

Technische Universität München

Department of Mathematics

Hawkes Processes in Insurance

Risk Modelling and Optimal Investment

Master Thesis

by

Gabriela Zeller

Supervisor (TUM):	Prof. Dr. Rudi Zagst
Supervisor (University of Calgary):	Prof. Dr. Anatoliy Swishchuk
Submission Date:	01.03.2019

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

Munich, 01.03.2019

Abstract

This master's thesis is about applications of self-exciting Hawkes processes in insurance mathematics. The Hawkes process framework allows to investigate an insurance risk model with an interesting type of dependence, namely a claim arrival process which displays endogenously caused temporal clustering (for instance caused by an outcoming claim payment that induces a stream of subsequent payments). We first extensively study properties of Hawkes processes, in particular the exponential specification which is the focus of this thesis. After reviewing results from classical risk theory and some of its extensions, we introduce a risk model based on a general compound Hawkes process. We derive its theoretical properties, in particular a law of large numbers and a central limit theorem which allows a pure diffusion approximation of the model and analytical calculation of finite-time and infinite-time ruin probabilities. Using empirical insurance data from the class of legal expenses insurance, we corroborate that the model is able to accurately reflect reality. Finally, we use the diffusion approximation of the model to study the optimal investment problem of an insurer whose claim arrival process displays temporal clustering using a mean-variance framework. We quantify the effect of the risk-enhancing clustering feature on the optimal investment strategy and attainable returns given a risk boundary using results for a holistic asset-liability management approach.

Zusammenfassung

Diese Masterarbeit beschäftigt sich mit der Anwendung von selbst-anregenden Hawkes Prozessen in der Versicherungsmathematik. Im Rahmen dieser Betrachtung lässt sich ein Risikomodell mit einer interessanten Art der Abhängigkeit, nämlich Schadenszahlungen, deren Ankunftszeiten einen endogen verursachten temporären Clustering-Effekt aufweisen (beispielsweise verursacht durch eine ausgehende Schadenszahlung, welche eine Reihe von folgenden Auszahlungen auslöst), studieren. Wir beschäftigen uns zunächst eingehend mit den Eigenschaften von Hawkes Prozessen, insbesondere der exponentiellen Spezifikation, welche den Fokus dieser Arbeit darstellt. Nachdem einige Resultate der klassischen Risikotheorie und einiger ihrer Erweiterungen rekapituliert werden, führen wir ein Risikomodell basierend auf einem allgemeinen zusammengesetzten Hawkes Prozess ein. Wir leiten seine theoretischen Eigenschaften, insbesondere ein Gesetz der großen Zahlen und einen zentralen Grenzwertsatz her, welcher eine Approximation des Modells durch einen reinen Diffusionsprozess und die analytische Berechnung von Ruinwahrscheinlichkeiten im endlichen und unendlichen Zeithorizont ermöglicht. Unter Verwendung empirischer Daten aus dem Bereich der Rechtsschutzversicherung bestätigen wir, dass das Modell in der Lage ist, den realen Sachverhalt angemessen abzubilden. Abschließend verwenden wir die obige Approximation des Modells, um das Investitions-Problem eines Versicherers, dessen Schadensankunftszeiten einen zeitlichen Clustering-Effekt aufweisen, im Rahmen der Portfoliotheorie nach Markowitz (Ertrags-Risiko-Effizienzlinie) zu untersuchen. Wir quantifizieren den Effekt der risikoerhöhenden Clustering-Eigenschaft auf die optimale Investmentstrategie sowie die erzielbare Rendite gegeben einer Risikolimitierung unter Verwendung eines Ansatzes des ganzheitlichen Bilanzstrukturmanagements.

Contents

List of symbols and abbreviations

a.e.	almost every(where)
a.s.	almost surely
càdlàg	right continuous with left limits (continue à droite, limite à gauche)
c.d.f.	cumulative distribution function
c.p.	ceteris paribus
(F)CLT	(functional) central limit theorem
i.i.d.	independent and identically distributed
LLN	law of large numbers
LQ	linear-quadratic
MLE	maximum likelihood estimate / estimation
MVE	mean-variance efficient
p.d.f.	probability distribution function
RMGCHP	risk model based on general compound Hawkes processes
r.p.	ruin probability
r.v.	random variable(s)
w.r.t.	with respect to
$\mathbb{N}_0$	Natural numbers including 0
$\mathbb{R}$	Real numbers
$D([0,\infty),\mathbb{R})$	Skorokhod space (collection of càdlàg functions mapping to $\mathbb{R}$) on $[0,\infty)$ equipped with Skorokhod's J_1 topology
$\mathcal{C}([0,T];\mathbb{R}^{n\times k})$	Continuous bounded deterministic functions on $[0,T]$ with values in $\mathbb{R}^{n\times k}$
$\mathcal{L}^2_{\mathcal{F}}([0,T];\mathbb{R}^m)$	$\mathbb{R}^m$-valued, progressively measurable and square integrable random variables on $[0,T]$ with $\mathbb{P}$-a.s. finite $\mathcal{L}^2_{\mathcal{F}}$-norm.

Chapter 1

Introduction

Ever since the introduction of the classical Cramer-Lundberg risk model in the early 20^{th} century, it has served as the theoretical basis for risk theory and it has been of interest to generalize it in ways that incorporate more of the challenges insurers face in the real world. In reality, insurance claims usually do not arrive evenly and independently over time, but display some sort of time-dependence or even clustering with periods of relative calm and short periods with a large number of occurrences. The causes for this phenomenon can be exogenous, for example seasonalities (e.g. wintery road conditions causing more car accidents) or occurrence of natural catastrophes (e.g. hurricane seasons), and have been studied using Cox (doubly stochastic Poisson) models (see e.g. Albrecher and Asmussen 2006, Dassios and Jang 2003). However, a maybe even more interesting type of dependence to study is endogenously caused clustering, where a claim occurrence itself triggers a stream of subsequent occurrences. This case can be investigated within the framework of self-exciting Hawkes processes, a versatile class of simple point processes that were introduced by Hawkes 1971 and have since found applications in various fields like seismology (Ogata 1999), neuroscience (Truccolo 2016) and finance (overview in Bacry, Mastromatteo, and Muzy 2015). The motivation for using Hawkes processes in insurance risk theory is manifold: For some classes of (re-)insurance claims (e.g. earthquake insurance or health insurance in regions subject to pandemics), the main interest is understanding the temporal clustering phenomenon in order to mitigate risk to the insurer's portfolio. This is gaining general importance in an increasingly interconnected, globalised world where risks more easily gain a systematic characteristic and where risks of natural catastrophes become more imminent due to climate change. In other cases, the main insight from using a Hawkes process may come from its interpretation as a branching process, where an initial claim occurrence triggers a stream of subsequent claim payments. It is of natural interest for an insurer to understand and estimate the temporal structure of these

payment streams, ideally from the moment of the first claim occurrence, for reservation purposes. While there are extensive amounts of research about Hawkes processes in finance, its applications to insurance problems are quite recent: The first to consider a risk model with Hawkes claims arrivals were Stabile and Torrisi 2010, who derive the asymptotic behavior of ruin probabilities and asymptotically efficient simulation laws assuming light-tailed claims. Their work was extended by Zhu 2013 who considered subexponential claims. Dassios and Zhao 2012 considered a risk process with claim arrivals following a dynamic contagion process, generalising the Hawkes process and the Cox process and thus including both self-excited and externally excited jumps. Jang and Dassios 2013 studied a bivariate shot noise self-exciting process for insurance, including a constant rate of exponential decay, which they show could be used for the modelling of discounted aggregate losses from catastrophic events. Very recently, Cheng and Seol 2018 derived a diffusion approximation and thus expressions for the ruin probabilities of a risk model with Hawkes claims arrivals. They found that the diffusion limit is a Gaussian process that can be decomposed into a centered Gaussian process and an independent Brownian motion. While all these works are theoretically very promising and provide extensions of classical theoretical results, they often only provide (asymptotic) estimates which are not necessarily easily applicable or understandable. Furthermore, to the best of our knowledge, there is no application of a Hawkes risk model to empirical insurance data to this date. Therefore, we build on the risk model based on a general compound Hawkes process introduced by Swishchuk 2017b, whose convenient theoretical properties include a law of large numbers, implying a net profit condition and premium principle, and a functional central limit theorem, allowing an approximation of the risk process by a jump diffusion process. We use these results to deduce a pure diffusion approximation by a simple Brownian motion with drift which allows the analytical calculation of ruin probabilities in closed form. As another application, we use this approximation to study the optimal investment problem of an insurer in an incomplete market whose claim arrival process follows a Hawkes process using a mean-variance framework and results from Xie, Li, and Wang 2008. We show how not only the expected number and size of claims, but also their temporal distribution and potential clustering has to be taken into account in an insurer's asset-liability management in order to adhere to given risk limitations. Furthermore, we check the applicability of the Hawkes risk model, the diffusion approximation and the ruin probability estimates on real empirical data from the class of legal expenses insurance.

The remainder of this thesis is structured as follows:

Chapter 2 summarizes some mathematical background knowledge which will be used throughout the thesis without further explanation. Chapter 3 provides an extensive overview of properties of Hawkes processes, in particular the exponential specification which is the focus of this thesis. The foundations of classical risk theory and its exten-

sions are studied in Chapter 4, before we combine Hawkes processes and risk models in Chapter 5. We first review some previous work (as mentioned above) and then introduce the risk model based on a general compound Hawkes process and its theoretical properties. Afterwards, the implementation of the model with empirical insurance data follows in Chapter 6. Lastly, the application to an optimal investment problem is studied in Chapter 7. The thesis is concluded by highlighting limitations and potential future research opportunities in Chapter 8.

Chapter 2

Preliminaries

In the following, we review some definitions and theorems which we assume the reader to be familiar with. They are well-known results and thus will be stated without proof here. More background information can be found in the recommendable books by Grimmett and Stirzaker 2006 (probability theory), Woess 2009 (Markov chains), Schmidli 2018 (risk theory) and Yong and Zhou 1999 (stochastic controls). For the scope of this thesis, the following section can act as a glossary for all basic definitions and results which appear without further explanation later on. As a prerequisite, we assume all random variables to be defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set, $\mathcal{F}$ a σ-algebra of subsets of Ω and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$.

Probability Theory

Definition 2.1 (Modes of convergence (Grimmett and Stirzaker 2006, p. 308))**.** Let $X, X_1, X_2, \cdots$ be random variables on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say

- $X_n \to X$ *almost surely*, written $X_n \overset{a.s.}{\to} X$, if the event $\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}$ has probability 1.
- $X_n \to X$ *in distribution*, written $X_n \overset{D}{\to} X$, if

$$F_n(x) := \mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x) =: F(x) \text{ as } n \to \infty$$

  for all points x at which the function $F(x)$ is continuous.

There are other modes of convergence (e.g. *in probability* or *in mean*) which we don't list here as they do not appear in this thesis.

Note that convergence in distribution is innately linked to Skorokhod spaces (sets of all càdlàg functions over a given domain) as any distribution function is by definition a càdlàg function. We will only use the space $D([0,\infty),\mathbb{R})$ here, i.e. the space of càdlàg functions on $\mathbb{R}_+$ equipped with Skorokhod's J_1 topology. For a background on this, see Skhorokhod 2014 or Billingsley 1999.

Definition 2.2 (Conditional probability (Grimmett and Stirzaker 2006, p. 9))**.** For two events A and B, if $\mathbb{P}(B) > 0$, then the conditional probability that A occurs given that B occurs is defined to be

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Theorem 2.3 (Law of total probability (Grimmett and Stirzaker 2006, p. 10))**.** Let $B_1, B_2, \cdots, B_n$ be a partition of Ω such that $\mathbb{P}(B_i) > 0$ for all i. Then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Theorem 2.4 (Probability integral transform (Angus 1994))**.** Let X be a random variable with distribution function F and quantile function F^{-1}. Then:

- Let $U \sim Unif(0,1)$. Then $F^{-1}(U) \overset{d}{=} X$ and
- $F(X) \overset{d}{=} U$ if and only if F is continuous,

where $\overset{d}{=}$ denotes equality in distribution.

Theorem 2.5 (Law of large numbers (Grimmett and Stirzaker 2006, p. 326))**.** Let $\{X_i\}$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_i] =: m_1$ and $\mathbb{E}[X_i^2] < \infty$ and let $Z_n := \sum_{i=1}^{n} X_n$. Then for $n \to \infty$,

$$\frac{Z_n}{n} \longrightarrow m_1$$

almost surely.

Theorem 2.6 (Central limit theorem (Grimmett and Stirzaker 2006, p. 194))**.** Let $\{X_i\}$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_i] =: m_1 < \infty$ and $\mathbb{V}ar[X_i] =: \sigma^2 \in (0,\infty)$, and let $Z_n := \sum_{i=1}^{n} X_n$. Then for $n \to \infty$,

$$\frac{Z_n - nm_1}{\sqrt{n\sigma^2}} \xrightarrow{D} Z$$

where Z is standard normally distributed.

Theorem 2.7 (Hahn's theorem / CLT for stochastic processes (Whitt 2002, p. 226f))**.** Let $\{X_i\} \equiv \{X_i(t) : t \geq 0\}$ be a sequence of i.i.d. random elements of $D \equiv D([0,\infty), \mathbb{R})$ with $\mathbb{E}[X_i(t)] = 0$ and $\mathbb{E}[X_i(t)^2] < \infty$ for all t. If for all $0 < T < \infty$ there exist continuous and non-decreasing real-valued functions g and f on $[0,T]$ and numbers $a > 1/2$ and $b > 1$ such that

$$\mathbb{E}[(X(u) - X(s))^2] \leq (g(u) - g(s))^a \tag{2.1}$$

and

$$\mathbb{E}[(X(u) - X(t))^2 (X(t) - X(s))^2] \leq (f(u) - f(t))^b \tag{2.2}$$

for all $0 \leq s \leq t \leq u \leq T$ with $u - s < 1$, then

$$Z_n(t) := \frac{\sum_{i=1}^{n} (X_i(t) - \mathbb{E}[X_i(t)])}{\sqrt{n}} \xrightarrow{D} Z$$

in D and Z is an a.s. continuous, mean-zero Gaussian process with the covariance function of X_1.

Theorem 2.8 (Wald equation (Wald 1944))**.** Let $\{X_i\}$ be a sequence of real-valued, i.i.d. random variables and let $N(t) \geq 0$ be an integer-valued r.v. independent of the sequence $\{X_i\}$. Suppose $\mathbb{E}[N(\cdot)] < \infty$ and $\mathbb{E}[X_i] < \infty$. Then

$$\mathbb{E}\left[\sum_{i=1}^{N(t)} X_i\right] = \mathbb{E}[X_1]\mathbb{E}[N(t)].$$

Theorem 2.9 (Law of total variance (Blitzstein and Hwang 2015, p. 401))**.** Let X and Y be random variables on the same probability space and assume $\mathbb{V}ar[Y] < \infty$. Then

$$\mathbb{V}ar[Y] = \mathbb{E}[\mathbb{V}ar(Y|X)] + \mathbb{V}ar(\mathbb{E}[Y|X]).$$

The last two results imply that if $\{X_i\}$ is a sequence of i.i.d. random variables and $N(t) \geq 0$ an integer-valued random variable independent of the sequence $\{X_i\}$, then it holds

$$\begin{aligned}
\mathbb{V}ar\left(\sum_{i=1}^{N(t)} X_i\right) =: \mathbb{V}ar(Y(t)) &= \mathbb{E}[\mathbb{V}ar(Y(t)|N(t)] + \mathbb{V}ar(\mathbb{E}[Y(t)|N(t)]) \\
&= \mathbb{E}[N(t)\mathbb{V}ar(X_1)] + \mathbb{V}ar(N(t)\mathbb{E}[X_1]) \\
&= \mathbb{V}ar(X_1)\mathbb{E}[N(t)] + \mathbb{E}[X_1]^2\mathbb{V}ar(N(t)).
\end{aligned}$$

Markov Chains

Definition 2.10 (Markov chain (Grimmett and Stirzaker 2006, p. 214))**.** Let $X = \{X_0, X_1, \cdots\}$ be a sequence of random variables which take values in some countable set S, called the *state space*. The process X is a *Markov chain* if it satisfies the *Markov condition*:

$$\mathbb{P}(X_n = x | X_0 = x_0, X_1 = x_1, \cdots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = x | X_{n-1} = x_{n-1})$$

for all $n \geq 1, x, x_0, \cdots, x_{n-1} \in S$. The chain X is called *homogeneous* if

$$\mathbb{P}(X_n = j | X_{n-1} = i) = \mathbb{P}(X_1 = j | X_0 = i) \quad \forall n, i, j.$$

The *initial distribution* ν is defined as the vector of probabilities of the initial state, i.e. $\nu_i = \mathbb{P}(X_0 = i)$ for all $i \in S$.

Definition 2.11 (Transition probabilities (Grimmett and Stirzaker 2006, p. 214f))**.** The *transition matrix* $P = (p_{ij})$ of a Markov Chain X is the $|S| \times |S|$ matrix of *transition probabilities*

$$p_{ij} = \mathbb{P}(X_n = j | X_{n-1} = i).$$

The *n-step transition matrix* $P(m, m+n)$ is the matrix of transition probabilities

$$p_{ij}(m, m+n) := \mathbb{P}(X_{m+n} = j | X_m = i).$$

By the Chapman-Kolmogorov equations it holds that $P(m, m+n) = P^n$.

Definition 2.12 (Characteristics of states (Grimmett and Stirzaker 2006, p. 220ff))**.**

- A state $i \in S$ is called *recurrent* if

  $$\mathbb{P}(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1,$$

  which means that the probability of eventually returning to state i, having started in state i, is 1.

- A recurrent state is called *positive recurrent*, if the mean recurrence time (the expected time it takes to return to state i) is finite.

- The *period* $d(i)$ of a state i is defined as the greatest common divisor of the number of steps after which returning to the same state i is possible, i.e. $d(i) = gcd\{n : p_{ii}(n) > 0\}$. A state $i \in S$ is called *aperiodic* if $d(i) = 1$ and *periodic* otherwise.

- A state i *communicates* with a state j ($i \to j$) if $p_{ij}(m) > 0$ for some $m \geq 0$, meaning that if the chain starts from i, it might visit state j with positive probability. If $i \to j$ and $j \to i$, we say i and j intercommunicate ($i \leftrightarrow j$).

Definition 2.13 (Irreducible Markov chain (Grimmett and Stirzaker 2006, p. 223ff))**.** A set $C \subset S$ is called *irreducible* if $i \leftrightarrow j$ for all $i, j, \in C$. If this holds for all $i, j \in S$, the whole Markov chain is called *irreducible*. This means that the chain, having started from any state $i \in S$, may ever visit any state $j \in S$ with positive probability.

Lemma 2.14 (Grimmett and Stirzaker 2006, p. 224)**.** If $i \leftrightarrow j$, then i is recurrent if and only if j is recurrent.

Lemma 2.15 (Grimmett and Stirzaker 2006, p. 225)**.** If S is finite, then at least one state is recurrent and all recurrent states are positive recurrent.

Definition 2.16 (Stationary distribution (Grimmett and Stirzaker 2006, p. 227))**.** The vector π^* is called a *stationary distribution* of a Markov chain if π^* has entries $(\pi_j^* : j \in S)$ such that:

- $\pi_j^* \geq 0$ for all j, and $\sum_{j \in S} \pi_j^* = 1$,
- $\pi^* = \pi^* P$, which means $\pi_j^* = \sum_{i \in S} \pi_i^* p_{ij}$ for all j.

Theorem 2.17 (Stationary distribution (Grimmett and Stirzaker 2006, p. 227))**.** An irreducible Markov Chain has a stationary distribution π^* if and only if all the states are positive recurrent. In this case, π^* is the unique stationary distribution.

Note that the results above imply that an irreducible Markov Chain on a finite state space S must automatically be positive recurrent, so an irreducible Markov chain on a finite state space always has a unique stationary distribution.

Theorem 2.18 (Ergodic theorem (Woess 2009, p. 69))**.** Let X be a positive recurrent, irreducible Markov chain with stationary distribution π^* and $f : X \to \mathbb{R}$ an integrable function w.r.t. π^*, that is $\sum\limits_{x \in S} |f(x)|\pi^*(x) < \infty$, then for any initial distribution it holds that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) = \sum_{x \in S} f(x)\pi^*(x) = \mathbb{E}_{\pi^*}[X_0] \text{ a.s.} \tag{2.3}$$

Therefore, a state $x \in S$ is called *ergodic* if it is positive recurrent and aperiodic and an irreducible Markov chain is called *ergodic* if all $x \in S$ are ergodic.

Definition 2.19 (Continuous-time Markov chain (Grimmett and Stirzaker 2006, p. 256))**.** Let $X = \{X(t) : t \geq 0\}$ be a family of random variables taking values in some countable state space S and indexed by the halfline $[0, \infty)$. The process X is called a *continuous-time Markov chain* if it satisfies the following Markov property:

$$\mathbb{P}(X(t_n) = x | X(t_1) = x_1, X(t_2) = x_2, \cdots, X(t_{n-1}) = x_{n-1}) = \mathbb{P}(X(t_n) = x | X(t_{n-1}) = x_{n-1})$$

for all $x, x_1, \cdots, x_{n-1} \in S$ and any sequence $t_1 < t_2 < \cdots < t_n$ of times.

Definition 2.20 (Markov process (Grimmett and Stirzaker 2006, p. 406))**.** The continuous-time process X, taking values in $\mathbb{R}$, is called a *Markov process* if it holds

$$\mathbb{P}(X(t_n) \leq x | X(t_1) = x_1, X(t_2) = x_2, \cdots, X(t_{n-1}) = x_{n-1}) = \mathbb{P}(X(t_n) \leq x | X(t_{n-1}) = x_{n-1})$$

for all $x, x_1, \cdots, x_{n-1}$ and any sequence $t_1 < t_2 < \cdots < t_n$ of times.

Stochastic processes and stochastic calculus

Definition 2.21 (Wiener Process / Brownian Motion (Grimmett and Stirzaker 2006, p. 516))**.** A stochastic process $W = \{W(t) : t \geq 0\}$ is called (standard) *Brownian motion* if it fulfills the following properties:

- $W(0) = 0$ almost surely,
- W has independent increments, i.e. the random variables $(W(t) - W(s))$ and $(W(v) - W(u))$ are independent for all $0 \leq u < v \leq s < t < \infty$,
- $W(s+t) - W(s) \sim N(0, \sigma^2 t)$ for all $s, t \geq 0$ where σ^2 is a positive constant,
- the paths of W are a.s. continuous.

We call the stochastic process $X(t) = \mu t + \sigma W(t)$ a Brownian motion with drift μ and volatility σ (or simply (μ, σ)-Brownian motion) for $\mu \in \mathbb{R}$ and $\sigma \in (0, \infty)$.

Definition 2.22 (Gaussian Process (Grimmett and Stirzaker 2006, p. 406))**.** A real-valued continuous-time process X is called a *Gaussian process* if each finite-dimensional vector $(X(t_1), X(t_2), \cdots, X(t_n))$ has the multivariate normal distribution $N(\mu(t), V(t))$ for some mean vector μ and some covariance matrix V which may depend on $t = (t_1, t_2, \cdots, t_n)$. A Gaussian process is called *centered* if its mean function $\mu(t)$ is identically zero.

Definition 2.23 (Stationarity (Grimmett and Stirzaker 2006, p. 361))**.** The process $X = \{X(t) : t \geq 0\}$, taking values in $\mathbb{R}$, is called *(strongly) stationary* if the families

$$\{X(t_1), X(t_2), \cdots, X(t_n)\} \text{ and } \{X(t_1 + h), X(t_2 + h), \cdots, X(t_n + h)\}$$

have the same joint distribution for all $t_1, t_2, \cdots, t_n$ and $h > 0$.

Theorem 2.24 (Itô formula (Grimmett and Stirzaker 2006, p. 545))**.** Let X be an *Itô diffusion*, i.e. an adapted process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ fulfilling the stochastic differential equation

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t)$$

where μ and σ are functions from $\mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and $W(t)$ is a Brownian motion. Let $Y(t) := f(t, X(t))$, where $f \in C^{1,2}$ is once resp. twice continuously differentiable in t resp. $X(t)$, then $Y(t)$ is also an Itô diffusion, given by

$$dY = \big[f_x(t, X)\mu(t, X) + f_t(t, X) + \frac{1}{2}f_{xx}(t, X)\sigma^2(t, X)\big]dt + f_x(t, X)\sigma(t, X)dW(t).$$

Corollary 2.25 (Product rule)**.** Let X and Y be two *Itô diffusions* fulfilling

$$dX(t) = \mu_X dt + \sigma_X dW(t) \text{ and}$$

$$dY(t) = \mu_Y dt + \sigma_Y dW(t)$$

respectively, where we suppress the arguments of μ and σ for notational convenience. Then the product $X(t)Y(t)$ is also an Itô diffusion, satisfying

$$\begin{aligned} d(X(t)Y(t)) &= X(t)dY(t) + Y(t)dX(t) + d\langle X, Y\rangle(t) \\ &= (Y(t)\mu_X + X(t)\mu_Y + \sigma_X\sigma_Y)dt + (Y(t)\sigma_X + X(t)\sigma_Y)dW(t). \end{aligned}$$

Theorem 2.26 (Multi-dimensional Itô formula)**.** Let $W = (W_1, \cdots, W_m)$ be an m-dimensional Brownian motion and $X = (X_1, \cdots, X_n)$ be an n-dimensional Itô process with

$$dX_i(t) = \mu_i(t, X(t))dt + \sum_{j=1}^{m} \sigma_{ij}(t, X(t))dW_j(t), \text{ for } i = 1, \cdots, n.$$

Let $Y(t) := f(t, X(t))$, where $f \in C^{1,2}$ is once resp. twice continuously differentiable in t resp. $X(t)$. Then $Y(t)$ is also an Itô diffusion, given by

$$dY = df(t, X(t)) = f_t(t, X(t))dt + \sum_{i=1}^{n} f_{X_i}dX_i(t) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} f_{X_iX_j}d\langle X_i, X_j\rangle(t),$$

where

$$d\langle X_i, X_j\rangle(t) = \sum_{k=1}^{m} \sigma_{ik}(t, X)\sigma_{jk}(t, X)dt.$$

Theorem 2.27 (Itô formula for jump diffusion processes (Privault 2019, p. 577))**.** Let X be an adapted process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ fulfilling the stochastic differential equation

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t) + \eta(t, X(t))dN(t)$$

where μ, σ and η are functions from $\mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$, $W(t)$ is a Brownian motion and $N(t)$ is a counting process. Let $Y(t) := f(t, X(t))$, where $f \in C^{1,2}$ is once resp. twice continuously differentiable in t resp. $X(t)$, then the SDE for $Y(t)$ is given by

$$\begin{aligned} dY(t) =& \big[f_x(t, X(t))\mu(t, X(t)) + f_t(t, X(t)) + \frac{1}{2} f_{xx}(t, X(t))\sigma^2(t, X(t))\big]dt + \\ & f_x(t, X(t))\sigma(t, X(t))dW(t) + (f(t, X(t)) - f(t_-, X(t_-)))dN(t), \end{aligned}$$

where $f(t_-) = \lim_{s \nearrow t} f(s)$ denotes the left limit.

Theorem 2.28 (First passage time for Brownian motion (Schmidli 2018, p. 109))**.** Let $W(t)$ be a (μ, σ^2)-Brownian motion with $\mu > 0$ and $\tau = inf\{t \geq 0 : u + W(t) < 0\}$ for $u \geq 0$. Then

$$\mathbb{P}(\tau \leq t) = 1 - \Phi\left(\frac{\mu t + u}{\sigma\sqrt{t}}\right) + e^{\frac{-2u\mu}{\sigma^2}} \Phi\left(\frac{\mu t - u}{\sigma\sqrt{t}}\right)$$

and

$$\mathbb{P}(\tau < \infty) = e^{\frac{-2u\mu}{\sigma^2}}.$$

Stochastic control theory

We introduce some basic definitions and results from Yong and Zhou 1999 that might help to understand the results in Section 7.2 in a broader context.

Definition 2.29 (State equation (Yong and Zhou 1999, p. 62f))**.** Given a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with an m-dimensional standard Brownian motion $W(\cdot)$, consider a *controlled stochastic differential equation*, also called the *state equation*:

$$\begin{aligned} dx(t) &= b(t, x(t), u(t))dt + \sigma(t, x(t), u(t))dW(t), \\ x(0) &= x_0 \in \mathbb{R}^n, \end{aligned} \tag{2.4}$$

where $b : [0, T] \times \mathbb{R}^n \times U \to \mathbb{R}^n, \sigma : [0, T] \times \mathbb{R}^n \times U \to \mathbb{R}^{n \times m}$, with U a given separable metric space and $T \in \mathbb{R}_+$ fixed. The function $u(\cdot)$ is called *control* and represents the policy

of the decision-makers. The policy at each timepoint t might only depend on available information from the past (that is stored in $\mathcal{F}_t$), therefore the *set of feasible controls* is given by

$$\mathcal{U}[0,T] = \{u : [0,T] \times \Omega \to U | u(\cdot) \text{ is } \{\mathcal{F}_t\}\text{-adapted}\}.$$

$x(\cdot)$ is called *state* and might be subject to some *state constraint*, for example of the form

$$x(t) \in S(t), \text{ for all } t \in [0,T], \mathbb{P}-a.s.,$$

where $S(t) : [0,T] \to 2^{\mathbb{R}^n}$. Finally, the *cost functional* is defined as

$$J(u(\cdot)) = \mathbb{E}\Big[\int_0^T f(t,x(t),u(t))dt + h(x(T))\Big]. \tag{2.5}$$

Definition 2.30 ((Weak) Admissibility and problem formulation (Yong and Zhou 1999, p. 64))**.** $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}, W(\cdot), u(\cdot))$ (or, if the context is clear, simply $u(\cdot)$) is called *(w)-admissible control*, and $(x(\cdot), u(\cdot))$ a *(w)-admissible pair*, if

- $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ is a filtered probability space,
- $W(\cdot)$ is an m-dimensional standard Brownian motion defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$,
- $u(\cdot)$ is an $\{\mathcal{F}_t\}$-adapted process on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in U,
- $x(\cdot)$ is the unique solution of (2.4) on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ under $u(\cdot)$,
- some state constraints (e.g. as above) are fulfilled,
- $f(\cdot, x(\cdot), u(\cdot)) \in \mathcal{L}^1_{\mathcal{F}}(0,T,\mathbb{R})$ and $h(x(T)) \in \mathcal{L}^1_{\mathcal{F}_T}(\Omega,\mathbb{R})$, where $\mathcal{L}^1_{\mathcal{F}}(0,T,\mathbb{R})$ here denotes the set of all $\{\mathcal{F}\}_t$-adapted $\mathbb{R}$-valued processes $X(\cdot)$ such that $\mathbb{E}\big[\int_0^T |X(t)|dt\big] < \infty$ and $\mathcal{L}^1_{\mathcal{F}_T}(\Omega,\mathbb{R})$ the set of all $\mathcal{F}_T$-measurable, $\mathbb{R}$-valued random variables X such that $\mathbb{E}[|X|] < \infty$.

The *set of all (w-)admissible controls* is denoted $\mathcal{U}^w_{ad}[0,T]$. The prefix w- stands for *weak* formulation and will be omitted from now on as we only consider this formulation here. The stochastic optimal control problem is then stated as:

Minimize (2.5) over $\mathcal{U}_{ad}[0,T]$, i.e. seek $\bar{\varphi} \in \mathcal{U}_{ad}[0,T]$ (if it exists) such that

$$J(\bar{\varphi}) = \inf_{\varphi \in \mathcal{U}_{ad}[0,T]} J(\varphi).$$

A particular case which will be of interest here are so-called *stochastic linear-quadratic (LQ) problems*, which have a linear (in both $x(\cdot)$ and $u(\cdot)$) state equation and a quadratic cost functional. Let $T > 0$ be given and for any $(s, y) \in [0, T) \times \mathbb{R}^n$, consider the non-homogeneous linear equation:

$$\begin{aligned} dx(t) &= [A(t)x(t) + B(t)u(t) + b(t)]dt \\ &\quad + \sum_{j=1}^{m} [C_j(t)x(t) + D_j(t)u(t) + \sigma_j(t)]dW_j(t), \quad t \in [s, T] \\ x(s) &= y, \end{aligned} \tag{2.6}$$

where $A, B, C_j, D_j, b, \sigma_j$ are deterministic, matrix-valued functions. Additionally, consider the quadratic cost functional

$$J(s, y, u(\cdot)) = \mathbb{E}\left[\frac{1}{2}\int_s^T [\langle Q(t)x(t), x(t)\rangle + 2\cdot\langle Sx(t), u(t)\rangle + \langle R(t)u(t), u(t)\rangle]dt + \frac{1}{2}\langle Gx(T), x(T)\rangle\right]. \tag{2.7}$$

Assume that the coefficients fulfil

$$\begin{aligned} &A, C_j \in \mathcal{L}^\infty(0, T; \mathbb{R}^{n\times n}),\ B, D_j, S' \in \mathcal{L}^\infty(0, T; \mathbb{R}^{n\times k}), \\ &Q \in \mathcal{L}^\infty(0, T; \mathcal{S}^n),\ R \in \mathcal{L}^\infty(0, T; \mathcal{S}^k),\ G \in \mathcal{S}^n, \\ &b, \sigma_j \in \mathcal{L}^2(0, T; \mathbb{R}^n),\ j = 1, 2, \cdots, m. \end{aligned} \tag{2.8}$$

Under these assumptions, (2.6) has a unique solution $x(\cdot) \in \mathcal{L}^2_{\mathcal{F}}(s, T; \mathbb{R}^n)$ and (2.7) is well-defined. For any $s \in [0, T)$, denote by $\mathcal{U}[s, T]$ the set of all $(\Omega, \mathcal{F}, \mathbb{P}, W(\cdot), u(\cdot))$ (or simply all $u(\cdot)$) satisfying

- $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space,
- $W(t)$ is an m-dimensional standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ over $[s, T]$ and $\mathcal{F}_t^s = \sigma\{W(r) : s \le r \le t\}$,
- $u(\cdot) \in \mathcal{L}^2_{\mathcal{F}}(s, T; \mathbb{R}^k)$,
- under $u(\cdot)$, for any $y \in \mathbb{R}^n$ (2.6) admits a unique solution $x(\cdot)$ on $(\Omega, \mathcal{F}, \{\mathcal{F}_t^s\}, \mathbb{P})$,
- the right-hand side of (2.7) is well-defined under $u(\cdot)$.

The *stochastic linear-quadratic optimal control problem* is then stated as

Definition 2.31 (Stochastic LQ problem (Yong and Zhou 1999, p. 301))**.** For each $(s, y) \in [0, T] \times \mathbb{R}^n$, find $\bar{u}(\cdot) \in \mathcal{U}[s, T]$ such that

$$(SLQ) \quad J(s, y; \bar{u}(\cdot)) = \inf_{u(\cdot)\in\mathcal{U}[s,T]} J(s, y; u(\cdot)) = V(s, y),$$

where V is called the *value function* of problem (SLQ).

The book by Yong and Zhou 1999 devotes the whole Chapter 6 (p. 281ff) to the solution of LQ problems and shows that in the LQ case, the approaches via the maximum principle and dynamic programming are equivalent and can lead to the *Riccati equation* which delivers an optimal control in a linear state feedback form.

Definition 2.32 (Stochastic Riccati equation associated with SLQ (Yong and Zhou 1999, p. 314))**.** Let $P(\cdot) \in C^1([s,T];\mathcal{S}^n)$ be the solution of

$$\begin{aligned}
&\dot{P} + PA + A'P + C'PC + Q \\
&- (B'P + S + D'PC)'(R + D'PD)^{-1}(B'P + S + D'PC) = 0, \quad a.e.\ t \in [s,T], \\
&P(T) = G, \\
&R + D'PD > 0, \quad a.e. t \in [s,T],
\end{aligned} \tag{2.9}$$

and $Z(\cdot) \in C^1([s,T];\mathbb{R}^n)$ be the solution of the backward SDE

$$\begin{aligned}
&\dot{Z} + [A - B(R + D'PD)^{-1}(B'P + S + D'PC)]'Z \\
&+ [C - D(R + D'PD)^{-1}(B'P + S + D'PC)]'PZ + Pb = 0, \quad a.e.\ t \in [s,T], \\
&Z(T) = 0.
\end{aligned} \tag{2.10}$$

Theorem 2.33 (Optimal control for SLQ (Yong and Zhou 1999, p. 315))**.** Let (2.8) hold, and let $P(\cdot) \in C^1([s,T];\mathcal{S}^n)$ and $Z(\cdot) \in C^1([s,T];\mathbb{R}^n)$ be the solutions of (2.9) and (2.10) respectively, for some $s \in [0,T)$, such that

$$\begin{aligned}
&B\Psi, D\Psi \in \mathcal{L}^\infty(s,T;\mathbb{R}^{n\times n}), \text{ where} \\
&\Psi = (R + D'PD)^{-1}[B'P + S + D'PC],
\end{aligned}$$

and

$$\begin{aligned}
&B\psi, D\psi \in \mathcal{L}^2(s,T;\mathbb{R}^n), \text{ where} \\
&\psi = (R + D'PD)^{-1}[B'Z + D'P\sigma].
\end{aligned}$$

Then problem (SLQ) is solvable at s with the optimal control $\bar{u}(\cdot)$ being of a state feedback form,

$$\bar{u}(t) = -\Psi(t)x(t) - \psi(t), \qquad t \in [s,T], \tag{2.11}$$

and

$$\begin{aligned}
V(s,y) = &\frac{1}{2}\langle P(s)y, y\rangle + \langle Z(s), y\rangle \\
&+ \frac{1}{2}\mathbb{E}\bigg[\int_s^T (2\langle Z, b\rangle + \langle P\sigma, \sigma\rangle - |(R + D'PD)^{\frac{1}{2}}\psi|^2)dt\bigg], \quad \forall y \in \mathbb{R}^n.
\end{aligned} \tag{2.12}$$

We will study a mean-variance portfolio selection problem as a specialised case of (SLQ) in Chapter 7.

Chapter 3

Hawkes processes

3.1 Definition and notation

Many commonly encountered questions in reality, however complex their background mechanisms might be, can essentially be described as understanding a sequence of random future events. The two central questions then are *how many* events will occur and *at what times* those occurrences will take place. Some examples include

- Customers entering a shop during a given day. It is in the shopkeeper's interest to estimate how many customers will arrive (to ensure sufficient supplies) and when they will do so (if there are peak hours, he might need to employ several clerks to avoid long waiting times for the customers).
- Earthquakes occuring in a given geographical region during a year. In this case, the number of occurrences might be less crucial than the occurrence times. If time points with an especially high likelihood of an earthquake occuring soon in the future could be identified, the population could be alerted or even evacuated in time and casualties might be avoided.
- Incoming buy- and sell-orders for a given underlying that will eventually drive the asset price.

Although these examples come from diverse areas and are influenced by very different environmental factors, they all have in common that they can be described by a sequence of points on a timeline (representing events, which are in this context always assumed to occur in an *instant*, e.g. the moment a customer enters the shop). Studying these problems comes down to understanding the temporal distribution and mutual influence of the points and finding a stochastic process that adequately models these point patterns.

Given a point in time t, assume that all information about the process up until this time is available, and this knowledge about all past event times is stored in the *history* $\mathcal{H}(t)$ and is the only information influencing the properties of the process at time t. In other words, what happens in the present moment may only depend on the (known) past, but *not* on (unknown) future events - this is sometimes called *evolutionarity* (see Rasmussen 2011). Furthermore, we focus on processes where no two events occur at the exact same time, so their occurrence times form a strictly increasing sequence. To formalize this idea, we now give a concise summary of some background knowledge on the theory of point processes that can be found in Laub, Taimre, and Pollett 2015. Basic knowledge about probability theory and statistics will be assumed. For an extensive treatment of general point processes and some mathematical background we refer to Daley and Vere-Jones 2003 and Daley and Vere-Jones 2008.

Definition 3.1 (Counting process)**.** A *counting process* is a stochastic process $(N(t) : t \geq 0)$ taking values in $\mathbb{N}_0$ that satisfies $N(0) = 0$, is almost surely (a.s.) finite, and is a right-continuous step function with increments of size $+1$. Further, denote by $\mathcal{H}(u), u \geq 0$, the history of the arrivals up to time u, i.e. $(\mathcal{H}(u) : u \geq 0)$ is a filtration.

The counting process $N(t)$ counts the number of *arrivals* up to time t. An interchangeable way to characterize a process is by the sequence of random *arrival times* $\mathbf{T} = \{\mathbf{t_1}, \mathbf{t_2}, ...\}$ describing the times at which the process $N(\cdot)$ has jumped.

Definition 3.2 (Point process)**.** If a sequence of random variables $\mathbf{T} = \{\mathbf{t_1}, \mathbf{t_2}, \cdots\}$, taking values in $[0, \infty)$, has $\mathbb{P}(0 \leq t_1 \leq \cdots) = 1$, and the number of points in a bounded region is a.s. finite, then $\mathbf{T}$ is a (simple) *point process*.

An example of a realisation of a point process $\mathbf{T} = \{\mathbf{t_1}, \mathbf{t_2}, \cdots\}$ and the corresponding counting process $N(t)$ is given in Figure 1.

A very intuitive way to characterize a point process is by its conditional intensity function:

Definition 3.3 (Conditional Intensity Function)**.** Consider a counting process $N(\cdot)$ with associated history $\mathcal{H}(\cdot)$. If a non-negative, $\mathcal{H}(t)$-measurable function $\lambda^*(t)$ exists such that

$$\lambda^*(t) = \lim_{h \to 0} \frac{\mathbb{E}[N(t+h) - N(t) | \mathcal{H}(t)]}{h}, \tag{3.1}$$

then it is called *conditional intensity function* of $N(\cdot)$.

Definition 3.4 (Compensator)**.** For a counting process $N(\cdot)$ with conditional intensity function $\lambda^*(\cdot)$, the non-decreasing function

$$\Lambda(t) = \int_0^t \lambda^*(s) ds \tag{3.2}$$

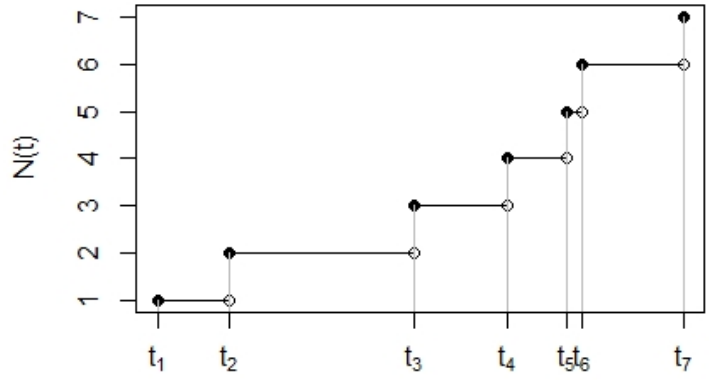


Figure 1: An example of a realisation of a point process $\mathbf{T} = \{\mathbf{t_1}, \mathbf{t_2}, \cdots, \mathbf{t_7}\}$ and the corresponding counting process $N(t)$.

is called the *compensator* of the counting process. It is technically defined as the unique, non-descreasing $\mathcal{H}(t)$-predictable function with $\Lambda(0) = 0$ such that the "compensated" counting process $N(t) - \Lambda(t) =: M(t)$ a.s. for $t \geq 0$ is a $\mathcal{H}(t)$ local martingale (Doob-Meyer-decomposition of N).

