2022-09-21

Merton Investment Problem for the Hawkes-based Risk Model

Nova, Mushfika Hossain

UNIVERSITY OF CALGARY

Merton Investment Problem for the Hawkes-based Risk Model

by

Mushfika Hossain Nova

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS

CALGARY, ALBERTA
SEPTEMBER, 2022

Abstract

We study the Merton investment problem in insurance where the risk process is based on the general compound Hawkes process. That means the arrival of claims modeled with a Hawkes process and the modeled claim sizes follow a finite number of fixed jump sizes governed by a Markov chain evolution. The Merton investment problem in insurance is an optimal control problem and we use the dynamic programming method to derive the stochastic Hamilton-Jacobi-Bellman (SHJB) equation satisfied by the value function. The stochastic HJB equation yields a means to obtain the optimal control and thus the optimally controlled stochastic differential equation. Finally, using the claim size from the empirical data set, we simulate the optimal investment portfolio and risk process.

Preface

This thesis is based on joint work with supervisor Dr. Jinniao Qiu and co-supervisor Dr. Anatoliy Swishchuk.

No part of this thesis has been previously published.

Acknowledgements

I would like to express my greatest gratitude to my supervisor Dr. Jinniao Qiu and co-supervisor Dr. Anatoliy Swishchuk for their invaluable patience and feedback. I could not have undertaken this journey without them, who generously provided knowledge, expertise, and motivation during my course work as well. I feel blessed to have worked with them.

I am also grateful to my classmates, especially my office mate Md. Hasib Uddin Molla, for his editing help, late-night feedback sessions, and moral support.

Lastly, I would be remiss in not mentioning my family, especially my mother, spouse, and daughter. Words cannot describe how fortunate I feel to have all of you. Your belief in me has kept my spirits high during this process.

Table of Contents

List of Figures

List of Tables

List of Symbols, Abbreviations, and Nomenclature

Symbol	Definition
HP	Hawkes process
$GCHP$	General compound Hawkes process
$\mathbb{N}_0$	The set of natural numbers including 0
$\mathbb{R}$	The set of real numbers
α	The first letter of the Greek alphabet
β	The second letter of the Greek alphabet
$a.s.$	almost surely
$càdlàg$	Right continuous with left limit
$c.d.f.$	Cumulative distribution function
$p.d.f.$	Probability distribution function
$i.i.d$	Independent and identically distributed
LLN	Law of large numbers
$FCLT$	Functional central limit theorem
MLE	Maximum likelihood estimation
$w.r.t$	With respect to
$w.l.o.g$	Without loss of generality

Chapter 1

Introduction

1.1 Overview

Any investors who invest money in the financial markets aim to maximize their expected returns while limiting the risk of losing money. Potential objects of investment can basically be divided into two categories: risky assets (stocks, derivatives, real estate, raw materials, etc.), which are subject to price volatility, and risk-free assets (bonds and T-bills), which are assets with a beforehand known future return. Normally risky assets have higher expected rate of return compared to risk-free assets. Depending on the degree of risk aversion, an investor will want to compose an investment portfolio as a mix of both risky and risk-free assets to match the level of risk the investor is comfortable with. This problem of finding optimal investment strategy is known as "Merton's portfolio problem". This problem can be modeled as a stochastic optimal control problem. In a stochastic optimal control problem, an investor wants to find the optimal control (optimal investment strategy) so that the criterion function(risk/revenue) is being optimized.

Under the Markovian setting, stochastic optimal control problem leads to the deterministic Hamilton–Jacobi–Bellman (HJB) equation. The HJB equation is a result of the theory of dynamic programming. For an optimal control problem, it is a prerequisite to solve the associated HJB equation to execute optimal control. In general, it is a nonlinear partial differential equation in the optimal value function, which means its solution is the value function itself. Under non-Markovian setting, the stochastic optimal control problem leads to the stochastic Hamilton–Jacobi–Bellman (SHJB) equation, which is a stochastic partial differential equation for optimal value function.

For processes that can be categorized as "Self-exciting", occurrence of each event increases the rate of future occurrence for certain amount of time. For example, each earthquake excites the process in the sense that the chance of aftershocks is increased for some time period after the first earthquake. The same thing happens for gang violence, trade order, or bank defaults. The Hawkes process (HP) is a point process named after Alan G. Hawkes (1971). The HP models the sequence of event occurrence (arrival) for "self-exciting" process. There is a few generalizations of the Hawkes process. General Compound Hawkes process (GCHP) was used to model for the mid-price by Swishchuk and Huffman (2019).

Insurance is a form of risk management which is used to hedge against the risk of a accidental or uncertain loss. An insurer or insurance company or an underwriter provides insurance. A person who buys insurance policy is known as a policyholder, while the person under the policy is called an insured. The insured receives the insurance policy which is a contract explains the conditions and circumstances under which the insurer will compensate the insured, or their designated beneficiary. The amount of money charged by the insurer to the policyholder in exchange for the coverage set forth in the insurance policy is called the premium. If the insured experiences a loss which is potentially covered by the insurance policy, the insured submits a claim to the insurer for processing by a claims adjuster. For example, a health insurance provides for the payments of benefits as a result of sickness or injury. It includes insurance for losses from accident, medical expense, disability, or accidental death and dismemberment. Sometimes a mandatory out-of-pocket expense required by an insurance policy before an insurer will pay a claim is called a deductible which is required by a health insurance policy. The insurer may hedge its own risk by taking out reinsurance, whereby another insurance company agrees to carry some of the risks, especially if the primary insurer assumes the risk is too large for it to carry. Since insurance companies are for-profit enterprise it must have a business model that earn more than it pays out as the form of claims. For this reason insurance companies want to earn revenues from the money collected on policy premiums, minus the money paid as claim payment. Besides this, like other financial institutions, insurance companies also make money by investing into financial markets.

In this research, We aim to study the Merton portfolio optimization problem in insurance for HP based model. In this setup, an insurance company wants to find an optimal investment strategy for their capital $R(t)$ at time t, where capital $R(t)$ and the amount of claims are described by the risk model based on the general compound Hawkes process.

1.2 Literature Review

The point process gained a significant amount of attention during the 1950s and 1960s. Cox introduced the notion of a doubly stochastic Poisson process (called the Cox process now) in [8] and in [3] Bartlett investigated statistical methods for point processes based on their power spectral densities. Lewis in [21] formulated a point process model which was a step in the direction of the Hawkes processes (HP). A nice introduction to the theory of point processes can be found in [5] and [11]. The HP was first introduced in [16]. Laub, Taimre and Pollett provide background, introduce historical development and discuss all major aspects (i.e, parameter estimation, simulation methods) of HP in [19]. The general compound Hawkes process (GCHP) was introduced by Swishchuk in [1] to model the risk process in insurance. For the first time, the general compound Hawkes processes were considered in [33] which explained FCLTs and LLNs for general compound Hawkes processes with dependent orders and regime-switching compound Hawkes process. Application of the risk model based on GCHP to empirical data and optimal investment problems were considered in [36]. In 2017, semi-Markov modeling of limit order books was first considered in [35] to model the mid-price change with two ticks by Swishchuk and Vadori. During the same year, the general semi-Markovian modeling of limit order books with n ticks was considered in [32]. We note that compound Hawkes processes were first applied to limit order books in [34]. [17] considered different types of GCHP and their diffusive limits to model the mid-price of six different stocks. We want to mention some sources which contain many useful results about Hawkes processes in finance. In 2007, Bowsher applied the HP to financial data modeling in [4]. A detailed description of the mathematical theory of Hawkes processes is given by Liniger in 2009 in his Ph.D. thesis [22]. In [7] Álvaro Cartea, Sebastian Jaimungal and Jason Ricci applied HP to model market order arrivals.

The problem where an agent wants to invest the wealth in order to maximize some utility function was first studied by Merton in [24]. The first applications of optimal control in insurance were found in [18], [23], [6]. The paper [31] about Merton investment problems by Swishchuk consists of new Hawkes-based models and the new optimal investment results in finance and insurance for those models.

1.3 Outline of the Thesis

The rest of this thesis is organized as follows. The Hawkes process and related definitions, its properties, and simulation methods are introduced in Chapter 2. Next in Chapter 3, stochastic optimal control problem, dynamic programming and Merton's investment problem in finance are presented. Then the implementation of the dynamic programming method to Merton investment problem for empirical data set containing claim

occurrences is discussed in Chapter 4. Finally, in Chapter 5 we conclude our thesis.

Chapter 2

Hawkes Process

The Hawkes process is a point process that models the self-exciting process and has a clustering effect and long memory. One can think of this as a non-markovian extension of the Poisson process. In this chapter, we will discuss in detail about Hawkes process, its various components and properties, and the simulation methods.

2.1 Basics of Hawkes Process

In this section, before introducing the formal definition of the Hawkes process, we will introduce some related definitions first.

Definition 2.1. Counting Process: A stochastic process $\{N(t), t \geq 0\}$ is said to be a counting process if $N(t)$ is the number of events occurred from time 0 up to and including time t. For a counting process $N(t)$:

- $N(0) = 0$.
- $N(t)$ take values in $\mathbb{N}_0$.
- If $0 \leq s < t$, then $N(t) - N(s)$ is the number of events occurred during the time interval $(s, t]$.

A counting process is almost surely finite and a right continuous step function which gets an increment of size $+1$ with each occurrence of events (arrival). The history of the arrival up to time t is denoted by $\mathcal{H}(t)$, which is a filtration. Another way to characterize such a process is to consider the sequence of random arrival times $\mathbf{T} = \{T_1, T_2, \cdots\}$ at which the counting process $N(\cdot)$ has jumped.

Definition 2.2. Point Process: If a sequence of random variables $\mathbf{T} = \{T_1, T_2, \cdots\}$ taking values in $[0, \infty)$, has $\mathbb{P}(0 \leq T_1 \leq T_2 \leq \cdots) = 1$, and the number of points in a bounded region is almost surely finite,

then **T** is a point process.

The counting and point process are interchangeable. One way to characterize a particular point process is to specify the distribution function of the next arrival time conditional on the past. But it is difficult to work with the conditional arrival distribution. In this case, another characterisation of the point process is used which is called conditional intensity function $\lambda^*(t)$, and can be defined as

$$\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)}, \tag{2.1}$$

where, $F^*(t)$ is the distribution function and $f^*(t)$ is the conditional c.d.f (or p.d.f). Generally, an intuitive representation as the expected rate of arrivals conditioned on $\mathcal{H}(t)$ is preferred over the above definition.

Definition 2.3. Conditional Intensity Function: Consider a counting process $N(t)$ with associated histories $\mathcal{H}(t), t > 0$. If a non-negative function $\lambda^*(t)$ exists such that

$$\lambda^*(t) = lim_{h \downarrow 0} \frac{E[N(t+h) - N(t)|\mathcal{H}(t)]}{h}, \tag{2.2}$$

which only relies on the information of $N(t)$ in the past. Originally, this function was called the hazard function.

By [11], 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 T and thus $\lambda^*(t)$ must be chosen carefully to avoid the event of an explosion, meaning the occurrence of an infinite number of events 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.

Definition 2.4. Compensator: For a counting process $N(t)$, the unique non-decreasing and $\mathcal{H}(t)$ predictable function

$$\Lambda(t) = \int_0^t \lambda^*(s) ds, \tag{2.3}$$

is called a compensator. Here, $\Lambda(0) = 0$ and $N(t) = M(t) + \Lambda(t)$ almost surely for $t > 0$. Also $M(t)$ is an $\mathcal{H}(t)$ local martingale whose existence is guaranteed by the Doob-Meyer decomposition theorem.

In the 1950s and 1960s, when point process get some attention in the field of statistics by Cox and Lewis [9] and Lewis [21], during the same period Bartlett introduced statistical methods for point processes based on their power spectral densities. Then in [16] Hawkes discussed the theoretical properties of a class of processes with particular reference to the point spectrum. The self-exciting point process is a particular result that came from Bartlett's spectral analysis approach by Hawkes. Now we will define the Hawkes process and its choice of different kinds of conditional intensity functions.

Definition 2.5. One-dimensional Hawkes process: Consider a counting process $\{N(t) : t \geq 0\}$, with associated history $\{\mathcal{H}(t) : t \geq 0\}$ which follows

$$\mathbb{P}(N(t+h) - N(t) = m|\mathcal{H}(t)) = \begin{cases} \lambda^*(t)h + o(h), & m = 1 \\ o(h), & m > 1 \\ 1 - \lambda^*(t)h + o(h), & m = 0 \end{cases}, \tag{2.4}$$

where $N(t)$ is characterized by the conditional intensity function with respect to its natural filtration is of the form

$$\lambda^*(t) = \lambda + \int_0^t \mu(t-s)dN(s), \tag{2.5}$$

for $\lambda > 0$ and $\mu : (0, \infty) \to [0, \infty)$ which are called background intensity and excitation function respectively. To avoid the trivial case of being a homogeneous Poisson process assume $\mu(\cdot) \neq 0$. Then the process $N(t)$ is called a Hawkes process.

Notice that the background intensity is constant. Arrivals occur according to λ before any arrivals even happened. If $\lambda = 0$ then the background intensity will no longer carry the process forward. The above form of Hawkes conditional intensity function is just a general form.

2.2 Hawkes Conditional Intensity Function

Let us consider a sequence $\{t_1, t_2, \cdots, t_k\}$ which denotes the sequence of past arrival times of the point process up to time t. Then the most intuitive form of Hawkes conditional intensity can be written as

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

The self-exciting property of the Hawkes process must be specified through the excitation function $\mu(\cdot)$ of the conditional intensity function, that means, if an arrival happens then the conditional intensity function must increase. Thus each new arrival increases the probability of having another new arrival and might leads to an explosion in the process. This behavior causes temporal clustering in $\mathbf{T}$ i.e, a lot of increments in a short period of time. For this reason, $\mu(\cdot)$ should be chosen in a way that avoids the explosion of the number of arrivals which is called stationary processes. An one-dimensional Hawkes process can be interpreted as stationary if the excitation function $\mu(\cdot)$ satisfies $\int_0^\infty \mu(s)ds < 1$ stated in [19].

Here we will introduce two types of excitation functions.

2.2.1 Exponential Decay Excitation Function

The exponential decay intensity function is one of the most common choices for the excitation function and has the form $\mu(t) = \alpha e^{-\beta t}$ first recommended by Hawkes in [16]. It has two constant parameters $\alpha, \beta > 0$. Then equation (2.5) can be expressed as

$$\lambda^*(t) = \lambda + \int_{-\infty}^{t} \alpha e^{-\beta(t-s)} dN(s) = \lambda + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}, \tag{2.6}$$

where the constant α represents the instantaneous increase of the arrival and β interprets the rate of arrival's influence decay over time. Differentiating the above equation gives us

$$d\lambda^*(t) = \beta(\lambda - \lambda^*(t))dt + \alpha dN(t), t \geq 0.$$

Here the initial condition $\lambda^*(0) = \lambda_0 = \lambda$. To solve this SDE we have to apply Itô's Lemma for $f(t, \lambda^*(t)) = e^{\beta t}\lambda^*(t)$ which will lead us

$$\begin{aligned} df(t, \lambda^*(t)) &= f_t dt + f_{\lambda^*} d\lambda^*(t) + \frac{1}{2} f_{\lambda^*\lambda^*} d < \lambda^*(t), \lambda^*(t) > \\ &= \beta e^{\beta t}\lambda^*(t)dt + e^{\beta t}(\alpha dN(t) + \beta(\lambda - \lambda^*(t))dt) + 0 \\ &= \beta e^{\beta t}(\lambda^*(t) + (\lambda - \lambda^*(t)))dt + e^{\beta t}\alpha dN(t) \\ &= \lambda\beta e^{\beta t}dt + e^{\beta t}\alpha dN(t). \end{aligned}$$

Integrating both side of the above equation from 0 to t gives us

$$\begin{aligned} f(t, \lambda^*(t)) &= f(0, \lambda^*(0)) + \int_0^t \lambda\beta e^{\beta u} du + \int_0^t \alpha e^{\beta u} dN(u) \\ &= \lambda_0 + \lambda(e^{\beta t} - 1) + \int_0^t \alpha e^{\beta u} dN(u). \end{aligned}$$

Finally, multiplying both sides with $e^{-\beta t}$ will lead us

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

which is nothing but equation (2.6).

The stationary condition for exponential decay gives us,

$$\int_0^\infty \mu(s)ds = \int_0^\infty \alpha e^{-\beta s} ds = \frac{\alpha}{\beta} < 1,$$
$$\Rightarrow \alpha < \beta.$$

2.2.2 Power-law Excitation Function

Another choice for the excitation function $\mu(\cdot)$ is the power-law function which has the form

$$\mu(t) = \frac{m}{(d + (t))^q}, \tag{2.7}$$

where, the parameters m, d and q are positive. Then conditional intensity function in (2.5) becomes

$$\lambda^*(t) = \lambda + \int_{-\infty}^t \frac{m}{(d + (t-s))^q} dN(s) = \lambda + \sum_{t_i < t} \frac{m}{(d + (t - t_i))^q}. \tag{2.8}$$

This definition of conditional intensity from power law excitation function $\lambda^*(t)$ was applied in the geological model called Omori's law and used to predict the rate of aftershocks caused by an earthquake.

For the power-law function stationary conditions yields that

$$\int_0^\infty \mu(s)ds = \int_0^\infty \frac{m}{(d+s)^q} ds = \frac{md^{1-q}}{q-1} < 1.$$

2.3 Some Properties of Hawkes Process

In this section, we will derive the mean and variance, the law of large numbers, and the central limit theorem for the Hawkes process with exponential excitation function. The main idea comes from [19] and [10].

