2022-03

Merton Problem in Insurance

Fooladamoli, Ehsan

UNIVERSITY OF CALGARY

Merton Problem in Insurance

by

Ehsan Fooladamoli

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

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

MARCH, 2022

Abstract

The goal of Insurance companies, like that of any other financial institution, is to maximize their wealth. In doing so, there are different parameters they have to consider, such as premium rate, number of claim arrivals, size of claim arrival, etc. Moreover, they can invest their money in risk-free and risky asset to earn some income from those resources as well. This thesis discusses the application of Merton problem in insurance and risk and how to solve it. That is, we design a trading strategy for an insurance company such that its utility is maximized over a given time horizon. We use General Compound Hawkes Process to model the insurance's risk and use the corresponding diffusion approximation to approximate the risk using a diffusion process. Then, we proceed with solving the problem by Hamilton-Jacobi-Bellman equation. Finally, we show some simulation results based on the calibration on data from insurance companies in Germany and their interpretations.

Acknowledgements

With great appreciation to my supervisor, Anatoliy Swishchuk, without whose insight and direction this thesis would not have been possible. I feel truly fortunate to have worked with him. He guided me with unending patience and never fell short of great ideas when we hit hurdles in our research together. My special thanks go to Dr. Wenyuan Liao and Dr. Rohana Ambagaspitiya for accepting to be my committee members. I am most grateful for all the love and support to my parents. Words cannot describe how lucky I feel to have you as my parents. I would also like to thank my brother, Hossein, who has always been not just a brother, but also a great friend and someone I could rely on. Last but not least, huge thanks to Fasiha Binat Zafar and Yang Yu who have been always there for me and are best friends one could ever wish for.

Table of Contents

List of Figures and Illustrations

List of Symbols and Abbreviations

Symbol or abbreviation	**Definition**
r.v	Random variabl
i.i.d	independent and identically distributed
a.s	Almost surely
Ω	Sample space
$\mathbb{N}$	natural numbers
LLN	Law of large numbers
CLT	Central limit theorem
p.d.f	Probability density function
c.d.f	Cumulative distribution function
$W(t)$	Weiner process
SDE	Stochastic differential equation
GCHP	General compound Hawkes process
RMGCHP	Risk model based on general compound Hawkes process
λ	background intensity
μ	self-excitig function
HJB	Hamilton − Jacobi − Bellman equation
DPP	Dynamic programming principle

Chapter 1

Introduction

1.1 Overview

Insurance companies are financial intermediaries whose job is to compensate partly or fully for a realized risk one may encounter, what is called a claim. In return , they receive some amount of money, namely, the premium, on a regular basis. The portion of the damage and the type of damages they cover is already determined in a contract between the company, namely the insurer, and the customer, namely the policy owner. For example, all else equal, a health insurance pays for prespecified health related expenses the policy holder may face, but not for a car accident, a life insurance in return, pays an prespecified amount of money to their named beneficiaries in the contract, when the policy holder dies. Since the insurer only pays if a policy holder faces a realized risk, in other words, the insurer faces a claim arrival, the main expense of the company is related to these claims. Also the inflow of the money is from the premiums the insurer receives. As a result, the company's goal is not only to be in profit, that is, the overall premium it receives exceeds the overall claim arrivals, but also it wishes to maximize this profit.

The way an insurance company works, is by what is called the risk pooling. In that, given that the company has an idea of how often a typical claim arrives and how large it is, they spread these expenses over all the policy holders, this is technically done by the premium the insurer receives. As a result, overestimating the overall claim occurrences, causes the insurer to set the premium high and thus, the demand for the insurance will decrease which results in decrease in potential income or even loss for the company. If the insurer underestimates the overall claim occurrences, it will result in a too low premium which may not really cover the claim occurrences and causes the insurer lose a lot of money or even go bankrupt. In this regard, some mathematical models regarding the risk the insurance companies are exposed to came into existence. The early types of risk models known as classical risk models, rely on individual risk policies individual by individual. In contrast, the collective risk theory, build the stochastic model based on all policies as a whole, so we do not care about the distribution of individual policies anymore but we encapsulate all these details in a distribution of all claim occurrences, regardless of which policy they are coming from. In this thesis, we use the collective risk theory to model the insurance risk (Burney and Hashmi, 2002). When it comes to model the claim occurrences, there are two things we need to take into account; the occurrence time and the size of the claims, this is because it's is not just the size of the claims that is random, but also the time of their occurrence is uncertain. The importance of the time of claim occurrences becomes more clear when we talk about the application of Merton problem in insurance. Once we have our model for the dynamic of the wealth of the company, we have the elements to theoretically measure the risk of the insurance company. A common measurement for an insurance company's risk, is the ruin probability, that is, the probability of the net wealth of the company becomes negative. Computing the ruin probability of a model in general is not usually simple in practice as we will see later except for special cases where we can derive a closed form solution.

As we mentioned earlier, the inflow of the money for the insurance company was from the premiums they receive and the outflow is from the claim occurrences they pay for. However, we know that big companies don't deposit all their money in banks in real life but they invest in financial markets too. Though risky, the expected gain from financial markets is usually more than that of a bank deposit otherwise, based on the no arbitrage principle, economic agents would invest their money to the bank account and there would be no demand for the financial markets other than risk-free securities. The final goal in this thesis is to design a trading strategy for the insurance company for which the expected utility function is maximized. In order to do so, we first need to model the risk which was as a stochastic process.

The next step, is to model the investment. We do so, by considering our derived stochastic process as a sum of two components at each time, one show the total amount invested in a risky asset and the other is the total amount of the wealth invested in a risky asset and we assume these two are all investment possibilities. Provided that we know the probability distribution of the price of the risky asset, we then define a control variable that show what portion of the whole wealth should be devoted to risky asset at each time in order to maximize utility up to some time T. This way, this control variable, if could be found, is a function of time and all the realization of our stochastic process up to each time t, gives us a trading strategy. Such problems of the optimal trading strategy is called the Merton's problem and finding an optimal function to maximize the expected value of a stochastic process is called an stochastic optimal control problem.

In order to solve an optimal control problem, we need to derive a stochastic differential equation of the response variable with respect to the optimal control function and then use Hamilton-Jacobi-Bellman equation to solve it (Cartea, Jaimungal, and Penalva, 2015). The HJB is a dynamic programming approach to solve a problem which will be discussed in more detail in this thesis. We will see that some complexities arise if we want to apply HJB directly to our model as the corresponding SDE becomes too complex to work with. Instead, we use a diffusion approximation, a model that asymptotically behaves the same way and is a function of Brownian motion and thus we could use the tools from stochastic calculus. Finally, we will show the simulation result based on the derived solutions and show their interpretations.

1.2 Literature review

As discussed earlier, we should start with a risk model before any further development. One of the earliest models of a collective risk theory was first proposed by Filip Lundberg in 1909. This risk model is as follows (Burney and Hashmi, 2002),(Lundberg, 1903):

$$\begin{aligned} R(t) &= s + ct - S(t) \\ S(t) &= X_1 + ... + X_{N(t)} = \sum_{i=1}^{N(t)} X_i \end{aligned} \tag{1.1}$$

Where the X_i denote the claim arrivals and are i.i.d r.v coming from some distribution Q and N(t) is a Poisson process with a constant intensity $\lambda > 0$. The parameter s is the initial capital of the company and c is the premium rate. Since the number of policy holders is usually large and the, the premium rate could be modelled as continuous inflow of money. As a result, If we let W_i denote the waiting time between two consecutive claim arrivals, then $W_i \sim Exp(\lambda)$. Later on, many contributions were made to collective risk theory by Cramer (Cramér, 1955), Segerbahl (Segerdahl, 1959), Arfwedson (Arfwedson, 1955),(Arfwedson, 1950) and many others. A more sophisticated model was later proposed by Andersen (Thorin, 1974) and E. Sparre, which is called the Sparre-Andersen model. In the Sparre-Andersen risk model, the waiting times between consecutive arrivals, could be any renewal process, whereas the claims sizes X_i are i.i.d and come from some arbitrary distribution. Note that the difference between Sparre-Andersen and the Lundberg risk process, is that the number of arrivals up to time t is a Poisson process which is indeed a renewal process but only one special case, while in the Sparre-Andersen any renewal in addition to Poisson process works as well. As a result, Lundberg risk model is a special case of Sparre-Andersen. Another risk developed is called the Markov-modulated risk process (Hipp, 2004). In the Markov-modulated model, the number of arrivals up to any time t, is a Poisson process but the corresponding λ is itself, stochastic. In that, the intensity at time t is $\lambda_M(t)$, where M(t) is a finite state Markov chain. These models in spite of working well in some cases, have a big disadvantage, they all assume the arrival of claims to be independent while in real world it is not usually the case. A more comprehensive model which takes into account the dependence of arrivals on all the previous ones was first developed by Stabile and Torrisi in 2010 (Stabile and Torrisi, 2010). In this model, the number of arrivals N(t) is a Hawkes process. We will talk in detail what a Hawkes process is and the model based on it. Finally, the most generalized version of the risk model was developed by Swishchuk in 2017 (Swishchuk, Remillard, et al., 2019) in which in addition to the number of arrivals that follow a Hawkes process, the claim sizes themselves are an ergodic Markov chain. This model is named RMGCHP. This allows that we model temporal clusters of arrivals and the dependence of the claim sizes themselve.

Now that we have seen the examples of collective risk models, we should have a measurement for the risk itself, since the models we have seen so far only show the probabilistic dynamic of the wealth of the company, not what exactly the risk itself is. Ruin probability is a metric for the risk of an insurance company and for a risk model $R(t) = s + ct - \sum_{i=1}^{N(t)} X_i$ is defined as:

$$\phi(s) = P\{R(t) < 0 \quad for \quad some \quad t \geq 0\} \tag{1.2}$$

Which means, the probability that the net wealth of the company becomes negative at some point in time. In general, deriving a closed form for ruin probability is very hard, if not impossible. For the Lundberg risk model, if the claim sizes come from an exponential distribution, the ruin probability is (Grandell, 1991)

$$\phi(s) = \frac{\lambda\mu \; exp(-Rs)}{c} \tag{1.3}$$

where λ, μ, R and c are the parameter Poisson distribution's parameter for arrivals, the average claim size and $R = (c - \lambda\mu)/c\mu$ and the premium rate respectively. For RMGCHP, Swischuk (Swishchuk, Remillard, et al., 2019) derived a diffusion approximation, in which the original risk model is approximated by a sequence of jump-diffusion processes that in the limit is a Brownian motion with drift, for the RMGCHP based on which he computed the ruin probability for special choices of the self-exciting function. The formula is mentioned later in the thesis. The approximation based on the Brownian motion gives us a lot of convenience both in computing the ruin probabilities and the stochastic optimal control of the process. With our stochastic models of insurance in hand, the goal is to solve an optimal control for the model. The first applications of optimal control in insurance could be found in (Brockett and Xiaohua, 1997),(Browne, 1995),(Martin-Löf, 1994). By optimal control, we simply mean that based on the information up to each time t, how should we change the so called control variables to optimize the expected value of some function of a stochastic random variable. By control variable, we mean literally any variable whose value at each time could be controlled and modified. For example, in our risk model $R(t) = s + ct - \sum_{i=1}^{N(t)} X_i$, we can change the premium rate at each time, or the initial capital. In our problem, we consider the premium rate to be constant, however, we aim to control the ratio of the whole wealth invested in risky asset and risk-free asset. These types of problems where we have a model for wealth and want to invest the money dynamically in some assets in order to maximize some utility function were first studied by Merton (Merton, 1969),(Merton, 1975). More details and examples of stochastic optimal control problems could be found in ((W. Fleming et al., 1975),Wendell H Fleming and Soner, 2006,(Karatzas et al., 1998),(Cartea, Jaimungal, and Penalva, 2015)). All the previous papers use the early models in insurance like Lundberg risk model to solve an optimal control problem. In this thesis, we solve the Merton problem for the most general case of insurance, namely, RMGCHP and then we show some numerical results and how they could be interpreted.

Chapter 2

Stochastic Processes

2.1 Definitions and examples

Using probability tools in any mathematical model that involves uncertainty is inevitable as probability is the language of uncertainty. Sometimes we are interested in measuring the uncertainty of a simple random experiment like what is the chance of winning in a lottery, what is the chance of getting a head when tossing a coin. In these examples, our believes about the outcome of the experiment does not usually change in time since the outcome will be determined at a specific moment and as the experiment could be done in a moment. In other cases however, we may keep updating our believes about the outcome of a random system and wish to be able to measure our beliefs of the uncertainties in that regard. For example, let's assume we are interested in the price of a stock in a period of time and at each point in time, based on the trend and patterns we have been observing up to that time, we decide whether we should buy or not. In this example, the trend of the stock itself so instead of single outcome of a experiment that happens in a moment we are interested in the way the system evolves and thus, we need some more advance tools than basic probability to be able to measure the uncertainties of different scenarios. Informally speaking, a stochastic processes is the name given to any probabilistic system that evolves in time. In this section we introduce some of the most useful stochastic processes models and their properties.

Stochastic Process (Ross, 2014)

A stochastic process is a collection of random variables X_t defined over some probability space (Ω , $\mathcal{F}$, P) where Ω is the sample space, $\mathcal{F}$ is a sigma algebra on the sample space and P is a probability measure and t indicates the time index and takes on values usually in $[0, \infty)$ or the integer numbers. If the time index comes from $[0, \infty)$, the stochastic process is said to be discrete-time and if the time index are only integers the stochastic process is said to be continuous-time.

Counting process (Higgins and Keller-McNulty, 1995)

Counting process a continuous-time stochastic process $N(t)$ is said to be a counting process if

- $N(0) = 0$ a.s
- $N(t)$ is a non-negative integer
- $N(t)$ is increasing a.s the increments are of size 1 and is a.s. finite

So a counting process is any process that says the number of occurrences of an event up to time t, so for example the number of people who visit a store on specific day in a week is a counting process or the number of car accidents from today till a year from now is a counting process.

Point process (Daley and Vere-Jones, 2003)

Point process a point process is a collection of random variables $\{t_i\}$ where i is an integer and $t_i \leq t_j$ if $i \leq j$ a.s.

Basically, a point process could be though of as the occurrence times of an even, so for example, if t_i denotes the i^{th} person who visits a specific store in a given day, then the collection t_i is a point process. We could see how any counting process $\{t_i\}$ could induce a point process by setting $N(t_i) = i$ and vice versa, that is, any counting process could also induce a point process by keeping record of the time of each occurrence in order and set t_i to be the time of the i^{th} occurrence.

Example: (Homogeneous Poisson process) A Poisson process is a counting process usually denoted as a collection of random variables $\{N(t), t \geq 0\}$ where $N(t)$ denotes the total number of occurrences of events up to time t and has three properties:

- $N(0) = 0$
- Has independent increments, namely, for any non-negative values t and s such that $t \geq s$, $N(t) - N(s)$ is independent of $N(s)$ and has the same distribution as $N(t-s)$
- The number of occurrences in any time interval of length t is a Poisson random variable with parameter λt.

The distribution of a homogeneous Poisson process with parameter λ is as follows:

$$P(N(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, k = 0, 1, 2... \tag{2.1}$$

There is another way to define the Poisson process and it could be shown that these two definitions are equivalent (Ross, 2014).

Poisson Process

A counting process is said to be a Poisson process with parameter λ if:

$$\begin{aligned} &N(0) = 0 \\ &P(N(t) = 1) = \lambda h + o(h) \\ &P(N(t) \geq 2) = o(h) \end{aligned} \tag{2.2}$$

A non-homogeneous Poisson process is the same as the homogeneous one except for the third property, the parameter depends on which time interval we are considering and not just the length of it.