Intuitively, $\lambda^*(t)$ can be interpreted as the *expected rate of arrivals* at time t of the process $N(t)$ conditioned on its history $\mathcal{H}(t)$. $\lambda^*(t)$ was originally called the *hazard function* and defined by using the distributions of the time intervals between events as

$$\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)}$$

where $F^*(t)$ and $f^*(t)$ are the conditional c.d.f. (and p.d.f.) of the next arrival time, say t_{k+1}, given the history until the last arrival time $t_k < t$, given by

$$F^*(t) := F^*(t|\mathcal{H}(t_k)) = \int_{t_k}^{t} \mathbb{P}(t_{k+1} \in [s, s + ds]|\mathcal{H}(t_k))ds =: \int_{t_k}^{t} f^*(s|\mathcal{H}(t_k))ds.$$

Intuitively we can see how these two definitions coincide: if we consider a very small time interval $[t, t + dt]$ and assume there have been exactly k arrivals before time t, it holds

$$\begin{aligned}
\frac{f^*(t)}{1 - F^*(t)} &= \frac{\mathbb{P}(t_{k+1} \in [t, t + dt]|\mathcal{H}(t))}{\mathbb{P}(t_{k+1} \geq t|\mathcal{H}(t))} = \frac{\mathbb{P}(t_{k+1} \in [t, t + dt], t_{k+1} \geq t|\mathcal{H}(t))}{\mathbb{P}(t_{k+1} \geq t|\mathcal{H}(t))} \\
&= \mathbb{P}(t_{k+1} \in [t, t + dt]|t_{k+1} \geq t, \mathcal{H}(t)) = \mathbb{P}(t_{k+1} \in [t, t + dt]|\mathcal{H}(t)) \\
&= 1 \cdot \mathbb{P}(t_{k+1} \in [t, t + dt]|\mathcal{H}(t)) + 0 \cdot \mathbb{P}(t_{k+1} > t + dt|\mathcal{H}(t))
\end{aligned}$$

$$= \mathbb{E}[N(dt)|\mathcal{H}(t)] = \lim_{h\to 0} \frac{\mathbb{E}[N(t+h) - N(t)|\mathcal{H}(t)]}{h}.$$

In the first and second line, we use that $t_{k+1} \in [t, t+dt]$ is of course a subset of $t_{k+1} \geq t$ and the definition of conditional probability and $\mathcal{H}(t)$ respectively. In the last line, $N(dt)$ denotes the number of arrivals in an interval of length dt and we recall that for a simple point process, no two arrival times can coincide and thus in an infitesimal timespan there can only be 0 or 1 arrival. 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. More generally, it can be used to classify certain classes of point processes. For instance:

- If $\lambda^*(t)$ is independent of $\mathcal{H}(t)$, the associated process is a *renewal process*.
- If an arrival causes $\lambda^*(t)$ to increase, the associated process is called *self-exciting*. This causes temporal clustering of $\mathbf{T}$ and thus $\lambda^*(t)$ must be chosen carefully to avoid the event of *explosion*, meaning the occurrence of an infinite number of event in a finite time interval.
- On the contrary, if an arrival causes $\lambda^*(t)$ to drop, the process is called *self-regulating* and thus the arrival times of new events appear quite regular.

Example (Poisson process)**.** A *Poisson process* $N(t)$ with intensity λ is a counting process with conditional intensity function $\lambda^*(t) = \lambda$ for all t. Equivalently, a Poisson process can be defined as

- A counting process $N(t)$ with independent increments, i.e. the random variables $(N(t) - N(s))$ and $(N(v) - N(u))$ are independent for all $0 \leq u < v \leq s < t < \infty$, and $(N(s+t) - N(s)) \sim Poi(\lambda t)$ for all $t > 0$.
- A point process $\{t_1, t_2, \cdots\}$ whose interarrival times are distributed $(t_k - k_{k-1} : k \geq 1) \sim Exp(\lambda)$ and independent, where $t_0 = 0$.

Note that the exponential distribution is the unique continuous distribution which has the feature of *memorylessness*, i.e. an interarrival time $\hat{t}$ fulfils

$$\mathbb{P}(\hat{t} > s + t|\hat{t} > s) = \mathbb{P}(\hat{t} > t) \text{ for all } s, t \geq 0.$$

As the exponential distribution and the Poisson process are intrinsically linked, the Poisson process likewise has the unique status of a point process with *memoryless* times between arrivals.

Example (Inhomogeneous Poisson process)**.** An inhomogeneous *Poisson process* $N(t)$ with rate $\lambda(t)$ is a counting process with independent increments such that $(N(t)-N(s)) \sim Poi(\Lambda(t)-\Lambda(s))$ for all $0 \leq s \leq t < \infty$, where $\Lambda(t) = \int_0^t \lambda(u)du$.

Example (Renewal process)**.** A simple point process $\{t_1, t_2, \cdots\}$ is called *ordinary renewal process* if the interarrival times $(t_k - k_{k-1} : k \geq 1)$ are i.i.d..

The assumption of conditionally independent interarrival times (or independent increments) in the last examples might not always be supportable in reality. In an interconnected economy, a credit default event in one company might very well affect other companies. An earthquake that recently occurred in one region usually increases the likelihood of more earthquakes (so-called aftershocks) following soon (Ogata 1988). To capture these phenomena, we now turn to the class of self-exciting processes, namely *Hawkes processes*. The Hawkes process was first introduced by Hawkes 1971 as a simple point process with self-exciting property, clustering effect and long memory. It has since been studied extensively and applied to model e.g. earthquake occurrences (Ogata 1999), trade orders (Da Fonseca and Zaatour 2013), credit defaults (Errais, Giesecke, and Goldberg 2010) or incoming insurance claims (Stabile and Torrisi 2010, Cheng and Seol 2018).

Definition 3.5 (One-dimensional Hawkes process)**.** Consider $N(t), t \geq 0$, a counting process with history $(\mathcal{H}(t), t \geq 0)$ that satisfies

$$\mathbb{P}(N(t+h) - N(t) = m|\mathcal{H}(t)) = \begin{cases} \lambda^*(t)h + o(h), & m = 1 \\ o(h) & m > 1. \\ 1 - \lambda^*(t)h + o(h), & m = 0 \end{cases} \tag{3.3}$$

Suppose its conditional intensity function is of the form

$$\lambda^*(t) = \lambda + \int_0^t \mu(t-s)dN(s) \tag{3.4}$$

where $\lambda > 0$ is called *background intensity* and the non-increasing function $\mu : (0, \infty) \to [0, \infty)$ 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 $(t_1, t_2, \cdots, t_k)$ up to time t, the conditional intensity can be written as

$$\lambda^*(t) = \lambda + \sum_{t_i < t} \mu(t - t_i). \tag{3.5}$$

The one-dimensional Hawkes process in general is very versatile, as theoretically there are countless options to choose the excitation function. However, as mentioned above, the conditional intensity function of a self-exciting process should be chosen in a way that avoids *explosion* of the number of arrivals. This is achieved by restricting our focus to so-called *stationary* processes, which (in the one-dimensional case) fulfill the following

Definition 3.6 (Stationarity condition)**.** A Hawkes process is called *stationary* if its excitation function $\mu(\cdot)$ fulfills

$$\hat{\mu} := \int_0^\infty \mu(s)ds < 1. \tag{3.6}$$

As shown e.g. in Massouli and Bremaud 1996, the above condition implies that there is a unique (asymptotic) stationary version of the Hawkes process $N(t)$, which in the context of simple point processes means that the distribution of $N(\cdot)$ is time-invariant. This implies that a stationary Hawkes process $N(\cdot)$ has asymptotically stationary increments, and the intensity process $\lambda^*(\cdot)$ is an asymptotically stationary process. In the following, we will present the two most common choices of excitation function.

Example (Exponential decay)**.** The *exponentially decaying* excitation function is of the form $\mu(t) = \alpha e^{-\beta t}$ with parameters $\alpha > 0, \beta > 0$. Hence the conditional intensity is given by

$$\lambda^*(t) = \lambda + \int_0^t \alpha e^{-\beta(t-s)} dN(s) = \lambda + \alpha \sum_{t_i<t} e^{-\beta(t-t_i)} \tag{3.7}$$

Then, the stationarity condition (3.6) corresponds to

$$\hat{\mu} = \int_0^\infty \alpha e^{-\beta s} ds = \alpha\big[-\frac{1}{\beta}e^{-\beta s}\big]_{s=0}^{s=\infty} = \frac{\alpha}{\beta}\big[-0+1\big] = \frac{\alpha}{\beta} < 1 \iff \alpha < \beta. \tag{3.8}$$

In this case, each new arrival makes the conditional intensity instantly jump up by α and as time passes, the influence of the arrival vanishes as the intensity decays back to the background rate exponentially with rate β. Thus, the larger the ratio $\frac{\alpha}{\beta}$, the stronger the influence of the self-exciting feature on the total number of arrivals. Figure 2 shows four sample paths of the conditional intensity function over time for different values of α and β to illustrate their influence. This is the most commonly used choice for the excitation function and will also be our focus in later chapters of this thesis. Section 3.5 will go into further detail about the analytical convenience and properties of this specification of excitation function.

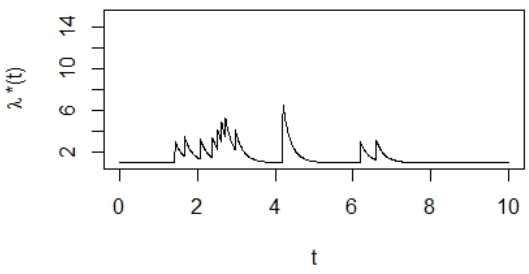
Example (Power Law)**.** The *power law* excitation function is of the form $\mu(t) = \frac{k}{(c+(t-s))^{1+\eta}}$ with parameters $c > 0, k > 0, \eta > 0$. Hence the conditional intensity writes as

$$\lambda^*(t) = \lambda + \int_0^t \frac{k}{(c+(t-s))^{1+\eta}} dN(s) = \lambda + \sum_{t_i<t} \frac{k}{(c+(t-t_i))^{1+\eta}} \tag{3.9}$$

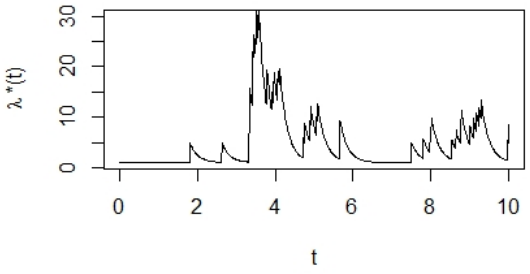
and the stationarity condition (3.6) corresponds to

$$\hat{\mu} = \int_0^\infty \frac{k}{(c+s)^{1+\eta}} ds = k\big[-\frac{1}{\eta}(c+s)^{-\eta}\big]_{s=0}^{s=\infty} = \frac{k}{\eta}\big[-0+c^{-\eta}\big] = \frac{kc^{-\eta}}{\eta} < 1 \iff k < \eta c^{\eta}.$$

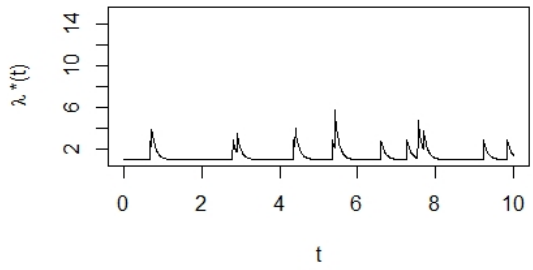
In this case, interpreting the parameters c, k and η is not as straightforward as in the last example, but roughly one could say that k corresponds to α describing the upward move of the intensity caused by an arrival (the magnitude of its influence), η corresponds to β determining how longlasting the impact of the arrival is and c describes a temporal shift to keep the intensity bounded when $(t - t_i)$ is close to 0 (see Rizoiu et al. 2017). Ogata 1988 used the power law specification for the prediction of aftershocks of an earthquake (captured in a geological model called Omori's law). Some authors have stated that e.g. for applications in finance a Hawkes process with power law decay might often reflect empirical data better than a Hawkes process with exponential decay (Bacry, Dayri, and Muzy 2012, Zhang 2016), but as the latter one is analytically and computationally much more convenient, it is often preferred.



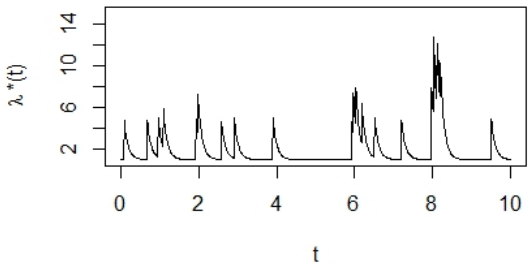
(a) $\lambda = 1, \alpha = 2, \beta = 5$



(b) $\lambda = 1, \alpha = 4, \beta = 5$



(c) $\lambda = 1, \alpha = 2, \beta = 10$



(d) $\lambda = 1, \alpha = 4, \beta = 10$

Figure 2: Plotting the conditional intensity $\lambda^*(t)$ for different realisations of exponentially decaying Hawkes processes illustrates the influence of α and β. In all cases, the background intensity is $\lambda = 1$ and we observe the intensity decaying back to this level in between events. Figure (a) shows that $\lambda^*(t)$ jumps up slightly at each arrival for a relatively low α value, but due to the relatively low β, the influence does not vanish until some time later. For instance, the cluster starting shortly before time 2 does not display very high intensity values, but is rather spread out over time. In Figure (b), the value of α is doubled (c.p.) leading to much stronger clustering as the weak dampening effect of β cannot match the strong self-exciting factor for each new arrival. Note that the y-scale even had to be adapted in this plot. Figure (c) pairs a low self-excitement factor α with a doubled dampening factor of β leading to small clusters that vanish quickly over time. This leads to more isolated spikes as opposed to e.g. (a) where the effect is more spread out over time. Finally, (d) combines relatively high α and β values, leading to "high spikes", i.e. a strong increase of $\lambda^*(t)$ at new arrivals that is quickly dampened again. The simulation method used to obtain the point processes underlying these sample paths of the corresponding conditional intensity $\lambda^*(t)$ is detailed in Section 3.3.

3.2 Immigration-birth-representation

In this section, we will introduce another representation of the Hawkes process, the so-called *immigration-birth* or *clustering* representation. It will be especially interesting for us due to its interpretation in the insurance context in Chapter 6. For background information on branching theory and the most well-known branching process, a Galton-Watson-process, we refer to Grimmett and Stirzaker 2006, Chapter 5.4. In order to understand the Hawkes process as a branching process, suppose we are studying the demographic development in a country where a person can arrive either via *immigration* or via *birth*. The arrival of an immigrant is called a *parent event* of generation 0 that initiates a cluster of *child events* (*descendants* of generations $1, 2, \cdots$ who arrive via birth). A Hawkes process $\{t_1, \cdots, t_k\}$ on an interval $[0, T]$ can be represented as a Poisson cluster process with the following structure:

Proposition 3.7 (Hawkes clustering representation, Bacry, Mastromatteo, and Muzy 2015)**.** Consider a (non-necessarily finite) time interval $[0, T]$ on which we define a sequence of events $\{t_i\}_{i=1}^{k}$ according to the following procedure:

- Consider a set of *immigrant events* $\{t_i^0\}_{i=1}^{k^0}$ according to a homogeneous Poisson process with intensity $\lambda > 0$ on the interval $[0, T]$.
- Each immigrant $t_{i'}^0$ generates a sequence of *first-generation* events $\{t_i^1\}_{i=1}^{k^1}$ following an inhomogeneous Poisson process with time-dependent rate $\mu(t - t_{i'}^0)$ on the interval $[t_{i'}^0, T]$.
- This procedure is iterated from generation $(n-1)$ to generation n to obtain the n^{th} *generation offspring* sequence $\{t_i^n\}_{i=1}^{k^n}$, until no more events are generated in $[0, T]$.

The union of all events $\bigcup_j \{t_i^j\}_{i=1}^{k^j}$ corresponds to the point process generated by a Hawkes process with intensity (3.4).

Note that this clustering representation can be equivalently taken as a definition for the Hawkes point process once one considers only the (ordered) union of all events without taking the generation into account. An illustration of the Hawkes process as a branching process can be seen in Figure 3, taken from Laub, Taimre, and Pollett 2015.

Viewing a Hakwes process as a branching process is often preferable to understand its stability properties. While the stationarity condition (3.6) might seem abstract when introduced as a requirement on the conditional intensity function, it has an easily understandable interpretation in the context of the clustering representation. If an immigrant enters the system at time $t_i \in \mathbb{R}$, they produce offspring at rate $\mu(t - t_i)$ at future times

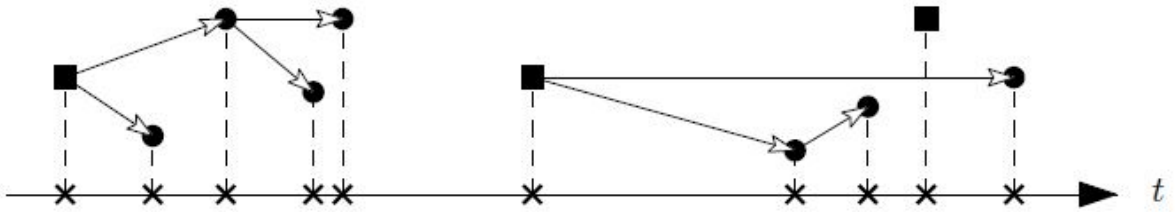


Figure 3: An example realisation of a Hawkes process represented as a collection of three clusters (immigration-birth-representation). Immigrants (parent events) are marked by squares, descendants (child events) by circles and the crosses on the time axis denote the Hawkes point process (ordered events, without taking the generation information into account).

$t > t_i$. Their offspring (first generation) again produces offspring (second generation) and so on - members of all generations are called descendants of the original arrival. Let Z_i denote the random number of offspring in the n^{th} generation (where $Z_0 = 1$ denotes the immigrant). Let the *branching ratio* be denoted as $\hat{\mu}$ and defined as introduced in (3.6)

$$\hat{\mu} = \int_0^\infty \mu(s)ds. \tag{3.10}$$

Then, the expected number of offspring of generation i for one immigrant is (Grimmett and Stirzaker 2006, Chapter 5.4, Lemma 2):

$$\mathbb{E}[Z_i] = \mathbb{E}[Z_1]^i = \hat{\mu}^i,$$

and the total expected number of descendants for one immigrant is given by

$$\mathbb{E}\big[\sum_{i=1}^{\infty} Z_i\big] = \sum_{i=1}^{\infty} \mathbb{E}[Z_i] = \sum_{i=1}^{\infty} \hat{\mu}^i = \sum_{i=0}^{\infty} \hat{\mu}^i - 1 = \begin{cases} \frac{1}{1-\hat{\mu}} - 1 = \frac{\hat{\mu}}{1-\hat{\mu}}, & \hat{\mu} < 1 \\ \infty & \hat{\mu} \geq 1 \end{cases}$$

where we use the convergence of the geometric series for $\hat{\mu} < 1$. Thus we can use the branching process terminology to describe the behaviour of a Hawkes model by the expected number of descendants generated by each immigrant (see Bacry, Mastromatteo, and Muzy 2015):

- $\hat{\mu} < 1$ corresponds to *sub-critical* or *stationary* case where each parent event generates on average less than one child event. This implies that the total number of offspring events of each parent event is a.s. finite.
- $\hat{\mu} > 1$ is the *super-critical* case in which more than one child event is generated by each parent event. In that case the expected total offspring of a parent is infinite leading to an *explosion* of the population.

- The case $\hat{\mu} = 1$ is the *critical* case, where the total number of offspring events has large fluctuations leading to a divergence of the expected number of generations and thus *explosion* of the expected population.

We can now see how the stationarity condition (3.6), $\hat{\mu} < 1$, is necessary to avoid explosion of the Hawkes process and from now on, we will always assume this condition to be fulfilled. Note that for $\hat{\mu} \in (0, 1)$, the branching ratio $\hat{\mu}$ can be interpreted as the probability that a random arrival was generated endogenously (a child event) as opposed to an exogenous arrival (a parent event). This can be seen from the ratio of the number of descendants to the size of the whole "family" (children + original parent), as it holds

$$\frac{\mathbb{E}\left[\sum_{i=1}^{\infty} Z_i\right]}{1 + \mathbb{E}\left[\sum_{i=1}^{\infty} Z_i\right]} = \frac{\frac{\hat{\mu}}{1-\hat{\mu}}}{1 + \frac{\hat{\mu}}{1-\hat{\mu}}} = \hat{\mu}.$$

In financial applications, this branching representation of the Hawkes process was first used by Filimonov and Sornette 2012 to quantify the endogeneity or *reflexivity* in financial markets. The idea was to differentiate price changes (e.g. in futures contracts) according to whether they occur due to actual exogenous news or are caused by endogenous feedback processes of the trading activity itself. They used this quantification of a *critical state* of the financial market (defined as a situation of diverging endogeneous trading activity in the absence of any external driving) to predict so-called *flash crashes* as had occurred in some US markets in 2010 supposedly due to the activity of high-frequency traders. Hardiman, Bercot, and Bouchaud 2013 analyze (and partly correct) their findings in a related work showing that financial market activity has been evolving around the critical state ($\hat{\mu} \approx 1$) for some time. Motivated by this result, Hardiman and Bouchaud 2014 introduce a model-independent approximation for the branching ratio based only on the mean and variance of the event count.
As we will see in Chapter 6, the branching ratio is a very interesting quantity to consider in insurance applications. The great advantage when working with an empirical insurance data set as in this thesis (as opposed to financial market data) is that it is possible to calculate an "empirical branching ratio" from the data in order to compare it with the estimate implied by the fitted Hawkes model.
Before moving on to the next section, we introduce two important results which were derived using the cluster representation of the Hawkes process:

Theorem 3.8 (Law of large numbers for Hawkes process (Laub, Taimre, and Pollett 2015, Daley and Vere-Jones 2008))**.** Let $N(t)$ be a stationary Hawkes process with conditional

intensity function $\lambda^*(t)$ as in (3.4). Let $0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1$. Then, as $t \to \infty$, it holds

$$\mathbb{E}[\lambda^*(t)] = \frac{\mathbb{E}[dN(t)]}{dt} \longrightarrow \frac{\lambda}{1-\hat{\mu}} \text{ a.s.}$$

and therefore

$$\frac{N(t)}{t} \longrightarrow \frac{\lambda}{1-\hat{\mu}}.$$

Proof. See Laub, Taimre, and Pollett 2015, Daley and Vere-Jones 2008. □

Theorem 3.9 (Central limit theorem for Hawkes process (Laub, Taimre, and Pollett 2015, Theorem 1))**.** If $N(t)$ is a stationary Hawkes process with conditional intensity function $\lambda^*(t)$ as in (3.4) and

$$0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1 \text{ and } \int_0^\infty s\mu(s)ds < \infty,$$

then the number of Hawkes process arrivals in $(0, t]$ is asymptotically normally distributed. More precisely

$$\mathbb{P}\left(\frac{N(t) - \frac{\lambda}{1-\hat{\mu}}t}{\sqrt{\frac{\lambda}{(1-\hat{\mu})^3}t}} \leq y\right) \overset{t\to\infty}{\longrightarrow} \mathbb{P}(Z \leq y),$$

where Z is a standard normally distributed r.v..

Proof. See Hawkes and Oakes 1974. □

3.3 Simulation

The ability to simulate a point process is useful for a variety of reasons. When studying a complex process, simulating a number of sample paths may give an idea of what this process typically looks like and help to estimate statistics that cannot be calculated analytically (e.g. the expected number of events on an interval of given length). Furthermore, simulated data can be used to check the functionality of methods to be used on empirical data sets. This section will introduce one method to procude Hawkes process realisations, the so-called *modified thinning algorithm*, in detail and give a brief overview of other approaches. All this information and a more extensive overview can be found in Laub, Taimre, and Pollett 2015. The method of *thinning* was first used by Lewis and Shedler 1979 to generate an inhomogeneous Poisson process and was later adapted by Ogata 1981

for Hawkes process simulation. The underlying idea is that a homogeneous Poisson process, say with rate λ, can be easily generated by drawing i.i.d. $Exp(\lambda)$-distributed random variables representing the interarrival times (see the example in Section 3.1). To generate an inhomogeneous Poisson process with time-dependent rate $\lambda(t)$, one chooses an upper bound $M \geq \lambda(\cdot)$ and generates a homogeneous process with rate M, i.e. interarrival times are drawn i.i.d. $Exp(M)$-distributed. As this process is "too fast" (i.e. it generally has too many events), points have to be removed probabilistically. This is accomplished by drawing a random variable $U_i \sim Unif(0, M)$ for each candidate point t_i and keeping the point only if $U_i \leq \lambda(t_i)$. Intuitively, this means that a large number of points are created evenly over time and the higher the intensity during some time period, the fewer points will be removed, so this period will finally have a relatively high number of events (and vice versa for periods with lower intensity). In fact, the two steps of generating and probabilistically removing arrivals are not conducted disjointly, but each generated arrival is immediately checked (and potentially removed). Essentially, the same approach can be used to simulate a Hawkes process, the difference being that one cannot set a fixed upper bound M beforehand as the intensity $\lambda^*(t)$ depends on the random realisation of the point process. However, one common assumption is that $\lambda^*(t)$ does not increase in periods without any arrivals. Thus one starts with a bound of e.g. $M^* = \lambda$ and subsequently for each arrival t_k that is generated, the upper bound M^* is adjusted to $\lambda^*(t_k + \epsilon)$, where $\epsilon > 0$ is chosen as a very small value. This reflects that the intensity is potentially highest just after a new arrival has been registered. In case the arrival candidate was kept, M^* will be adjusted upward (due to the *self-exciting* feature that a new arrival will increase the intensity), in case it was rejected, the adjustment will be downward to reflect the passage of time without a new arrival (and the effect of the last accepted arrival will meanwhile have diminished). The rest of the procedure is identical to the former case. The formal algorithms are given below and an example (taken from Laub, Taimre, and Pollett 2015) is displayed in Figure 4. This also serves to illustrate the main difference in the time-dependence of the intensity between a non-homogeneous Poisson process and a Hawkes process. In the former case, the time-dependence of $\lambda(t)$ is *deterministically* given by some external factor and not influenced by the realisation of the process itself. The interarrival times are generated as $Exp(M)$, where M is a fixed value set before the beginning of the simulation. This could represent a situation where the probability of an event depends on *seasonalities* over time, e.g. the likelihood of a hurricane occuring during a year usually depends on the season. In the Hawkes case on the other hand, $\lambda^*(t)$ *stochastically* depends on the realisation of the process over time, and the distributions of each interarrival time (drawn as $Exp(M^*)$) always depends on all the other arrivals preceeding it. As described above, this is more suitable for a situation such as the occurrence of earthquakes, which do not display fixed seasonalities, but rather depend on former earthquake occurrences in

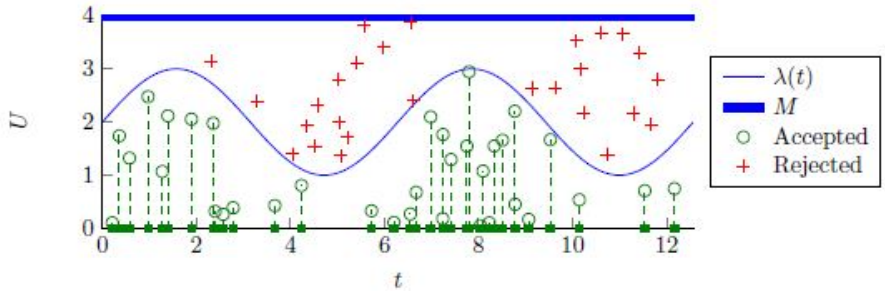
the same region.

Input: T, $\lambda(\cdot)$, M such that $\lambda(\cdot) < M$ on [0,T]
initialization t = 0, NHPP = ();
while $t < T$ **do**
    temp = Exp(M);
    t = t + temp;
    U = Unif(0,M);
    if $t < T$ ***and*** $U \leq \lambda(t)$ **then**
        NHPP = (NHPP,t);
    end
end
return *NHPP*

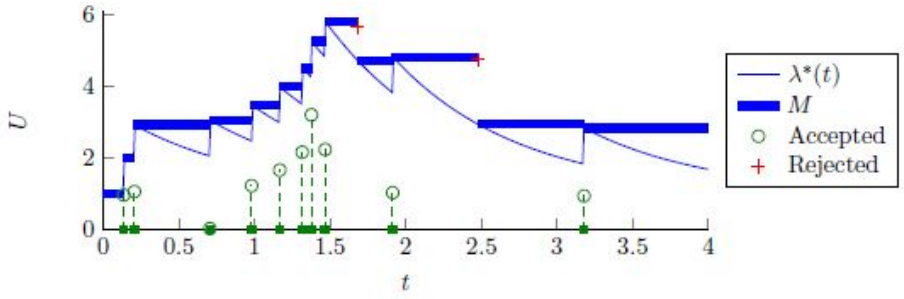
Algorithm 1: Generate a non-homogeneous Poisson process by thinning.

Input: T, $\lambda^*(\cdot)$ with $\lambda^*(\cdot)$ non-increasing in periods without arrivals
initialization t = 0, HP = ();
$\epsilon = 10^{-10}$;
while $t < T$ **do**
    $M^* = \lambda^*(t + \epsilon)$;
    temp = Exp(M^*);
    t = t + temp;
    U = Unif(0,M^*);
    if $t < T$ ***and*** $U \leq \lambda^*(t)$ **then**
        HP = (HP,t);
    end
end
return *HP*

Algorithm 2: Generate a Hawkes process by modified thinning.



(a) Non-homogeneous Poisson process with $\lambda(t) = 2 + sin(t)$, with $M = 4$.



(b) Exponential Hawkes process with $(\lambda, \alpha, \beta) = (1, 1, 1.1)$.

Figure 4: Examples of the (modified) thinning algorithm to generate a non-homogeneous Poisson process and an exponential Hawkes process. All candidates for arrivals are plotted together with the value of their corresponding $Unif(0, M)$ variable, and marked with a green circle in case the arrival was kept and a red cross in case it was deleted. The resulting point process is marked with green squares on the time axis. This illustrates how the probability of an arrival being removed depends on the current state of the intensity function. The blue line marking the upper bound M emphasizes the difference between the two cases: In the upper picture, it is constant over the whole simulation, and in the lower picture, it has to be adjusted after each potential arrival.

There are several alternative simulation methods which are described briefly in Laub, Taimre, and Pollett 2015. The *inverse compensator method* relies on the converse of the random time change theorem (introduced in Section 3.4 as Theorem 3.11). Essentially, the inverse of the compensator defined in 3.4 can be used to transform a unit rate Poisson process $\{t_1^*, t_2^*, \cdots\}$ into any general point process $\{t_1, t_2, \cdots\}$ corresponding to the compensator, by iteratively solving

$$t_1^* = \int_0^{t_1} \lambda^*(s)ds, \qquad t_{k+1}^* - t_k^* = \int_{t_k}^{t_{k+1}} \lambda^*(s)ds.$$

This is described in detail e.g. in Daley and Vere-Jones 2003 (Algorithm 7.4.III) and was first used in a modified version for Hawkes processes in Ozaki 1979. Another method is based on the *clustering representation* of the Hawkes process introduced in Section 3.2. It generates a homogeneous Poisson process with background rate λ for the arrivals representing *immigrants* and for each immigrant, say at time t_i, it generates

an inhomogeneous Poisson process with time-dependent intensity $\mu(t - t_i)$ on (t_i, ∞) representing its *descendants*.

3.4 Parameter estimation and goodness of fit testing

Apart from being able to simulate a typical realisation of a process, another essential tool when working with data is inference or parameter estimation. Given a set of data points which are suspected to follow a pattern described by a parametric model, the goal is to find the set of parameters that best decribe the data at hand. Goodness of fit methods are then used to check how closely the proposed model (with the obtained set of parameters) actually matches the observed data set, in the hope of determining (or at least getting an indication of) whether the model is suitable. The basic principles and a variety of methods for parameter estimation are studied in any statistics course and will not be reiterated at this point, but a recommendable source for an extensive overview is e.g. Czado and Schmidt 2011. One of the most common methods is *maximum likelihood estimation*, which relies on finding the likelihood function $L(\theta)$ (as a function of the parameter set θ and the observed data, which we drop as an argument here) and maximizing it over all feasible sets of parameter values $\theta \in \Theta$, i.e. finding the parameter set that makes the observed data point realisation most *likely*. Formally, the maximum likelihood estimate is defined as $\hat{\theta} = argmax_{\theta \in \Theta} L(\theta)$. Equivalently, the log-likelihood function $l := log(L)$ can be used and is often computationally more convenient as it replaces multiplication with summation. As will be the case for the Hawkes process, often the (log-)likelihood function cannot be optimized analytically, therefore it is often helpful to express it in a way that facilitates numerical optimization. We will now recapitulate the likelihood function for a general point process and then derive the log-likelihood function explicitly for two Hawkes processes, namely exponentially decaying and power law.

Theorem 3.10 (Point process likelihood (Daley and Vere-Jones 2003, Proposition 7.2.III))**.** Let $N(\cdot)$ be a regular point process on $[0, T]$ for some finite $T > 0$, and let $\{t_1, \cdots, t_k\}$ denote a realisation of $N(\cdot)$ over $[0, T]$. Then, the likelihood L of $N(\cdot)$ is expressible in the form

$$L = \big[\prod_{i=1}^{k} \lambda^*(t_i)\big] exp\bigg(- \int_0^T \lambda^*(u) du \bigg). \tag{3.11}$$

Proof. See Daley and Vere-Jones 2003. □

(3.11) implies that the log-likelihood can be derived as

$$l = \sum_{i=1}^{k} log(\lambda^*(t_i)) - \int_0^T \lambda^*(u)du = \sum_{i=1}^{k} log(\lambda^*(t_i)) - \Lambda^*(T), \tag{3.12}$$

where $\Lambda^*(\cdot)$ denotes the compensator.

Example (Exponential decay log-likelihood)**.** In this case the conditional intensity is given as $\lambda^*(t) = \lambda + \alpha \sum_{t_i<t} e^{-\beta(t-t_i)}$. Thus the compensator for an interval $[0, T]$ with arrivals $\{t_1, \cdots, t_k\}$ can be computed as

$$\begin{aligned}
\Lambda(T) &= \int_0^T \lambda^*(u)du = \int_0^{t_1} \lambda^*(u)du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda^*(u)du + \int_{t_k}^{T} \lambda^*(u)du \\
&= \int_0^{t_1} \lambda du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \left(\lambda + \alpha \sum_{t_j<u} e^{-\beta(u-t_j)}du\right) + \int_{t_k}^{T} \left(\lambda + \alpha \sum_{t_j<u} e^{-\beta(u-t_j)}du\right) \\
&= \lambda T + \alpha \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \sum_{j=1}^{i} e^{-\beta(u-t_j)}du + \alpha \int_{t_k}^{T} \sum_{j=1}^{k} e^{-\beta(u-t_j)}du \\
&= \lambda T + \alpha \sum_{i=1}^{k-1} \sum_{j=1}^{i} \int_{t_i}^{t_{i+1}} e^{-\beta(u-t_j)}du + \alpha \sum_{j=1}^{k} \int_{t_k}^{T} e^{-\beta(u-t_j)}du \\
&= \lambda T - \frac{\alpha}{\beta}\left[\sum_{i=1}^{k-1}\sum_{j=1}^{i}(e^{-\beta(t_{i+1}-t_j)} - e^{-\beta(t_i-t_j)}) + \sum_{j=1}^{k}(e^{-\beta(T-t_j)} - e^{-\beta(t_k-t_j)})\right] \\
&= \lambda T - \frac{\alpha}{\beta}\left[\sum_{i=1}^{k}(e^{-\beta(t_k-t_i)} - e^{-\beta(t_i-t_i)}) + \sum_{i=1}^{k}(e^{-\beta(T-t_i)} - e^{-\beta(t_k-t_i)})\right] \\
&= \lambda T - \frac{\alpha}{\beta}\left[\sum_{i=1}^{k}(e^{-\beta(T-t_i)} - 1)\right]
\end{aligned} \tag{3.13}$$

where we first split up the integral into the segments between the arrivals and insert the specification of $\lambda^*(t)$. After exchanging summation and integration and computing the integrals, we realize that the double summation part forms a kind of telescopic series (in i) and thus many terms cancel. We thus change j to i and further observe that including the summand for $i = k$ (whose value is 0) allows to combine the two sums. Inserting $\lambda^*(t)$ and $\Lambda(T)$ into (3.12) yields

$$l = \sum_{i=1}^{k} log\left(\lambda + \alpha \sum_{j=1}^{i-1} e^{-\beta(t_i-t_j)}\right) - \lambda T + \frac{\alpha}{\beta}\left[\sum_{i=1}^{k}(e^{-\beta(T-t_i)} - 1)\right].$$

Unfortunately, the double summation in the first term makes optimizing this form of the log-likelihood function computationally inefficient. However, Ozaki 1979 introduced a way of circumventing this problem by using a recursion: For $i \in \{1, \cdots, k\}$, let $A(i)$ be defined as

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

Then, l can be expressed as

$$l = \sum_{i=1}^{k} log(\lambda + \alpha A(i)) - \lambda T + \frac{\alpha}{\beta}\Big[\sum_{i=1}^{k}(e^{-\beta(T-t_i)} - 1)\Big]. \tag{3.14}$$

Note that this recursion implies that the joint process $(\lambda^*(t), N(t))$ is Markovian for this choice of excitation function (see Liniger 2009, Remark 1.22), a property that will be essential in Section 3.5.

Example (Power law log-likelihood)**.** In the power law case $\lambda^*(t)$ is given by $\lambda^*(t) = \lambda + \sum_{t_i<t} \frac{k}{(c+t-t_i)^{1+\eta}}$. Thus, analogously to above, the compensator for an interval $[0, T]$ with arrivals $\{t_1, \cdots, t_k\}$ is given by

$$\begin{aligned}
\Lambda(T) &= \int_0^T \lambda^*(u)du = \int_0^{t_1} \lambda^*(u)du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda^*(u)du + \int_{t_k}^{T} \lambda^*(u)du \\
&= \lambda T + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \sum_{j=1}^{i} \frac{k}{(c+u-t_j)^{1+\eta}}du + \int_{t_k}^{T} \sum_{j=1}^{k} \frac{k}{(c+u-t_j)^{1+\eta}}du \\
&= \lambda T + k\sum_{i=1}^{k-1}\sum_{j=1}^{i} \int_{t_i}^{t_{i+1}} \frac{1}{(c+u-t_j)^{1+\eta}}du + k\sum_{j=1}^{k} \int_{t_k}^{T} \frac{1}{(c+u-t_j)^{1+\eta}}du \\
&= \lambda T + k\sum_{i=1}^{k-1}\sum_{j=1}^{i} -\frac{1}{\eta}\big[(c+t_{i+1}-t_j)^{-\eta} - (c+t_i-t_j)^{-\eta}\big] \\
&+ k\sum_{j=1}^{k} -\frac{1}{\eta}\big[(c+T-t_j)^{-\eta} - (c+t_k-t_j)^{-\eta}\big] \\
&= \lambda T - \frac{k}{\eta}\Bigg(\sum_{i=1}^{k}\big((c+t_k-t_i)^{-\eta} - (c+t_i-t_i)^{-\eta}\big) + \sum_{i=1}^{k}(c+T-t_i)^{-\eta} - (c+t_k-t_i)^{-\eta}\Bigg) \\
&= \lambda T - \frac{k}{\eta}\sum_{i=1}^{k}\big((c+T-t_i)^{-\eta} - c^{-\eta}\big) = \lambda T + \frac{k}{\eta}\sum_{i=1}^{k}\big(c^{-\eta} - (c+T-t_i)^{-\eta}\big)
\end{aligned}$$

Thus, the log-likelihood is given by

$$l = \sum_{i=1}^{k} log\Bigg(\lambda + \sum_{j=1}^{i-1} \frac{k}{(c+t-t_i)^{1+\eta}}\Bigg) - \lambda T - \frac{k}{\eta}\sum_{i=1}^{k}\big(c^{-\eta} - (c+T-t_i)^{-\eta}\big),$$

where unfortunately in this case the double summation in the first term cannot be avoided, thus the estimation is computationally very slow for large data samples.

As the log-likelihood function in neither case can be optimized analytically, we rely on numerical optimization by using the *Nelder-Mead optimization* implemented in the *R*-package *lme4* to minimize the negative log-likelihood function. Table 1 in the Appendix gives a few examples of the estimation results for simulated exponential Hawkes processes with different parameter values (λ, α, β). We observe that generally the maximum likelihood method is very effective on simulated, sufficiently large data sets and the consistency, asymptotic normality and efficiency of the estimator were proven by Ogata 1978. However, for empirical data sets some concerns about numerical disadvantages of the method, such as bias for small sample sizes, dependence on the initial parameter values leading to get stuck in local optima and performance issues for very large samples (especially for high-frequency trading applications) have been raised e.g. in Da Fonseca and Zaatour 2013 and Filimonov and Sornette 2015. This motivated the derivation of a generalized method of moments for paramater estimation of exponential Hawkes processes in Da Fonseca and Zaatour 2013 based on their results introduced in Section 3.5. Other recent works on parameter estimation for Hawkes processes include an Expectation Maximization based technique based on the cluster representation of the Hawkes process in Veen and Schoenberg 2008 and different non-parametric approaches, e.g. in Bacry, Dayri, and Muzy 2012 and Lewis and Mohler 2011. However, for the purposes of this thesis we found the maximum likelihood method to be suitable and sufficiently fast for the exponential case and thus this method will be used for the empirical data sets in Chapter 6.

When working with empirical point data, after fitting a Hawkes model the next step is to determine the appropriateness of the model for the observed data. Laub, Taimre, and Pollett 2015 give an overview of approaches to test the *goodness of fit* of a Hawkes model. The basic tool for all of these approaches is the following theorem which allows to transform any point process realisation $\{t_1, t_2, \cdots, t_k\}$ into a unit rate Poisson process using the compensator of the point process.

Theorem 3.11 (Random Time Change Theorem (Daley and Vere-Jones 2003, Proposition 7.4.IV))**.** Let $\{t_1, t_2, \cdots, t_k\}$ be a realisation 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 $\{t_1^*, \cdots, t_k^*\} = \{\Lambda(t_1), \cdots, \Lambda(t_k)\}$ form a Poisson process with unit rate. $\Lambda(\cdot)$ denotes the compensator of the point process.

Proof. See Daley and Vere-Jones 2003. □

As in our case the closed form of the compensator for an exponential Hawkes process is

known from (3.13), one can test the quality of the parameter estimation by transforming the original timepoints and performing standard fitness tests (see e.g. Laub, Taimre, and Pollett 2015 or Cox and Lewis 1966 for an extensive overview) for a unit rate Poisson process on the transformed datapoints. A simple test is to check whether the interarrival times $\{\tau_1^*, \tau_2^*, \cdots\} := \{t_1^*, t_2^* - t_1^*, \cdots\}$ are i.i.d. $Exp(1)$ distributed. This can be done by creating a quantile-quantile plot comparing theoretical and empirical quantiles, where a good fit would be indicated by points closely following the identity line. A qualitative independence test can be conducted by plotting the series of transformed points (U_k, U_{k+1}), where $U_k := 1 - e^{-(t_k^* - t_{k-1}^*)}$, on the unit square $[0,1] \times [0,1]$ and visually examining if the points are evenly scattered (theoretically, by the probability integral transform, they should be uniformly distributed). The results for the example estimations from Table 1 are displayed in Figure 29.

We have hinted several times that the Hawkes process with exponentially decaying intensity has been most commonly used in applications (e.g. Embrechts, Liniger, and Lin 2011, Muni Toke and Pomponio 2011, Rambaldi, Pennesi, and Lillo 2015) and will be the focus on this work due to its analytical convenience. In this section, in particular we have seen this convenience in the recursive expression of the log-likelihood function that allows fast numerical optimization. The next section will go into more detail about the analytical properties of the exponential Hawkes process that increase the tractability of this model in a way that often outweighs the disadvantages of choosing such a fixed, quite restrictive parametric model specification.

