal wealth $X^*(T)$ under the modified jump–diffusion liability $\tilde{\hat{L}}(t)$ for each case. Comparing these values with the empirical mean and variance boundary of $X^*(T)$, we observe that they are matched quite closely. Keeping the optimal investment in the risky asset fixed and calculating the wealth process $\hat{X}^*(T)$ under the jump–diffusion liability $\hat{L}(t)$, we observe that the original variance restriction $\bar{v} = 8$ is met. Note that in the case $\rho \neq (0, 0)$ some inaccuracy is observable due to the theoretical assumption of independent jumps.

Process	$\hat{\bar{v}}$	$\mathbb{E}[X^*(T)]$ (theor.)	Modified jump–diffusion $\tilde{\hat{L}}(t)$ $\mathbb{E}[X^*(T)]$	$\mathbb{V}ar[X^*(T)]$	Jump–diffusion $\hat{L}(t)$ $\mathbb{E}[\hat{X}^*(T)]$	$\mathbb{V}ar[\hat{X}^*(T)]$
Poisson	7.375	12.7862	12.7981	7.4314	11.5215	8.0951
Hawkes 1	6.9755	12.7694	12.7625	6.8574	11.4647	7.9095
Hawkes 2	6.5152	12.7493	12.7396	6.4614	11.4333	7.9706

(a) $\rho = (0.65, 0.1)$

Process	$\hat{\bar{v}}$	$\mathbb{E}[X^*(T)]$ (theor.)	Modified jump–diffusion $\tilde{\hat{L}}(t)$ $\mathbb{E}[X^*(T)]$	$\mathbb{V}ar[X^*(T)]$	Jump–diffusion $\hat{L}(t)$ $\mathbb{E}[\hat{X}^*(T)]$	$\mathbb{V}ar[\hat{X}^*(T)]$
Poisson	7.375	12.6770	12.7012	7.3738	11.4026	8.0085
Hawkes 1	6.9755	12.6598	12.6805	7.0197	11.3795	8.0556
Hawkes 2	6.5152	12.6393	12.5743	6.5046	11.3495	8.0080

(b) $\rho = (0, 0)$

Appendix

A.1. Explanatory note on data set choice

One challenge about working with the data set provided to us by the insurance company was its overall size with a high average number of claims per day. This is problematic due to the aforementioned problem of non-unique timestamps (granularity of arrival times is per day) which have to be artificially distributed over the arrival day in order to fit a simple point process. For a high number of arrival times per day, the majority of interarrival times would thus be generated artificially, not due to the actual arrival pattern, which renders the significance of fitting an arrival process potentially meaningless. The first approach to thin the portfolio was to filter for the year of the *claim occurrence*, e.g. only consider claims which occurred in the year 2010, as we have the full "reporting picture" for them. The problem with this approach is that it produces a fairly skewed overall picture with a high number of payments in the first one to two years after the occurrence and fewer payments afterwards. Naturally, the majority of claims is reported and (at least partly) settled within the first years after its occurrence. Fitting an arrival process to this dataset would lead to an estimation that essentially tries to unite several different time periods into one picture and while doing a fair job of this, falls short to capture either of them accurately. At this point, it could of course be considered to fit different years separately. However, one general numerical challenge for insurance data (again due to the high timestamp granularity) is the short overall time horizon (e.g. compared to financial data, where often millisecond time intervals over one trading day are considered). Thus, we by all means wanted to use a dataset that makes use of the whole time horizon provided to us by the empirical data without having to divide it. Thus, we decided to classify claims according to their *delay in reporting*, thus from each of the reporting years we would consider claims that had occurred a fixed number of years before their reporting. This way the claim payments include claims *occurred* in several different years which mostly avoids the skewed picture mentioned before. We found claims with a *three year delay* in reporting to provide a good overall number of claims.

A.2. Proof of Theorem 4

Proof. We know from Theorem 3 that

$$\lim_{n\to\infty} \frac{R(nt) - (cnt - a^*N(nt))}{\sqrt{n}} \overset{D}{=} \hat{\sigma} W(t)$$

where $\hat{\sigma} = \sigma^* \sqrt{\lambda/(1-\hat{\mu})}$ is defined in Theorem 3. Now let us replace in the above equation $N(nt)$ by its expected value $\frac{\lambda}{1-\hat{\mu}} nt$ and thus look at

$$\lim_{n\to\infty} \frac{R(nt) - (cnt - a^* \frac{\lambda}{1-\hat{\mu}} nt)}{\sqrt{n}}.$$

Now we add and subtract the term $a^*N(nt)$ which yields

$$\frac{R(nt) - (cnt - a^*N(nt))}{\sqrt{n}} + \frac{a^* \frac{\lambda}{1-\hat{\mu}} nt - a^*N(nt)}{\sqrt{n}}.$$

We know by Theorem 3 that the first term converges to $\hat{\sigma} W(t)$ as $n \to \infty$ with $\hat{\sigma}$ as above and $W(t)$ a standard Wiener process.