Example (Renewal process): Let $\{N(t), t \geq 0\}$ be a counting process and let tn denote the time between the $(n-1)$ and nth occurrence. If the collection of random variables $\{t_i, i \in N\}$, are i.i.d , then the counting process $N(t)$ is said to be a renewal process. In fact, renewal process is a generalization of the homogeneous Poisson process except for the fact that the waiting time between two consecutive arrivals does not have to be exponential, so a renewal process whose waiting times are exponentially distributed is Poisson.

In all previous counting process examples, we characterized the processes by the distribution of their waiting times. Another way to do so, is by what is called the conditional intensity function (Hawkes, 1971):

Conditional Intensity

Let $N(t)$ be a counting process and let $H(t)$ denote the history of arrivals up to time t. If the random variable $\lambda^*(t)$ exists such that :

$$\lambda^*(t) = \lim_{h \to 0} \frac{E[N(t+h) - N(t)|H(t)]}{h} \tag{2.3}$$

Then $\lambda_{*(t)}$ is said to be the conditional intensity function of the counting process $N(t)$.

Based on the linearity of expectation, we can take the h inside the expectation. By doing so, it could be seen that the conditional intensity translates to the expected rate of arrivals at time t, given the history of arrivals $H(t)$ up to time t. Thus, $\lambda^*(t)$ itself could be random as the history of arrivals up to time t is random. it was shown that the conditional intensity function of a counting process $N(t)$ uniquely determines the distribution of $N(t)$. Now let's see what this function is for a Poisson process with parameter lambda:

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

$$= \lim_{h\to 0} \frac{E[N(h)]}{h}$$
$$= \lim_{h\to 0} \frac{I_x\big(\lambda h + o(h)) + o(h)\big)}{h} \tag{2.4}$$
$$= \lambda$$

Where the second equation resulted from the independent increments, the third equation resulted from the stationary property of Poisson process. As we mentioned earlier, the conditional intensity function is the expected rate of arrival at time t given the history of arrivals up to that time. So the parameter lambda is exactly the expected rate of arrivals. Since in this case the conditional intensity function turned out to be a constant, it means that the expected rate of arrivals does not depend on the time t nor the history of arrivals $H(t)$. This should not come at any surprise, since the independence of time results from the stationary property of the Poisson process and the independence from the $H(t)$ is a result of independent increments.

As we saw earlier, stochastic processes are a way to model how a system evolves in time from the probabilistic point of view. Some models were mentioned as well. For example, Poisson processes are a good fit when we want to model the number of arrivals in a system where the waiting time between any two consecutive arrivals are independent of all previous arrival times, say, starting from some point in time, the number of people who have entered a specific bant up to some time t. Weiner processes are very suitable when one wants to model a noise in a system in continuous time. In all previous examples however, a new arrival did not depend on the full history or at least more than one previous arrival, in other words, the history of arrivals does "excite" or "regulate" the future ones. In real world problems though, we may encounter systems where this does not seem like a realistic model. For example, when an earthquake hits, we expect some more around the time of the first one, in other words, the more frequent earthquakes happens in a period time, the more likely they will happen a gin in near future (Ogata, 1988). Another example would be limit orders of some trending stock, when a stock becomes trending and keeps increasing in price, more and more people get interested in investing on it and more and more limit orders will happen (Swishchuk, Remillard, et al., 2019), this was also seen in the cryptocurrencies market in 2020 as the price of bitcoin was increasing more and more people got interested in this markets and started to trade cryptos. In what follows, we go through another important probability model called Hawkes process (Hawkes, 1971), where it takes into account the effect of previous occurrences on the future ones. In fact, Hawkes process is and extension of the Poisson process. As we saw earlier, the conditional intensity function is the expected rate of arrivals and we saw that this quantity is always a constant for the Poisson process regardless of the history of arrivals or the time t. In the Hawkes process however, we aim to manipulate the definition of the Poisson process in such a way that this time instead of having a constant expected arrival rate, we get expected arrival rates at each point of t by incorporating all the history of past as all of them have an effect on future arrivals more or less. Also we want the recent arrivals to have bigger impact on next arrivals. We see how we can encapsulate these ideas in what's called the self-exciting function. The job of self-exciting function is to quantify the effect each arrival has on the next potential arrivals, both in terms of how much more probable next arrivals will be and the amount of time this effect lasts. We then show some famous examples of Hawkes process. Though Hawkes process is really sophisticated and makes use of a lot of information, using it in practice can sometimes be cumbersome. That being said, there are many useful theorems regarding Hawkes process which let us take advantage of this stochastic process in modelling different systems with such processes or somehow approximate the formulation of systems who evolve like a Hawkes process by more well-behaved processes. In fact, we will see that how the self-exciting functions can also play an important role in encoding the important information when we want to approximate a Hawkes process (Hawkes, 1971).

Hawkes process

Let $N(t)$ be a counting process with history of arrivals $H(t)$ up to time t such that:

$$P(N(t+h) - N(t) = k) = \begin{cases} \lambda^*(t)h + o(h) & k = 1 \\ o(h) & k \geq 2 \\ 1 - \lambda^*(t)o(h) + o(h) & k = 0 \end{cases} \tag{2.5}$$

Where $\lambda^*(t)$ is the conditional intensity function. Also assume that $\lambda^*(t)$ has the following form:

$$\lambda^*(t) = \lambda + \int_0^t \mu(t-s)dN(s) \tag{2.6}$$

Where λ is a positive constant called the background intensity and $\mu(t)$ is a non-increasing function $(0, \infty) \to [0, \infty)$. Then $N(t)$ is a Hawkes process.

The function $\mu(t)$ in the definition, is called the self-exciting function and as it comes from its name, this is the function used to model the "self-excitement" of the arrivals and incorporate all the past information. In the Poisson process, we do not expect any arrival affect the future ones, thus, intuitively speaking, we expect the self-exciting function to be indifferent about the history of arrivals and the integral stays the same no matter how the arrivals were arranged. This is only possible if $\mu(t)$ is always zero. By setting the self-exciting function it could easily be seen that the conditional intensity stays a constant as we would have expected. Since Hawkes process is a counting process, the integral term in the definition could be simplified as follows:

$$\begin{aligned}\lambda^*(t) &= \lambda + \int_0^t \mu(t-s)dN(s) \\ &= \lambda + \sum_{t_i<t} \mu(t-t_i)\end{aligned} \tag{2.7}$$

Now it seems reasonable why the self-exciting function has to be non-increasing. We want the effect of the recent arrivals be more than others, so if for example $t_1 < t_2$ are two arrival times and let t be such that $t_1 < t_2 < t$, then the effect of the arrival at time t_2 on the probability of arrival at time t is somehow porportional to $\mu(t-t_2)$ and the effect of the arrival time t_1 on the probability of arrival at time t is somehow proportional to $\mu(t-t_1)$ and since t_2 is considered to be more recent, then we expect its effect on the arrival at time t to be as much as that of t_1 if note more, and all these translates to:

$$\begin{aligned}&t_2 > t_1 \iff t-t_2 < t-t_1 \\ &\mu(t-t_1) \leq \mu(t-t_2) \implies \mu \text{ non-increasing.}\end{aligned} \tag{2.8}$$

So every Hawkes process is determined by a background intensity constant λ and a self-exciting function.

Based on the definition of self-exciting function, any choice of a non-negative non-increasing function will do. In real applications however, when we model a probabilistic system as a counting process we usually would not expect a typical realization of the number of arrivals to be infinite, what is called the stationarity property of a counting process.

Stationary condition for Hawkes process (Bremaud and Massoulie, 1996)

Let $N(t)$ be a Hawkes process and $\mu(t)$ be its associated self-exciting function. Then $N(t)$ is said to be stationary if:

$$\hat{\mu} = \int_0^\infty \mu(s)ds < 1 \tag{2.9}$$

We will show later how this property makes sense in the Immigration-birth representation of the Hawkes process section.

Example(Power Law self-exciting function (Laub, Taimre, and Pollett, 2015)): The Power Law self-exciting function is a function $\mu(t)$ of the form :

$$\mu(t) = \frac{k}{(c + (t - s))^{1+n}} \tag{2.10}$$

Where $k > 0$, $c > 0$ and $n > 0$ are the parameters of the function. Herby the corresponding function would be:

$$\begin{aligned} \lambda^*(t) &= \lambda + \mu(t) \\ &= \lambda + \int_0^t \mu(t) d\mu(t) \\ &= \lambda + \sum_{t_i < t} \frac{k}{(c + (t - t_i))^{1+n}} \end{aligned} \tag{2.11}$$

Where t_i denote the arrival times up to time t. In order for Hawkes process corresponding to the power law excitation function to be stationary we should have:

$$\begin{aligned} \int_0^\infty \mu(t) < 1 &\iff \int_0^\infty \frac{k}{(c+t)^{1+n}} dt < 1 \\ &\iff -\frac{k}{n(c+t)^n}\Big|_{t=0}^{t=\infty} = 0 + \frac{k}{nc^n} < 1 \\ &\iff \frac{k}{nc^n} < 1 \end{aligned} \tag{2.12}$$

Example (Exponential decay self-exciting function (Laub, Taimre, and Pollett, 2015)): The exponential decay self-exciting function is a function $\mu(t) = \alpha e^{-\beta t}$ where α and β are non-negative parameters of the function. The stationary condition would be:

$$
\begin{aligned}
\int_0^\infty \mu(t) < 1 \iff & \int_0^\infty \alpha e^{-\beta t} dt < 1 \\
\iff & -\frac{\alpha e^{-\beta t}}{\beta}\Big|_{t=0}^{t=\infty} < 1 \\
\iff & \frac{\alpha}{\beta} < 1 \\
\iff & \alpha < \beta
\end{aligned}
\tag{2.13}
$$

The corresponding conditional intensity is of the form:

$$
\lambda^*(t) = \lambda + \alpha \sum_{t<t_i} e^{-\beta(t-t_i)} \tag{2.14}
$$

Intuitively speaking, α represents the magnitude of the effect of each arrival and β represents how fast the effect goes away, so all else equal, the larger the alpha is, the more likely each arrival makes the next ones and the smaller the β is the more the effect of each arrival lasts.

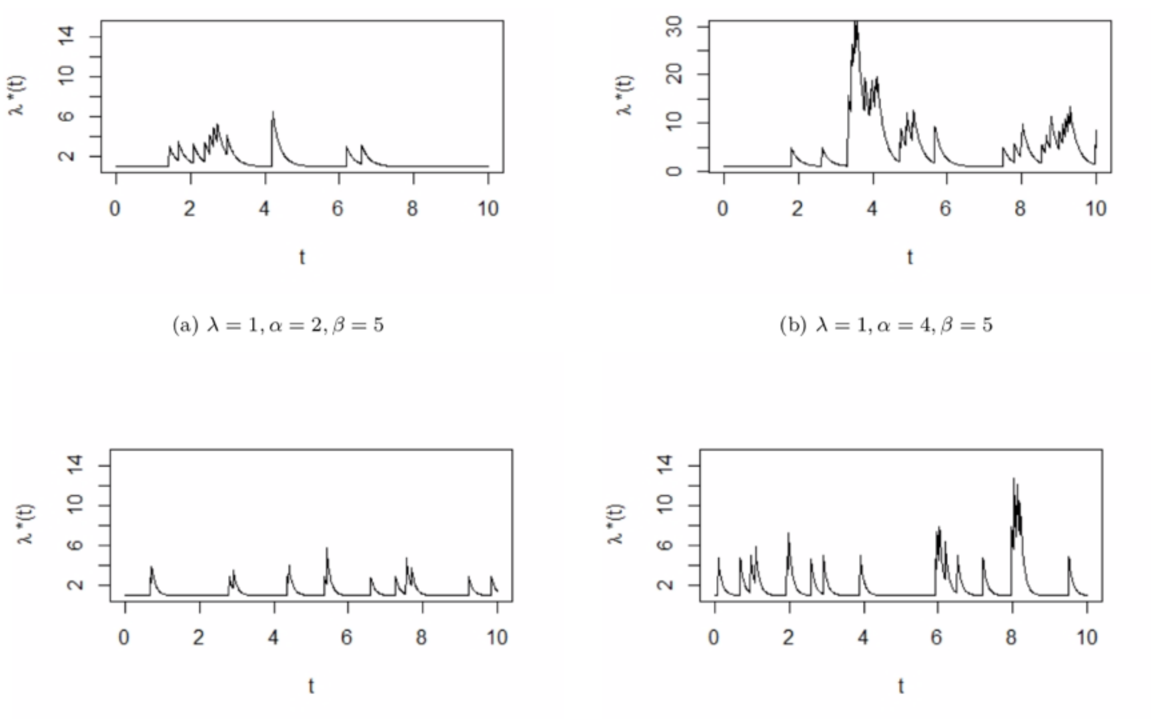
Since the exponential decay is linear with respect to α and exponential with respect to β, by scaling α and β by a same number, the intensity will increase much more as oppose to the case of. β. This is also illustrated in the following graphs.

Plotting some realizations of the conditional intensity by exponential decay self-exciting function. The background intensity in all the four figures is the same and equals to 1.

2.2 Immigration – birth representation of the Hawkes process

One way of looking at the binomial distribution is to see it as a set of Bernoulli trials. We do this because then we could use the linearity of expected value function and working with Bernoulli distribution is easy as it gets. In this chapter we do something similar for Hawkes

Figure 2.1: Some trajectories of Hawkes process with exponential self exciting funtction (G. F. Zeller, 2018)



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

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

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

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

process by presenting an equivalent definition of Hawkes process. In one hand, the best thing about Hawkes processes is that it takes all the information into account and in many real world application it is really the case that a lot of past information is important, but the downside is, however, that using the previous definition it gets really hard sometimes to gain insight about the process and some computations get really cumbersome. As mentioned earlier, Hawkes process is an extension of the Poisson process, so if we aim to think of it as a set of simpler processes, the Poisson process should sound like a good candidate. We do this by representing the Hawkes process as a collection of in-homogeneous Poisson processes. This is in particular useful as it gives us an idea as to how to simulate the Hawkes process. Also, it helps gaining intuition about where the stationary condition comes from.

Proposition: Immigrant – birth representation of Hawkes process (Bacry, Mastromatteo, and Muzy, 2015)

Let λ be a real number and μ be a self - exciting function. Consider a point process on a time interval $[0, T]$, where $T \in (0, \infty]$, such that:

- A set of immigrants $\{t_i^0\}$ arrive according to a Poisson process with parameter λ on the interval $[0, T]$.
- Each immigrant t_i^0 , independently from other ones, generates a sequence of arrivals itself $\{t_i^{0j}\}$ according to a in-homogeneous Poisson process with the time dependent parameter lambda $\mu(t - t_i^{0j})$. We call this generated sequence of arrivals by the immigrants, the first generation.
- This procedure carries from one generation to the next one until there are no more arrivals in the interval $[0, T]$. Now if we consider the set of all arrival times regardless of their generations, this corresponds to a Hawkes process with the background intensity λ and the self-exciting function $\mu(t)$.

