2-25-2022 2:00 PM

Risk theory: data-driven models

Yang Miao, *The University of Western Ontario*

Supervisor: Sendova, Kristina P., *The University of Western Ontario*
Co-Supervisor: Jones, Bruce L., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

Abstract

Ruin theory studies the riskiness of an insurance portfolio by investigating the evolution of an insurer's surplus. The existing models often assume stationary increments of the surplus process, which is not always appropriate to describe the actual experience. In this thesis, we consider some modifications that are inspired by a real-life one-year term auto insurance data set to the existing risk models and investigate how these modifications affect ruin theory results.

We first explore potential surplus modelling improvements by investigating how well the available models describe an insurance risk process. To this end, we obtain and analyze a real-life data set that is provided by an anonymous insurer. Based on our analysis, we discover that both the purchasing process and the corresponding claim process have seasonal fluctuations. Some special events, such as public holidays, also have impact on these processes. In the existing literature, the seasonality is often stressed in the claim process, while the cash inflow usually assumes simple forms. We further suggest a possible way of modelling the dependence between these two processes. A preliminary analysis of the impact of these patterns on the surplus process is also conducted. As a result, we propose a surplus process model which utilizes a non-homogeneous Poisson process for premium counts and a Cox process for claim counts that reflect the specific features of the data.

Next, we study a risk model with stochastic premium income. It is assumed that both the premium arrival process and the claim arrival process are modelled by homogeneous Poisson processes, and that the premium amounts are modelled by independent and identically distributed random variables. After reviewing various known results of this model, a simulation approach for obtaining the probability of ultimate ruin based on importance sampling is derived. We demonstrate this approach by examples where the distribution of the sampling random variable can be identified. We then give other examples where we use fast Fourier transform to obtain an approximation of the sampling random variable. The simulated results are compared with known results in the existing literature.

In the last part of the thesis, we consider a risk model where both the premium income and the claim process have seasonal fluctuations. We obtain the probability of ruin based on the simulation approach presented in Morales (2004). We also discuss the conditions that must be satisfied for this approach to work. We give both a numerical example that is based on a simulation study and an example using a real-life auto insurance data set. Various properties of this risk model are also discussed and compared with the existing literature.

Keywords: Insurance, premium, risk theory, seasonality, simulation, surplus process.

Summary for Lay Audience

Insurance companies operate by charging their customers premiums, investing these premiums in the financial market, setting aside a sufficient amount of funds to pay out claims if the customers suffer certain losses. An insurance company must hold sufficient reserve so that it can settle claims at any given time. If the insurance company is not able to pay a claim, we say the insurance company is ruined. *Ruin theory* is a branch of actuarial science that studies an insurer's insolvency risk. A key component of such studies is to build a model to describe how the available funds to the insurer, or the *surplus*, changes over time. In the existing literature, many models have been built and various conclusions have been obtained.

In this thesis, we investigate how well the existing models fit the actual data. To this end, we obtained and analyzed a real-life insurance data set. We discovered that both the premium income and the claim payment have seasonal fluctuations. Some special events, such as major public holidays, also have impact on the cash flow of an insurer. We demonstrate how these features affect the insurance company's insolvency risk via a simulation example. We then study two different models, each featuring one of the characteristics of the data set. Specifically, we study a model under which the premium income is stochastic, and a model under which both the premium income and the claim payment have seasonal fluctuations. We show how to obtain the probability of ruin under these models via simulation. The behaviours of these models are demonstrated by numerical examples.

Co-authorship Statement

This work was completed under the supervision of Dr. Kristina Sendova. Materials from three jointly authored papers are contained in this thesis.

- Chapter 2 is based on a manuscript titled *Some observations on the temporal patterns in the surplus process of an insurer*, co-authored with Dr. Wei Huang, Dr. Bruce Jones, and Dr. Kristina Sendova. This manuscript has been submitted and is currently under review.
- Chapter 3 is based on a manuscript in progress titled *A simulation approach to a risk model with stochastic premium income*, co-authored with Dr. Kristina Sendova.
- Chapter 4 is based on a manuscript titled *On a risk model with dual seasonalities*, co-authored with Dr. Kristina Sendova and Dr. Bruce Jones. This manuscript has been peer reviewed by *North American Actuarial Journal* and is currently under revision.

I certify that I am the principal contributor for all these manuscripts. I would like to thank Dr. Kristina Sendova and Dr. Bruce Jones for their immense help with formulating ideas, conducting research, and writing. I would like to thank Dr. Wei Huang for providing an insurance data set.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Dr. Kristina Sendova and Dr. Bruce Jones. As my supervisor for both M.Sc. and Ph.D. programs, Dr. Kristina Sendova has provided me with invaluable support and inspiration. Her optimistic disposition helped me to face the many obstacles and setbacks during my study. The conversations with her were both informative and agreeable. This thesis would not have been possible without her guidance. Dr. Bruce Jones has provided me with immense help with both research and professional development. An erudite scholar, he often helped formulate new ideas during our discussions, which greatly broadened my horizon.

I would also like to thank my examining committee, Dr. Jiandong Ren, Dr. Douglas Woolford, Dr. Guangyong Zou, and Dr. Cary Chi-Liang Tsai, for their careful read of my thesis. Their suggestions have improved this thesis greatly.

Special thanks to Dr. Zhong Li who helped procure the data set. I wish to thank Mia Seong for her suggestions that improved the readability of this thesis. Sincere thanks to the professors and the staffs at the Department of Statistical and Actuarial Sciences at the University of Western Ontario for their kindness and assistance. Heartfelt thanks to Prof. Mary Millard, who introduced me to actuarial science and always encouraged me to be a better researcher and actuary. I wish to thank Dr. Ričardas Zitikis for the opportunity to participate in a research project partnered with an insurance company.

I wish to thank my family and friends. I am especially thankful to my parents for their patience, understanding, and support. Special thanks to my friends and fellow students, Lingzhi Chen, Junhe Chen, Ang Li, Yifan Li, Shanshan Liu, and Zhenxian Gong for the shared experiences that I will always treasure. Thanks to Junxiu Liu, Yumeng Sun, and Yuyao Lian for being my longtime friends who make life delightful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Uncertainty is an inherent trait of human lives. An individual could suffer loss or injury when adverse events happen. The possibility of such consequences is called *risk*. The practice of pooling and redistributing risks to protect against contingent financial loss has a history of many centuries. With the development and maturity of the insurance industry followed the need for rigorous study of an insurer's financial risk, that is the possibility of loss due to disadvantageous events. An insurer's insolvency risk is especially of interest, that is the possibility that the insurer is unable to completely fulfill its future financial obligations. These studies are now known collectively as *ruin theory* or *risk theory*. The earliest attempt in this field is attributed to Lundberg (1903) who proposed what is now known as the *classical compound-Poisson risk model*. Under this model, the funds available to an insurer, or the *surplus*, are modelled by a stochastic process

$$U(t) = u + ct - \sum_{k=1}^{N(t)} Y_k, \tag{1.1}$$

where u is the initial surplus, c is the premium rate, $N(t)$ is a homogeneous Poisson process that counts the number of claim arrivals up to time t, and Y_k is the amount of the kth claim arrival. No investment income is considered under this model. Figure 1.1 gives a sample trajectory of this process.

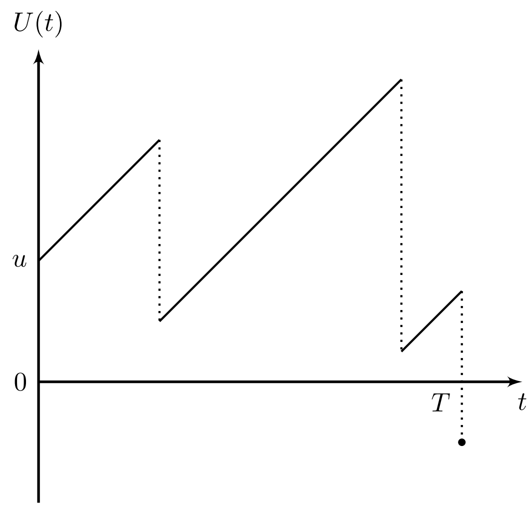


Figure 1.1: A sample trajectory of the classical compound-Poisson model

A negative surplus at any given time implies that the insurer does not have sufficient funds to settle the last claim. In this case, we say the insurer experiences *ruin*. The time of ruin is defined as $T := \inf\{t > 0 : U(t) < 0\}$. A key quantity of interest is the probability of ruin,

$$\psi(u) = \mathbb{P}\{\, T < \infty | U(0) = u \,\}.$$

Given that insurers usually charge a sufficiently large premium, the surplus process (1.1) has a positive drift and hence the probability of ruin is less than 1. Many ruin theory results under the classical model have been derived using probabilistic arguments. Over the years, many researchers attempted to generalize the surplus process and make it more pertinent to the actual practice. One notable attempt in this direction was made by Andersen (1957), where a renewal process is used to model the arrival of claims. More recently, many other extensions to the classical model have also been considered, see, for example, Schlegel (1998) who considered a risk model with a perturbation process, Morales (2004) who considered a risk model with periodic claim arrival intensity, Albrecher et al. (2011) derived explicit ruin formulas for models with dependence among claim sizes and among claim inter-occurrence times, Sendova et al. (2018) where the authors studied the impact of a constant dividend barrier under different risk processes with Parisian ruin, Yang et al. (2020) studied threshold dividend strategy under the dual Lévy risk model, among many others.