- **Mean and Variance:**

  Assume, $g(t) = E[\lambda^*(t)]$. By taking definition of $\lambda^*(t)$ from equation (2.6) we can get,

$$g(t) = \lambda + \int_0^t \alpha e^{-\beta(t-u)} E[dN(u)]. \tag{2.9}$$

Now take expectation on the both side of equation (2.2) to get,

$$E[dN(t)] = E[\lambda^*(t)]dt. \tag{2.10}$$

Then by using (2.10) in (2.9) we get,

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

Taking Laplace transformation on both sides of the above equation yields

$$\begin{aligned}\Phi(s) = \mathcal{L}[E(\lambda^*(t))] &= \int_0^\infty [E(\lambda^*(t))]e^{-st}dt = \frac{\lambda}{s} + \frac{\alpha}{s+\beta}\Phi(s) \\ \Rightarrow \Phi(s) &= \frac{\lambda}{\beta-\alpha}(\frac{\beta}{s} - \frac{\alpha}{s+\beta-\alpha}).\end{aligned}$$

After taking inverse Laplace transformation we get,

$$E[\lambda^*(t))] = \frac{\lambda}{\beta-\alpha}(\beta - \alpha e^{-(\beta-\alpha)t}). \tag{2.11}$$

Finally from equation (2.10) we can write,

$$\begin{aligned}E[N(t)] &= \int_0^t E[\lambda^*(u)]du \\ &= \int_0^t [\frac{\lambda}{\beta-\alpha}(\beta - \alpha e^{-(\beta-\alpha)u})]du \\ &= \frac{\lambda\beta t}{\beta-\alpha} + \frac{\alpha\lambda}{(\beta-\alpha)^2}[e^{-(\beta-\alpha)t} - 1],\end{aligned} \tag{2.12}$$

which is the mean or first moment of the Hawkes process.

The idea of covariance came from Hawkes process power spectral density theorem stated in [19]. So the covariance density $R(\tau)$ can be written as,

$$\begin{aligned}R(\tau) &= \overline{\lambda^*}\mu(\tau) + \int_{-\infty}^{\tau} \mu(\tau - v)R(v)dv \\ &= \overline{\lambda^*}\mu(\tau) + \int_0^{\tau} \mu(\tau - v)R(v)dv + \int_0^{\infty} \mu(\tau + v)R(v)dv,\end{aligned} \tag{2.13}$$

where, $\overline{\lambda^*} = E[\lambda^*(t)] = \frac{E[dN(t)]}{dt} = \frac{\lambda\beta}{\beta-\alpha}$. The relation between variance and covariance of the Hawkes

process can be expressed as

$$Var[N(t)] = \overline{\lambda^*}t + 2\int_0^t \int_0^{t_1} R(t_2 - t_1)dt_1 dt_2. \tag{2.14}$$

From equation (2.13) we can calculate covariance, where the equation can be written as

$$R(t) = \frac{\lambda\alpha\beta e^{-\beta t}}{\beta - \alpha} + \alpha e^{-\beta t}\Big[\int_0^t e^{\beta v}R(v)dv + \int_0^\infty e^{-\beta v}R(v)dv\Big].$$

Now by using Laplace and inverse Laplace transformation we can get the covariance,

$$R(\tau) = \frac{\lambda\alpha\beta}{2(\beta - \alpha)^2}(2\beta - \alpha)e^{-(\beta-\alpha)\tau}. \tag{2.15}$$

After that, by putting the value of covariance $R(t)$ in equation (2.14) and completing proper integration we can get the variance of the one-dimensional Hawkes process which is,

$$Var[N(t)] = \frac{\lambda\beta^3 t}{(\beta - \alpha)^3} - \frac{\lambda\alpha\beta}{(\beta - \alpha)^4}(2\beta - \alpha)(1 - e^{-(\beta-\alpha)t}). \tag{2.16}$$

Now we are going to state the LLN and CLT for HP which is described in [2]. The convergences are in the weak sense for the Skorokhod topology.

- **Law of Large Numbers for Hawkes Process:**

  Let $N(t)$ be the one-dimensional Hawkes process which is defined above in (2.5). If the excitation function satisfies $0 < \int_0^\infty \mu(s)ds < 1$ which is the stationary condition then

  $$\lim_{t\to\infty} \frac{N(t)}{t} \to \frac{\lambda\beta}{\beta - \alpha}.$$

  Or, can be written as

  $$\lim_{t\to\infty} \frac{N(t)}{t} \to \frac{\lambda}{1 - \alpha/\beta}. \tag{2.17}$$

  By LLN, $N(t) \approx \frac{\lambda}{1-\alpha/\beta}t$ for large t.

- **Functional Central Limit Theorem for Hawkes Process:**

  If $0 < \int_0^\infty \mu(s)ds < 1$ and $\int_0^\infty s\mu(s)ds < \infty$ then the number of Hawkes process arrivals in $(0, t]$ is asymptotically normally distributed. So for $t \to \infty$

  $$P(\frac{N(t) - \frac{\lambda\beta t}{\beta-\alpha}}{\sqrt{\lambda\beta^3 t/(\beta - \alpha)^3}} < y) \to \Phi(y).$$

Or, can be written as

$$P(\frac{N(t) - \frac{\lambda t}{1-\alpha/\beta}}{\sqrt{\lambda t/(1-\alpha/\beta)^3}} < y) \to \Phi(y). \tag{2.18}$$

where $\Phi(y)$ is the c.d.f of the standard normal distribution.

By FCLT, $N(t) \approx \frac{\lambda}{1-\alpha/\beta}t + \sqrt{\lambda t/(1-\alpha/\beta)^3}W(t)$ for large t. Here, $W(t)$ is a standard Wiener process.

Definition 2.6. Non-linear Hawkes process: Consider a counting process $N(t)$ with conditional intensity function of the form

$$\lambda^*(t) = h\Big(\lambda + \int_0^t \mu(t-s)dN(s)\Big), \tag{2.19}$$

where, $h : \mathbb{R} \to [0, \infty)$ is a nonlinear function. Some frequent choices for h are $h(x) = \mathbf{1}_{x\in\mathbb{R}^+}$ and $h(x) = e^x$.

Linear HPs are much better studied than non-linear HPs. We will not touch on the non-linear case in this research and focus exclusively on the linear form. However, we would refer to [37] for further reading about non-linear HPs.

2.4 Parameter Estimation

For a process realization, the important tool while working with data set is the estimation of the parameter. If the points of the data set are supposed to follow a pattern, then the aim is to find the set of parameters that is able to describe the data at hand properly. Then the goodness of fit method gives a sign of whether the model is suitable or not after checking the obtained parameters with the proposed model. The method we will consider here is maximum likelihood estimation (MLE). This is achieved by maximizing the likelihood function or minimizing the negative likelihood function. If the likelihood function is differentiable, the derivative test for determining maxima can be applied. In some cases, the first-order conditions of the likelihood function can be solved explicitly. Under most circumstances, numerical methods will be necessary to find the maximum likelihood function. The log-likelihood function is used in the same way. Because it replaces multiplication with summation which is convenient for mathematical computation.

In this section, we will discuss how to estimate the parameters of the conditional intensity function of the Hawkes process. Say we want to estimate the parameter θ of conditional intensity function as $\hat{\theta}$ given some finite set of arrival times $\mathbf{t} = \{t_1, t_2, \cdots, t_k\}$, where $\hat{}$ (hat) symbol denotes an estimated value. In the subsequent part of this section we will omit the arguments $\hat{\theta}$ and $\mathbf{t}$ to avoid redundancy which will lead to the following functions: $L = L(\hat{\theta}; \mathbf{t})$, $l = l(\hat{\theta}; \mathbf{t})$, $\lambda^*(t) = \lambda^*(t; \mathbf{t}, \hat{\theta})$, and $\Lambda(t) = \Lambda(t; \mathbf{t}, \hat{\theta})$. The following derivation of likelihood function estimation for the Hawkes process is due to [11, Proposition 7.2.III] .

Let $N(t)$ be a regular point process on the interval $[0, T]$ for some finite positive T and let $(t_1, t_2, \cdots, t_k)$ denote a realization of $N(t)$ during the period. Now, since the likelihood function L of the function $f(t)$ can be described from joint density and it is assumed that the process is observed up to the k-th arrival time, the likelihood function can be written as

$$L = f(t_1, t_2, \cdots, t_k) = \prod_{i=1}^{k} f^*(t_i). \tag{2.20}$$

Therefore, we will isolate $f^*(t)$ and write the above function in terms of conditional intensity function stated in equation(2.1). Also, we know that $f^*(t) = \frac{d}{dt}F^*(t)$. Then rearranging (2.1) will give us

$$\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)} = \frac{\frac{d}{dt}F^*(t)}{1 - F^*(t)} = -\frac{d}{dt}\log(1 - F^*(t)). \tag{2.21}$$

As we will consider the last known event time before t as t_k, integrating both sides over the interval (t_k, t) gives

$$\int_{t_k}^{t} \lambda^*(u)du = \int_{t_k}^{t} -\frac{d}{dt}\log(1 - F^*(t))dt = -\log(1 - F^*(t)) + \log(1 - F^*(t_k)). \tag{2.22}$$

For a simple point process like the Hawkes process, multiple arrivals can not occur at the same time. Hence, $F^*(t_k) = 0$, as $T_{k+1} > t_k$. Therefore

$$\int_{t_k}^{t} \lambda^*(u)du = -\log(1 - F^*(t)). \tag{2.23}$$

After taking exponential on both sides and isolating $F^*(t)$ gives us

$$F^*(t) = 1 - \exp\bigg(-\int_{t_k}^{t} \lambda^*(u)du\bigg), \tag{2.24}$$

and then after differentiating both sides we get

$$\frac{d}{dt}F^*(t) = f^*(t) = 0 - \bigg(\big(-\lambda^*(t)\exp(-\int_{t_k}^{t} \lambda^*(u)du\big)\bigg) = \lambda^*(t)\exp\bigg(-\int_{t_k}^{t} \lambda^*(u)du\bigg). \tag{2.25}$$

Thus by using equation (2.20) and equation (2.25), the likelihood becomes

$$\begin{aligned} L = \prod_{i=1}^{k} f^*(t_i) &= \prod_{i=1}^{k} \lambda^*(t_i) \exp\Big(- \int_{t_{i-1}}^{t_i} \lambda^*(u)du \Big) \\ &= \lambda^*(t_1) \exp\Big(- \int_{t_0}^{t_1} \lambda^*(u)du \Big) \ \lambda^*(t_2) \exp\Big(- \int_{t_1}^{t_2} \lambda^*(u)du \Big) \ \cdots \\ &= \Big[\prod_{i=1}^{k} \lambda^*(t_i) \Big] \exp\Big(- \int_{t_0}^{t_1} \lambda^*(u)du - \int_{t_1}^{t_2} \lambda^*(u)du - \int_{t_2}^{t_3} \lambda^*(u)du \ \cdots \Big) \\ &= \Big[\prod_{i=1}^{k} \lambda^*(t_i) \Big] \exp\Big(- \int_{0}^{t_k} \lambda^*(u)du \Big), \end{aligned} \tag{2.26}$$

as the integral limits are a telescoping series where the lower limit is t_0 and the upper limit is t_k thus we can take the limit starting from 0 and ending at t_k.

Consider the process is observed over some time period $[0, T] \supset [0, t_k]$ which implies that the likelihood will include the possibility of noticing no arrivals in the time interval $(t_k, T]$. So we have

$$L = \Big[\prod_{i=1}^{k} f^*(t_i) \Big] (1 - F^*(T)). \tag{2.27}$$

Using the formula from equation (2.24) to equation (2.27) gives us

$$L = \Big[\prod_{i=1}^{k} \lambda^*(t_i) \Big] \exp\Big(- \int_{0}^{T} \lambda^*(u)du \Big), \tag{2.28}$$

which is the expression for the likelihood function of $N(t)$.

The log-likelihood function l of the function $f(t)$ can be written as

$$\begin{aligned} l &= \log\Big(\prod_{i=1}^{k} f^*(t_i) \Big) = \sum_{i=1}^{k} \log\big(f^*(t_i) \big) \\ &= \sum_{i=1}^{k} \log(\lambda^*(t_i)) - \int_{0}^{T} \lambda^*(u)du \\ &= \sum_{i=1}^{k} \log(\lambda^*(t_i)) - \Lambda(T), \end{aligned} \tag{2.29}$$

where $\Lambda(\cdot)$ is the compensator function defined in (2.3). Now we will derive the log-likelihood function for the exponential decay and power law intensity functions explicitly.

2.4.1 Log-likelihood for Exponential Decay Intensity

For exponential decay the conditional intensity function is $\lambda^*(t) = \lambda + \alpha \sum_{t_i<t} e^{-\beta(t-t_i)}$. So the compensator for interval $[0,T]$ can be separated into segments $[0,t_1], (t_1,t_2], \cdots, (t_{k-1},t_k], (t_k,T]$ and therefore can be evaluated as

$$\begin{aligned}
\Lambda(T) &= \int_0^{t_1} \lambda^*(u)du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda^*(u)du + \int_{t_k}^{T} \lambda^*(u)du \\
&= \int_0^{t_1} \lambda du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda du + \int_{t_k}^{T} \lambda du + \alpha \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \sum_{j=1}^{i} e^{-\beta(u-t_j)}du + \alpha \int_{t_k}^{T} \sum_{j=1}^{k} e^{-\beta(u-t_j)}du \\
&= \int_0^{T} \lambda du + \alpha \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \sum_{j=1}^{i} e^{-\beta(u-t_j)}du + \alpha \int_{t_k}^{T} \sum_{j=1}^{k} e^{-\beta(u-t_j)}du \\
&= \int_0^{T} \lambda du + \alpha \sum_{i=1}^{k-1} \sum_{j=1}^{i} \int_{t_i}^{t_{i+1}} e^{-\beta(u-t_j)}du + \alpha \sum_{j=1}^{k} \int_{t_k}^{T} e^{-\beta(u-t_j)}du.
\end{aligned}$$

After exchanging summation with integration and working through the integral we are left with

$$\Lambda(T) = \lambda T - \frac{\alpha}{\beta} \sum_{i=1}^{k-1} \sum_{j=1}^{i} [e^{-\beta(t_{i+1}-t_j)} - e^{-\beta(t_i-t_j)}] - \frac{\alpha}{\beta} \sum_{j=1}^{k} [e^{-\beta(T-t_j)} - e^{-\beta(t_k-t_j)}].$$

Now if we swap the order of the summation, we can keep j fixed and vary i to k. This way many of the terms will cancel and we will get

$$\begin{aligned}
\Lambda(T) =& \lambda T - \frac{\alpha}{\beta} \sum_{i=1}^{k} [e^{-\beta(t_k-t_i)} - e^{-\beta(t_i-t_i)}] - \frac{\alpha}{\beta} \sum_{i=1}^{k} [e^{-\beta(T-t_i)} - e^{-\beta(t_k-t_i)}] \\
=& \lambda T - \frac{\alpha}{\beta} \sum_{i=1}^{k} [e^{-\beta(T-t_i)} - 1], \qquad (2.30)
\end{aligned}$$

which is our final expression for $\Lambda(T)$. Now substituting $\lambda^*(t)$ and $\Lambda(t)$ from equation (2.30) into equation (2.29) yields

$$l = \sum_{i=1}^{k} \log[\lambda + \alpha \sum_{j=1}^{i-1} e^{-\beta(t_i-t_j)}] - \lambda T + \frac{\alpha}{\beta} \sum_{i=1}^{k} [e^{-\beta(T-t_i)} - 1]. \qquad (2.31)$$

A summation within a summation entails $\mathcal{O}(k)$ complexity. So computing this term of the log-likelihood function is not feasible. For $i = 1, 2, \cdots, k$, $A(i)$ can be defined as

$$\begin{aligned} A(i) =& \sum_{j=1}^{i-1} e^{-\beta(t_i - t_j)} = e^{-\beta t_i + \beta t_{i-1}} \sum_{j=1}^{i-1} e^{-\beta t_{i-1} + \beta t_j} \\ =& e^{-\beta t_i + \beta t_{i-1}} (1 + \sum_{j=1}^{i-2} e^{-\beta(t_{i-1} - t_j)}) \\ =& e^{-\beta(t_i - \beta t_{i-1})} (1 + A(i-1)), \end{aligned} \tag{2.32}$$

with $A(1) = 0$. Finally, the log-likelihood function can be rewritten as

$$l = \sum_{i=1}^{k} \log[\lambda + \alpha A(i)] - \lambda T + \frac{\alpha}{\beta} \sum_{i=1}^{k} [e^{-\beta(T - t_i)} - 1]. \tag{2.33}$$

By using equation(2.32) and (2.33) we can determine estimation for background intensity λ, arrival intensity α and arrival influence decay β. Also, the recursion implies that the joint process $(N(t), \lambda^*(t))$ is Markovian.

2.4.2 Log-likelihood of Power Law Intensity

In this case the intensity function is $\lambda^*(t) = \lambda + \sum_{t_i < t} \frac{m}{(d + (t - t_i))^q}$. So we can compute the compensator for the interval $[0, T]$ as

$$\begin{aligned} \Lambda(T) =& \int_0^{t_1} \lambda^*(u) du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda^*(u) du + \int_{t_k}^{T} \lambda^*(u) du \\ =& \int_0^T \lambda du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \sum_{j=1}^{i} \frac{m}{(d + (t - t_j))^q} du + \int_{t_k}^{T} \sum_{j=1}^{k} \frac{m}{(d + (t - t_j))^q} du. \end{aligned}$$

Rearranging and working through the integral will give us

$$\begin{aligned}
\Lambda(t) =& \lambda T + \sum_{i=1}^{k-1}\sum_{j=1}^{i}\int_{t_i}^{t_{i+1}} \frac{m}{(d+(u-t_j))^q}du + \sum_{j=1}^{k}\int_{t_k}^{T}\frac{m}{(d+(u-t_j))^q}du \\
&= \lambda T + \frac{m}{1-q}\sum_{i=1}^{k-1}\sum_{j=1}^{i}(d+(u-t_j))^{-q+1}\Big|_{u=t_i}^{u=t_{i+1}} + \frac{m}{1-q}\sum_{j=1}^{k}(d+(u-t_j))^{-q+1}\Big|_{u=t_k}^{u=T} \\
&= \lambda T + \frac{m}{1-q}\sum_{i=1}^{k-1}\sum_{j=1}^{i}[(d+(t_{i+1}-t_j))^{-q+1} - (d+(t_i-t_j))^{-q+1}] \\
&\quad + \frac{m}{1-q}\sum_{j=1}^{k}[(d+(T-t_j))^{-q+1} - (d+(t_k-t_j))^{-q+1}] \\
&= \lambda T + \frac{m}{1-q}\sum_{i=1}^{k}[(d+(t_k-t_i))^{-q+1} - (d+(t_i-t_i))^{-q+1}] \\
&\quad + \frac{m}{1-q}\sum_{i=1}^{k}[(d+(T-t_i))^{-q+1} - (d+(t_k-t_i))^{-q+1}] \\
&= \lambda T + \frac{m}{1-q}\sum_{i=1}^{k}[(d+(T-t_i))^{-q+1} - d^{-q+1}].
\end{aligned}$$

Finally using the values of $\lambda^*(t)$ and $\Lambda(t)$ in equation (2.29) gives us the likelihood function

$$l = \sum_{i=1}^{k}\log[\lambda + \sum_{j=1}^{i-1}\frac{m}{(d+(t-t_i))^q}] - \lambda T - \frac{m}{1-q}\sum_{i=1}^{k}[(d+(T-t_i))^{-q+1} - d^{-q+1}]. \tag{2.34}$$

For the above log-likelihood function the summation within summation cannot be avoided. So the computation of estimation will be very slow.