Now that we have represented the Hawkes process as a branching process, the number of over all arrivals is equal to the number of the immigrants plus the number of their first generation offsprings plus the number of the second generation offsprings and so on. If we denote by Z_i the number of offsprings of generation i of an immigrant, the expected number of total offsprings of this immigrant would be $Z_1 + Z_2 + \cdots = \sum_i^{\infty} Z_I$. Now based on the linearity of the expected value function we have :

$$E[\sum_{i=1}^{\infty} Z_i] = \sum_{i=1}^{\infty} E[Z_i] \tag{2.15}$$

It can be shown that (Grimmett and Stirzaker, 2020)

$$E[Z_i] = \hat{\mu}^i \tag{2.16}$$

Where $\hat{\mu} = \int_0^\infty \mu(t)dt$ and $\mu(t)$ is the corresponding self-exciting function. Note that this quantity $\hat{\mu}$ has been seen before in previous section. Recall that the sufficient condition for stationarity of Hawkes process was $\hat{\mu} < 1$. Now we could interpret this condition as the following

$$\begin{aligned} E[\sum Z_i] &= \sum_{i=0}^{\infty} E[Z_i] \\ &= \sum_{i=0}^{\infty} \hat{\mu}^i \\ &= \begin{cases} \frac{\hat{\mu}}{1-\hat{\mu}} & \text{if } \hat{\mu} < 1 \\ \infty & \text{if } \hat{\mu} \geq 1 \end{cases} \end{aligned} \tag{2.17}$$

Since the immigrants arrive based on a Poisson process with parameter *lambda* and the number of arrivals in a finite interval for a Poisson process is a.s finite, then if each immigrant has a finite number of off-springs on average, then the overall number of arrivals in a finite interval will be finite and this is what stationarity means. On the other hand, based on what we just showed, the number of off-springs of each immigrant is finite only when $\hat{\mu} < 1$ and $\hat{\mu} < 1$ if and only if $\hat{\mu}^I = E[Z_i] < 1$ for each $i \geq 1$,that is, if each offspring gives birth to less than one child on average.

Finally we state two important theorems for Hawkes process in this chapter that were derived from the immigrant-birth representation of the Hawkes process.

Theorem (Law of large numbers of Hawkes process) (Laub, Taimre, and Pollett, 2015),(Valkeila, 2008)

Let $N(t)$ be a Hawkes process such that $N(t)$ satisfies the stationarity condition. Then

$$\lim_{h\to\infty} \frac{N(t)}{t} \to \frac{\lambda}{1-\hat{\mu}} \tag{2.18}$$

Central limit theorem for Hawkes process (Laub, Taimre, and Pollett, 2015)

Let $N(t)$ be a stationary Hawkes process with conditional intensity $\lambda^*(t)$ such that:

$$0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1 \text{ and } \int_0^\infty s\mu(s)ds < \infty, \tag{2.19}$$

then the number of Hawkes arrival in $(0, t]$ is asymptotically normally distributed, that is:

$$\mathbb{P}\Big(\frac{N(t) - \frac{\lambda}{1-\hat{\mu}}t}{\sqrt{\frac{\lambda}{(1-\hat{\mu})^3}t}} \leq y\Big) \xrightarrow{t\to\infty} \mathbb{P}(Z \leq y), \tag{2.20}$$

where Z is a standard normal random variable.

Chapter 3

Risk Theory

This section is devoted to an introduction to risk theory, in particular, risk related to insurance companies and how it could be mathematically modelled. The insurance company sells a contract based on which the person who buys the contract, known as the policy holder, pays a premium to the company on a regular basis (monthly, yearly, etc.) and in return, the company guarantees to compensate for the loss that the policy holder might run into. The amount of money that the company is willing to pay and the circumstances under which the insurer will compensate for the loss are all mentioned in the contract.

At its most abstract level, the net wealth of an insurance company is a function of three factors; the initial capital, the inflow of income from the risk premium the company receives and the overall claims the company has to pay for. In what follows we introduce the rigorous mathematical model of insurance companies and some classic models and then we proceed to a recently developed model called risk model with general compound Hawkes process.

3.1 Classical models

The general model for the net wealth of the insurance at time t could be denoted as (Hipp, 2004):

$$R(t) = s + ct - \sum_{i=1}^{N(t)} X_i \tag{3.1}$$

Where $R(t)$ denotes the net wealth of the insurance company at time t, s is the initial capital, c is the premium rate, $N(t)$ is the number of claim arrivals up to time t and X_i is the size of the $\mathrm{i_{th}}$ claim arrival. Note that in this model instead of summing all the premiums received up to time t, we can consider the income money from the premium as a continuous inflow of money and thus c could be thought of as how much money per time unit the company receives. Moreover, $N(t)$ and X is are all random variables that cannot be fully controlled by the company.

The premium rate and the initial capital of the company are two factors that can be accurately set by the company. On the other hand, the number of claim arrivals and the size of the claims is cannot be predicted beforehand. That being said, it is extremely helpful to have an idea of the distribution of $N(t)$ and X_i, that is why, the main difference between various risk models boils down to what probability distribution do $N(t)$ and X_i follow. In what follows, $R(t)$, s, c and X_i are defined as above unless state otherwise.

3.1.1 The Lundberg Risk Model

The Lundberg risk model is as follows (Lundberg, 1903):

$$R(t) = s + ct – S(t) \tag{3.2}$$

Where $S(t) = X_1+X_2+\cdots+X_{N(t)}$ is a compound Poisson process and denotes the aggregated claim sizes up to time t. Thus, in this model, $N(t)$ is a homogeneous Poisson Process with some patameter λ and X_is are iid random variabales that come from some distribution F and $N(t)$ and X_is are independent. So If we denote the waiting time between two consecutive claim arrivals by W_i , then $W_i \sim \exp(\lambda)$ and M_is are independent.

3.1.2 Sparre – Andersen Risk Model

The Lundberg risk model seems to be too simplistic to be applied in real scenarios. Moreover, it restricts the waiting times between consecutive claim arrivals to follow an exponential distribution. The Sparre – Andersen risk model is (Hipp, 2004):

$$R(t) = s + ct - \sum_{i=1}^{N(t)} X_i \tag{3.3}$$

Where $N(t)$ is a renewal process and X_is are iid random variables. Note that the Lundberg risk model is a special case of Sparre – Andersen model.

3.1.3 Markov-Modulated Risk Model

In the Markov-modulated risk model (Rolski et al., 2009), $N(t)$ is an inhomogeneous Poisson process with stochastic intensities $\lambda_1, \lambda_2, \cdots \lambda_l$ the intensities are themselves a continuous-time Markov chain with state space $\{\lambda_1, \lambda_2, \cdots, \lambda_k\}$. The parameters of a Markov-modulated risk model, in addition s,c and the distribution for X_i, is a transition probability matrix for the intensities.

3.2 Risk model with general compound Hawkes process

In all the previous models, the claim arrivals are assumed to be independent. However, in real life, this is usually not the case. For example, consider a situation where a natural disaster happens, clearly in this case the insurance companies may face more dense clusters of claim arrivals as compare to other situations. In general, in any case where an insurance company sometimes observes some clusters of claim arrivals, the previous models may not be helpful anymore. As mentioned before, the main difference between different risk models boils down to how the distribution of $N(t)$ and $X(t)$ is modelled. Thus, in order to incorporate the idea of potential dependence between arrivals we need a more comprehensive model. In chapter 2, we discussed Hawkes process as a counting process which incorporates all history of past in order to determine the distribution of future arrivals. In this regard, Hawkes process seems like a good candidate for modelling $N(t)$. In what follows, we give the definition of a recent model developed by Swishchuk (Swishchuk, Zagst, and G. Zeller, 2021) which is much more comprehensive than previous ones and in fact is a generalization of them.

General Compound Hawkes Porcess (GCHP)

Consider the stochastic process

$$S(t) = S_0 + \sum_{i=1}^{N(t)} a(X_i) \tag{3.4}$$

Where S_0 is a constant, $a(x)$ is any bounded and continuous function and X_i is an ergodic continuous-time and an at most countable state Markov chain, independent of $N(t)$. If $N(t)$ is a one-dimensional Hawkes process, then $S(t)$ is said to be a General Compound Hawkes Process or GCHP.

Risk model based on general compound Hawkes process (RMGCHP)

Consider the stochastic process

$$R(t) = r + ct - \sum_{i=1}^{N(t)} a(X_i) \tag{3.5}$$

Where X_i is a continuous-time Markov chain, $N(t)$ is a Hawkes process and independent of X_i, $a(x)$ is a continuous and bounded function and the rest of the parameters defined as before. Then, $R(t)$ is said to be a risk model based on GCHP. If the state space of claim sizes is finite, then $R(t)$ is called a finite state RMGCHP. If state space of the claim sizes is countably infinite, the $R(t)$ is called an infinite state RMGCHP.

Note that RMGCHP could be thought of as a generalization of the Lundberg risk model. This follows from the fact that Hawkes process itself is a generalization of Poisson process. Thus, by taking the X_i to be some i.i.d random variables and $N(t)$ a Hawkes process with self-exciting function zero, we get the Lundberg risk model.

There are two extremely useful theorems regarding RMGCHP whose implications will be used through this thesis many times and in fact, is the heart of the model we develop to solve Merton problem in insurance in section 4. In what follows we mention the Law of Large Numbers (LLN)and the Functional Central Limit theorem for RMGCHP.

Theorem.(LLN for RMGCHP (Swishchuk, 2017))

Let $R(t)$ be a RGCHP and $R(t)$, with premium rate c, the corresponding self-exciting function μ and its background intensity λ and X_i be an ergodic Markov chain with state space S with stationary probabilities π_i^* and let $a(x)$ be a continuous and bounded function corresponding to $R(t)$. Then we have:

$$\lim_{t\to\infty} \frac{R(t)}{t} = c - a^* \frac{\lambda}{1-\hat{\mu}} \tag{3.6}$$

Where $a^* = \sum_{k\in S} a(k)\pi_k^*$ and $0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1$.

Theorem(FCLT for RMGCHP (Swishchuk, 2017))

Let $R(t)$ be a RMGCHP and X_i be an ergodic Markov chain with stationary probabilities π^*. Then

$$\lim_{t\to\infty} \frac{R(t) - \big(u + ct - a^* N(t)\big)}{\sqrt{t}} \stackrel{D}{=} \hat{\sigma}\Phi(0,1) \tag{3.7}$$

where $\Phi(.,.)$ is the Normal c.d.f and:

$$\begin{aligned}
&\hat{\sigma} := \sigma^* \sqrt{\lambda/(1-\hat{\mu})}, \quad (\sigma^*)^2 := \sum_{i\in S} \pi_i^* v(i) \\
&a^* := \sum_{i\in S} \pi_i^* a(i), \quad b(i) := a^* - a(i) \\
&v_i := b(i)^2 + \sum_{j\in S} \Big(g(j) - g(i)\Big)^2 P(i,j) - 2b(i) \sum_{j\in S} \Big(g(j) - g(i)\Big) P(i,j) \\
&g := \Big(P + \Pi^* - I\Big)^{-1} \Big(b(1), \cdots, b(n)'\Big)
\end{aligned} \tag{3.8}$$

Where P is the transition matrix of X_I and Π^* be a matrix whose rows are π^*.

3.2.1 Ruin Probability

So far we talked about how to model the cash flow of a risk company. The whole purpose of these models is to enable us to measure the risk of the company based on them. But how should we exactly define the risk? One of the measurements for risk of an insurance company are the ruin probability and ultimate ruin probability. The ruin probability up to time is defined to be

$$\phi(r, \tau) = p(R(t) < 0 \quad \text{for some } t < \tau), \tag{3.9}$$

where r is the initial capital. In other words, the ruin probability up to time and starting from initial capital u, is the probability that the net wealth of the company becomes negative at least once before . The ultimate ruin probability is

$$\phi(r) = \lim_{\tau\to\infty} \phi(r,\tau) \tag{3.10}$$

That is, the probability that the net wealth of the company becomes negative at some point in time.

In general finding a closed form for the ruin probability is quite difficult. For the Lundberg risk model, the ruin probability could be obtained by solving the following equation for $\phi(r)$

$$\begin{aligned} \phi(r) &= \frac{\lambda}{c}\Big(\int_u^{\infty} \big(1 - Q(x)\big)dx + \int_0^{x} \phi(r-x)\big(1-Q(x)\big)dx\Big) \\ \phi(0) &= \frac{\lambda m}{c} \end{aligned} \tag{3.11}$$

Where $X_i \sim Q$ and Q is some generic distribution (Rolski et al., 2009). For example, the ruin probability for the case that claim sizes follow an exponential distribution, that is, $X_i \exp(\theta)$ for some $\theta > 0$, is

$$\phi(r) = \frac{\lambda}{c\theta}\exp\Big(\frac{\lambda}{c} - \theta\Big)r \tag{3.12}$$

A nice corollary of FCLT for RMGCHP is the ruin probability for RMCHP could be computed using the following formula (Swishchuk, 2017):

$$\psi(u) = 1 - \phi(u) = P(T_u < +\infty) = e^{-\frac{2(c-a^*\lambda/(1-\hat{\mu}))}{\sigma^2}u} \tag{3.13}$$

where a^*, λ and $\hat{\mu}$ are average claim size, background intensity and branching ratio respectively. σ is defined as in the FCLT theorem for RMGCHP. One application of ruin probability is choosing a value for the premium rate for the insurance company. For the Lundberg risk model we have the following lemma by Schmidli (Schmidli, 2017).

Lemma (Net Profit Condition).

Consider the Lundberg risk model

$$R(t) = r + ct - \sum_{i=1}^{N(t)} X_i \tag{3.14}$$

Let $N(t)$ be a Poisson process with parameter lambda and $m = E[X_i]$. Then we have the followings

- If $c < \lambda m$, then $\lim_{t\to\infty} R(t) = -\infty$, which means $\phi(r) = 1$.
- If $c = \lambda m$, then $\limsup R(t) = -\liminf R(t) = \infty$ implying $\phi(r) = 1$
- If $c > \lambda m$, then $\lim_{t\to\infty} R(t) = \infty$ and $P\big(R(t) > 0\big) \quad \forall t$, that is, the probability that the net wealth of the company never becomes negative is positive.

(Expected Value principle (Schmidli, 2017))

Let m and λ be as defined in the lemma. The expected value principle sets the price of the premium rate c as:

$$c = (1 + \theta)\lambda m \tag{3.15}$$

where $\theta > 0$ is called the safety loading. An application of LLN for RMGCHP is the net profit condition whose corollary gives us a way to set premium for RMGCHP.

Corollary (Swishchuk, 2017): Consider a RMGCHP $R(t) = r + ct - \sum_{i=1}^{N(t)} a(X_i)$. Then the net profit condition for the corresponding RMGCHP is :

$$
\begin{aligned}
&c > a^* \frac{\lambda}{1-\hat{\mu}} \\
&\text{where } a^* = \sum_{k \in X} a(k)\pi_k^*
\end{aligned}
\tag{3.16}
$$

where θ is some positive number and called the safety loading.

According to the Expected Value Principle, the premium for RMGHP is :

$$
(1+\theta)\frac{a^*\lambda}{1-\hat{\mu}} \tag{3.17}
$$

For some $\theta > 0$.

Note that the expected value principle based on the lemma. This results from the fact that if we set the premium rate any number less than equal to λm, then the ruin is unavoidable.

We will finish this chapter by mentioning the diffusion approximation for classical risk models as well as RMGCHP. Though all the risk models we saw earlier, despite all the advantages have a downside. The probabilistic part of these models is a counting function, it does not allow us to easily use the powerful tools we saw in stochastic calculus. We will see in chapter 4 why it is important to have a model whose probabilistic part is modeled by a function of Brownian motion in order to solve the Merton problem. The idea of diffusion approximation is the construct a sequence of random processes whose limit , in the weak convergence sense, is a diffusion process.

Proposition (Schmidli, 2017)

Let $R^{(n)}(t)$ be a sequence of Lundberg risk processes with initial capital r and claim arrivals $X_i^{(n)}$. Let $\lambda^{(n)}$ be rate of claim arrivals an $Q^{(n)}(x) = Q(\sqrt{(n)})(x)$ be the distribution from which $X_i^{(n)}$ comes from. Also, let the premium rates be as follows:

$$c^{(n)} = \Big(1 + \frac{c - \lambda m_1}{\lambda m_1 \sqrt{n}}\Big)\lambda^{(n)} m_1^{(n)} = c + \big(\sqrt{n} - 1\big)\lambda m_1 \tag{3.18}$$

Denote the first and second moments of claim sizes by m_1 and m_2 respectively. Then:

$$R^{(n)}(t) \xrightarrow{d} \Big(u + W(t)\Big) \tag{3.19}$$

where $W(t)$ is a $(c\lambda m_1, \lambda m_2)$-Brownian motion.