Since the advent of risk theory more than a century ago, many external factors have changed fundamentally. The insurance market has evolved and matured, and many new insurance products emerged. These new insurance products cover distinct perils and have drastically different cash flow patterns. The financial market has developed. New financial instruments have been designed and standardized, giving insurers higher flexibility to invest and to manage their financial risk. Consumer behaviour today is different from the 1900s, which affects all aspects of the economy. The changes in the physical environment have a strong impact on insurance practice. For instance, climate change alters the frequency and the severity of natural disasters, posing challenges to catastrophe insurance and property and casualty insurance. See, for example, the Actuaries Climate Index[1] for a monitoring tool of climate trends. Furthermore, the regulatory system is constantly developing. The goal of insurance regulation is focused primarily on solvency. This requires insurers to have a clear understanding of their assets and liabilities. All these changes require us to have better models to describe the operation of an insurer to better understand the risks an insurer takes.

The rapid advancements of information technology enabled storing and analyzing insurance data. One would be able to investigate how well the existing risk models describe the actual insurance practice with the help of data. Regrettably, most insurance data sets remain proprietary. Openly accessible insurance data sets are scarce and often contain only partial information on the business. For example, a compilation of insurance data sets is given in Dutang and Charpentier (2020). However, most insurance data sets contained in this `R` package have only claim information of an insurer. Furthermore, the data sets are often aggregated on an annual basis, meaning the timing of the claim

[1] url: https://actuariesclimateindex.org/home/

occurrences is not known. From a risk theory point of view, the granularity of these data sets is not sufficiently fine for validating existing risk models. Assuredly, an appropriate insurance data set containing all the necessary information is valuable to a researcher and can inspire new research. Some recent examples of research supported by real-life data set include a study of a fire insurance data set using a Cox process in Albrecher et al. (2021), and a study of an auto insurance data set using a Markov-modulated non-homogeneous Poisson process in Avanzi et al. (2021a).

We start this thesis by analyzing a real-life insurance data set. We obtained the data set from an anonymous insurance company. The data set contains one-year auto insurance records from the year 2013 to the year 2015. The sizes and the arrival times are recorded for both premiums and claims, making this data set appropriate for fitting a risk model. To investigate how well the existing models describe the actual data, we conducted an exploratory data analysis, which revealed the following characteristics:

- Seasonal effects are observed for both premium income and claim payments. For this specific data set, both the volume of insurance sales and the volume of claims are higher in winter and lower in summer. The length of one cycle is approximately one year. This indicates that it may be appropriate to consider a periodic structure in the risk model.

- The variability among premium amounts is high. This suggests that another stochastic process may be utilized to model the premium income.

- Both the premium and the claim processes exhibit increasing intensities, indicating that the insurance company grew during the study period.

- Extreme values are observed for both premium amounts and claim sizes. This means that both distributions have heavy tail property, meaning extreme values are likely to happen.

- Some other external factors, such as major public holidays, have impact on the surplus process. For this data set, we observe elevated volume of sales prior to major public holidays.

To the authors' knowledge, explicit solutions under risk models featuring these characteristics are usually difficult to obtain. In fact, the derivation of quantities of interest under mathematically tractable models often relies on the fact that the surplus process has stationary and independent increments, meaning it essentially restarts at each claim arrival with a new initial surplus. Given the features of the data set, these models may not always be appropriate.

For models with the aforementioned features, some results have been derived in the existing literature. These works are reviewed in Section 3.1 and Section 4.1 of this thesis. These theoretical results usually concern the bounds and some asymptotic properties of the probability of ruin. Some of these works studied the Gerber-Shiu function under different models and derived equations satisfied by the Gerber-Shiu function. Explicit expressions are derived for special cases, such as the case with exponentially distributed claim sizes. Although such results provide a great deal of insight into how these models behave, it remains challenging to implement the results in practice. Practitioners may find it difficult to compare different risk models and fine-tune them to fit their needs, much less use these models to assist risk management. In this thesis, we attempt to incorporate different characteristics of the data set and provide intuitive methods to investigate the impact of these features.

The rest of this thesis is structured as follows. In Chapter 2, we introduce and analyze a real-life insurance data set. Based on both exploratory analysis and quantitative methods, we summarize the characteristics of the data set. We suggest a possible way of modelling the dependence between the premium arrivals and the claim arrivals. We demonstrate how these features affect the riskiness of a portfolio by a simulation study of the probability of ruin within finite time horizon. Inspired by the findings here, we consider two different risk models in Chapter 3 and Chapter 4, each incorporating one feature of the data set. In Chapter 3, we consider a risk model with stochastic premium income. After reviewing the results in the existing literature, we derive an estimator based on importance sampling to obtain the probability of ruin. For cases where the importance sampling distribution is not easily identifiable, we propose to use fast Fourier transform as an approximation. Numerical examples are given to illustrate various results. In Chapter 4, we propose a risk model whose premium rate is periodic. After proving the equivalence of this risk model and the risk model proposed by Morales (2004), we demonstrate how to obtain the probability of ruin using the method therein. We also identify the condition that must be satisfied for this approach to work. We also fit different risk models to the data set for comparison. Finally, we end with conclusions and discussions of future research topics in Chapter 5.

Chapter 2

Some observations on the temporal patterns in the surplus process of an insurer

2.1 Introduction

Modelling the cashflows of an insurer is the foundation of a substantial amount of actuarial research. Among many different branches of actuarial science that require a surplus process model, ruin theory is a branch that studies the risks leading to and resulting from possible insolvency of an insurer. The classical ruin model, also known as the *compound-Poisson risk model*, is given by

$$U(t) = u + ct - \sum_{i=1}^{N(t)} Y_i, \quad u \geq 0,$$

where u represents the initial surplus of the insurer, c is the continuous premium rate, $N(t)$ is a counting process with initial value $N(0) = 0$ which counts the number of claims up to time t, and Y_i is the ith claim severity. Under the classical model, $N(t)$ is a homogeneous Poisson process. The homogeneous Poisson process has some properties that make mathematical analysis simpler compared to other counting processes. This is why it has been thoroughly studied. See, for example, Lundberg (1903), Cramér (1955), Dickson (1992), Gerber and Shiu (1998), and Lin and Willmot (2000) among many others.

Over the years, this model has been extended in many directions. In particular, Cramér (1955) introduced an aggregate premium process to replace the deterministic premiums. In 1957, Sparre Anderson proposed to use a renewal process for the claim counting process. Asmussen (1989) introduced a Markovian environment where the claim inter-arrival times, the claim sizes and the premiums were influenced by an external Markovian process. Embrechts and Schmidli (1994) studied a risk model where borrowing, investment and inflation were incorporated. Yang and Zhang (2001) proposed to model the insurer's surplus by a spectrally-negative Lévy process. Lin et al. (2003) introduced an absorbing barrier to model a dividend strategy. Subsequently, Lin and Pavlova (2006) extended this model to a threshold dividend strategy. Albrecher and Boxma (2004) and Boudreault et al. (2006) explored dependence structures linking interclaim times and claim amounts. Dassios and Wu (2008) allowed the insurer to have negative surplus for a fixed amount of time called *Parisian delay*. These are all theory-driven extensions of the classical ruin model.

In this chapter, we offer some data-driven extensions of the classical model. We study two data sets containing premium payments and claim payments for auto insurance policies. The goal of our study is to verify to what extent the assumptions in the classical ruin model and its variants, especially the assumptions on the counting processes, reflect the main features exhibited by the data.

The first data-driven modification of the classical model that we propose is related to the aggregate premiums. Namely, under the compound Poisson surplus process, the cumulative premium at time t is assumed to be equal to ct, which implies that the insurer collects premiums at a constant rate. This condition is also assumed in most of the ruin theory research. In practice, business growth is standard, i.e., insurers sell policies at increasing rates. This is supported by the data that we analyze in this paper. Moreover, our studies reveal seasonality in the premium process. These characteristics would alter the rate at which the insurer collects premiums. These findings agree with findings presented in previous works, such as Ellis (1974), where the author analyzes the time patterns of an American life insurance product. To model these features, it may be appropriate to choose another stochastic process for the premium income. The major theoretical advances in this direction come from the model proposed by Cramér (1955) where the premium process is a compound-Poisson process. This model has been studied in further depth by Boikov (2003), Labbé et al. (2011), Zhao and Yin (2012) and several others. After analyzing our data, we propose to generalize the aggregate premium process by replacing the homogeneous Poisson premium-counting process by a non-homogeneous Poisson process.