3.5 Properties of exponentially decaying Hawkes processes

This section is mainly based on the work of Da Fonseca and Zaatour 2013 who review the analytical properties of the exponential Hawkes process and derive a way to compute the moments of the number of jumps of such a process over a given time interval in closed form. We will particularly focus on the first two moments, that is the expectation and variance of the number of jumps, as well as the autocorrelation function. Based on their results, Da Fonseca and Zaatour 2013 even develop a new method of parameter estimation (essentially a generalized method of moments) whose main advantages over the maximum likelihood method is its extremely fast computation speed and thus the ability to roll the estimation procedure to study parameter stability. These advantages are crucial in their field of application, namely analyzing high-frequency trading activity, where data is available on a millisecond scale and decisions have to be made instantaneously. For

the applications of this thesis, namely insurance data, we rather use their results as another test of the goodness of fit of the Hawkes model by comparing the moments of the number of jumps obtained from the empirical data with the theoretical quantities given the maximum likelihood estimate. Recall that according to (3.7) the conditional intensity of an exponential Hawkes process is given by

$$\lambda^*(t) = \lambda + \int_0^t \alpha e^{-\beta(t-s)} dN(s). \tag{3.15}$$

Differentiating this equation leads to

$$\begin{aligned} d\lambda^*(t) &= \alpha dN(t) + (\alpha \int_0^t -\beta e^{-\beta(t-s)} dN(s))dt \\ &= \alpha dN(t) - \beta(\alpha \int_0^t e^{-\beta(t-s)} dN(s))dt \\ &= \alpha dN(t) - \beta(\lambda^*(t) - \lambda)dt, \end{aligned} \tag{3.16}$$

where we can add an initial condition $\lambda^*(0) = \lambda_0$. So far we have implicitly assumed $\lambda_0 = \lambda$, i.e. at time 0 the process already starts in its stationary regime. Da Fonseca and Zaatour 2013 use the more general formulation of a process that might start at an arbitrary level λ_0, but then study the limit for $t \to \infty$ to get rid of the dependence of the starting value λ_0. Thus, they denote as λ_∞ the background level λ which the process assumes after running a sufficiently long time to arrive in its stationary regime. In any case, an underlying assumption here is that the stationarity condition (3.8) is fulfilled, i.e. $\alpha < \beta$. In order to solve the above SDE, applying Itô's Lemma for jump diffusion processes on $f(t, \lambda^*(t)) := e^{\beta t}\lambda^*(t)$ yields

$$\begin{aligned} df(t, \lambda^*(t)) &= f_t dt + f_{\lambda^*}(-\beta(\lambda^*(t) - \lambda))dt + (f(t, X(t)) - f(t_-, X(t_-)))dN(t) \\ &= \beta e^{\beta t}\lambda^*(t)dt + e^{\beta t}(-\beta(\lambda^*(t) - \lambda))dt + \alpha e^{\beta t} dN(t) \\ &= \lambda\beta e^{\beta t} dt + e^{\beta t}\alpha dN(t). \end{aligned}$$

Integration from 0 to t gives

$$\begin{aligned} f(t, \lambda^*(t)) &= f(0, \lambda^*(0)) + \int_0^t \lambda\beta e^{\beta s} ds + \int_0^t \alpha e^{\beta s} dN(s) \\ &= \lambda_0 + \lambda[e^{\beta s}]_0^t + \alpha \int_0^t e^{\beta s} dN(s) \\ &= \lambda_0 + \lambda(e^{\beta t} - 1) + \alpha \int_0^t e^{\beta s} dN(s), \end{aligned}$$

and multiplying with $e^{-\beta t}$ finally yields

$$\begin{aligned}\lambda^*(t) &= \lambda_0 e^{-\beta t} + \lambda(1 - e^{-\beta t}) + \alpha \int_0^t e^{-\beta(t-s)} dN(s) \\ &= e^{-\beta t}(\lambda_0 - \lambda) + \lambda + \int_0^t \alpha e^{-\beta(t-s)} dN(s).\end{aligned} \tag{3.17}$$

Remark. On the first glance, (3.16) might remind the reader of the equation defining an *Ornstein-Uhlenbeck-process* (see Uhlenbeck and Ornstein 1930) $x(t)$ which is given by

$$dx(t) = \theta(\mu - x(t))dt + \sigma dW(t), \qquad x(0) = a, \tag{3.18}$$

where $\theta > 0$, μ and $\sigma > 0$ are the parameters and $W(t)$ is a standard Brownian motion. μ (corresponding to λ) denotes the *mean reversion level*, θ (corresponding to β) the *mean reversion speed* and σ (corresponding to α) the influence of the stochastic (noise) $W(t)$. However, there are a few fundamental differences between the processes described by (3.16) and (3.18) respectively:

- The Ornstein-Uhlenbeck process behaves symmetrically around its mean level μ, in particular the process $x(t)$ can take values above and below μ making the drift term $\theta(\mu - x(t))$ negative or positive respectively. On the contrary, for the exponential Hawkes process λ denotes the background rate, and in the stationary regime it always holds $\lambda^*(\cdot) > \lambda$, thus λ is not a mean level, but rather a lower bound on the process and the drift (reversion to this level) $-\beta(\lambda^*(t) - \lambda)$ is always negative.
- Likewise, the stochastic influence of the noise term $\sigma dW(t)$ is symmetric around 0 whereas the stochastic influence of $\alpha dN(t)$ is always positive.
- Most importantly, in the Ornstein-Uhlenbeck case the stochastic term $\sigma dW(t)$ is governed by a process $W(t)$ which is *independent* of $x(t)$. In the Hawkes process case, we have a process $X(t) = (\lambda^*(t), N(t))$ were the dynamic of the intensity $d\lambda^*(t)$ is stochastically influenced by the jump process $N(t)$ which in turn is probabilistically governed by the intensity (recall the first part of Definition 3.5).

This originates the self-exciting feature of the Hawkes process where the evolution of the intensity stochastically depends on the process itself whereas in the Ornstein-Uhlenbeck case the stochastic influence can be interpreted as noise whose evolution is independent from the observed process.

An essential observation is that while the Hawkes process in general in a non-Markovian extension of the Poisson process, choosing the specification with an exponentially decaying

intensity will lead to $X(t) = (\lambda(t), N(t))$ being a Markov process on the state space $\mathbb{R}_+ \times \mathbb{N}$ (see Oakes 1975 or Proposition 2 of Bacry, Mastromatteo, and Muzy 2015). This property gives access to the *Dynkin formula*, which will be an essential tool:

Theorem 3.12 (Dynkin formula (adapted from Revuz and Yor 1999, p. 284))**.** Let $X(t)$ be a Markov process on the state space S. Let $\mathcal{L}$ denote the *infinitesimal generator* of the process, that is let $f : S \to \mathbb{R}$ be a sufficiently regular function (e.g. $f \in C_0^2(\mathbb{R}^n)$) and $\mathcal{L}$ the operator acting on f such that

$$\mathcal{L}f(x) = \lim_{h \to 0} \frac{\mathbb{E}_t^x[f(X(t+h))] - f(x)}{h}$$

where $\mathbb{E}_t^x[\cdot] = \mathbb{E}^x[\cdot|\mathcal{H}(t)]$ and $X(t) = x$. Then, for every function f in the domain of the infinitesimal generator, the process

$$M(t) := f(X(t)) - f(X(0)) - \int_0^t \mathcal{L}f(X(u))du$$

is a martingale relative to its natural filtration and thus, for $s > t$:

$$\mathbb{E}_t\left[f(X(s)) - \int_0^s \mathcal{L}f(X(u))du\right] = f(X(t)) - \int_0^t \mathcal{L}f(X(u))du.$$

This leads to the *Dynkin formula*:

$$\mathbb{E}_t[f(X(s))] = f(X(t)) + \mathbb{E}_t\left[\int_t^s \mathcal{L}f(X(u))du\right] \tag{3.19}$$

which allows computation of conditional expectations of the Markov process $X(t)$.

First, we compute the infinitesimal generator for the exponential Hawkes process $X(t) = (\lambda^*(t), N(t))$:

$$\begin{aligned}
\mathbb{E}_t^x[f(X(t+h))] &= \mathbb{E}[f(X(t+h))|X(t) = x] \\
&= \lambda^*(t)hf(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h + \alpha, N(t)+1) \\
&+ (1 - \lambda^*(t)h)f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t)) \\
&= \lambda^*(t)h[f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h + \alpha, N(t)+1) - f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t))] \\
&+ f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t))
\end{aligned}$$

where we use the first part of Definition 3.5 already keeping in mind that there can only be no jump or exactly one jump in an infitesimal time step. The first term corresponds to the case where one jump occurs (with probability $\lambda^*(t)h$) and the second term to the

case without a jump (with probability $(1-\lambda^*(t)h)$). The second step only rearranges the terms.

$$\begin{aligned}
\Rightarrow \mathcal{L}f(x) &= \lim_{h\to 0} \lambda^*(t)[f(\lambda^*(t)+\beta(\lambda-\lambda^*(t))h+\alpha, N(t)+1) - f(\lambda^*(t)+\beta(\lambda-\lambda^*(t))h, N(t))] \\
&\quad + \lim_{h\to 0} \frac{f(\lambda^*(t)+\beta(\lambda-\lambda^*(t))h, N(t)) - f(\lambda^*(t), N(t))}{h} \\
&= \lambda^*(t)[f(\lambda^*(t)+\alpha, N(t)+1) - f(\lambda^*(t), N(t))] + \beta(\lambda-\lambda^*(t))\frac{\partial f(\lambda^*(t), N(t))}{\partial \lambda^*} \\
&= \beta(\lambda-\lambda^*(t))\frac{\partial f(x)}{\partial \lambda^*} + \lambda^*(t)[f(\lambda^*(t)+\alpha, N(t)+1) - f(x)]. \qquad (3.20)
\end{aligned}$$

where we use that $x = X(t) = (\lambda^*(t), N(t))$ and the definition of the partial derivative with respect to λ^*. The knowledge about the infinitesimal generator in (3.20) and the Dynkin formula in (3.19) allow to derive the following set of differential equations for the first two moments of the intensity and the number of jumps:

Lemma 3.13 (Da Fonseca and Zaatour 2013)**.** Given a Hawkes process $X(t) = (\lambda^*(t), N(t))$ with dynamics given by (3.16), $\mathbb{E}[N(t)], \mathbb{E}[\lambda^*(t)], \mathbb{E}[N(t)^2], \mathbb{E}[\lambda^*(t)^2]$ and $\mathbb{E}[\lambda^*(t)N(t)]$ satisfy the following set of ODE:

$$d\mathbb{E}[N(t)] = \mathbb{E}[\lambda^*(t)]dt \qquad (3.21)$$

$$d\mathbb{E}[\lambda^*(t)] = (\beta\lambda + (\alpha-\beta)\mathbb{E}[\lambda^*(t)])dt \qquad (3.22)$$

$$d\mathbb{E}[N(t)^2] = 2\mathbb{E}[\lambda^*(t)N(t)]dt + \mathbb{E}[\lambda^*(t)]dt \qquad (3.23)$$

$$d\mathbb{E}[\lambda^*(t)N(t)] = \beta\lambda\mathbb{E}[N(t)]dt + (\alpha-\beta)\mathbb{E}[\lambda^*(t)N(t)]dt + \mathbb{E}[\lambda^*(t)^2]dt + \alpha\mathbb{E}[\lambda^*(t)]dt \qquad (3.24)$$

$$d\mathbb{E}[\lambda^*(t)^2] = (\alpha^2+2\beta\lambda)\mathbb{E}[\lambda^*(t)]dt + 2(\alpha-\beta)\mathbb{E}[\lambda^*(t)^2]dt. \qquad (3.25)$$

Proof. The steps for deriving each differential equation are the same: Choose $f(X(t))$ suitably, write the infinitesimal generator using (3.20) and apply Dynkin's formula (3.19). Then apply Fubini's Theorem to exchange the order of integration and taking expectation and differentiate the resulting equation w.r.t t. We show these steps explicitly for (3.21) and (3.22) here, the rest follows analogously and can be found in Appendix B. To get (3.21), let $f(X(t)) \equiv N(t)$. Then the infinitesimal generator becomes

$$\begin{aligned}
\mathcal{L}f(X(t)) &= \beta(\lambda-\lambda^*(t))\frac{\partial f(X(t))}{\partial \lambda^*} + \lambda^*(t)[f(\lambda^*(t)+\alpha, N(t)+1) - f(\lambda^*(t), N(t))] \\
&= \beta(\lambda-\lambda^*(t))0 + \lambda^*(t)[N(t)+1-N(t)] = \lambda^*(t)
\end{aligned}$$

Dynkin's formula (with $s=t$ and $t=0$) yields

$$\mathbb{E}[N(t)] = \mathbb{E}[f(X(t))] = f(X(0)) + \mathbb{E}\left[\int_0^t \mathcal{L}f(X(u))du\right] = N(0) + \mathbb{E}\left[\int_0^t \lambda^*(u)du\right]$$

Exchanging integration and expectation and differentiating with respect to t yields (3.21):

$$\mathbb{E}[N(t)] = N(0) + \int_0^t \mathbb{E}[\lambda^*(u)]du$$
$$\Rightarrow d\mathbb{E}[N(t)] = \mathbb{E}[\lambda^*(t)]dt.$$

For (3.22), let $f(X(t)) \equiv \lambda^*(t)$, thus

$$\begin{aligned}\mathcal{L}f(X(t)) &= \beta(\lambda - \lambda^*(t))\frac{\partial f(X(t))}{\partial \lambda^*} + \lambda^*(t)[f(\lambda^*(t) + \alpha, N(t) + 1) - f(\lambda^*(t), N(t))] \\ &= \beta(\lambda - \lambda^*(t))1 + \lambda^*(t)[\lambda^*(t) + \alpha - \lambda^*(t)] = \beta\lambda + (\alpha - \beta)\lambda^*(t).\end{aligned}$$

Dynkin's formula (with $s = t$ and $t = 0$) and Fubini's Theorem yield

$$\mathbb{E}[\lambda^*(t)] = \lambda^*(0) + \mathbb{E}\bigg[\int_0^t (\beta\lambda + (\alpha - \beta)\lambda^*(u))du\bigg] = \lambda_0 + \beta\lambda t + (\alpha - \beta)\int_0^t \mathbb{E}[\lambda^*(u)]du.$$

Finally, differentiating gives

$$d\mathbb{E}[\lambda^*(t)] = (\beta\lambda + (\alpha - \beta)\mathbb{E}[\lambda^*(t)])dt.$$

The remaining parts of the proof are given in Appendix B. □

Lemma 3.13 allows to derive the first and second moment of the number of jumps of an exponential Hawkes process in its stationary regime for an interval of given length. The results are summarized in the following proposition.

Proposition 3.14 (Da Fonseca and Zaatour 2013)**.** Given a Hawkes process $X(t) = (\lambda^*(t), N(t))$ with dynamics given by (3.16), the long-run expected value of the number of jumps during an interval of length τ is given by:

$$\lim_{t\to\infty} \mathbb{E}[N(t + \tau) - N(t)] = \frac{\lambda}{1 - \alpha/\beta}\tau = \lim_{t\to\infty} \mathbb{E}[\lambda^*(t)]\tau. \tag{3.26}$$

The variance of the number of jumps is given by:

$$\begin{aligned}\mathbb{V}ar(\tau) &= \lim_{t\to\infty} \mathbb{E}[(N(t + \tau) - N(t))^2] - \mathbb{E}[N(t + \tau) - N(t)]^2 \\ &= \frac{\lambda}{1 - \alpha/\beta}\bigg(\tau\bigg(\frac{1}{1 - \alpha/\beta}\bigg)^2 + \bigg(1 - \bigg(\frac{1}{1 - \alpha/\beta}\bigg)^2\bigg)\frac{1 - e^{-\tau(\beta-\alpha)}}{\beta - \alpha}\bigg).\end{aligned} \tag{3.27}$$

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

$$\mathbb{C}ov(\tau, \delta) = \lim_{t\to\infty} \mathbb{E}[(N(t + \tau) - N(t))(N(t + 2\tau + \delta) - N(t + \tau + \delta))]$$

$$
\begin{aligned}
&- \mathbb{E}[N(t+\tau) - N(t)]E[N(t+2\tau+\delta) - N(t+\tau+\delta)] \\
&= \frac{\lambda\beta\alpha(2\beta-\alpha)(e^{(\alpha-\beta)\tau}-1)^2}{2(\alpha-\beta)^4} e^{(\alpha-\beta)\delta}. \qquad (3.28)
\end{aligned}
$$

Note again that taking the limit for $t \to \infty$ is equivalent to assuming that the process has reached its stationary regime with background intensity λ. Thus dependence with respect to the initial value λ_0 vanishes. Taking this limit requires the stability of the process, thus it is implied that $\alpha < \beta$.

Proof. The proof essentially relies on solving and combining the set of ODE derived in Lemma 3.13. We will demonstrate the calculation of the first moment $\mathbb{E}[N(t+\tau) - N(t)]$ here, the rest follows analogously from the other equations and the details can be found in Appendix C or in Da Fonseca and Zaatour 2013. Given the initial condition $\mathbb{E}[\lambda^*(0)] = \lambda_0$, we can solve the differential equation (3.22) by applying standard techniques: Multiplying both sides of the equation with the integrating factor $e^{(\beta-\alpha)t}$ and rearranging terms gives

$$
\begin{aligned}
e^{(\beta-\alpha)t} d\mathbb{E}[\lambda^*(t)] + (\beta-\alpha)e^{(\beta-\alpha)t}\mathbb{E}[\lambda^*(t)]dt &= \beta\lambda e^{(\beta-\alpha)t}dt \\
d\big(e^{(\beta-\alpha)t}\mathbb{E}[\lambda^*(t)]\big) &= \beta\lambda e^{(\beta-\alpha)t}dt
\end{aligned}
$$

Integrating from 0 to t and inserting the initial condition leads to

$$
e^{(\beta-\alpha)t}\mathbb{E}[\lambda^*(t)] - 1\lambda_0 = \frac{\beta\lambda}{\beta-\alpha}\big(e^{(\beta-\alpha)t} - 1\big).
$$

Finally, solving for $\mathbb{E}[\lambda^*(t)]$ gives

$$
\mathbb{E}[\lambda^*(t)] = \frac{\beta\lambda}{\alpha-\beta}\big(e^{(\alpha-\beta)t} - 1\big) + \lambda_0 e^{(\alpha-\beta)t}.
$$

Inserting this result into the integral form of (3.21) gives

$$
\begin{aligned}
\mathbb{E}[N(t)] &= N(0) + \int_0^t \left(\frac{\beta\lambda}{\alpha-\beta}\big(e^{(\alpha-\beta)s} - 1\big) + \lambda_0 e^{(\alpha-\beta)s}\right) ds \\
&= N(0) + \frac{\beta\lambda}{\alpha-\beta}\left[\frac{1}{\alpha-\beta}e^{(\alpha-\beta)s}\right]_0^t - \frac{\beta\lambda t}{\alpha-\beta} + \frac{\lambda_0}{\alpha-\beta}\left[e^{(\alpha-\beta)s}\right]_0^t \\
&= N(0) + \frac{\lambda\beta(e^{(\alpha-\beta)t} - 1 - (\alpha-\beta)t)}{(\alpha-\beta)^2} + \lambda_0\frac{e^{(\alpha-\beta)t}-1}{\alpha-\beta}.
\end{aligned}
$$

Thus, the expected number of jumps during an interval of length τ is given by

$$
\begin{aligned}
&\mathbb{E}[N(t+\tau) - N(t)] \\
&= \lambda_0\frac{e^{(\alpha-\beta)(t+\tau)} - e^{(\alpha-\beta)t}}{\alpha-\beta} + \frac{\beta\lambda}{(\alpha-\beta)^2}\left(e^{(\alpha-\beta)(t+\tau)} - 1 - (\alpha-\beta)(t+\tau) - e^{(\alpha-\beta)t} + 1 + (\alpha-\beta)t\right)
\end{aligned}
$$

$$= \frac{\lambda_0}{\alpha-\beta} e^{(\alpha-\beta)t}\big(e^{(\alpha-\beta)\tau}-1\big) + \frac{\beta\lambda}{(\alpha-\beta)^2}\Big(-(\alpha-\beta)\tau + e^{(\alpha-\beta)(t+\tau)} - e^{(\alpha-\beta)t}\Big)$$
$$= \frac{-\tau\lambda\beta}{\alpha-\beta} + e^{(\alpha-\beta)t}\frac{\lambda_0(\alpha-\beta)e^{(\alpha-\beta)\tau} - \lambda_0(\alpha-\beta) + \lambda\beta e^{(\alpha-\beta)\tau} - \lambda\beta}{(\alpha-\beta)^2}.$$

Now considering $\lim_{t\to\infty}$ makes the second term vanish (as $\lim_{t\to\infty} e^{(\alpha-\beta)t} = 0$ for $\alpha < \beta$) and thus eliminates the dependence on the initial value λ_0. Therefore in the stationary regime the expected number of jumps in an interval of length τ is given by

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

The other results of the proposition require similar, albeit more lengthy, computations involving the other ODEs from Lemma 3.13 and will not be detailed here. □

For empirical applications, it might be more convenient to work with the autocorrelation function of the number of jumps instead of the autocovariance.

Proposition 3.15 (Da Fonseca and Zaatour 2013)**.** Given a Hawkes process $X(t) = (\lambda^*(t), N(t))$ with dynamic given by (3.16) the autocorrelation function of the number of jumps on two intervals of length τ separated by a time lag of $\delta > 0$ is

$$Acf(\tau,\delta) = \frac{e^{-2\beta\tau}(e^{\alpha\tau}-e^{\beta\tau})^2\alpha(\alpha-2\beta)}{2(\alpha(\alpha-2\beta)(e^{(\alpha-\beta)\tau}-1)+\beta^2\tau(\alpha-\beta))} e^{(\alpha-\beta)\delta}. \tag{3.29}$$

Note that the above expression is always positive for $\alpha < \beta$, independent of the background intensity λ and exponentially decaying with the lag δ. This emphasizes the typical feature of an exponential Hawkes process, namely that the influence of events which have occurred in the past decays exponentially as time passes.

Proof. This follows straightforward from (3.28) and (3.27) in Proposition 3.14. □

These results derived by Da Fonseca and Zaatour 2013 make working with an exponential Hawkes process for empirical data very convenient, as it is now possible to theoretically compute the moments and autocorrelation of the number of jumps over intervals of given length and compare them with the empirical values for the observed arrival process. In particular testing for the typical exponentially decaying form of the autocorrelation of the number of jumps between intervals as a function of the time lag can serve as a first indicator of the *self-exciting* characteristics of an observed data set. Furthermore, comparing empirical and theoretical moments can serve as another measure of the goodness of fit of a Hawkes model. This is to say that in most cases the well-known goodness of

fit measures will show some imperfections in the model fit for empirical data, as is to be expected if the data does not by chance happen to closely follow the quite restrictive model specification of an exponential Hawkes process. Even in this case, the model might be suitable to describe the pattern of event occurrences in the data set sufficiently well by providing close estimates for the expected value, variance and autocorrelation of the number of events in given intervals. For the exponential Hawkes process, we are now equipped with the theoretical knowledge to assess the model fit given these characteristics and will do so in Chapter 6 for an insurance data set.

3.6 More general Hawkes processes

In the last sections, we have defined a one-dimensional Hawkes process, reviewed its properties and explained why the rest of this thesis will focus on a Hawkes process with exponential decay. This restriction is suitable for the scope of this thesis as it is quite easily understandable and its analytical tractability is very convenient for applications with empirical data. However, we would like to emphasize that self-exciting Hawkes processes in general are a very wide and versatile class of point processes and much recent work has been devoted to more general and arbitrarily sophisticated Hawkes models. We will introduce some selected generalisations here as to not deviate too much from the focus of this thesis, but to give a glimpse into the wide spectrum of Hawkes process applications available.

3.6.1 Multi-dimensional Hawkes processes

Of course the one-dimensional Hawkes process introduced in Definition 3.5 can be extended to a multivariate version:

Definition 3.16 (Multi-dimensional Hawkes process (Bacry, Mastromatteo, and Muzy 2015))**.** A *D-dimensional Hawkes process* is a D-variate counting process $N(t) = \{N^i(t)\}_{i=1}^D$ whose associated intensity vector $\lambda^*(t) = \{\lambda^{i*}(t)\}_{i=1}^D$ can be written as

$$\lambda^{i*}(t) = \lambda^i + \sum_{j=1}^{D} \int_0^t \mu^{ij}(t-u) dN^j(u)$$

where $\lambda = \{\lambda^i\}_{i=1}^D$ is a vector of (exogenous) background intensities, and $\Phi(t) = \{\mu^{ij}(t)\}_{i,j=1}^D$ is a matrix-valued kernel which is component-wise positive ($\mu^{ij}(t) \geq 0$ for each $1 \leq i, j \leq D$), component-wise causal (if $t < 0$, $\mu^{ij}(t) = 0$ for each $1 \leq i, j \leq D$) and each component $\mu^{ij}(t)$ belongs to the space of L^1-integrable functions.

Analogously to above, the intensity can be written using the past event times $\{t_1, \cdots, t_k\}$. In this case, for each event time it has to be indicated which component of the process jumped at this time; this can be done by using the couples $\{(t_m, c_m)\}_{m=1}^k$, where t_m indicates the event time and $c_m \in [1, \cdots, D]$ indicates the component. Thus, the conditional intensity at time t can be written as

$$\lambda^{i*}(t) = \lambda^i + \sum_{m=1}^{k} \mu^{i,c_m}(t - t_m). \tag{3.30}$$

Figure 5, taken from Bacry, Mastromatteo, and Muzy 2015, shows an example realisation of a multivariate Hawkes process, where different colours are used to indicate the components.

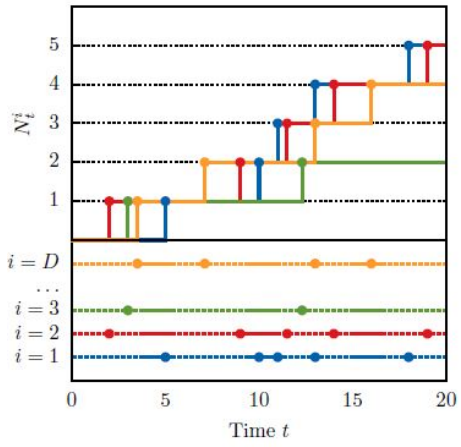


Figure 5: An example realisation of a multivariate Hawkes process where the lower part shows the point process for each component (arrival times are marked as dots), and the upper part displays the corresponding counting processes.

Of course, a clustering representation of the Hawkes process can be derived analogously in the multivariate case, see Proposition 6 of Bacry, Mastromatteo, and Muzy 2015. In the multi-dimensional case, the stationarity condition (3.6) generalizes to

Proposition 3.17 (Bacry, Mastromatteo, and Muzy 2015, Proposition 2)**.** The process $N(t)$ has asymptotically stationary increments and $\lambda^*(t)$ is asymptotically stationary if its matrix kernel has spectral radius[1] $\|\Phi\| = \{\|\mu^{ij}\|\}_{i,j=1}^D$ smaller than 1.

Example (Two-dimensional exponential Hawkes process)**.** Consider a bivariate Hawkes process with kernel matrix

$$\Phi(t) = \begin{pmatrix} \mu^{(s)}(t) & \mu^{(c)}(t) \\ \mu^{(c)}(t) & \mu^{(s)}(t) \end{pmatrix} = \begin{pmatrix} \alpha^{(s)} e^{-\beta^{(s)} t} & \alpha^{(c)} e^{-\beta^{(c)} t} \\ \alpha^{(c)} e^{-\beta^{(c)} t} & \alpha^{(s)} e^{-\beta^{(s)} t} \end{pmatrix} \qquad \forall t > 0.$$

[1]The spectral radius of a square matrix is defined as the largest absolute value of its eigenvalues, see also Appendix D.

The parameters $\alpha^{(\cdot)}$ and $\beta^{(\cdot)}$ can generally be interpreted analogously to the one-dimensional case. However, now the process has two components which are both *self-exciting* and *mutually exciting*. This means that an incoming event both influences the intensity of its own component (according to the parameters $\alpha^{(s)}$ and $\beta^{(s)}$, where the superscript (s) indicates *s*elf-excitement) as well as the other component (via the parameters $\alpha^{(c)}$ and $\beta^{(c)}$, where the superscript (c) indicates *c*ross-excitement). Due to the symmetry of the kernel matrix, the stability condition in this case is simply given by (see Appendix D)

$$\|\Phi\| = \|\mu^{(s)}\| + \|\mu^{(c)}\| = \frac{\alpha^{(s)}}{\beta^{(s)}} + \frac{\alpha^{(c)}}{\beta^{(c)}} < 1.$$

In Bacry et al. 2013, this specification is used to describe the upward and downward movements of an asset price at the tick-by-tick level in order to study microstructure noise. As they assume the price at microstructure level to display a mean-reverting behaviour, they consider a *purely mean-reverting* scenario, that is $\mu^{(s)}(t) = 0$ and $\mu^{(c)}(t)$ as above. This means that instead of self-excitement, we observe mutual *cross-excitement* between the processes of upward and downward jumps respectively. The conditional intensities for upward and downward jumps are given as

$$\lambda_u^*(t) = \lambda + \int_0^t \alpha^{(c)} e^{-\beta^{(c)}(t-s)} dN_d(s),$$

$$\lambda_d^*(t) = \lambda + \int_0^t \alpha^{(c)} e^{-\beta^{(c)}(t-s)} dN_u(s).$$

The arrival of an upward price jump at time s increases the intensity of a downward jump via the cross-excitement term $\alpha^{(c)} e^{-\beta^{(c)}(t-s)}$ and vice versa, thus leading to a self-regulating behaviour of the price. The opposite case of a *purely trend-following* scenario where $\mu^{(c)}(t) = 0$ and $\mu^{(s)}(t)$ is considered in Da Fonseca and Zaatour 2013. However, this choice essentially describes two independent self-exciting processes and does not consider any mutual excitement (which could be considered the most interesting feature of a multi-dimensional specification in the first place). As an extension of their model, Bacry et al. 2013 even consider an analogous 4-dimensional model to describe the joint price dynamics of a pair of assets.

Other applications of multi-dimensional Hawkes processes include Aït-Sahalia, Cacho-Diaz, and Laeven 2015 who use such a model to describe the dynamics of asset returns in geographically different markets across the world. A jump in one region increases the intensity of both jumps in the same region (self-excitement) as well as other regions (cross-excitement) leading to periods of highly clustered jumps across world markets imitating

periods of crises characterized by contagion. Embrechts, Liniger, and Lin 2011 use a multivariate Hawkes model to analyze daily stock market index data, where the arrivals describe extreme positive (above the 90%-quantile) and negative (below the 10%-quantile) exceedances of the daily log-return.

3.6.2 Marked Hawkes processes

Another feature that can be incorporated into a point process are *marks* which can be used to differentiate different types of events. For example in a series of earthquake occurences, it might be of interest to describe not only the occurrence time, but also the magnitude of the earthquake as this has a substantial influence on the development of aftershocks. Likewise, in a model describing credit default events, it might be sensible to differentiate between defaults caused by temporary illiquidity (where the firm has a chance of recovering in the future) and defaults caused by insolvency. This motivates defining a *marked multivariate Hawkes process* as a sequence of event times, components and marks $\{(t_m, c_m, \xi_m)\}_{m=1}^k$, where usually the marks ξ_m are supposed to be i.i.d. random variables drawn with each event time. It is assumed that events with different marks influence the conditional intensity differently, thus (3.30) becomes

$$\lambda^{i*}(t) = \lambda^i + \sum_{m=1}^{k} \mu^{i,c_m}(t - t_m, \xi_m).$$

Typically, it is assumed that the kernel can be factorized into the self-/cross-excitement of the processes and the effect of the marks, i.e. $\mu^{ij}(t, \xi) = \mu^{ij}(t)\chi^{ij}(\xi)$.
Such a model was originally used for earthquakes times and magnitudes in the ETAS (epidemic type aftershock sequence) model by Ogata 1988. As to finance applications, Fauth and Tudor 2012 use a 4-dimensional marked Hawkes model to describe the increase/decrease of the bid/ask price where each incoming transaction is marked by its transaction volume. Embrechts, Liniger, and Lin 2011 extended their work on multivariate Hawkes models to analyze stock market data by using a marked three-dimensional model to study an equally weighted portfolio of three indices, where events are considered to be days with very small (below 1%-quantile) or very large (above 99%-quantile) returns, and the marks are used to code the excess of each index w.r.t. its quantile.

For more details on applications of Hawkes processes in finance, Bacry, Mastromatteo, and Muzy 2015 provides a very comprehensive overview.

Chapter 4

Risk theory

In this chapter, we will recapitulate some basic definitions and results of classical risk theory and briefly introduce some extensions of the classical Cramer-Lundberg risk model. This will give us the necessary background knowledge in order to combine risk theory with Hawkes processes in the subsequent chapter and relate the results obtained therein to classical ones. This chapter is based on Schmidli 2018 and Rolski 2001 who provide very extensive and recommendable reads for an introduction to risk theory.
The business of an insurance company is usually based on a multitude of different portfolios, such as policies covering theft, fire, automobile or earthquake damage. Each portfolio is characterized by a different source of *insurance risk* and has deterministic as well as stochastic characteristics, such as:

- The observed time period $[0, T]$, where the time horizon T depends on the planning or bookkeeping period of the company, usually 1 to 5 years.
- The *initial capital* or *initial reserve* u of a portfolio can be interpreted as the amount of capital that is set aside at time 0 to cover expenses from this portfolio until premium payments have been received. As we will see later, it can also be understood as a risk buffer determined by regulatory requirements.
- The premium income $C(t)$ describes the total premium amount earned from the portfolio during the time period $[0, t]$. Setting an adequate premium rate for each policy is one of the most crucial tasks of any insurance company and we will see the principles that come into play here shortly.
- The number of claims up to time t, described by a counting process $N(t)$ or equivalenty by a sequence of claim arrival times $0 \leq t_1 < t_2 < \cdots < t_k \leq T$. Of course, this is a random process, thus choosing the right stochastic model to describe the

arrivals of incoming claims is crucial in order to estimate risk correctly.

- Each incoming claim, say at time t_i, has a claim size Y_i. The sequence of claim sizes $\{Y_i\}$ is often assumed to be formed of i.i.d. random variables. Thus the *aggregate claim amount* $\sum_{i=1}^{N(t)} Y_i$ is a random sum of random variables that captures all the stochastic aspects of the risk process.

In the following, we will introduce how the above concepts are formalized in a risk model.

4.1 The classical risk model

4.1.1 Definition and notation

First, it is natural to assume that the whole risk of a portfolio can be split into several independent single risks, e.g. for a portfolio of car insurances, each policy covers one individual customer who might be involved in an accident. Equivalently, the time interval $[0, T]$ can be split into independent, equal smaller intervals such that there can be at most one claim per policy and interval. If the probability of a claim occuring is then p, the number of claims per policy is then a Bernoulli random variable and the number of claims of the whole portfolio of n individual contracts is $N \sim Bin(n, p)$. Now assume that the number of policies in the portfolio (n) is rather large and the probability of a claim occuring (p) is rather small. Let $\lambda := np$, then it follows

$$N \sim Bin(n, p) = Bin(n, \lambda/n) \overset{n\to\infty}{\longrightarrow} Poi(\lambda).$$

This intuition naturally motivated Filip Lundberg in his fundamental work Lundberg 1903 to introcude a risk model where the number of incoming claims is decribed by a Poisson process. Furthermore, assume that the premium payments are spread over the whole time period and can thus be approximated as a continuous income stream for a large portfolio of policies. This model, also known as the *Cramer-Lundberg process*, is the foundation of modern risk theory and almost any more complicated model can be traced back to it. Rolski 2001 emphasizes the importance of the Poisson process in risk theory by stating that the exponential distribution (which is intrinsically linked to the Poisson process) *"plays a similar crucial role in actuarial applications as the normal distribution does in statistics"*.

Definition 4.1 (Classical risk process)**.** The *classical risk process* describes the surplus of an insurance portfolio as

$$R(t) = u + ct - \sum_{i=1}^{N(t)} Y_i$$

where u denotes the initial capital, c denotes the (continuous) premium rate and the number of claims in the interval $[0, t)$ is a homogeneous Poisson process $N(t)$ with rate λ. The claims Y_i are i.i.d. positive random variables with distribution function G and mean m_1 independent of $N(t)$.

Remark (Compound Poisson process)**.** Note that the random sum of random variables $\sum_{i=1}^{N_t} Y_i$, where $N(t)$ is a Poisson process and Y_i is an i.i.d. sequence of random variables is called a *compound Poisson process*.

For an insurance company, probably the most important consideration is to ensure that the surplus $R(t)$ does not fall below a certain critical level, w.l.o.g. 0.

Definition 4.2 (Ruin time)**.** The *ruin time* is defined as

$$\tau = \inf\{t > 0 : R(t) < 0\}$$

where $inf\emptyset = \infty$.

Definition 4.3 (Probability of ruin)**.** The *probability of ruin* in a finite time interval $(0, T]$ is defined as

$$\Psi(u, t) = \mathbb{P}(\tau \leq t | R(0) = u) = \mathbb{P}(\inf_{0<s\leq t} R(s) < 0 | R(0) = u) \tag{4.1}$$

and the *probability of ultimate ruin* as

$$\Psi(u) = \lim_{t\to\infty} \Psi(u, t) = \mathbb{P}(\inf_{t>0} R(t) < 0 | R(0) = u). \tag{4.2}$$

Conversely, $\delta(u) = 1 - \Psi(u)$ is called *survival probability*.

Note that the word "ruin" is only a technical term as it does not refer to actual ruin in reality for two reasons: In practice, the critical limit would never be set as 0, but rather a certain level given by regulatory requirements. If the surplus of the portfolio decreased too close towards the critical level, the company would react before the level was actually breached (e.g. by raising premiums). Furthermore, the event of "ruin" in one portfolio does not equate to "bankruptcy" of the whole company which is usually built on a diversified set of various portfolios. Therefore, the probability of ruin does not refer to an actual positive probability of bankruptcy but should rather be seen as a useful tool in decision making, i.e. calculating premiums, reserving capital as risk buffers for certain portfolios or determining reinsurance levels. Nonetheless, we will use the terms "ruin" and "probability of ruin" without further comment in the remainder of this thesis. An example realisation of a classical risk process $R(t)$ is given in Figure 6.

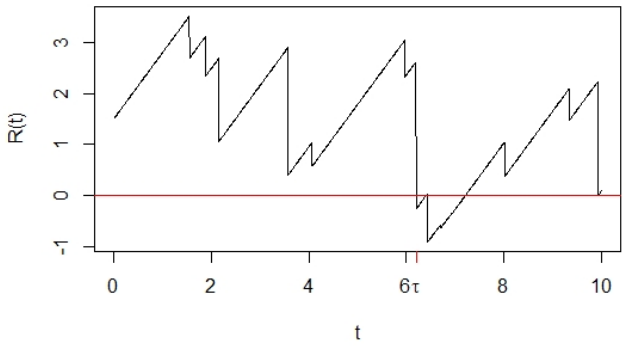


Figure 6: An example realisation of the classical risk model over the time period $[0, T] = [0, 10]$. The initial capital is set as $u = 1.5$, the premium rate as $c = 1.3$ and incoming claims follow a Poisson process with rate $\lambda = 1$ and their sizes are i.i.d. $\sim Exp(\gamma)$ with $\gamma = 1$. In this example, ruin occurs within the observed time interval, namely at $\tau \approx 6.2$. One can either consider the process as terminated once ruin occurs, or one might consider that the company might be able to carry on with the portfolio through borrowing from other branches of the business until the portfolio becomes profitable again or until it has been restructured.

We start with the question of how the premium for a portfolio should be set. Clearly, it should be the primary interest of an insurance company to avoid a situation where ruin occurs with probability 1. It holds:

Lemma 4.4 (Net Profit Condition (adapted from Schmidli 2018, p. 21))**.** Let $R(t) = u + ct - \sum_{i=1}^{N_t} Y_i$ be given as in Definition 4.1. Then

- If $c < \lambda m_1$, then $R(t) \xrightarrow{t\to\infty} -\infty$ a.s. That means ruin is unavoidable, i.e. $\Psi(u) = 1$.
- If $c = \lambda m_1$, then $\limsup_{t\to\infty} R(t) = -\liminf_{t\to\infty} R(t) = \infty$, also implying $\Psi(u) = 1$.
- If $c > \lambda m_1$, then $R(t) \xrightarrow{t\to\infty} \infty$ a.s. and $\mathbb{P}(R(t) \geq 0 \quad \forall t) > 0$.

Proof. The first and third statement follow directly from the law of large numbers. Consider

$$\begin{aligned}
\lim_{t\to\infty} \frac{R(t)}{t} &= \lim_{t\to\infty} \left(\frac{u}{t} + c - \frac{1}{t} \sum_{i=1}^{N(t)} Y_i \right) \\
&= c - \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{N(t)} Y_i \\
&= c - \lim_{t\to\infty} \frac{N(t)}{t} \cdot \frac{1}{N(t)} \sum_{i=1}^{N(t)} Y_i
\end{aligned}$$

Applying the law of large numbers to each factor yields

$$\lim_{t\to\infty} \frac{R(t)}{t} = c - \lambda m_1 \text{ a.s.},$$

which imply the statements of the lemma.
The second statement requires some deep theory of random walks (e.g. Rolski 2001, p. 234f) and will not be detailed here. □

This obviously implies that the premium rate c must be chosen such that $c > \lambda m_1$, this is called the *net profit condition*. There are many approaches to setting the premium, the easiest and most commonly used is the following:

Definition 4.5 (Expected value principle (adapted from Schmidli 2018, p. 21))**.** The *expected value principle* sets the premium rate c as

$$c = (1+\theta)\lambda m_1,$$

where $\theta > 0$ is called the *safety loading*.

In general, finding the ruin probability is a non-trivial problem that can be solved explicitly only for a few choices of distribution function G even in the (simplest) case of the classical risk model. In general, one can find that $\Psi(u)$ fulfils the following defective renewal equation

Theorem 4.6 (Rolski 2001, p. 164)**.**

$$\Psi(u) = \frac{\lambda}{c}\left(\int_u^\infty (1-G(x))dx + \int_0^u \Psi(u-x)(1-G(x))dx\right), \tag{4.3}$$
$$\Psi(0) = \frac{\lambda m_1}{c}.$$

Example. If claim sizes are assumed to be i.i.d. $\sim Exp(\gamma)$, the ruin probability can be explicitly calculated as

$$\Psi(u) = \frac{\lambda}{c\gamma} e^{(\frac{\lambda}{c}-\gamma)u}.$$

This convenience is one of the reasons why the prime example of light-tailed claims is often a sequence of i.i.d. $\sim Exp(\gamma)$ random variables.

As (4.3) can unfortunately not be solved explicitly for general choices of G, a substantial amount of research has been devoted to finding bounds and approximations for the ruin probability. Lundberg derived the *adjustment coefficient*, which allows finding an upper bound on $\Psi(u)$ (called *Lundberg's inequality*) and deriving the *Cramer-Lundberg approximation* on $\Psi(u)$ for $u \to \infty$ (e.g. Schmidli 2018, p. 92ff). Alternatively, for a specific model, the ruin probability could of course always be estimated numerically by Monte-Carlo simulation. Another approach is to find a suitable approximation on $R(t)$ for which the ruin probability can be computed (or approximated better) and using the result as an approximation for $\Psi(u)$ in the original model. This will be detailed in the following section.

4.1.2 Diffusion approximation

In this section, we will introduce the diffusion approximation for the classical risk model. The idea is to contruct a sequence of classical risk processes that converges weakly to a *diffusion limit*. We will see that in this case the limit is a Brownian motion, which is very convenient as it gives access to many well-known results and applications.

Proposition 4.7 (Schmidli 2018, p. 108f)**.** Let $\{R^{(n)}(t)\}$ be a sequence of Cramer-Lundberg processes with initial capital u, claim arrival rates $\lambda^{(n)} = n\lambda$, claim sizes $\{Y_i^{(n)}\}$ with distributions $G^{(n)}(x) = G(\sqrt{n}x)$ (and mean $\mathbb{E}[Y_1^{(n)}] = \frac{m_1}{\sqrt{n}} =: m_1^{(n)}$) and premium rates

$$c^{(n)} = \left(1 + \frac{c - \lambda m_1}{\lambda m_1 \sqrt{n}}\right)\lambda^{(n)} m_1^{(n)} = c + (\sqrt{n} - 1)\lambda m_1.$$

Let $m_1 = \mathbb{E}[Y_1] = \int_0^\infty y dG(y)$ and assume that $m_2 = \mathbb{E}[Y_1^2] = \int_0^\infty y^2 dG(y) < \infty$.
Then

$$R^{(n)}(t) \xrightarrow{d} (u + W(t)) \tag{4.4}$$

where $W(t)$ is a $(c - \lambda m_1, \lambda m_2)$-Brownian motion and $\xrightarrow{d}$ denotes convergence in distribution.

Proof. See Grandell 1977. □

The intuitive idea is that we start with the risk process $R(t) = R^{(1)}(t)$ and then increase the average number of claims while decreasing their average size in a way that the expectation stays at $\mathbb{E}[R^{(n)}(t) - u] = (c - \lambda m_1)t$ and the distribution of $R^{(n)}(t) - u$ tends to a normal distribution. The procedure is illustrated in Figure 7.