Theorem 8 (Swishchuk, 2017)

Let $R(t)$ be a RMGCHP and let X_i be an ergodic Markov chain with wtationary probabilities π^*. Provided that $0 < \hat{\mu} = \int_0^\infty \mu(s)ds < 1$ and $\int_0^\infty s\mu_{(s)}ds < \infty$, then

$$\begin{aligned} \lim_{t\to\infty} \frac{R(t) - \big(u + ct - a^* \frac{\lambda}{1-\hat{\mu}} t\big)}{\sqrt{t}} &\stackrel{D}{=} \overline{\sigma}\Phi(0,1) \\ \overline{\sigma} &= \sqrt{\hat{\sigma}^2 + \big(a^* \sqrt{\frac{\lambda}{(1-\hat{\mu})^3}}\big)^2} \end{aligned} \tag{3.20}$$

One implication from this theorem is that, for large enough t, we can approximate $R(t)$ by:

$$R(t) \approx u + ct - a^* \frac{\lambda}{1 - \hat{\mu}} t + \overline{\sigma} w(t) \tag{3.21}$$

To explain more about the idea of diffusion approximation, let's define the random variable $A(t) = R(t) - \mu$. If we denote by N(n)(t) the arrival time distributions of claims for R(n)(t), then the average arrival time is $E[N(n)(t)] = n\lambda t$. On the other hand, the expected value of claim sizes is $E[X_1^n] = \frac{1}{\sqrt{n}}$. Finally, the expected value of $A^n(t) = E[c^n t - \sum_{i=1}^{N(t)} X_i] = c - \lambda m_1 t$. This means that, we are starting from a risk process R1 and increase its arrival times(this alone increases R(t) on average) and then decrease the average claim sizes(this alone decrease R(t) on average) in such a way that offsets the effect of increase in arrivals and the expected value of Ri(t) remains constant. Now for a fixed t, we can apply the central limit theorem and thus Rn(t) is normally distributed for large enough n.

Chapter 4

Stochastic Optimal Control

In this chapter, we talk about stochastic optimal control, one of the most trending topics in finance. The goal of stochastic optimal control is to design a "strategy" for a system that evolves in time stochastically, by "tuning" variables that at each point in time are under our control, in order to maximize some objective function. So the difference between stochastic optimal control and other classical optimization problems such as linear or convex optimization is that first, we are dealing with some stochastic processes so the system is not deterministic but instead we have an idea of how probable different scenarios are, and second, the space over which we are searching is a set of functions rather than real valued vectors as in other optimization problems. The heart of stochastic optimal control problems is Hamilton-Jacobi-Bellman equation, the equation that was derived from the Dynamic Programming principle. The key is to map the original problem to a larger space in which the original problem would be a special case and try to solve the mapped problem. We will talk about how this equation is derived in this chapter. The idea of dynamic programming principle for optimization problems is to break the original problem into two optimization subproblems, and try to solve one of them given that the other is already optimal and try to derive a relation between the original problem with the solution to its subproblem. We will talk about dynamic programming in more detail and with some examples. Before we dive into the technicalities, let's see some famous applications of stochastic optimal control in financial markets.

4.1 Merton Problem

This problem, which is a portfolio optimization problem was introduced by Merton(1971). The problem is as follows: consider an economic agent (an individual, household, company, etc.) who has some initial wealth denoted by x, and wishes to maximize its utility at some time T by designing a self-financing strategy. At each point of time t where $0 \leq t \leq T$, the agent could allocate a portion at time t (denoted by $\pi(t)$) of its wealth to a risky asset, denoted by S, and the rest to a risk-free asset, denoted by B. The gain from the risk free asset, as it comes from the name is deterministic and fixed. The "potential" gain from the risky asset, on the other hand, could be higher or lower or the agent may even lose some money by investing in the risky asset. Moreover, let's assume we know the dynamic of the price of the risky asset in a probabilistic framework, that is, we know the stochastic differential equation based on which the price of the risky asset behaves . Also, the reason we use a utility function for modelling the objective rather than the amount of the wealth itself, is because the effect the agent gets from losing one dollar and gaining one dollar depending on how much their wealth is, is different, as a result, this affects how much willing the agent is to take risks based on their wealth at each point in time and thus, we incorporate all these into our model by using a utility function (Cartea, Jaimungal, and Penalva, 2015).

Self-financing strategy

Let $S(t)$ denote the value of a risky asset per unit at time t and $B(t)$ denote that of a risk-free asset. Let $P(t) = (n_s(t), n_B(t))$ denote the number of risky asset and risk free asset held at time t respectively and let $\pi(t)$ denote the value of the portfolio at time t, that is, $\pi(t) = S(t)n_s(t) + B(t)n_B(t)$.

Then the portfolio strategy $P(t)$ is said to be self-financing if:

$$S(t)dn_s(t) + B(t)dn_B(t) = 0. \tag{4.1}$$

Another equivalent definition is, $P(t)$ is said to be self-financing if the following holds:

$$d\pi(t) = n_s(t)dS(t) + n_B(t)dS(t). \tag{4.2}$$

Intuitively speaking, a self-financing portfolio strategy is one that does not need any exogenous resource or money withdrawal in order to change its state, namely, changing the numbers n_s and n_B to the desired one(based on the designed strategy), could be done merely by whatever the old value of portfolio was so the portfolio could finance itself at each time t. This definition also makes more sense if we consider it in a discrete case. So let's say for example, we want to have $n_s(1)$ of the risky asset and $n_B(1)$ of the risk-free asset at time 1 and the number of risky assets and risk free assets we have at time 0 is $n_s(0)$ and $n_B(0)$ respectively. Now if the portfolio strategy is self-financing, what ever we are holding from time 0 to 1, should be enough to change the portfolio from $(n_s(0), n_B(0))$ to $(n_s(1), n_B(1))$. The value of the portfolio at time 1 and before changing its state is :

$$n_s(0)S(1) + n_B B(1) \tag{4.3}$$

The value of the portfolio if we change its state, at time 1 would be:

$$n_s(1)S(1) + n_B B(1) \tag{4.4}$$

So this portfolio is self-financing if:

$$
\begin{aligned}
& n_s S(1) + n_B(0)B(1) = n_s(1)S(1) + n_B(1)B(1) \\
\implies & \big(n_s(1) - n_s(0)\big)S(1) + \big(n_B(1) - n_B(0)\big)B(1) = 0 \\
\implies & \Delta n_s S(1) + \Delta n_B B(1) = 0
\end{aligned}
\tag{4.5}
$$

Now since in continuous time instead of difference (Δ) we work with differentials, by replacing the deltas in the formula above by differentials n_s, n_B we get the definition of a self-financing strategy in continuous time. To see how the two definitions coincide note that:

$$
\begin{aligned}
& dn_s(t)S(t) + dn_B(t)B(t) = 0 \\
\iff & dS(t)n_s(t) + dB(t)n_B(t) + dn_s(t)S(t) + dn_B(t)B(t) = dS(t)n_s(t) + dB(t)n_B(t) \\
\iff & \alpha\Pi(t) = dS(t)n_s(t) + dB(t)n_B(t)
\end{aligned}
\tag{4.6}
$$

Now let's get back to the Merton Problem. Let $R(t)$ denote the total wealth of the economic agent at time t. Let $R_s(t)$ and $R_B(t)$ denote the amount of money invested in the risky asset and risk-free asset respectively. Let $n_s(t)$ and $n_B(t)$ be the number of risky asset and risk free assets at time t, respectively and let $S(t)$ and $B(t)$ denote the price of risky asset and risk free asset with interest rate r at time t. Also assume that the price of the risky asset behaves according to the geometric Brownian motion with drift μ and volatility σ. We have:

$$
\begin{aligned}
& R(t) = R_s(t) + R_B(t) = n_s(t)S(t) + n_B(t)B(t) \\
\implies & dR(t) = n_s(t)dS(t) + n_B(t)dB(t) \\
& dS(t) = \mu S_t dt + \sigma S_t dW_t \\
& dB(t) = rB(t)dt
\end{aligned}
\tag{4.7}
$$

Where the W_t denotes the wiener process and the first equation resulted from the fact that the portfolio was assumed to be self- financing .Now if we denote by $\pi(t)$ the portion of the whole wealth invested in the risky asset we could proceed as:

$$\begin{aligned} dR(t) &= n_s(t)S(t)\big(\mu dt + \sigma dW_t\big) + n_B(t)rB(t)dt \\ &= R_s(t)\big(\mu dt + \sigma dW_t\big) + R_B(t)rdt \\ &= \frac{R_s(t)}{R(t)} \times R(t)\big(\mu dt + \sigma dW_t\big) + \frac{R_B(t)}{R(t)} \times R(t)rdt \\ &= \pi(t)R(t)\big(\mu dt + \sigma dW_t\big) + \big(1 - \pi(t)\big)R(t)rdt \end{aligned} \tag{4.8}$$

Now let $U(x,t)$ be some functions that takes as input the amount of total wealth x and time t, and outputs the utility the agent gets at time t if he has total wealth of x. Our problem is now ready in the form of a stochastic optimal control problem. It is modelled as follows:

$$\begin{aligned} &\sup E[U_{\pi\in A}\big(R^\pi(X_T)\big)] \\ &\textbf{s.t.} \\ &dR^\pi(t) = \pi(t)R(t)\big(\mu dt + \sigma dW_t\big) + \big(1 - \pi(t)\big)R(t)rdt \\ &R(0) = x \end{aligned} \tag{4.9}$$

The reason that we use the expected value in our objective function rather than the utility function itself, is that the amount of wealth by time T, X_T , is a random variable so we need a somehow deterministic criteria, so instead we try to maximize the utility the agent gets "on average". Note that Merton problem is suitable when it comes to designing strategies where the transactions do not happen very frequently and the orders placed are not in large amounts each time, otherwise we have to incorporate the induced effect on the value of the asset. The next example is suitable for such problems.

4.2 The Optimal Liquidation Problem (Cartea, Jaimungal, and Penalva, 2015)

As mentioned earlier, sometimes we aim to design a trading strategy for the case where orders occur in large amounts and short period of times. In such situations, it would be unrealistic to ignore the effect of such orders on the market price of the asset. The liquidation problem is an example. The problem is as follows. Consider an economic agent who wishes to sell a large amount of a specific asset they have. Since the size of their order is considerable, if they try to sell it all at once, such a large supply of the asset in the market in a short period of time, will decrease the price of the asset and thus, the agent will not sell their asset at a good price. Moreover, assume that there is a penalty for holding their inventory, for example, assume that the agent has rented a stockroom and as a result, as long as there are some inventories there, the agent must pay for its rent. Thus, the agent keeps losing money if he does not liquidate the inventories fast enough. So the idea is to spread out his orders over time in such a way that the expected utility she gets up to time T is maximized. So the formal setting would be as follows:

$$
\begin{aligned}
&\sup_{\pi \in A} E\Big[X_T^\pi + (Q_T^\pi \times S_T^\pi)\Big] \\
&\text{s.t} \\
&dQ_t^\Pi = -\pi_t dt \qquad\qquad\quad Q_0^\pi = q \\
&dS_t^\pi = -f(\pi_t)dt + \sigma dw_t \quad S_0^\pi = s \\
&dX_t^\pi = \pi_t S_t^\pi dt
\end{aligned}
\tag{4.10}
$$

Where:

- $Q(t)$: represents the amount of the inventory at time t.
- $\pi(t)$: represents the rate at which the inventory should be liquidated at time t.

- $S(t)$: represents the price of the inventory at time t.
- f : is a real valued (negative) function that affects the drift of the inventory's price based on the amount the agents try to liquidate at time t.
- $X(t)$: represents the total cash earned by time t through liquidating the inventory.

4.3 Control For Diffusion Problems (Cartea, Jaimungal, and Penalva, 2015)

In this section we show how a general control for diffusion problem looks. Before doing so, let's consider the main components a control problem is consist of. We are always dealing with one or more state processes X_i, these are usually what we aim to maximize a function over and we model them by some stochastic differential equations . In the Liquidation problem for example, there is no constrain on the number of different assets we want to liquidate, so if our goal was to liquidate two or more, say "n" assets, each of them would have their own price process. So in the case of liquidating two assets, the Liquidation problem could be written as:

$$\begin{aligned} dX_1^\pi(t) &= \mu_1\Big(t, X_1^\pi(t), \pi_1(t)\Big)dt \\ &\quad + \sigma_1\Big(t, X_1^\pi, \pi_1(t)\Big)dw_1 \\ dX_2^\pi &= \mu_2\Big(t, X_2^\pi, \pi_2(t)\Big)dt \\ &\quad + \sigma_2\Big(t, X_2^\pi, \pi_2(t)\Big)dw_2 \end{aligned} \tag{4.11}$$

Or more concisely as:

$$
\begin{aligned}
dX_t^\pi(t) = \mu\Big(t, X_t^\pi(t), \pi(t)\Big)dt \\
+ \sigma\Big(t, X^\pi, \pi(t)\Big)dw
\end{aligned}
\tag{4.12}
$$

Where $\vec{X} = (X_1, X_2)$, $\vec{\mu} = (\mu_1, \mu_2)$, $\vec{\pi} = (\pi_1, \pi_2)$ and $\vec{\sigma} = (\sigma_1, \sigma_2)$. Another component, is about how we model the noise in a control problem, we consider the cases only where the noise is modelled by a Wiener process. Let's get back to the Merton example, we modelled the noise in the price of the risky asset using the Brownian motion. Again, it is possible to do the problem in the case where there is more than one risky asset and each of them has their own noise modelled by a Brownian motion. So for example in the case where the agent could invest their wealth in two risky assets, the problem would be as follows:

The last element, is the utility function that takes the time index and the state at each time, and outputs the obtained utility accordingly, and the goal is to maximize this function up to a given time.

As we mentioned earlier, in order to model a problem as a stochastic optimal control, all we need is to derive a SDE for our state processes in terms of the control function π_t . The initial state and a utility function. So the formal problem could be written as follows:

$$
\begin{aligned}
&\sup_{\pi \in A} E\Big[G\big(X_T^\pi\big) + \int_0^\infty F(t, X_t^\pi, \pi_t)dt\Big] \\
&\text{s.t} \\
&dX_t^\pi(t) = \mu_t\Big(t, X_t^\pi(t), \pi(t)\Big)dt \\
&+ \sigma\Big(t, X_t^\pi, \pi_1(t)\Big)dw_t \\
&X_0^u = x
\end{aligned}
\tag{4.13}
$$

Note that in the last line we may or may not impose a constrain for our control functions because in practice, not any control function makes sense or is acceptable. For example, in the Merton problem, the π_T represents the portion of the whole wealth to be invested in the risky asset, it is reasonable to assume that $\pi_t \leq 1$ since it makes no sense to invest more than what we already have in something. Also, if the short selling is not allowed, we should also assume that $\pi_t \geq 0$. Also for the Liquidation problem, the amount of the asset we liquidate cannot exceed the initial amount as again it is not possible to liquidate more than the amount of asset we already have. The set P of functions that satisfy all the conditions is called the Admissible set. So Admissible set is the set of all candidates of the possible solutions among which we look for the best. The admissible set seems to be too large to search for the best solution in by brute force techniques, so the need for a systematic way of finding the best optimal function seems to be inevitable. The key to this problem is what is called the Hamilton-Jacobi-Bellman equation and is derived based on what is called the Dynamic programming principle. In what follows we will elaborate more on this.