Secondly, we show that similar behaviours are also exhibited in the claim process. Consequently, we explore how this model should be further adapted to reflect the specific features of the data. This choice of a claim-counting process is supported by the studies of Lu and Garrido (2005) who demonstrate that such a model is an appropriate fit to hurricane data. Moreover, as noted in Beard et al. (1984) and Daykin et al. (1994), the insurer's risk process is often affected by long-term trends and short-term variations: features that are also prominent in our data, which exhibits a notable upward trend and seasonality. Our choice of claim-counting process is further supported by Morales (2004) who demonstrates how seasonality may be incorporated in a non-homogeneous Poisson process.

Finally, incorporating the policy purchasing process and the claim process, we propose a new surplus process model. Special consideration is given to the dependence between the purchasing and the claims. Under this new framework, the claim counting process becomes a *Cox process*, or *doubly stochastic Poisson process*, which has been constructed via different approaches in the existing literature, see for example Asmussen (1989), Guillou et al. (2015), Albrecher et al. (2021), and Avanzi et al. (2021a). This framework provides an intuitive way of incorporating the risk exposure born by the insurer into the surplus process model, and allows for separate consideration of time patterns in the purchasing process and the claim process. A comparison of goodness-of-fit between different candidate models shows that the proposed model captures more of the exhibited features in our data sets than other models that have been considered in the past. A simulation study is also conducted to evaluate some quantities of interest under the new model.

It should be noted that we are not aiming at providing an in-depth statistical analysis of the data; this should be done in a subsequent study. Instead, we want to verify to what extent theory-driven ruin models are supported by the data that we have. Subsequently, we suggest ways of improving the existing models that would account for specific features of the data. Again, further theoretical study of this model is delegated to subsequent works.

this chapter is structured as follows. The data set is described briefly in Section 2.2. In Section 2.3, we study temporal patterns of the premium-counting and claim-counting processes implied by the data sets. We propose a new surplus process model to incorporate the characteristics that we find. In Section 2.4, we obtain some ruin theory results using simulation. We also compare the proposed model with some existing models. Conclusions are drawn in Section 2.5.

2.2 Concise description of the data set

We obtained two data sets for the purpose of this chapter. Both data sets are provided by a regional insurance company who wishes to remain anonymous.

The first data set contains detailed information on the timing of the cash flows, including the timing of premium payments and claim payments. Consequently, the first data set is useful for building counting processes, which is the main goal of this chapter. For this reason, the first data set is used extensively in this chapter. We present this data set in this section.

The second data set is from a different region. It contains more details on the premium, such as when the premium was paid, and when the coverage started. Since it lacks necessary details on the claim payments, it is not used for the purpose of building counting processes in this chapter. Instead, it is used to validate the assumptions we use in the modelling process. This data set is described in Subsection 2.3.3.

The first data set contains 54,218 records of a one-year auto-insurance policy. The coverages of these policies started between January 1, 2013 and December 31, 2015. The policies are bundles of compulsory third-party liability coverage and additional coverage chosen by the policyholder. The covered perils include damage suffered by a third party and damage suffered by the policyholder.

The recorded information is:

- *Vehicle identifier* that uniquely identifies the vehicle.
- *Premium* is the single premium for the policy. The premium is paid in a lump sum.
- *Premium date* is the date when the premium was collected.
- *Accident date* is the date when an accident occurred. There are multiple records associated with the same policy if multiple accidents occur during the effective period.
- *Claim date* is the date when the claim was paid out.
- *Claim amount* is the full or partial amount that was paid to the policyholder to cover accident-related expenses.

The exact effective dates of these policies are not known, only the years in which these policies became effective are recorded. We use the premium date as a proxy to calculate the exposure at any given time.

This data set provides detailed information on premium date and claim date. Hence it is useful for building models for the premium-counting process and the claim-counting process. Although it is not the goal of this chapter and is not performed in this chapter, we may also build models for premium sizes and claim severities from recorded premium amounts and claim sizes. From ruin theory perspective, this should be sufficient to fit a risk model.

A brief summary of the first data set is given in Table 2.1.

Table 2.1: Summary of the first data set

Year	**Number of records**	**Number of claims**
2013	13,180	4,442
2014	17,632	5,527
2015	23,406	4,792

It may be further deduced that, based on the vehicle identifier, 6,152 customers who purchased this policy in 2013 also purchased a policy in 2014; 9,223 customers who purchased this policy in 2014 continued their coverage in 2015. There are 4,808 customers who appeared in all three years.

2.3 Characteristics deduced from the data sets

The second data set provides much more details on the timing of premiums, but little information on accident time and claim sizes. As a result, unless specifically stated, the following conclusions are all deduced from the first data set.

Specifying the premium process and the claim process should be sufficient for modelling the surplus process as the latter can be derived directly from the first two processes.

2.3.1 Characteristics of the premium arrival process

To verify how appropriate the classical ruin model is with respect to real premium processes, we first examine the premium process of our data. In the classical model, this process is simply represented by a straight line ct. Using the data, the cumulative premium over time may be easily obtained. A plot is given in Figure 2.1.

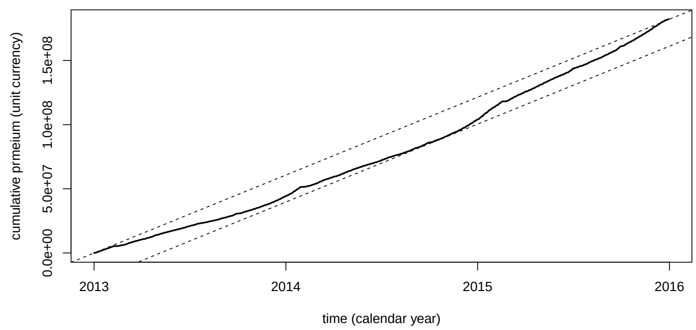


Figure 2.1: Cumulative premium with dashed auxiliary lines to show the convexity of the plot

Two dashed parallel auxiliary lines are added to emphasize the convexity of the cumulative premium. The convexity in the plot suggests that the insurer collects premium at an increasing rate. Increasing premium amount and/or increasing purchasing rate may cause the convexity that is observed on the graph. To investigate whether there is a change in the premium amount, we group the data by policy year and analyze the premium distribution for different years. The empirical distributions of the premiums are given in Figure 2.2.

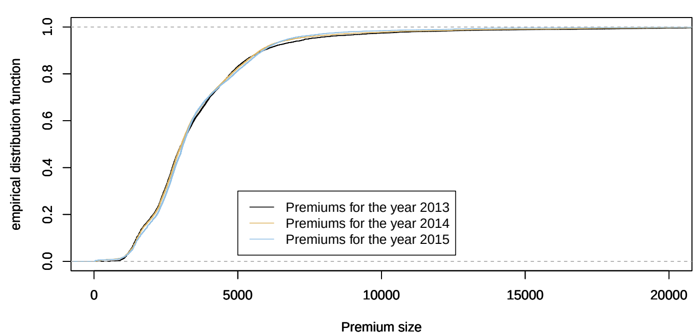


Figure 2.2: The empirical distribution of the premium sizes for different years

There is no evidence from the empirical distribution that the premiums are different for different years. More specifically, a summary of the data is given in Table 2.2.

Table 2.2: Summary of premium sizes by year

Year	1^{st} quartile	Median	Mean	3^{rd} quartile
2013	2282	3047	3668	4392
2014	2290	3039	3593	4363
2015	2370	3084	3571	4364

We may conclude then that the premium amount does not change over time, as all these statistics essentially point in that direction. It is also noteworthy that the inflation rate in the region where the insurer operates was relatively low during the studied period ($1\% - 2\%$ annual inflation rate, source: World Bank). The impact of inflation on the premium amounts over such a short period is negligible.

To check the trend in the purchasing process, we first investigate the daily sales of the policy which are illustrated in Figure 2.3. Since most sales happened during workdays, we see a strong weekly pattern. To show the yearly fluctuation, a 7-day moving average is added. The figure shows significant time structures, suggesting a more flexible premium income model would be more realistic.