2.5 Simulation Methods

The simulation of the point process is useful in many ways. Simulating a few sample paths of a process shows how the process evolves under different conditions over time. It may also help to estimate values that are impossible to generate analytically. In this section, we will describe two fundamental approaches to producing the realizations of HPs in detail. The first one is a modified thinning algorithm and the second one is called the transformation method. Also, we will present a brief overview of a third approach.

2.5.1 Thinning algorithm

The thinning algorithm is first introduced by Lewis and Shedler in 1979 [20] for the inhomogeneous Poisson process. The standard way of generating an inhomogeneous Poisson process is governed by intensity function

$\lambda(\cdot)$ via thinning. For this algorithm first we need to choose the upper bound where the first process' rate M is always bigger than the intensity $\lambda(\cdot)$ over the interval $[0, T]$. As this method is very fast and generates too many events so some points should be removed probabilistically which can be done by taking a random variable $U_i \sim U(0, M)$ for each candidate point t_i and keeping if the point satisfies $U_i \leq \lambda(t_i)$. These two steps are merging simultaneously which implies each generated point is checked immediately and removed if necessary. Finally, the remaining points satisfy the time-varying intensity $\lambda(\cdot)$.

Hawkes process generation is very similar to inhomogeneous Poisson process generation. In 1981 Ogata [27] adapted it for Hawkes process simulation. The only difference is the conditional intensity $\lambda^*(\cdot)$ does not have an a.s. asymptotic upper bound. The upper bound M^* is adjusted $\lambda^*(t + \epsilon)$, where $\epsilon > 0$ is a very small number which explains that the upper bound will be updated during each simulation and the intensity is always higher just after a new arrival. The rest of the procedure is the same as the former one. This simulation process is called Ogata's modified thinning algorithm and was used by Ogata in 1981 for Hawkes process simulation. Both algorithms are stated below:

Algorithm 1 Generating an inhomogeneous Poisson process by thinning

Input $T, \lambda(\cdot), M$.
Require: $\lambda(\cdot) \leq M$ on $[0, T]$.
$P \leftarrow []$,
$t \leftarrow 0$.
while $t < T$ **do**
  $E \leftarrow Exp(M)$.
  $t \leftarrow t + E$. (Generating next candidate point)
  $U \leftarrow Unif(0, M)$. (keeping it with some probability)
  if $t < T$ **then** and $U \leq \lambda(t)$
    $P \leftarrow [P, t]$.
  end if
end while
return P

In the first figure, an inhomogeneous Poisson process is generated by thinning with intensity function $\lambda(t) = 1 + sint$ and bounded above by $M = 2.2$. In the second figure, with $(\lambda, \alpha, \beta) = (1, 1, 1.1)$, a Hawkes process is generated via thinning. In both figures, *, +, and o signs indicate accepted points, rejected points, and the resulting point processes. Also, each (t, U) point describes a suggested arrival at time t whose U value is in Algorithm 1 and Algorithm 2.

Algorithm 2 Generating a Hawkes process by modified thinning

Input $T, \lambda^*(\cdot), \epsilon$.
Require: $\lambda^*(\cdot)$ non-increasing in period of no arrivals.
$P \leftarrow []$,
$t \leftarrow 0$.
$\epsilon \leftarrow 10^{-100}$ (very small positive value)
while $t < T$ **do**
  $M^* \leftarrow \lambda^*(t + \epsilon)$. (Finding new upper bound)
  $E \leftarrow Exp(M^*)$.
  $t \leftarrow t + E$. (Generating next candidate point)
  $U \leftarrow Unif(0, M^*)$. (Keeping it with some probability)
  if $t < T$ **then** and $U \leq \lambda^*(t)$
    $P \leftarrow [P, t]$.
  end if
end while
return P

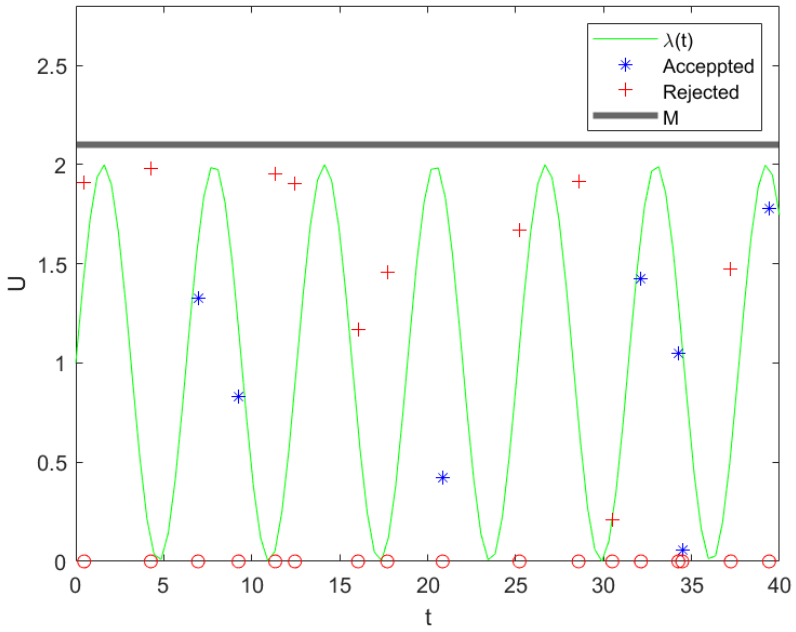

Figure 2.1: Inhomogeneous Poisson process generated by thinning

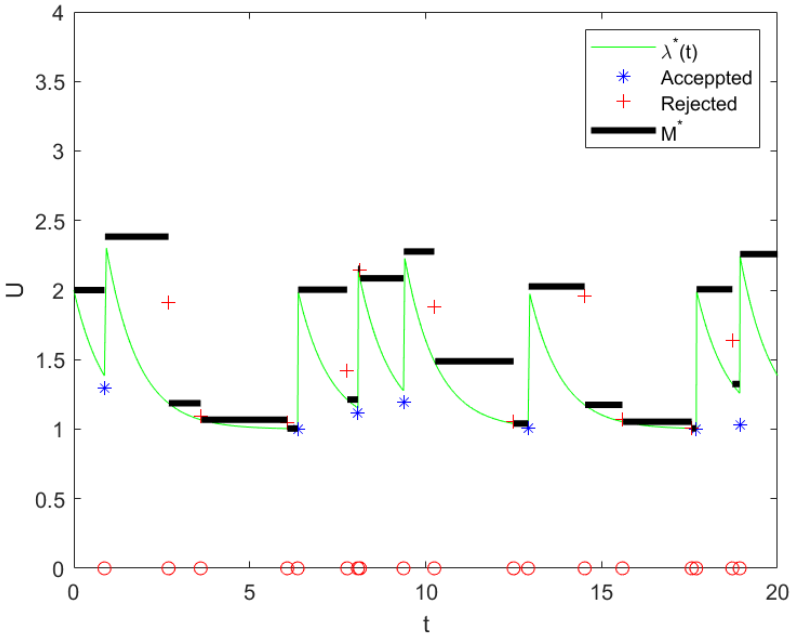

Figure 2.2: Hawkes process generated by thinning

2.5.2 Transformation method

Theorem 2.7. ***(Random time change theorem)*** *Say* $\{t_1, t_2, \cdots, t_k\}$ *is 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 point* $\{\Lambda(t_1), \Lambda(t_2), \cdots, \Lambda(t_k)\}$ *form a Poisson process with unit rate.*

For any general point process, a simulation algorithm relies on the converse of the random change theorem stated above. By using the inverse compensator method, the arrival times for the unit rate Poisson process $\{t_1^*, t_2^*, \cdots\}$ are transformed by the inverse compensator $\Lambda(\cdot)^{-1}$ into any general point process defined by the compensator. The method iteratively solves the equations

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

for realisation of the desired point process $\{t_1, t_2, \cdots\}$. This is described in [15] and Algorithm 7.4.III of [11]. Ozaki in [28] suggested the algorithm for HPs where he did not state explicitly any relation to time changes. Also being a simple point process, multiple arrivals can not occur at the same time for the HP. Hence over the interval (t_k, t) we can write

$$-\int_{t_k}^{t} \lambda^*(u)du = log(1 - F^*(t)),$$

where $F^*(t)$ is the distribution function of the next arrival time. The above equation relates the conditional

c.d.f of the next arrival to the previous history of arrivals $\{t_1, t_2, \cdots, t_k\}$ and the specified conditional intensity function $\lambda^*(t)$. From this relation, we can state that by using the inverse transform method we can easily figure out the next arrival time t_{k+1}. In this process we will take $U \sim Unif[0,1]$ then t_{k+1} can found by solving

$$\int_{t_k}^{t_{k+1}} \lambda^*(u)du = -log(U). \tag{2.35}$$

Case I: Exponential Decay Excitation Function

For an exponentially decaying intensity, the equation (2.35) can be written as

$$\begin{aligned}
&\int_{t_k}^{t_{k+1}} (\lambda + \sum_{i=1}^{k} \alpha e^{-\beta(u-t_i)})du = -log(U) \\
\Rightarrow\ & log(U) + \lambda \int_{t_k}^{t_{k+1}} du + \alpha \sum_{i=1}^{k} \int_{t_k}^{t_{k+1}} e^{-\beta(u-t_i)}du = 0 \\
\Rightarrow\ & log(U) + \lambda(t_{k+1} - t_k) + \alpha \sum_{i=1}^{k} \int_{t_k}^{t_{k+1}} e^{-\beta(u-t_i)}du = 0
\end{aligned}$$

By approximation we get

$$log(U) + \lambda(t_{k+1} - t_k) + \frac{\alpha}{\beta}[e^{-\beta(t_{k+1}-t_k)} - 1] \sum_{i=1}^{k} e^{-\beta(t_k - t_i)} = 0 \tag{2.36}$$

If we put $k = 0$ equation (2.36) will give us: $t_1 = t_0 - \frac{1}{\lambda} log(U)$.

After applying Taylor expansion for each e^{-x}, for $k = 1, 2, 3, 4, \cdots$ we will get,

$$\begin{aligned}
t_2 &= \frac{(\lambda + \alpha)t_1 - log(U)}{\lambda + \alpha}, \\
t_3 &= \frac{[\lambda + \alpha \sum_{i=1}^{2} e^{-\beta(t_2 - t_i)}]t_2 - log(U)}{\lambda + \alpha \sum_{i=1}^{2} e^{-\beta(t_2 - t_i)}}, \\
t_4 &= \frac{[\lambda + \alpha \sum_{i=1}^{3} e^{-\beta(t_3 - t_i)}]t_3 - log(U)}{\lambda + \alpha \sum_{i=1}^{3} e^{-\beta(t_3 - t_i)}}, \\
t_5 &= \frac{[\lambda + \alpha \sum_{i=1}^{4} e^{-\beta(t_4 - t_i)}]t_4 - log(U)}{\lambda + \alpha \sum_{i=1}^{4} e^{-\beta(t_4 - t_i)}}, \cdots
\end{aligned}$$

In this way the $(k+1)$-th term can be written as

$$t_{k+1} = \frac{[\lambda + \alpha \sum_{i=1}^{k} e^{-\beta(t_k - t_i)}]t_k - log(U)}{\lambda + \alpha \sum_{i=1}^{k} e^{-\beta(t_k - t_i)}}.$$

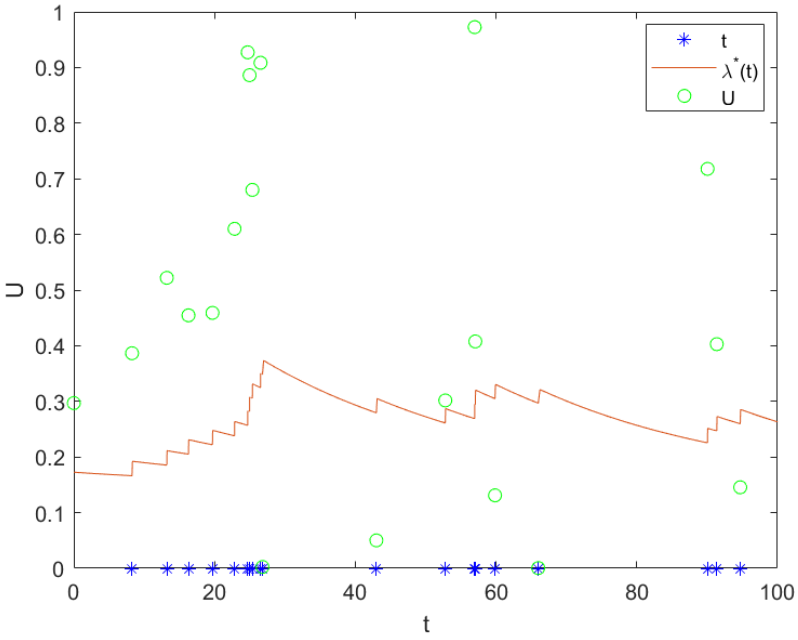

Figure 2.3: HP generated by transformation method (exponential decay intensity)

Case II: Power Law Excitation Function

For power utility function equation (2.35) can be written as

$$\int_{t_k}^{t_{k+1}} (\lambda + \sum_{i=1}^{k} \frac{m}{(d+(u-t_i))^q}) du = -log(U)$$

$$\Rightarrow log(U) + \lambda \int_{t_k}^{t_{k+1}} du + m \sum_{i=1}^{k} \int_{t_k}^{t_{k+1}} (d+(u-t_i))^{-q} du = 0$$

$$\Rightarrow log(U) + \lambda(t_{k+1} - t_k) + m \sum_{i=1}^{k} \int_{t_k}^{t_{k+1}} (d+(u-t_i))^{-q} du = 0$$

By approximation we get

$$log(U) + \lambda(t_{k+1} - t_k) + m \sum_{i=1}^{k} [d+(t_k - t_i)]^{-q} (t_{k+1} - t_k) = 0 \tag{2.37}$$

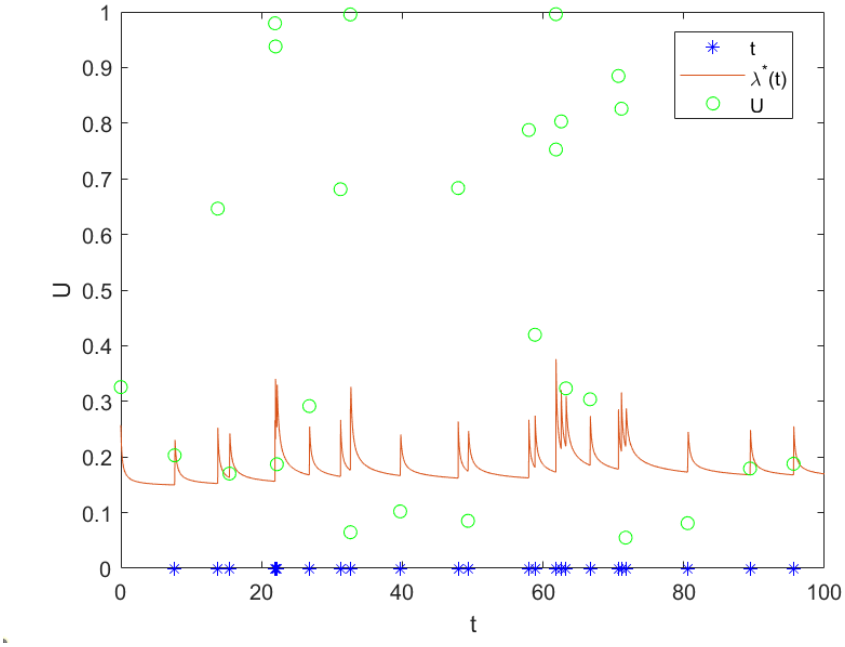

Figure 2.4: HP generated by transformation method (power law intensity)

If we put $k = 0$ equation (2.37) will give us: $t_1 = t_0 - \frac{1}{\lambda} log(U)$.

After applying Taylor expansion for each $(a + x)^{-m}$, for $k = 1, 2, 3, 4, \cdots$ we will get,

$$
\begin{aligned}
t_2 &= \frac{t_1(\lambda + md^{-q}) - log(U)}{\lambda + md^{-q}} \\
t_3 &= \frac{t_2[\lambda + m\sum_{i=1}^{2}(d + t_2 - t_i)^{-q}] - log(U)}{\lambda + m\sum_{i=1}^{2}(d + t_2 - t_i)^{-q}}, \\
t_4 &= \frac{t_3[\lambda + m\sum_{i=1}^{3}(d + t_3 - t_i)^{-q}] - log(U)}{\lambda + m\sum_{i=1}^{3}(d + t_3 - t_i)^{-q}}, \\
t_5 &= \frac{t_4[\lambda + m\sum_{i=1}^{4}(d + t_4 - t_i)^{-q}] - log(U)}{\lambda + m\sum_{i=1}^{4}(d + t_4 - t_i)^{-q}}, \cdots
\end{aligned}
$$

In this way the $(k + 1)$-th term can be written as

$$
t_{k+1} = \frac{t_k[\lambda + m\sum_{i=1}^{k}(d + t_k - t_i)^{-q}] - log(U)}{\lambda + m\sum_{i=1}^{k}(d + t_k - t_i)^{-q}}.
$$

Another simulation method is based on the clustering representation of HP. Over the interval $[0, T]$, immigrants form a homogeneous Poisson process of rate λ. Their arrival times $C_1, C_2, \cdots, C_k$ are uniform i.i.d. random variables, while the number of immigrants is distributed $Poi(\lambda T)$. On the other hand, descendants of each immigrant form an inhomogeneous Poisson process and are distributed $Poi(\alpha/\beta)$. At time C_i the

arrival intensity of descendants is $\mu(t - C_i)$ for $t > C_i$. Then at times $(C_i + E_1, C_i + E_2, \cdots, C_i + E_{D_i})$ the descendants of the i-th immigrant arrived. The algorithm for the superposition of Poisson processes will be stated below.