For the second term we apply the Central Limit Theorem for Hawkes processes (see Laub et al., 2015, Theorem 1) which yields

$$\lim_{n\to\infty} \frac{a^* \frac{\lambda}{1-\hat{\mu}} nt - a^*N(nt)}{\sqrt{n}} \overset{D}{=} a^* \sqrt{\frac{\lambda}{(1-\hat{\mu})^3}} \bar{W}(t)$$

where $\bar{W}(t)$ is a standard Wiener process independent of $W(t)$.

So for the sum of the two limits we obtain

$$\lim_{n\to\infty} \frac{R(nt) - (cnt - a^* \frac{\lambda}{1-\hat{\mu}} nt)}{\sqrt{n}} = \sqrt{\hat{\sigma}^2 + (a^*)^2 \frac{\lambda}{(1-\hat{\mu})^3}} \hat{W}(t)$$

where $\hat{W}(t)$ is a standard Wiener process independent of $W(t)$ and $\bar{W}(t)$ and a^* and $\hat{\sigma}$ are defined in Theorem 3.

References

Aït-Sahalia, Y., Cacho-Diaz, J., Laeven, R.J., 2015. Modeling financial contagion using mutually exciting jump processes. J. Financ. Econom. 117 (3), 585–606. http://dx.doi.org/10.1016/j.jfineco.2015.03.002.

Albrecher, H., Asmussen, S., 2006. Ruin probabilities and aggregrate claims distributions for shot noise cox processes. Scand. Actuar. J. 2006 (2), 86–110. http://dx.doi.org/10.1080/03461230600630395.

Albrecher, H., Boxma, O.J., 2004. A ruin model with dependence between claim sizes and claim intervals. Insurance Math. Econom. 35 (2), 245–254. http://dx.doi.org/10.1016/j.insmatheco.2003.09.009.

Albrecher, H., Teugels, J.L., 2006. Exponential behavior in the presence of dependence in risk theory. J. Appl. Probab. 43 (1), 257–273. http://dx.doi.org/10.1239/jap/1143936258.

Asmussen, S., Albrecher, H., 2010. Ruin probabilities (2nd edition). Advanced Series On Statistical Science And Applied Probability, World Scientific Publishing Company.

Bacry, E., Delattre, S., Hoffmann, M., Muzy, J.F., 2013. Modelling microstructure noise with mutually exciting point processes. Quant. Finance 13 (1), 65–77. http://dx.doi.org/10.1080/14697688.2011.647054.

Bacry, E., Mastromatteo, I., Muzy, J.-F., 2015. Hawkes processes in finance. Available on arXiv: https://arxiv.org/abs/1502.04592..

Basu, R., 2016. Diffusion approximations for insurance risk processes. Stoch. Models 32 (1), 52–76. http://dx.doi.org/10.1080/15326349.2015.1083445.

Boudreault, M., Cossette, H., Landriault, D., Marceau, E., 2006. On a risk model with dependence between interclaim arrivals and claim sizes. Scand. Actuar. J. 2006 (5), 265–285. http://dx.doi.org/10.1080/03461230600992266.

Boumezoued, A., Devineau, L., 2017. Individual claims reserving: a survey. https://hal.archives-ouvertes.fr/hal-01643929, Preprint.

Brown, E.N., Barbieri, R., Ventura, V., Kass, R.E., Frank, L.M., 2002. The time-rescaling theorem and its application to neural spike train data analysis. Neural Comput. 14 (2), 325–346. http://dx.doi.org/10.1162/08997660252741149.

Browne, S., 1995. Optimal investment policies for a firm with a random risk process: Exponential utility and minimizing the probability of ruin. Math. Oper. Res. 20 (4), 937–958. http://dx.doi.org/10.1287/moor.20.4.937.

Chavez-Demoulin, V., McGill, J.A., 2012. High-frequency financial data modeling using Hawkes processes. J. Bank. Financ. 36 (12), 3415–3426. http://dx.doi.org/10.1016/j.jbankfin.2012.08.011.