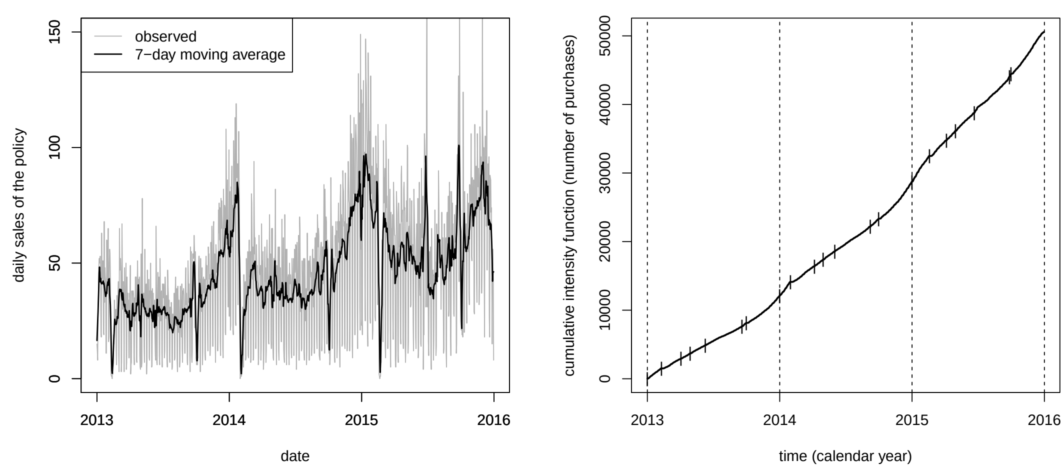


Figure 2.3: Left: Observed daily sales of the policy with 7-day moving average; Right: Observed cumulative sales of the policy, vertical short lines indicate public holidays

One way to extend the classical risk model is to use another compound Poisson process for the premium income. This model was proposed at the beginning of the 21st century, and has since generated some research. See, for example, Boikov (2003), Labbé and Sendova (2009), and Temnov (2014). We extend the model considered in these works by using a *non-homogeneous Poisson process* to allow for seasonal variations.

Different algorithms are developed to estimate the intensity function of a non-homogeneous Poisson process. For early references, see for example, Leemis (1991), Arkin and Leemis (2000), and Henderson (2003). Asymptotic properties of these estimators are derived in those works. Chernobai et al. (2007) demonstrated that when dealing with one realization of a non-homogeneous Poisson process, the cumulative number of arrivals can be used as an estimate of the cumulative intensity function. A detailed algorithm is also given therein. Using daily data, the plot of the estimated cumulative intensity function is given in the right panel of Figure 2.3. On this plot, we use vertical short solid lines to indicate public holidays in the region where the insurance company operates.

Some patterns of the data set are immediately noticeable on Figure 2.3. There is a clear yearly cycle in the sales of the policy, which means that the premium income is not uniform in a year. The estimated cumulative intensity function is convex, suggesting that the corresponding intensity function is increasing. Peaks are also observed around major public holidays, suggesting that these events affect the premium income.

We may also carry out quantitative analysis to investigate the temporal patterns of the purchasing process. We want to emphasize that the goal here is to simply observe whether there is any temporal pattern in the purchasing process. Using more sophisticated predictive models may be useful in prediction problems, but the complexity of such models blurs the characteristics of the data.

To estimate the cumulative intensity function, we first apply a polynomial regression model. We obtain the smoothly increasing component of the cumulative intensity function, and by subtracting this growth component from the overall intensity, we obtain the remaining cyclical component. To this end, we fit the curve

$$f(t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2$$

by minimizing the squared error.

We also apply certain time-series techniques to separate the long-term trend and the seasonality. Detrending is a topic that is explored in the time-series literature and that is often needed in practice. Different algorithms are developed, and many of them are readily available in various statistical programming languages. We use the *mFilter* package in R for our analysis. The filters we use are

- Hodrick-Prescott filter. This model assumes that a time series y_t can be viewed as the sum of a trend component g_t, and a cyclical component c_t,

  $$y_t = g_t + c_t + \epsilon_t, \quad t = 1, \ldots, T.$$

  In addition, the trend component is assumed to be smooth. The objective is to find the trend component

  $$g_t = \operatorname{argmin}\left\{\sum_{t=1}^{T}(y_t - g_t)^2 + \lambda \cdot \sum_{t=2}^{T-1}[(g_{t+1} - g_t) - (g_t - g_{t-1})]^2\right\},$$

  where the positive parameter λ controls the smoothness of g_t. A closed-form expression for g_t exists and may be expressed as matrix calculation. For the detailed algorithm, see Hodrick and Prescott (1997). Ravn and Uhlig (2002) showed that the parameter λ should be adjusted according to the fourth power of the frequency of observations. Based on this result, we choose $\lambda = 1600 \times 90^4 = 1.04976 \times 10^{11}$ for our daily data.

- Christiano-Fitzgerald filter. This is a special case of the *band-pass filters*. It is an approximation to the ideal infinite band pass filter. This method analyzes cycles with different frequencies in a time series data set. By setting cutoff frequencies, we may separate short-term shock and long-term trend. For details, see Christiano and

Fitzgerald (2003). As it is clear from Figure 2.3 that the period of the seasonality is approximately one year, we define cycles with period greater than 365 days to be long-term trend, and other cycles to be seasonality.

The seasonality and long-term trend of the cumulative intensity function captured by different algorithms are given in Figure 2.4.

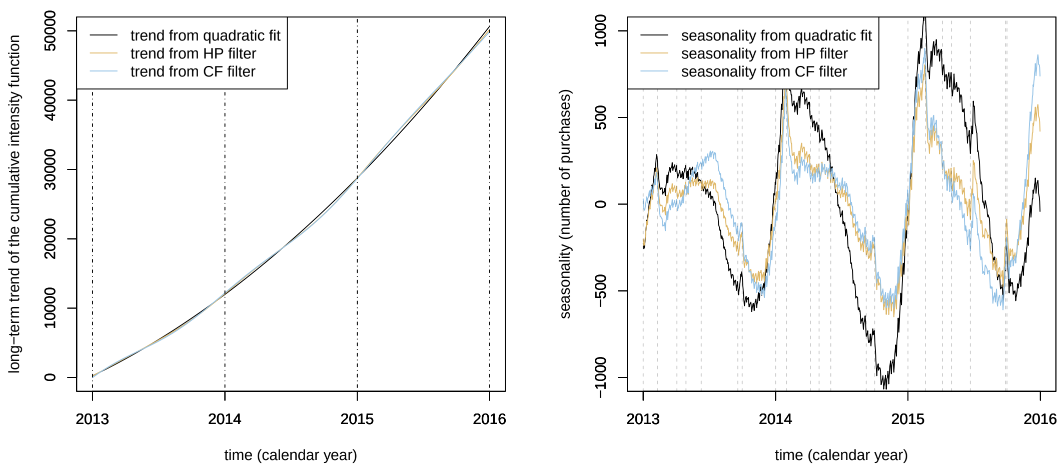


Figure 2.4: Left: Comparison of the long-term trend of the cumulative intensity function of insurance policy purchases captured by different algorithms. Right: comparison of the seasonalities of the cumulative intensity function of insurance policy purchases captured by different algorithms, doted vertical lines mark the public holidays in the region where the insurance company operates.

The long-term trends captured by different algorithms are practically identical. The convexity of the trend implies that the insurance company sells policies at an increasing rate. Both the HP filter and the CF filter yield virtually identical seasonalities, while the polynomial regression model yields a slightly different result. This is expected since the HP filter and the CF filter are non-parametric estimates and hence, they tend to be more flexible. All three curves have similar shapes. The derivatives of these curves are positive at the beginning and at the end of a year, and are negative in the middle of a year. Since the derivative of the cumulative intensity function is the intensity function, this result means that less policies are sold in the middle of a year compared to other times of the year.

We observe that there are peaks on the seasonality curve in Figure 2.4. Although public holidays usually fall on the same dates from year to year and hence, are themselves periodic, we may separate these events from the overall seasonality and quantify their impact. Two types of holidays are observed in the region where the insurance company operates. Most public holidays are 3 days in length, i.e., long weekends, while two holidays are 7 days in length. We only consider major public holidays that are 7 days in length. As these holidays are longer, they have larger impact on consumer behaviours.

We study the change of policy purchases for three periods around a holiday: 7 days prior to a public holiday, during public holidays, and 7 days immediately after a public holiday. We choose 7 days to eliminate possible weekly fluctuation in purchases. Figure 2.5 gives an illustration of these three periods assuming that October 1 – October 7 is a public holiday.