There are several methods that are stated in Laub, Taimre, and Pollett. Dassion and Zhao in [12] and

Algorithm 3 Generating a Hawkes process by clusters

$P \leftarrow \{\}$,
$k \leftarrow Poi(\lambda T)$. (number of immigrants)
for $i = 1, 2, \cdots, k$,$C_i \leftarrow Unif(0, T)$ which is i.i.d. (arrival times of immigrants)
for $i = 1, 2, \cdots, k$,$D_i \leftarrow Poi(\alpha/\beta)$ which is i.i.d. (number of descendants)
for $i \leftarrow 1$ to k **do**
  if $D_i > 0$ **then**
    $E_i \leftarrow Exp(\beta)$ which is i.i.d. for $i = 1, 2, \cdots, k$.
    $P \leftarrow P \cup \{C_i + E_1, \cdots, C_i + E_{D_i}\}$
  end if
end for
$P \leftarrow \{P_i : P_i \in P, P_i \leq T\}$ (removing descendants outside $[0, T]$).
$P \leftarrow Sort(P \cup \{C_1, C_2, \cdots, C_k\})$. (adding and sorting immigrants)
return P

Moller and Rasmussen in [25] gave an alternative to the methods which are not stated above.
The data sets gathered from the simulation of point processes are applied to check the functionality of methods to be used on empirical sets of data which is important for the rest thesis work.

2.6 General Compound Hawkes Process

In this section, we will define the general compound Hawkes process with different state-dependent order [1].

Definition 2.8. General Compound Hawkes Process with N-state Dependent Orders:
Suppose that X_k is an ergodic continuous-time Markov chain, independent of $N(t)$, with state space $X := \{1, 2, 3, \cdots, n\}$, $N(t)$ is a one-dimensional Hawkes process and $a(x)$ is any bounded and continuous function on X. The General Compound Hawkes Process with N-state Dependent Orders can be expressed as:

$$S_t = S_0 + \sum_{k=1}^{N(t)} a(X_k). \tag{2.38}$$

Definition 2.9. General Compound Hawkes Process with Two-state Dependent Orders:
Suppose that X_k is an ergodic continuous-time Markov chain, independent of $N(t)$ with two states $\{1, 2\}$. Then, equation (2.38) becomes

$$S_t = S_0 + \sum_{k=1}^{N(t)} a(X_k), \tag{2.39}$$

where, $a(X_k)$ takes only the values $a(1)$ and $a(2)$.

Definition 2.10. General Compound Hawkes Process with Dependent Orders:
Suppose that, $X_k \in \{-\delta, \delta\}$ and that $a(x) = x$, then equation (2.39) becomes

$$S_t = S_0 + \sum_{k=1}^{N(t)} X_k. \tag{2.40}$$

This type of process can be modeled for the mid-price which is defined as the average of best ask and best bid prices. Here, δ is a fixed tick size and $N(t)$ is the number of arrivals up to time t. We can call this process General Compound Hawkes Process with Dependent Orders. This is generalization of the previous process obtained by letting $a(1) = -\delta$ and $a(2) = \delta$.

Filimonov and Sornette [13] had worked with a data set with timestamps accurate to a second and sometimes this led to multiple arrivals nominally at the same time as order arrivals and cancellations are very frequent. With many applications occurring at the millisecond time scale, Lorenzen formed a measure of high-frequency trading activity leading to an interesting correlation between this activity and n over the observed period where n can be 1000, 10000 making nt large with respect to t and then we can consider the scale nt instead of t.

Chapter 3

Merton Investment Problems in Insurance under a Hawkes Process-based Setting

Merton's portfolio problem is a well-known problem in continuous-time finance and insurance. The problem was formulated and solved by Robert C. Merton in 1969 both for a finite lifetime and for the infinite case. Research has continued to extend and generalize the model to include factors like transaction costs and bankruptcy. In this chapter, first we will discuss stochastic optimal control problem and the dynamic programming method, and the Merton investment problem in finance and insurance. Finally we use the dynamic programming method to obtain the optimal investment portfolio in a Hawkes process setting.

3.1 Stochastic Optimal Control Problem

Let us consider a controlled stochastic differential equation of the form:

$$\begin{cases} dR(t) &= b\left(t, R(t), \pi(t)\right) dt + f\left(t, R(t), \pi(t)\right) dW(t),\ 0 \leq t \leq T, \\ R(0) &= x_0, \end{cases} \tag{3.1}$$

with the initial value $x_0 \in \mathbb{R}^n$, and the functions

$$b : \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n,$$
$$f : \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^{n \times d}.$$

Here the n-dimensional process R is called the state process which is controlled by the k-dimensional control process π. W is a d-dimensional Brownian motion. Also consider the value function

$$J(\pi(\cdot)) = E\bigg[\int_0^T F(t, R(t), \pi(t))dt + U(R(T))\bigg], \tag{3.2}$$

where F is the instantaneous utility function and U is the utility function for terminal time. Now we can formally state the stochastic optimal control problem as:

$$\max_{\pi} E\bigg[\int_0^T F(t, R^\pi(t), \pi(t))dt + U(R^\pi(T))\bigg], \tag{3.3}$$

where R^π is the unique solution of the following SDE:

$$\begin{cases} dR^\pi(t) &= b\left(t, R^\pi(t), \pi(t)\right) dt + f\left(t, R^\pi(t), \pi(t)\right) dW(t),\ 0 \le t \le T, \\ R^\pi(0) &= x_0. \end{cases} \tag{3.4}$$

Here the notation R^π refers to the dependence of the state process R on the control process π. Another notation, $R^{\pi;t,x}$ for the state process R will indicate further that the initial time for the state process is $t; 0 \le t \le T$ with the initial state x. With this notation, $R^\pi = R^{\pi;0,x_0}$.

For any given adapted π, if the SDE (3.4) admits a unique solution, the process π is called an admissible control. If

$$J(\pi^*(\cdot)) = \max_{\pi} J(\pi(\cdot)),$$

then π^* is the optimal control and R^{π^*} is the optimal state.

3.1.1 Dynamic Programming (DP) Method and HJB Equation

For a stochastic optimal control problem, two questions arise naturally. The first one is whether there exists an optimal control law or not and the second one is if the optimal control exists then what is the way to find it. To deal with the second concern, the dynamic programming method is used very frequently. The main

idea is to break down the main problem into sub-problems in a recursive manner. This is helpful in the sense that the original complex problems may result in simpler sub-problems. Then tie all these sub-problems with the so-called Hamilton-Jacobi-Bellman (HJB) equation which is a nonlinear second-order partial differential equation of the optimal value function of the main optimal control problem. The stochastic optimal control problem is then equivalent to finding the solution of the HJB equation which we may solve for the optimal value function and the optimal feedback control that is indexed by the initial time $t \in [0, T]$ and the initial state $x \in \mathbb{R}^n$.

For example, let us consider the optimal control problem (3.4) − (3.3). Now if we denote the value function by V which is indexed by the initial time $t \in [0, T]$ and the initial state $x \in \mathbb{R}^n$, then

$$V(t,x) = E\bigg[\int_t^T F(s, R^{\pi;t,x}(s), \pi(s))ds + U(R^{\pi;t,x}(T))\bigg], \tag{3.5}$$

where $R^{\pi;t,x}$ is the unique solution of the following stochastic differential equation:

$$\begin{cases} dR^{\pi;t,x}(t) &= b\left(t, R^{\pi;t,x}(t), \pi(t)\right) dt + f\left(t, R^{\pi;t,x}(t), \pi(t)\right) dW(t),\ 0 \le t \le T, \\ R^{\pi;t,x}(t) &= x. \end{cases} \tag{3.6}$$

Then under the Markovian setting, the value function $V(t,x)$ satisfies the partial differential equation

$$\begin{cases} -\frac{\partial V}{\partial t} &= \sup_\pi LV(t,x), \\ V(T,x) &= U(x), \end{cases} \tag{3.7}$$

know as the Hamilton-Jacobi-Bellman (HJB) equation, where the operator L is defined as:

$$LV(t,x) := \bigg(\sum_{i=1}^n b_i(t,x,\pi)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^n f_{ij}(t,x,\pi)\frac{\partial^2}{\partial x_i \partial x_j}\bigg)V(t,x), \tag{3.8}$$

and known as the infinitesimal generator for the Markov state process $R^{\pi,t,x}$ defined in (3.6).

3.2 Merton Investment Problem

3.2.1 Merton Investment under Black-Scholes model

Consider an economic agent over a fixed time interval $[0, T]$ where the agent has the limited choice of investing his/her wealth in only two different assets: a risky asset and a risk-free asset. At the time $t = 0$ the agent is endowed with initial wealth x_0 and Merton's goal is to determine how the agent should allocate and

reallocate the wealth at each time point in order to maximize the expected utility of the terminal wealth $R(T)$ at horizon T. Suppose the price process $S(t)$ for the risky asset follows geometric Brownian motion and the risk-free asset $B(t)$ has interest rate r, then they have the dynamics:

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sigma S(t)dW(t), \\ dB(t) &= rB(t)dt, \end{aligned}$$

where $(W(t))_{t\geq 0}$ is a standard Brownian motion, μ is the expected rate of return and σ is the volatility of the stock. Assume the amount of wealth invested in risky and risk-free assets are $n_S(t)$ and $n_B(t)$ respectively at time t. Then, after restricting the agent's investment strategies to be self-financing the dynamics of the total wealth R of the agent may be given by

$$\begin{aligned} dR(t) &= n_B(t)\frac{dB(t)}{B(t)} + n_S(t)\frac{dS(t)}{S(t)} \\ &= n_B(t)rdt + n_S(t)(\mu dt + \sigma dW(t)). \end{aligned} \tag{3.9}$$

Let $\pi(t) = n_s(t)/R(t)$ be relative portfolio weights of wealth invested in the risky asset at time t. Then equation (3.9) can be rewritten as

$$\begin{aligned} dR(t) &= \frac{n_B(t)}{R(t)}R(t)rdt + \frac{n_S(t)}{R(t)}R(t)(\mu dt + \sigma dW(t)) \\ &= (1-\pi(t))R(t)rdt + \pi(t)R(t)(\mu dt + \sigma dW(t)) \\ &= R(t)[(1-\pi(t))rdt + \pi(t)\mu dt + \pi(t)\sigma dW(t)]. \end{aligned} \tag{3.10}$$

Finally, the associated stochastic optimal control problem for the Merton investment problem becomes:

$$\max_{\pi} E[U(R^{\pi}(T))], \tag{3.11}$$

subject to,

$$\begin{cases} dR^{\pi}(t) &= R^{\pi}(t)\left[(1-\pi(t))rdt + \pi(t)\mu dt + \pi(t)\sigma dW(t)\right], \\ R^{\pi}(0) &= x_0, \end{cases} \tag{3.12}$$

where $U(\cdot)$ is the utility function and x_0 is the initial wealth.

Now we can explain Merton's investment problem in finance for the Hawkes process-based model. Here,

the risk-free asset follows the same dynamics as before, but the risky asset follows the dynamics, $dS(t) = S(t)dG(t)$. In this case, the model for the stock price (risky asset) $S(t)$ is based on the general compound Hawkes process instead of geometric Brownian motion. If $n(t) = (n_B(t), n_S(t))$ is the investor portfolio and $\pi(t) = n_s(t)/R(t)$ is same as before and the portfolio is admissible, i.e., $R(t) \geq 0$ almost surely, and self-financing, i.e., there is no exogenous infusion or withdrawal of money, then the associated stochastic control problem becomes

$$\max_{\pi} E[U(R^{\pi}(T))], \tag{3.13}$$

subject to,

$$\begin{cases} dR^{\pi}(t) = R^{\pi}(t)\big[(1-\pi(t))rdt + \pi(t)dG(t)\big], \\ R^{\pi}(0) = x_0. \end{cases} \tag{3.14}$$

Note that the problem (3.13)-(3.14) will not be addressed in this thesis.

3.2.2 Merton Investment problem in Insurance under a Hawkes Process-based Setting

When an insurance company makes a business arrangement with an individual or company or organization, they promise to pay a certain amount of money (claim) to the insured for a specific asset loss. In exchange, the insured makes regular payments (premium) to the insurance company.

Suppose, u is the initial capital for an insurance company and $c > 0$ is constant premium intensity, then the capital $R(t)$ at any time $t > 0$ for the insurance company is modeled in [31] based on GCHP as

$$R(t) = u + ct - \sum_{i=1}^{N(t)} a(X_i), \tag{3.15}$$

where X_i is a finite state Markov chain, $a(x)$ is a continuous and bounded function on $X = \{1, 2, ..., N\}$-space state for X_i, $a(X_i)$ are claim sizes and $N(t)$ is a one-dimensional Hawkes process. The process $R(t)$ is also called the risk process.

To increase their revenues, insurance companies also invest their capital into financial markets. Let $A(t)$ be the amount invested in a risky asset and suppose that the price process $S(t)$ for the risky asset follows geometric Brownian motion, then

$$dS(t) = S(t)(\mu dt + \sigma dW(t)). \tag{3.16}$$

Now, if the rest positive amount $R(t) - A(t)$ is invested in the risk-free asset (bank account, bonds, etc.) at

the deterministic short interest rate $r > 0$, then we have

$$d(R(t) - A(t)) = r(R(t) - A(t))dt. \tag{3.17}$$

Set

$$G(t) = \sum_{i=1}^{N(t)} a(X_i), \quad t \geq 0,$$

and then the position $R(t)$ of the insurer at any time $t > 0$ has the dynamics

$$\begin{aligned} dR(t) &= A(t)(\mu dt + \sigma dW(t)) + r(R(t) - A(t))dt + cdt - dG(t) \\ &= rR(t)dt + A(t)(\mu dt + \sigma dW(t)) - rA(t)dt + cdt - dG(t). \end{aligned} \tag{3.18}$$

Next, we let $\pi(t) = A(t)/R(t)$ be the fraction of the total wealth $R(t)$ invested in the risky assets.

At this stage let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a complete filtered probability space with the filtration $(\mathscr{F}_t)_{t\in[0,T]}$ being the augmented filtration generated by the 1-dimensional Wiener process W and the process G. Denote by $(\mathscr{F}_t^G)_{t\in[0,T]}$ the augmented filtration generated by the process G. Throughout this thesis, the control/policy process $(\pi(t))_{t\in[0,T]}$ is said to be admissible if it is an $(\mathscr{F}_t)_{t\in[0,T]}$-adapted real-valued process such that the stochastic differential equation (SDE) (3.21) admits a unique (strong) solution $(R^\pi(t))_{t\in[0,T]}$ in $L^2(\Omega\times[0,T])$ and we denote by $\mathcal{A}$ the set of all the admissible controls/policies.

With the notation $\pi(t) = A(t)/R(t)$, we can rewrite the equation for $R(t)$, the position of the insurer at time $t > 0$, as

$$\begin{aligned} dR^\pi(t) &= rR^\pi(t)dt + A(t)(\mu dt + \sigma dW(t)) - rA(t)dt + cdt - dG(t) \\ &= rR^\pi(t)dt + A(t)[(\mu - r)dt + \sigma dW(t)] + cdt - dG(t) \\ &= R^\pi(t)\left[r + (\mu - r)\pi(t)\right]dt + R^\pi(t)\pi(t)\sigma dW(t) + cdt - dG(t). \end{aligned} \tag{3.19}$$

Thus the utility maximization of Merton's investment problem for the insurance company becomes

$$\max_{\pi\in\mathcal{A}} E\left[U(R^\pi(T))\right], \tag{3.20}$$

subject to,

$$\begin{cases} dR^{\pi}(t) &= R^{\pi}(t)\left[r + (\mu - r)\pi(t)\right]dt + R^{\pi}(t)\pi(t)\sigma dW(t) + cdt - dG(t), \\ R^{\pi}(0) &= u. \end{cases} \tag{3.21}$$

Further using the basic properties of derivatives we can have

$$\begin{aligned} d\left(e^{-rt}R^{\pi}(t)\right) &= e^{-rt}dR^{\pi}(t) - re^{-rt}R^{\pi}(t)dt \\ &= e^{-rt}\left\{R^{\pi}(t)\left[r + (\mu - r)\pi(t)\right]dt + R^{\pi}(t)\pi(t)\sigma dW(t) + cdt - dG(t)\right\} - re^{-rt}R^{\pi}(t)dt \\ &= e^{-rt}R^{\pi}(t)(\mu - r)\pi(t)dt + e^{-rt}R^{\pi}(t)\pi(t)\sigma dW(t) + ce^{-rt}dt - e^{-rt}dG(t). \end{aligned}$$

Then if we set

$$\begin{aligned} H(t) &= \frac{c}{r}\left(1 - e^{-rt}\right) - \int_0^t e^{-rs}dG(s), \\ \widetilde{R}^{\pi}(t) &= e^{-rt}R^{\pi}(t) - H(t), \end{aligned}$$

the utility maximization of Merton's investment problem for the insurance company can also be written as

$$\max_{\pi \in \mathcal{A}} E\left[U\left(e^{rT}(\widetilde{R}^{\pi}(T) + H(T))\right)\right], \tag{3.22}$$

subject to,

$$\begin{cases} d\widetilde{R}^{\pi}(t) &= (\widetilde{R}^{\pi}(t) + H(t))(\mu - r)\pi(t)dt + (\widetilde{R}^{\pi}(t) + H(t))\pi(t)\sigma dW(t), \quad t \in [0, T]; \\ \widetilde{R}^{\pi}(0) &= u. \end{cases} \tag{3.23}$$

The stochastic optimal control problem $(3.22) - (3.23)$ is just the transformed version of $(3.20) - (3.21)$, and it will be addressed in this thesis.

3.3 Optimal Investment Portfolio with DP Method

To find the optimal investment portfolio with dynamic programming method, in the next part of this section we will consider the stochastic optimal control problem $(3.22) - (3.23)$ for the Merton investment problem in insurance with G being a GCHP.