Cheng, Z., Seol, Y., 2018. Gaussian approximation of a risk model with stationary Hawkes arrivals of claims. Available on arXiv: https://arxiv.org/abs/1801.07595.

Da Fonseca, J., Zaatour, R., 2013. Hawkes process: Fast calibration, application to trade clustering and diffusive limit. SSRN Electron. J. http://dx.doi.org/10.2139/ssrn.2294112.

Daley, D.J., Vere-Jones, D. (Eds.), 2003. An Introduction to the Theory of Point Processes. In: Probability and its Applications, Springer-Verlag, New York, http://dx.doi.org/10.1007/b97277.

Dassios, A., Jang, J.-W., 2003. Pricing of catastrophe reinsurance and derivatives using the cox process with shot noise intensity. Finance Stoch. 7 (1), 73–95. http://dx.doi.org/10.1007/s007800200079.

Dassios, A., Zhao, H., 2012. Ruin by dynamic contagion claims. Insurance Math. Econom. 51 (1), 93–106. http://dx.doi.org/10.1016/j.insmatheco.2012.03.006.

Dassios, A., Zhao, H., 2013. A risk model with delayed claims. J. Appl. Probab. 50 (03), 686–702. http://dx.doi.org/10.1239/jap/1378401230.

Embrechts, P., Liniger, T., Lin, L., 2011. Multivariate Hawkes processes: an application to financial data. J. Appl. Probab. 48 (A), 367–378. http://dx.doi.org/10.1239/jap/1318940477.

Embrechts, P., Schmidli, H., Grandell, J., 1993. Finite-time lundberg inequalities in the Cox case. Scand. Actuar. J. 1993 (1), 17–41. http://dx.doi.org/10.1080/03461238.1993.10413911.

Errais, E., Giesecke, K., Goldberg, L.R., 2010. Affine point processes and portfolio credit risk. SIAM J. Financial Math. 1 (1), 642–665. http://dx.doi.org/10.1137/090771272.

Fauth, A., Tudor, C.A., 2012. Modeling first line of an order book with multivariate marked point processes. Available on arXiv: http://arxiv.org/pdf/1211.4157v1.

Filimonov, V., Sornette, D., 2012. Quantifying reflexivity in financial markets: toward a prediction of flash crashes. Phys. Rev. E 85 (5 Pt 2), 056108. http://dx.doi.org/10.1103/PhysRevE.85.056108.

Filimonov, V., Sornette, D., 2015. Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. Quant. Finance 15 (8), 1293–1314. http://dx.doi.org/10.1080/14697688.2015.1032544.

Gao, X., Zhu, L., 2017. Limit theorems for Markovian hawkes processes with a large initial intensity. Available on arXiv: http://arxiv.org/pdf/1512.02155v3.

Grandell, J., 1977. A class of approximations of ruin probabilities. Scand. Actuar. J. 1977 (sup1), 37–52. http://dx.doi.org/10.1080/03461238.1977.10405071.

Hardiman, S.J., Bercot, N., Bouchaud, J.-P., 2013. Critical reflexivity in financial markets: a Hawkes process analysis. Eur. Phys. J. B 86 (10), 421. http://dx.doi.org/10.1140/epjb/e2013-40107-3.

Hawkes, A.G., 1971. Spectra of some self-exciting and mutually exciting point processes. Biometrika 58 (1), 83–90. http://dx.doi.org/10.1093/biomet/58.1.83.

Iglehart, L.D., 1969. Diffusion approximations in collective risk theory. J. Appl. Probab. 6 (02), 285–292. http://dx.doi.org/10.2307/3211999.

Jang, J., Dassios, A., 2013. A bivariate shot noise self-exciting process for insurance. Insurance Math. Econom. 53 (3), 524–532. http://dx.doi.org/10.1016/j.insmatheco.2013.08.003.

Karabash, D., Zhu, L., 2015. Limit theorems for marked Hawkes processes with application to a risk model. Stoch. Models 31 (3), 433–451. http://dx.doi.org/10.1080/15326349.2015.1024868.

Lallouache, M., Challet, D., 2016. The limits of statistical significance of hawkes processes fitted to financial data. Quant. Finance 16 (1), 1–11. http://dx.doi.org/10.1080/14697688.2015.1068442.

Laub, P.J., Taimre, T., Pollett, P.K., 2015. Hawkes processes. Available on arXiv: https://arxiv.org/abs/1507.02822.

Lewis, P.A.W., Shedler, G.S., 1979. Simulation of nonhomogeneous poisson processes by thinning. Nav. Res. Logist. Q. 26 (3), 403–413. http://dx.doi.org/10.1002/nav.3800260304.