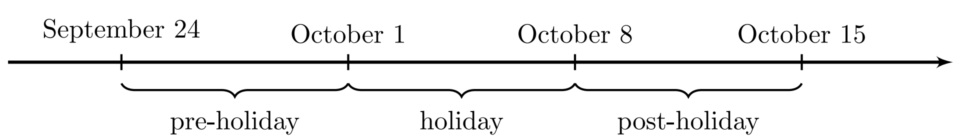


Figure 2.5: Illustration of three time periods around a public holiday, assume October 1 - October 7 are public holidays

Analyzing the impact of a specific event is frequently conducted across different disciplines. For count data, a generalized linear model is often used. Since we use a non-homogeneous Poisson process to model the sales of the policy, it is natural to use a Poisson regression model. This approach and its variations are explored in many works, see for example Chang et al. (2018). In light of our findings so far, we incorporate in our analysis components that reflect the long-term trend, the seasonality, the impact of weekends, together with the impact of public holidays. We assume

$$\begin{cases} H(t) \sim Poisson(\zeta(t)) \\ \log \zeta(t) = \beta_0 I_0(t) + \beta_1 I_1(t) + \beta_2 I_2(t) + \beta_3 I_3(t) + \gamma_0 + \gamma_1 t + \gamma_{cos} \cos\left(\frac{2\pi}{365} t\right) + \gamma_{sin} \sin\left(\frac{2\pi}{365} t\right), \end{cases} \tag{2.1}$$

where $H(t)$ is the number of policies sold in day t, and I_0, I_1, I_2, I_3 are indicator functions defined as

$$I_0(t) = \begin{cases} 1 & \text{day } t \text{ is in a pre-holiday period} \\ 0 & \text{otherwise} \end{cases},$$

$$I_1(t) = \begin{cases} 1 & \text{day } t \text{ is holiday} \\ 0 & \text{otherwise} \end{cases},$$

$$I_2(t) = \begin{cases} 1 & \text{day } t \text{ is in a post-holiday period} \\ 0 & \text{otherwise} \end{cases},$$

$$I_3(t) = \begin{cases} 1 & \text{day } t \text{ is weekend} \\ 0 & \text{otherwise} \end{cases}.$$

The maximum likelihood estimate*s* of these parameters are given in Table 2.3. The p-values for the null hypothesis that the coefficients equal to 0 are also present in this table.

The values of $\beta_i, i = 0, 1, 2, 3$, capture the impact of public holidays and weekends. On average, there is a 31% increase of the daily policy arrivals compared to the base level prior to a holiday, an 84% decrease during a holiday, and an 8% decrease after a holiday. These fluctuations explain the peaks that we observe in Figure 2.4. We also notice that there is a 67% decrease on weekends, which explains the pattern observed in Figure 2.3.

The phase of the seasonality may be obtained by the estimates of γ_{cos} and γ_{sin}. By

Table 2.3: Estimated results with corresponding significance level obtained using Poisson regression. Significance level: *** p-value < 0.1%, ** p-value < 1%, * p-value < 5%.

Coefficients	Estimate	exp(Estimate)	p-value	Significance level
β_0	0.2701	1.310039	$< 2 \times 10^{-16}$	***
β_1	-1.8147	0.162894	$< 2 \times 10^{-16}$	***
β_2	-0.0872	0.916525	0.000175	***
β_3	-1.1192	0.326527	$< 2 \times 10^{-16}$	***
γ_0	3.5940	36.37954	$< 2 \times 10^{-16}$	***
γ_1	0.0008	1.000787	$< 2 \times 10^{-16}$	***
γ_{cos}	0.2486	1.282246	$< 2 \times 10^{-16}$	***
γ_{sin}	0.0062	1.006248	0.114185	

trigonometry, we have

$$\hat{\gamma}_{cos}\cos\left(\frac{2\pi}{365}t\right) + \hat{\gamma}_{sin}\sin\left(\frac{2\pi}{365}t\right) = \alpha\cos\left(\frac{2\pi}{365}t + \omega\right),$$

where

$$\begin{aligned} \omega &= \arctan\left(-\frac{\hat{\gamma}_{sin}}{\hat{\gamma}_{cos}}\right) = -0.0249 \approx -0.0079\pi, \\ \alpha &= \frac{\hat{\gamma}_{cos}}{\cos(\omega)} = 0.2487. \end{aligned} \tag{2.2}$$

Equation (2.2) indicates that the seasonality component is a cosine function. This is consistent with the seasonality obtained in Figure 2.4.

Using the estimates in Table 2.3, the intensity function (2.1) becomes

$$\begin{aligned} \log\zeta(t) = 0.27I_0(t) - 1.81I_1(t) + -0.09I_2(t) - 1.12I_3(t) \\ + 3.5940 + 0.0008t + 0.2487\sin\left(\frac{2\pi}{365}(t + 89.8)\right). \end{aligned}$$

In this section, we considered different components in the purchasing process. As shown in Figure 2.3, Figure 2.4, and subsequent analysis, the intensity function of the Poisson process should incorporate long-term trend, seasonalities, and impact of public holidays. This analysis could be performed periodically to ensure the estimate reflects the latest trends.

2.3.2 Characteristics of the claim count process

We next analyze the patterns exhibited by the claim process over time. In this subsection, we study the claim process as a stand-alone process. In Subsection 2.3.3, we explore a possible relationship between the purchasing process and the claim process. Analyzing and improving the claim process of a risk model is the focus of a great deal of research. As shown in many papers, the seasonal trend in the claim process is prominent. As the technique we use here is identical to that in Subsection 2.3.1, we omit the technical details and simply state the results.

Table 2.4 summarizes the claims by year and season. As shown in this table, more claims happened in winter. A chi-square test on the null hypothesis that the claim arrivals are uniformly distributed throughout a year yields a test statistic of 117.02 with 3 degrees of freedom. The corresponding p-value is almost 0. We reject the null hypothesis that the claims are uniformly distributed within a year and believe that the claim frequency varies by season.

Table 2.4: A summary of the claims by season, spring: March–May, summer: June–August, fall: September–November, winter: December–February

Year	**Spring**	**Summer**	**Fall**	**Winter**	**Total**
2013	1,072	1,003	1,094	1,273	4,442
2014	1,334	1,221	1,335	1,637	5,527
2015	1,088	1,165	1,211	1,328	4,792
Total	3,494	3,389	3,640	4,238	14,761
Percentage	23.67%	22.96%	24.66%	28.71%	

Using the same algorithms, the estimated cumulative intensity function and the seasonality are plotted in Figure 2.6. The cumulative intensity function is again convex, which means that claims are paid more and more frequently. A possible explanation is explored in Subsection 2.3.3. Also, there are more claims in winter than in summer, which is consistent with Table 2.4.

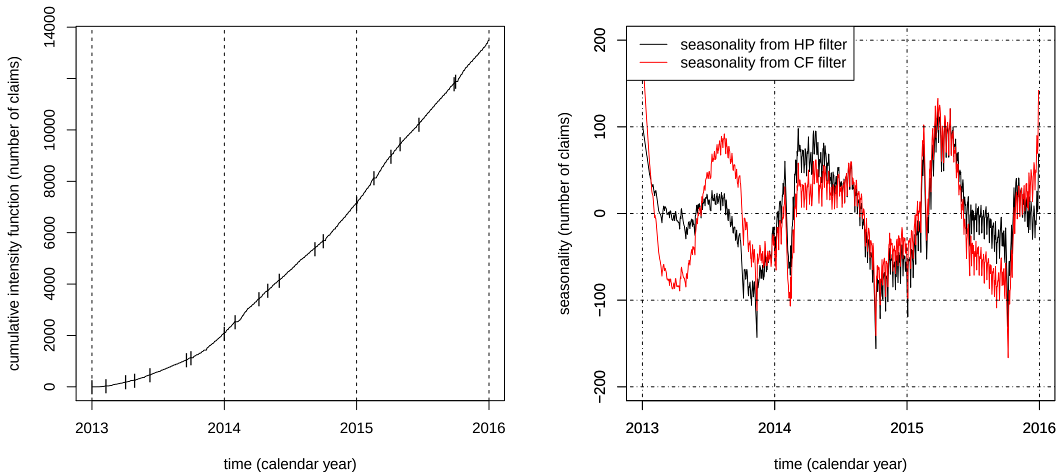


Figure 2.6: Left: estimated cumulative intensity function. Right: seasonal patterns of the claim process

We notice that the seasonal pattern in the first year is different from that in the last two years. As mentioned in Section 2.2, only policies that became effective after the year 2013 are included in the data set. As a result, the claim information for the policies that became effective in the year 2012 but extended to the year 2013 are not available. In other words, the data set does not contain all the claims for the year 2013. The incomplete claim information needs to be considered in tandem with the purchasing information to be reasonable. We commence this analysis in the next section.

2.3.3 Relations between the two processes

In this subsection, we investigate a possible relation between the two counting processes. We discover in Subsections 2.3.1 and 2.3.2 that both counting processes have increasing intensity, and that seasonal patterns are present in both processes. It is natural that both the arrival of premiums and the arrival of claims become more frequent as the business grows. On the other hand, it is not immediately evident whether the same driver causes the seasonalities in both premiums and claims. Indeed, if the insurer is responsible for more policyholders for some time of a year, then one would expect more claims in that period. In view of this, we explore the evolution of the *exposure* of the insurer over time. This is a natural extension to the *collective risk model*, see Chapter 9 of Klugman et al. (2012).