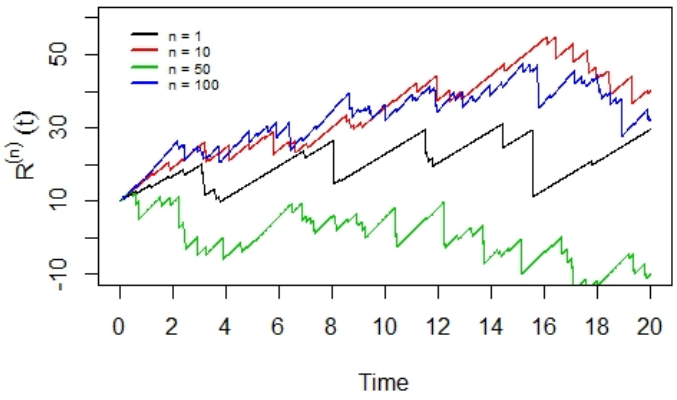


Figure 7: Illustration of the idea of approximating the classical risk model by a diffusion process, where the original risk process $R^{(1)}(t)$ has $u = 10$, $\theta = 0.1$ and claim arrivals following a Poisson process with rate $\lambda = 1$ and claim sizes $\sim Exp(0.25)$. For different values of n, the plots show realisations of $R^{(n)}(t)$ over the period $[0, 20]$. We can clearly see that with increasing n, the downward jumps representing claim arrivals get more frequent while their sizes decrease such that the risk process starts resembling a diffusion process.

Remark.

- By the *Wald equation*, for a compound Poisson process it holds

$$\mathbb{E}\left[\sum_{i=1}^{N(t)} Y_i\right] = \mathbb{E}[N(t)]\mathbb{E}[Y_1] = \lambda t m_1,$$
$$\mathbb{V}ar\left[\sum_{i=1}^{N(t)} Y_i\right] = \lambda t m_2,$$

  where $N(t)$ follows a Poisson process with rate λ and $\{Y_i\}$ are i.i.d. r.v. with distribution G.

- The claim sizes $\{Y_i\}$ and $\{Y_i^{(n)}\}$ have distribution functions $G(x)$ and $G^{(n)}(x) = G(\sqrt{n}x)$ respectively and thus belong to a *location-scale family*. Therefore $Y_1^{(n)} \overset{d}{=} \frac{Y_1}{\sqrt{n}}$, $\mathbb{E}[Y_1^{(n)}] = \frac{m_1}{\sqrt{n}}$ and $\mathbb{E}[(Y_1^{(n)})^2] = \frac{m_2}{n}$.

- Thus it holds

$$\mathbb{E}[R^{(n)}(t) - u] = c^{(n)}t - \mathbb{E}\left[\sum_{i=1}^{N^{(n)}(t)} Y_i\right] = ct + \lambda m_1 t(\sqrt{n} - 1) - \lambda nt\frac{m_1}{\sqrt{n}}$$
$$= (c - \lambda m_1)t$$
$$\mathbb{V}ar[R^{(n)}(t) - u] = \mathbb{V}ar\left[\sum_{i=1}^{N^{(n)}(t)} Y_i\right] = \lambda nt\mathbb{E}[(Y_1^{(n)})^2] = \lambda nt\frac{m_2}{n} = \lambda t m_2$$

- By subtracting u and $(c-\lambda m_1)t$ and dividing by $\sqrt{\lambda m_2 t}$ in (4.4), we obtain

$$\frac{(R^{(n)}(t)-u)-(c-\lambda m_1)t}{\sqrt{\lambda m_2 t}} \xrightarrow{d} W(t)$$

  where $W(t)$ is a standard $(0,1)$-Brownian motion and we can see the resemblance to the standard central limit theorem.

The following proposition tells us that we can use the ruin probability of the diffusion limit as an approximation for the ruin probability of the original model.

Proposition 4.8 (Schmidli 2018, p. 109)**.** Let $R^{(n)}(t)$ and $W(t)$ be as above and let $\tau^{(n)}$ denote the ruin time of $\{R^{(n)}(t)\}$ and $\tau = inf\{t \geq 0 : u + W(t) < 0\}$ the ruin time of the diffusion limit. Then

$$\lim_{n\to\infty} \mathbb{P}[\tau^{(n)} \leq t] = \mathbb{P}[\tau \leq t]$$

and

$$\lim_{n\to\infty} \mathbb{P}[\tau^{(n)} \leq \infty] = \mathbb{P}[\tau \leq \infty]$$

Proof. The first result (finite time horizon) is a special case of Whitt 1970, Theorem 9 (see also Grandell 1977). The second result (infinite time horizon) can be found in Schmidli 1994. □

Thus it remains to find the ruin probability for the diffusion limit, which can be done by applying well-known results for the Brownian motion.

Theorem 4.9 (Adapted from Schmidli 2018, p. 109)**.** The finite-time and infinite-time ruin probabilities for the classical Poisson risk model are given by

$$\mathbb{P}(\tau \leq t) = 1 - \Phi\left(\frac{(c-\lambda m_1)t+u}{\sqrt{\lambda m_2 t}}\right) + e^{\frac{-2u(c-\lambda m_1)}{\lambda m_2}} \Phi\left(\frac{(c-\lambda m_1)t-u}{\sqrt{\lambda m_2 t}}\right) \tag{4.5}$$

and

$$\mathbb{P}(\tau < \infty) = e^{\frac{-2u(c-\lambda m_1)}{\lambda m_2}}. \tag{4.6}$$

Proof. This follows directly from Theorem 2.28 with $\mu = c - \lambda m_1$ and $\sigma = \sqrt{\lambda m_2}$. □

It has to be remarked that the accuracy of such diffusion approximations always depends on the parameters of the model, e.g. in the above case the approximation is only accurate if $c \approx \lambda m_1$, i.e. the safety loading is relatively small. There exist some approaches for corrected diffusion approximations, e.g. Asmussen 1984. At the same time, it is obviously only possible to give exact evaluations about the accuracy of the approximation in the

case where the ruin probability can be computed exactly (that is, for $Exp(\gamma)$ claims in the classical model) and thus the approximation is redundant. In any case, a diffusion approximation can be a very convenient tool for many applications if one keeps in mind that it stays an *approximation* whose accuracy depends on the model and its specific parameters.

4.2 Extensions of the classical model

In the last section, we have introduced the classical risk model which serves as the foundation of actuarial risk theory. Modelling the sum of incoming claims as a compound Poisson process is theoretically convenient and many results can only be obtained in closed form for this basic model. Naturally the model makes many simplifying assumptions which might not hold true in reality. Thus, ever since its introduction, a lot of work has been devoted to adding more sophisticated features into the classical model in order to include phenomena from reality. We will give a short overview of the scope of more general risk models here.

Maybe the most natural generalisation of the classical model is to replace the Poisson process for claim arrivals by a more general process, e.g. a general *renewal* process. The resulting risk model is then called *renewal risk model* or *Sparre Andersen model* as it was first introduced in Andersen 1957.

Definition 4.10 (Renewal risk model)**.** The *renewal risk model* describes the surplus of an insurance portfolio as

$$R(t) = u + ct - \sum_{i=1}^{N(t)} Y_i$$

where u denotes the initial capital, c denotes the (continuous) premium rate and the number of claims in the interval $[0, t)$ is a renewal process $N(t)$ with arrival times $\{t_1, t_2, \cdots\}$ whose interarrival times have distribution function F. The claims $\{Y_i\}$ are i.i.d. positive random variables with distribution function G and mean m_1 independent of $N(t)$.

This generalizes the classical model in the sense that the lengths of periods between arrivals are still independent, but the distribution of the *time* of the next claim is dependent on the time of the last claim. As stated in Asmussen 2010 (p. 132f), in reality the mechanism generating a renewal arrival process for claims is hard to understand and therefore the relevance of the renewal risk model has been questioned. However, one tangible example of a renewal process would be an environment with two states and Markovian switching between them, where each state induces Poisson arrivals with a different rate.

A more commonly used approach is to generalize the homogeneous Poisson process $N(t)$ by introducing a time-dependent intensity $\lambda(t)$. For the simple case that $\lambda(t)$ is deterministic, this leads to an inhomogeneous Poisson process. This could be suitable to describe e.g. a situation in car insurance, where the number of accidents depends on the time during the year or the week (see Grandell 1991). A more interesting and commonly used setting is considering a Poisson process with a *stochastic* (time-dependent) intensity $\lambda(t)$, a so-called *doubly stochastic Poisson process* or *Cox process*. This "double randomization" allows for more flexibility:

Definition 4.11. The *Cox risk model* describes the surplus of an insurance portfolio as

$$R(t) = u + ct - \sum_{i=1}^{N(t)} Y_i$$

where u denotes the initial capital, c denotes the (continuous) premium rate and the number of claims in the interval $[0, t)$ is a counting process $N(t)$ with stochastic intensity $\lambda(t)$. The claims $\{Y_i\}$ are i.i.d. positive random variables with distribution function G and mean m_1 independent of $N(t)$.

The most common specification for $\lambda(t)$ is a *Poisson shot noise intensity* given by

$$\lambda(t) = \lambda + \sum_{s_i<t} h(t - s_i, Z_i) + \nu(t)$$

where $\{s_i\}$ are the arrivals of a homogeneous Poisson process with rate γ, $\{Z_i\}$ are i.i.d. random variables with distribution function F independent of $\{s_i\}$ and $h(t, x)$ is a non-negative function. $\nu(t) \geq 0$ is a stochastic term independent of $h(t - s_i, Z_i)$ which could represent some initial conditions or a noise term, but is likewise often left out for modelling purposes. If we consider so-called multiplicative shots, $\lambda(t)$ takes the form

$$\lambda(t) = \lambda + \sum_{s_i<t} Z_i \cdot h(t - s_i) + \nu(t) \tag{4.7}$$

where everything is as above and $h(t)$ is again a non-negative function. This allows for a more intuitive interpretation of the model, as used in Dassios and Jang 2003 to price catastrophe reinsurance: There is a stream of "normal" claims following a homogeneous Poisson process with rate λ, but from time to time an external event such as a natural catastrophe occurs (following a Poisson process with rate γ and arrival times $\{s_i\}$) and causes a dramatic increase in the number of claims. The random variables $\{Z_i\}$ describe the severity of the catastrophes. As it might take some time to report and settle all those "extraordinary" claims, the increase in overall intensity caused by an event at time s_i develops according to $h(t - s_i)$. Albrecher and Asmussen 2006 study aggregate claims

distributions and asymptotic estimates for ruin probabilites for the same Cox model. As another application, modelling the claim arrival process directly as a Cox process is often used to incorporate delayed claims settlement into the risk model (see Klüppelberg, Mikosch, and Kluppelberg 1995, Macci and Torrisi 2004, Brémaud 2000). An overview of shot-noise modeling in insurance applications is also given in Kühn 2014.

Note that (4.7) quite closely resembles (3.5), in fact Dassios and Jang 2003 even use $\nu(t) \equiv 0$ and $h(t - s_i) = e^{-\delta(t-s_i)}$, giving the intensity $\lambda(t)$ the form we have seen in Figure 2. However, there is a fundamental difference: For the Cox process, there is an independent Poisson process $\{s_i\}$ describing stochastic occurrences of external events. This process influences the intensity $\lambda(t)$ which describes the rate of event occurrences $\{t_i\}$ in the claims process $N(t)$. In the Hawkes process case, the claims process $N(t)$ itself takes the role of the formerly independent Poisson process, thus the arrivals of the claims process itself determine its underlying intensity.

Of course, there are numerous other ways of generalizing the classical model apart from changing the arrival process $N(t)$. Some results for the classical model were at first only available for light-tailed claim distributions G, thus research has been devoted to obtain the same or adjusted results for heavy-tailed claim size distributions. Some authors have generally challenged the assumption of i.i.d. claim sizes $\{Y_i\}$ (Albrecher and Kantor 2002) and the independence between claim arrival times $N(t)$ and their sizes $\{Y_i\}$ (Albrecher and Boxma 2004, Meng, Zhang, and Guo 2008, Boudreault et al. 2006). Other work has been devoted to incorporating interest rate or inflation into the model, for example by considering discounted claims (Léveillé and Hamel 2018, Jang 2004) or by including the possibility of borrowing and investment into the model (Schmidli 1994).

All these generalisations contribute to the development of risk models that describe real-world phenomena as closely as possible and are by themselves interesting topics to study. However, the focus of this thesis will be to combine the previous two chapters, namely Hawkes processes and risk theory in order to describe certain classes of insurance claims which can be suspected to display *self-exciting* and clustering behaviour.

Chapter 5

Risk model with Hawkes processes

In this chapter, we will combine the two previous topics by considering a risk model where claims arrive according to a Hawkes process. This is interesting in a scenario where claim arrivals cannot be assumed to arrive independently from one another and where there is no indication that the time-dependence of claim arrival rates is seasonal or caused by exogeneous events like catastrophes. We will see an example of such a type of insurance in the numerical calculations in Chapter 6.
We will first give an overview over recent work and existing results on risk models with Hawkes processes before introducing the *risk model with general compound Hawkes processes* which will be our focus. This model was introduced by Swishchuk 2017b and is very convenient as it allows the derivation of a law of large numbers and functional central limit theorem. The latter allows to construct a diffusion approximation which gives access to many analytical results, such as closed formulas for finite and infinite horizon ruin probabilities.

5.1 Previous work

The first work to consider a risk model with Hawkes claims arrivals was Stabile and Torrisi 2010 who introduce the following:

Definition 5.1 (Hawkes risk model)**.** Let the surplus of an insurance portfolio be given by

$$R(t) = u + ct - \sum_{i=1}^{N(t)} Y_i \tag{5.1}$$

where u denotes the initial capital, c denotes the (continuous) premium rate and the number of claims in the interval $(0, t]$ is denoted $N(t)$ and follows a stationary Hawkes process

with intensity (3.5). The claims $\{Y_i\}$ are i.i.d. positive random variables independent of $N(t)$ with $m_1 := \mathbb{E}[Y_1] < \infty$ and $m_2 := \mathbb{E}[Y_1^2] < \infty$.

Stabile and Torrisi 2010 use the clustering representation of the Hawkes process to interpret the risk model as follows: there are *standard claims* which occur according to a homogeneous Poisson process with rate λ and *trigger* secondary claims according to the branching structure of the Hawkes process (see Proposition 3.7). The decreasing nature of the excitation function $\mu(\cdot)$ of the Hawkes process (see Definition 3.5) reflects that recent events affect the occurrence of triggered claims more than events in the distant past. Stabile and Torrisi 2010 even compare their model to the shot-noise Cox model in Albrecher and Asmussen 2006, pointing out the difference between the "external events" structure there and the internally self-exciting branching structure of the Hawkes model. Using the theory of large deviations and assuming light-tailed claims, Stabile and Torrisi 2010 study the asymptotic behavior of infinite and finite horizon ruin probabilities and manage to derive asymptotically efficient simulation laws. Their work represents an important first step for applying Hawkes processes in risk theory and was extended by Zhu 2013 who considered the same problems assuming (subexponential) heavy tailed claims. However, their results are of a very theoretical nature and only allow estimations of ruin probabilities by essentially Monte Carlo simulation and only in the limit $u \to \infty$.

Another important stream of contributions includes a risk model based on a *dynamic contagion process* introduced by Dassios and Zhao 2011. This process generalises the Hawkes process and the Cox process with shot noise intensity by including both self-excited and externally excited jumps. Dassios and Zhao 2012 use the dynamic contagion process as the claim arrival process in their risk model to derive generalisations of the Cramer-Lundberg approximation, the Lundberg inequality and bounds for the probability of ruin with special attention on the case of exponential jumps. They provide some numerical examples of ruin probabilities via Monte-Carlo simulation.

Jang and Dassios 2013 study a bivariate extension of the shot noise self-exciting process introduced in Dassios and Zhao 2012 with externally excited joint jumps (shot noise Cox process) and two separate self-excited jumps (branching structure of a Hawkes process with exponential rate). Furthermore, a constant rate of exponential decay that could be interpreted as the time value of money is included in the process. They derive its theoretical distributional properties and provide some simple numerical examples to show that this point process could be used for the modelling of discounted aggregate losses from catastrophic events. They argue that a process with self-exciting features is suitable to capture the clustering arrival of losses caused by the increasing frequency and intensity of natural and man-made disasters observed in practice (caused by such as global warming,

climate change and terrorism). The bivariate process reflects that a single (catastrophic) event might induce losses for two different lines of the insurance business. For example, a flood or an earthquake with tsunami could result in simultaneous joint losses of automobiles and properties. This model is theoretically very interesting, but probably impossible to corroborate with empirical data as it would be very hard to find an adequate data set and conduct the model fitting.

A very recent and interesting result was derived in Cheng and Seol 2018, namely a diffusion approximation and asymptotics for the ruin probabilities of a risk model with stationary Hawkes claims arrivals. They use the risk model from Definition 5.1 to define the following series of risk models, where we slightly change the presentation in order to emphasize the analogy to the classical approximation in Proposition 4.7. Let

$$R^{(n)}(t) = u + c^{(n)}t - \sum_{i=1}^{N^{(n)}(t)} \frac{Y_i}{\sqrt{n}}$$

be a sequence of risk processes from (5.1), where $\{Y_i\}$ are i.i.d. with $m_1 := E[Y_1] < \infty$ and $m_2 := E[Y_1^2] < \infty$. $N^{(n)}(\cdot)$ is a Hawkes process with intensity given in (3.5), background rate $\lambda^{(n)} = \lambda n$ and non-increasing excitation function $\mu(\cdot)$. The premium rate is given by

$$c^{(n)} = \sqrt{n}\frac{\lambda m_1}{1-\hat{\mu}} + k = c + (\sqrt{n}-1)\frac{\lambda m_1}{1-\hat{\mu}}$$

for a constant $k > 0$ and $\hat{\mu}$ given in (3.6). Note that if the premium rate of the original model is denoted as c, k must be given by $k = c - \frac{\lambda m_1}{1-\hat{\mu}}$. The exact value depends on the premium principle which is chosen for the model, for the *expected value principle* with safety loading θ it follows $k = \frac{\lambda\theta m_1}{1-\hat{\mu}}$.
Analogously to the classical case, we illustrate the procedure in Figure 8.

Cheng and Seol 2018 derive the following results:

Theorem 5.2 (Cheng and Seol 2018)**.** Let $X^{(n)}(t) = \sum_{i=1}^{N^{(n)}(t)} Y_i$ be the aggregate claims process and assume that $\mu(\cdot)$ is a decreasing function with $\int_0^\infty t\mu(t)dt < \infty$. Then, as $n \to \infty$,

$$\frac{X^{(n)}(t) - \frac{\lambda m_1 t}{1-\hat{\mu}} n}{\sqrt{n}} \xrightarrow{D} G \tag{5.2}$$

in distribution where G is a mean-zero almost surely continuous Gaussian process with covariance function ($t \geq s$),

$$\mathbb{C}ov(G(t), G(s)) = m_1^2 \int_s^t \int_0^s \phi(u-v)dvdu + \frac{m_2 s}{1-\hat{\mu}} + 2m_1^2 \int_0^s \int_0^u \phi(u-v)dvdu \tag{5.3}$$

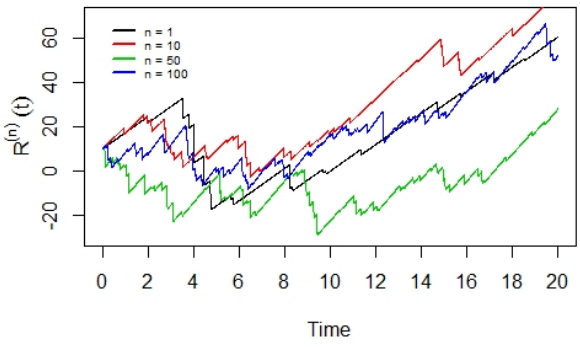


Figure 8: Illustration of the idea of approximating the Hawkes risk model by a diffusion process, where the original risk process $R^{(1)}(t)$ has $u = 10$, $\theta = 0.1$ and claim arrivals following a Hawkes process with parameters $(\lambda, \alpha, \beta) = (1, 2, 5)$ and claim sizes $\sim Exp(0.25)$. For different values of n, the plots show realisations of $R^{(n)}(t)$ over the period $[0, 20]$. We can clearly see that with increasing n, the risk process starts resembling a continuous diffusion process.

where $\phi : [0, \infty) \to [0, \infty)$ satisfies the integral equation

$$\phi(t) = \frac{\mu(t)}{1 - \hat{\mu}} + \int_0^\infty \mu(t + v)\phi(v)dv + \int_0^t \mu(t - v)\phi(v)dv. \tag{5.4}$$

As a result,

$$\begin{aligned} R^{(n)}(t) &= u + \left(\sqrt{n}\frac{\lambda m_1}{1 - \hat{\mu}} + k\right)t - \frac{X^{(n)}(t)}{\sqrt{n}} \\ &= u + kt - \frac{X^{(n)}(t) - \frac{\lambda n m_1 t}{1-\hat{\mu}}}{\sqrt{n}} \xrightarrow{D} u + kt - G(t). \end{aligned} \tag{5.5}$$

Proof. As will be detailed below, we use the clustering representation of the Hawkes process to write $X^{(n)}(t) = \sum_{i=1}^{n} X^{(1)}(t) =: \sum_{i=1}^{n} X(t)$, i.e. to decompose the aggregate claims sum from a process with background rate λn into the sums resulting from n independent processes with background rate λ.

Let $\tilde{N}_i(t) := N_i(t) - \mathbb{E}[N_i(t)] = N_i(t) - \frac{\lambda t}{1-\hat{\mu}}$, then $\tilde{N}_i$ are i.i.d. random elements of $D([0, \infty), \mathbb{R})$ with $\mathbb{E}[\tilde{N}_i(t)] = 0$ and $\mathbb{E}[\tilde{N}_i(t)^2] < \infty$ for any t. Analogously, define $\tilde{X}_i(t) = X_i(t) - \frac{\lambda t m_1}{1-\hat{\mu}}$. Then, assuming some regularity conditions on the second moments of $\tilde{X}_i(\cdot)$ (see (2.1) and (2.2) in Theorem 2.7) it follows by Hahn's Theorem that

$$\frac{\sum_{i=1}^{n}(X_i(t) - \frac{\lambda t m_1}{1-\hat{\mu}})}{\sqrt{n}} \xrightarrow{D} G$$

in $D([0,\infty),\mathbb{R})$ where Z is an a.s. continuous, mean-zero Gaussian process with the covariance function of $\tilde{X}_1$. One can indeed show that (2.1) and (2.2) are fulfilled (by mainly using the tower property - the rather lengthy calculations will not be detailed here and can be seen in Cheng and Seol 2018), thus it remains to compute the covariance function. It holds by Wald's equation and the law of total variance

$$\begin{aligned}
\mathbb{E}[X^{(n)}(t)] &= m_1\mathbb{E}[N^{(n)}(t)] \\
\mathbb{E}[(X^{(n)}(t))^2] &= \mathbb{V}ar((X^{(n)}(t))) + \mathbb{E}[X^{(n)}(t)]^2 \\
&= \mathbb{V}ar[Y_1]\mathbb{E}[N^{(n)}(t)] + \mathbb{E}[Y_1]^2\mathbb{V}ar[N^{(n)}(t)] + m_1^2\mathbb{E}[N^{(n)}(t)]^2 \\
&= (m_2 - m_1^2)\mathbb{E}[N^{(n)}(t)] + m_1^2(\mathbb{E}[(N^{(n)}(t))^2] - \mathbb{E}[N^{(n)}(t)]^2) + m_1^2\mathbb{E}[N^{(n)}(t)]^2 \\
&= (m_2 - m_1^2)\mathbb{E}[N^{(n)}(t)] + m_1^2\mathbb{E}[(N^{(n)}(t))^2].
\end{aligned}$$

Furthermore, Cheng and Seol 2018 infer that

$$\mathbb{E}[(X^{(n)}(t) - X^{(n)}(s))X^{(n)}(s)] = m_1^2\mathbb{E}[(N^{(n)}(t) - N^{(n)}(s))N^{(n)}(s)]$$

to conclude

$$\begin{aligned}
\mathbb{C}ov(X^{(n)}(t), X^{(n)}(s)) &= \mathbb{E}[X^{(n)}(t)X^{(n)}(s)] - \mathbb{E}[X^{(n)}(t)]\mathbb{E}[X^{(n)}(s)] \\
&= \mathbb{E}[(X^{(n)}(t) - X^{(n)}(s))X^{(n)}(s)] + \mathbb{E}[(X^{(n)}(s))^2] - \mathbb{E}[X^{(n)}(t)]\mathbb{E}[X^{(n)}(s)] \\
&= m_1^2\mathbb{E}[(N^{(n)}(t) - N^{(n)}(s))N^{(n)}(s)] + \mathbb{E}[(X^{(n)}(s))^2] - m_1^2\mathbb{E}[N^{(n)}(t)]\mathbb{E}[N^{(n)}(s)] \\
&= m_1^2\mathbb{E}[N^{(n)}(t)N^{(n)}(s)] - m_1^2\mathbb{E}[(N^{(n)}(s))^2] + \mathbb{E}[N^{(n)}(s)](m_2 - m_1^2) \\
&+ m_1^2\mathbb{E}[(N^{(n)}(s))^2] - m_1^2\mathbb{E}[N^{(n)}(t)]\mathbb{E}[N^{(n)}(s)] \\
&= m_1^2\mathbb{C}ov(N^{(n)}(t), N^{(n)}(s)) + \mathbb{E}[N^{(n)}(s)](m_2 - m_1^2).
\end{aligned}$$

Inserting the covariance density of the Hawkes process given in (5.4) (whose derivation goes beyond the scope of this proof and can be found e.g. in Laub, Taimre, and Pollett 2015 (Theorem 2) for an exponential Hawkes process or Appendix E for a special case) yields (5.3). □

Proposition 5.3 (Cheng and Seol 2018)**.** It holds in distribution that

$$G(t) = m_1 \int_0^t H(s)ds + \frac{m_2}{1-\hat{\mu}}W(t)$$

where $H(s)$ is a centered stationary Gaussian process with

$$\mathbb{C}ov(H(t), H(s)) = \phi(t-s), \qquad t \geq s,$$

and $W(t)$ is a standard Brownian motion independent of $H(t)$.

Remark. We can clearly see the difference between the classical Poisson case and the more involved Hawkes case. In the classical case, increasing the rate λ of jumps leads to a simple Brownian motion as a diffusion limit, whereas in the Hawkes case the increase of the background rate λ does not capture the self-excitation function $\mu(\cdot)$ and therefore the limit process is more complicated. In order to still understand the result from a classical central limit theorem point of view, we can use the immigration-birth-representation of the Hawkes process: Recall that the immigrant events arrive as a homogeneous Poisson process according to the background rate and produce children independently according to the excitation function. Thus we can decompose a stationary Hawkes process with background rate λn and excitation function $\mu(\cdot)$ into n independent processes, each with background rate λ. $N(t)$ is the total number of events from each such process, and $N^{(n)}(t)$ the aggregated number from n independent processes. Therefore, we would expect $N^{(n)}(t)$ to follow some sort of Gaussian law as $n \to \infty$.

In the general case, (5.4) and thus (5.3) cannot be solved explicitly. However, for the exponential Hawkes process, the expressions simplify and allow some more explicit insights. Inserting $\mu(t) = \alpha e^{-\beta t}$ (where $\alpha < \beta$) into (5.4) leads to

$$\phi(t) = \frac{\alpha e^{-\beta t}}{1 - \alpha/\beta} + \alpha e^{-\beta t} \int_0^\infty e^{-\beta v} \phi(v) dv + \alpha e^{-\beta t} \int_0^t e^{\beta v} \phi(v) dv$$

which in fact is the equation describing the *covariance density* of an exponential Hawkes process with background rate $\lambda = 1$. Appendix E highlights how to solve this equation to obtain

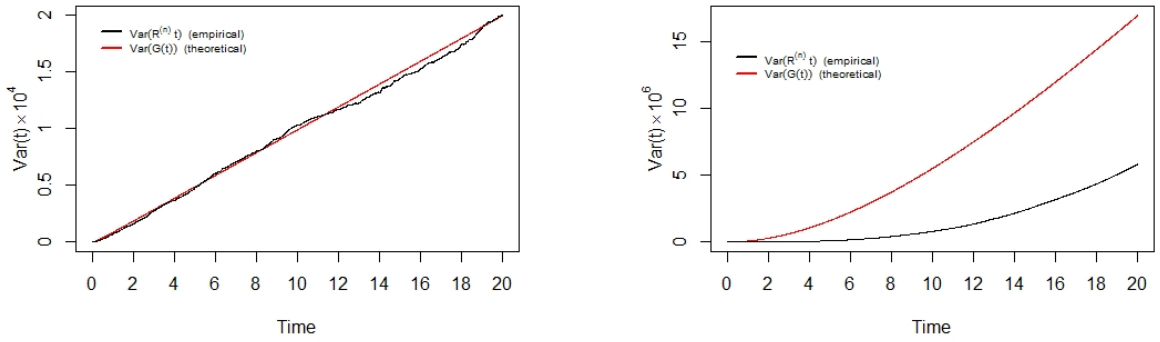
$$\phi(t) = \frac{\alpha\beta(2\beta - \alpha)}{2(\beta - \alpha)^2} e^{-(\beta - \alpha)t}, \text{ for all } t \geq 0.$$

Applying (5.3) for $s = t$ yields the variance of the Gaussian limit as

$$\begin{aligned}
\mathbb{V}ar(G(t)) = \mathbb{C}ov(G(t), G(t)) &= \frac{m_2 t}{1 - \hat{\mu}} + 2m_1^2 \int_0^t \int_0^{t_2} \phi(t_2 - t_1) dt_1 dt_2 \\
&= \frac{m_2 t}{1 - \alpha/\beta} + \frac{2m_1^2 \alpha\beta(2\beta - \alpha)}{2(\beta - \alpha)^2} \int_0^t \int_0^{t_2} e^{-(\beta - \alpha)(t_2 - t_1)} dt_1 dt_2 \\
&= \frac{m_2 t}{1 - \alpha/\beta} + \frac{m_1^2 \alpha\beta(2\beta - \alpha)}{(\beta - \alpha)^2} \int_0^t e^{-(\beta - \alpha)t_2} \cdot \left[\frac{e^{(\beta - \alpha)t_2} - 1}{\beta - \alpha} \right] dt_2 \\
&= \frac{m_2 t}{1 - \alpha/\beta} + \frac{m_1^2 \alpha\beta(2\beta - \alpha)}{(\beta - \alpha)^3} \int_0^t \left(1 - e^{-(\beta - \alpha)t_2} \right) dt_2 \\
&= \frac{m_2 t}{1 - \alpha/\beta} + \frac{m_1^2 \alpha\beta(2\beta - \alpha)}{(\beta - \alpha)^3} \left[t + \frac{e^{-(\beta - \alpha)t} - 1}{\beta - \alpha} \right]
\end{aligned}$$

$$= \left(\frac{\beta}{\beta - \alpha} m_2 + \frac{\alpha\beta(2\beta - \alpha)}{(\beta - \alpha)^3} m_1^2 \right) t - m_1^2 \frac{\alpha\beta(2\beta - \alpha)}{(\beta - \alpha)^4} \left(1 - e^{-(\beta - \alpha)t}\right). \quad (5.6)$$

Figure 9 compares the theoretical variance of the limit process $G(t)$ to the empirical variance of a risk process $R^{(n)}(t)$ for $n = 50$ over 1000 simulation runs.



(a) $\beta = 5$ (b) $\beta = 2.1$

Figure 9: In order to illustrate the explicit result for the exponential Hawkes process case, we simulate 1000 runs of a risk process $R^{(50)}(t)$ on a grid of step size $\delta = 0.01$ over the period $[0, 20]$ and calculate the empirical variance at each time point. In (a), the underlying parameters of $R^{(1)}(t)$ are again $u = 10$, $\theta = 0.1$, $(\lambda, \alpha, \beta) = (1, 2, 5)$ (for the Hawkes claim arrival process) and $\gamma = 0.25$ (for the exponentially distributed claim sizes). We compare the results with the theoretical variance in (5.6) as a function of t and find that the values match very closely. It has to be remarked that in this case, as $\alpha << \beta$, the variance function is almost linear in time. In (b), where we leave all parameters unchanged but set $\beta = 2.1$, we clearly observe the non-linearity in time and the match between the theoretical and empirical values deteriorates.

We can see that the variance function of the diffusion limit is *non-linear* in t, which is different from the classical case of the Poisson process that yields as standard Brownian motion. Unfortunately, this complicates the application of the diffusion approximation for the calculation of ruin probabilities. Cheng and Seol 2018 derive some asymptotic results for letting $u \to \infty$ in both the finite- and infinite-horizon case (which are outside the scope of this thesis and will not be presented), and use numerical simulations for the limit process $G(t)$ in the special case of an exponential Hawkes process and exponentially or Gamma-distributed claim sizes to approximate ruin probabilities.
In summary, while their results are theoretically very interesting and very recent, so far their empirical applicability is quite restricted due to the complicated nature of the limit process. Therefore, we now turn to a different Hawkes model which allows for a more convenient result and application.

5.2 Risk model with general compound Hawkes processes

5.2.1 Definition and properties

Swishchuk 2017b introduces the following risk model:

Definition 5.4 (Risk model with general compound Hawkes process (Swishchuk 2017b))**.** Let $N(t)$ be a one-dimensional Hawkes process as in Definition 3.5. Let $X = (X_i)$ be an ergodic continuous-time Markov chain on a finite (or countably infinite) state space $S = \{1, 2, \cdots, n\}$ (or $S = \{1, 2, \cdots, n, \cdots\}$) independent of $N(t)$, and let $a(x)$ be any continuous and bounded function on S. Then a general compound Hawkes process is defined as

$$H(t) = H(0) + \sum_{i=1}^{N(t)} a(X_i). \tag{5.7}$$

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

$$R(t) = u + ct - \sum_{i=1}^{N(t)} a(X_i) \tag{5.8}$$

where u is the initial capital, c is the premium rate and $N(t), (X_i)$ and $a(x)$ are as above. We will abbreviate this *risk model based on general compound Hawkes processes* as *RMGCHP*.

A related model has been successfully used for modelling Limit Order Book dynamics in Swishchuk 2017a where the $a(X_i)$ are interpreted as a conditionally dependent sequence of price changes which are not fixed to one-tick movements. In our insurance context, we will adapt the interpretation of the Markov chain as described in Chapter 6 in order to appropriately reflect incoming insurance claim sizes.

Remark. Note that as a sequence of i.i.d. random variables can be seen as a special case of a Markov chain (by choosing the transition probabilities such that the conditional dependence vanishes) and a Poisson process can be seen as a special case of a Hawkes process (with excitation function $\mu(\cdot) \equiv 0$), *RMGCHP* generalizes both the Hawkes model in Definition 5.1 as well as the classical risk model in Definition 4.1. The only limitation is that the claim sizes in *RMGCHP* can take at most countably many values whereas the other models usually assume claim sizes to follow a continuous distribution function G. However, as we will see in detail in Chapter 6, this is not a very relevant restriction for empirical purposes, as we can approximate any empirical distribution of claim sizes arbitrarily well with our model by increasing the number of states $|S|$.

Analogously to the classical model, Swishchuk 2017b proves a law of large numbers for *RMGCHP* and uses it to derive a net profit condition and a premium principle based on the expected value principle.

Theorem 5.5 (Law of large numbers for RMGCHP (Swishchuk 2017b))**.** Let $R(t)$ be the risk model from Definition 5.4, and let $X = (X_i)$ be an ergodic Markov chain with state space S and stationary probabilities π_n^*. We suppose that $0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1$. Then

$$\lim_{t\to\infty} \frac{R(t)}{t} = c - a^* \frac{\lambda}{1-\hat{\mu}}, \tag{5.9}$$

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

Proof. The proof works analogously to the classical case in Lemma 4.4. Consider

$$\begin{aligned}\lim_{t\to\infty} \frac{R(t)}{t} &= c - \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{N(t)} a(X_i) \\ &= c - \lim_{t\to\infty} \frac{N(t)}{t} \cdot \frac{1}{N(t)} \sum_{i=1}^{N(t)} a(X_i)\end{aligned}$$

Applying the LLN for Hawkes processes (Theorem 3.8) to the first factor gives

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{\lambda}{1-\hat{\mu}},$$

while the ergodic theorem for Markov chains yields

$$\lim_{t\to\infty} \frac{1}{N(t)} \sum_{i=1}^{N(t)} a(X_i) = \sum_{i\in S} a(i)\pi_i^* =: a^*$$

for the second factor, where π^* is the unique stationary distribution. This proves the statement of the theorem. □

Remark. Note that if $a(X_i) = X_i$ are i.i.d. random variables, then simply $a^* = \mathbb{E}[X_1] =: m_1$.

The previous theorem implies the following net profit condition and premium principle.

Corollary 5.6 (Net profit condition and premium principle for RMGCHP (Swishchuk 2017b))**.** The net profit condition for *RMGCHP* is given as

$$c > a^* \frac{\lambda}{1-\hat{\mu}}, \tag{5.10}$$

and the premium principle for *RMGCHP*, based on the expected value principle, is

$$c = (1+\theta)a^* \frac{\lambda}{1-\hat{\mu}} \tag{5.11}$$

where θ denotes the safety loading.

5.2.2 Diffusion approximations

In the following, we will review a functional central limit theorem by Swishchuk 2017b which allows to approximate the risk model by a jump-diffusion process. We further use this jump diffusion approximation to construct an approximation by a pure diffusion process. Conveniently, the diffusion limit is found to be a Brownian motion with drift, which admits analytical calculation of estimates for finite-time and infinite-horizon ruin probabilities and the application of other well-known results.

Approximation by jump diffusion process

Theorem 5.7 (FCLT, Approximation by jump diffusion process (Swishchuk 2017b))**.** Let $R(t)$ be the risk model defined from Definition 5.4, and (X_i) be an ergodic Markov Chain with stationary distribution π^*. Suppose that $0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1$ and $\int_0^\infty s\mu(s)ds < \infty$. Then:

$$\lim_{t\to\infty} \frac{R(t) - (u + ct - a^* N(t))}{\sqrt{t}} \overset{D}{=} \hat{\sigma}\Phi(0,1) \tag{5.12}$$

or in Skorokhod topology (see Skhorokhod 2014)

$$\lim_{n\to\infty} \frac{R(nt) - (u + cnt - a^* N(nt))}{\sqrt{n}} = \hat{\sigma} W(t) \tag{5.13}$$

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

$$\begin{aligned}
\hat{\sigma} &:= \sigma^* \sqrt{\lambda/(1-\hat{\mu})}, \quad (\sigma^*)^2 := \sum_{i\in S} \pi_i^* \nu(i), \\
a^* &:= \sum_{i\in S} \pi_i^* a(i), \quad b(i) := a^* - a(i), \\
\nu_i &:= b(i)^2 + \sum_{j\in S} (g(j) - g(i))^2 P(i,j) - 2b(i) \sum_{j\in S} (g(j) - g(i)) P(i,j), \\
g &:= (P + \Pi^* - I)^{-1} (b(1), ..., b(n))',
\end{aligned} \tag{5.14}$$

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

Proof. The proof relies heavily on the martingale method developed in Vadori and Swishchuk 2015 and we will only sketch the main idea here. We start with the second statement (5.13). By Definition 5.4

$$\frac{R(nt) - (u + cnt - a^* N(nt))}{\sqrt{n}} = \frac{u + cnt - \sum_{i=1}^{N(nt)} a(X_i) - (u + cnt - a^* N(nt))}{\sqrt{n}}$$

$$= \frac{\sum_{i=1}^{N(nt)} (a^* - a(X_i))}{\sqrt{n}}$$

Thus, we need to find

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{N(nt)} (a^* - a(X_i))}{\sqrt{n}}. \tag{5.15}$$

For $n \in \mathbb{N}_0$, consider

$$R_n^* = \sum_{i=1}^{n} (a(X_i) - a^*)$$

and let $U_n^*(t)$ be defined as in Vadori and Swishchuk 2015, (3.52), to be

$$U_n^*(t) = \frac{(1 - (nt - \lfloor nt \rfloor))R^*_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)R^*_{\lfloor nt \rfloor + 1}}{\sqrt{n}},$$

where $\lfloor \cdot \rfloor$ is the *floor* function $\lfloor x \rfloor = \max\{m \in \mathbb{Z} : m \leq x\}$. Note that $U_n^*(t)$ is the continuous function which we get by linearly extending $\frac{R^*_{nt}}{\sqrt{n}}$ to values $nt \notin \mathbb{N}_0$. Theorem 3.17 in Vadori and Swishchuk 2015 then yields that

$$\lim_{n\to\infty} U_n^*(t) \overset{D}{=} \sigma^* W(t) \tag{5.16}$$

where σ^* is given in (5.14) and $W(t)$ is a standard Brownian motion. The derivation of this result is quite involved and can be reviewed in Vadori and Swishchuk 2015.
For a fixed value of t, we now change the time scale to $\frac{N(nt)}{n}$ and therefore consider $U_n^*(\frac{N(nt)}{n})$ which is the continuous version of

$$\frac{R^*_{n\cdot\frac{N(nt)}{n}}}{\sqrt{n}} = \frac{R^*_{N(nt)}}{\sqrt{n}} = \frac{\sum_{i=1}^{N(nt)} (a(X_i) - a^*)}{\sqrt{n}}.$$

This expression coincides with the negative of the expression whose limit we seek in (5.15). From Theorem 3.8 we know that

$$\lim_{n\to\infty} \frac{N(nt)}{n} = \frac{\lambda t}{1 - \hat{\mu}},$$

therefore it holds

$$\lim_{n\to\infty} U_n^*\left(\frac{N(nt)}{n}\right) \overset{D}{=} \sigma^* W\left(\frac{\lambda t}{1 - \hat{\mu}}\right).$$

As for the standard Brownian motion it holds that $W(at) \overset{D}{=} \sqrt{a}W(t)$ for all $a \in [0, \infty)$, it follows

$$\lim_{n\to\infty} U_n^*\left(\frac{N(nt)}{n}\right) \overset{D}{=} \sigma^* \sqrt{\frac{\lambda}{1 - \hat{\mu}}} W(t).$$

As furthermore $W(t) \stackrel{D}{=} -W(t)$, it follows that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{N(nt)} (a^* - a(X_i))}{\sqrt{n}} = \sigma^* \sqrt{\frac{\lambda}{1-\hat{\mu}}} W(t).$$

This proves the statement of (5.13). To get (5.12), set $n = t$ and $t = 1$ in (5.13). □

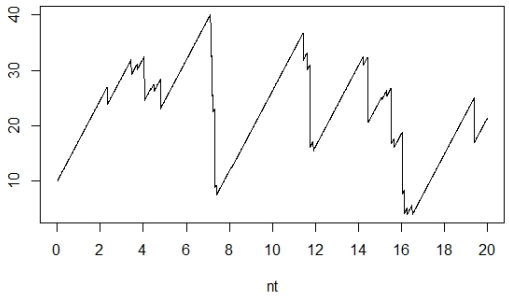
Remark (Swishchuk 2017b, Remark 7)**.** If $a(X_i) = X_i$ are i.i.d. random variables, then

$$(\sigma^*)^2 = \mathbb{V}ar[X_1]$$

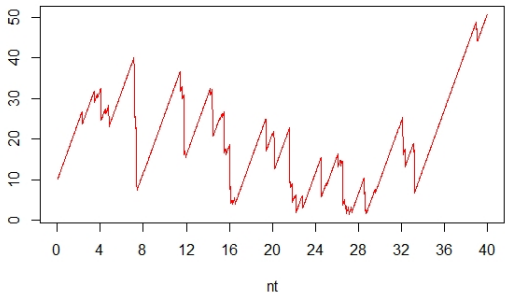
and thus

$$\sigma^* = \sqrt{\mathbb{V}ar[X_1]} = \sqrt{\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2} =: \sqrt{m_2 - m_1^2}.$$

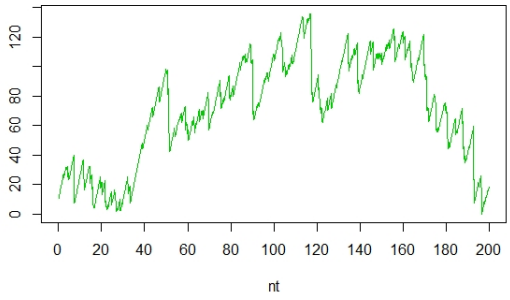
The approximation can be interpreted as follows: Letting $n \to \infty$ in (5.13) corresponds to changing the time scale from the original time steps to a larger scale and therefore observing the process at fewer time points. For example, instead of observing the process each hour or each day, we only observe it monthly. Therefore, we would assume that in between observations, more claims arrive and therefore the distribution of their sizes starts resembling the stationary distribution. This means that the actual aggregated claims sum $\sum_{i=1}^{N(nt)} a(X_i)$ can be more closely approximated by $a^* N(nt)$ together with a random term following a standard Brownian motion. We illustrate this procedure in Figure 10. Swishchuk et al. 2017 uses an analogous model to describe the link between price volatility and order flow in limit order books. When studying limit order book data on the original timescale of milliseconds, very frequent tick-sized jumps occur (order arrivals and cancellations through high-frequency trading) that can be described e.g. by a Hawkes process. By changing the timescale e.g. to tens of seconds or minutes, one can study the order flow and thus the price process which is then assumed continuous and described by a diffusion process.