3.3.1 Stochastic HJB Equation of the Value Function

For our optimal control problem, define the value function

$$\widetilde{V}(t,x) = esssup_{\pi\in\mathcal{A}} E\left[U\left(e^{rT}(\widetilde{R}^{\pi;t,x}(T)+H(T))\right)\big|\mathscr{F}_t^G\right], \quad (t,x)\in[0,T]\times\mathbb{R}. \tag{3.24}$$

Inspired by the dynamic programming principle (see [29, 30] for instance), one may expect the value function to satisfy the following stochastic Hamilton-Jacobi-Bellman (SHJB) equation:

$$\begin{aligned}
-d\widetilde{V}(t,x) =& \max_{\pi\in\mathbb{R}}\left\{\frac{1}{2}(x+H(t))^2\pi^2\sigma^2 D_{xx}\widetilde{V} + (x+H(t))(\mu-r)\pi D_x\widetilde{V}\right\}(t,x)dt - d\widetilde{M}(t,x)\\
=& \max_{\pi\in\mathbb{R}}\left\{\frac{1}{2}\sigma^2(x+H(t))^2 D_{xx}\widetilde{V}\left[\pi^2 + 2\pi\frac{(\mu-r)D_x\widetilde{V}}{\sigma^2(x+H(t))D_{xx}\widetilde{V}}\right]\right\}(t,x)dt - d\widetilde{M}(t,x)\\
=& \max_{\pi\in\mathbb{R}}\left\{\frac{1}{2}\sigma^2(x+H(t))^2 D_{xx}\widetilde{V}\left(\pi + \frac{(\mu-r)D_x\widetilde{V}}{\sigma^2(x+H(t))D_{xx}\widetilde{V}}\right)^2 - \frac{(\mu-r)^2|D_x\widetilde{V}|^2}{2\sigma^2 D_{xx}\widetilde{V}}\right\}(t,x)dt - d\widetilde{M}(t,x),\\
=& -\frac{(\mu-r)^2|D_x\widetilde{V}(t,x)|^2}{2\sigma^2 D_{xx}\widetilde{V}(t,x)}dt - d\widetilde{M}(t,x),
\end{aligned} \tag{3.25}$$

with $\widetilde{V}(T,x) = U(\kappa_T(x+H(T)))$ where $\kappa_T = e^{rT}$ and the pair $(\widetilde{V}(t,x),\ \widetilde{M}(t,x))$ is unknown with $\left(\widetilde{M}(t,x)\right)_{t\in[0,T]}$ being an $(\mathscr{F}_t^G)_{t\geq 0}$-adapted martingale parameterized by $x\in\mathbb{R}$. Observing that the above equation does not depend on the unknown random field $\widetilde{M}(t,x)$, it can be equivalently written as

$$\begin{aligned}
\widetilde{V}(s,x) = E\Big[&U(\kappa_T(x+H(T)))\\
&+\int_s^T \max_{\pi\in\mathbb{R}}\left\{\frac{1}{2}(x+H(t))^2\pi^2\sigma^2 D_{xx}\widetilde{V} + (x+H(t))(\mu-r)\pi D_x\widetilde{V}\right\}(t,x)\,dt\Big|\mathscr{F}_s^G\Big],
\end{aligned} \tag{3.26}$$

for $s\in[0,T]$.

Lemma 3.1. *Let $u(t,x)$, $(t,x)\in[0,T]\times\mathbb{R}$ be a random field satisfying:*

(1) For each $x\in\mathbb{R}$, $u(t,x)$ is càdlàg (right continuous with left limits) in time t a.s. and $(u(t,\cdot))_{t\in[0,T]}$ is a $C_b^2(\mathbb{R};\mathbb{R})$-valued $(\mathscr{F}_t^G)_{t\in[0,T]}$-adapted process with

$$E\left[\sup_{(t,x)\in[0,T]\times\mathbb{R}}|u(t,x)|^2\right] < \infty;$$

(2) For each $(t,x)\in[0,T]\times\mathbb{R}$, it holds that

$$u(t,x) = E\left[u(T,x) + \int_t^T f(s,x)\,ds\big|\mathscr{F}_t^G\right], \quad a.s.,$$

where the process $(f(s,\cdot))_{s\in[0,T]}$ is valued in $C_b(\mathbb{R};\mathbb{R})$ and càdlàg (right continuous with left limits) in time s a.s. and for each $x\in\mathbb{R}$, $(f(s,x))_{s\in[0,T]}$ is $(\mathscr{F}_t^G)_{t\in[0,T]}$-adapted process lying in $L^2(\Omega\times[0,T];\mathbb{R})$.

Then it holds that for each $t\in[0,T]$,

$$u(t,\widetilde{R}^\pi(t)) = E\Big[u(T,\widetilde{R}^\pi(T)) - \int_t^T \Big\{\frac{1}{2}(\widetilde{R}^\pi(s)+H(s))^2\pi^2\sigma^2 D_{xx}u(s,\widetilde{R}^\pi(s)) + (\widetilde{R}^\pi(s)+H(t))(\mu-r)\pi D_x u(s,\widetilde{R}^\pi(s)) - f(s,\widetilde{R}^\pi(s))\Big\}\,ds\big|\mathscr{F}_t^G\Big], \quad a.s. \tag{3.27}$$

Proof. W.l.o.g., we only prove (3.27) for the case when $t=0$. For each $J\in\mathbb{N}^+$ with $J>2$, let $t_i=\frac{iT}{J}$ for $i=0,1,\cdots,J$. Then, we get a partition of $[0,T]$ with $0=t_0<t_1<\cdots<t_{N-1}<t_N=T$.

For each $i\in\{0,\cdots,J-1\}$,

$$\begin{aligned} E\Big[u(t_{i+1},\widetilde{R}^\pi(t_{i+1})) - u(t_i,\widetilde{R}^\pi(t_i))\Big] &= E\Big[u(t_{i+1},\widetilde{R}^\pi(t_i)) - u(t_i,\widetilde{R}^\pi(t_i)) + u(t_{i+1},\widetilde{R}^\pi(t_{i+1})) - u(t_{i+1},\widetilde{R}^\pi(t_i))\Big] \\ &:= I_1^i + I_2^i, \end{aligned}$$

where

$$I_1^i = E\Big[u(t_{i+1},\widetilde{R}^\pi(t_i)) - u(t_i,\widetilde{R}^\pi(t_i))\Big] = -E\left[\int_{t_i}^{t_{i+1}} f(s,\widetilde{R}^\pi(t_i))\,ds\right],$$

and by the integration by parts formula

$$\begin{aligned} I_2^i &= E\Big[u(t_{i+1},\widetilde{R}^\pi(t_{i+1})) - u(t_{i+1},\widetilde{R}^\pi(t_i))\Big] \\ &= E\Big[\int_{t_i}^{t_{i+1}} \Big\{\frac{1}{2}(\widetilde{R}^\pi(s)+H(s))^2\pi^2\sigma^2 D_{xx}u(t_{i+1},\widetilde{R}^\pi(s)) + (\widetilde{R}^\pi(s)+H(t))(\mu-r)\pi D_x u(t_{i+1},\widetilde{R}^\pi(s))\Big\}\,ds\Big]. \end{aligned}$$

Recall that

$$E\Big[u(T,\widetilde{R}^\pi(T)) - u(0,\widetilde{R}^\pi(0))\Big] = \sum_{i=0}^{N-1} I_1^i + I_2^i. \tag{3.28}$$

Then, the dominated convergence, the time continuity of $\widetilde{R}^\pi$, and the right-continuity of $u(t,x)$ in time t imply that the expected sum of Lebesgue integrals converges. Finally, we have

$$\begin{aligned} &E\Big[u(T,\widetilde{R}^\pi(T)) - u(0,\widetilde{R}^\pi(0))\Big] \\ &= E\Big[\int_0^T \Big\{\frac{1}{2}(\widetilde{R}^\pi(s)+H(s))^2\pi^2\sigma^2 D_{xx}u(s,\widetilde{R}^\pi(s)) + (\widetilde{R}^\pi(s)+H(t))(\mu-r)\pi D_x u(s,\widetilde{R}^\pi(s)) - f(s,\widetilde{R}^\pi(s))\Big\}\,ds\Big]. \end{aligned}$$

Analogously, we may prove (3.27) for all $t \in [0, T]$. □

3.3.2 Finding Optimal Value Function and Optimal Control

Here, we consider the exponential utility function $U(x) = 1 - e^{-px}$ for some $p > 0$. In view of such an exponential structure, a standard argument suggests a multiplicative decomposition of the value function which gives

$$\widetilde{V}(t,x) = 1 - e^{-p\kappa_T x} Y(t),$$

with the stochastic process Y satisfying

$$Y(t) = E\Big[e^{-p\kappa_T H(T)} - \int_t^T \frac{(\mu - r)^2}{2\sigma^2} Y(s)\, ds \Big| \mathscr{F}_t^G\Big], \quad t \in [0, T]. \tag{3.29}$$

Solving the linear BSDE (3.29) gives

$$Y(t) = E\left[e^{-p\kappa_T H(T) - \frac{(\mu-r)^2(T-t)}{2\sigma^2}} \big| \mathscr{F}_t^G\right], \quad t \in [0, T].$$

Here, we have, $D_x\widetilde{V} = p\kappa_T e^{-p\kappa_T x} Y(t)$ and $D_{xx}\widetilde{V} = -p^2\kappa_T^2 e^{-p\kappa_T x} Y(t)$. Also from the last line of equation (3.25) we get maximum by taking $\pi + \frac{(\mu-r)D_x\widetilde{V}}{\sigma^2(x+H(t))D_{xx}\widetilde{V}} = 0$. Then, putting the values of $D_x\widetilde{V}$ and $D_{xx}\widetilde{V}$ in that equation we obtain the optimal control:

$$\begin{aligned} \pi^*(t,x) &= -\frac{(\mu - r) D_x \widetilde{V}(t,x)}{\sigma^2 (x + H(t)) D_{xx} \widetilde{V}(t,x)} \\ &= \frac{(\mu - r)}{\sigma^2 (x + H(t)) p \kappa_T}. \end{aligned} \tag{3.30}$$

Theorem 3.2. *When $U(x) = 1 - e^{-px}$ for $p > 0$, the value function defined in* (3.24) *has the expression:*

$$\widetilde{V}(t,x) = 1 - e^{-p\kappa_T x} E\left[e^{-p\kappa_T H(T) - \frac{(\mu-r)^2(T-t)}{2\sigma^2}} \big| \mathscr{F}_t^G\right], \quad \textit{for } (t,x) \in [0,T] \times \mathbb{R}, \tag{3.31}$$

and the optimal (feedback) policy is given by (3.30)*.*

Proof. It is easy to check that the feedback policy π^* defined by (3.30) belongs to $\mathcal{A}$. Set

$$u(t,x) = 1 - e^{-p\kappa_T x} E\left[e^{-p\kappa_T H(T) - \frac{(\mu-r)^2(T-t)}{2\sigma^2}} \big| \mathscr{F}_t^G\right],$$

which is the right hand side of (3.31). It is straightforward to verify that u satisfies the stochastic HJB

equation (3.26). In particular, using the (feedback) policy given by (3.30) yields that for each $s \in [0,T]$,

$$\begin{aligned}
&u(s,x)\\
&= E\Bigg[U(\kappa_T(x+H(T)))\\
&\quad + \int_s^T \left\{\frac{1}{2}(x+H(t))^2|\pi^*|^2\sigma^2 D_{xx}u + (x+H(t))(\mu-r)\pi^* D_x u\right\}(t,x)\,dt\Big|\mathscr{F}_s^G\Bigg] \qquad (3.32)\\
&= E\Bigg[U(\kappa_T(x+H(T)))\\
&\quad + \int_s^T \max_{\pi\in\mathbb{R}}\left\{\frac{1}{2}(x+H(t))^2|\pi|^2\sigma^2 D_{xx}u + (x+H(t))(\mu-r)\pi D_x u\right\}(t,x)\,dt\Big|\mathscr{F}_s^G\Bigg]. \qquad (3.33)
\end{aligned}$$

On the one hand for each $\pi \in \mathcal{A}$, applying Lemma 3.1 with $\tau_k := \inf\left\{t > s : \widetilde{R}^{\pi;s,x}(t) < -k\right\}$ for $k > 0$, we have

$$\begin{aligned}
&u(s,x)\\
&= E\Bigg[u(\tau_k, \widetilde{R}^{\pi;s,x}(\tau_k))\\
&\qquad + \int_s^{\tau_k} \max_{\pi\in\mathbb{R}}\left\{\frac{1}{2}(\widetilde{R}^{\pi;s,x}(t)+H(t))^2|\pi(t)|^2\sigma^2 D_{xx}u + (x+H(t))(\mu-r)\pi(t) D_x u\right\}(t,\widetilde{R}^{\pi;s,x}(t))\,dt\\
&\qquad - \int_s^{\tau_k} \left\{\frac{1}{2}(\widetilde{R}^{\pi;s,x}(t)+H(t))^2|\pi_t|^2\sigma^2 D_{xx}u + (x+H(t))(\mu-r)\pi(t) D_x u\right\}(t,\widetilde{R}^{\pi;s,x}(t))\,dt\Big|\mathscr{F}_s^G\Bigg]\\
&\geq E\left[u(\tau_k, \widetilde{R}^{\pi;s,x}(\tau_k))\Big|\mathscr{F}_s^G\right], \quad \text{a.s.}, \qquad (3.34)
\end{aligned}$$

which by letting $k \to \infty$ implies that $u(s,x) \geq \widetilde{V}(s,x)$, a.s. for each $(s,x) \in [0,T]\times\mathbb{R}$, where we recall the value function $\widetilde{V}(s,x)$ in (3.24). On the other hand, we have the equality for $\pi = \pi^*$ in (3.34) particularly and thus it holds that $u(s,x) \leq \widetilde{V}(s,x)$, a.s. for all $(s,x) \in [0,T]\times\mathbb{R}$. Therefore, we have $u(s,x) = \widetilde{V}(s,x)$, a.s. for all $(s,x) \in [0,T]\times\mathbb{R}$, and π^* given by (3.30) is optimal. □

Moreover, in Section 3.2, for Merton investment problem in insurance for Hawkes process based model we assumed, $\widetilde{R}^{\pi}(t) = e^{-rt}R^{\pi}(t) - H(t)$. After rearranging and isolating $R^{\pi}(t)$ we get, $\widetilde{R}^{\pi}(t) + H(t) = e^{-rt}R^{\pi}(t)$. Putting this in equation (3.30) we can get,

$$\begin{aligned}
\pi^*(t,\bar{x}) =& \frac{(\mu-r)e^{rt}}{\sigma^2\bar{x}\kappa_T p}\\
=& \frac{(\mu-r)e^{-r(T-t)}}{\sigma^2\bar{x}p},
\end{aligned} \qquad (3.35)$$

where $\bar{x}$ represents $R^{\pi}(t)$ and $\kappa_T = e^{rT}$.

3.4 Application in the case where the conditional intensity is exponential decay for HP

Here we consider the stochastic optimal control problem $(3.22)-(3.23)$ from the Merton investment problem in insurance. From equation (2.6) we can recall Hawkes conditional intensity function for exponential decay:

$$\lambda^*(t) = \lambda + \int_{-\infty}^{t} \alpha e^{-\beta(t-s)} dN(s) = \lambda + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)},$$

Here the self-exciting function $\mu(t)$ of the intensity $\lambda^*(t)$ of the Hawkes process $N(t)$ is exponential:

$$\mu(t) = \alpha e^{-\beta t},$$

where $\alpha, \beta > 0$ are some parameters. In this case $(\lambda^*(t), N(t))$ is Markovian.

3.4.1 Infinitesimal generator

For any Markov processes $(Y(t))_{t\geq 0}$ infinitesimal generator L is defined as

$$Lf(y) = lim_{h\downarrow 0} \frac{E[f(Y(t+h)) - f(y)|Y(t) = y]}{h},$$

where the function $f(y)$ is sufficiently smooth and restricted to the domain D of L where this limit exists. When D is linear, the operator L is linear and the infinitesimal generator does not depend on t if the Markov process is time homogeneous.

Since the two-component process $Y(t) := (\lambda^*(t), N(t))$ is Markovian, its infinitesimal generator is

$$Lf(\lambda^*(t), N(t)) = Lf(Y(t)) = \lim_{h\to 0} \frac{E_t^y[f(Y(t+h)] - f(y)}{h},$$

where $y = Y(t) = (\lambda^*(t), N(t))$, $E_t^y[\cdot] := E^y[\cdot|\mathcal{F}_t]$, and $\mathcal{F}_t$ is a σ-algebra generated by $N(t)$. If we choose exponential decay function for our self exciting process, then from the definition of Hawkes process we get

$$d\lambda^*(t) = \beta(\lambda - \lambda^*(t))dt + \alpha dN(t).$$

The descretised equation of the above equation is

$$\lambda^*(t+h) = \lambda^*(t) + \beta(\lambda - \lambda^*(t))h + \alpha \triangle N(t),$$

where, $\triangle N(t) = N(t+h) - N(t)$. And $\triangle N(t) = 1$ for new arrivals and $\triangle N(t) = 0$ for other cases. Now, from the definition of infinitesimal generator we get the term $E_t^y[f(Y(t+h))]$ which is mentioned in [14] and can be explained as,

$$\begin{aligned}
E_t^y[f(Y(t+h))] =& \lambda^*(t)hf(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h + \alpha, N(t)+1) + (1 - \lambda^*(t)h)f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t)) \\
=& \lambda^*(t)h[f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h + \alpha, N(t)+1) - f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t))] \\
&+ f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t)).
\end{aligned}$$

Here, each jump occurs with probability $\lambda^*(t)h$ and the probability of no jump is $1 - \lambda^*(t)h$. Now, we can bring limit on the both sides of above equation and after dividing the equation by h and we get,

$$\begin{aligned}
\lim_{h\to 0} \frac{E_t^y[f(Y(t+h)) - f(Y(t))]}{h} =& \lim_{h\to 0} \lambda^*(t)[f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h + \alpha, N(t)+1) - f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t))] \\
&+ \lim_{h\to 0} \frac{f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t)) - f(\lambda^*(t), N(t))}{h}.
\end{aligned}$$

So the infinitesimal generator for Hawkes process can be written as

$$\begin{aligned}
Lf(Y(t)) &= \lim_{h\to 0} \lambda^*(t)[f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h + \alpha, N(t)+1) - f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t))] \\
&\quad + \lim_{h\to 0} \left[\frac{f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t))}{h} - \frac{f(\lambda^*(t), N(t))}{h}\right] \\
&= \lim_{h\to 0} \lambda^*(t)[f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h + \alpha, N(t)+1) - f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t))] \\
&\quad + \lim_{h\to 0} \frac{f(\lambda^*(t) + \beta(\lambda - \lambda^*(t))h, N(t)) - f(\lambda^*(t), N(t))}{h} \\
&= \lambda^*(t)[f(\lambda^*(t) + \alpha, N(t)+1) - f(\lambda^*(t), N(t))] + \beta(\lambda - \lambda^*(t))\frac{\partial f(\lambda^*(t), N(t))}{\partial \lambda^*(t)}.
\end{aligned}$$

From here it follows that the generator of $G(t)$ is:

$$L^{G,x}f(y) = \beta(\lambda - \lambda^*(t))\tfrac{\partial f(y)}{\partial \lambda^*(t)} + \lambda^*(t)[\textstyle\int_X (f(\lambda^*(t) + \alpha, N(t) + a(u)) - f(y))P(x, du)],$$

where X is the state space of Markov chain X_i and $P(x, dy)$ are transition probabilities of X_i, $X_0 = x$. Thus, if our initial risk process $\bar{R}(t)$ is

$$\bar{R}(t) = u + ct - G(t), \quad \bar{R}(0) = u,$$

then the generator of $(\bar{R}(t), \lambda^*(t), N(t))$, which is Markov process as well, is

$$\begin{aligned}
Af(y,u) &= c\tfrac{\partial f(y,u)}{\partial u} + \beta(\lambda - \lambda^*(t))\tfrac{\partial f(y,u)}{\partial \lambda^*(t)} \\
&+ \lambda^*(t)E_t[f(y, u - a(X)) - f(y,u)],
\end{aligned}$$

where $a(X)$ is defined in $G(t) = \sum_{i=1}^{N(t)} a(X_i), \quad t \geq 0$.