Let $M(t)$ denote the non-homogeneous Poisson process for the premium arrivals, let $\mu(t)$ be the intensity function of $M(t)$. Based on the results in Subsection 2.3.2, let

$$\mu(t) = \kappa + g(t) + s(t),$$

where κ is a constant representing the initial business scale, $g(t)$ and $s(t)$ are two generic functions representing the growth component and the seasonality component respectively. Let $s(t)$ have a period of 1, i.e., $s(t+1) = s(t)$ for all $t \geq 0$. Let ℓ be the term of the insurance policy. We further assume that the policy is in force immediately upon purchasing. The exposure of the insurer at time t, denoted by $\xi(t)$, is then

$$\xi(t) = \begin{cases} M(t), & t \leq \ell \\ M(t) - M(t-\ell), & t > \ell \end{cases}. \tag{2.3}$$

Notice that the exposure is again a stochastic process that is driven by $M(t)$. The expected exposure at time t and its derivative is then

$$\mathbb{E}[\xi(t)] = \begin{cases} \int_0^t \mu(r)\mathrm{d}r, & t \leq \ell \\ \int_{t-\ell}^t \mu(r)\mathrm{d}r, & t > \ell \end{cases},$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\xi(t)] = \begin{cases} \kappa + g(t) + s(t), & t \leq \ell \\ [g(t) - g(t-\ell)] + [s(t) - s(t-\ell)], & t > \ell \end{cases}. \tag{2.4}$$

Consider $[s(t) - s(t-\ell)]$ in equation (2.4). This component equals 0 if and only if the term of the insurance policy is an integer. The result may be easily extended to the scenario where $s(t)$ has a period other than 1. We conclude that the exposure process does not have seasonality if the term of the insurance policy is a multiplier of the period of the seasonality exhibited in the premium arrival process. Given that the period of both the premium arrival process and the claim process is 1 year in the data set, the exposure process does not have seasonal fluctuation. Figure 2.7 is the observed exposure from the data set. A linear function is fitted to the estimated exposure using the least-squared-error estimates. There is no apparent yearly fluctuation.

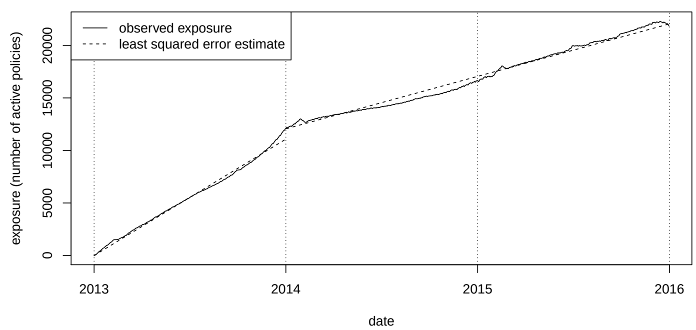


Figure 2.7: Estimated exposure, using policy purchasing dates as effective dates

Another simplification assumption we make is that an insurance policy becomes effective immediately upon purchasing. Usually a policy is purchased some days prior to when it becomes effective. We note that the date of purchasing is usually a good proxy for the effective date. To this end, we examine the second data set which contains both the date when a policy is sold and the date when the policy becomes effective. Figure 2.8 gives the histogram of the difference between these two dates. More precisely, 56% of the policies became effective on the same day or the second day of purchasing, while 80% of the policies became effective within one week.

In conclusion, if the seasonal fluctuation remains the same for different years, and the term of the insurance policy is a multiplier of the period of the seasonality of the premium process, then the exposure does not have seasonal fluctuations. Otherwise, the seasonality in the premium process will impact the claim process. One may track different causes of seasonality for different processes and determine whether there is an interaction between them based on the specification of the portfolio.

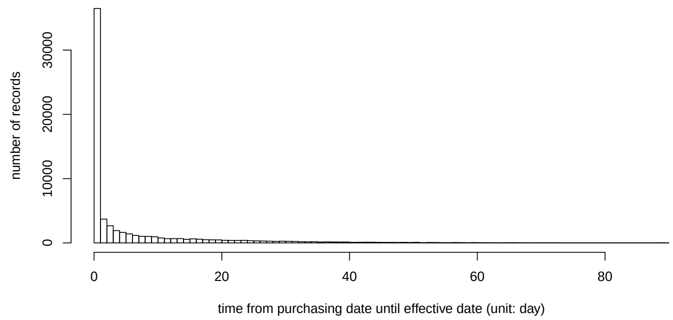


Figure 2.8: Histogram of the time difference between purchasing and effective date

2.3.4 A Cox process modeling the claim arrivals

We argue in Subsection 2.3.3 that the claim process needs to be considered in tandem with the purchasing process. We now explore an alternative model for the claim process which encompasses the intrinsic interactions between the purchasing of insurance and the resulting claims, namely a Cox process model.

We assume that the claim experiences for different policyholders are independent, and that the claim arrivals for each exposure follows a non-homogeneous Poisson process with intensity function $r(t)$. As before, suppose the exposure at time t is $\xi(t)$. Let $N(t)$ be the total number of claims at time t for the entire portfolio. We have the following proposition.

Proposition 2.1. *Suppose $N_1, N_2, \cdots$ are independent non-homogeneous Poisson processes with common intensity function $r(t)$, and let $\xi(t)$ be a stochastic process with integer values. Then $\sum_{i=1}^{\xi(t)} N_i$ is a Cox process.*

Proof. It is known that the superposition of independent non-homogeneous Poisson processes is a non-homogeneous Poisson process. Suppose $\xi(t) = k$, then the claims resulting from these k exposures follow a non-homogeneous Poisson process with intensity function $k \cdot r(t)$. The claim-counting process, conditional on the exposure $\xi(t)$, is a Poisson process, and thus, the unconditional process is a Cox process. □

To compare the performance of different models, various models are fitted to the first data set introduced in Section 2.2. The models for the claim process are 1) compound Poisson process (HPP) 2) non-homogeneous Poisson process with seasonal claim rate (NHPP) 3) Cox process. Notice that among the three models, only the Cox process model allows for the adjustment to the exposure. Since we only have partial information for the exposure for year 2013, the other two models would be unsuitable. For comparison reasons, all three models are fitted to the data from the last two years that are available and then projected to all three years.

The estimates are obtained by maximizing the likelihood function in a similar way as was done the analysis in Subsection 2.3.1. While fitting the Cox process, the observed exposure, as shown in Figure 2.7, is used as the offset. Figure 2.9 compares the outputs of the three different models. The Cox model is sufficiently flexible for modelling both the increasing trend and the seasonality in the claim process. Furthermore, the Cox model predicts more variability in the claim process caused by the fluctuations in the exposure. Among these three different models, the Cox model provides the closest fit to the data.

Some basic statistics are provided in Table 2.5. We compare the sum squared error and the Akaike Information Criteria (AIC) for different models. As shown by these statistics, the Cox process yields the smallest error, and hence, is the closest to the data.

Table 2.5: Summary statistics for model comparison

	HPP	NHPP	Cox
$\sum e_i^2$	72361	71664	68724
AIC	6250	6221	6083

(a) statistics for 2014-2015

	HPP	NHPP	Cox
$\sum e_i^2$	120737	119788	80018
AIC	11205	11154	8123

(b) statistics for 2013-2015

Remark 2.2. *The differences between any two models in terms of error statistics for this data set is relatively small. This is because the data set covers a short time period. The increment in the exposure is relatively small compared to its magnitude, and as a result, all three models provide a reasonable fit. We choose the Cox model because it is closest*

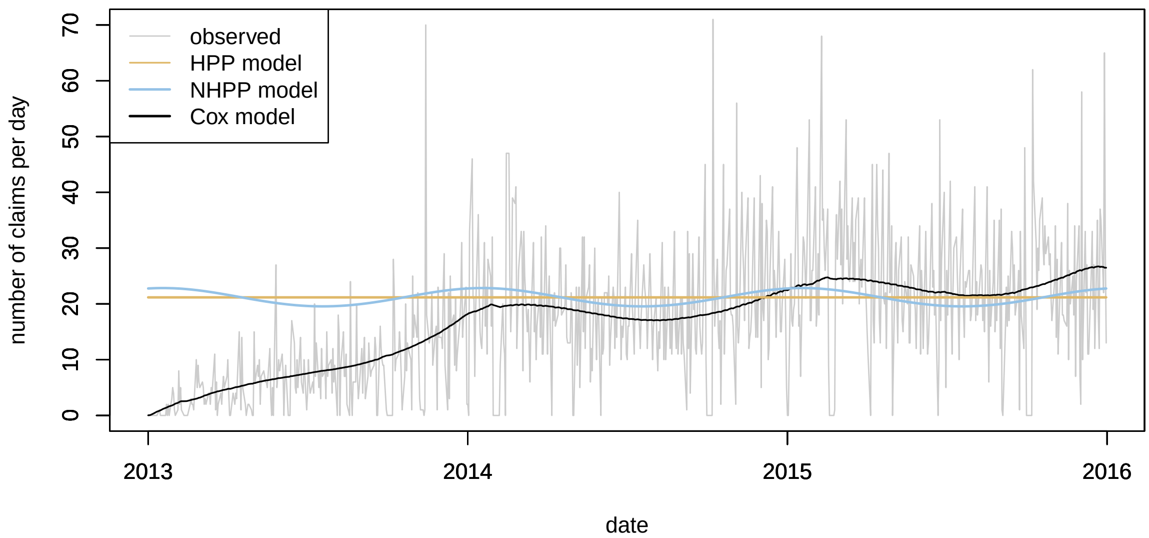


Figure 2.9: Comparison of three different numbers of claims. Three models are fitted to the data, then the predicted number of claims are calculated and plotted for each model.