4.4 The Dynamic Programming Principle

Dynamic programming was developed by Richard Bellman mainly for the purpose of mathematical optimization problems. Nowadays it is also widely used as one of the most powerful tools in algorithm design related areas in computer science. The idea of dynamic programming, is to break the original problem into smaller subproblems and try to find the optimal solution of subproblem and then using the suboptimal solutions, solve the original problem by some recursive approach. If a problem could be solved this way, it is said to have an optimal substructure. We will prove that the optimal control problem for diffusion processes has an optimal substructure.

Definition: Stopping time

Let $X = \{X_n : n \geq 0\}$ be a stochastic process. A stopping time with respect to X is a random time such that for each $n \geq 0$, the event $\{\tau = n\}$ is completely determined by (at most) the total information known up to time n, $\{X_0, \cdots, X_n\}$.

This definition was taken from

`http://www.columbia.edu/~ks20/stochastic-I/stochastic-I-ST.pdf`

Theorem (Cartea, Jaimungal, and Penalva, 2015)

Let $H(t,x) = \sup H^{\pi}_{\pi \in A}(t,x)$ and $H^{\pi}(t,x) = E_{t,x}\Big[G(X^{\pi}_T) + \int_t^T F(t, X^{\pi}_t, \pi_t)dt\Big]$. Let τ be any stopping time between 0 and T. Then, the following holds

$$H(t,x) = \sup E_{t,x}\Big[H(\tau, X^{\pi}_T) + \int_t^{\tau} F(S, X^{\pi}_S, \pi_S)dS\Big] \tag{4.14}$$

Though dynamic programming principle gives us an idea of suboptimality of the optimal control for diffusion problems, yet, it cannot be used directly to solve the problem. In that regard, The following theorem which is a result of dynamic programming principle and was shown in a paper by Richard Bellman (Richard, 1954), gives us a systematic way of solving the optimal control problem for diffusion processes. The Hamilton-Jacobi-Bellman equation, denoted by HJM, is as follows

$$\begin{aligned} &\partial_t H(t,x) + \sup_{\pi \in A}\Big(A^{\pi}_t H(t,x) + F(t,x,u)\Big) = 0 \\ &H(T,x) = G(x) \end{aligned} \tag{4.15}$$

Where ∂_t denotes the partial derivative with respect to t and A^{π} denotes the infinitesimal generator where for an Ito process $dX^{\pi}_t = \mu(t,x,\pi)dt + \sigma(t,x,\pi)dw_t$

$$A^{\pi} = \mu(t, X(0), \pi)\frac{\partial}{\partial x} + \sigma^2(t, X(0), \pi)\frac{\partial^2}{\partial x^2} \tag{4.16}$$

4.4.1 Solving Merton Problem for RMGCHP

In this section we show how to solve the Merton problem for RMGCHP. From now on, instead of writing the whole formula of the RMGHCP, we simply show it as $R(t) = u + ct -$ GCHP where GCHP is the corresponding General Compound Hawkes process. Consider an insurance company whose wealth at each time t follows a RMGCHP. Suppose that at each time t, the company can invest part of its money in a risky asset and the remaining in a risk-free asset. Let $S(t)$ denote the price of a risky asset at time t and assume that it is a geometric Brownian motion with mean μ and drift σ. Let $\pi(t)$, $R_s(t)$, $R_b(t)$, $n_s(t)$ and $n_b(t)$ be as defined in the Merton problem previously. Now we have

$$dR(t) = dR_s(t) + dR_B(t) = cdt - d(\text{RGCHP}) \tag{4.17}$$

In order to apply HJM, we need to obtain the infinitesimal generator for this SDE. Unfortunately, we cannot do so due to the term $d(\text{GCHP})$. However, we know the infinitesimal generator for a diffusion process. Thus, we can use the pure diffusion approximation for RMGCHP mentioned earlier and then use HJM.

$$\begin{aligned} dR^{\pi} &= \pi(t)R^{\pi}(t)\big(\mu dt + bdw_t\big) + (1-\pi_t)R^{\pi}(t)rdt + cdt - a^*\frac{\lambda}{1-\hat{\mu}}dt - \overline{\sigma}dw_t \\ &= R^{\pi}(t)\Big(r\pi(t)(\mu - r) + c - \frac{a^*\lambda}{1-\hat{\mu}}\Big)dt + \sqrt{\pi^2(t)b^2R^2(t) + \overline{\sigma}^2}dw_t \\ &= R(0) = u \end{aligned} \tag{4.18}$$

Now before applying the HJM equation, we need to first set some utility function. In our model, we used the exponential utility function $U(x) = -e^{(-ax)}$. Thus, we have:

Based on our utility function now, we set an parametrized ansatz $H(t,x) = -e^{-px}$ (p has to be found)

Now based on HJB we have:

$$\frac{\partial H(t,x)}{\partial x} + \sup_{\pi}[A^{\pi}H(t,x)] = 0$$

$$\Rightarrow \quad \sup_{\pi}[A^{\pi}H(t,x)] = 0 \qquad (1)$$

$$A^{\pi} = [u(r + (\mu - r)\pi(t)) + (c - a^{*}\frac{\lambda}{1-\hat{\mu}})]\frac{\partial}{\partial x} + [\frac{(u^2b^2\pi^2(t) + \sigma^2)}{2}]\frac{\partial}{\partial x^2} \Rightarrow$$

$$A^{\pi}(H(t,x)) = A^{\pi}(-e^{-px}) = [u(r+(\mu-r))\pi(t)+(c-a^{*}\frac{\lambda}{1-\hat{\mu}})]e^{-px}+[\frac{(u^2b^2\pi^2(t)+\sigma^2)}{2}]\times(-p^2e^{-px})$$

Now in order to compute $\sup_{\Pi}[A^{\pi}]H(t,x)$ we take the derivative of $A^{\pi}H(t,x)$ with respect to $\pi(t)$ and set it equal to zero:

$$\frac{d\ A^{\pi}H(t,x)}{d\ \pi(t)} = u(\mu - r)pe^{-px} - u^2b^2\pi(t)p^2e^{-px} = 0 \Rightarrow \pi(t) = \frac{(\mu - r)}{ub^2p}$$

Now we plug in this $\Pi(t)$ to (1):

$$\sup_{\pi}[A^{\pi}H(t,x)] = [u(r + (\mu - r)\frac{(\mu - r)}{ub^2p}) + (c - a^{*}\frac{\lambda}{1-\hat{\mu}})]\frac{\partial - e^{-px}}{\partial x}$$

$$+[(u^2b^2\frac{(\mu - r)^2}{u^2b^4p^2} + \sigma^2)/2]\frac{\partial^2 - e^{-px}}{\partial x^2} = [u(r + \frac{(\mu - r)^2}{ub^2p}) + (c - a^{*}\frac{\lambda}{1-\hat{\mu}})] \times (pe^{-px})$$

$$+[\frac{((\mu - r)^2}{b^2p^2} + \sigma^2)/2] \times (-p^2e^{-px})$$

$$(ur + (c - a^{*}\frac{\lambda}{1-\hat{\mu}})pe^{-px} + (\frac{\sigma^2}{2}) \times (-p^2e^{-px}) +$$

$$(\frac{(\mu-r)^2}{b^2}-\frac{(\mu-r)^2}{2b^2})e^{-px}=(ur+(c-a^*\frac{\lambda}{1-\hat{\mu}}))pe^{-px}+$$

$$(\frac{\sigma^2}{2})p^2+(ur+c-a^*\frac{\lambda}{1-\hat{\mu}})p+\frac{(\mu-r)^2}{2b^2}=0\Rightarrow$$

$$p=\frac{-ur-c+a^*\frac{\lambda}{1-\hat{\mu}}\pm\sqrt{(ur+(c-a^*\frac{\lambda}{1-\hat{\mu}}))^2+\frac{\sigma^2(\mu-r)^2}{b^2}}}{-\sigma^2}$$

$$=\frac{-ur-c+a^*\frac{\lambda}{1-\hat{\mu}}\pm\sqrt{u^2r^2+c^2+a^{*2}\frac{\lambda^2}{(1-\hat{\mu})^2}-2ca^*\frac{\lambda}{1-\hat{\mu}}+\frac{\sigma^2(\mu-r)^2}{b^2}+2ur(c-\frac{a^*\lambda}{1-\hat{\mu}})}}{-\sigma^2}$$

$$\pi(t)=\frac{(\mu-r)}{ub^2\frac{-ur-c+a^*\frac{\lambda}{1-\hat{\mu}}\pm\sqrt{u^2r^2+c^2+a^{*2}\frac{\lambda^2}{(1-\hat{\mu})^2}-2ca^*\frac{\lambda}{1-\hat{\mu}}+\frac{\sigma^2(\mu-r)^2}{b^2}+2ur(c-\frac{a^*\lambda}{1-\hat{\mu}})}}{-\sigma^2}}$$

Chapter 5

Simulations , Graphical illustrations and Interpretations

In previous sections, we went through the theory needed to construct the model, as well as the construction of the model itself. Though all the steps were explained, working merely by numbers alone, won't help one in gaining insight about the model. Moreover, though we may not be able to easily say if a model is going to work properly but we could easily say it does not work correctly by checking a few things that we know they must hold and once we see the model show otherwise, it is a warning sign that it should be reconsidered. Simulations and plotting their graphs, is the best way to communicate the results of a model not just to people who have a mathematical background, but to those who come from any other disciplines so that everyone could have a better grasp of how changing different parameters are going to affect the behavioure of a system. In this section, we are going to talk about the algorithm for simulating a Hawkes process and then show the graphs of how the ruin probability and the optimal value could be affected by different parameters of the model. All numbers used in this section are based on the calibration on real insurance company data (G. F. Zeller, 2018). For the ruin probability section, the parameters investigated are initial capital(u), average claim size (a^*), background intensity (λ), premium rate (c) and branching ration ($\hat{\mu}$) and they are all based on the formula derived earlier . For the optimal control section, the parameters investigated are the drift of the risky asset (μ), interest rate of the risk-free asset (r), average claims size a^*, background intensity (λ), α, β where are all these parameters come from the related formula we derived for optimal control earleir. Each simulation is based on changing the corresponding parameter fixed and all the other parameters in the related formula fixed. So for example, for the graph of the ruin probability based on initial capital, we change the initial capital while holding all the other parameters fixed in the ruin probability formula we had earlier.

5.1 Ruin probability dependencies

In this section we show how ruin probability depends on different parameters of the model and explain how the graphs match our intuition. The ruin probability formula used here is:

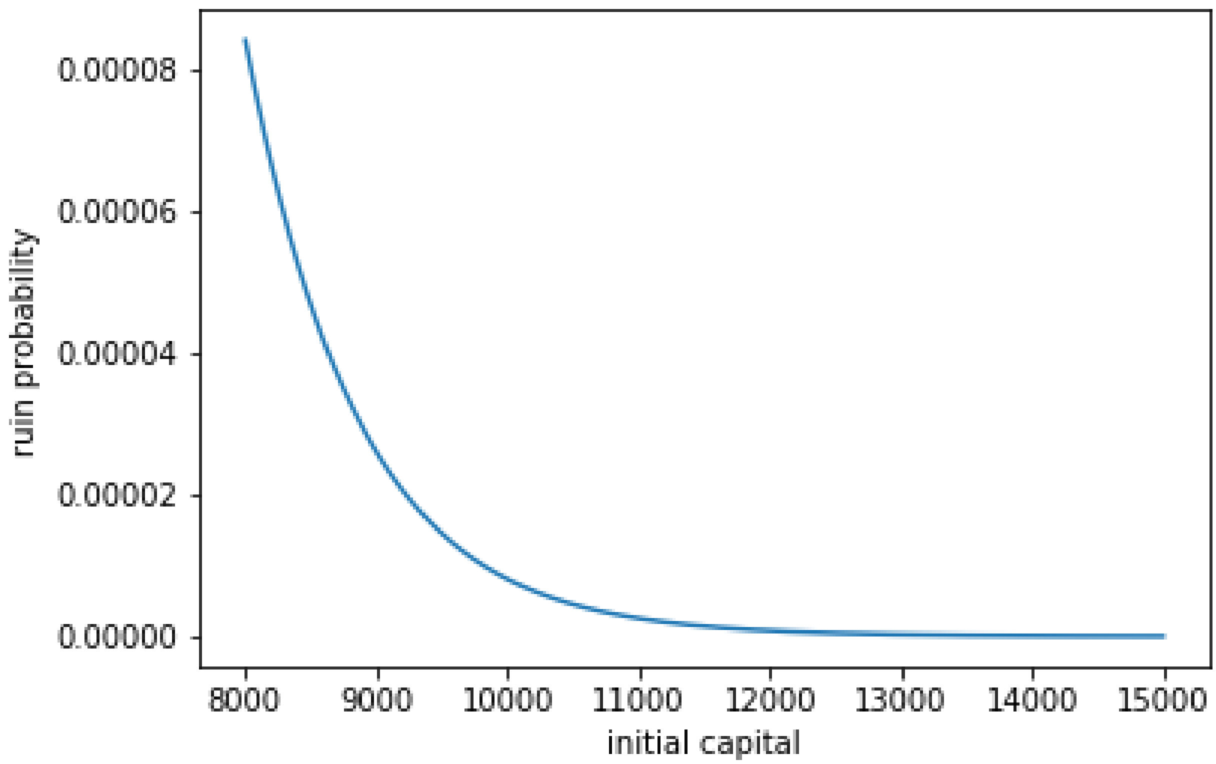
$$\psi(u) = e^{-\frac{2(c-a^*\lambda/(1-\hat{\mu}))}{\sigma^2}u} \tag{5.1}$$

Where c is the premium rate, a^* is the average claim size, λ is the background intensity and $\hat{\mu}$ is the branching ratio.