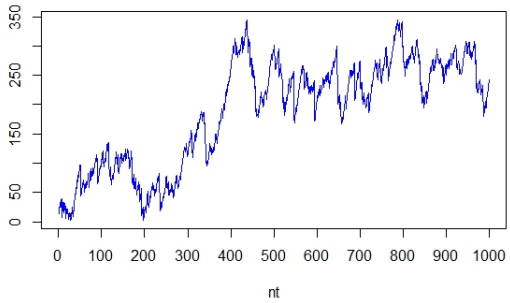
(a) $n = 1$



(b) $n = 2$



(c) $n = 10$



(d) $n = 50$

Figure 10: Illustration of the idea of changing the time scale for *RMGCHP*: We start with a risk process over the period $[0, 1000]$ (say days) with $u = 10$, $\theta = 0.1$ and claim arrivals following a Hawkes process with parameters $(\lambda, \alpha, \beta) = (1, 2, 5)$ and i.i.d. claim sizes with $m_1 = 4$ and $\mathbb{V}ar[X_1] = 16$. Now assume we want to have an "observation window" of length 20. In the original timescale ($n = 1$), this of course corresponds to the window of the first 20 days $[0, 20]$. If we now change the timescale to e.g. $nt = 10$, we observe the value of the process only every 10 days, i.e. at the original timepoints $10, 20, \cdots$. This corresponds to a new observation window of $[0, 200]$ days. We plot the *same* underlying risk process for four different values of n above. We can see that with increasing n, the individual jumps become less significant and the observed risk process starts resembling a diffusion. We should mention that the y-scales differ between plots, as naturally the process can take a higher range of values if we observe it over a longer time.

Theorem 5.7 implies that the risk model in (5.8) for large t can be approximated by the jump-diffusion process

$$R(t) \approx R_j(t) = u + ct - a^* N(t) + \hat{\sigma} W(t) \tag{5.17}$$

where a^* and $\hat{\sigma}$ are defined above, $N(t)$ is a Hawkes process and $W(t)$ is a standard Wiener process. The subscript j denotes the *jump* diffusion approximation. Assuming an exponential Hawkes process, the variance of the approximative process $R_j(t)$ is given by

$$\mathbb{V}ar[R_j(t)] = (a^*)^2 \mathbb{V}ar[N(t)] + \hat{\sigma}^2 t$$

$$= (a^*)^2\left(\frac{\lambda}{1-\alpha/\beta}t\left(\frac{1}{1-\alpha/\beta}\right)^2 + (*)\right) + (\sigma^*)^2\frac{\lambda}{1-\alpha/\beta}t$$
$$= t\left((\sigma^*)^2\frac{\lambda}{1-\alpha/\beta} + (a^*)^2\frac{\lambda}{(1-\alpha/\beta)^3}\right) - (**) \tag{5.18}$$

where

$$(**) = -(a^*)^2\frac{\lambda}{1-\alpha/\beta}\left(1-\left(\frac{1}{1-\alpha/\beta}\right)^2\right)\left(\frac{1-e^{-t(\beta-\alpha)}}{\beta-\alpha}\right) > 0, \tag{5.19}$$

where we have inserted the variance of the number of jumps $N(t)$ of a stationary Hawkes process from (3.27). Note again that we have one part of the variance which is linear in t, and another part which is non-linear. For the classical Poisson case, we would set $\alpha = 0$ implying $(**) = 0$, leaving only the linear part. We will examine this further in the next section. Note that in general it is very difficult to estimate analytically how fast (or well) the diffusion process approximates the original process. For details about the rate of approximation for random evolutions, see Swishchuk 2000. Therefore, we will check the accuracy of this (and the next) approximation using an empirical method in Section 6.3.

Approximation by pure diffusion process and ruin probabilities

While the last result is already quite convenient, many applications are not available or more restricted when dealing with a diffusion process that includes jumps. Therefore, we would now like to construct an approximation of $R(t)$ from (5.8) by a continuous (pure) diffusion process.

Theorem 5.8 (FCLT 2, Approximation by pure diffusion process)**.** Let $R(t)$ be the risk model from Definition 5.4, and let (X_i) be an ergodic Markov Chain with stationary probabilities π^*. We suppose that $0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1$ and $\int_0^\infty s\mu(s)ds < \infty$. Then

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

or in Skorokhod topology (see Skhorokhod 2014)

$$\lim_{n\to\infty}\frac{R(nt) - (u + cnt - a^*\frac{\lambda}{1-\hat{\mu}}nt)}{\sqrt{n}} = \bar{\sigma}W(t) \tag{5.20}$$

where $\Phi(\cdot,\cdot)$ is the standard Normal c.d.f., $W(t)$ is a standard Wiener process and $\bar{\sigma} = \sqrt{\hat{\sigma}^2 + (a^*\sqrt{\frac{\lambda}{(1-\hat{\mu})^3}})^2}$ where a^* and $\hat{\sigma}$ are defined in Theorem 5.7.

Proof. We only prove (5.20). We know from Theorem 5.7 that

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

where $\hat{\sigma} = \sigma^* \sqrt{\lambda/(1-\hat{\mu})}$ and σ^* is defined in Theorem 5.7. Let us replace in the above equation $N(nt)$ by its expected value $\frac{\lambda}{1-\hat{\mu}}nt$ and thus look at

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

Adding and subtracting the term $a^* N(nt)$ yields

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

We know by Theorem 5.7 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 CLT for Hawkes processes (Theorem 3.9) 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 $W(t)$. Thus for the sum of the two limits we obtain

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

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 5.7. □

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

$$R(t) \approx R_p(t) = u + ct - a^* \frac{\lambda}{1-\hat{\mu}} t + \bar{\sigma} W(t) \tag{5.21}$$

where a^* and $\bar{\sigma}$ are defined above and $W(t)$ is a standard Wiener process. The subscript p denotes the *pure* diffusion approximation. The variance of the approximative process $R_p(t)$ is thus given by

$$\mathbb{V}ar[R_p(t)] = \bar{\sigma}^2 t = \left((\sigma^*)^2 \frac{\lambda}{1-\hat{\mu}} + (a^*)^2 \frac{\lambda}{(1-\hat{\mu})^3} \right) t,$$

which we actually observe to be the linear part in (5.18). Thus, the second approximation naturally gets rid of the non-linear contribution to the variance caused by the Hawkes jump process by approximating the risk process by a Brownian motion with drift. We would naturally assume this approximation to be less accurate (where the severity of the estimation error depends on the specific parameters of the Hawkes process, compare e.g. Figure 9), but it is very convenient nevertheless as it gives access to several analytical results, such as ruin probability estimates in closed form.

Theorem 5.9 (Ruin probabilities for RMGCHP)**.** The ruin probability in the interval $(0, T]$ for a risk model as in Theorem 5.8 is given by

$$\Psi(u,T) = \Phi\bigg(-\frac{u+(c-a^*\lambda/(1-\hat{\mu}))T}{\bar{\sigma}\sqrt{T}}\bigg) + e^{-\frac{2(c-a^*\lambda/(1-\hat{\mu}))}{\bar{\sigma}^2}u}\Phi\bigg(-\frac{u-(c-a^*\lambda/(1-\hat{\mu}))T}{\bar{\sigma}\sqrt{T}}\bigg), \tag{5.22}$$

and the ultimate ruin probability is given by

$$\Psi(u) = e^{-\frac{2(c-a^*\lambda/(1-\hat{\mu}))}{\bar{\sigma}^2}u}. \tag{5.23}$$

Proof. Note that

$$R_p(t) \overset{D}{=} u + W\bigg(\bigg(c - a^*\frac{\lambda}{1-\hat{\mu}}\bigg)t, \bar{\sigma}t\bigg).$$

Then the result follows directly from Theorem 2.28 by inserting $\mu = c - a^*\frac{\lambda}{1-\hat{\mu}}$, $\sigma = \bar{\sigma}$ and using that $1 - \Phi(x) = \Phi(-x)$. □

Remark. We find our results to be in accordance with the classical results in Proposition 4.7: If we insert above the values for a Poisson arrival process ($\hat{\mu} = 0$) and i.i.d. claim sizes ($a^* = m_1, \sigma^* = \sqrt{m_2 - m_1^2}$) we find

$$\bar{\sigma} = \sqrt{(\sigma^*)^2\lambda + m_1^2\lambda} = \sqrt{(m_2 - m_1^2)\lambda + m_1^2\lambda} = \sqrt{m_2\lambda}$$

and thus

$$R_p(t) \overset{D}{=} u + W((c - m_1\lambda)t, m_2\lambda t).$$

The next part of the thesis will be devoted to implementing the *RMGCHP* with empirical data in order to corroborate its applicability and check the accuracy of the theoretical results.

Chapter 6

Implementation with empirical data

Whereas the previous work on Hawkes risk models was focused on theoretical results and sometimes numerical illustrations with simulated data, we would like to extend the scope of *RMGCHP* to an application with real empirical data. This chapter is devoted to explaining the characteristics of the empirical data set and fitting a *RMGCHP* to empirical claim arrivals and sizes. Furthermore, we check the goodness of the diffusion approximations and ruin probability estimates introduced in Chapter 5.

6.1 Properties of the empirical data set

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 suspect 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. Therefore we suspect a Hawkes process might be suitable to model claim arrivals for this class of insurance claims. The empirical data set is divided according to the *reporting year* of a claim, i.e. the year the claim has been reported to the insurance company and thus has come to their knowledge. 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, Guo, and Ng 2005, Boumezoued and Devineau 2017. We observe that the majority of the claims reported within a year have occurred the same or the previous year, but there is a non-negligible number of claims

with longer delay in reporting. An example extract of the original data set is given in Figure 11 and its features are summarized in Table 6.1. The figure illustrates that for the same *claim occurrence* we often observe multiple *claim payment dates* corresponding to cash outflows for the insurance company. It is of great interest for the company to understand the characteristics of these cash outflows. Suppose a claim has been reported and the first payment made, but the case is not typically closed after a single payment. It is beneficial to already be able to estimate the amount (and ideally the temporal structure) of the total cash outflows which have to be expected from this claim in the future. This is essential for the purpose of *reserving* an adequate amount of capital to cover future liabilities arising from this known case. As we have seen in Section 3.2, the Hawkes process has an intuitive interpretation as an immigration-birth-process whose parameters provide information about the expected *number of children* of each exogenous immigrant event and thus the overall share of endogenously triggered events. We now see that this has an empirically very relevant interpretation, namely the share of outgoing payments which are actually *subsequent* payments in known open cases.

VNR	SCHD_LNR	SCHD_DAT	ZAHLU	ZHLG_DATU	LEISTUNGSART	STATUS	MELD_JAH
000001738	9865	28.03.2002	745.78	06.05.2010	Privat-Vertrags-RS	geschlossen	2010
000001738	9865	28.03.2002	502.78	06.07.2010	Privat-Vertrags-RS	geschlossen	2010
000001738	9865	28.03.2002	4.43	11.08.2010	Privat-Vertrags-RS	geschlossen	2010
000001738	9897	01.01.2004	514.08	13.01.2011	Privat-Versicherungs	geschlossen	2010
000001738	9897	01.01.2004	147.08	19.01.2011	Privat-Versicherungs	geschlossen	2010
000001738	9933	01.01.2004	254.42	15.03.2013	Privat-Schadenersatz	geschlossen	2010
000001738	9933	01.01.2004	1242.36	06.03.2015	Privat-Schadenersatz	geschlossen	2010
000001738	9933	01.01.2004	3722.75	31.07.2015	Privat-Schadenersatz	geschlossen	2010
000001738	9871	20.07.2007	1025.3	24.02.2010	Privat-Vertrags-RS	geschlossen	2010
000001738	9871	20.07.2007	1377.07	30.09.2010	Privat-Vertrags-RS	geschlossen	2010
000001738	9963	01.01.2008	17.85	11.11.2010	Privat-Vertrags-RS	geschlossen	2010
000001738	9885	01.01.2009	17.85	02.03.2010	Privat-Schadenersatz	geschlossen	2010
000001738	9939	01.01.2009	1286.46	25.10.2010	Privat-Arbeits-RS	geschlossen	2010

Figure 11: An extract from the data set of the reporting year 2010. We can see that for example the first three rows belong to the same policy and claim occurrence (on 28.03.2002). The claim was reported in 2010 and has led to three payments by the insurer on future dates (06.05.,06.07. and 11.08.2010) with different claim sizes (745.78, 502.78 and 4.43). The case has been closed since.

Out of the five extensive data sets, one for each of the *reporting years* 2010 until 2014 provided by the insurance company, the question arises how to extract a meaningful yet usable subset. As a starting point, we concentrate on one subclass of legal expenses insurance for our one-dimensional model, with the future idea of potentially fitting a multivariate marked Hawkes model to compare different kinds of incoming claims and the characteristics of their evolution. As there is no indication that one subclass is particularly more suited than others, we select a subclass with a relatively high number of incoming claims, indicating its relevance in the overall portfolio of the insurance company. The next challenge is to thin the claim portfolio, as due to its sheer overall size, the average number of claims per day would be very high. This would pose a problem as we see in Figure 11, only the day of each claim payment is recorded as any finer granurality is not of particular

Column	Meaning	Comment
VNR	Policy Number	One policy contract, for example one client.
SCHDLNR	Running number of claim per policy	A client can have several claims over the years on the same policy.
SCHDDAT	Date of claim occurrence	This is the original occurrence date which can be earlier than the reporting year and does not immediately induce a payment.
ZAHLUNG	Size of claim payment	
ZHLGDATUM	Date of the claim payment	One claim can result in several payments over time whose the temporal structure is main point of interest.
LEISTUNGSART	Subclass of insurance	There are many types of legal expenses insurance, in total 44 classes in the whole dataset.
STATUS	Status	This indicates whether the case has been closed or remains open.
MELDJAHR	Reporting year	Data sets are divided according to this, we have data for the years 2010 to 2014.

Table 6.1: Explanation of the columns of the original data set.

interest to the insurance company. As the one-dimensional Hawkes process is a simple point process, multiple arrivals with the same timestamp are theoretically not possible. Therefore, for the parameter estimation, we artificially distribute the indistiguishable arrivals uniformly over their arrival day in order to generate distinct timestamps. For a high average number of arrivals 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. Likewise, a very low number of overall claim arrivals (e.g. working on individual client portfolios) simply does not contain enough data points to conduct a stable model fitting. A first approach to thinning the portfolio would be 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 their 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 low 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 want to use a data set that makes use of the whole time horizon provided to us by the empirical data. Thus, we decided to divide claim occurences according to their *delay in reporting*, thus from each of the reporting years we consider claims that had occurred a fixed number of years ago. Characterising claims this way leads to a complete division of the overall portfolio as can be seen in Figure 12 (if all the cases (diagonals) are considered separately). This way the claim payments include claims that have 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 across several subclasses of insurance. This means that the data set under consideration in the following does not include all claims of the overall portfolio, but only those along the marked diagonal in Figure 12. In order to analyse the whole data set, the other cases (along the diagonals) would have to be taken into account analogously.

Incurred \ Reported	2010	2011	2012	2013	2014
2007	3	4	5	6	7
2008	2	3	4	5	6
2009	1	2	3	4	5
2010	0	1	2	3	4
2011	-----	0	1	2	3
2012	-----	-----	0	1	2
2013	-----	-----	-----	0	1
2014	-----	-----	-----	-----	0

Figure 12: Combinations of occurence and reporting year leading to a certain delay in reporting. By aggregating arrivals along the diagonals, the whole data set could be reconstructed. However, in the following, we only consider claims coming from the marked diagonal, i.e. with a three-year delay in reporting.

6.2 Model fitting

6.2.1 Claim payment times

The data set used in the following comes from the subclass of *legal insurance against damage compensation in consequence of incidents related to traffic*. It 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. We remark that the original data set contained sporadic payments afterwards, but we choose to restrict our focus to this period of length $T = 2400$ in order to guarantee a comparable process over the whole period. Furthermore, there were several days with more than one outgoing payment. In these cases, we distribute the indistiguishable 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). Note that the method of adding uniform random increments to distinguish arrivals with the same timestamp was e.g. used in Bowsher 2007 for trading data. However, as we find adding random timestamps to produce very unstable estimation results, we resort to deterministic uniform distribution. After preparing the data set, we essentially follow the methods used by Da Fonseca and Zaatour 2013 to analyze and quantify trade clustering for stock and futures data. The first step is to test whether the claim arrivals can be described by a memoryless Poisson distribution as assumed by the classical model. In Figure 13, 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. Figure 14 displays the number of payments per week (7 days) over the whole time period of 2400 days, where some clustering of payment occurrences over time can be observed.

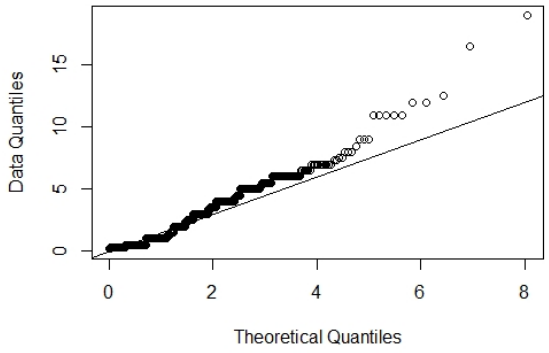


Figure 13: The plot of empirical interarrival times against an exponential distribution indicates that a Poisson model would not be a suitable fit.

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

$$AC(\tau,\delta) = \frac{\mathbb{E}[(N_{t+\tau} - N_t)(N_{t+2\tau+\delta} - N_{t+\tau+\delta})] - \mathbb{E}[(N_{t+\tau} - N_t)]\mathbb{E}[(N_{t+2\tau+\delta} - N_{t+\tau+\delta})]}{\sqrt{\mathbb{V}ar(N_{t+\tau} - N_t)\mathbb{V}ar(N_{t+2\tau+\delta} - N_{t+\tau+\delta})}}, \tag{6.1}$$

where we choose the interval length as $(7, 14, 21, 28)$ days and let the lag range from 0 to

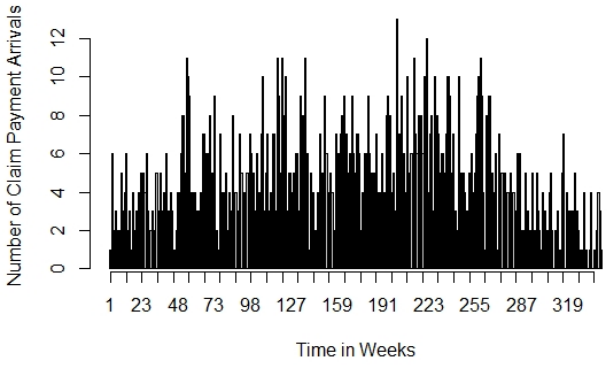


Figure 14: The number of claim payments per week during the time period 01 Jan 2010 to 28 July 2016 gives further indication as to the presence of clustering.

240 days by steps of 1. We plot the resulting autocorrelation values as a function of the time lag δ in Figure 15. 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.

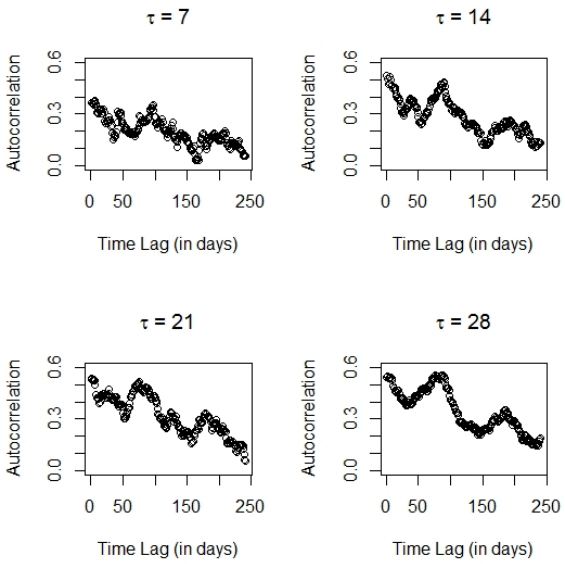


Figure 15: Empirical autocorrelation function, as in (6.1), of the number of claim payments on intervals of length τ as a function of the time lag δ between the intervals. The overall trend is decreasing for all interval lengths. Clearly, the number of claim payment arrivals on consecutive intervals can not be assumed independent.

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. To this end, we estimate the parameter set (λ, α, β) from (3.7) using the maximum likelihood estimation described in Section 3.4. 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 6.2.

Parameter	Maximum Likelihood Estimate
λ	0.1467
α	0.0260
β	0.0334
$\mathbb{E}[N(0,1]]$	0.6621
$\widehat{\mathbb{E}[N(0,1]]}$	0.6483

Table 6.2: Parameter estimates $\hat{\lambda}, \hat{\alpha}, \hat{\beta}$ from MLE as well as the estimated number of arrivals on a unit interval $\mathbb{E}[N(0,1]] = \frac{\hat{\lambda}}{1-\hat{\alpha}/\hat{\beta}}$ compared with the empirical expected value $\mathbb{E}[\widehat{N(0,1]}]$. We observe that the expected number of arrivals is matched quite closely. Note that both $\hat{\alpha}$ and $\hat{\beta}$ are estimated quite close to 0.

In order to test the goodness of fit of the Hawkes process, in Figure 16 we first plot the transformed interarrival times against a unit Exponential distribution as described in Section 3.4. We observe an improvement over the fit of a classical Poisson model in Figure 13. The qualitative independence test in Figure 17a clearly shows some patterns, but we acknowledge that it is generally difficult to get a good fit for empirical data here, so we at least corroborate that the transformed interarrival times seem to be uncorrelated in Figure 17b.

Furthermore, using formulas (3.26), (3.27) and (3.29), 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 6.2 with the corresponding values of the empirical arrivals in Table 6.2, Table 6.3 and Figure 18 respectively.

Overall, these tests lead us to conclude that a risk model with claim arrivals according to a stationary Hawkes process with exponentially decaying intensity is suitable for our data set. Next, we turn our attention to accurately describing the sizes of outgoing claim payments.

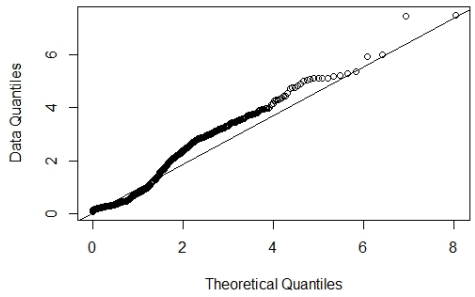
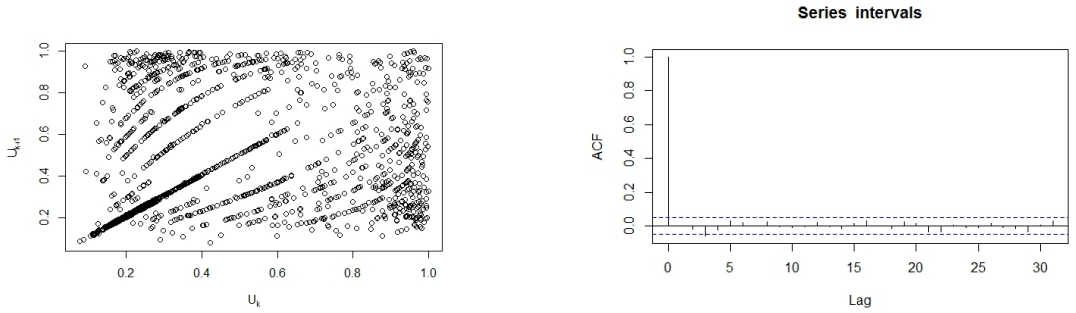


Figure 16: The plot of transformed interarrival times against a unit exponential distribution indicates that an exponential Hawkes model with parameters $(\hat{\lambda}, \hat{\alpha}, \hat{\beta})$ from Table 6.2 would be an acceptable fit.



(a) Qualitative independence test.

(b) Qualitative autocorrelation test.

Figure 17: The independence plot of the transformed interarrival times for an exponential Hawkes model with parameters $\hat{\lambda}, \hat{\alpha}, \hat{\beta}$ from Table 6.2 clearly shows some patterns. However, note that in general it is very difficult to get a good fit for empirical data for this qualitative independence test and the autocorrelation function over time of the transformed interarrival times at least indicates that they are (almost) uncorrelated.

6.2.2 Claim payment sizes

As we have described in Chapter 4, in insurance literature the claim sizes $\{Y_i\}$ are usually supposed to be i.i.d. with distribution G having finite first two moments $E[Y_1] = m_1$ and $E[Y_1^2] = m_2$. 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 the example for Theorem 4.6). Thus, we could first check whether the observed empirical claim sizes can be described by an exponential distribution. The comparison of the empirical and theoretical distribution functions and the corresponding QQ-plot in Figure 19 indicate that it would not be an acceptable approximation for this data set.

Furthermore, we have seen in Section 5.1 that for previously studied risk models with Hawkes arrivals and i.i.d. claim sizes, theoretical results are not easily applicable to em-

Interval length (days)	Theoretical Variance	Empirical Variance
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

Table 6.3: Comparison of the theoretical variance of the number of arrivals of a Hawkes process with parameters $\hat{\lambda} = 0.1467, \hat{\alpha} = 0.026, \hat{\beta} = 0.0334$ according to (3.27) to the corresponding empirical variance.

pirical data and ruin probabilities can only be obtained numerically. Thus, we would like to make use of the theoretical results for *RMGCHP* which assumes claim sizes to follow a finite number of fixed jump sizes governed by the evolution of a Markov chain. For the original use of this type of model in a limit order book context, it is natural to assume dependent jump sizes (see Swishchuk 2017a). 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 here. Thus, we reinterpret the Markov Chain (X_i) and the function $a(x)$ from Definition 5.4 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 $(b_1, b_2, \cdots, b_N = B)$ and define $\pi^* = (\pi_1^*, \cdots, \pi_N^*)$ as

$$\begin{aligned} \pi_1^* &= \hat{G}(b_1) \\ \pi_2^* &= \hat{G}(b_2) - \pi_1^* \\ &\cdots \\ \pi_N^* &= \hat{G}(b_N) - \sum_{i=1}^{N-1} \pi_i^* = 1 - \sum_{i=1}^{N-1} \pi_i^*, \end{aligned} \tag{6.2}$$

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

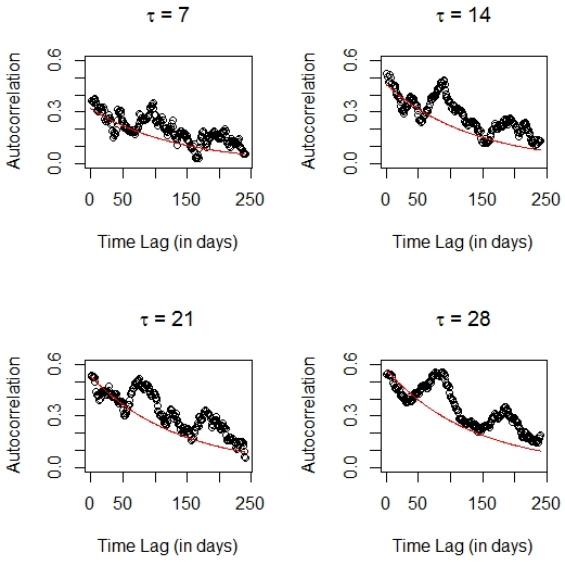


Figure 18: Empirical autocorrelation function, as in (6.1), of the number of claim payments on intervals of length τ as a function of the time lag δ between the intervals compared to the corresponding theoretical value from Proposition 3.15 for an exponential Hawkes process with parameters $\hat{\lambda} = 0.1467, \hat{\alpha} = 0.026, \hat{\beta} = 0.0334$.

Let (X_i) be a Markov Chain on $S = \{1, \cdots, N\}$ 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 by Theorem 2.17 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^*:

$$\begin{aligned} \pi^* P &= (\pi_1^*, \cdots, \pi_N^*) \begin{pmatrix} \pi_1^* & \cdots & \pi_N^* \\ \cdots & \cdots & \cdots \\ \pi_1^* & \cdots & \pi_N^* \end{pmatrix} = (\pi_1^* \sum_{i=1}^N \pi_i^*, \cdots, \pi_N^* \sum_{i=1}^N \pi_i^*) \\ &= (\pi_1^*, \cdots, \pi_N^*) = \pi^*. \end{aligned} \tag{6.3}$$

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

$$\mathbb{P}(X_{i+1} = k \mid X_i = j) = \mathbb{P}(X_{i+1} = k \mid X_i = l) = \pi_k^* \qquad \forall j, l \in S, i \in \mathbb{N} \tag{6.4}$$

and by the Markov property and the law of total probability

$$\mathbb{P}(X_{i+1} = k) = \sum_{j \in S} \mathbb{P}(X_{i+1} = k \mid X_i = j) \mathbb{P}(X_i = j)$$

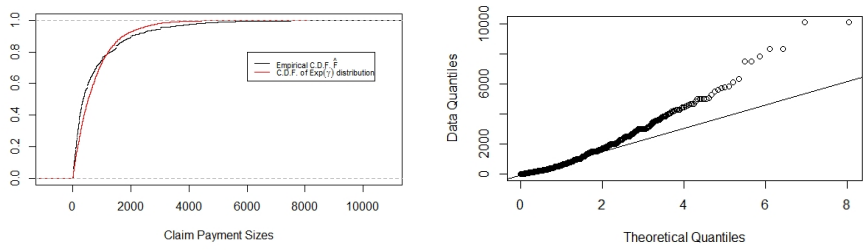


Figure 19: The plot of the empirical distribution function against an exponential $Exp(\gamma)$ distribution with matching mean and the corresponding QQ-plot indicate that empirical claim sizes should not be assumed to be i.i.d. exponentially distributed in this case.

$$= \pi_k^* \sum_{j \in S} \mathbb{P}(X_i = j) = \pi_k^* \qquad \forall k \in S, i \in \mathbb{N} \tag{6.5}$$

Thus, the probability of realizing one state is independent of the previous state and (X_i) describes an i.i.d. sequence.
Now, let Y be a r.v. with c.d.f. $\hat{G}$ and set

$$a(i) = \mathbb{E}[Y | b_{i-1} < Y \leq b_i] =: \mathbb{E}[Y|A_i] = \frac{\mathbb{E}[Y \mathbf{1}_{A_i}]}{\mathbb{P}(A_i)} = \frac{\mathbb{E}[Y \mathbf{1}_{A_i}]}{\pi_i^*}. \tag{6.6}$$

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 $N \to \infty$. The idea is illustrated graphically in Figure 20 and for the empirical example, Table 6.4 gives the values of the equidistant boundaries $b = (b_1, ..., b_N = B)$, 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. To replicate empirical claim sizes, usually significantly more states would be used, for Figure 20 and Table 6.4 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.

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

Table 6.4: Boundaries $b = (b_1, ...b_N)$, 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.

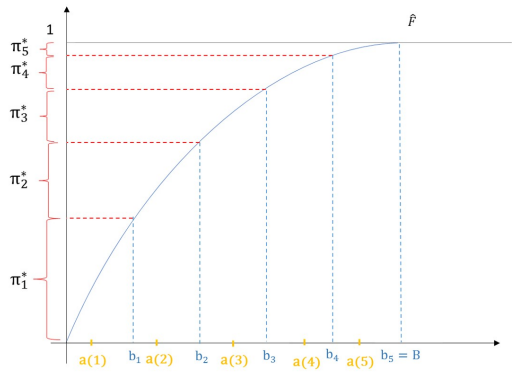


Figure 20: This figure illustrates the approach: Equidistant boundaries $(b_1, ..., b_N)$ are set, the corresponding increments of the empirical distribution function chosen as stationary distribution π^* and a Markov chain with corresponding transition probabilities is constructed. The state values $a(i)$ are set as the conditional expectations in each segment.

To corroborate that the generated claim sizes indeed describe claims with distribution function $\hat{G}$, we compare in Figure 21 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.

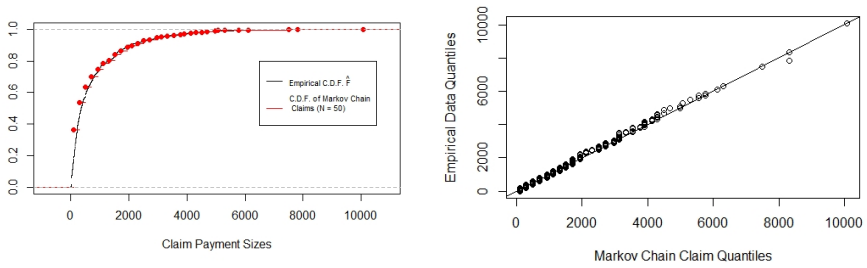


Figure 21: The distribution function of claim sizes generated by the Markov chain approach for a 50-state Markov chain against the empirical distribution function of claim sizes indicates that the approach replicates claim sizes as assumed. This is corroborated by the corresponding QQ-plot.

Overall, we conclude that empirical claim sizes are replicated reasonably well within the framework of *RMGCHP*.

6.2.3 Risk process

In order to simulate the complete risk model from Definition 5.4, 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 5.6 with a safety loading of $\theta = 0.2$ and obtain

$$c = (1+\theta)a^* \frac{\lambda}{1-\alpha/\beta} = 633.5552. \tag{6.7}$$

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 incur a claim, providing additional premium income for the company which is not considered in our data set. 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 indicating 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 $\hat{\lambda} = 0.1467, \hat{\alpha} = 0.026, \hat{\beta} = 0.0334$ and claim sizes being generated by a Markov chain with 50 states according to the procedure described above. Figure 22 compares the underlying empirical risk process with the first 50 simulations.

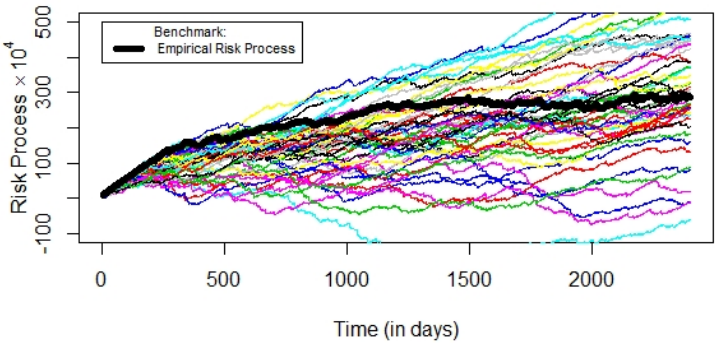


Figure 22: Plotting the empirical risk process with parameters $u = 8000$ and $c = 633.5552$ against 50 simulated paths of *RMGCHP* with arrivals following an exponential Hawkes process with $\hat{\lambda} = 0.1467, \hat{\alpha} = 0.026, \hat{\beta} = 0.0334$ and claims following a Markov chain with 50 states as described in Section 6.2.2. 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.

In order to further 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

$$\hat{S}(L) = \frac{1}{L}\sum_{i=1}^{L} \frac{\max(\hat{R}_i(t)) - \min(\hat{R}_i(t))}{\max(R(t)) - \min(R(t))} \tag{6.8}$$

Parameter	Value
$\hat{S}$	1.2032
$\hat{F}$	1.1151
$R(T)$ (Empirical)	287,829
$\mathbb{E}[\hat{R}(T)]$ (Simulation)	320,970
$\sqrt{\mathbb{V}ar[\hat{R}(T)]}$ (Simulation)	144,571

Table 6.5: 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.

$$\hat{F}(L) = \frac{1}{L}\sum_{i=1}^{L}\frac{\hat{R}_i(T)}{R(T)} \tag{6.9}$$

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 6.5 gives an overview of the results.

Overall, we conclude that *RMGCHP* is able to describe the empirical data set reasonably well.

6.2.4 Branching ratio

As mentioned before, a convenient feature of using a Hawkes process is its interpretation as an immigration-birth-process. Given the estimated parameters of our model and the empirical data set, we can compare the theoretical branching ratio $\frac{\alpha}{\beta}$ (the ratio of endogenously triggered events, i.e. secondary payments) with its empirical counterpart. Recall that for an exponential Hawkes process the branching ratio is equal to $n = \int_0^\infty \mu(s)ds = \frac{\alpha}{\beta}$. Given the estimated parameters for our empirical data, we obtain an estimated branching ratio of $\frac{\hat{\alpha}}{\hat{\beta}} = \frac{0.026}{0.0334} = 0.7784$, indicating that around 78% of incoming claims are actually generated endogenously. The "empirical branching ratio" (which can be found simply by looking at the original data set including client and claim policy number and counting secondary payments) is found to be 0.5119, indicating that the real branching ratio was overestimated by the Hawkes process. Note however that due to the "capped" time horizon some claims in the dataset are still open, which means that 0.5119 is a lower bound for the branching ratio we would obtain by observing the process until all claims are fully settled. The reason for the overestimation might be the following: An exponential Hawkes process assumes a strictly decreasing excitation function, thus the probability

of a subsequent payment being triggered strictly decreases over time. This assumption is questionable for an empirical data set as delays in payments naturally occur due to processing times, operational delays etc. making the autocorrelation decay slower than it would be without these "real-world side effects". In order to match the interarrival times and the slower autocorrelation decay, the Hawkes process estimate might tend to favor estimates with smaller values of $\hat{\beta} - \hat{\alpha}$, distorting the ratio $\frac{\hat{\alpha}}{\hat{\beta}}$ to become larger than the actual branching ratio of the original process.
In fact, some authors suggest that for empirical data (in finance), the exponential kernel is not the most favourable choice and i.e. a (long-range) power law kernel should be used (e.g. Zhang 2016). Another promising approach would be to use a model-independent branching ratio approximation as developed in Hardiman and Bouchaud 2014. For the moment, we conclude that the exponential Hawkes process is able to capture the development of arrival times over time quite accurately but tends to attribute too many events to endogenous clustering potentially due to real-world (delay) effects.

6.3 Diffusion approximations and ruin probabilities

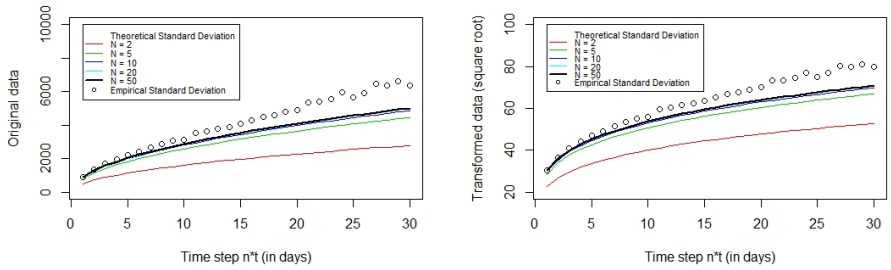
After reassuring that the *RMGCHP* is suitable for the empirical data set, we proceed to check the accuracy of the diffusion approximations introduced in Section 5.2.2. We start with Theorem 5.7, more precisely (5.13), and proceed as suggested in Swishchuk et al. 2017: We compare the standard deviation on the right-hand side of (5.13) multiplied by $\sqrt{n}$ to its empirical counterpart on the left-hand side, that is the standard deviation of

$$R(nt) - (u + cnt - a^* N(nt)) = \sum_{k=1}^{N(nt)} (a(X_k) - a^*). \tag{6.10}$$

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

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

thus considering intervals of length e.g. one day in the first step. We compare the standard deviation of these values to the standard deviation theoretically obtained on the right-hand side of (5.13) multiplied by $\sqrt{n}$, that is $\hat{\sigma}\sqrt{nt}$. 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 (6.10) is only based on few observations. To ensure statistical significance, we thus choose 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 Figure 23 and we can see that for small n they look quite accurate.



(a) Original Data. (b) $\sqrt{\cdot}$-transformed data to stabilize variance.

Figure 23: Error estimation of jump diffusion approximation by comparison of standard deviations of empirical process in (6.10) and corresponding theoretical values from Theorem 5.7. 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.

We proceed analogously for Theorem 5.8, (5.20), with $\hat{\mu} := \frac{\alpha}{\beta}$ for the exponential Hawkes process. This time we compare the standard deviation of

$$R(nt) - (u + cnt - a^* \frac{\lambda}{1 - \alpha/\beta} nt) = \left(\sum_{k=1}^{N(nt)} a(X_k) - a^* \frac{\lambda}{1 - \alpha/\beta} nt \right) \tag{6.11}$$

with its counterpart on the right-hand side in (5.20), that is $\sqrt{nt}\bar{\sigma}$. The results can be seen on the left-hand side of Figure 24. 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. As we have seen in Section 5.2.2, the variance of the pure diffusion and the jump diffusion approximation are given by

$$\mathbb{V}ar[R_p(t)] = (\bar{\sigma})^2 t$$

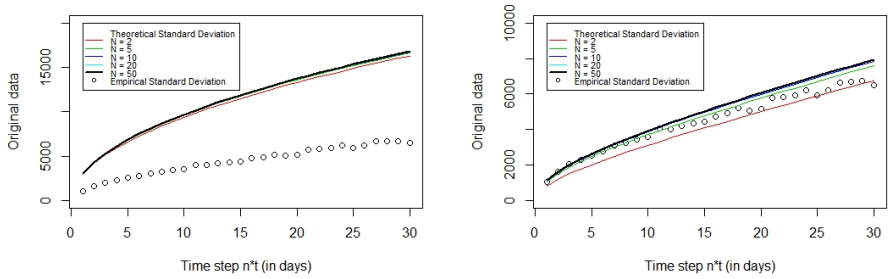
and

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

respectively, where

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

We again note that the pure diffusion approximation naturally does not capture the (negative) non-linear influence $(**)$ on the variance. Indeed, if we plot the standard deviation of (6.11) against $\sqrt{\bar{\sigma}^2 n - (**)(nt)}$ (the theoretical jump-diffusion standard deviation) on the right-hand side of Figure 24, we see a very close match. However, the absolute value



(a) Original Data.

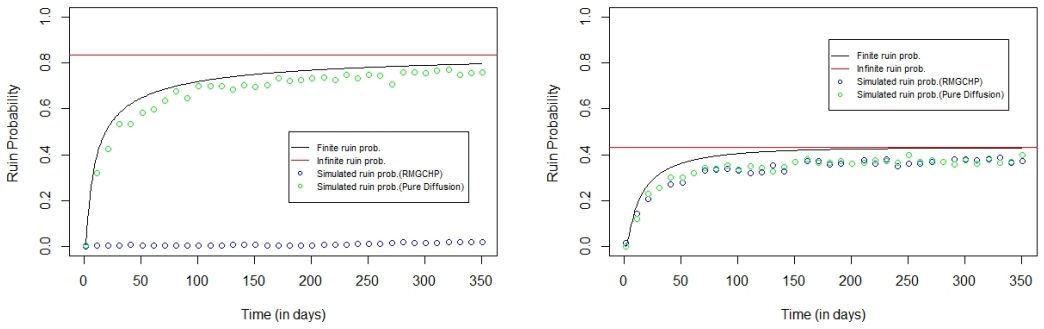
(b) Original Data, with correction term (∗∗) from (5.19).

Figure 24: (a) Error estimation of pure diffusion approximation by comparison of standard deviations of empirical process in (6.11) and corresponding theoretical values from Theorem 5.8. 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 5.8, corrected by (∗∗) from (5.19), shows a very close match of theoretical and empirical values.

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.