to the data, and it allows us to investigate the dependence between the purchasing process and the claim process. We also point out that by allowing the claim rate to increase with time, we could obtain better results under the NHPP model. But doing so implies that the claim experience is deteriorating indefinitely, which is not realistic. □

2.4 A surplus process model with dual seasonalities and simulation studies

Integrating our findings in the previous sections, we propose to modify the classical ruin model in the following way to reflect the patterns exhibited in the data set:

$$U(t) = u + \sum_{k=1}^{M(t)} X_k - \sum_{i=1}^{N(t)} Y_i, \quad u \geq 0, \tag{2.5}$$

where u is the initial surplus, $M(t)$ is a non-homogeneous Poisson process that counts the number of policies sold by time t, X_k is the premium charged for the k-th policy, $N(t)$

is a Cox process that counts the number of claims by time t, and Y_i is the size of the ith claim. Denote the intensity functions of $M(t)$ and $N(t)$ by $\mu(t)$ and $\nu(t)$, respectively. The dependence structure between $M(t)$ and $N(t)$ is given by

$$\nu(t) = r(t) \cdot \xi(t) = r(t) \cdot [M(t) - M(t-\ell)],$$

where $\xi(t)$ is the exposure at time t, ℓ is the duration of the insurance policy, and $r(t)$ is a periodic function that accounts for different claim rates in a year.

We illustrate some properties of this model by employing Monte-Carlo simulation. We first investigate how the proposed model affects quarterly risk measures. We consider three different surplus process models: Model 1 (M1) has dual seasonalities and Cox claim arrivals, Model 2 (M2) uses stochastic premiums but the seasonality is only present in the claim process, and Model 3 (M3) has dual seasonalities and uses deterministic premium income and non-homogeneous Poisson claim arrivals. For simulation purposes, we use the previous estimate (4.12) without the terms representing growth or impact of holidays. An estimate of the claim rate function $r(t)$ is also obtained from the data set. More specifically, the intensity functions used in the simulation are

$$\mu(t) = 365 \exp(3.5940 + 0.2487 \cdot \sin(2\pi(t + 0.246027))), \tag{2.6}$$

$$\nu(t) = \xi(t) \cdot [0.488972 + 0.074706 \cdot \sin(2\pi(t + 0.120373))], \tag{2.7}$$

where $\xi(t)$ is the exposure at time t. Notice that to determine the evolution of the exposure, we need to know when the existing policies expire which depends on the premium arrivals in the previous year. To this end, for models using Cox process, we simulate the premium arrivals for the interval $[-1, 1]$ in order to obtain a sample path of the exposure on $[0, 1]$. Finally, we use the empirical premium-size distribution and the empirical claim-size distribution. We simulate 1,000,000 sample paths for each surplus model. To show how the seasonalities change the risk within a year, we obtain the quarterly risk measures from the simulation. The risk measures considered are value at risk (VaR) and tail value at risk (TVaR). These risk measures have intuitive interpretations and are used in many regulation systems. The results are given in Table 2.6.

Table 2.6: VaR and TVaR of loss in millions at quarter ends under different models

quarter	measure	M1	M2	M3
Q1	VaR(99.5)	- 2.790	-0.849	-2.863
	TVaR(99.5)	-2.354	-0.432	-2.450
Q2	VaR(99.5)	-4.690	-4.619	-4.804
	TVaR(99.5)	-4.081	-3.424	-4.197
Q3	VaR(99.5)	-7.874	-9.729	-7.993
	TVaR(99.5)	-5.964	-5.690	-6.019
Q4	VaR(99.5)	-13.855	-13.876	-13.975
	TVaR(99.5)	-8.499	-7.205	-8.559

Since the insurer charges sufficiently high premiums for this policy, all the risk measures are negative. The differences between Model 1 and Model 2 are due to the effect of the seasonality in the premium arrivals, while the differences between Model 1 and Model 3 are due to the effect of using a stochastic premium process. We observe that the seasonality in the premium arrivals causes differences in the risk measures. The impact of using a stochastic premium process is mild in this case. This is because the intensity functions used in the simulation have very large values. Consequently, the non-homogeneous Poisson premium arrivals may be well approximated by a deterministic function. For further discussion, see Temnov (2004), Section 5. We note that the differences between Model 1 and Model 3 may be greater for other types of insurance whose premium arrivals and claim arrivals are less frequent.

We may also consider the probability of ruin, that is the probability that an insurer is depleted of available funds to settle claims. For the following simulation, we allow the seasonal components of equations (2.6) and (2.7) to shift horizontally. In other words, the intensity functions we use in the following simulation are

$$\mu(t) = 365 \exp(3.5940 + 0.2487 \cdot \sin(2\pi(t-a))),$$

$$\nu(t) = \xi(t) \cdot [0.488972 + 0.074706 \cdot \sin(2\pi(t-b))],$$

where $\xi(t)$ is the exposure at time t calculated by equation (2.3) . We allow the sinusoidal functions to shift horizontally to capture the impact of different combinations of seasonalities. Although in the data set that we analyzed, premium arrivals and claim arrivals have peak seasons around the same time in a year, it is possible that the two seasonalities are overall unsynchronized. By shifting the functions representing the seasonalities along the horizontal axis, we are able to accommodate for such a difference. The initial surplus u is assumed to be 0 in the simulation, and empirical distributions are used for premium sizes and claim sizes. We define the time of ruin τ as the first passage time when the surplus drops below 0, i.e.,

$$\tau = \inf\{t : U(t) < 0\}.$$

The one-year ruin probability is then

$$\Psi(u; 1) = \mathbb{P}\{\, \tau \leq 1 | U(0) = u \,\}.$$

We consider different combinations of a and b. The one-year ruin probability is given in Figure 2.10. Previously, similar work has been done for the case where seasonality is only present in the claim process. See, for example, Morales (2004). Figure 2.10 shows that the seasonality in the premium process also has impact on the riskiness of the business.

2.4.1 Discussions and comparisons

We note that the seasonality in the claim process is well observed, and has been the focus of many industry studies and research papers. For example, an industry study jointly conducted by the Casualty Actuarial Society, the Property Casualty Insurers Association of America, and the Society of Actuaries in 2018[1] examines the drivers of

[1] Auto Loss Cost Drivers: Physical Damage, August 2018, url: https://www.soa.org/globalassets/assets/files/resources/research-report/2018/auto-loss-cost-drivers.pdf

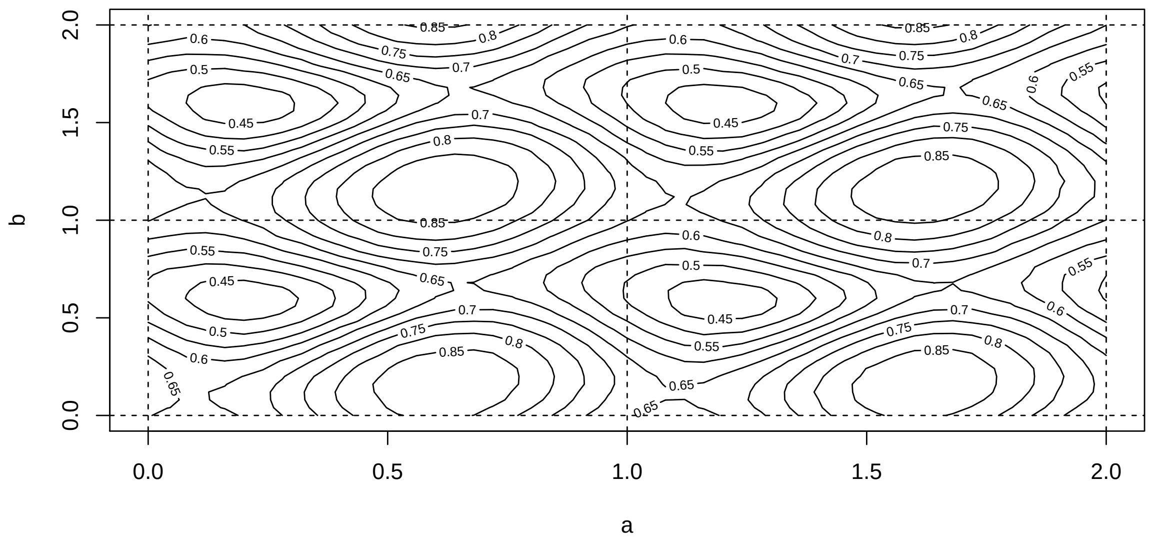


Figure 2.10: contour plot of one-year ruin probability with different combinations of seasonalities: a represents the initial season of the premium process, b represents the initial season of the claim process.

collision and comprehensive frequency and severity. A number of externals factors, such as natural disasters and hailstorms, are identified to have impact on the claim process. Clear seasonal fluctuations are also documented. Many researchers use non-homogeneous Poisson process to model these characteristics. For example Lu and Garrido (2005) use a NHPP with both long-term trend and short-term fluctuation to model hurricane arrivals. Morales (2004) considers a risk process where the claim arrivals are modelled by a periodic NHPP and derives a simulation method to obtain the probability of ruin.