3.4.2 HJB Equation and Optimal Control

Let $(\lambda^*(t), N(t)) = (\lambda^*, n)$. Inspired by equation (3.26) and Markovian framework we may find function $\hat{V}$ on $[0,T] \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+$ such that

$$\widetilde{V}(s, x - H(s)) = \hat{V}(s, x, n, \lambda^*).$$

And we can go further by taking expectations on both sides:

$$\begin{aligned}\hat{V}(s,x,N(t),\lambda^*(t)) =& E\Bigg[U(\kappa_T x) \\ &+ \int_s^T \max_{\pi\in\mathbb{R}} \left\{\frac{1}{2}x^2\pi^2\sigma^2 D_{xx}\hat{V} + x(\mu - r)\pi D_x \hat{V}\right\}(t,x,N(t),\lambda^*(t))\,dt\Bigg]. \end{aligned} \tag{3.36}$$

Now by using dynamic programming principle we can have,

$$\begin{aligned}\hat{V}(s,x,N(s),\lambda^*(s)) =& E\Bigg[\hat{V}(s+h,x,N(s+h),\lambda^*(s+h)) \\ &+ \int_s^{s+h} \max_{\pi\in\mathbb{R}} \left\{\frac{1}{2}x^2\pi^2\sigma^2 D_{xx}\hat{V} + x(\mu - r)\pi D_x \hat{V}\right\}(t,x,N(t),\lambda^*(t))\,dt\Bigg]. \end{aligned} \tag{3.37}$$

We apply the expectation operator E to the above equation then divide both sides by h, move h within the expectations and let h tends to 0^+. We assume enough regularity to allow us to take the limit within the expectation. It follows that

$$\begin{aligned}&\left[\int_s^{s+h} \max_{\pi\in\mathbb{R}} \left\{\frac{1}{2}x^2\pi^2\sigma^2 D_{xx}\hat{V} + x(\mu - r)\pi D_x \hat{V}\right\}(t,x,N(t),\lambda^*(t))\,dt\right] \\ &\qquad = -E\left[\hat{V}(s+h,x,N(s+h),\lambda^*(s+h)) - \hat{V}(s,x,N(s),\lambda^*(s))\right] \\ \Rightarrow &\frac{1}{h}\left[\int_s^{s+h} \max_{\pi\in\mathbb{R}} \left\{\frac{1}{2}x^2\pi^2\sigma^2 D_{xx}\hat{V} + x(\mu - r)\pi D_x \hat{V}\right\}(t,x,N(t),\lambda^*(t))\,dt\right] \\ &\qquad = -\frac{1}{h}E\left[\hat{V}(s+h,x,N(s+h),\lambda^*(s+h)) - \hat{V}(s,x,N(s),\lambda^*(s))\right] \\ \Rightarrow &\max_{\pi\in\mathbb{R}} \left\{\frac{1}{2}x^2\pi^2\sigma^2 D_{xx}\hat{V} + x(\mu - r)\pi D_x \hat{V}\right\}(t,x,n,\lambda^*) = (-\frac{\partial \hat{V}}{\partial s} - A\hat{V})(s,x,n,\lambda^*). \end{aligned} \tag{3.38}$$

In this way we get the deterministic HJB equation for the value function $\hat{V}(t,x,n,\lambda^*)$:

$$\begin{cases} -\frac{\partial \hat{V}}{\partial t} & = A\hat{V} + \max_{\pi\in\mathbb{R}}\left\{\frac{1}{2}x^2\pi^2\sigma^2 D_{xx}\hat{V} + x(\mu-r)\pi D_x\hat{V}\right\}(t,x,n,\lambda^*) \\ & = A\hat{V} + \max_{\pi\in\mathbb{R}}\left\{\frac{1}{2}\sigma^2x^2D_{xx}\hat{V}\left[\pi^2 + 2\pi\frac{(\mu-r)D_x\hat{V}}{\sigma^2xD_{xx}\hat{V}}\right]\right\}(t,x,n,\lambda^*) \\ & = A\hat{V} + \max_{\pi\in\mathbb{R}}\left\{\frac{1}{2}\sigma^2x^2D_{xx}\hat{V}\left(\pi + \frac{(\mu-r)D_x\hat{V}}{\sigma^2xD_{xx}\hat{V}}\right)^2 - \frac{(\mu-r)^2|D_x\hat{V}|^2}{2\sigma^2D_{xx}\hat{V}}\right\}(t,x,n,\lambda^*), \\ & = A\hat{V} - \frac{(\mu-r)^2|D_x\hat{V}(t,x,n,\lambda^*)|^2}{2\sigma^2D_{xx}\hat{V}(t,x,n,\lambda^*)} \\ \hat{V}(T,x,n,\lambda^*) & = 1-e^{-p\kappa_T x}. \end{cases} \tag{3.39}$$

To solve the above HJB equation we consider the form $\hat{V}(t,x,n,\lambda^*) = 1 - e^{-p\kappa_T x}\Phi(t,n,\lambda^*)$ which will give us a transformed HJB:

$$\begin{cases} -\frac{\partial\Phi}{\partial t} = & A\Phi - \frac{(\mu-r)^2}{2\sigma^2}\Phi, \\ \Phi(T) = & 1. \end{cases} \tag{3.40}$$

Let $\eta = -\frac{(\mu-r)^2}{2\sigma^2}$. Then $\Phi(t) = e^{\eta(T-t)} = e^{-\frac{(\mu-r)^2}{2\sigma^2}(T-t)}$. So, we can get our value function $\hat{V}(t,x,n,\lambda^*) = 1 - e^{-p\kappa_T x}e^{-\frac{(\mu-r)^2}{2\sigma^2}(T-t)}$. Therefore, $D_x\hat{V} = p\kappa_T e^{-p\kappa_T x}e^{-\frac{(\mu-r)^2}{2\sigma^2}(T-t)}$ and $D_{xx}\hat{V} = -p^2\kappa_T^2 e^{-p\kappa_T x}e^{-\frac{(\mu-r)^2}{2\sigma^2}(T-t)}$. Finally, we can get $\pi^*(t,x)$ as

$$\begin{aligned} \pi^*(t,x) =& -\frac{\mu-r}{x\sigma^2}\frac{D_x\hat{V}}{D_{xx}\hat{V}} \\ =& \frac{\mu-r}{x\sigma^2 p\kappa_T}, \end{aligned} \tag{3.41}$$

which coincides with the optimal control obtained in (3.35). This is a direct approach to deal with Hawkes process with exponential conditional intensity.

Chapter 4

Implement of Risk Model Results

To implement the results of the risk model for insurance which is based on GCHP, we will use empirical data for claim occurrence. A large German insurance company supplied the data set consisting of claim occurrences from the class of legal insurance expenses. There were 44 classes in the whole data set which refers to insurance protection covering the cost of a legal dispute. The data set comprises claims which take place in the years from 2007 to 2011. But they are reported after three years from 2010 to 2014 with a delay and have corresponding payment dates from 1st January 2010 to 28th July 2016. The day of the payment is documented for each claim payment.

In our research, we did not work with the empirical data firsthand. Rather we used the results from [36], which processed and estimated the parameters from the empirical data. Also, we used their value for other parameters that are not derivable from the empirical data set. Most of the calculations for simulation of claim size in the next section are due to [36].

4.1 Simulation of Claim Size

The claim sizes for insurance companies are supposed to be i.i.d. with a distribution, say F, for which the first two moments are finite. From the existing insurance literature, it is evident that exponential distribution with rate parameter $\gamma > 0$ is a common choice for F. We want to simulate modeled claim sizes to approximate an i.i.d. sequence which will follow the empirical distribution of observed claim sizes. For this purpose, we assume that the claim sizes will follow a finite number of fixed jump sizes governed by the Markov chain evolution, which is a theoretical result for the risk model with GCHP, where no dependence between subsequent claim sizes is assumed.

Let $\hat{F}$ be the empirical distribution function of the claim sizes Y_i, with $\hat{F}$ having finite two moments:

$$E[Y] = \frac{1}{\gamma} \text{ and } E\left[Y^2\right] = \frac{2}{\gamma^2} \tag{4.1}$$

Also, let B be the maximum observed claim size and hence $\hat{F}(B) = 1$. The simulation of the sequence of claim sizes to approximate the empirical distribution $\hat{F}$ can be made arbitrarily well by increasing $\hat{N}$, the number of states of the Markov chain.

From the empirical data set, the maximum observed claim size $B = 10071$. With equidistant boundaries $(b_1, b_2, \cdots, b_{\hat{N}} = B)$, we define $\pi^* = (\pi_1^*, \pi_2^*, \cdots, \pi_{\hat{N}}^*)$ as

$$\begin{aligned} \pi_1^* =& \hat{F}(b_1), \\ \pi_2^* =& \hat{F}(b_2) - \pi_1^*, \\ & \cdots \\ \pi_{\hat{N}}^* =& \hat{F}(b_{\hat{N}}) - \sum_{i=1}^{\hat{N}-1} \pi_i^* = 1 - \sum_{i=1}^{\hat{N}-1} \pi_i^*, \end{aligned} \tag{4.2}$$

where $\sum_{i=1}^{\hat{N}} \pi_i^* = 1$.

Let (X_i) be a Markov Chain on the state space $X = \{1, 2, ..., \hat{N}\}$ with transition matrix,

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

Being an irreducible Markov chain on a finite state space, (X_i) has a unique stationary distribution [26]. Also we can verify that π^* is indeed the stationary distribution. Since,

$$\begin{aligned} \pi^* P =& \begin{pmatrix} \pi_1^* & \pi_2^* & \cdots & \pi_{\hat{N}}^* \end{pmatrix} \begin{pmatrix} \pi_1^* & \pi_2^* & \cdots & \pi_{\hat{N}}^* \\ \cdots & \cdots & \cdots & \cdots \\ \pi_1^* & \pi_2^* & \cdots & \pi_{\hat{N}}^* \end{pmatrix} \\ =& \begin{pmatrix} \pi_1^* \sum_{i=1}^{\hat{N}} \pi_i^*, & \pi_2^* \sum_{i=1}^{\hat{N}} \pi_i^*, & \cdots, & \pi_{\hat{N}}^* \sum_{i=1}^{\hat{N}} \pi_i^* \end{pmatrix} \\ =& \begin{pmatrix} \pi_1^*, & \pi_2^*, & \cdots, & \pi_{\hat{N}}^* \end{pmatrix} \\ =& \pi^*. \end{aligned} \tag{4.3}$$

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

$$\mathbb{P}(X_{i+1} = k | X_i = j) = \mathbb{P}(X_{i+1} = k | X_i = l) = \pi_k^*, \tag{4.4}$$

for all $j, l \in X, i \in \mathbb{N}$. By Markov property and the law of total probability

$$\begin{aligned} \mathbb{P}(X_{i+1} = k) =& \sum_{j \in X} \mathbb{P}(X_{i+1} = k | X_i = j) \mathbb{P}(X_i = j) \\ =& \pi_k^* \mathbb{P}_{j \in X}(X_i = j) \\ =& \pi_k^*, \end{aligned} \tag{4.5}$$

for all $k \in X, i \in \mathbb{N}$.

The probability of realizing one state is not dependent on the previous state and the Markov Chain (X_i) represents an i.i.d. sequence. Let Y be a random variable with c.d.f. $\hat{F}$, $A_i := \{\omega : Y(\omega) \in (b_{i-1}, b_i]\}$, and set

$$\begin{aligned} a(i) =& \mathbb{E}[Y | b_{i-1} < Y \leq b_i] \\ =& \mathbb{E}[Y | A_i] \\ =& \frac{\mathbb{E}[Y 1_{A_i}]}{\mathbb{P}(A_i)} \\ =& \frac{\mathbb{E}[Y \mathbf{1}_{A_i}]}{\mathbb{P}(A_i)} \\ =& \frac{\mathbb{E}[Y \mathbf{1}_{A_i}]}{\pi_i^*}. \end{aligned} \tag{4.6}$$

Now using the equation (4.6) we get

$$a^* = \sum_{i=1}^{\hat{N}} \pi_i^* a(i) = \sum_{i=1}^{\hat{N}} \pi_i^* \frac{\mathbb{E}[Y \mathbf{1}_{A_i}]}{\pi_i^*} = \sum_{i=1}^{\hat{N}} \mathbb{E}[Y \mathbf{1}_{A_i}] = \mathbb{E}[Y], \tag{4.7}$$

and thus $a^* = \frac{1}{\gamma}$. Here $a(X_i)$ describes an i.i.d. sequence that approximates the distribution $\hat{F}$ arbitrarily close as the number of states $\hat{N} \to \infty$.

For the case of 20-state Markov Chain, Table 4.1 gives us the equidistant boundaries, state values $a_i := a(i)$, the distribution π_i^* and the value of a^*.

Through Figure 4.1, 4.2 and 4.3 we represent the distribution function of the empirical claims with its counterpart from claims generated by the Markov chain approach with number of states $\hat{N} = 5, 20, 50$ respectively. We observe that as $\hat{N}$ increases, the empirical claim sizes are replicated reasonably within the

Parameter	Value
$(b_1, b_2, ..., b_{20} = B)$	$(503.55, 1007.1, 1510.65, 2014.2, 2517.75, 3021.3, 3524.85, 4028.4, 4531.95, 5035.5,$ $5539.05,\ 6042.6,\ 6546.15,\ 7049.7,\ 7553.25,\ 8056.8,\ 8560.35,\ 9063.9,\ 9567.45,\ 10071.0)$
$(\pi_1^*, \pi_2^*, ..., \pi_{20}^*)$	$(0.46821489, 0.24898971, 0.13240902, 0.07041314, 0.03744466, 0.01991251, 0.01058918,$ $0.00563117,\ 0.00299457,\ 0.00159247,\ 0.00084685,\ 0.00045034,\ 0.00023949,\ 0.00012735,$ $0.00006773,\ 0.00003602,\ 0.00001915,\ 0.00001018,\ 0.00000542,\ 0.00000288)$
$(a_1, a_2, ..., a_{20})$	$(225.4495, 728.9995, 1232.5495, 1736.0995, 2239.6495, 2743.1995, 3246.7495,$ $3750.2995,\ 4253.8495,\ 4757.3995,\ 5260.9495\ 5764.4995,\ 6268.0495\ 6771.5995,$ $7275.1495,\ 7778.6995,\ 8282.2495,\ 8785.7995,\ 9289.3495,\ 9792.8995\ )$
a^*	797.3672
$E[X_1]$	797.3672

Table 4.1: Boundaries $b = (b_1, b_2, ..., b_{\hat{N}})$, stationary distribution π^*, state values $a(i)$ and expected value a^* under the stationary distribution for a 20-state Markov Chain and empirical claim sizes.

$\alpha = 0.0260$	$\beta = 0.0334$	$\lambda = 0.1467$

Table 4.2: Parameters for the Exponential HP

framework of the risk model with GCHP.

4.2 Simulation of the Risk Process using Exponential Decay Excitation for HP

To simulate the risk process $R^\pi(t)$, first, we simulate the Hawkes point process to model claim arrival time by the inverse compensator method and then simulate the claim sizes from a Markov chain with the number of states $\hat{N} = 20$ using the procedure described in the previous section. The parameters of the exponential Hawkes process are given in Table 4.2. All these parameters were not inferable from the empirical data set but they are taken from [36].

Then we will simulate the process

$$G(t) = \sum_{i=1}^{N(t)} a(X_i), \quad t \geq 0,$$

which is the aggregated claim size until time $t > 0$. At initial time when $t_0 = 0$ then $G(0) = 0$. If the first point of Hawkes process is t_1, then $G(t_1) = a(X_1)$. Then with the second point t_2 of the Hawkes process, $G(t_2) = G(t_1) + a(X_2)$. Thus for each point t_i of the Hawkes process we get $G(t_i) = G(t_{i-1}) + a(X_i)$. Recall from Merton investment problem in insurance in chapter 3, that, $H(t) = \frac{c}{r}(1 - e^{-rt}) - \int_0^t e^{-rs} dG(s)$. So, once we have the process $G(t)$, we can simulate $H(t)$.