5.1.1 Dependency on Initial Capital

It seems reasonable to believe that the more the initial capital is, the less the chances of the overall wealth of the company to get into negative would be. This is because in order for the wealth to become negative, the outflow of the money has to outweigh all the wealth up to the time, and if the capital is more in amount, it requires more outflow of money for the wealth to become negative and more outflow of money means more claim arrivals as claim arrivals are the main reason why the company loses money. The graph of the dependency of the ruin probability on the initial capital is as follows:

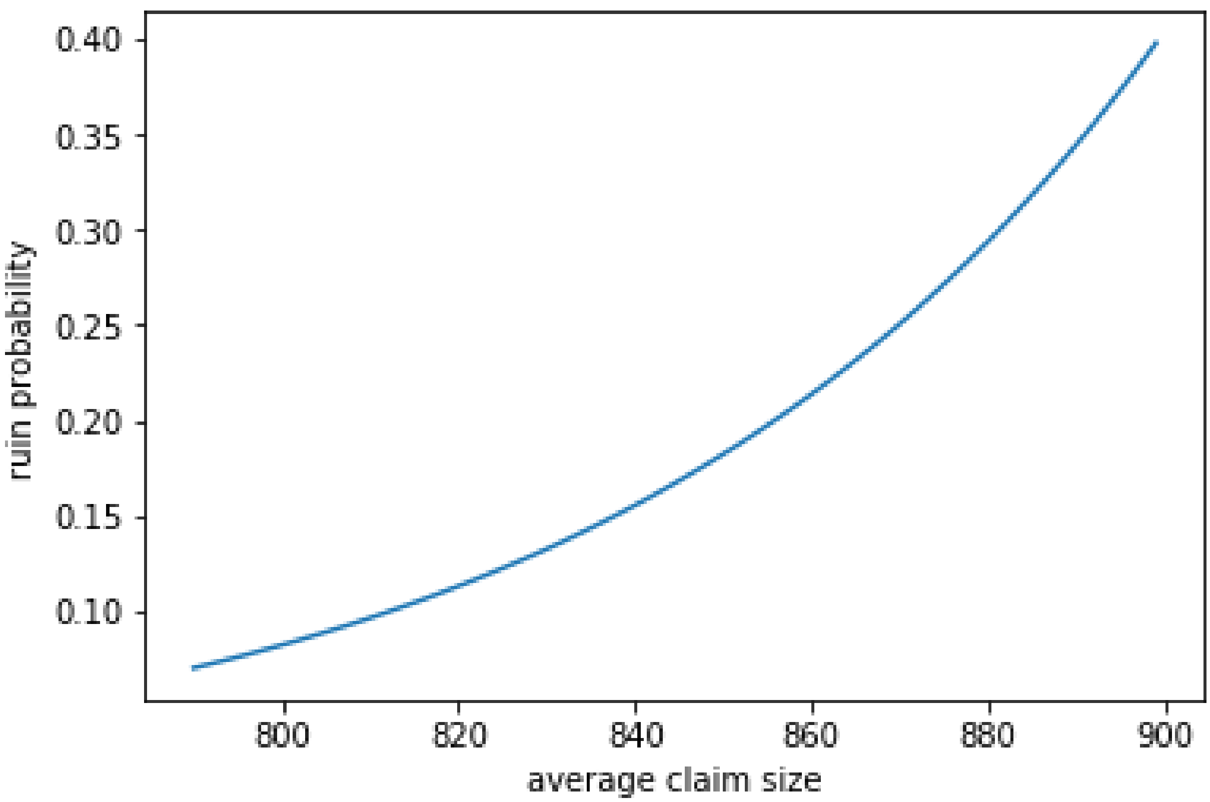
Figure 5.1: Dependency of ruin probability on initial capital



5.1.2 Dependency on average claim's size

The larger the claim size is, the more the company will lose money and thus it will be more likely for the overall total wealth to become negative. As a result we expect the ruin probability to be strictly increasing in terms of the average claim size. The graph is as follows.

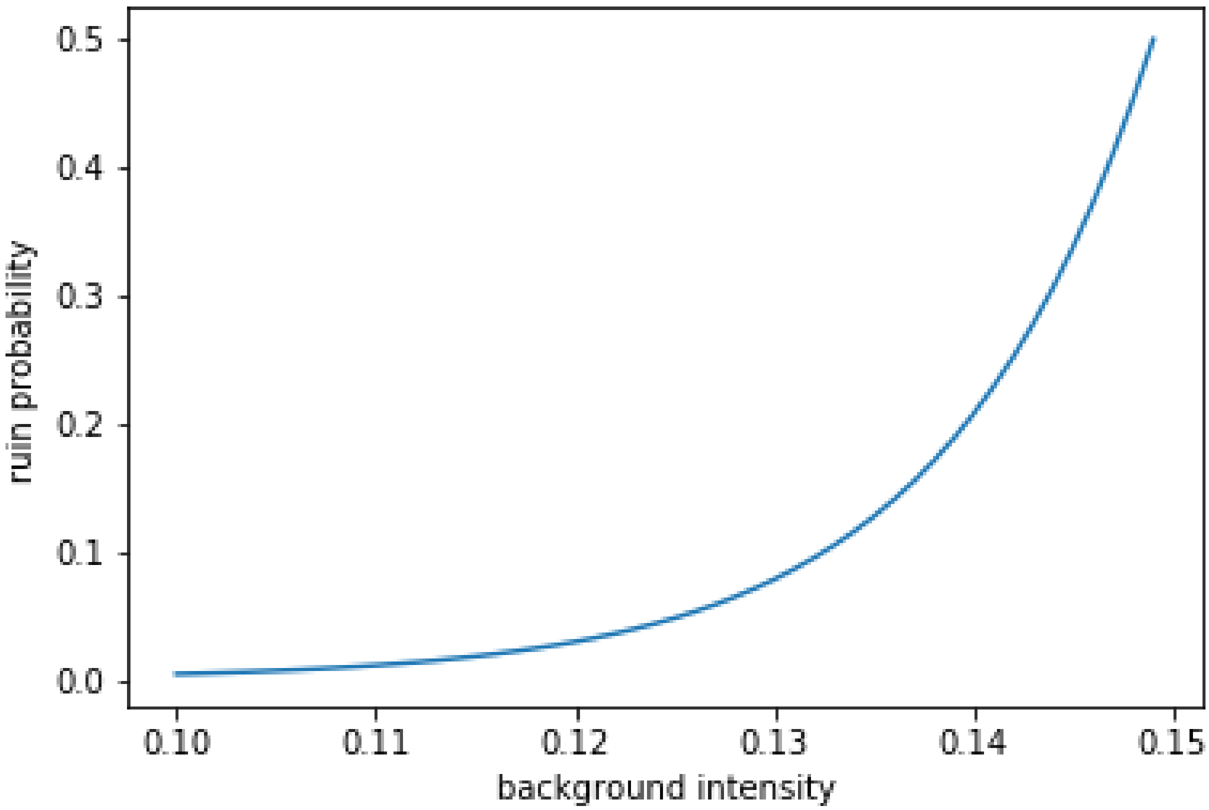
Figure 5.2: Dependency of ruin probability on average claim size



5.1.3 Dependency on the background intensity

As discussed earlier, one of the parameters of the Hawkes process is the background intensity, which could be thought of the default expected rate of arrivals even if there is no self-exciting. So the larger this parameter gets, it means that we should expect more claim arrivals and that means more loss of money and this leads to lower overall wealth which increases the probability of ruin. This could be seen in the graph as follows:

Figure 5.3: Dependency of ruin probability on background intensity



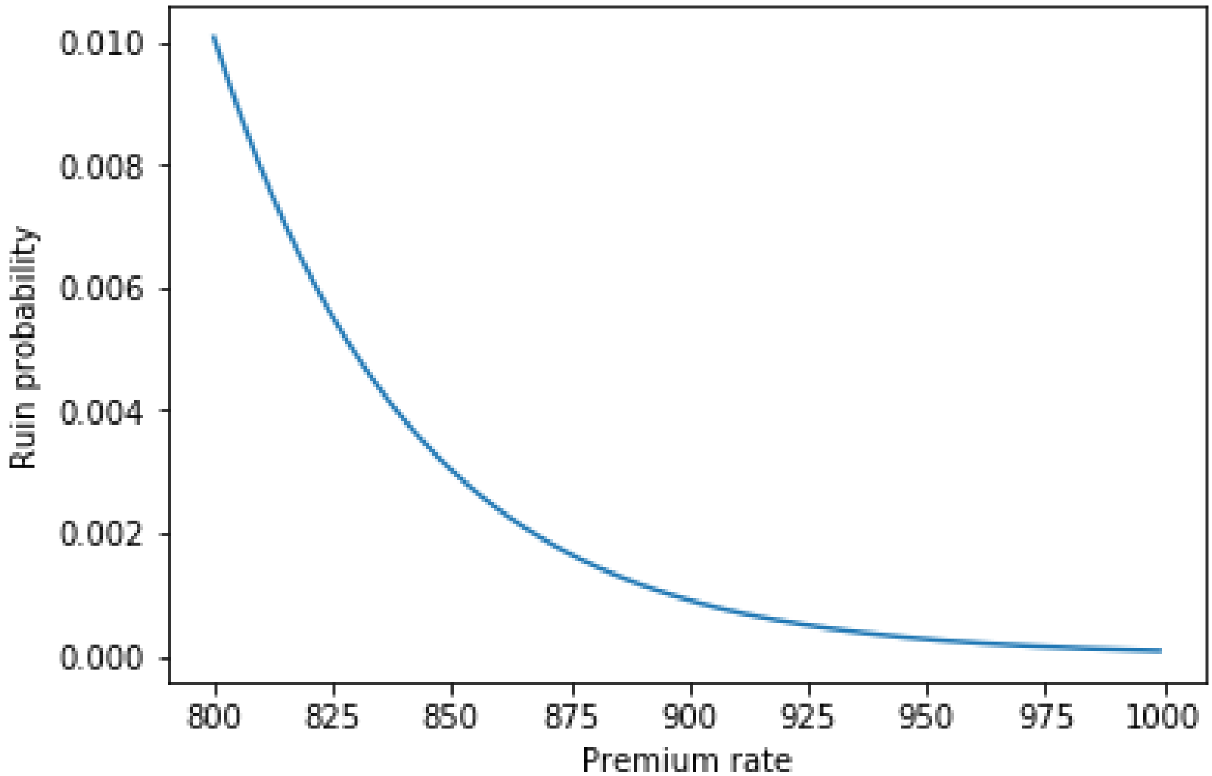
5.1.4 Dependency on Premium rate

The cash inflow of the company depends on 3 factors, the premium rate of the company , the potential profit of the risky asset and the discount rate earned from the risk-free asset. So all else equal, if we increase the premium rate the cash inflow will increase and thus it will be less likely for the cash outflow to outweigh all the accumulated wealth so the ruin probability is expected to be decreasing with respect to the premium rate. The graph is as follows:

5.1.5 Dependency on the branching ratio

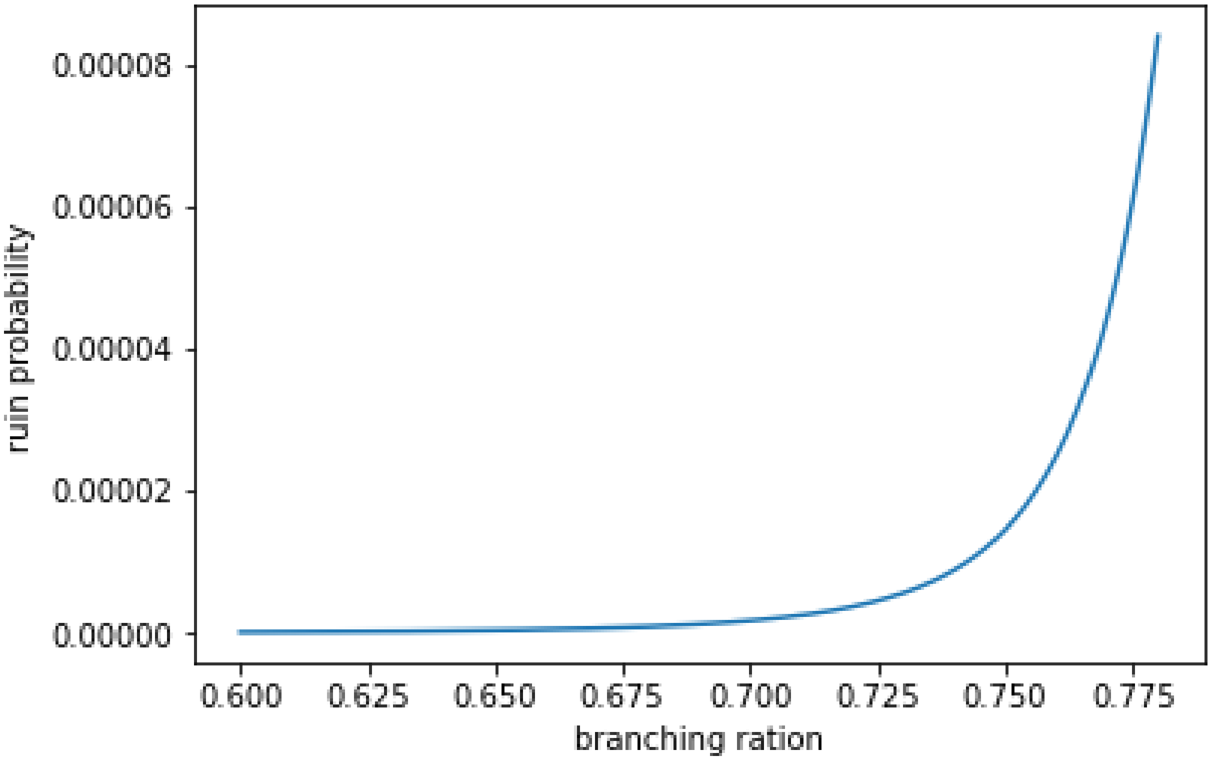
As we saw in the immigration-birth representation of Hawkes process, the more the branching ratio is, the more the average of the number of descendants of an immigrant will be. In the

Figure 5.4: Dependency of ruin probability on premium rate



RGCHP terminology, the branching ratio indicates the average number of claim arrivals due to a "first generation" claim arrival, though in reality the claims may not exactly happen in this order but we showed that using the immigration-birth representation we can think of them this way. So if the branching ratio is larger, it means each first generation claim arrival generates more claim arrivals in the future on average and this means more cash outflow from the company which makes it more likely for the wealth of the company become negative. So we expect that the ruin probability is increasing in the branching ratio. The graph is as follows:

Figure 5.5: Dependency of ruin probability on branching ratio



5.2 Optimal Value Dependencies

In the previous section we showed how the ultimate ruin probability depends on different factor. In this section, we show how the optimal value π depends on the different parameters. We will see that in this section unlike the previous one, merely by setting all but one parameter, we may not get same results and the intuition behind these different results will be discussed. The formula based on which the graphs are is what we derived earlier in the optimal control section:

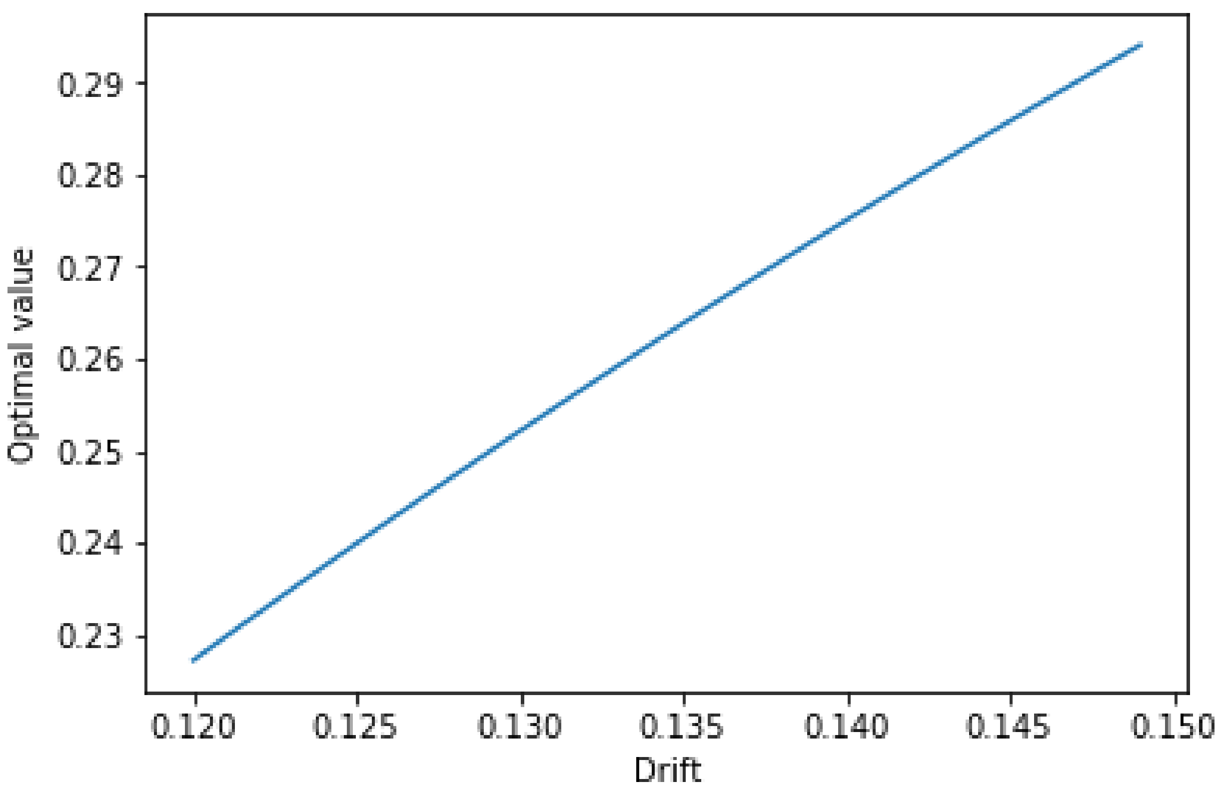
$$\pi(t) = \frac{(\mu - r)}{ub^2 \frac{-ur-c+a^* \frac{\lambda}{1-\hat{\mu}} \pm \sqrt{u^2r^2+c^2+a^{*2}\frac{\lambda^2}{(1-\hat{\mu})^2} - 2ca^*\frac{\lambda}{1-\hat{\mu}} + \frac{\sigma^2(\mu-r)^2}{b^2} + 2ur(c-\frac{a^*\lambda}{1-\hat{\mu}})}}{-\sigma^2}}$$

Where μ is the drift of the risky asset, r it the risk free return, c is the premium rate, a^* is the average claim size, λ is is the background intensity, $\hat{\mu}$ is the branching ratio and b is the volatility of the risky asset.