Lundberg, F., 1903. I. Approximerad framstallning af sannolikhetsfunktionen: Ii. Aterforsakring af kollektivrisker. Almqvist Wiksell.

Norris, J.R., 2009. Markov Chains, 15. printing In: Cambridge Series on Statistical and Probabilistic Methods, vol. 2, Cambridge Univ. Press, Cambridge.

Ogata, Y., 1981. On Lewis' simulation method for point processes. IEEE Trans. Inform. Theory 27 (1), 23–31. http://dx.doi.org/10.1109/TIT.1981.1056305.

Ogata, Y., 1999. Seismicity analysis through point-process modeling: A review. In: Wyss, M., Shimazaki, K., Ito, A. (Eds.), Seismicity Patterns, their Statistical Significance and Physical Meaning. Birkhäuser Basel, Basel, pp. 471–507. http://dx.doi.org/10.1007/978-3-0348-8677-2{_}14.

Schmidli, H., 1994. Diffusion approximations for a risk process with the possibility of borrowing and investment. Commun. Stat. Stoch. Models 10 (2), 365–388. http://dx.doi.org/10.1080/15326349408807300.

Seal, H.L., 1983. The poisson process: Its failure in risk theory. Insurance Math. Econom. 2 (4), 287–288. http://dx.doi.org/10.1016/0167-6687(83)90027-6.

Skhorokhod, A.V., 2014. Studies in the Theory of Random Processes. Dover Publications, Newburyport.

Stabile, G., Torrisi, G.L., 2010. Risk processes with non-stationary Hawkes claims arrivals. Methodol. Comput. Appl. Probab. 12 (3), 415–429. http://dx.doi.org/10.1007/s11009-008-9110-6.

Swishchuk, A., 2000. Random Evolutions and their Applications: New Trends. Springer Netherlands, Dordrecht, http://dx.doi.org/10.1007/978-94-015-9598-8.

Swishchuk, A., 2017a. Risk model based on general compound Hawkes processes. Available on arXiv: https://arxiv.org/abs/1706.09038.

Swishchuk, A., 2017b. General compound Hawkes processes in limit order books. Available on arXiv: http://arxiv.org/pdf/1706.07459v2.

Swishchuk, A., Remillard, B., Elliott, R., Chavez-Casillas, J., 2019. Compound Hawkes processes in limit order books.. In: Financial Mathematics, Volatility and Covariance Modelling. Edited by: Julien Chevallier, Stéphane Goutte, David Guerreiro, Sophie Saglio and Bilel Sanhaji. Abingdon upon Thames: Routledge Advances in Applied Financial Econometrics. pp. 191–214.

Whitt, W., 1970. Weak convergence of probability measures on the function space $C[0, \infty)$. Ann. Math. Stat. 41 (3), 939–944. http://dx.doi.org/10.1214/aoms/1177696970.

Xie, S., Li, Z., Wang, S., 2008. Continuous-time portfolio selection with liability: Mean–variance model and stochastic LQ approach. Insurance Math. Econom. 42 (3), 943–953. http://dx.doi.org/10.1016/j.insmatheco.2007.10.014.

Yong, J., Zhou, X.Y., 1999. Stochastic Controls: Hamiltonian Systems and HJB Equations. In: Applications of Mathematics, vol. 43, Springer, New York, NY.

Yuen, K.C., Guo, J., Ng, K.W., 2005. On ultimate ruin in a delayed-claims risk model. J. Appl. Probab. 42 (01), 163–174. http://dx.doi.org/10.1239/jap/1110381378.

Zhang, C., 2016. Modeling high frequency data using hawkes processes with power-law kernels. Procedia Comput. Sci. 80, 762–771. http://dx.doi.org/10.1016/j.procs.2016.05.366.

Zheng, B., Roueff, F., Abergel, F., 2014. Modelling bid and ask prices using constrained Hawkes processes: Ergodicity and scaling limit. SIAM J. Financial Math. 5 (1), 99–136. http://dx.doi.org/10.1137/130912980.

Zhou, X.Y., 2000. Continuous-time mean-variance portfolio selection: A stochastic LQ framework. Appl. Math. Optim. 42 (1), 19–33. http://dx.doi.org/10.1007/s002450010003.

Zhu, L., 2013. Ruin probabilities for risk processes with non-stationary arrivals and subexponential claims. Insurance Math. Econom. 53 (3), 544–550. http://dx.doi.org/10.1016/j.insmatheco.2013.08.008.