Finally we use the formula of feedback control $\pi^*(t, x)$ from equation (3.35) and simulate $R^\pi(t)$ using

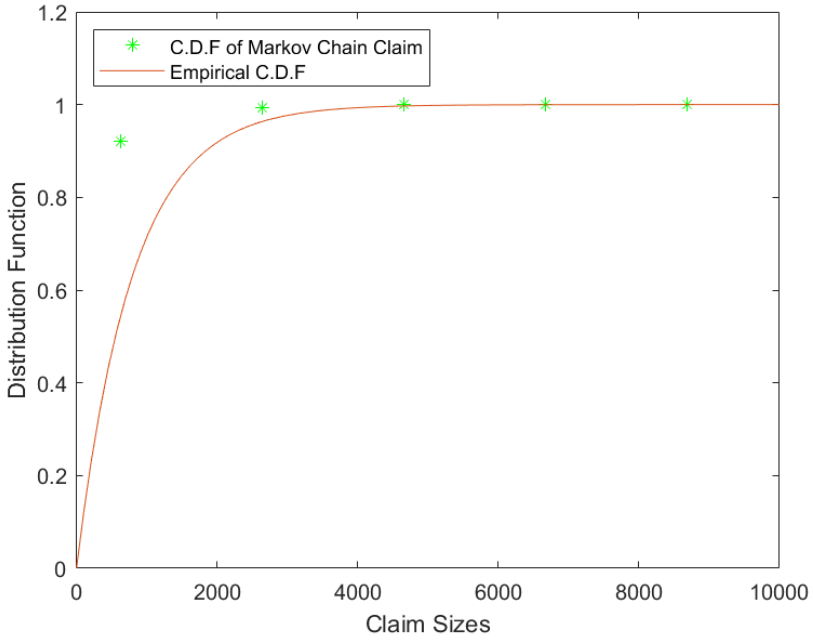

Figure 4.1: Claim sizes from 5-States of Markov Chain

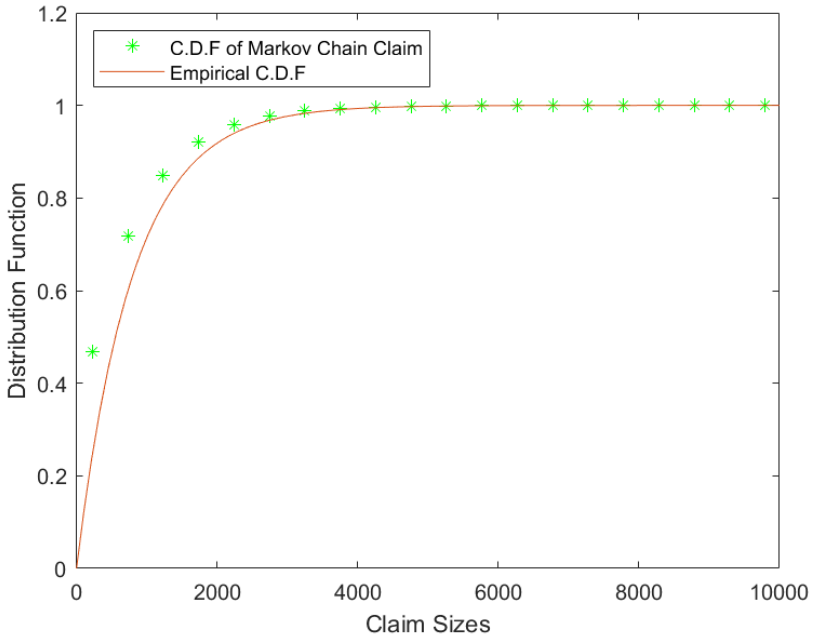

Figure 4.2: Claim sizes from 20-States of Markov Chain

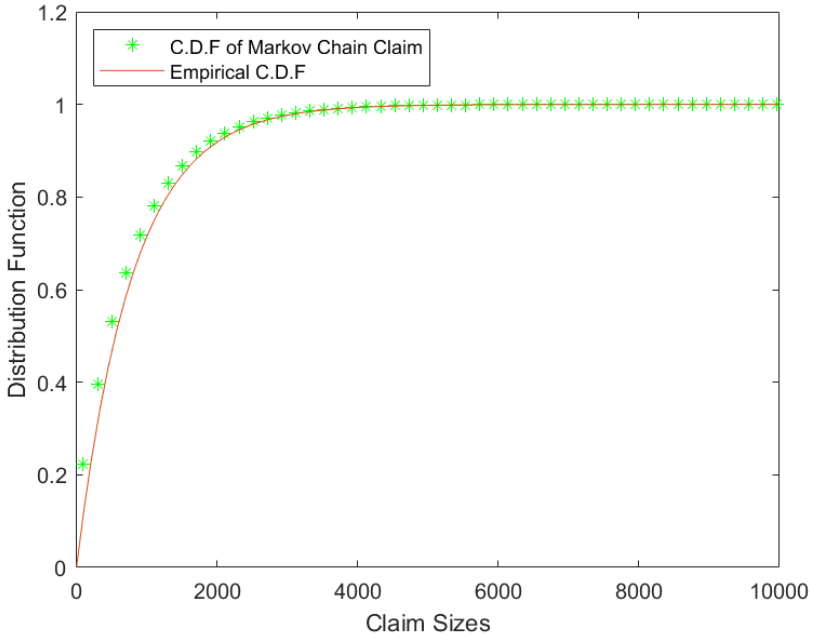

Figure 4.3: Claim sizes from 50-States of Markov Chain

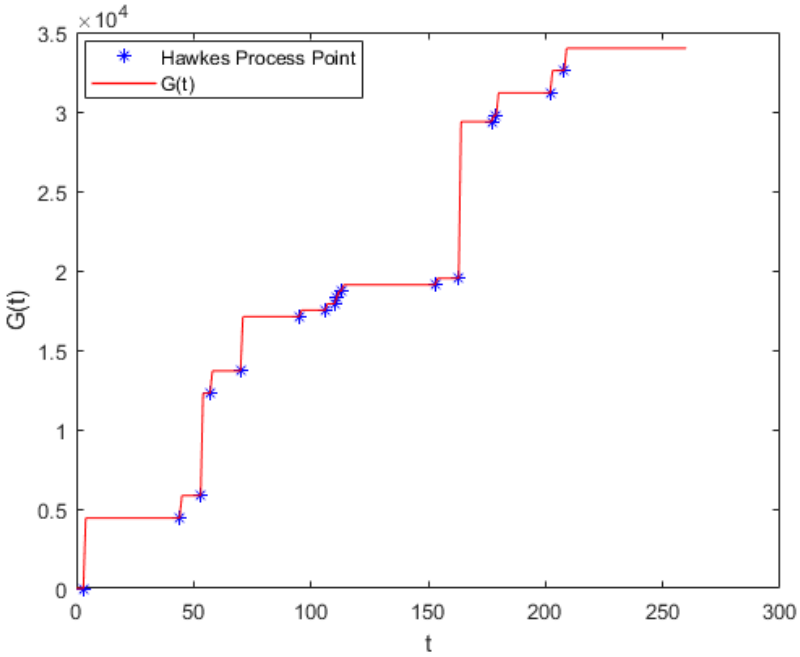

Figure 4.4: Sample path of $G(t)$

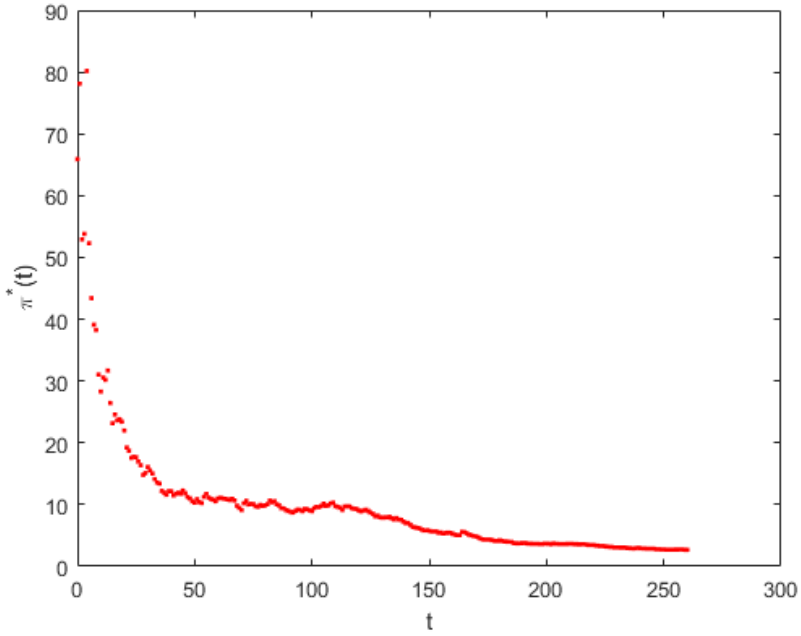

Figure 4.5: Simulations of $\pi^*(t, x(t))$

$r = 0.06$	$\mu = 0.20$	$\sigma = 0.05$	$p = 0.0001$	$u = 8000$	$c = 633.5552$

Table 4.3: Parameters for the Risk process

stochastic differential equation (SDE) (3.21). For the simulation of an SDE, we used Euler-Maruyama Method with MATLAB. In Itô calculus, the Euler–Maruyama method is used to approximate the numerical solution of a stochastic differential equation (SDE). It is an extension of the Euler method for ordinary differential equations (ODE) to stochastic differential equations (SDE). It is named after Leonhard Euler and Gisiro Maruyama. To apply the numerical method to (3.21), first we discretize the time interval $[0, T]$. Let $\Delta t = dt = (T - 0)/N$ for some positive integer N, and denote $\tau_j = j\Delta t; j = 0, 1, \cdots, N$. Here we are simulating the risk process for the period of 1 year (260 days) and also we choose $N = 260$. The numerical approximation to $R^\pi(\tau_j)$ will be denoted by R^π_j. Then the Euler–Maruyama (EM) method gives us the time discretization of (3.21) as:

$$R^\pi_{j+1} = R^\pi_j + rR^\pi_j dt + (\mu - r)^2 e^{r*t_j}/(\sigma^2 \kappa_T p)dt + (\mu - r)e^{rt_j}/(\sigma\kappa_T p)(W_{j+1} - W_j) + cdt - (G_{j+1} - G_j).$$

The values of the parameters for the risk process are given in Table 4.3. Figure 4.4 shows us a sample path of $G(t)$ for a 20-state Markov Chain and the corresponding paths of $\pi^*(t)$ and $R^\pi(t)$ is shown on Figure 4.5 and 4.6 respectively.

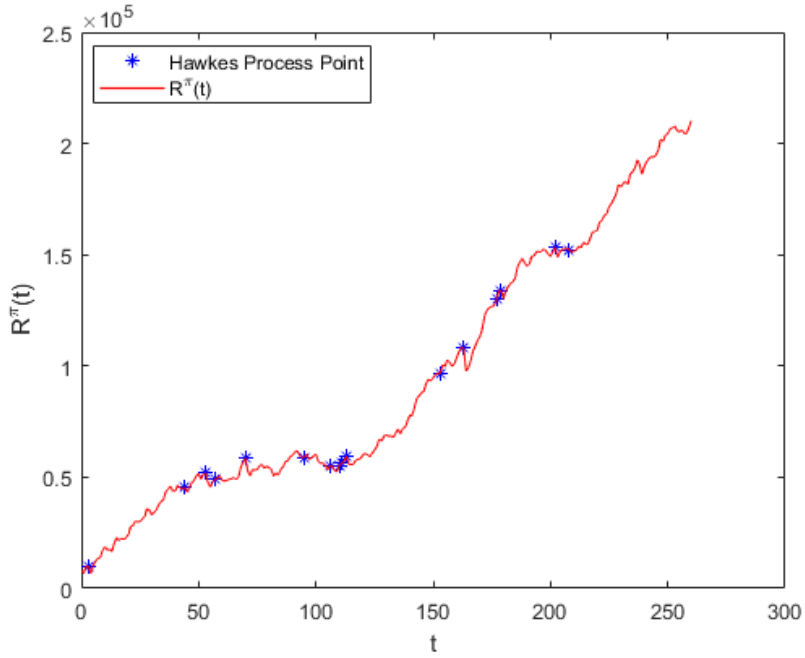

Figure 4.6: Simulations of $R^{\pi}(t)$

4.3 Simulation of the Risk Process using Power-law Excitation for HP

In the previous section, we simulate the risk process using exponential decay excitaion function for the conditional intensity of the HP. Note that, when the conditional intensity of HP is exponential type, the joint process $(\lambda^*(t), N(t))$ is Markovian leading to a stochastic optimal control problem of Markovian type as seen in Section 3.4. However, when the conditional intensity of HP is power type, the corresponding stochastic optimal control problem is non-Markovian and we would stress that the theory established in Chapter 3 is working even for non-Markovian cases. In this section, we will simulate the risk process using Hawkes process with the power-law excitation function for the conditional intensity. From Chapter 2 we can recall that the excitation function for power law is $\mu(t) = \frac{m}{(d+(t))^q}$. So, the conditional intensity function described in equation (2.8) is:

$$\lambda^*(t) = \lambda + \int_{-\infty}^{t} \frac{m}{(d+(t-s))^q} dN(s) = \lambda + \sum_{t_i<t} \frac{m}{(d+(t-t_i))^q}.$$

Table 4.4 gives us the values of the parameters for the HP with Power-law excitaion function. The parameters for the risk process in this case remains the same from the previous section. Figure 4.7 shows us a sample path of $G(t)$ for a 20-state Markov Chain and the corresponding paths of $\pi^*(t)$ and $R^{\pi}(t)$ is shown on Figures

$m = 0.005$	$q = 2$	$d = 0.5$

Table 4.4: Parameters for the Power-law HP

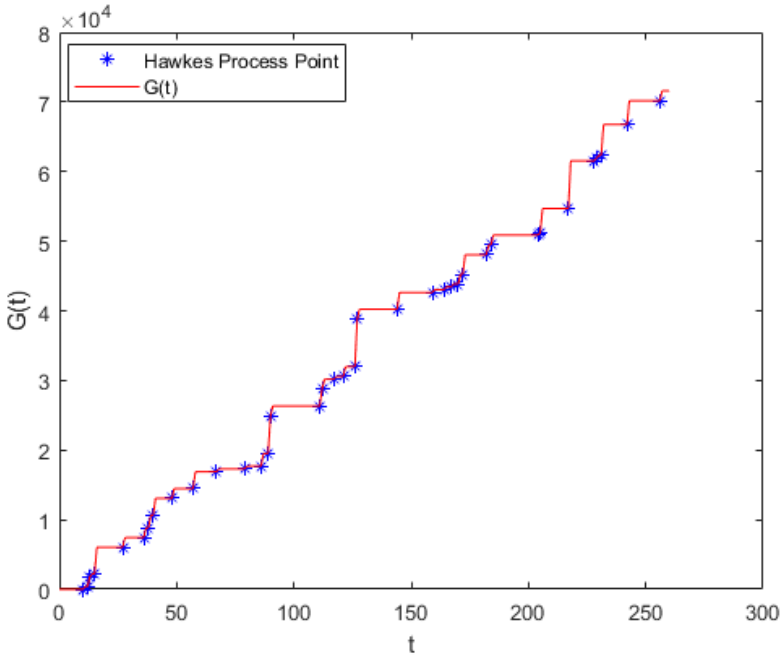

Figure 4.7: Sample path of $G(t)$

4.8 and 4.9 respectively.

Observing the pair of Figures 4.5 and 4.6 from HP with exponential excitation and pair of Figures 4.8 and 4.9 from HP with power-law excitation, we see that when the capital (risk process) increase, the investment in risky assets of the insurance company decrease, that is a company does not want to take more risk when their capital keeps increasing. On the other hand, whenever the capital decreases, the investment in the risky assets increases.

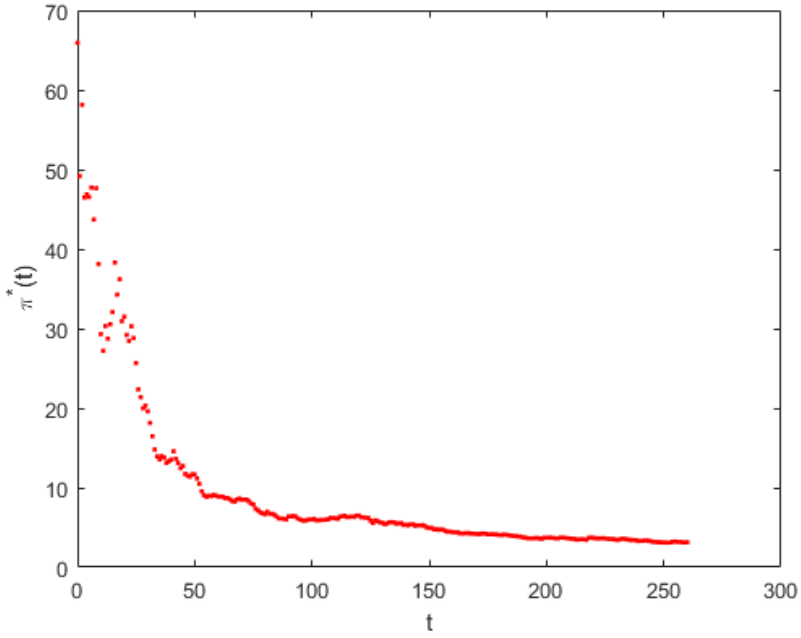

Figure 4.8: Simulations of $\pi^*(t, x(t))$

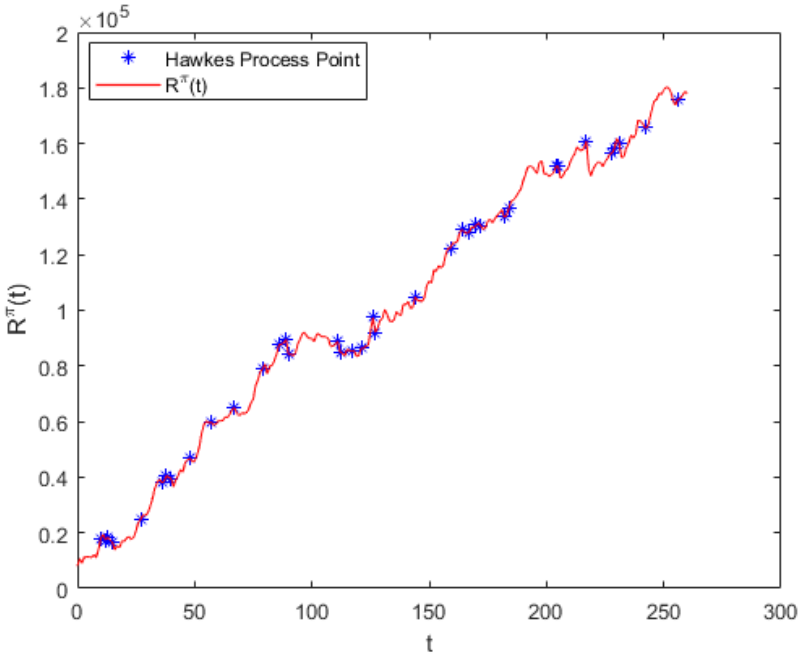

Figure 4.9: Simulations of $R^\pi(t)$

Chapter 5

Conclusion

In this thesis, we study the Merton investment problem for insurance company, where the arrival of claims follow a Hawkes process and the amount of claims follow a finite number of fixed jump governed by a Markov chain, that means, the risk model is based on general compound Hawkes process. Also the insurance company wants to invest their capital into financial market to increase their revenue. Investment in the risk-free assets give a constant amount of return over time with no risk whereas risky assets always have some risk exposure and uncertainty but expected rate of return may be higher than risk-free assets. So the insurance company wants to find out what is the optimal allocation of their capital between risky and risk-free assets that would maximize their expected utility at terminal time. When the aggregated claim size increases, it translates to an outflow of the money from the company which decreases the utility and to cover up the loss the company may need to adjust their strategies.