5.2.1 The Drift

Intuitively speaking, the drift of a geometric Brownian motion, represents the trend of the random variable. In the case of a stock, it could be interpreted as the mean return of the stock price. Thus, the larger the drift is, the higher return we expect on the risky asset. As a result, we expect the optimal value be increasing in drift as larger drift means higher return which leads to higher utility. The graph is as follows

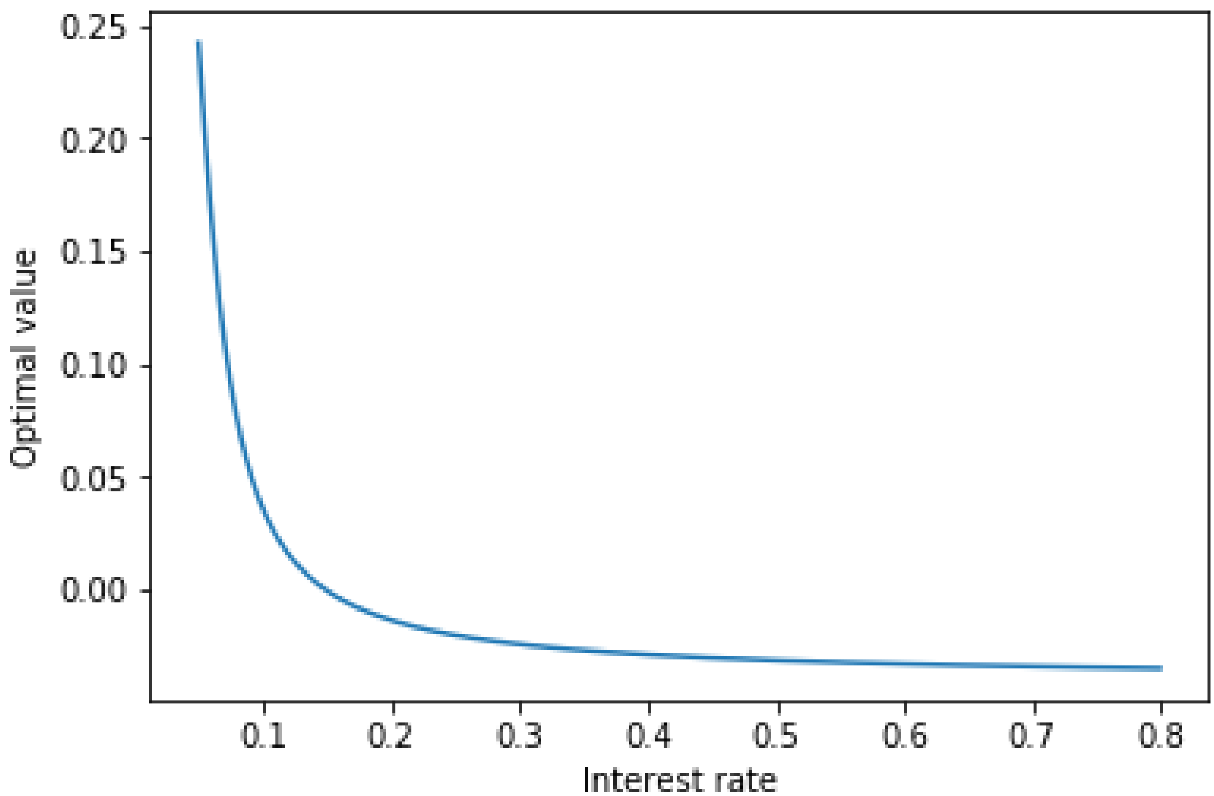
Figure 5.6: Dependency of optimal value on drift



5.2.2 Discount rate of the risk-free asset

The more the company invests in the risky asset, the less will be left to be invested in the risk- free asset. Since the objective is to maximize the expected wealth, if the risk-free asset has higher discount rate, then it is more tempting to invest more in it as the return is higher. So we expect that the parameter π to be decreasing in risk-free asset's discount rate:

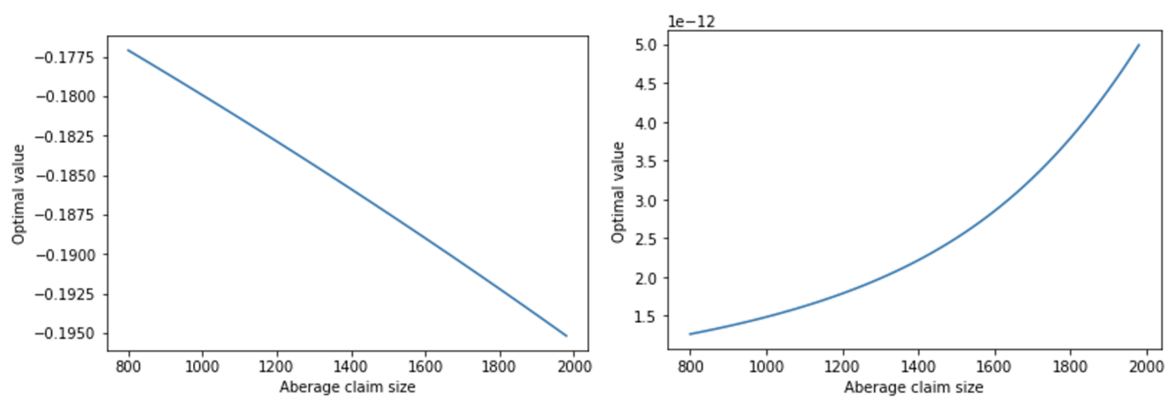
Figure 5.7: Dependency of optimal value on interest rate



5.2.3 Average Claim Size

Larger average claim size means higher cost for the company. So in order for the company to loss money as much as possible, it has two ways to compensate for the claim size, the potential return they get from their investment in the risky asset and the interest rate they earn from risk-free asset. This case, unlike previous one is less trivial since even if we set all the parameters constant and change the average claim size, the behavior of the optimal value still depends on what values we have set the other parameters to be. This could be justified intuitively as well; assume that the interest rates of the risk-free asset is very low but the claim sizes are really large, so if the company relies merely on the risk-free assets, the return would not be enough to compensate for the claim costs. On the other hand, the potential return of the risk free asset is not bounded so at some point it will worth to take some risk for the potential high return of the risky asset. The relation between the average claim size and the optimal π is illustrated in the graphs below:

Figure 5.8: Dependency of optimal value on average claim size

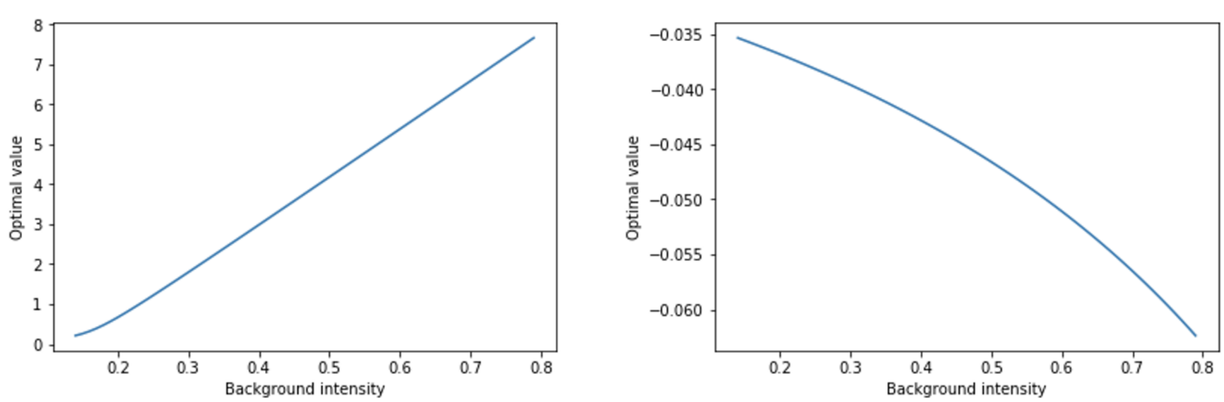


The graph on the left, is for the where the interest rate is 0.05 and the one in the right is where the interest rate is 0.9. In the application however, the interest rate 0.9 is too high to be real. That being said, the reason the graphs were shown for both of these numbers, were to make a point; if the interest rate is not high enough and the claim sizes are large, then the risk will be worthy to take. On the other hand though, if the interest rate is so high, then not only taking long position in stock market is not reasonable, but also taking short position would make more sense as in the worse case scenario, we could use the high return of the risk free asset to compensate almost any potential loss due to the risky asset.

5.2.4 Dependency on Background Intensity

The more the background intensity is, the more we claim arrivals we expect. Since the claims have been on the expense of the company, the company needs a way to pay them off. If the interest rate of the risk-free asset is very low, then the company needs to take more risk and spend more money on the risky assets as the return of the risky assets could be very high. On the other hand, if the interest rate is one can earn through risk-free asset is very high, then the company can even use the opportunity to short sell in the risky asset and deposit the money in the risk-free asset and then give back the asset. Thus, for low interest rates, the optimal value is increasing in the background intensity as the risk-free return is not enough to pay off the claims and for the high interest rate, the optimal value is decreasing and the optimal values are negative, which means the company should short-sell more risky assets. The graphs below show the optimal value with respect to the background intensity for %5 and %90 interest rates from left to right respectively.

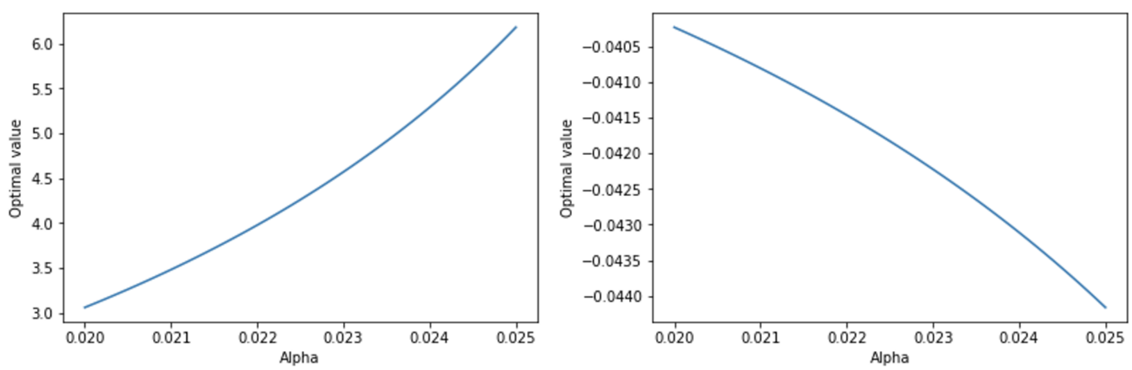
Figure 5.9: Dependency of optimal value on background intensity



5.2.5 Dependency on Alpha

As we saw earlier, the parameter alpha of the exponential decay function indicates the magnitude of the impact of each arrival on next potential arrivals. Thus, the larger alpha in RMGHP, means more expected claim arrivals. As before, if the interest rate of the risk-free asset is not high enough, then the company has to invest more in the risky assets to be able to pay off the claims and if the interest rates are too high, the company can short sell the risky asset and earn interest on it by depositing it in the risk free asset. The following graphs are optimal value with respect to alpha parameter for %5 and %90 interest rates from left to right respectively.

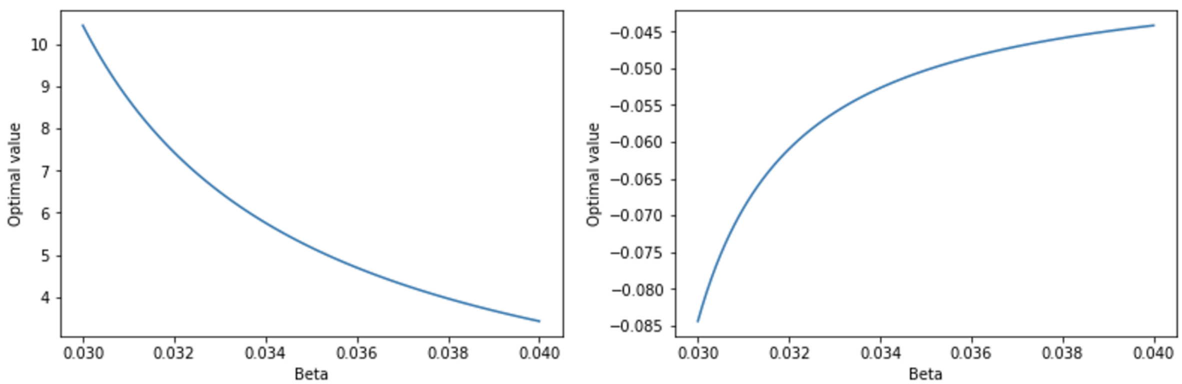
Figure 5.10: Dependency of optimal value on alpha



5.2.6 Dependency on Beta

As discussed in chapter 2, the parameter beta shows how fast the impact of each arrival on next potential arrivals disappears. So, all else equal, the larger the beta is, the less arrivals we should expect. As a result, similar as before, the optimal value should be decreasing in beta for low interest rate and increasing for high interest rates. The following graphs are optimal value with respect to beta from left to right respectively.

Figure 5.11: Dependency of optimal value on beta



Chapter 6

Conclusion

In this thesis, we derived a optimal control formula based on a sophisticated and more generalized risk model as compare to previous ones, namely the general compound Hawkes process for insurance companies. In the last chapter, we illustrated the graphs of the numerical simulations and their interpretations as well. Some results from the simulations were quite intuitive, for example, we saw that if all else equal, the interest rate increases, the optimal value, that is, the amount of money the insurance company needs to invest in the risky asset, decreases. It is indeed intuitive since if everything other than the risk-free rate of return is kept constant, then investing more in the risk-free asset with higher rate of return means higher overall rate of return with the same risk exposure for the case where risk-free interest rate is lower and all else is the same. But, investing more in the risk free return means spending less money on the risky assets, that is, smaller pi^*. Dependency of the optimal value on parameters that determine the exposure of the insurance company to the risks of claim arrivals however, was less intuitive. When a parameter causes an increase in the aggregated claim size, it translates to an outflow of the money from the company and thus, decreasing the utility. In order for the company to compensate for the losing money, they have to choose between one of the followings:

- Allocate more of the money to risky assets and thus, increasing pi^*

- Allocate more of the money to the risk-free asset, thus, decreasing pi^*

Based on the formulas and calibration we used, it seems like spending more money on the risk-free asset is not a good idea. This is somewhat reasonable, as it could be thought of this way: if we are losing more money, in order to stay close to the optimal point where less money was being lost, we basically need a more return. More return would be gained usually through risky-assets since the return of the risky asset is more than that of the risk-free asset because of the risk-premium, the expected extra rate of return that gives the market participants enough incentive to buy the risky asset and bare its risk. This idea of taking more risk by buying more risky asset in order to be able to compensate for the increased risk-exposure due to claim arrivals, is very consistent in the numerical results as could be seen, any place that a variable causes the overall claims to increase, whether by making arrivals more frequent or by increasing the size of a typical claim, results in an increase in the optimal value as well. It is also true that, any factor that decreases the expected overall claims, results in decrease in the optimal value and this is because the optimal value and this is due to the risk aversion that was modelled by the utility function. In that, though investing more in the risky assets has expected higher return, but since it also has higher risk, by investing more in the risky asset the utility will decrease as the risk to bare is not worth the extra expected return.