In Section 5.2.2, we have seen how the pure diffusion approximation can be used to estimate finite and infinite time ruin probabilities and would finally like to check the accuracy of those results empirically. In Figure 25 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 and compare them with values obtained from $L = 1000$ simulations of the corresponding *RMGCHP* and of the pure diffusion approximation process. For the original data set, clearly the calculated ruin probabilities overestimate the simulated ones, an effect that originates from the large estimation error of the pure diffusion approximation. 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 as displayed in Figure 25.

We conclude that the pure diffusion approximation of *RMGCHP* is generally suitable for calculating ruin probabilities for a risk model with Hawkes 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



(a) $(\lambda, \alpha, \beta) = (0.1467, 0.026, 0.0334)$ (b) $(\lambda, \alpha, \beta) = (1, 2, 5)$

Figure 25: (a) We calculate finite-time ruin probabilities for intervals $[0, t)$, where t ranges from 1 to 350 days by steps of 1, and the infinite-time ruin probability according to Theorem 5.9. 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 *RMGCHP* and the pure diffusion approximation process. For this case, there is a large discrepancy between *RMGCHP* and the pure diffusion approximation, as seen in Figure 24, due to the small difference in β and α. (b) Changing the parameters of the Hawkes arrival process to $(\lambda, \alpha, \beta) = (1, 2, 5)$ and leaving everything else equal (apart from the premium rate, which has to be calculated anew as $c = 1594.734$) leads to a significant increase in the accuracy of the ruin probability estimation. This is due to the fact that with increasing $\beta - \alpha$, the influence of the inaccuracy $(**)$ diminishes.

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. In the next chapter, we will look at another application of the pure diffusion approximation of *RMGCHP*.

Chapter 7

Optimal investment

7.1 Background and market model

The goal of this chapter is to apply the results for *RMGCHP* in order to investigate 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. In general, portfolio selection refers to the challenge of allocating the available wealth of an individual or company (from now on: an investor) into different assets with the goal of optimizing some criterion. This criterion depends on the investor's preferences and restraints, but typically involves maximizing factors like the expected return or expected utility of final wealth or minimizing financial risk that occurs as the future developments of most assets are uncertain. Modern portfolio theory is based on the work of Markowitz 1952 who was the first to highlight the importance of *diversification* for efficient portfolio selection. Markowitz uses the expected return (assuming the investor c.p. always deems more return better than less) as well as the variance of returns (which is assumed as non-desirable and should be limited) as the criteria for portfolio selection. Therefore, for each given expected return value, the portfolio with the smallest variance, i.e. the *efficient* portfolio is searched (or vice versa, for each given risk value, the portfolio with the highest expected return). Plotting all efficient combinations of risk and return in the mean-variance plane gives the *mean-variance efficient (MVE) frontier*. Ever since the introduction of this concept and the work of Markowitz 1956 who found the expression of the MVE frontier for a single-period setting, a substantial amount of research has been devoted to generalisations and extensions. Examples would be the introduction of consumption (Merton 1975) or labor income (Koo 1998) into the model. Another important step was to generalize beyond a one-period setting. In multi-period portfolio selection, it has been very common to consider maximizing expected utility functions of the terminal

wealth, i.e. $\mathbb{E}[U(X(T))]$, where U is a utility function describing the investor's preferences, e.g. of a power, log or exponential form. This is due to the fact that dynamic programming principles are accessible for these formulations due to the *tower property*, implying $\mathbb{E}[\mathbb{E}[U(X(T))|\mathcal{F}_m]\mathcal{F}_n] = \mathbb{E}[U(X(T))|\mathcal{F}_n]$, where $\{\mathcal{F}_k\}$ is the underlying filtration and $n \leq m$. However, an expected utility setting often makes the tradeoff between risk and expected return more implicit and thus the investment decision less intuitive. To use an explicit mean-variance formulation, a term of the form $U(\mathbb{E}[X(T)])$, where U is a nonlinear (specifically quadratic) utility function, is involved. This is an important point as it renders the tower property inaccessible. Due to this "inconvenience" of the mean–variance objective function for dynamic programming, it has first been challenging to generalize Markowitz' result to a multi-period or continuous-time setting. Analytical solutions for the mean-variance setting were finally derived by respectively Li and Ng 2000 and Zhou and Li 2000, who use the framework of stochastic LQ control. Li, Chan, and Ng 1998 generalized the setting to consider safety-first portfolio selection while Zhu, Li, and Wang 2004 and Bielecki et al. 2005 considered bankruptcy control or bankruptcy prohibition for the mean–variance portfolio selection in multi-period or continuous-time settings, respectively. When working in the insurance setting, it is important to keep in mind that any portfolio management decisions should be subject to risk management considerations, in particular keeping in mind potential future cash outflows from the insurance business that have to be covered. The fields of *asset liability management (ALM)* and *liability-driven investment strategies (LDI)* deal with this challenge and provide a holistic framework that takes both sides of a company's balance sheet into account when determining investment decisions. For an overview, the book by Zenios and Ziemba 2007 is recommended. It is clear that if we want to study optimal investment in the insurance context, considering a liability should be included. At the same time, the mean-variance approach has the advantage of being intuitively understandable and explicitly displaying the trade-off between risk and reward. Furthermore, Bäuerle 2005 highlights the relevance of the mean-variance criterion in insurance by using it to find the optimal dynamic proportional reinsurance strategy. Therefore, we turn our attention to research on mean-variance portfolio selection with liability, which was first investigated in a single-period setting by Sharpe and Tint 1990. Based on the work of Li and Ng 2000, Leippold, Trojani, and Vanini 2004 derived explicitly the optimal strategy and the MVE frontier for a multi-period asset–liability management problem. Chiu and Li 2006 worked on the analogous problem in a continuous-time setting. Xie, Li, and Wang 2008 generalize their results by considering an incomplete market where the evolution of assets and liability are possibly correlated. They use an embedding technique similar to the one in Zhou and Li 2000, but based on the more general stochastic LQ control technique from Yong and Zhou 1999. As their result is the more general one, we will work with the setting of Xie, Li, and Wang

2008, where in particular the liability is supposed to follow a Brownian motion with drift. Thus, the idea is to use the pure diffusion approximation of *RMGCHP* to represent the (negative) liability and then employ the results from Xie, Li, and Wang 2008 to show how the MVE frontier, optimal strategy and attainable expected return change under a Hawkes process given a fixed risk constraint (i.e. a boundary on the variance of terminal wealth). This could be viewed as an approach to quantifying the impact of clustering in the claims arrival process on the whole (insurance and investment management) business of the company.

We start by formalizing the market model and the optimization problem at hand as in Xie, Li, and Wang 2008. 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 $\{(W_0(t), W_1(t), \cdots, W_n(t))' : t \in [0,T]\}$ 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 $\mathbb{R}^{n\times k}$. Consider a financial market with $(m+1)$ assets being traded continuously, where $m \leq n$ and the assets are labelled $i = 0, 1, ..., m$, where $i = 0$ refers to the riskfree asset. The price of the risk-free asset $S_0(t)$ evolves according to the ODE

$$dS_0(t) = r(t)S_0(t)dt, \qquad S_0(0) = 1, \tag{7.1}$$

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

$$\begin{aligned} dS_i(t) &= \mu_i(t)S_i(t)dt + \sigma_i(t)S_i(t)dW(t), \qquad S_i(0) = s_i \in \mathbb{R} \\ &i = 1, \cdots, m \end{aligned} \tag{7.2}$$

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

$$\sigma(t) := (\sigma_1(t), \cdots, \sigma_m(t))' \in \mathcal{C}([0,T]; \mathbb{R}^{m\times n}) \tag{7.3}$$

and let

$$\mu(t) := (\mu_1(t), \cdots, \mu_m(t))' \in \mathcal{C}([0,T]; \mathbb{R}^{m\times 1}) \tag{7.4}$$

denote the *rate of return* of the risky assets. As described before, the company also has to take into a account a stochastic liability which is modelled as a Brownian motion with drift. Denote the company's cumulative liability at time t by $L(t)$ and assume $L(t)$ evolves according to the SDE

$$dL(t) = g(t)dt + \upsilon(t)dB(t), \qquad L(0) = l, \tag{7.5}$$

where $B(t)$ is a standard Brownian motion.
Denote by $\rho_j(t)$ the correlation coefficient between $B(t)$ and $W_j(t)$ for $j = 1, \cdots, n$ and let

$$\rho(t) := (\rho_1(t), \cdots, \rho_n(t))' \in \mathcal{C}([0,T]; \mathbb{R}^{n\times 1}) \tag{7.6}$$

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

$$B(t) = \rho(t)'W(t) + \sqrt{1-\rho(t)'\rho(t)}W_0(t) \tag{7.7}$$

where $B(t), W_0(t), W_1(t), \cdots, W_n(t)$ are standard Brownian motions and $\rho(t)'\rho(t) \leq 1$ for all $t \in [0,T]$. Combining (7.5) and (7.7), $L(t)$ evolves according to

$$dL(t) = g(t)dt + \upsilon(t)\rho(t)'dW(t) + \upsilon(t)\sqrt{1-\rho(t)'\rho(t)}dW_0(t), \qquad L_0 = l \tag{7.8}$$

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 $i = 1, ..., m$ and $j = 1, ..., n$, $r(t), \mu_i(t), \sigma(t)^{ij}, g(t), \upsilon(t), \rho(t)^j$ are deterministic functions of t, $\mu_i(t) > r(t)$, 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.

Definition 7.1 (Trading strategy)**.** 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 $\varphi(t) := (\varphi_1(t), \cdots, \varphi_m(t))'$, then we call the process $\varphi := \{\varphi(t) : t \in [0,T]\}$ a *trading strategy*.

Remark. One could alternatively define $\eta(t) := (\eta_1(t), \cdots, \eta_m(t))'$ as a trading strategy, but the above formulation is the commonly used one within the framework of stochastic LQ problems as it allows to deduce the state equation (wealth) in a convenient form.

We assume that the company can dynamically adjust its investment portfolio during the time period $[0,T]$ without incurring transaction fees or short-selling restrictions. Furthermore, we only consider *self-financing* trading strategies, which means that no money can be taken out (no consumption) or additionally added (no other income) to the investment portfolio during the observed period. This implies the following:

Definition 7.2 ((Net) Wealth process)**.** Let $X(t)$ be the *net wealth* of the company at time t, 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 $x = w - l > 0$. Then $X(t)$ fulfils

$$dX(t) = \sum_{i=0}^{m} \eta_i(t)dS_i(t) - dL(t), \qquad X_0 = x > 0. \tag{7.9}$$

Inserting (7.1),(7.2) and (7.8), (7.9) is equivalent to

$$\begin{aligned}
dX(t) &= \sum_{i=0}^{m} \eta_i(t) dS_i(t) - dL(t) \\
&= \eta_0(t) r(t) S_0(t) dt + \sum_{i=1}^{m} \eta_i(t) S_i(t) (\mu_i(t) dt + \sigma_i(t) dW(t)) - g(t) dt - \upsilon(t) \rho(t)' dW(t) \\
&\quad - \upsilon(t) \sqrt{1 - \rho(t)' \rho(t)} dW_0(t) \\
&= \varphi_0(t) r(t) dt + \sum_{i=1}^{m} \varphi_i(t) (\mu_i(t) dt + \sigma_i(t) dW(t)) - g(t) dt - \upsilon(t) \rho(t)' dW(t) \\
&\quad - \upsilon(t) \sqrt{1 - \rho(t)' \rho(t)} dW_0(t) + \sum_{i=1}^{m} \varphi_i(t) r(t) dt - \sum_{i=1}^{m} \varphi_i(t) r(t) dt \\
&= X(t) r(t) dt + \varphi(t) (\mu(t) - r(t) \mathbb{1}) dt - g(t) dt + \varphi(t)' \sigma(t) dW(t) - \upsilon(t) \rho(t)' dW(t) \\
&\quad - \upsilon(t) \sqrt{1 - \rho(t)' \rho(t)} dW_0(t) \\
&= (r(t) X(t) + b(t) \varphi(t) - g(t)) dt + (\varphi(t)' \sigma(t) - (\upsilon(t) \rho(t))') dW(t) \\
&\quad - \upsilon(t) \sqrt{1 - \rho(t)' \rho(t)} dW_0(t),
\end{aligned}$$

where $b(t) = \mu(t) - r(t)\mathbb{1}$ and $\mathbb{1}$ denotes the unit vector of length m.
In summary, (7.9) becomes

$$\begin{aligned}
dX(t) &= (r(t) X(t) + (\mu(t) - r(t) \mathbb{1})' \varphi(t) - g(t)) dt + (\varphi(t)' \sigma(t) - (\upsilon(t) \rho(t))') dW(t) \\
&\quad - \upsilon(t) \sqrt{1 - \rho(t)' \rho(t)} dW_0(t) \\
X(0) &= x.
\end{aligned} \tag{7.10}$$

Remark. If $\varphi(t)$ is a self-financing trading strategy s.t. the wealth process is positive $\mathbb{P}$-a.s. for all $t \in [0, T]$, we can define a *self-financing relative portfolio process* as $\pi(t) := (\pi_1(t), \cdots, \pi_m(t))'$ where

$$\pi_i(t) = \frac{\varphi_i(t)}{X(t)},$$

and $\pi_0(t) := 1 - \pi(t)'\mathbb{1}$. Thus $\pi(t)$ describes the fraction of the total wealth invested in each asset at time t. This can sometimes be more convenient as the weights add up to 1 and thus strategies are more easily comparable across assets with different price processes. In the classical market model (complete market without liability), (7.10) can be expressed in terms of π, which turns out not to be convenient in the present market model.

We now limit the set of all trading strategies to so-called *admissible* ones.

Definition 7.3 (Set of admissible trading strategies)**.** The *set of admissible trading strategies* for initial wealth x is defined as

$$\mathcal{A}(x) := \left\{ \varphi : \varphi(t) \in \mathcal{L}^2_{\mathcal{F}}([0, T]; \mathbb{R}^m), (X(t), \varphi(t)) \text{ satisfies (7.10)} \right\} \tag{7.11}$$

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

$$\left(\mathbb{E}\left[\int_0^T |\epsilon(t)|^2 dt\right]\right)^{\frac{1}{2}} < \infty, \qquad \forall \epsilon(t) \in \mathcal{L}^2_{\mathcal{F}}([0,T];\mathbb{R}^m).$$

We are now ready to introduce the optimization problem whose solution describes the mean-variance efficient frontier in our market model. A strategy $\varphi \in \mathcal{A}(x)$ is considered *optimal* if it solves the optimization problem

$$P(\chi) \quad \min_{\varphi \in \mathcal{A}(x)} \; (-\mathbb{E}[X(T)] + \chi \mathbb{V}ar[X(T)]) \tag{7.12}$$

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. Define

$$\varphi_{P(\chi)} := \{\varphi : \varphi \text{ is an optimal strategy of } P(\chi)\}.$$

Unfortunately, $P(\chi)$ is not a standard stochastic optimal control problem and is hard to solve directly due to the term $\mathbb{E}[X(T)]^2$ in its cost function. We therefore review the results of Zhou and Li 2000 who show how to infer from the original problem (7.12) an auxiliary problem which is a standard stochastic optimal linear-quadratic problem. The optimal solution of (7.12) can then be located using the solution of the auxiliary problem. This allows us to obtain the mean-variance efficient frontier for the original problem in closed form as shown in Xie, Li, and Wang 2008.

7.2 Optimal strategy and mean-variance efficient frontier

In this section, we will review the solution of the problem $P(\chi)$ in order to obtain the mean-variance efficient frontier in closed form for the market model in Section 7.1. This relies on the work of Zhou and Li 2000 and Xie, Li, and Wang 2008. To give an overview of general stochastic optimal control theory (see Yong and Zhou 1999) would be outside the scope of this thesis, but the basic terms that are needed here are reviewed in Chapter 2. Xie, Li, and Wang 2008 introduce the *auxiliary problem*

$$A(\chi,\omega) \quad \min_{\varphi \in \mathcal{A}(x)} \; \mathbb{E}[\chi X^2(T) - \omega X(T)] \tag{7.13}$$

and define

$$\varphi_{A(\chi,\omega)} := \{\varphi : \varphi \text{ is an optimal strategy of } A(\chi,\omega)\},$$

where $\omega \in \mathbb{R}$ is given beforehand. The following theorem shows the relationship between problems $P(\chi)$ and $A(\chi, \omega)$.

Theorem 7.4 (Zhou and Li 2000)**.** For any $\chi > 0$, it holds that

$$\varphi_{P(\chi)} \subseteq \bigcup_{\omega \in \mathbb{R}} \varphi_{A(\chi,\omega)}.$$

Moreover, if $\varphi^* \in \varphi_{P(\chi)}$, then $\varphi^* \in \varphi_{A(\chi,\omega)}$ with $\omega^* = 1 + 2\chi\mathbb{E}[X^*(T)]$, where $X^*(t)$ is the wealth process associated with the strategy φ^*.

Proof. See Zhou and Li 2000. □

The theorem implies that any optimal solution of $P(\chi)$ (as long as it exists) can be found by solving problem $A(\chi, \omega)$. The auxiliary problem $A(\chi, \omega)$ is now a standard stochastic optimal control problem parameterized by (χ, ω) which has an objective function of the form $\mathbb{E}[U(X(T))]$ as well as a LQ structure. To solve the auxiliary problem, first define

$$\gamma := \frac{\omega}{2\chi} \text{ and } Y(t) := X(t) - \gamma.$$

Thus, (7.10) becomes

$$\begin{aligned} dY(t) &= (r(t)Y(t) + (\mu(t) - r(t)\mathbb{1})'\varphi(t) + r(t)\gamma - g(t))dt + (\varphi(t)'\sigma(t) - (\upsilon(t)\rho(t))')dW(t) \\ &\quad - \upsilon(t)\sqrt{1 - \rho(t)'\rho(t)}dW_0(t) \\ Y(0) &= y := x - \gamma, \end{aligned} \tag{7.14}$$

and the objective of (7.13) becomes

$$\begin{aligned} \mathbb{E}[\chi(Y(T) + \gamma)^2 - \omega(Y(T) + \gamma)] &= \mathbb{E}\left[\chi Y^2(T) + 2\chi Y(T)\frac{\omega}{2\chi} + \chi\frac{\omega^2}{4\chi^2} - \omega Y(T) - \frac{2 \cdot \omega^2}{2 \cdot 2\chi}\right] \\ &= \mathbb{E}\left[\chi Y^2(T) - \frac{\omega^2}{4\chi}\right]. \end{aligned}$$

Thus, the auxiliary problem is equivalent to

$$A(\gamma) \quad \min_{\varphi \in \Lambda(y)} \mathbb{E}\left[\frac{\chi}{2}Y^2(T)\right] =: \min_{\varphi \in \Lambda(y)} J(\varphi, \gamma), \tag{7.15}$$

where the set of admissible strategies is given as

$$\Lambda(y) := \{\varphi : \varphi(t) \in \mathcal{L}^2_{\mathcal{F}}\big([0,T]; \mathbb{R}^m\big), (Y(t), \varphi(t)) \text{ fulfils (7.14)}\}.$$

The *optimal cost functional* of (7.15) is defined as

$$J^* := J(\varphi^*, \gamma) = \inf_{\varphi \in \Lambda(y)} J(\varphi, \gamma),$$

where φ^* is the optimal solution of $A(\gamma)$.
So now the first step is to solve $A(\gamma)$. To this end, Xie, Li, and Wang 2008 combine the results from Zhou and Li 2000 and general stochastic LQ optimal control theory in Yong and Zhou 1999 to obtain the following result.

Theorem 7.5 (Optimal feedback control (Xie, Li, and Wang 2008))**.** Consider the Riccati differential equation

$$\begin{aligned} &\frac{d\varrho(t)}{dt} + (2r(t) - \varsigma(t))\varrho(t) = 0, \\ &\varrho(T) = \chi, \\ &\varrho(t)\sigma(t)\sigma(t)' > 0 \text{ for a.e. } t \in [0,T], \end{aligned} \tag{7.16}$$

along with the adjoint ODE

$$\begin{aligned} &\frac{d\phi(t)}{dt} + (r(t) - \varsigma(t))\phi(t) + (r(t)\gamma + \kappa(t))\varrho(t) = 0, \\ &\phi(T) = 0, \end{aligned} \tag{7.17}$$

where

$$\tau(t) := (\sigma(t)\sigma(t)')^{-1}(\mu(t) - r(t)\mathbb{1})' \in \mathcal{C}([0,T];\mathbb{R}^{m\times 1}), \tag{7.18}$$

$$\zeta(t) := (\sigma(t)\sigma(t)')^{-1}\sigma(t)(-\upsilon(t)\rho(t)) \in \mathcal{C}([0,T];\mathbb{R}^{m\times 1}), \tag{7.19}$$

$$\varsigma(t) := (\mu(t) - r(t)\mathbb{1})\tau(t) \in \mathcal{C}([0,T];\mathbb{R}^{+}), \tag{7.20}$$

$$\kappa(t) := -(\mu(t) - r(t)\mathbb{1})\zeta(t) - g(t) \in \mathcal{C}([0,T];\mathbb{R}). \tag{7.21}$$

If (7.16) and (7.17) have solutions $\varrho(t) \in \mathcal{C}([0,T];\mathbb{R}^{+})$ and $\phi(t) \in \mathcal{C}([0,T];\mathbb{R})$ respectively, then the stochastic LQ problem $A(\gamma)$ has an optimal feedback control

$$\varphi^*(Y(t)) = -\tau(t)(Y(t) + \vartheta(t)) - \zeta(t), \qquad t \in [0,T]. \tag{7.22}$$

The corresponding optimal cost functional is given by

$$\begin{aligned} J^* = \frac{1}{2}\int_0^T &[-(\tau(t)\vartheta(t) + \zeta(t))'(\sigma(t)\sigma(t)')(\tau(t)\vartheta(t) + \zeta(t))\varrho(t) \\ &+ 2(r(t)\gamma - g(t))\phi(t) + \upsilon^2(t)\varrho(t)]dt + \frac{1}{2}\varrho(0)y^2 + \phi(0)y, \end{aligned} \tag{7.23}$$

where $\vartheta(t) := \frac{\phi(t)}{\varrho(t)}$.

Remark. Note that problem $A(\gamma)$ is a special case of (SLQ) in Definition 2.31 with coefficients

$$A(t) = r(t),\ B(t) = \mu(t) - r(t)\mathbb{1},\ b(t) = \gamma r(t) - g(t)$$

$$C_j(t) = 0,\ D_j(t) = \sigma_j(t),\ \sigma_j(t) = -\nu(t)\rho_j(t),\ \forall j = 1, \cdots, n$$
$$Q(t), R(t), S(t) \equiv 0,\ G = \chi,\ (s, y) = (0, x - \gamma).$$

Inserting these coefficients into (2.9) and (2.10) yields (7.16) and (7.17) respectively. Analogously, (2.11) becomes (7.22) and (2.12) simplifies to (7.23). The mean-variance problem as a special case of a stochastic LQ problem is further studied in Yong and Zhou 1999 (Chapter 6, Section 8). We will now prove the theorem.

Proof. The first step is to show that $\varphi^*(t)$ is an admissible solution. Inserting (7.22) into (7.14) yields

$$\begin{aligned}
dY(t) &= (r(t)Y(t) + (\mu(t) - r(t)\mathbb{1})'(-\tau(t)(Y(t) + \vartheta(t)) - \zeta(t)) + r(t)\gamma - g(t))dt \\
&\quad + ((-\tau(t)(Y(t) + \vartheta(t)) - \zeta(t))'\sigma(t) - (\upsilon(t)\rho(t))')dW(t) - \upsilon(t)\sqrt{1 - \rho(t)'\rho(t)}dW_0(t) \\
&= ((r(t) - (\mu(t) - r(t)\mathbb{1})'\tau(t))Y(t) - (\mu(t) - r(t)\mathbb{1})'\tau(t)\vartheta(t) - (\mu(t) - r(t)\mathbb{1})'\zeta(t) \\
&\quad + r(t)\gamma - g(t))dt + (-\tau(t)'\sigma(t)Y(t) - \tau(t)'\sigma(t)\vartheta(t) - \zeta(t)'\sigma(t) - (\upsilon(t)\rho(t))')dW(t) \\
&\quad - \upsilon(t)\sqrt{1 - \rho(t)'\rho(t)}dW_0(t) \\
&= [(r(t) - \varsigma(t))Y(t) - \varsigma(t)\vartheta(t) + \kappa(t) + r(t)\gamma]dt + (-\tau(t)'\sigma(t)Y(t) - \tau(t)'\sigma(t)\vartheta(t) \\
&\quad - \zeta(t)'\sigma(t) - (\upsilon(t)\rho(t))')dW(t) - \upsilon(t)\sqrt{1 - \rho(t)'\rho(t)}dW_0(t) \\
Y(0) &= y.
\end{aligned} \tag{7.24}$$

As $\varrho(t) \in \mathcal{C}([0, T]; \mathbb{R}^+)$ and $\phi(t) \in \mathcal{C}([0, T]; \mathbb{R})$, it is also $\vartheta(t) \in \mathcal{C}([0, T]; \mathbb{R})$. Therefore, the drift and diffusion coefficient in (7.24) are Lipschitz continuous and by Theorem 2.9 in Karatzas and Shreve 1998 it follows that (7.24) admits a unique strong solution $Y^*(t)$ such that

$$\mathbb{E}[\sup_{t\in[0,T]} |Y^*(t)|^2] \leq K(T)(1 + |y|^2),$$

where $K(T) > 0$ is constant associated with the terminal time. Thus, $\varphi^*(Y(t)) \in \Lambda(y)$ is an admissible control. For any $\varphi \in \Lambda(y)$, let $Y(t)$ be the state variable associated with the control φ. We now show that $\varphi^*(Y(t))$ is an optimal feedback control of state variable $Y(t)$. We first use Itô's formula to obtain $\frac{1}{2}d(\varrho(t)Y^2(t))$ and $d(\phi(t)Y(t))$. First note that by Itô's formula with $f(t, x(t)) = x^2$,

$$\begin{aligned}
d(Y^2(t)) = \big(&2Y(t)(r(t)Y(t) + (\mu(t) - r(t)\mathbb{1})'\varphi(t) + r(t)\gamma - g(t)) \\
&+ 0.5 \cdot 2[(\varphi(t)'\sigma(t) + \delta(t)')^2 + (\delta_0(t))^2]\big)dt \\
&+ 2Y(t)(\varphi(t)'\sigma(t) + \delta(t)')dW(t) + 2Y(t)\delta_0(t)dW_0(t),
\end{aligned}$$

where $\delta(t) := -(\upsilon(t)\rho(t))$ and $\delta_0(t) := -\upsilon(t)\sqrt{1 - \rho(t)'\rho(t)}$. By (7.16) we know $d\varrho(t) = -(2r(t) - \varsigma(t))\varrho(t)dt$, therefore by the product rule it holds

$$\frac{1}{2}d(\varrho(t)Y^2(t)) = \frac{1}{2}\big[-Y^2(t)(2r(t) - \varsigma(t))\varrho(t) + \varrho(t)[2r(t)Y^2(t) + 2Y(t)(\mu(t) - r(t)\mathbb{1})'\varphi(t)$$

$$
\begin{aligned}
&+ 2Y(t)r(t)\gamma - 2Y(t)g(t) + (\varphi(t)'\sigma(t) + \delta(t)')^2 + (\delta_0(t))^2]\big]dt + \\
&+ \frac{1}{2}\varrho(t)\cdot 2\cdot Y(t)(\varphi(t)'\sigma(t) + \delta(t)')dW(t) + \frac{1}{2}\varrho(t)\cdot 2\cdot Y(t)\delta_0(t)dW_0(t) \\
&= \frac{1}{2}\Big[\varsigma(t)\varrho(t)Y^2(t) + 2Y(t)\varrho(t)\big((\mu(t) - r(t)\mathbb{1})'\varphi(t) + r(t)\gamma - g(t)\big) \\
&+ \big(\varphi(t)'\sigma(t)\sigma(t)'\varphi(t) + 2\varphi(t)'\sigma(t)\delta(t) + \delta(t)'\delta(t) + (\delta_0(t))^2\big)\varrho(t)\Big]dt \\
&+ \varrho(t)Y(t)(\varphi(t)'\sigma(t) + \delta(t)')dW(t) + \varrho(t)Y(t)\delta_0(t)dW_0(t).
\end{aligned}
$$

Likewise, as from (7.17) we know $d\phi(t) = -(r(t) - \varsigma(t))\phi(t) - (r(t)\gamma + \kappa(t))\varrho(t)$ and thus by the product rule

$$
\begin{aligned}
d(\phi(t)Y(t)) = &\big[\phi(t)(r(t)Y(t) + (\mu(t) - r(t)\mathbb{1})'\varphi(t) + r(t)\gamma - g(t)) + Y(t)(-(r(t) - \varsigma(t))\phi(t) \\
&- (r(t)\gamma + \kappa(t))\varrho(t))\big]dt + \phi(t)(\varphi(t)'\sigma(t) + \delta(t)')dW(t) + \phi(t)\delta_0(t)dW_0(t) \\
&= \big[Y(t)\varsigma(t)\phi(t) - (r(t)\gamma + \kappa(t))\varrho(t)Y(t) + \phi(t)\big((\mu(t) - r(t)\mathbb{1})'\varphi(t) \\
&+ r(t)\gamma - g(t)\big)\big]dt + \phi(t)(\varphi(t)'\sigma(t) + \delta(t)')dW(t) + \phi(t)\delta_0(t)dW_0(t).
\end{aligned}
$$

Note that integrating the above equations from 0 to T, taking expectations and adding them together yields

$$
\begin{aligned}
&\mathbb{E}\Big[\int_0^T \frac{1}{2}d(\varrho(t)Y^2(t))\Big] + \mathbb{E}\Big[\int_0^T d(\phi(t)Y(t))\Big] = \\
&= \mathbb{E}\Big[\frac{1}{2}\underbrace{\varrho(T)}_{=\chi}Y^2(T) - \frac{1}{2}\varrho(0)Y^2(0)\Big] + \mathbb{E}[\underbrace{\phi(T)}_{=0}Y(T) - \phi(0)Y(0)] \\
&= \mathbb{E}\Big[\frac{\chi}{2}Y^2(T)\Big] - \frac{1}{2}\varrho(0)y^2 - \phi(0)y \\
&= J(\varphi,\gamma) - \frac{1}{2}\varrho(0)y^2 - \phi(0)y
\end{aligned}
$$

where we have used the terminal values in (7.16) and (7.17) and the definition of $J(\varphi,\gamma)$. Inserting the expressions from above (where we use that for our integrands the expectation of an integral w.r.t. a Brownian motion is 0) gives

$$
\begin{aligned}
&J(\varphi,\gamma) - \frac{1}{2}\varrho(0)y^2 - \phi(0)y \\
&= \frac{1}{2}\mathbb{E}\Big[\int_0^T \Big(\varsigma(t)\varrho(t)Y^2(t) + 2Y(t)\varrho(t)\big((\mu(t) - r(t)\mathbb{1})'\varphi(t) + r(t)\gamma - g(t)\big) \\
&+ \big(\varphi(t)'\sigma(t)\sigma(t)'\varphi(t) + 2\varphi(t)'\sigma(t)\delta(t) + \delta(t)'\delta(t) + (\delta_0(t))^2\big)\varrho(t) \\
&+ 2Y(t)\varsigma(t)\phi(t) - 2(r(t)\gamma + \kappa(t))\varrho(t)Y(t) + 2\phi(t)\big((\mu(t) - r(t)\mathbb{1})'\varphi(t) + r(t)\gamma - g(t)\big)\Big)dt\Big]
\end{aligned}
$$

$$= \frac{1}{2}\mathbb{E}\Big[\int_0^T \Big(\varsigma(t)\varrho(t)Y^2(t) + 2Y(t)\varrho(t)\big((\mu(t) - r(t)\mathbb{1})'\varphi(t) - \kappa(t) - g(t)\big)$$
$$+ \big(\varphi(t)'\sigma(t)\sigma(t)'\varphi(t) + 2\varphi(t)'\sigma(t)\delta(t) + \delta(t)'\delta(t) + (\delta_0(t))^2\big)\varrho(t)$$
$$+ 2Y(t)\varsigma(t)\phi(t) + 2\phi(t)\big((\mu(t) - r(t)\mathbb{1})'\varphi(t) + r(t)\gamma - g(t)\big)\Big)dt\Big]$$

Resubstituting the expressions for $\varsigma(t) = \tau(t)'(\sigma(t)\sigma(t)')\tau(t)$, $(\mu(t) - r(t)\mathbb{1}) = (\sigma(t)\sigma(t)')\tau(t)$, $\phi(t) = \vartheta(t)\varrho(t)$ and $\kappa(t), \delta(t)$ and $\delta_0(t)$ as above yields

$$J(\varphi,\gamma) - \frac{1}{2}\varrho(0)y^2 - \phi(0)y$$
$$= \frac{1}{2}\mathbb{E}\Big[\int_0^T \Big(\tau(t)'(\sigma(t)\sigma(t)')\tau(t)\varrho(t)Y^2(t) + 2\varrho(t)Y(t)\tau(t)'(\sigma(t)\sigma(t)')\varphi(t)$$
$$+ 2Y(t)\varrho(t)\big(\tau(t)'(\sigma(t)\sigma(t)')\zeta(t) + g(t) - g(t)\big) + 2Y(t)\vartheta(t)\varrho(t)\tau(t)'(\sigma(t)\sigma(t)')\tau(t)$$
$$+ 2\vartheta(t)\varrho(t)\varphi(t)'(\sigma(t)\sigma(t)')\tau(t) + 2\phi(t)r(t)\gamma - 2\phi(t)g(t) + \varphi(t)'(\sigma(t)\sigma(t)')\varphi(t)\varrho(t)$$
$$+ 2\varphi(t)'\sigma(t)\delta(t)\varrho(t) + \upsilon(t)\rho(t)'\rho(t)\upsilon(t)\varrho(t) + \upsilon(t)^2(1 - \rho(t)'\rho(t))\varrho(t)\Big)dt\Big].$$

Rearranging the terms, using that $\zeta(t) = (\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t)$ and adding and substracting the term $(\tau(t)\vartheta(t) + \zeta(t))'(\sigma(t)\sigma(t)')(\tau(t)\vartheta(t) + \zeta(t))\varrho(t)$ yields

$$J(\varphi,\gamma) - \frac{1}{2}\varrho(0)y^2 - \phi(0)y$$
$$= \frac{1}{2}\mathbb{E}\Big[\int_0^T \Big(\tau(t)'(\sigma(t)\sigma(t)')\tau(t)\varrho(t)Y^2(t) + 2\tau(t)'(\sigma(t)\sigma(t)')\varphi(t)\varrho(t)Y(t)$$
$$+ 2Y(t)\varrho(t)\tau(t)'(\sigma(t)\sigma(t)')\big(\tau(t)\vartheta(t) + \zeta(t)\big) + 2\varphi(t)'(\sigma(t)\sigma(t)')\big(\tau(t)\vartheta(t) + \zeta(t)\big)\varrho(t)$$
$$+ \varphi(t)'(\sigma(t)\sigma(t)')\varphi(t)\varrho(t) + (\tau(t)\vartheta(t) + \zeta(t))'(\sigma(t)\sigma(t)')(\tau(t)\vartheta(t) + \zeta(t))\varrho(t)$$
$$- (\tau(t)\vartheta(t) + \zeta(t))'(\sigma(t)\sigma(t)')(\tau(t)\vartheta(t) + \zeta(t))\varrho(t) + 2\phi(t)(r(t)\gamma - g(t)) + \upsilon(t)^2\varrho(t)\Big)dt\Big].$$

Realizing that the first part of the integrand can be written as a quadratic form gives

$$J(\varphi,\chi) - \frac{1}{2}\varrho(0)y^2 - \phi(0)y = \frac{1}{2}\mathbb{E}\Big[\int_0^T (\tau(t)Y(t) + \varphi(t) + \tau(t)\vartheta(t) + \zeta(t))'(\sigma(t)\sigma(t)')\varrho(t)$$
$$\times (\tau(t)Y(t) + \varphi(t) + \tau(t)\vartheta(t) + \zeta(t))dt\Big]$$
$$+ \frac{1}{2}\int_0^T \big[- (\tau(t)\vartheta(t) + \zeta(t))'(\sigma(t)\sigma(t)')(\tau(t)\vartheta(t) + \zeta(t))\varrho(t)$$
$$+ 2(r(t)\gamma - g(t))\phi(t) + \upsilon^2(t)\varrho(t)\big]dt.$$

Note that the first summand (whose integrand is a quadratic form) is ≥ 0 due to the assumption $\varrho(t)(\sigma(t)\sigma(t)') > 0$ in (7.16) (which is automatically fulfilled due to $\sigma(t)\sigma(t)' \geq \epsilon\mathbb{I}_m$). Therefore

$$J(\varphi,\chi) - \frac{1}{2}\varrho(0)y^2 - \phi(0)y \geq \frac{1}{2}\int_0^T \big[-(\tau(t)\vartheta(t)+\zeta(t))'(\sigma(t)\sigma(t)')(\tau(t)\vartheta(t)+\zeta(t))\varrho(t) + 2(r(t)\gamma - g(t))\phi(t) + \upsilon^2(t)\varrho(t)\big]dt,$$

with equality if and only if the first summand above equals 0, i.e.

$$\begin{aligned} &\tau(t)Y(t) + \varphi(t) + \tau(t)\vartheta(t) + \zeta(t) \equiv 0 \\ \iff &\varphi(t) = -\tau(t)(Y(t)+\vartheta(t)) - \zeta(t). \end{aligned}$$

This implies that the feedback control given in (7.22) is indeed an optimal control and adding $\frac{1}{2}\varrho(0)y^2 + \phi(0)y$ above gives the corresponding optimal cost functional given in (7.23). □

Solving the resulting ODE for $\vartheta(t)$ using (7.16) and (7.17) (see Appendix F) in this case yields an explicit expression for $\vartheta(t)$ as

$$\vartheta(t) = \int_t^T \kappa(s) e^{\int_s^t r(z)dz} ds + \gamma\left(1 - e^{-\int_t^T r(z)dz}\right). \tag{7.25}$$

As the auxiliary problems $A(\gamma)$ and $A(\chi,\omega)$ are equivalent, we directly obtain the optimal feedback control of the auxiliary problem $A(\chi,\omega)$ as

$$\varphi^*(X(t)) = -\tau(t)(X(t)+\vartheta(t)-\gamma) - \zeta(t), \qquad t \in [0,T], \tag{7.26}$$

where $\tau(t), \zeta(t)$ and $\vartheta(t)$ are given in (7.18),(7.19) and (7.25) respectively.
It now remains to deduce the optimal strategy and the efficient frontier of the original problem $P(\chi)$. Substituting (7.26) into (7.10) (where $X^*(t)$ denotes the wealth process under the optimal strategy) gives

$$\begin{aligned} dX^*(t) &= \big[r(t)X^*(t) + (\mu(t) - r(t)\mathbb{1})'(-\tau(t)(X^*(t)+\vartheta(t)-\gamma) - \zeta(t)) - g(t)\big]dt \\ &\quad + \big[(-\tau(t)(X^*(t)+\vartheta(t)-\gamma) - \zeta(t))'\sigma(t) + \delta(t)'\big]dW(t) + \delta_0(t)dW_0(t) \\ &= \big[(r(t) - (\mu(t)-r(t)\mathbb{1})'\tau(t))X^*(t) - (\mu(t)-r(t)\mathbb{1})'\tau(t)\vartheta(t) + (\mu(t)-r(t)\mathbb{1})'\tau(t)\gamma \\ &\quad - (\mu(t)-r(t)\mathbb{1})'\zeta(t) - g(t)\big]dt \\ &\quad + \big[-\tau(t)'\sigma(t)(X^*(t)+\vartheta(t)-\gamma) - \zeta(t)'\sigma(t) + \delta(t)'\big]dW(t) + \delta_0(t)dW_0(t) \\ &= \big[(r(t)-\varsigma(t))X^*(t) - \varsigma(t)\vartheta(t) + \varsigma(t)\gamma + \kappa(t)\big]dt \\ &\quad + \big[-\tau(t)'\sigma(t)(X^*(t)+\vartheta(t)-\gamma) - \zeta(t)'\sigma(t) + \delta(t)'\big]dW(t) + \delta_0(t)dW_0(t), \end{aligned}$$

$X^*(0) = x.$

Taking expectation on both sides yields

$$\begin{aligned} d\mathbb{E}[X^*(t)] &= \big[(r(t) - \varsigma(t))\mathbb{E}[X^*(t)] - \varsigma(t)\vartheta(t) + \varsigma(t)\gamma + \kappa(t)\big]dt, \\ \mathbb{E}[X^*(0)] &= x. \end{aligned}$$

We can solve the above ODE by applying the standard technique of multiplying both sides with $e^{\int_0^s (\varsigma(u) - r(u))du}$, integrating both sides from 0 to t and solving for $\mathbb{E}[X^*(t)]$ (detailed in Appendix F). This gives

$$\mathbb{E}[X^*(t)] = xe^{\int_0^t (r(u) - \varsigma(u))du} + \int_0^t (\varsigma(s)\gamma + \kappa(s) - \varsigma(s)\vartheta(s))e^{\int_s^t (r(u) - \varsigma(u))du} ds \tag{7.27}$$

for all $t \in [0, T]$. Inserting $t = T$ and (7.25) yields

$$\begin{aligned} \mathbb{E}[X^*(T)] = xe^{\int_0^T (r(u) - \varsigma(u))du} &+ \int_0^T \varsigma(s)\gamma e^{\int_s^T (r(u) - \varsigma(u))du} ds \\ &\underbrace{+ \int_0^T \kappa(s)e^{\int_s^T (r(u) - \varsigma(u))du} ds - \int_0^T \varsigma(s)e^{\int_s^T (r(u) - \varsigma(u))du} \int_s^T \kappa(z)e^{\int_z^s r(u)du} dzds}_{I_2} \\ &- \int_0^T \varsigma(s)\gamma e^{\int_s^T (r(u) - \varsigma(u))du} ds + \gamma \underbrace{\int_0^T \varsigma(s)e^{-\int_s^T r(u)du} e^{\int_s^T (r(u) - \varsigma(u))du} ds}_{I_1}. \end{aligned}$$

We calculate I_1 and I_2 as follows:

$$I_1 = \int_0^T \varsigma(s)e^{\int_T^s \varsigma(u)du} ds = \left[e^{\int_T^s \varsigma(u)du}\right]_0^T = \left[1 - e^{-\int_0^T \varsigma(u)du}\right],$$

and

$$\begin{aligned} I_2 &= \int_0^T \kappa(s)e^{\int_s^T (r(u) - \varsigma(u))du} ds - \int_0^T \int_s^T \varsigma(s)\kappa(z)e^{\int_s^T (r(u) - \varsigma(u))du} e^{\int_z^s r(u)du} dzds \\ &= \int_0^T \kappa(s)e^{\int_s^T (r(u) - \varsigma(u))du} ds - \int_0^T \int_0^z \varsigma(s)\kappa(z)e^{\int_s^T (r(u) - \varsigma(u))du} e^{\int_z^s r(u)du} dsdz \\ &= \int_0^T \kappa(s)e^{\int_s^T (r(u) - \varsigma(u))du} ds - \int_0^T \int_0^z \varsigma(s)\kappa(z)e^{\int_z^T r(u)du} e^{-\int_s^T \varsigma(u)du} dsdz \\ &= \int_0^T \kappa(s)e^{\int_s^T (r(u) - \varsigma(u))du} ds - \int_0^T \kappa(z)e^{\int_z^T r(u)du} \int_0^z \varsigma(s)e^{\int_T^s \varsigma(u)du} dsdz \\ &= \int_0^T \kappa(s)e^{\int_s^T (r(u) - \varsigma(u))du} ds - \int_0^T \kappa(z)e^{\int_z^T r(u)du} \left[e^{\int_T^s \varsigma(u)du}\right]_0^z dz \end{aligned}$$

$$
\begin{aligned}
&= \int_0^T \kappa(s) e^{\int_s^T (r(u)-\varsigma(u))du} ds - \int_0^T \kappa(z) e^{\int_z^T r(u)du} e^{-\int_z^T \varsigma(u)du} dz + \int_0^T \kappa(z) e^{\int_z^T r(u)du} e^{-\int_0^T \varsigma(u)du} dz \\
&= \int_0^T \kappa(z) e^{\int_z^T r(u)du} e^{-\int_0^T \varsigma(u)du} dz.
\end{aligned}
$$

In summary, we can conclude that one can write $\mathbb{E}[X^*(T)]$ as

$$\mathbb{E}[X^*(T)] = a + b\gamma, \tag{7.28}$$

with

$$
\begin{aligned}
a &:= \int_0^T \kappa(t) e^{\int_t^T r(u)du - \int_0^T \varsigma(u)du} dt + x e^{\int_0^T (r(u)-\varsigma(u))du}, \\
b &:= 1 - e^{-\int_0^T \varsigma(u)du}.
\end{aligned} \tag{7.29}
$$

Analogously, we proceed with $\mathbb{E}[X^*(t)^2]$, where we first have to apply Itô's formula with $f(t, x(t)) = x^2$ to obtain

$$
\begin{aligned}
dX^*(t)^2 &= 2X^*(t)\big[(r(t)-\varsigma(t))X^*(t) - \varsigma(t)\vartheta(t) + \varsigma(t)\gamma + \kappa(t)\big]dt \\
&\quad + \frac{1}{2}\cdot 2 \cdot \big[\big(-\tau(t)'\sigma(t)(X^*(t)+\vartheta(t)-\gamma) - \zeta(t)'\sigma(t) + \delta(t)'\big)^2 + \delta_0(t)^2\big]dt \\
&\quad + 2X^*(t)\big[(-\tau(t)'\sigma(t)(X^*(t)+\vartheta(t)-\gamma) - \zeta(t)'\sigma(t) + \delta(t)'\big]dW(t) + 2X^*(t)\delta_0(t)dW_0(t) \\
&= \big[2X^*(t)^2(r(t)-\varsigma(t)) - 2X^*(t)\varsigma(t)\vartheta(t) + 2X^*(t)\varsigma(t)\gamma + 2X^*(t)\kappa(t)\big]dt \\
&\quad + \big[\tau(t)'\sigma(t)\sigma(t)'\tau(t)\big(X^*(t)^2 + 2X^*(t)(\vartheta(t)-\gamma) + (\vartheta(t)-\gamma)^2\big) \\
&\quad + \zeta(t)'\sigma(t)\sigma(t)'\zeta(t) + \delta(t)'\delta(t) + \delta_0(t)^2\big]dt \\
&\quad + 2\big[(\tau(t)'\sigma(t)\sigma(t)'\zeta(t) - \tau(t)'\sigma(t)\delta(t))\big(X^*(t)+\vartheta(t)-\gamma\big) - \zeta(t)'\sigma(t)\delta(t)\big]dt \\
&\quad + 2X^*(t)\big[(-\tau(t)'\sigma(t)(X^*(t)+\vartheta(t)-\gamma) - \zeta(t)'\sigma(t) + \delta(t)'\big]dW(t) + 2X^*(t)\delta_0(t)dW_0(t) \\
&= \big[2X^*(t)^2(r(t)-\varsigma(t)) - 2X^*(t)\varsigma(t)\vartheta(t) + 2X^*(t)\varsigma(t)\gamma + 2X^*(t)\kappa(t)\big]dt \\
&\quad + \big[\varsigma(t)X^*(t)^2 + 2X^*(t)\varsigma(t)(\vartheta(t)-\gamma) + \varsigma(t)(\vartheta(t)-\gamma)^2 \\
&\quad + \delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t) + \delta(t)'\delta(t) + \delta_0(t)^2\big]dt \\
&\quad + \big[2(X^*(t)+\vartheta(t)-\gamma)[\tau(t)'\sigma(t)\sigma(t)'\zeta(t) - \tau(t)'\sigma(t)\delta(t)] - 2\delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t)\big]dt \\
&\quad + 2X^*(t)\big[(-\tau(t)'\sigma(t)(X^*(t)+\vartheta(t)-\gamma) - \zeta(t)'\sigma(t) + \delta(t)'\big]dW(t) + 2X^*(t)\delta_0(t)dW_0(t) \\
&= \big[X^*(t)^2(2r(t)-\varsigma(t)) + 2X^*(t)\kappa(t) + \varsigma(t)(\vartheta(t)-\gamma)^2 + \delta(t)'\delta(t) \\
&\quad - \delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t) + \delta_0(t)^2\big]dt \\
&\quad + 2X^*(t)\big[(-\tau(t)'\sigma(t)(X^*(t)+\vartheta(t)-\gamma) - \zeta(t)'\sigma(t) + \delta(t)'\big]dW(t) + 2X^*(t)\delta_0(t)dW_0(t)
\end{aligned}
$$

where we have used that $\varsigma(t) = \tau(t)'\sigma(t)\sigma(t)'\tau(t)$, $\zeta(t) = (\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t)$ and $\tau(t)'\sigma(t)\sigma(t)'\zeta(t) - \tau(t)'\sigma(t)\delta(t) = \tau(t)'\sigma(t)\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t) - \tau(t)'\sigma(t)\delta(t) = 0$.