In order to solve the Merton investment problem which is a stochastic optimal control problem, we use the dynamic programming problem to derive the stochastic Hamilton-Jacobi-Bellman(SHJB) equation for the value function. The SHJB equation provides a way of solving for the optimal feedback control (optimal investment portfolio), which then can be used to solve the controlled stochastic differential equation (SDE) for the risk process.

To implement the idea for solving Merton investment problem through the dynamic programming method and the SHJB equation, we use empirical data set for claim occurrence. First we simulate the sequence of claim sizes which follows the distribution of empirical claim sizes and Hawkes point process to model the arrival process of claims. Then using the optimal investment portfolio obtained from dynamic programming method and SHJB equation we simulate the controlled SDE for the risk process.

The future work will be devoted to simulation of p for the utility function $U(x) = 1 - e^{-px}$ by using

particle swarm optimization (PSO) algorithm. PSO is a computational method that optimizes a problem by iteratively trying to improve a candidate solution. This algorithm searches the space of an objective function by adjusting the trajectories of particles. When a particle finds a location that is better than the previous one then PSO updates that location as the new current best for the mentioned particle. The aim is to find the global best among all the individual best solution until the objective no longer improves. As we will not use diffusion approximation for $R(t)$ like stated in [31] so that we can use only numerical methods.

Bibliography

[1] Swishchuk Anatoliy. Risk Model based on General Compound Hawkes Process. *arXiv*, 2018.

[2] E. Bacry, S. Delattre, M. Hoffmann, and J. Muzy. Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123:2475–2499, 2013.

[3] M. S. Bartlett. The Spectral Analysis of Point Processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25(2):264–296, 1963.

[4] Clive G. Bowsher. Modelling Security Market Events in Continuous time: Intensity based, Multivariate Point Process Models. *Journal of Econometrics*, 141(2):876–912, 2007.

[5] David R. Brillinger. Measuring the Association of Point Processes: A Case History. *American Mathematical Monthly*, 83:16–22, 1976.

[6] Patrick Brockett and Xiaohua Xia. Operations Research in Insurance: A Review. *Transactions of Society of Actuaries 1995*, 47, 1995.

[7] Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci. Buy Low, Sell High: A High Frequency Trading Perspective. *SIAM Journal on Financial Mathematics*, 5(1):415–444, 2014.

[8] D. R. Cox. Some Statistical Methods Connected with Series of Events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(2):129–164, 1955.

[9] David Cox and P. A. W Lewis. *The Statistical Analysis of Series of Events*. 1966.

[10] Lirong Cui, Alan Hawkes, and He Yi. An Elementary Derivation of Moments of Hawkes Processes. *Advances in Applied Probability*, 52(1):102–137, 2019.

[11] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer New York, 2006.

[12] Angelos Dassios and Hongbiao Zhao. Exact Simulation of Hawkes Process with Exponentially Decaying Intensity. *Electronic Communications in Probability*, 18:1–13, 2013.

[13] V. Filimonov and D. Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Phys. Rev. E*, 85, 2012.

[14] Jose Da Fonseca and Riadh Zaatour. Hawkes Process: Fast Calibration, Application to Trade Clustering, and Diffusion Limit. *Journal of Futures Markets*, 34:1–42, 2013.

[15] Kay Giesecke and Pascal Tomecek. Dependent Events and Changes of Time. 08 2022.

[16] Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90, 1971.

[17] Qiyue He and Anatoliy Swishchuk. Quantitative and Comparative Analyses of Limit Order Books with General Compound Hawkes Processes. *Risks*, 7(4), 2019.

[18] Christian Hipp. Optimal Control for Insurance and Finance. 2016.

[19] Patrick J. Laub, Homas Taimre, and Philip K. Pollett. Hawkes Processes. 2015.

[20] P. A. W Lewis and Gerald S Shedler. Simulation of Nonhomogeneous Poisson Processes by Thinning. *Naval research logistics quarterly*, 26(3):403–413, 1979.

[21] Peter A. W. Lewis. A Branching Poisson Process Model for the Analysis of Computer Failure Patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(3):398–456, 1964.

[22] Thomas Josef Liniger. *Multivariate Hawkes Processes.* PhD thesis, Swiss Fed. Inst. Tech., Zurich, 2009.

[23] Anders Martin-LöF. Lectures on the Use of Control Theory in Insurance. *Scandinavian Actuarial Journal*, 1994(1):1–25, 1994.

[24] Robert C. Merton. Lifetime Portfolio Selection under Uncertainty: The Continuous-time Case. *The Review of Economics and Statistics*, 51(3):247–257, 1969.

[25] Jesper Møller and Jakob G Rasmussen. Perfect Simulation of Hawkes Processes. *Advances in applied probability*, 37(3):629–646, 2005.

[26] James R Norris and James Robert Norris. *Markov Chains.* Number 2. Cambridge university press, 1998.

[27] Yosihiko Ogata. On Lewis' Simulation Method for Point Processes. *IEEE transactions on information theory*, 27(1):23–31, 1981.

[28] T. Ozaki. Maximum Likelihood Estimation of Hawkes' Self-exciting Point Processes. *Annals of the Institute of Statistical Mathematics*, 31:145–155, 1979.

[29] Shige Peng. Stochastic Hamilton-Jacobi-Bellman Equations. *SIAM J. Control Optim.*, 30:284–304, 1992.

[30] Jinniao Qiu. Viscosity Solutions of Stochastic Hamilton–Jacobi–Bellman Equations. *SIAM J. Control Optim.*, 56(5):3708–3730, 2018.

[31] Anatoliy Swishchuk. Merton Investment Problems in Finance and Insurance for the Hawkes-based Models. *arXiv*, 2021.

[32] Anatoliy Swishchuk, Katharina Cera, Julia Schmidt, and Tyler Hofmeister. General Semi-Markov Model for Limit Order Books: Theory, Implementation and Numerics. *arXiv*, 2016.

[33] Anatoliy Swishchuk and Aiden Huffman. General Compound Hawkes Processes in Limit Order Books. *arXiv*, 2020.

[34] Anatoliy Swishchuk, Bruno Remillard, Robert Elliott, and Jonathan Chavez-Casillas. Compound Hawkes Processes in Limit Order Books. *arXiv*, 2017.

[35] Anatoliy Swishchuk and Nelson Vadori. A Semi-Markovian Modeling of Limit Order Markets. *arXiv*, 2016.

[36] Anatoliy Swishchuk, Rudi Zagst, and Gabriela Zeller. Hawkes Processes in Insurance: Risk Model, Application to Emperical Data and Optimal Investment. *Insurance: Mathematics and Economics*, 101(PA):107–124, 2021.

[37] Lingjiong Zhu. Nonlinear Hawkes Processes. *Journal of Applied Probability 2013*, 2013.

Appendix A

MATLAB code for simulations and graphs

```
% Simulation of Risk Process with Exponential Decay Intensity
"++++++++++++    Start of the program  +++++++++++++++"
N_hat=10; %10-state Markov chain
T= 260; % days in a year without weekends
N=260;  % Time discretization steps
L=1; %Number of simulation
astar=797.3672; gamma=1/astar;
r=0.06/T; mu=0.2/T; sigma=0.05/sqrt(T);
p=0.0001;
kappa=exp(r*T);
c=633.5552;
Rzero=8000; %initial capital
dt=(T-0)/N;
t=linspace(0,T,N+1);
G_sim=zeros(1,N+1);
G=zeros(L,N+1);
H=zeros(L,N+1);
lambda=0.1467; alpha=0.0260/T; beta=0.0334*T;
t0=0; ti(1)=t0;
```

```
%% finding equidistance boundary for claim state
b0=0; bn=10071;
diff=(bn-b0)/N_hat;
bi=zeros(1,N_hat);s=0;
bi(1)=diff;
for i=1:N_hat-1
    bi(i+1)=bi(i)+diff;
end
vpa(sym(bi), 6)
%% finding transition probability (pi)
pi=zeros(1,N_hat);
pi(1)=1- exp(-gamma*bi(1));
s1=0;
for j=1:N_hat-1
    s1=s1+pi(j);
    Gcap=(1-exp(-gamma*bi(j+1)));
    pi(j+1)=Gcap-s1;
end
format long
%pi
%sum(pi)
%% finding a(i), state value for claim size
fun=@(y) y.*gamma.*exp(-gamma.*y);
ai_1= (integral(fun,b0,bi(1)))/pi(1);
ai=zeros(N_hat,1);
ai(1)=ai_1;
for k=1:N_hat-1
    ai(k+1)= (integral(fun,bi(k),bi(k+1)))/pi(k+1);
end
vpa(sym(ai), 8)
%% L Simulations of Hawkes Process, Markov Chain, Claim Sizes and thus G
for sim=1:L
```

```
% Single Simulation of Hawkes Point Process HPp
HPp=[];  %Hawkes Point Process
U=rand;U1=[U];
ti(2)=t0-log(U)/lambda;
k=2;
while ti(k)<T
    HPp=[HPp,ti(k)];
    s1= lambda;
    for i=1:k
        s1=s1+alpha*exp(-beta*(ti(k)-ti(i)));
    end
    U2=rand;
    ti(k+1)=(-log(U2)+ti(k)*(s1))/(s1); %finding points from HP
    U1=[U1,U2];
    k=k+1;
end
HPp
HPp_0= [t0,HPp];   %Hawkes Point Process including 0
lenH= length(HPp);

% Simulation of claim sizes
aX_i=zeros(1,lenH);
for k=1:lenH
    bk=exprnd(astar);
    for j=1:N_hat
        if bk< b(j)
            aX_i(k)=ai(j);
            break;
        end
    end
end
aX_i
```

```
    % Single Simulation of process G(t)
    for j=1:N+1
        S=0;k=1;
        while t(j)> HPp(k)
          S=S+aX_i(k);
          k=k+1;
          if k>lenH
              break
          end
        end
        G_sim(j)=S;
    end
    G_sim;
    G(sim,:)=G_sim;
end
%% Figure for L simulations of G

figure()

y=zeros(1,lenH);
Hp=zeros(1,lenH);
for i=1:lenH
    y(i)=G(ceil(HPp(i)));
    Hp(i)=floor(HPp(i));
end
plot(Hp,y,"blue*")
hold on
for i=1:L
    plot(t,G(i,:),'r-','LineWidth',1)
end
legend('Hawkes Process Point','G(t)','AutoUpdate','off')
hold off
```

```
xlabel("t")
ylabel("G(t)")
%plot(t,G(6,:))
%% L Simulations of the process H(t)
for i=1:N+1
    s1=(c/r)*(1-exp(-r*t(i)));
    s2=transpose(zeros(1,L));
    for j=2:i
        %s3=G(:,j)-G(:,j-1);
        %s4=exp(-r*t(j))*s3;
        s2=s2+exp(-r*t(j))*G(:,j)-G(:,j-1);
    end
    H(:,i)=s1-s2;
end
%% Finding pistar & R_pi(t)from G(t)
dW=sqrt(dt)*randn(L,N);
R_pi=zeros(L,N+1);
pistar=zeros(L,N+1);
R_pi(:,1)=Rzero;
pistar(:,1)=((mu-r)*exp(r*t(1)))/((sigma^2)*R_pi(:,1)*kappa*p);
St_i=zeros(L,N+1);
St_i(:,1)=R_pi(:,1).*pistar(:,1);
for j=1:N
    R_pi(:,j+1) = R_pi(:,j) + r*R_pi(:,j)*dt +
    (((mu-r)^2)*exp(r*t(j))/((sigma^2)*kappa*p))*dt +
    (((mu-r)*exp(r*t(j)))/(sigma*kappa*p))*dW(:,j) + c*dt -(G(:,j+1)-G(:,j));
    pistar(:,j+1)=((mu-r)*exp(r*t(j+1)))*transpose(ones(1,L))
    ./((sigma^2)*R_pi(:,j+1)*kappa*p);
end
figure()

y=zeros(1,lenH);
```

```
Hp=zeros(1,lenH);
for i=1:lenH
    y(i)=R_pi(ceil(HPp(i)));
    Hp(i)=floor(HPp(i));
end
plot(Hp,y,"blue*")
hold on
for i=1:L
    plot(t,R_pi(i,:),'r-','LineWidth',1)
end
legend('Hawkes Process Point','R^\pi(t)','AutoUpdate','off')
hold off
xlabel("t")
ylabel("R^\pi(t)")

figure()
for i=1:L
    plot(t,pistar(i,:),'r.')
    hold on
end
hold off
xlabel("t")
ylabel("\pi^*(t)")

"++++++++++++    End of the program  +++++++++++++"


% Simulation of Risk Process with Power Law Intensity
"++++++++++++    Start of the program  +++++++++++++"
N_hat=10; %10-state Markov chain
T= 260; % 20 days in a year
N=260;          % Time discretization steps
L=1;    %Number of simulation
```

```
astar=797.3672; gamma=1/astar;
r=0.06/T; mu=0.2/T; sigma=0.05/sqrt(T);
p=0.0001;
kappa=exp(r*T);
c=633.5552;
Rzero=8000; %initial capital
dt=(T-0)/N;
t=linspace(0,T,N+1);
G_sim=zeros(1,N+1);
G=zeros(L,N+1);
H=zeros(L,N+1);
lambda=0.1467; m=0.005; d=0.5; q=2;
t0=0; ti(1)=t0;
%% finding equidistance boundary for claim state
b0=0; bn=10071;
diff=(bn-b0)/N_hat;
bi=zeros(1,N_hat);s=0;
bi(1)=diff;
for i=1:N_hat-1
    bi(i+1)=bi(i)+diff;
end
vpa(sym(bi), 6)
%% finding transition probability (pi)
pi=zeros(1,N_hat);
pi(1)=1- exp(-gamma*bi(1));
s1=0;
for j=1:N_hat-1
    s1=s1+pi(j);
    Gcap=(1-exp(-gamma*bi(j+1)));
    pi(j+1)=Gcap-s1;
end
format long
```

```
%pi
%sum(pi)
%% finding a(i), state value for claim size
fun=@(y) y.*gamma.*exp(-gamma.*y);
ai_1= (integral(fun,b0,bi(1)))/pi(1);
ai=zeros(N_hat,1);
ai(1)=ai_1;
for k=1:N_hat-1
    ai(k+1)= (integral(fun,bi(k),bi(k+1)))/pi(k+1);
end
vpa(sym(ai), 8)
%% L Simulations of Hawkes Process, Markov Chain, Claim Sizes and thus G
for sim=1:L
    % Single Simulation of Hawkes Point Process HPp
    HPp=[];  %Hawkes Point Process
    U=rand;U1=[U];
    ti(2)=t0-log(U)/lambda;
    k=2;
    while ti(k)<T
    HPp=[HPp,ti(k)];
    s1=0;
    for i=1:k
        s1=s1+m*(d+ti(k)-ti(i))^(-q);
    end
    U2=rand;
    ti(k+1)=(-log(U2)+ti(k)*(lambda+s1))/(lambda+s1); %finding Hawkes's process
    U1=[U1,U2];
    k=k+1
end
HPp
HPp_0= [t0,HPp];  %Hawkes process including 0
lenH= length(HPp);
```

```
    % Simulation of claim sizes
     aX_i=zeros(1,lenH);
     for k=1:lenH
         bk=exprnd(astar);
         for j=1:N_hat
             if bk< b(j)
                 aX_i(k)=ai(j);
                 break;
             end
         end
     end
     aX_i
     % Single Simulation of process G(t)
     for j=1:N+1
         S=0;k=1;
         while t(j)> HPp(k)
           S=S+aX_i(k);
           k=k+1;
           if k>lenH
               break
           end
         end
         G_sim(j)=S;
     end
     G_sim;
     G(sim,:)=G_sim;
end
%% Figure for L simulations of G
y=zeros(1,lenH);
Hp=zeros(1,lenH);
figure()
```

```
for i=1:lenH
    y(i)=G(ceil(HPp(i)));
    Hp(i)=floor(HPp(i));
end
plot(Hp,y,"blue*")
hold on
for i=1:L
    plot(t,G(i,:))
end
legend('Hawkes Process Point','G(t)','AutoUpdate','off')
hold off
xlabel("t")
ylabel("G(t)")
%plot(t,G(6,:))
%% L Simulations of the process H(t)
for i=1:N+1
    s1=(c/r)*(1-exp(-r*t(i)));
    s2=transpose(zeros(1,L));
    for j=2:i
        %s3=G(:,j)-G(:,j-1);
        %s4=exp(-r*t(j))*s3;
        s2=s2+exp(-r*t(j))*G(:,j)-G(:,j-1);
    end
    H(:,i)=s1-s2;
end
%% Finding pistar & R_pi(t)from G(t)
dW=sqrt(dt)*randn(L,N);
R_pi=zeros(L,N+1);
pistar=zeros(L,N+1);
R_pi(:,1)=Rzero;
pistar(:,1)=((mu-r)*exp(r*t(1)))/((sigma^2)*R_pi(:,1)*kappa*p);
St_i=zeros(L,N+1);
```

```
St_i(:,1)=R_pi(:,1).*pistar(:,1);
for j=1:N
    R_pi(:,j+1) = R_pi(:,j) + r*R_pi(:,j)*dt
    + (((mu-r)^2)*exp(r*t(j))/((sigma^2)*kappa*p))*dt
    + (((mu-r)*exp(r*t(j)))/(sigma*kappa*p))*dW(:,j) + c*dt -(G(:,j+1)-G(:,j));
    pistar(:,j+1)=((mu-r)*exp(r*t(j+1)))*transpose(ones(1,L))
    ./((sigma^2)*R_pi(:,j+1)*kappa*p);
end
figure()

y=zeros(1,lenH);
Hp=zeros(1,lenH);
for i=1:lenH
    y(i)=R_pi(ceil(HPp(i)));
    Hp(i)=floor(HPp(i));
end
plot(Hp,y,"blue*")
hold on
for i=1:L
    plot(t,R_pi(i,:),'r-','LineWidth',1)
end
legend('Hawkes Process Point','R^\pi(t)','AutoUpdate','off')
hold off
xlabel("t")
ylabel("R^\pi(t)")

figure()
for i=1:L
    plot(t,pistar(i,:),'r.')
    hold on
end
hold off
```

```
xlabel("t")
ylabel("\pi^*(t)")

"++++++++++++      End of the program  ++++++++++++++"
```