Bibliography

Arfwedson, Gerhard (1950). "Some problems in the collective theory of risk". In: *Scandinavian Actuarial Journal* 1950.1-2, pp. 1–38.

— (1955). "Research in collective risk theory: Part II". In: *Scandinavian Actuarial Journal* 1955.1-2, pp. 53–100.

Bacry, Emmanuel, Iacopo Mastromatteo, and Jean-François Muzy (2015). "Hawkes processes in finance". In: *Market Microstructure and Liquidity* 1.01, p. 1550005.

Bremaud, Pierre and Laurent Massoulie (1996). "Stability of Nonlinear Hawkes Processes". In: *The Annals of Probability* 24.3, pp. 1563–1588. ISSN: 00911798. URL: `http://www.jstor.org/stable/2244985`.

Brockett, Patrick and Xia Xiaohua (1997). "Operations Research In Insurance: A Review." In: *Insurance Mathematics and Economics* 2.19, p. 154.

Browne, Sid (1995). "Optimal investment policies for a firm with a random risk process: exponential utility and minimizing the probability of ruin". In: *Mathematics of operations research* 20.4, pp. 937–958.

Burney, SM Aqil and MA Hashmi (2002). *Risk Theory and Insurance: A stochastic approach.* Bureau of Composition, Compilation & Translation, University of Karachi.

Cartea, Álvaro, Sebastian Jaimungal, and José Penalva (2015). *Algorithmic and high-frequency trading.* Cambridge University Press.

Cramér, Harald (1955). *Collective risk theory: A survey of the theory from the point of view of the theory of stochastic processes.* Esselte Stockholm.

Daley, Daryl J and David Vere-Jones (2003). *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer.

Fleming, Wendell H and Halil Mete Soner (2006). *Controlled Markov processes and viscosity solutions*. Vol. 25. Springer Science & Business Media.

Fleming, WH et al. (1975). "Applications of mathematics". In: *Deterministic and Stochastic Optimal Control*.

Grandell, Jan (1991). "Aspects of risk theory Springer-Verlag". In: *New York*.

Grimmett, Geoffrey and David Stirzaker (2020). *Probability and random processes*. Oxford university press.

Hawkes, Alan G (1971). "Spectra of some self-exciting and mutually exciting point processes". In: *Biometrika* 58.1, pp. 83–90.

Higgins, James J and Sallie Keller-McNulty (1995). *Concepts in probability and stochastic modeling*. Duxbury Resource Center.

Hipp, Christian (2004). "Stochastic control with application in insurance". In: *Stochastic methods in finance*. Springer, pp. 127–164.

Karatzas, Ioannis et al. (1998). *Methods of mathematical finance*. Vol. 39. Springer.

Laub, Patrick J, Thomas Taimre, and Philip K Pollett (2015). "Hawkes processes". In: *arXiv preprint arXiv:1507.02822*.

Lundberg, Fillip (1903). "Approximerad framställning av sannolikhetsfunktionen. aterförsäkring av kollektivrisker. Akad". PhD thesis. Afhandling. Almqvist och Wiksell, Uppsala.

Martin-Löf, Anders (1994). "Lectures on the use of control theory in insurance". In: *Scandinavian Actuarial Journal* 1994.1, pp. 1–25.

Merton, Robert C (1969). "Lifetime portfolio selection under uncertainty: The continuous-time case". In: *The review of Economics and Statistics*, pp. 247–257.

— (1975). "Optimum consumption and portfolio rules in a continuous-time model". In: *Stochastic optimization models in finance*. Elsevier, pp. 621–661.

Ogata, Yosihiko (1988). "Statistical models for earthquake occurrences and residual analysis for point processes". In: *Journal of the American Statistical association* 83.401, pp. 9–27.

Richard, Bellman (1954). "The theory of dynamic programming". In: *Bulletin of the American Mathematical Society* 60.6, pp. 503–516.

Rolski, Tomasz et al. (2009). *Stochastic processes for insurance and finance*. Vol. 505. John Wiley & Sons.

Ross, Sheldon M (2014). *Introduction to probability models*. Academic press.

Schmidli, Hanspeter (2017). "Risk Models". In: *Risk Theory*. Springer, pp. 1–33.

Segerdahl, CARL-OTTO (1959). "A survey of results in the collective theory of risk". In: *The Harald Cramer Volume*.

Stabile, Gabriele and Giovanni Luca Torrisi (2010). "Risk processes with non-stationary Hawkes claims arrivals". In: *Methodology and Computing in Applied Probability* 12.3, pp. 415–429.

Swishchuk, Anatoliy (2017). "Risk model based on general compound hawkes process". In: *arXiv preprint arXiv:1706.09038*.

Swishchuk, Anatoliy, Bruno Remillard, et al. (2019). "Compound Hawkes processes in limit order books". In: *Financial Mathematics, Volatility and Covariance Modelling*. Routledge, pp. 191–214.

Swishchuk, Anatoliy, Rudi Zagst, and Gabriela Zeller (2021). "Hawkes processes in insurance: Risk model, application to empirical data and optimal investment". In: *Insurance: Mathematics and Economics*.

Thorin, Olof (1974). "Some comments on the Sparre Andersen model in the risk theory". In: *ASTIN Bulletin: The Journal of the IAA* 8.1, pp. 104–125.

Valkeila, Esko (2008). *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure, by Daryl J. Daley, David Vere-Jones*.

Zeller, Gabriela FIM (2018). "Hawkes Processes in Insurance: Risk Modelling and Optimal Investment". In.

Appendix A

Python code for simulations and graphs

```
import numpy as np

import math

import scipy.stats

import matplotlib.pyplot as plt

def compute_a_star(pi_star,a):

    return sum(pi_star*a)

def compute_b(a_star,a):

    return a_star-a
```

```python
def compute_g(pi_star,b):

    P=[]

    for i in range(len(pi_star)):

        P.append(pi_star)

    P=np.array(P)

    a=np.linalg.inv(2*P-np.eye(len(pi_star)))

    return np.matmul(a,b.T)

def compute_v(b,g,pi_star):

    v=np.zeros((1,len(b)))

    v+=b*b

    for i in range(len(v)):

        for j in range(len(v)):

            v[i]+=((g[j]-g[i])**2)*pi_star[j]-2*b[i]*(g[j]-g[i])
            *pi_star[j]
```

```
    return v

def compute_sigma_hat(pi_star,lambbda,miu_hat,v):

    sigm_star=math.sqrt(np.sum(pi_star*v))

    sigma=sigm_star*math.sqrt(lambbda/(1-miu_hat))

    return sigma

def compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a):

    b=x*r+c-a_star*(lambbda/(1-alpha/beta))

    a=-sigma**2/2

    c=(G_a-r)**2/(2*G_b**2)

    result1=(-b+math.sqrt(b**2-4*a*c))/(2*a)

    result2=(-b-math.sqrt(b**2-4*a*c))/(2*a)

    return (result1,result2)

a_star=797.3672
```

```
pi_star=[0.9017, 0.0720, 0.0206, 0.0032, 0.0026]

pi_star=np.array(pi_star)

a=[499.5056, 2821.8888, 4743.6872, 7049.5920, 9199.8750]

a=np.array(a)

lambbda=0.1467

alpha=0.0260

beta=0.0334

b=compute_b(a_star,a)

g=compute_g(pi_star,b)

v=compute_v(b,g,pi_star)

miu_hat=alpha/beta

sigma=compute_sigma_hat(pi_star,lambbda,miu_hat,v)

G_a=0.12#drift dS=adt+bdw

G_b=0.15#volatility
```

```
c=633.5552

x=8000

r=0.05

p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)

theta=0.2

u=8000

pi=(G_a-r)/(x*(G_b**2)*p[1])

#dependency on G_a

r=0.05

G_a_dependencies=[]

for i in np.arange(0.12,0.15,0.001):

    G_a=i

    p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)
```

```
    pi=(G_a-r)/(x*(G_b**2)*p[1])

    G_a_dependencies.append(pi)

plt.xlabel("Drift")

plt.ylabel("Optimal value")

plt.plot(np.arange(0.12,0.15,0.001),G_a_dependencies)

#dependency on G_b

G_b_dependencies=[]

for i in np.arange(0.12,0.15,0.0001):

    G_b=i

    p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)

    pi=(G_a-r)/(x*(G_b**2)*p[1])

    G_b_dependencies.append(pi)

plt.xlabel("Volatility")

plt.ylabel("Optimal value")
```

```
plt.plot(np.arange(0.12,0.15,0.0001),G_b_dependencies)

r=0.05

a_star_dependencies=[]

for i in np.arange(800,2000,20):

    a_star=i

    p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)

    pi=(G_a-r)/(x*(G_b**2)*p[1])

    a_star_dependencies.append(pi)

plt.xlabel("Aberage claim size")

plt.ylabel("Optimal value")

plt.plot(np.arange(800,2000,20),a_star_dependencies)

x_dependencies=[]

for i in np.arange(8000,15000,20):
```

```
    x=i
    p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)
    pi=(G_a-r)/(x*(G_b**2)*p[1])
    x_dependencies.append(pi)
plt.xlabel("Initial capital")
plt.ylabel("Optimal value")
plt.plot(np.arange(8000,15000,20),x_dependencies)
r_dependencies=[]
for i in np.arange(0.05,0.8,0.001):
    r=i
    p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)
    pi=(G_a-r)/(x*(G_b**2)*p[1])
    r_dependencies.append(pi)
plt.xlabel("Interest rate")
```

```
plt.ylabel("Optimal value")

plt.plot(np.arange(0.05,0.8,0.001),r_dependencies)

lambda_dependencies=[]

for i in np.arange(0.14,0.8,0.01):

    lambbda=i

    p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)

    pi=(G_a-r)/(x*(G_b**2)*p[1])

    lambda_dependencies.append(pi)

plt.xlabel("Background intensity")

plt.ylabel("Optimal value")

plt.plot(np.arange(0.14,0.8,0.01),lambda_dependencies)

r=.05

c_dependecies=[]
```

```
r=.05
for i in np.arange(633,900,1):
    c=i
    p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)
    pi=(G_a-r)/(x*(G_b**2)*p[1])
    c_dependecies.append(pi)
plt.xlabel("Premium Rate")
plt.ylabel("Optimal value")
plt.plot(np.arange(633,900,1),c_dependecies)
#dependency on alpha
#try with different r 0.8 0.005
r=0.05
alpha_dependencies=[]
for i in np.arange(0.02,0.025,0.0001):
```

```
    alpha=i
    p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)
    pi=(G_a-r)/(x*(G_b**2)*p[1])
    alpha_dependencies.append(pi)
plt.xlabel("Alpha")
plt.ylabel("Optimal value")
plt.plot(np.arange(0.02,0.025,0.0001),alpha_dependencies)
#dependency on beta
r=0.9
beta_dependencies=[]
for i in np.arange(0.03,0.04,0.0001):
    beta=i
    p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)
```

```
    pi=(G_a-r)/(x*(G_b**2)*p[1])
    beta_dependencies.append(pi)
plt.xlabel("Beta")
plt.ylabel("Optimal value")
plt.plot(np.arange(0.03,0.04,0.0001),beta_dependencies)
r=.9
sigma_dependencies=[]
for i in np.arange(815,2000,1):
    sigma=i
    p=compute_p(x,r,a_star,lambbda,c,alpha,beta,sigma,G_b,G_a)
    pi=(G_a-r)/(x*(G_b**2)*p[0])
    sigma_dependencies.append(pi)
plt.xlabel("sigma")
plt.ylabel("Optimal value")
```

```python
plt.plot(np.arange(815,2000,1),sigma_dependencies)

def Ultimate_ruin_probability(c,a_star,lambbda,miu_hat,u,sigma):

    return np.exp(-(2*(c-a_star*lambbda/(1-miu_hat))*u)/sigma**2)

Ultimate_ruin_probability(c,a_star,lambbda,miu_hat,u,sigma)

#ruin_dependencies_c

c_ruin=[]

for i in range(800,1000):

    c=i

    c_ruin.append(Ultimate_ruin_probability(c,a_star,lambbda,
    miu_hat,u,sigma))

plt.xlabel("Premium rate")

plt.ylabel("Ruin probability")

plt.plot(np.arange(800,1000,1),c_ruin)

a_star_ruin=[]
```

```
for i in range(790,900):

    a_star=i

    a_star_ruin.append(Ultimate_ruin_probability(c,a_star,
    lambbda,miu_hat,u,sigma))

plt.xlabel("average claim size")

plt.ylabel("ruin probability")

plt.plot(np.arange(790,900,1),a_star_ruin)

#ruin_dependencies_lambda

lambda_ruin=[]

for i in np.arange(0.1,0.15,0.001):

    lambbda=i

    lambda_ruin.append(Ultimate_ruin_probability(c,a_star,
    lambbda,miu_hat,u,sigma))

plt.xlabel("background intensity")
```

```
plt.ylabel("ruin probability")

plt.plot(np.arange(0.1,0.15,0.001),lambda_ruin)

#ruin_dependencies_miu_hat

miu_hat_ruin=[]

for i in np.arange(0.6,0.78,0.001):

    miu_hat=i

    miu_hat_ruin.append(Ultimate_ruin_probability(c,a_star,
    lambbda,miu_hat,u,sigma))

plt.xlabel("branching ration")

plt.ylabel("ruin probability")

plt.plot(np.arange(0.6,0.78,0.001),miu_hat_ruin)

#ruin_dependencies_u

u_ruin=[]

for i in np.arange(8000,15000,1):
```

```
    u=i
    u_ruin.append(Ultimate_ruin_probability(c,a_star,
    lambbda,miu_hat,u,sigma))
plt.xlabel("initial capital")
plt.ylabel("ruin probability")
plt.plot(np.arange(8000,15000,1),u_ruin)
sigma_ruin=[]
for i in np.arange(800,1000,1):
    sigma=i
    sigma_ruin.append(Ultimate_ruin_probability(c,a_star,
    lambbda,miu_hat,u,sigma))
plt.plot(np.arange(800,1000,1),sigma_ruin)
def Ruin_probability(c,a_star,lambbda,miu_hat,u,sigma,tao):
    return scipy.stats.norm(0,1).pdf(-(u+(c-a_star*lambbda/
    (1-miu_hat)))*tao/(sigma*math.sqrt(tao)))+
    Ultimate_ruin_probability(c,a_star,lambbda,miu_hat,u,sigma)*
```

```
    scipy.stats.norm(0,1).pdf(-(u-(c-a_star*lambbda/(1-miu_hat)))
    *tao/(sigma*math.sqrt(tao)))

import Hawkes as hp

from matplotlib import pyplot as plt

model=hp.simulator()

model.set_kernel('exp')

model.set_baseline('const')

para={'mu':0.1467,'alpha':0.0260,'beta':0.0334}

model.set_parameter(para)

itv=[0,100]

T=model.simulate(itv)

model.plot_l()
```