Thus, in summary

$$\begin{aligned}dX^*(t)^2 = &\big[(2r(t)-\varsigma(t))X^*(t)^2 + 2\kappa(t)X^*(t) + \varsigma(t)(\vartheta(t)-\gamma)^2 + \delta(t)'\delta(t)\\ &- \delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t) + \delta_0(t)^2\big]dt\\ &+ 2X^*(t)[-\tau(t)'\sigma(t)(X^*(t)+\vartheta(t)-\gamma) - \zeta(t)'\sigma(t) + \delta(t)']dW(t) + 2X^*(t)\delta_0(t)dW_0(t),\\ X^*(0)^2 = &\, x^2.\end{aligned}$$

Taking expectation on both sides gives

$$\begin{aligned}d\mathbb{E}[X^*(t)^2] = &\big[(2r(t)-\varsigma(t))\mathbb{E}[X^*(t)^2] + 2\kappa(t)\mathbb{E}[X^*(t)] + \varsigma(t)(\vartheta(t)-\gamma)^2 + \delta(t)'\delta(t)\\ &- \delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t) + \delta_0(t)^2\big]dt,\\ \mathbb{E}[X^*(0)^2] = &\, x^2.\end{aligned}$$

Solving the above ODE by applying standard methods (see Appendix F) yields

$$\begin{aligned}\mathbb{E}[X^*(t)^2] = x^2 e^{\int_0^t (2r(u)-\varsigma(u))du} &+ \int_0^t \big(2\kappa(s)\mathbb{E}[X^*(s)] + \varsigma(s)(\vartheta(s)-\gamma)^2 + \delta(s)'\delta(s)\\ &- \delta(s)'\sigma(s)'(\sigma(s)\sigma(s)')^{-1}\sigma(s)\delta(s) + \delta_0(s)^2\big)e^{\int_s^t (2r(u)-\varsigma(u))du}ds.\end{aligned} \tag{7.30}$$

For $t = T$, according to Xie, Li, and Wang 2008, analogously to above $\mathbb{E}[X^*(T)^2]$ can be written as

$$\mathbb{E}[X^*(T)^2] = \eta + b\gamma^2$$

where b is given in (7.29) and η is given by

$$\begin{aligned}\eta := &\, 2xe^{\int_0^T (r(u)-\varsigma(u))du}\int_0^T \kappa(t)e^{\int_t^T r(u)du}dt + e^{\int_0^T \varsigma(u)du}\left(\int_0^T \kappa(t)e^{\int_t^T r(u)du - \int_0^T \varsigma(u)du}dt\right)^2\\ &+ x^2 e^{\int_0^T (2r(u)-\varsigma(u))du} + \int_0^T \big[\delta(t)'\delta(t) - \delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t) + \delta_0(t)^2\big]e^{\int_t^T (2r(u)-\varsigma(u))du}dt.\end{aligned} \tag{7.31}$$

We know by Theorem 7.4 that any optimal solution of the original problem $P(\chi)$ can be obtained via the solution of the auxiliary problem $A(\chi, \omega^*)$ where $\omega^* = 1 + 2\chi\mathbb{E}[X^*(T)]$, and we know by (7.28) that $\mathbb{E}[X^*(T)] = a + b\gamma^*$ with $\gamma^* = \frac{\omega^*}{2\chi}$ by definition. Therefore

$$\begin{aligned}&\omega^* = 1 + 2\chi(a + b\gamma^*) = 1 + 2\chi\left(a + b\frac{\omega^*}{2\chi}\right) = 1 + 2\chi a + b\omega^*\\ &\iff \omega^* = \frac{1+2\chi a}{1-b}\end{aligned}$$

and

$$\gamma^* = \frac{\omega *}{2\chi} = \frac{1+2\chi a}{2\chi(1-b)} = \frac{1}{2\chi(1-b)} + \frac{a}{1-b}. \tag{7.32}$$

Thus, the optimal strategy (i.e. the optimal feedback control of $P(\chi)$) is given by (7.26) with $\gamma = \gamma^*$ given in (7.32).
The variance of the terminal wealth is given by

$$\begin{aligned}\mathbb{V}ar(X^*(T)) &= \mathbb{E}[X^*(T)^2] - (\mathbb{E}[X^*(T)])^2 \\ &= \eta + b(\gamma^*)^2 - (a+b\gamma^*)^2 = b(1-b)(\gamma^*)^2 - 2ab\gamma^* + \eta - a^2. \end{aligned} \tag{7.33}$$

This leads to the following result.

Theorem 7.6 (Mean-variance efficient frontier (Xie, Li, and Wang 2008))**.** The efficient frontier of the mean-variance portfolio selection problem $P(\chi)$, if it ever exists, is given by

$$\begin{aligned}\mathbb{V}ar[X^*(T)] &= \frac{e^{-\int_0^T \varsigma(z)dz}}{1-e^{-\int_0^T \varsigma(z)dz}} \left[\mathbb{E}[X^*(T)] - \left(xe^{\int_0^T r(z)dz} + \int_0^T \kappa(t) e^{\int_t^T r(z)dz} dt\right)\right]^2 \\ &+ \int_0^T \left[(\delta_0(t))^2 + \delta(t)'\delta(t) - \delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t)\right] e^{\int_t^T (2r(z)-\varsigma(z))dz} dt \\ &=: \frac{e^{-\int_0^T \varsigma(z)dz}}{1-e^{-\int_0^T \varsigma(z)dz}} [\mathbb{E}[X^*(T)] - \mathcal{D}_1]^2 + \mathcal{D}_2 \end{aligned} \tag{7.34}$$

for $\mathbb{E}[X^*(T)] \geq \mathcal{D}_1$, where $\varsigma(t)$ and $\kappa(t)$ are defined in (7.20) and (7.21) respectively, $\delta(t) := -\upsilon(t)\rho(t) \in \mathcal{C}([0,T];\mathbb{R}^{n\times 1})$ and $\delta_0(t) := -\upsilon(t)\sqrt{1-\rho(t)'\rho(t)}$.

Proof. The formula for $\mathbb{V}ar[X^*(T)]$ follows from the above calculations by using $\gamma^* = \frac{\mathbb{E}[X^*(T)]-a}{b}$ and inserting the explicit expression for a, b and η from (7.29) and (7.31). First, rewrite (7.33) as

$$\begin{aligned}\mathbb{V}ar[X^*(T)] &= b(1-b)(\gamma^*)^2 - 2ab\gamma^* + \eta - a^2 \\ &= \frac{1-b}{b}\left(b^2(\gamma^*)^2 - \frac{2ab^2\gamma^*}{1-b} + \frac{b(\eta-a^2)}{1-b}\right) \\ &= \frac{1-b}{b}\left((b\gamma^*+a)^2 - \frac{2ab\gamma^*}{1-b} + \frac{b\eta-a^2}{1-b}\right). \end{aligned}$$

Then substitute $b\gamma^* = \mathbb{E}[X^*(T)] - a$ from (7.28) and use square addition to obtain

$$\mathbb{V}ar[X^*(T)] = \frac{1-b}{b}\left(\mathbb{E}[X^*(T)]^2 - \frac{2a}{1-b}(\mathbb{E}[X^*(T)] - a) + \frac{b\eta-a^2}{1-b}\right)$$

$$= \frac{1-b}{b}\left(\left(\mathbb{E}[X^*(T)] - \frac{a}{1-b}\right)^2 - \frac{a^2}{(1-b)^2} + \frac{2a^2}{1-b} + \frac{b\eta - a^2}{1-b}\right)$$
$$= \frac{1-b}{b}\left(\mathbb{E}[X^*(T)] - \underbrace{\frac{a}{1-b}}_{=:y_1}\right)^2 \underbrace{- \frac{a^2}{b(1-b)} + \frac{2a^2}{b} + \frac{b\eta - a^2}{b}}_{=:y_2}.$$

Clearly, it holds that

$$\frac{1-b}{b} = \frac{e^{-\int_0^T \varsigma(z)dz}}{1 - e^{-\int_0^T \varsigma(z)dz}},$$

therefore it remains to be shown that y_1 and y_2 coincide with $\mathcal{D}_1$ and $\mathcal{D}_2$ as they are stated in the theorem. First consider y_1, where by definition of a and b in (7.29) it holds

$$y_1 = \frac{a}{1-b} = \frac{\int_0^T \kappa(t) e^{\int_t^T r(u)du - \int_0^T \varsigma(u)du} dt + x e^{\int_0^T (r(u)-\varsigma(u))du}}{e^{-\int_0^T \varsigma(u)du}}$$
$$= x e^{\int_0^T r(u)du} + \int_0^T \kappa(t) e^{\int_t^T r(u)du} dt.$$

Furthermore, for y_2 it holds

$$y_2 = -\frac{a^2}{b(1-b)} + \frac{2a^2}{b} + \frac{b\eta - a^2}{b} = -\frac{a^2}{b(1-b)} + \frac{b\eta + a^2}{b}. \tag{7.35}$$

We first compute a^2 and $b\eta$:

$$a^2 = \left(\int_0^T \kappa(t) e^{\int_t^T r(u)du - \int_0^T \varsigma(u)du} dt\right)^2 + 2x \int_0^T \kappa(t) e^{\int_t^T r(u)du - \int_0^T (2\varsigma(u) - r(u))du} dt$$
$$+ x^2 e^{\int_0^T 2(r(u)-\varsigma(u))du}$$
$$b\eta = 2x e^{\int_0^T (r(u)-\varsigma(u))du} \int_0^T \kappa(t) e^{\int_t^T r(u)du} dt + e^{\int_0^T \varsigma(u)du} \left(\int_0^T \kappa(t) e^{\int_t^T r(u)du - \int_0^T \varsigma(u)du} dt\right)^2$$
$$+ x^2 e^{\int_0^T (2r(u)-\varsigma(u))du} + \mathcal{D}_2$$
$$- 2x e^{\int_0^T (r(u)-2\varsigma(u))du} \int_0^T \kappa(t) e^{\int_t^T r(u)du} dt - \left(\int_0^T \kappa(t) e^{\int_t^T r(u)du - \int_0^T \varsigma(u)du} dt\right)^2$$
$$- x^2 e^{\int_0^T 2(r(u)-\varsigma(u))du} - (1-b)\mathcal{D}_2.$$

Therefore, it follows that

$$\begin{aligned} a^2 + b\eta &= b\mathcal{D}_2 + 2xe^{\int_0^T (r(u)-\varsigma(u))du} \int_0^T \kappa(t) e^{\int_t^T r(u)du} dt \\ &\quad + e^{\int_0^T \varsigma(u)du} \left(\int_0^T \kappa(t) e^{\int_t^T r(u)du - \int_0^T \varsigma(u)du} dt \right)^2 + x^2 e^{\int_0^T (2r(u)-\varsigma(u))du} \\ &= b\mathcal{D}_2 + \frac{a^2}{e^{-\int_0^T \varsigma(u)du}} = b\mathcal{D}_2 + \frac{a^2}{1-b}. \end{aligned}$$

Inserting this result into (7.35) yields

$$\begin{aligned} y_2 &= -\frac{a^2}{b(1-b)} + \frac{b\eta + a^2}{b} = -\frac{a^2}{b(1-b)} + \frac{b\mathcal{D}_2 + \frac{a^2}{1-b}}{b} \\ &= -\frac{a^2}{b(1-b)} + \frac{a^2}{b(1-b)} + \mathcal{D}_2 = \mathcal{D}_2. \end{aligned}$$

This concludes the proof.
We note that

$$\frac{e^{-\int_0^T \varsigma(z)dz}}{1 - e^{-\int_0^T \varsigma(z)dz}} > 0, \quad e^{\int_0^T r(z)dz} > 1, \quad e^{\int_t^T (2r(z)-\varsigma(z))dz} > 0,$$

and furthermore it can be shown that $\mathcal{D}_2 \geq 0$ (see Xie, Li, and Wang 2008). □

Note that (7.34) reveals the trade-off between risk (measured in terms of variance) and return (mean) of the terminal wealth: For a given expected return level, the risk that must be accepted is given by (7.34), and vice versa. The *minimum-variance-portfolio* is reached for $\mathbb{E}[X^*(T)] = \mathcal{D}_1$ and entails a variance of $\mathcal{D}_2 \geq 0$. In particular, the only case where in the presence of liability it holds $\mathcal{D}_2 = 0$, 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 $\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t) = \mathbb{I}_m$ and thus

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

This is the case when the financial market is complete and the risk from the liability can be completely hedged by trading the assets.

Remark. Note that from the above result in (7.34), we can deduce the formula (e.g. known from the lecture notes of the TUM lecture *Investment Strategies*, p. 42) for the dynamic mean-variance efficient frontier in the "classical" case where there is no liability

(i.e. $g(t) \equiv 0, \upsilon(t) \equiv 0$), the market is complete (i.e. $m = n$), the initial wealth is $x = 1$ and all coefficients are constant (i.e. $r(t) \equiv r, \mu(t) \equiv \mu, \sigma(t) \equiv \sigma$). This implies above

$$\begin{aligned}\tau(t) \equiv \tau &= (\sigma\sigma')^{-1}(\mu - r\mathbb{1}), \qquad \varsigma(t) \equiv \varsigma = (\mu - r\mathbb{1})\tau \\ \zeta(t) \equiv 0, \qquad \kappa(t) &\equiv 0, \qquad \vartheta(t) = \gamma\big(1 - e^{-r(T-t)}\big).\end{aligned}$$

Denote for now $\mathbb{E}[X^*(T)] =: \bar{m}, \mathbb{V}ar[X^*(T)] =: \bar{s}^2$, then it follows from (7.34)

$$\bar{s}^2(\bar{m}) = \frac{e^{-\varsigma T}}{1 - e^{-\varsigma T}}\big[\bar{m} - e^{rT}\big]^2 = \frac{1}{e^{\varsigma T} - 1}\big[\bar{m} - e^{rT}\big]^2.$$

This is equivalent to

$$\begin{aligned}\bar{s}(\bar{m}) &= \frac{1}{\sqrt{e^{||\bar{\gamma}||^2 T} - 1}} \cdot (\bar{m} - e^{rT}) \\ \iff \bar{m}(\bar{s}) &= e^{rT} + \sqrt{e^{||\bar{\gamma}||^2 T} - 1} \cdot \bar{s},\end{aligned}$$

where $\bar{\gamma} = \sigma^{-1}(\mu - r\mathbb{1})$. Therefore, we can see that the results of Xie, Li, and Wang 2008 describe a more general case in accordance with known results.

In their paper, Xie, Li, and Wang 2008 give numerical examples to show how the incompleteness of the market and the introduction of a liability influence the efficient frontier. They further show the influence of varying the parameters $T, x, g(t), \upsilon(t), \rho(t)$ on the efficient frontier and the optimal strategy. For our purposes, particularly varying the diffusion $\upsilon(t)$ of the liability is relevant. Thus, in the next section we apply the results from this section in order to study how the optimal investment decision of an insurance company is altered when its liability follows a risk process with Hawkes arrivals instead of Poisson arrivals.

7.3 Application to Hawkes risk model

For the assets, we use the parameters from the numerical example in Xie, Li, and Wang 2008 which are summarized in Table 7.1. This means we consider an incomplete market with one riskfree asset and one risky asset whose evolution is governed by a two-dimensional Brownian motion. However, this type of market incompleteness is not an essential factor in our example and could likewise be simplified to a one-dimensional Brownian motion.

To describe the company's *liability*, we use the pure diffusion approximation of the (negative) risk process, i.e. following (5.21) we set

$$\begin{aligned}L_p(t) := -R_p(t) &= -u + \left(\frac{\lambda}{1 - \alpha/\beta}a^* - c\right)t + \bar{\sigma}W(t) \\ &= -u - \theta a^* \frac{\lambda}{1 - \alpha/\beta}t + \bar{\sigma}W(t)\end{aligned} \tag{7.36}$$

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

Table 7.1: Asset parameters, taken from Xie, Li, and Wang 2008. The initial wealth x was adapted to fit this example.

where λ, α and β are the parameters of the Hawkes arrival process, θ is the safety loading of the premium, a^* and $\bar{\sigma}$ are given in Theorem 5.8 and 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 analyze the difference in risk arising from the variance of the number of arrivals. Table 7.2 gives an overview of these quantities for Poisson and Hawkes processes.

Process	Parameters	$\mathbb{E}[N(0,t)]$	$\mathbb{V}ar[N(0,t)]$
Poisson	λ_P	$\lambda_P t$	$\lambda_P t$
Hawkes	λ, α, β	$\frac{\lambda}{1-\alpha/\beta} t$	$\frac{\lambda}{1-\alpha/\beta}(t(\frac{1}{1-\alpha/\beta})^2 + (1 - (\frac{1}{1-\alpha/\beta})^2)\frac{1-e^{-t(\beta-\alpha)}}{\beta-\alpha})$

Table 7.2: 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 derived in Section 3.5.

Remark. Note in Table 7.2 that for a Poisson process $\mathbb{E}[N(0,t)] = \mathbb{V}ar[N(0,t)]$ whereas for a Hawkes process $\mathbb{E}[N(0,t)] < \mathbb{V}ar[N(0,t)]$ for any $t > 0$. We can show this by realizing that in the Hawkes case for

$$\mathbb{V}ar[N(0,t)] = \frac{\lambda}{1-\alpha/\beta} \underbrace{\left(t\left(\frac{1}{1-\alpha/\beta}\right)^2 + \underbrace{\left(1 - \left(\frac{1}{1-\alpha/\beta}\right)^2\right)}_{:=v_1} \frac{1-e^{-t(\beta-\alpha)}}{\beta-\alpha} \right)}_{=:v_2}$$

it holds that $v_1 < 0$ and thus $v_2 \geq t$. This is sufficient as $\frac{\lambda}{1-\alpha/\beta} > 0$.
As $0 < \frac{\alpha}{\beta} < 1$, also $0 < 1 - \frac{\alpha}{\beta} < 1$, and thus $\frac{1}{1-\frac{\alpha}{\beta}} > 1$ and $v_1 < 0$.
We know from the Taylor series of e^x that $\forall x \in \mathbb{R}$ it holds

$$e^x \geq 1 + x \iff 1 - e^x \leq -x.$$

Thus it holds for $x = -t(\beta - \alpha)$ that

$$\frac{1-e^{-t(\beta-\alpha)}}{\beta-\alpha} \leq \frac{t(\beta-\alpha)}{\beta-\alpha} = t. \tag{7.37}$$

As $v_1 < 0$, it holds

$$v_2 = t\left(\frac{1}{1-\alpha/\beta}\right)^2 + \underbrace{\left(1-\left(\frac{1}{1-\alpha/\beta}\right)^2\right)}_{<0}\underbrace{\frac{1-e^{-t(\beta-\alpha)}}{\beta-\alpha}}_{\leq t \text{ by (7.37)}}$$
$$\geq t\left(\frac{1}{1-\alpha/\beta}\right)^2 + \left(1-\left(\frac{1}{1-\alpha/\beta}\right)^2\right)t = t.$$

and this shows the claim.

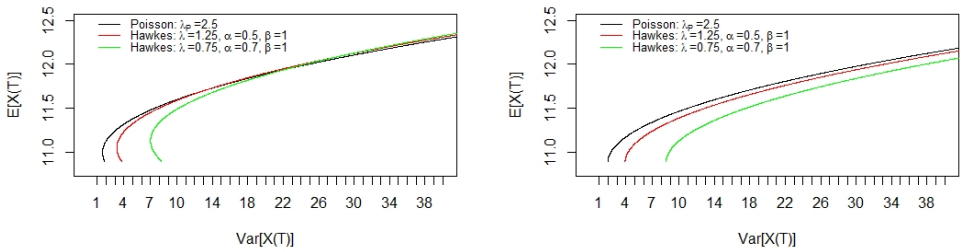
Table 7.3 gives the parameter values chosen for our numerical example where we compare a Poisson and two Hawkes processes.

Process	(λ, α, β)	$\mathbb{E}[N(0,1)]$	$\mathbb{V}ar[N(0,1)]$	$g(t) \equiv -\theta m_1$ $\mathbb{E}[N(0,1)]$	$\upsilon(t) \equiv \bar{\sigma}$	$\mathbb{V}ar(L_p(T))$
Poisson	$(2.5, -, -)$	2.5	2.5	-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 7.3: Liability parameters, where the safety loading for the premium calculation is $\theta = 0.2$. Parameters are chosen such that $\mathbb{E}[N(0,1)]$ 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 endogeneous events α/β, which corresponds to stronger influence of self-excitement and clustering.

In Figure 26 we plot the mean-variance efficient frontiers for all three cases using (7.34). 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 7.4 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 Figure 26). As effectively we increase the diffusion of the liability while leaving other parameters untouched, these results are consistent with Xie, Li, and Wang 2008, Figure 1e.

We will now consider optimal strategies in different contexts. In this section, we have used $\varphi(X(t))$ to emphasize the connection between the *control* φ and the corresponding *state variable* $X(t)$. In the following, we change the notation to $\varphi(t)$ for convenience, as we like to think of the strategy as dynamically evolving over *time*. However, keep in mind that φ still also depends on $X(t)$.



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

Figure 26: Comparison of mean-variance efficient frontiers for parameters given in Tables 7.1 and 7.3. This is consistent with the observation in Figure 1e of Xie, Li, and Wang 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.

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

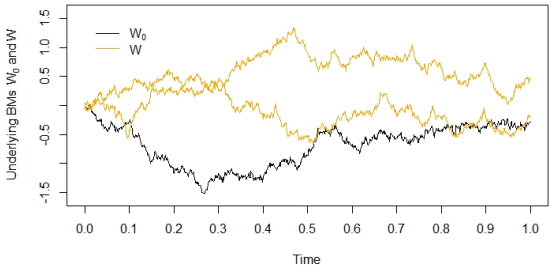
Table 7.4: Values of $\mathcal{D}_1$ and $\mathcal{D}_2$ for the three processes and two different choices of ρ. Note that $\mathcal{D}_1$ corresponds to the expected terminal wealth of the minimum-variance portfolio and $\mathcal{D}_2$ to its variance. As shown in Xie, Li, and Wang 2008, $\mathcal{D}_2 > 0$ 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.

7.3.1 Optimal Strategy for given expected return level

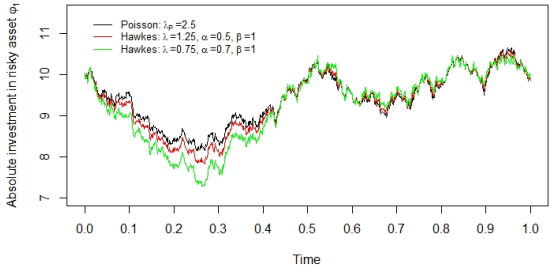
As in Xie, Li, and Wang 2008, first assume the company would like to obtain a certain expected return level, say 15%, i.e. $\mathbb{E}[X(T)^*] = 11.5$. Table 7.5 shows the minimum attainable variance of terminal wealth and the optimal strategy at the initial time 0 for all three cases. As could be expected from Figure 26, 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 realisation of the optimal investment in the risky asset $\varphi_1^*(t)$ over time in Figure 27. 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 (7.10) and (7.26) respectively.

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	$(-0.0029, 10.0029)$

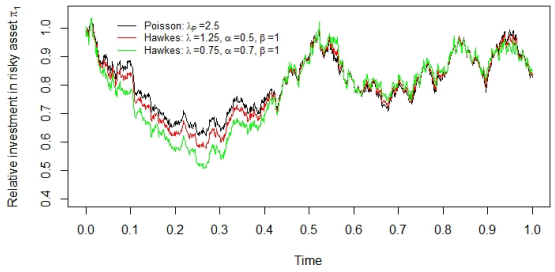
Table 7.5: Optimal strategies $\varphi^*(t)$ at $t = 0, X(t) = X_0 = x$ from (7.26) for a given expected return level of 15%, thus $\mathbb{E}[X^*(T)] = 11.5$. Note that in the Hawkes cases, more is initially invested in the risky asset in the case $\rho = (0.65, 0.1)$. For the case $\rho = (0, 0)$, the strategy does not depend on the process (as can be seen from (7.19) and (7.26) 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 Figure 26 that for attaining this level of $\mathbb{E}[X^*(T)]$, a higher risk $\mathbb{V}ar[X^*(T)]$ must be accepted.



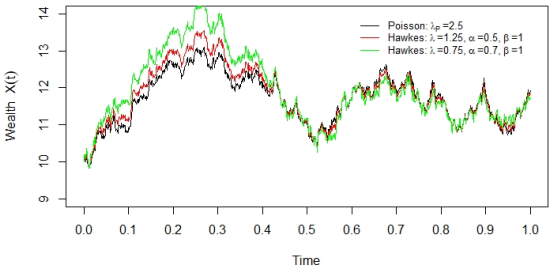
(a) Underlying standard Brownian Motions (W_0, W).



(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.

Figure 27: One realisation 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 (7.10) and (7.26).

7.3.2 Optimal Strategy for given risk level

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 in variance of terminal wealth) to $\bar{v} = 8$ and list in Table 7.6 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 7.3.

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	$(2.5680, 7.4320)$	–	–

Table 7.6: Optimal strategies $\varphi^*(t)$ at $t = 0, X(t) = X_0 = x$ from (7.26) 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.

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 6.3, 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 (5.17) it is given by

$$\begin{aligned} L_j(t) := -R_j(t) &= -u - ct + \hat{\sigma}W(t) + a^*N(t) \\ &= -u - (1+\theta)m_1\frac{\lambda}{1-\alpha/\beta}t + \hat{\sigma}W(t) + m_1N(t) \\ &=: -u + \hat{g}t + \hat{\sigma}W(t) + m_1N(t) \end{aligned} \tag{7.38}$$

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 $\hat{l} = -u$, $\hat{g}(t) \equiv -(1+\theta)m_1\frac{\lambda}{1-\alpha/\beta}$ and $\hat{v}(t) \equiv \hat{\sigma}$, where the values in this case are given in Table 7.7.

As naturally Theorem 7.5 is only applicable for a liability following a Brownian motion with drift (without jumps), we need an approximative approach in order to use $L_j(t)$ instead of $L_p(t)$ for the liability process. In order to find an *approximative optimal strategy*

Process	(λ, α, β)	$\mathbb{E}[N(0,1)]$	$\mathbb{V}ar[N(0,1)]$	$\hat{g}(t) \equiv -(1+\theta)m_1 \mathbb{E}[N(0,1)]$	$\hat{v}(t) \equiv \hat{\sigma}$	$\mathbb{V}ar[L_j(T)]$
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

Table 7.7: Liability parameters for the jump diffusion approximation, where the safety loading for the premium calculation is $\theta = 0.2$. Note that $\hat{v}$ 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 7.3).

for a given level of accepted risk $\bar{v}$, we initially omit the jump part of $L_j(t)$ and use Theorem 7.5 with the liability process given by

$$\hat{L_j}(t) = u + \hat{g}t + \hat{v}\rho' W(t) + \hat{v}\sqrt{1-\rho'\rho}W_0(t)$$

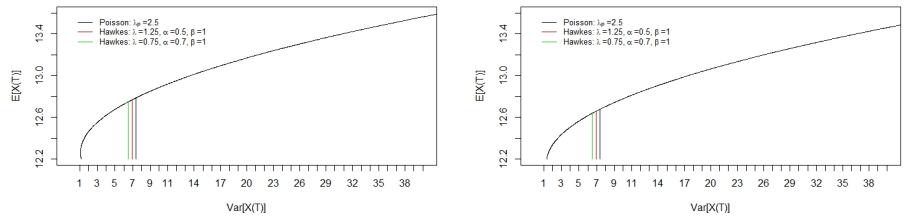
where $\hat{g}$ and $\hat{v}$ are given in Table 7.7. This implies that the jumps are considered to be a part of the liability which is independent from the assets and thus is not taken into consideration when calculating the optimal strategy. This is an assumption that could be interpreted in practice 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

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

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 depicted in Figure 28 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 7.6, the expectation of the *jump part*, that is $m_1 \mathbb{E}[N(1)] = 1.25$ has to be subtracted from the estimate given by the MVE frontier obtained using $\hat{L_j}(t)$. The results for parameters in Table 7.7 are given in Table 7.8.

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 $\hat{\bar{v}}$. We simulate the



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

Figure 28: Mean-variance efficient frontier for parameters given in Tables 7.1 and 7.7. 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.

continuous part of $L_j(t)$ on a grid with step size $\delta = 0.001$ over $[0, T] = [0, 1]$ according to

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

then calculate $\hat{X}^*(t)$ and $\hat{\varphi}_1^*(t)$ iteratively using (7.10) and (7.26) respectively and set $\hat{\varphi}_0^*(t) = \hat{X}^*(t) - \hat{\varphi}_1^*(t)$. Over all simulation runs, we calculate empirical mean and variance of $\hat{X}^*(T)$ and compare them with the theoretical values in Table 7.8. We then simulate $L_j(t)$ including jumps as

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

where for each run, the same realisations 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 $(0, t]$. For the simulation of the jumps it is worth noting that instead of simulating a point process on $[0, T]$ for each of the K simulations, we simulate one process on $[0, KT]$ and map the realisation of jumps on the interval $((k-1)T, kT]$ to the interval $(0, T]$ to be used in the k^{th} 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 (7.26), is kept fixed as calculated under $\hat{L}_j(t)$ (thus $\varphi_1^*(t) = \hat{\varphi}_1^*(t)$), but the wealth process $X^*(t)$ is calculated anew according to (7.10) with $L_j(t)$ in place of $\hat{L}_j(t)$ and the investment in the riskless asset is adjusted accordingly as $\varphi_0^*(t) = X^*(t) - \varphi_1^*(t)$. The results of the simulation are given in Table 7.9, 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.

Process	$m_1^2\mathbb{V}ar[N(T)]$	$\hat{\bar{v}}$	$\mathbb{E}[\hat{X}^*(T)](-1.25)$	$\hat{\varphi}^*(0)$
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	$m_1^2\mathbb{V}ar[N(T)]$	$\hat{\bar{v}}$	$\mathbb{E}[\hat{X}^*(T)](-1.25)$	$\hat{\varphi}^*(0)$
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) $\rho = (0, 0)$

Table 7.8: 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 $\bar{v} = 8$, 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 7.6, 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.

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. We have highlighted the economic implications of a claim process where claims tend to display clustering as opposed to claims occuring 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. This 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.

			Modified jump diffusion $\hat{L}_j(t)$		Jump diffusion $L_j(t)$	
Process	$\hat{\bar{v}}$	$\mathbb{E}[\hat{X}^*(T)]$ (theor.)	$\mathbb{E}[\hat{X}^*(T)]$	$\mathbb{V}ar[\hat{X}^*(T)]$	$\mathbb{E}[X^*(T)]$	$\mathbb{V}ar[X^*(T)]$
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) $\rho = (0.65, 0.1)$

			Modified jump diffusion $\hat{L}_j(t)$		Jump diffusion $L_j(t)$	
Process	$\hat{\bar{v}}$	$\mathbb{E}[\hat{X}^*(T)]$ (theor.)	$\mathbb{E}[\hat{X}^*(T)]$	$\mathbb{V}ar[\hat{X}^*(T)]$	$\mathbb{E}[X^*(T)]$	$\mathbb{V}ar[X^*(T)]$
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) $\rho = (0, 0)$

Table 7.9: For $K = 10000$ simulations, we give the shifted variance boundaries and the theoretically attainable expected terminal wealth $\hat{X}^*(T)$ under the modified jump diffusion liability $\hat{L}_j(t)$ for each case. Comparing these values with the empirical mean and variance boundary of $\hat{X}^*(T)$, we observe that they are matched quite closely. Keeping the optimal investment in the risky asset fixed and calculating the wealth process $X^*(T)$ under the jump diffusion liability $L_j(t)$, we observe that the original variance restriction $\bar{v} = 8$ is met. Note that in the case $\rho \neq (0, 0)$ some inaccurary is observable due to the theoretical assumption of independent jumps.

Chapter 8

Conclusion

In this thesis, we have introduced a risk model with claim arrivals based on a general compound Hawkes process 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 analytical calculation of ruin probabilities and the 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. Our results complement and extend current research on risk theory with self-exciting Hawkes processes. They allow to incorporate the real-world phenomenon of claims triggering subsequent claims and the peril of a resulting "contagion effect" into ruin probability estimations and asset-liability management considerations. Within the current extensive framework of regulatory requirements, our results can give indications about the enhanced risk introduced by claim arrival clustering and how to take it into account in order to adhere to given risk constraints.

At the same time, there are many approaches to generalize the present work in potential future research: As highlighted in Chapter 3, we have currently restricted our focus to a one-dimensional, exponential Hawkes process. As the knowledge about theoretical properties of Hawkes processes is steadily being advanced, more general specifications might become available for empirical usage. These could include processes with other excitation functions (e.g. long-range power law or mixed kernels) to incorporate real-world delay effects or multi-dimensional, marked Hawkes processes to model the evolution and interdependence of several classes of insurance claims. The status quo of Hawkes risk models summarized in Chapter 4 shows that there have been quite recent advancements (in particular the work of Cheng and Seol 2018) which can serve as a basis for further empirical implementation. In Chapter 5, we derived two approximations for the Hawkes risk model and have remarked that their accuracy (in particular for the pure diffusion

approximation) heavily depends on the model parameters. The analytical estimation of the rate of approximation is generally very complicated, but could be studied in future research, where Swishchuk 2000 could serve as background knowledge. With respect to the applications to empirical data in Chapter 6, we started by studying data from legal expenses insurance as we suspected a Hawkes process to be adequate here and the data was readily available with convenient granularity. However, there are obviously other classes of insurance where an empirical study would be of interest, in particular e.g. earthquake (re-)insurance or health insurance in regions with pandemics. Furthermore, as we have seen that the main interest for an insurer might be the branching structure of a Hawkes arrival process, it might be of interest to use empirical data for independent branching ratio approximation. The last chapter investigated an optimal investment problem under a mean-variance framework. It would naturally be of interest to study other objectives of insurers, e.g. studying an optimal strategy that minimizes the ruin probability by using an exponential utility function (see Browne 1995). Another generalisation would be to use an example with various risky assets with time-dependent deterministic (or even stochastic) parameters.

Bibliography

Aït-Sahalia, Yacine, Julio Cacho-Diaz, and Roger J.A. Laeven (2015). "Modeling financial contagion using mutually exciting jump processes". In: *Journal of Financial Economics* 117.3, pp. 585–606.

Albrecher, Hansjörg and Sören Asmussen (2006). "Ruin probabilities and aggregrate claims distributions for shot noise Cox processes". In: *Scandinavian Actuarial Journal* 2006.2, pp. 86–110.

Albrecher, Hansjörg and Onno J. Boxma (2004). "A ruin model with dependence between claim sizes and claim intervals". In: *Insurance: Mathematics and Economics* 35.2, pp. 245–254.

Albrecher, Hansjörg and Josef Kantor (2002). "Simulation of ruin probabilities for risk processes of Markovian type". In: *Monte Carlo Methods and Applications* 8.2.

Andersen, E. Sparre. (1957). "On the collective theory of risk in case of contagion between claims." In: *Transactions of the XVth International Congress of Actuaries* 6.

Angus, John E. (1994). "The Probability Integral Transform and Related Results". In: *SIAM Review* 36.4, pp. 652–654.

Asmussen, Sören (1984). "Approximations for the probability of ruin within finite time". In: *Scandinavian Actuarial Journal* 1984.1, pp. 31–57.

– (2010). *Ruin probabilities*. 2nd ed. Vol. v. 14. Advanced series on statistical science & applied probability. Singapore and Hackensack, N.J: World Scientific Pub. Co. ISBN: 978-981-4282-52-9.

Bacry, Emmanuel, K. Dayri, and J. F. Muzy (2012). "Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data". In: *The European Physical Journal B* 85.5, p. 1.

Bacry, Emmanuel, Iacopo Mastromatteo, and Jean-François Muzy (2015). *Hawkes processes in finance*. URL: `http://arxiv.org/pdf/1502.04592v2`.

Bacry, Emmanuel et al. (2013). "Modelling microstructure noise with mutually exciting point processes". In: *Quantitative Finance* 13.1, pp. 65–77.

Bäuerle, Nicole (2005). "Benchmark and mean-variance problems for insurers". In: *Mathematical Methods of Operations Research* 62.1, pp. 159–165.

Bielecki, Tomasz R. et al. (2005). "Continuous-time mean-variance portfolio selection with bankruptcy prohibition". In: *Mathematical Finance* 15.2, pp. 213–244.

Billingsley, Patrick (1999). *Convergence of probability measures*. 2. ed. Wiley series in probability and statistics Probability and statistics section. New York, NY: Wiley. ISBN: 0471197459.

Blitzstein, Joseph K. and Jessica Hwang (2015). *Introduction to probability*. Texts in statistical science. Boca Raton, FL: CRC Press. ISBN: 9781466575578.

Boudreault, Mathieu et al. (2006). "On a risk model with dependence between interclaim arrivals and claim sizes". In: *Scandinavian Actuarial Journal* 2006.5, pp. 265–285.

Boumezoued, Alexandre and Laurent Devineau (2017). *Individual claims reserving: a survey*. URL: `https://hal.archives-ouvertes.fr/hal-01643929`.

Bowsher, Clive G. (2007). "Modelling security market events in continuous time: Intensity based, multivariate point process models". In: *Journal of Econometrics* 141.2, pp. 876–912.

Brémaud, Pierre (2000). "An insensitivity property of Lundberg's estimate for delayed claims". In: *Journal of Applied Probability* 37.3, pp. 914–917.

Browne, Sid (1995). "Optimal Investment Policies for a Firm With a Random Risk Process: Exponential Utility and Minimizing the Probability of Ruin". In: *Mathematics of Operations Research* 20.4, pp. 937–958.

Cheng, Zailei and Youngsoo Seol (2018). *Gaussian Approximation of a Risk Model with Stationary Hawkes Arrivals of Claims*. URL: `http://arxiv.org/pdf/1801.07595v1`.

Chiu, Mei Choi and Duan Li (2006). "Asset and liability management under a continuous-time mean–variance optimization framework". In: *Insurance: Mathematics and Economics* 39.3, pp. 330–355.

Cox, D. R. and P. A. W. Lewis (1966). *The Statistical Analysis of Series of Events*. Dordrecht: Springer Netherlands. ISBN: 978-94-011-7803-7.

Czado, Claudia and Thorsten Schmidt (2011). *Mathematische Statistik*. Statistik und ihre Anwendungen. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. ISBN: 9783642172601.

Da Fonseca, José and Riadh Zaatour (2013). "Hawkes Process: Fast Calibration, Application to Trade Clustering and Diffusive Limit". In: *SSRN Electronic Journal*.

Daley, Daryl J. and David Vere-Jones, eds. (2003). *An Introduction to the Theory of Point Processes*. Probability and its Applications. New York: Springer-Verlag. ISBN: 0-387-95541-0.

– (2008). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. 2nd ed. Probability and Its Applications, A Series of the Applied Probability Trust. New York, NY: Applied Probability Trust. ISBN: 978-0-387-21337-8.

Dassios, Angelos and Ji-Wook Jang (2003). "Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity". In: *Finance and Stochastics* 7.1, pp. 73–95.

Dassios, Angelos and Hongbiao Zhao (2011). "A dynamic contagion process". In: *Advances in Applied Probability* 43.03, pp. 814–846.

– (2012). "Ruin by dynamic contagion claims". In: *Insurance: Mathematics and Economics* 51.1, pp. 93–106.

– (2013). "A Risk Model with Delayed Claims". In: *Journal of Applied Probability* 50.03, pp. 686–702.

Embrechts, Paul, Thomas Liniger, and Lu Lin (2011). "Multivariate Hawkes processes: an application to financial data". In: *Journal of Applied Probability* 48.A, pp. 367–378.

Errais, Eymen, Kay Giesecke, and Lisa R. Goldberg (2010). "Affine Point Processes and Portfolio Credit Risk". In: *SIAM Journal on Financial Mathematics* 1.1, pp. 642–665.

Fauth, Alexis and Ciprian A. Tudor (2012). *Modeling First Line Of An Order Book With Multivariate Marked Point Processes*. URL: `http://arxiv.org/pdf/1211.4157v1`.

Filimonov, V. and D. Sornette (2015). "Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data". In: *Quantitative Finance* 15.8, pp. 1293–1314.

Filimonov, Vladimir and Didier Sornette (2012). "Quantifying reflexivity in financial markets: toward a prediction of flash crashes". In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 85.5 Pt 2, p. 056108.

Grandell, Jan (1977). "A class of approximations of ruin probabilities". In: *Scandinavian Actuarial Journal* 1977.sup1, pp. 37–52.

– (1991). *Aspects of Risk Theory*. Springer Series in Statistics, Probability and its Applications. New York, NY: Springer. ISBN: 978-1-4613-9058-9.

Grimmett, Geoffrey and David R. Stirzaker (2006). *Probability and random processes*. 3. ed., reprinted, with corrections. Oxford: Oxford Univ. Press. ISBN: 9780198572237.

Hardiman, Stephen J., Nicolas Bercot, and Jean-Philippe Bouchaud (2013). "Critical reflexivity in financial markets: a Hawkes process analysis". In: *The European Physical Journal B* 86.10, p. 421.

Hardiman, Stephen J. and Jean-Philippe Bouchaud (2014). "Branching ratio approximation for the self-exciting Hawkes process". In: *Physical Review E* 90.6, p. 438.

Hawkes, Alan (1971). "Spectra of some self-exciting and mutually exciting point processes". In: *Biometrika* 58.1, pp. 83–90.

Hawkes, Alan and David Oakes (1974). "A cluster process representation of a self-exciting process". In: *Journal of Applied Probability* 11.03, pp. 493–503.

Jang, Ji-Wook (2004). "Martingale Approach for Moments of Discounted Aggregate Claims". In: *Journal of Risk & Insurance* 71.2, pp. 201–211.

Jang, Ji-Wook and Angelos Dassios (2013). "A bivariate shot noise self-exciting process for insurance". In: *Insurance: Mathematics and Economics* 53.3, pp. 524–532.

Karatzas, Ioannis and Steven E. Shreve, eds. (1998). *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. New York, NY: Springer New York. ISBN: 978-0-387-97655-6.

Klüppelberg, Claudia, Thomas Mikosch, and Claudia Kluppelberg (1995). "Explosive Poisson Shot Noise Processes with Applications to Risk Reserves". In: *Bernoulli* 1.1/2, p. 125.

Koo, Hyeng Keun (1998). "Consumption and Portfolio Selection with Labor Income: A Continuous Time Approach". In: *Mathematical Finance* 8.1, pp. 49–65.

Kühn, Christoph (2014). "Shot-Noise Processes". In: *Wiley StatsRef: Statistics Reference Online*. Ed. by N. Balakrishnan et al. Vol. 37. Chichester, UK: John Wiley & Sons, Ltd, p. 914. ISBN: 9781118445112.

Laub, Patrick J., Thomas Taimre, and Philip K. Pollett (2015). *Hawkes Processes*. URL: `http://arxiv.org/pdf/1507.02822v1`.

Leippold, Markus, Fabio Trojani, and Paolo Vanini (2004). "A geometric approach to multiperiod mean variance optimization of assets and liabilities". In: *Journal of Economic Dynamics and Control* 28.6, pp. 1079–1113.

Léveillé, Ghislain and Emmanuel Hamel (2018). "Conditional, Non-Homogeneous and Doubly Stochastic Compound Poisson Processes with Stochastic Discounted Claims". In: *Methodology and Computing in Applied Probability* 20.1, pp. 353–368.

Lewis, E. and G. Mohler (2011). "A Nonparametric EM algorithm for Multiscale Hawkes Processes". In: *Joint Satistical Meetings*.

Lewis, P. A. W. and G. S. Shedler (1979). "Simulation of nonhomogeneous poisson processes by thinning". In: *Naval Research Logistics Quarterly* 26.3, pp. 403–413.

Li, Duan, Tsz-Fung Chan, and Wan-Lung Ng (1998). "Safety-first dynamic portfolio selection." In: *Dynamics of Continuous, Discrete and Impulsive Systems Series B: Application and Algorithm* 4.4, pp. 585–600.

Li, Duan and Wan-Lung Ng (2000). "Optimal Dynamic Portfolio Selection: Multiperiod Mean-Variance Formulation". In: *Mathematical Finance* 10.3, pp. 387–406.

Liniger, Thomas Josef (2009). *Multivariate Hawkes processes*.

Lundberg, Filip (1903). "Approximerad Framställning av Sannolikehetsfunktionen: Återförsäkering av Kollektivrisker". In:

Macci, Claudio and Giovanni Luca Torrisi (2004). "Asymptotic results for perturbed risk processes with delayed claims". In: *Insurance: Mathematics and Economics* 34.2, pp. 307–320.

Markowitz, Harry (1952). "Portfolio Selection". In: *The Journal of Finance* 7.1, p. 77.

Markowitz, Harry (1956). "The optimization of a quadratic function subject to linear constraints". In: *Naval Research Logistics Quarterly* 3.1-2, pp. 111–133.

Massouli, Laurent and Pierre Bremaud (1996). "Stability of nonlinear Hawkes processes". In: *The Annals of Probability* 24.3, pp. 1563–1588.

Meng, Qingbin, Xin Zhang, and Junyi Guo (2008). "On a risk model with dependence between claim sizes and claim intervals". In: *Statistics & Probability Letters* 78.13, pp. 1727–1734.

Merton, Robert (1975). "Optimum Consumption and Portfolio Rules in a Continuous-Time Model". In: *Stochastic Optimization Models in Finance*. Elsevier, pp. 621–661. ISBN: 9780127808505.

Muni Toke, Ioane and Fabrizio Pomponio (2011). "Modelling Trades-Through in a Limited Order Book Using Hawkes Processes". In: *SSRN Electronic Journal*.

Oakes, David (1975). "The Markovian self-exciting process". In: *Journal of Applied Probability* 12.01, pp. 69–77.

Ogata, Yoshiko (1978). "The asymptotic behaviour of maximum likelihood estimators for stationary point processes". In: *Annals of the Institute of Statistical Mathematics* 30.1, pp. 243–261.

– (1981). "On Lewis' simulation method for point processes". In: *IEEE Transactions on Information Theory* 27.1, pp. 23–31.

– (1988). "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes". In: *Journal of the American Statistical Association* 83.401, p. 9.

– (1999). "Seismicity Analysis through Point-process Modeling: A Review". In: *Seismicity Patterns, their Statistical Significance and Physical Meaning*. Basel: Birkhäuser Basel, pp. 471–507. ISBN: 978-3-7643-6209-6.

Ozaki, Tohru (1979). "Maximum likelihood estimation of Hawkes' self-exciting point processes". In: *Annals of the Institute of Statistical Mathematics* 31.1, pp. 145–155.

Privault, Nicolas (2019). *Lecture Notes on Stochastic Calculus for Jump Processes*. URL: `http://www.ntu.edu.sg/home/nprivault/MA5182/stochastic-calculus-jump-processes.pdf`.

Rambaldi, Marcello, Paris Pennesi, and Fabrizio Lillo (2015). "Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach". In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 91.1, p. 012819.

Rasmussen, Jakob G. (2011). "Temporal point processes: the conditional intensity function." In: URL: `http://people.math.aau.dk/~jgr/teaching/punktproc11/tpp.pdf`.

Revuz, Daniel and Marc Yor (1999). *Continuous Martingales and Brownian Motion*. Corrected Third Printing of the Third Edition. Vol. 293. Grundlehren der mathematischen Wissenschaften, A Series of Comprehensive Studies in Mathematics. Berlin and Heidelberg: Springer. ISBN: 978-3-642-08400-3.

Rizoiu, Marian-Andrei et al. (2017). *A Tutorial on Hawkes Processes for Events in Social Media*. URL: `http://arxiv.org/pdf/1708.06401v2`.

Rolski, Tomasz (2001). *Stochastic processes for insurance and finance*. Reprinted. Wiley series in probability and statistics. Chichester: Wiley. ISBN: 0471959251.

Schmidli, Hanspeter (1994). "Diffusion approximations for a risk process with the possibility of borrowing and investment". In: *Communications in Statistics. Stochastic Models* 10.2, pp. 365–388.

– (2018). *Risk theory*. Springer actuarial lecture notes. Cham: Springer. ISBN: 978-3-319-72004-3.

Sharpe, William F. and Lawrence G. Tint (1990). "Liabilities— A New Approach". In: *The Journal of Portfolio Management* 16.2, pp. 5–10.

Skhorokhod, Anatoliy V. (2014). *Studies in the Theory of Random Processes*. Newburyport: Dover Publications. ISBN: 9780486642406.

Stabile, Gabriele and Giovanni Luca Torrisi (2010). "Risk Processes with Non-stationary Hawkes Claims Arrivals". In: *Methodology and Computing in Applied Probability* 12.3, pp. 415–429.

Swishchuk, Anatoliy (2000). *Random Evolutions and their Applications: New Trends*. Vol. 504. Mathematics and Its Applications. Dordrecht: Springer. ISBN: 978-90-481-5441-8.

– (2017a). *General Compound Hawkes Processes in Limit Order Books*. URL: `http://arxiv.org/pdf/1706.07459v2`.

– (2017b). *Risk Model Based on General Compound Hawkes Processes*. URL: `http://arXiv.org/pdf/1706.09038`.

Swishchuk, Anatoliy et al. (2017). *Compound Hawkes Processes in Limit Order Books*. URL: `http://arxiv.org/pdf/1712.03106v1`.

Truccolo, Wilson (2016). "From point process observations to collective neural dynamics: Nonlinear Hawkes process GLMs, low-dimensional dynamics and coarse graining". In: *Journal of physiology, Paris* 110.4 Pt A, pp. 336–347.

Uhlenbeck, G. E. and L. S. Ornstein (1930). "On the Theory of the Brownian Motion". In: *Physical Review* 36.5, pp. 823–841.

Vadori, Nelson and Anatoliy Swishchuk (2015). "Strong Law of Large Numbers and Central Limit Theorems for Functionals of Inhomogeneous Semi-Markov Processes". In: *Stochastic Analysis and Applications* 33.2, pp. 213–243.

Veen, Alejandro and Frederic P. Schoenberg (2008). "Estimation of Space–Time Branching Process Models in Seismology Using an EM–Type Algorithm". In: *Journal of the American Statistical Association* 103.482, pp. 614–624.

Wald, Abraham (1944). "On Cumulative Sums of Random Variables". In: *The Annals of Mathematical Statistics* 15.3, pp. 283–296.

Whitt, Ward (1970). "Weak Convergence of Probability Measures on the Function Space $C[0, \infty)$". In: *The Annals of Mathematical Statistics* 41.3, pp. 939–944.

– (2002). *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer Series in Operations Research and Financial Engineering. New York, NY: AT&T. ISBN: 0387953582.

Woess, Wolfgang (2009). *Denumerable Markov chains: Generating functions, boundary theory, random walks on trees*. EMS Textbooks in mathematics. Zürich: European Mathematical Society. ISBN: 978-3-03719-071-5.

Xie, Shuxiang, Zhongfei Li, and Shouyang Wang (2008). "Continuous-time portfolio selection with liability: Mean–variance model and stochastic LQ approach". In: *Insurance: Mathematics and Economics* 42.3, pp. 943–953.

Yong, Jiongmin and Xun Yu Zhou (1999). *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Vol. 43. Applications of Mathematics, Stochastic Modelling and Applied Probability. New York, NY: Springer. ISBN: 978-1-4612-7154-3.

Yuen, Kam C., Junyi Guo, and Kai W. Ng (2005). "On Ultimate Ruin in a Delayed-Claims Risk Model". In: *Journal of Applied Probability* 42.01, pp. 163–174.

Zenios, Stauros Andrea and William T. Ziemba (2007). *Applications and case studies*. Vol. 2. Handbooks in finance. Amsterdam: Elsevier. ISBN: 978-0444508751.

Zhang, Changyong (2016). "Modeling High Frequency Data Using Hawkes Processes with Power-law Kernels". In: *Procedia Computer Science* 80, pp. 762–771.

Zhou, X. Y. and D. Li (2000). "Continuous-Time Mean-Variance Portfolio Selection: A Stochastic LQ Framework". In: *Applied Mathematics and Optimization* 42.1, pp. 19–33.

Zhu, Lingjiong (2013). "Ruin probabilities for risk processes with non-stationary arrivals and subexponential claims". In: *Insurance: Mathematics and Economics* 53.3, pp. 544–550.

Zhu, S.-S., D. Li, and S.-Y. Wang (2004). "Risk Control Over Bankruptcy in Dynamic Portfolio Selection: A Generalized Mean-Variance Formulation". In: *IEEE Transactions on Automatic Control* 49.3, pp. 447–457.

Appendix

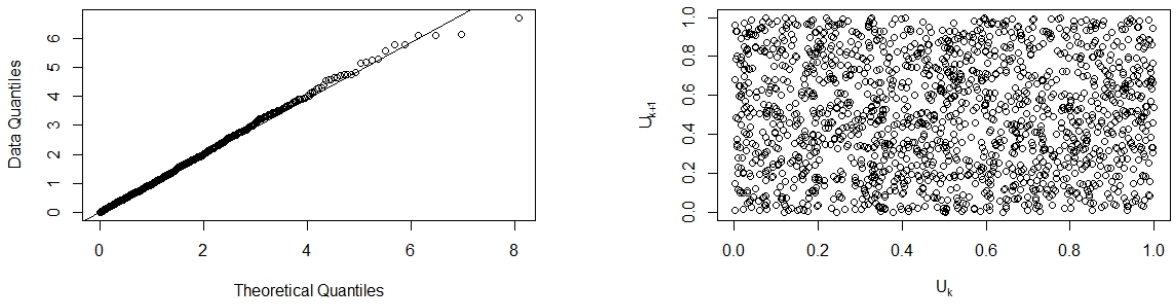
A Parameter estimation and goodness of fit testing for simulated data

For four different parameter sets (λ, α, β), we simulate an exponential Hawkes process on the interval $[0, 1000]$ and then use the maximum likelihood method to estimate the parameters $(\hat{\lambda}, \hat{\alpha}, \hat{\beta})$ for the realisation of each process $\{t_1, t_2, \cdots\}$. In each case, the length of the observed interval leads to more than 1600 arrivals which can be considered a large enough data set to produce reliable estimates. We use the implementation of the *Nelder-Mead* algorithm from the *R* package *mle4* to minimize the negative value of the log-likelihood function given in (3.14) with bounds $[0, 500]$ on each parameter value. As concerns about dependence of the result on the initial parameter set have been raised, we use the four sets of (λ, α, β) as starting parameters in each case. We observe that apart from in one case, the algorithm gives the same, quite accurate results for all sets of initial parameter values. In the case where different initial parameters lead to different results, the result closest to the true parameter values turns out to be the one with the lowest value of the objective function (negative log-likelihood). This leads us the approach of using several sets of different starting values and picking the result with the lowest value of the objective when working with empirical data sets.

(λ, α, β)	$N(T)$	Initial parameters $(\lambda_0, \alpha_0, \beta_0)$	Maximum likelihood estimates $(\hat{\lambda}, \hat{\alpha}, \hat{\beta})$	Value of objective $-l(\hat{\lambda}, \hat{\alpha}, \hat{\beta})$
$(1, 2, 5)$	1621	$(1, 2, 5)$	$(0.9887, 2.0201, 5.1987)$	597.2086
		$(1, 2, 2.1)$	$(0.9887, 2.0201, 5.1987)$	597.2086
		$(1, 0.2, 0.5)$	$(0.9887, 2.0201, 5.1987)$	597.2086
		$(10, 200, 400)$	$(0.9887, 2.0201, 5.1987)$	597.2086
$(1, 2, 2.1)$	24476	$(1, 2, 5)$	$(1.0364, 1.9653, 2.0518)$	-65126.02
		$(1, 2, 2.1)$	$(1.0364, 1.9653, 2.0518)$	-65126.02
		$(1, 0.2, 0.5)$	$(1.0364, 1.9653, 2.0518)$	-65126.02
		$(10, 200, 400)$	$(1.0364, 1.9653, 2.0518)$	-65126.02
$(1, 0.2, 0.5)$	1594	$(1, 2, 5)$	$(1.1079, 0.1394, 0.4559)$	834.6009
		$(1, 2, 2.1)$	$(1.1079, 0.1394, 0.4559)$	834.6009
		$(1, 0.2, 0.5)$	$(1.1079, 0.1394, 0.4559)$	834.6009
		$(10, 200, 400)$	$(1.5940, 0, 490.3411)$	850.8030
$(10, 200, 400)$	19956	$(1, 2, 5)$	$(9.8339, 198.561, 391.4656)$	-56373.87
		$(1, 2, 2.1)$	$(9.8339, 198.561, 391.4656)$	-56373.87
		$(1, 0.2, 0.5)$	$(9.8339, 198.561, 391.4656)$	-56373.87
		$(10, 200, 400)$	$(9.8339, 198.561, 391.4656)$	-56373.87

Table 1: Results of parameter estimation using the maximum likelihood method on simulated exponential Hawkes processes with four different parameter sets.

We give an example for the goodness of fit testing method using the random time change theorem introduced in Section 3.4. For the first set of true parameters $(\lambda, \alpha, \beta) = (1, 2, 5)$ above, we transform the 1621 simulated arrival times using the compensator of the exponential Hawkes process given in (3.13) given the estimated parameters $(\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = (0.9887, 2.0201, 5.1987)$. We then compare the interarrival times of the transformed process $\{t_1^*, t_2^*, \cdots\}$ against an $Exp(1)$-distribution using a QQ-plot. Furthermore, we plot the transformed points (U_k, U_{k+1}), where $U_k := 1 - e^{-(t_k^* - t_{k-1}^*)}$, on the unit square $[0, 1] \times [0, 1]$ to check for independence according to the probability integral transform. The results are displayed in Figure 29.



(a) (b)

Figure 29: (a) shows the QQ-plot of interarrival times of the transformed point process given $(\hat{\lambda}, \hat{\alpha}, \hat{\beta})$ against a $Exp(1)$ distribution. We observe a very good fit indicated by the data points following the QQ-line very closely. (b) shows the series (U_k, U_{k+1}) plotted in the unit square, where they are scattered quite evenly without any noticeable patterns indicating independence of the transformed interarrival times.

B Proof of Lemma 3.13 continued

Proof of Lemma 3.13 continued.
For (3.23) choose $f(X(t)) \equiv N(t)^2$, then

$$\mathcal{L}f(X(t)) = \beta(\lambda - \lambda^*(t))0 + \lambda^*(t)[(N(t)+1)^2 - N(t)^2] = \lambda^*(t)[2N(t)+1] = 2N(t)\lambda^*(t) + \lambda^*(t)$$

Dynkin's formula (with $s = t$ and $t = 0$) and Fubini's Theorem yield

$$\begin{aligned}\mathbb{E}[N(t)^2] &= N(0)^2 + \mathbb{E}\Big[\int_0^t 2N(u)\lambda^*(u) + \lambda^*(u)du\Big] \\ &= N(0)^2 + 2\int_0^t \mathbb{E}[N(u)\lambda^*(u)]du + \int_0^t \mathbb{E}[\lambda^*(u)]du\end{aligned}$$

Finally, differentiating gives

$$d\mathbb{E}[N(t)^2] = 2\mathbb{E}[N(t)\lambda^*(t)]dt + \mathbb{E}[\lambda^*(t)]dt$$

For (3.24), let $f(X(t)) \equiv \lambda^*(t)N(t)$. Analogously to above,

$$\begin{aligned}\mathcal{L}f(X(t)) &= \beta(\lambda - \lambda^*(t))N(t) + \lambda^*(t)[(\lambda^*(t)+\alpha)(N(t)+1) - \lambda^*(t)N(t)] \\ &= \beta\lambda N(t) - \beta\lambda^*(t)N(t) + \lambda^*(t)[\lambda^*(t) + \alpha N(t) + \alpha] \\ &= \beta\lambda N(t) - \beta\lambda^*(t)N(t) + \lambda^*(t)^2 + \lambda^*(t)\alpha N(t) + \lambda^*(t)\alpha \\ &= (\alpha - \beta)\lambda^*(t)N(t) + \beta\lambda N(t) + \lambda^*(t)\alpha + \lambda^*(t)^2\end{aligned}$$

$$\Rightarrow \mathbb{E}[\lambda^*(t)N(t)] = \lambda_0 N(0) + (\alpha-\beta)\int_0^t \mathbb{E}[\lambda^*(u)N(u)]du + \beta\lambda\int_0^t \mathbb{E}[N(u)]du$$
$$+ \alpha\int_0^t \mathbb{E}[\lambda^*(u)]du + \int_0^t \mathbb{E}[\lambda^*(u)^2]du$$
$$\Rightarrow d\mathbb{E}[\lambda^*(t)N(t)] = (\alpha-\beta)\mathbb{E}[\lambda^*(t)N(t)]dt + \beta\lambda\mathbb{E}[N(t)]dt + \alpha\mathbb{E}[\lambda^*(t)]dt + \mathbb{E}[\lambda^*(t)^2]dt.$$

For (3.25), let $f(X(t)) \equiv \lambda^*(t)^2$. As above,

$$\begin{aligned}\mathcal{L}f(X(t)) &= \beta(\lambda-\lambda^*(t))2\lambda^*(t) + \lambda^*(t)[(\lambda^*(t)+\alpha)^2 - \lambda^*(t)^2]\\ &= 2\beta\lambda\lambda^*(t) - 2\beta\lambda^*(t)^2 + \alpha^2\lambda^*(t) + 2\alpha\lambda^*(t)^2\\ &= (\alpha^2 - 2\beta\lambda)\lambda^*(t) + 2(\alpha-\beta)\lambda^*(t)^2\end{aligned}$$
$$\Rightarrow \mathbb{E}[\lambda^*(t)^2] = \lambda_0^2 + (\alpha^2-2\beta\lambda)\int_0^t \mathbb{E}[\lambda^*(u)]du + 2(\alpha-\beta)\int_0^t \mathbb{E}[\lambda^*(u)^2]du$$
$$\Rightarrow d\mathbb{E}[\lambda^*(t)^2] = (\alpha^2-2\beta\lambda)\mathbb{E}[\lambda^*(t)]dt + 2(\alpha-\beta)\mathbb{E}[\lambda^*(t)^2]dt.$$

□

C Proof of Proposition 3.14 continued

Proof. We proceed analogously for the second moment by additionally making use of (3.23),(3.24) and (3.25). Define

$$I = \mathbb{E}[(N(t_2)-N(t_1))^2] = \mathbb{E}[\mathbb{E}_{t_1}[N(t_2)^2] - 2N(t_1)\mathbb{E}_{t_1}[N(t_2)] + N(t_1)^2],$$

where we have used the tower property and $\mathbb{E}_{t_1}[\cdot] := \mathbb{E}[\cdot|\mathcal{F}(t_1)]$. By ODE (3.23) and (3.21) respectively, it holds that

$$\mathbb{E}_{t_1}[N(t_2)^2] = N(t_1)^2 + 2\int_{t_1}^{t_2} \mathbb{E}_{t_1}[\lambda^*(u)N(u)]du + \int_{t_1}^{t_2} \mathbb{E}_{t_1}[\lambda^*(u)]du,$$

and

$$\mathbb{E}_{t_1}[N(t_2)] = N(t_1) + \int_{t_1}^{t_2} \mathbb{E}_{t_1}[\lambda^*(u)]du,$$

which implies

$$I = \underbrace{\int_{t_1}^{t_2} \mathbb{E}[\lambda^*(u)N(u)]du}_{=:I_1} + \int_{t_1}^{t_2} \mathbb{E}[\lambda^*(u)]du - 2\underbrace{N(t_1)\int_{t_1}^{t_2} \mathbb{E}_{t_1}[\lambda^*(u)]du}_{=:I_2}.$$

Using (3.24) it holds

$$I_1 = \int_{t_1}^{t_2} e^{(\alpha-\beta)(u-t_1)}\mathbb{E}[\lambda^*(t_1)N(t_1)]du + \int_{t_1}^{t_2}\int_{t_1}^{u} e^{(\alpha-\beta)(u-s)}\big[\beta\lambda\mathbb{E}[N(s)] + \mathbb{E}[(\lambda^*(s))^2] + \alpha\mathbb{E}[\lambda^*(s)]\big]dsdu,$$

whereas using (3.22) for I_2 yields

$$I_2 = \mathbb{E}[\lambda^*(t_1)N(t_1)]\int_{t_1}^{t_2} e^{(\alpha-\beta)(u-t_1)}du + \beta\lambda\mathbb{E}[N(t_1)]\int_{t_1}^{t_2}\int_{t_1}^{u} e^{(\alpha-\beta)(u-s)}dsdu.$$

Using that $\mathbb{E}[N(s)] = \mathbb{E}[N(t_1)] + \int_{t_1}^{s}\mathbb{E}[\lambda^*(u)]du$ by (3.21), substituting I_1 and I_2 into I and rearranging the terms yields

$$I = \int_{t_1}^{t_2}\mathbb{E}[\lambda^*(u)]du + 2\int_{t_1}^{t_2}\int_{t_1}^{u} e^{(\alpha-\beta)(u-s)}\big[\beta\lambda\int_{t_1}^{s}\mathbb{E}[\lambda^*(u)]du + \mathbb{E}[(\lambda^*(s))^2] + \alpha\mathbb{E}[\lambda^*(s)]\big]dsdu.$$

Thus, we calculate the expected second moment of the number of jumps during an interval of length τ by first conditioning on $\mathcal{F}(t)$ and considering $\mathbb{E}[(N(t+\tau)-N(t))^2]$, which as we see above depends on $\mathbb{E}[\lambda^*(t)]$ and $\mathbb{E}[(\lambda^*(t))^2]$. The dependence of these expressions w.r.t. the starting value λ_0 is again eliminated by considering the limit as $t\to\infty$, thus taking the process to its stationary regime. Denote for now $\lim_{t\to\infty}\mathbb{E}[\lambda^*(t)] = \Lambda$ and $\lim_{t\to\infty}\mathbb{E}[(\lambda^*(t))^2] = \Lambda_2$. Using the above expression for I with $t_1 = t$ and $t_2 = t+\tau$ and letting $t\to\infty$ then yields

$$\begin{aligned}\lim_{t\to\infty}\mathbb{E}[(N(t+\tau)-N(t))^2] = \tau\Lambda + 2\beta\lambda\Lambda\lim_{t\to\infty}\int_{t}^{t+\tau}\int_{t}^{u} e^{(\alpha-\beta)(u-s)}\int_{t}^{s} drdsdu \\ + 2(\Lambda_2+\alpha\Lambda)\int_{t}^{t+\tau}\int_{t}^{u} e^{(\alpha-\beta)(u-s)}dsdu,\end{aligned}$$

where

$$\int_{t}^{t+\tau}\int_{t}^{u} e^{(\alpha-\beta)(u-s)}\int_{t}^{s} drdsdu = -\frac{\tau^2}{2(\alpha-\beta)} - \frac{\tau}{(\alpha-\beta)^2} + \frac{e^{(\alpha-\beta)\tau}-1}{(\alpha-\beta)^3}$$

$$\int_{t}^{t+\tau}\int_{t}^{u} e^{(\alpha-\beta)(u-s)}dsdu = -\frac{\tau}{\alpha-\beta} + \frac{e^{(\alpha-\beta)\tau}-1}{(\alpha-\beta)^2}.$$

Substracting $\lim_{t\to\infty}\mathbb{E}[N(t+\tau)-N(t)]^2$ (which is known from (3.26)) and combining the terms finally yields (3.27).

The autocovariance function is deduced analogously by considering $\mathbb{E}[(N(t_1)-$

$N(t))(N(t_3) - N(t_2))]$ where $t < t_1 < t_2 < t_3$ and $\tau_1 := t_1 - t, \delta := t_2 - t_1, \tau_2 := t_3 - t_2$. By successive application of the tower property, we obtain

$$\mathbb{E}[(N(t_1) - N(t))(N(t_3) - N(t_2))] = \mathbb{E}[\mathbb{E}_t[\mathbb{E}_{t_1}[\mathbb{E}_{t_2}[(N(t_1) - N(t))(N(t_3) - N(t_2))]]]].$$

Using that $N(t)$ and $N(t_1)$ are $\mathcal{F}_{t_2}$-measurable and the computations for the first moment (on $(N(t_3) - N(t_2))$) yields for the innermost expectation

$$\mathbb{E}_{t_2}[(N(t_1) - N(t))(N(t_3) - N(t_2))] = (N(t_1) - N(t)) \cdot \left[\frac{\lambda\beta(e^{(\alpha-\beta)\tau_2} - 1 - (\alpha-\beta)\tau_2)}{(\alpha-\beta)^2} + \lambda^*(t_2)\frac{e^{(\alpha-\beta)\tau_2} - 1}{\alpha-\beta}\right].$$

Everything in the above expression is $\mathcal{F}_{t_1}$-measurable apart from $\lambda^*_{t_2}$, so for the next conditioning step, we need to remember that by the calculations for the first moment we also know

$$\mathbb{E}_{t_1}[\lambda^*(t_2)] = \frac{\beta\lambda}{\alpha-\beta}\big(e^{(\alpha-\beta)\delta} - 1\big) + \lambda^*(t_1)e^{(\alpha-\beta)\delta}.$$

So for the last conditioning (on $\mathcal{F}_t$) we need to remember likewise

$$\mathbb{E}_t[\lambda^*(t_1)] = \frac{\beta\lambda}{\alpha-\beta}\big(e^{(\alpha-\beta)\tau_1} - 1\big) + \lambda^*(t)e^{(\alpha-\beta)\tau_1},$$

and $\mathbb{E}_t[\lambda^*(t_1)N(t_1)]$ which is known from the previous calculations.
Finally, letting $t \to \infty$ and setting $\tau_1 = \tau_2 = \tau$ yields

$$\lim_{t\to\infty}\mathbb{E}[(N(t+\tau) - N(t))(N(t+2\tau+\delta) - N(t+\tau+\delta))] = \frac{\lambda\beta\alpha(2\beta-\alpha)(e^{(\alpha-\beta)\tau} - 1)^2}{2(\alpha-\beta)^4}e^{(\alpha-\beta)\delta} + \frac{\lambda^2\beta^2}{(\beta-\alpha)^2}\tau^2,$$

where substracting $\mathbb{E}[N(\tau)]^2$ yields (3.28). □

D Stability of two-dimensional exponential Hawkes process

Recall that the matrix kernel for our bivariate exponential Hawkes process is given by

$$\Phi(t) = \begin{pmatrix}\mu^{(s)}(t) & \mu^{(c)}(t)\\ \mu^{(c)}(t) & \mu^{(s)}(t)\end{pmatrix} = \begin{pmatrix}\alpha^{(s)}e^{-\beta^{(s)}t} & \alpha^{(c)}e^{-\beta^{(c)}t}\\ \alpha^{(c)}e^{-\beta^{(c)}t} & \alpha^{(s)}e^{-\beta^{(s)}t}\end{pmatrix} \qquad \forall t > 0.$$

and the stability condition in this case is that that the spectral radius of $\|\Phi\| = \{\|\mu^{ij}\|\}_{i,j=1}^D$ should be smaller than 1, where the spectral radius is simply the largest (absolute) eigenvalue. We know from the one-dimensional-case that $\|\Phi\|$ is given by

$$\|\Phi\| = \begin{pmatrix}\alpha^{(s)}/\beta^{(s)} & \alpha^{(c)}/\beta^{(c)}\\ \alpha^{(c)}/\beta^{(c)} & \alpha^{(s)}/\beta^{(s)}\end{pmatrix} =: \begin{pmatrix}\eta_{(s)} & \eta_{(c)}\\ \eta_{(c)} & \eta_{(s)}\end{pmatrix}.$$

The eigenvalues μ_1, μ_2 of $\|\Phi\|$ are then given by the solutions of $det(\mu \mathbf{1} - \|\Phi\|) = 0$, where $\mathbf{1}$ is the (2×2)-identity matrix. Thus

$$det \begin{pmatrix} \mu - \eta_{(s)} & -\eta_{(c)} \\ -\eta_{(c)} & \mu - \eta_{(s)} \end{pmatrix} = (\mu - \eta_{(s)})^2 - \eta_{(c)}^2 = \mu^2 - 2\eta_{(s)}\mu + \eta_{(s)}^2 - \eta_{(c)}^2 = 0$$

has solutions

$$\mu_{1,2} = \frac{2\eta_{(s)} \pm \sqrt{4\eta_{(c)}^2}}{2} = \eta_{(s)} \pm \eta_{(c)}.$$

As both $\eta_{(s)}$ and $\eta_{(c)}$ are positive, the spectral radius of $\|\Phi\|$ is $\eta_{(s)} + \eta_{(c)}$, which yields the stability condition.

E Exponential Hawkes process covariance density

We would like to compute the covariance density $\phi(t)$ of an exponential Hawkes process with background rate $\lambda = 1$, which is described by the equation

$$\phi(t) = \underbrace{\frac{\alpha e^{-\beta t}}{1 - \alpha/\beta}}_{=:g_1(t)} + \underbrace{\alpha e^{-\beta t} \int_0^\infty e^{-\beta v} \phi(v) dv}_{=:g_2(t)} + \underbrace{\alpha e^{-\beta t} \int_0^t e^{\beta v} \phi(v) dv}_{=:g_3(t)}.$$

This is a *Wiener-Hopf type* integral equation and can be solved by using the *Laplace transform* as described for the general case in Laub, Taimre, and Pollett 2015, Theorem 2. Recall that the *Laplace transform* of a function $f(t)$ is given by

$$\mathcal{L}\{f(t)\}(s) = \int_0^\infty f(t) e^{-st} dt, \qquad \forall t \geq 0.$$

Let n denote the *branching ratio*, in this case $n = \frac{1}{1-\alpha/\beta}$. Thus, the Laplace transforms of $g_1(t), g_2(t)$ and $g_3(t)$ are given by

$$\mathcal{L}\{g_1(t)\}(s) = \frac{\alpha}{1 - \alpha/\beta} \int_0^\infty e^{(\beta+s)t} dt = \frac{\alpha n}{\beta + s},$$

$$\begin{aligned} \mathcal{L}\{g_2(t)\}(s) &= \int_0^\infty \left(\alpha e^{-\beta t} \int_0^\infty e^{-\beta v} \phi(v) \right) e^{-st} dt = \alpha \int_0^\infty e^{-\beta v} \phi(v) \int_0^\infty e^{-(s+\beta)t} dt dv \\ &= \frac{\alpha}{s + \beta} \int_0^\infty e^{-\beta v} \phi(v) dv = \frac{\alpha}{s + \beta} \mathcal{L}\{\phi(t)\}(\beta), \end{aligned}$$

and

$$\begin{aligned}\mathcal{L}\{g_3(t)\}(s) &= \int_0^\infty \left(\alpha e^{-\beta t}\int_0^t e^{\beta v}\phi(v)\right)e^{-st}dt = \alpha\int_0^\infty e^{-(\beta+s)t}\int_0^t e^{\beta v}\phi(v)dvdt \\ &= \frac{\alpha}{s+\beta}\mathcal{L}\{\phi(t)\}(s).\end{aligned}$$

Putting everything together using the linearity of the Laplace transform yields

$$\mathcal{L}\{\phi(t)\}(s) = \frac{\alpha}{s+\beta}\big(n + \mathcal{L}\{\phi(t)\}(\beta) + \mathcal{L}\{\phi(t)\}(s)\big),$$

where choosing $s = \beta$ implies

$$\begin{aligned}\mathcal{L}\{\phi(t)\}(\beta) &= \frac{\alpha}{2\beta}\big(n + 2\mathcal{L}\{\phi(t)\}(\beta)\big) \\ \iff \mathcal{L}\{\phi(t)\}(\beta) &= \frac{\alpha n}{2(\beta-\alpha)}.\end{aligned}$$

Thus for general s it holds

$$\begin{aligned}\mathcal{L}\{\phi(t)\}(s) &= \frac{\alpha}{s+\beta}\left(n + \frac{\alpha n}{2(\beta-\alpha)} + \mathcal{L}\{\phi(t)\}(s)\right) \\ \iff \mathcal{L}\{\phi(t)\}(s) &= \frac{\frac{\alpha}{s+\beta}\big(n + \frac{\alpha n}{2(\beta-\alpha)}\big)}{1 - \frac{\alpha}{s+\beta}} = \frac{\alpha n(2\beta-\alpha)}{2(\beta-\alpha)(s+\beta-\alpha)}.\end{aligned}$$

Therefore, the covariance density $\phi(t)$ is given by

$$\begin{aligned}\phi(t) &= \mathcal{L}^{-1}\left(\frac{\alpha n(2\beta-\alpha)}{2(\beta-\alpha)(s+\beta-\alpha)}\right) = \frac{\alpha n(2\beta-\alpha)}{2(\beta-\alpha)}\mathcal{L}^{-1}\left(\frac{1}{s+\beta-\alpha}\right) \\ &= \frac{\alpha\beta(2\beta-\alpha)}{2(\beta-\alpha)^2}e^{-(\beta-\alpha)t},\end{aligned}$$

where $\mathcal{L}^{-1}$ denotes the inverse Laplace transform and we have resubstituted $n = \frac{\beta}{\beta-\alpha}$.

F Auxiliary calculations for proof of Theorem 7.5

We would like to solve the following linear ODE for $\mathbb{E}[X^*(t)]$:

$$\begin{aligned}d\mathbb{E}[X^*(t)] &= \big[(r(t)-\varsigma(t))\mathbb{E}[X^*(t)] - \varsigma(t)\vartheta(t) + \varsigma(t)\gamma + \kappa(t)\big]dt, \\ \mathbb{E}[X^*(0)] &= x.\end{aligned}$$

We first rewrite the ODE in standard form:

$$\frac{d\mathbb{E}[X^*(t)]}{dt} + (\varsigma(t) - r(t))\mathbb{E}[X^*(t)] = -\varsigma(t)\vartheta(t) + \varsigma(t)\gamma + \kappa(t)$$

$$\mathbb{E}[X^*(0)] = x.$$

The standard technique of multiplying both sides with the integrating factor $e^{\int_0^s (\varsigma(u)-r(u))du}$ yields

$$d\left(\mathbb{E}[X^*(s)]e^{\int_0^s (\varsigma(u)-r(u))du}\right) = (-\varsigma(s)\vartheta(s) + \varsigma(s)\gamma + \kappa(s))e^{\int_0^s (\varsigma(u)-r(u))du}.$$

Integrating both sides of the equation from 0 to t and using the initial condition $\mathbb{E}[X^*(0)] = x$ yields

$$\mathbb{E}[X^*(t)]e^{\int_0^t (\varsigma(u)-r(u))du} - x = \int_0^t (-\varsigma(s)\vartheta(s) + \varsigma(s)\gamma + \kappa(s))e^{\int_0^s (\varsigma(u)-r(u))du} ds.$$

Finally, solving for $\mathbb{E}[X^*(t)]$ yields (7.27):

$$\mathbb{E}[X^*(t)] = xe^{\int_0^t (r(u)-\varsigma(u))du} + \int_0^t (\varsigma(s)\gamma + \kappa(s) - \varsigma(s)\vartheta(s))e^{\int_s^t (r(z)-\varsigma(z))dz} ds.$$

We would like to solve the following ODE for $\mathbb{E}[X^*(t)^2]$:

$$\begin{aligned} d\mathbb{E}[X^*(t)^2] = \big[&(2r(t) - \varsigma(t))\mathbb{E}[X^*(t)^2] + 2\kappa(t)\mathbb{E}[X^*(t)] + \varsigma(t)(\vartheta(t) - \gamma)^2 + \delta(t)'\delta(t) \\ &- \delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t) + \delta_0(t)^2\big]dt, \\ \mathbb{E}[X^*(0)^2] = x^2.& \end{aligned}$$

Again, we can rewrite the equation in standard form

$$\begin{aligned} \frac{d\mathbb{E}[X^*(t)^2]}{dt} + (\varsigma(t) - 2r(t))\mathbb{E}[X^*(t)^2] =& 2\kappa(t)\mathbb{E}[X^*(t)] + \varsigma(t)(\vartheta(t) - \gamma)^2 + \delta(t)'\delta(t) \\ &- \delta(t)'\sigma(t)'(\sigma(t)\sigma(t)')^{-1}\sigma(t)\delta(t) + \delta_0(t)^2, \end{aligned}$$

where we abbreviate the right-hand side as $C(t)$ for now. Multiplying both sides with the integrating factor $e^{\int_0^s (\varsigma(u)-2r(u))du}$ yields

$$d\left(\mathbb{E}[X^*(s)^2]e^{\int_0^s (\varsigma(u)-2r(u))du}\right) = C(s)e^{\int_0^s (\varsigma(u)-2r(u))du}.$$

Integrating from 0 to t and inserting the initial condition $\mathbb{E}[X^*(0)^2] = x^2$ yields

$$\mathbb{E}[X^*(t)^2]e^{\int_0^t (\varsigma(u)-2r(u))du} - x^2 = \int_0^t C(s)e^{\int_0^s (\varsigma(u)-2r(u))du} ds.$$

By solving the equation for $\mathbb{E}[X^*(t)^2]$ and resubstituting $C(s)$, we finally obtain (7.30):

$$\mathbb{E}[X^*(t)^2] = x^2 e^{\int_0^t (2r(u)-\varsigma(u))du} + \int_0^t \big(2\kappa(s)\mathbb{E}[X^*(s)] + \varsigma(s)(\vartheta(s)-\gamma)^2 + \delta(s)'\delta(s) - \delta(s)'\sigma(s)'(\sigma(s)\sigma(s)')^{-1}\sigma(s)\delta(s) + \delta_0(s)^2\big) e^{\int_s^t (2r(u)-\varsigma(u))du} ds.$$

To calculate the expression for $\vartheta(t)$, recall that $\varrho(t)$ and $\phi(t)$ are given by the equations

$$\begin{aligned} &\dot{\varrho}(t) + (2r(t) - \varsigma(t))\varrho(t) = 0, \\ &\varrho(T) = \chi, \\ &\varrho(t)\sigma(t)\sigma(t)' > 0 \text{ for a.e. } t \in [0,T], \end{aligned}$$

and

$$\begin{aligned} &\dot{\phi}(t) + (r(t) - \varsigma(t))\phi(t) + (r(t)\gamma + \kappa(t))\varrho(t) = 0, \\ &\phi(T) = 0, \end{aligned}$$

where $\dot{()}$ denotes the derivative w.r.t. t.

First, note that in this case the equation for $\varrho(t)$ simplifies to a linear ODE which can be easily solved by the standard method of multiplying both sides with the integrating factor $e^{\int(\varsigma(s)-2r(s))ds}$, integrating the resulting equation from t to T, using the terminal condition and solving for $\varrho(t)$. This clearly yields

$$\varrho(t) = \chi e^{-\int_t^T (\varsigma(s)-2r(s))ds}.$$

Let $\vartheta(t) := \frac{\phi(t)}{\varrho(t)}$, then by the quotient rule

$$\begin{aligned} \dot{\vartheta}(t) &= \frac{\varrho(t)\dot{\phi}(t) - \phi(t)\dot{\varrho}(t)}{\varrho^2(t)} \\ &= \frac{\varrho(t)(\varsigma(t) - r(t))\phi(t) - \varrho^2(t)(\gamma r(t) + \kappa(t)) - \phi(t)(\varsigma(t) - 2r(t))\varrho(t)}{\varrho^2(t)} \\ &= \frac{\phi(t)\varrho(t)r(t) - \varrho^2(t)(\gamma r(t) + \kappa(t))}{\varrho^2(t)} \\ &= \vartheta(t)r(t) - (\gamma r(t) + \kappa(t)), \\ \vartheta(T) &= 0. \end{aligned}$$

This is again a linear ODE for $\vartheta(t)$ which we can solve by multiplying both sides with $e^{-\int r(s)ds}$, integrating the resulting equation from t to T, inserting the terminal condition and solving for $\vartheta(t)$. This yields (7.25).