If the claim process is assumed to be directed by an observable or unobservable driver, then a more flexible counting process than NHPP is needed. Specifically, if the driver is itself stochastic, then a Cox process is a natural choice for modelling. Albrecher et al. (2021) construct a Cox process by using a subordinator and demonstrate the success of this model using Dutch fire insurance claims. Avanzi et al. (2021a) construct a Markov-modulated Poisson process to account for both known information and unobservable driver where the underlying environmental process that impacts the event arrival intensity is assumed to be unobservable and is modelled by a continuous-time Markov chain, while the known exposure process serves as input in the hidden Markov chain calibration. While the model is different, similar results are obtained in this chapter. Other research projects are dedicated to improve premium modelling. The seasonality in the premium is discussed by Asmussen and Rolski (1994) where the constant premium income is replaced by a deterministic periodic function. The authors point out that one may obtain an equivalent risk model with constant premium rate by using a change of timeline technique. This approach works if the premium is deterministic. Some recent papers study the stochastic premium model, see for example Boikov (2003), Labbé and Sendova (2009), and Temnov (2014). These studies use homogeneous Poisson processes to model both the premium arrival and the claim arrival, and some theoretical results are obtained. While this approach extends the classical model, the premium arrival is assumed stationary, and the premium and the claims are assumed independent. Consequently, these models are unable to capture more variability in the risk process, such as seasonalities and dependence between premiums and claims.

By using the proposed model (2.5), we are able to incorporate the characteristics of the data set in the risk model, as well as to explicitly connect the claim process with the premium. The dependence has impact on the riskiness of the portfolio, especially when the premium arrivals have larger variation. For example, assume the intensity function for the premium arrivals is given by

$$\mu(t) = 100 + 25\cos(2\pi(t-a)) + 25\cos\left(\frac{2\pi t}{5}\right),$$

where an additional periodic function is added to represent economic cycles. This additional component adds more variation to the premium arrivals. Recall that in this case, the term of the insurance policy is not a multiplier of the period of the seasonality, and hence, the exposure itself has fluctuations. Consider two different claim arrival process:

$$\text{Cox model:} \quad \nu(t) = \xi(t) \cdot [0.1 + 0.05\cos(2\pi(t-b))],$$

$$\text{NHPP model:} \quad \nu(t) = 100 \cdot [0.1 + 0.05 \cos(2\pi(t-b))],$$

i.e., the dependence between the premium and the claim is only considered in the Cox model. The 10-year ruin probabilities using these two models are given in Figure 2.11. In this scenario, the model with dependence is able to capture the additional risk.

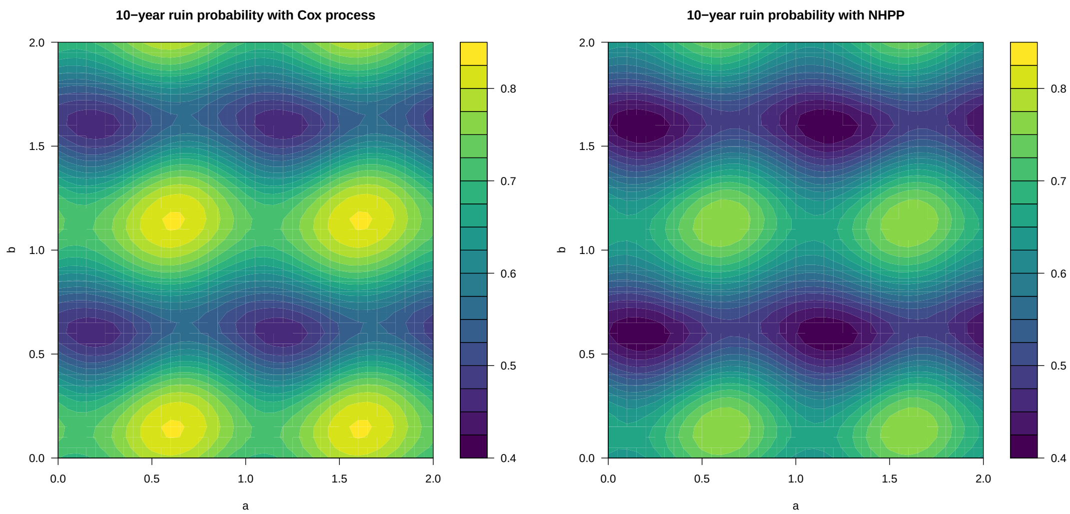


Figure 2.11: contour plot of 10-year ruin probability with different combinations of seasonality of the premiums (a) and seasonality of the claims (b). Left: the claim arrivals form a Cox process, right: the claim arrivals form a NHPP.

One possible application of the proposed model is to provide insights into how to determine the capital that an insurer is required to hold. Insurance companies are subject to the regulations applicable in the jurisdiction where they operate. For instance, Solvency II codifies the European Union insurance regulations, insurers in the United States are required to meet the risk-based capital (RBC) requirements, while the Life Insurance Capital Adequacy Test (LICAT) developed by Canadian regulators measures the capital adequacy of an insurer. These insurance regulations are focused primarily on solvency. Using Solvency II as an example, insurers are required to hold eligible own funds covering the Solvency Capital Requirement (SCR). An insurer may use full or partial internal models, upon approval from supervisory authorities, to better align the SCR calculation to its operation. The proposed model in this paper may contribute to the understanding of various components of the calculation. For example, the model directly contributes to the understanding of "the risk of loss resulting from fluctuations in the timing, frequency and severity of insured events, and in the timing and amount of claim settlements" (Solvency II, Article 105). The proposed model may also serve to link different components of the SCR. For example, insurers are required to consider the operational risk. During peak seasons, due to the elevated pressure on the resources needed for processing new policy purchases or settling claims, the insurer might face higher operational risk. An industry study[2] found that among insurance companies who use internal models for operational risk, a significant proportion of them are taking the approach of modelling frequency and severity separately. The proposed model allows an insurance company to investigate the correlation between the operational risk and other risks, which improves the accuracy of the internal models.

2.5 Conclusion

In this chapter, we study the time patterns of the premium and the claim processes of an insurer. We find that both processes exhibit increasing intensities with seasonal fluctuations. Major public holidays also have impact on these intensities. Further, we find that under certain conditions, the seasonality in the claim process is independent of the seasonality in the purchasing process. Based on these characteristics exhibited in the data set, we propose a new model for the surplus process that utilizes both a non-homogeneous Poisson process and a Cox process as counting processes.

The model suggested in this chapter allows one to gain more flexibility in modelling the surplus process. The special choice of non-homogeneous Poisson process for the purchasing process reflects the arrival of purchasing more closely. The Cox process used in the claim process takes into consideration the change of exposure over time and therefore

[2] Operational risk modelling: common practices and future development, url:https://www.theirm.org/media/6809/irm_operational-risks_booklet_hi-res_web-2.pdf

is capable of modelling more variability. As this model is intuitive and each component in it has a direct interpretation, the parameters of this model are also easy to estimate from the data.

A possible application of this result is to help insurers better manage their assets. Insurers need to set aside reserves according to regulations to meet their future obligations. By studying the time patterns of cash inflow and cash outflow, insurers are in a better position to optimally manage their assets, both to achieve higher profitability and financial security.

Due to the lack of appropriate data, we did not examine data from different regions. Although the proposed model is general enough to handle different situations, it might be the case that the specific time patterns are different in different regions. To this end, more data sets should be examined. Theoretical results are yet to be derived under this model. Simulation techniques may be employed to obtain the results of interest.

Chapter 3

A simulation approach to a risk model with stochastic premium income

3.1 Introduction

Modelling the cashflow of an insurer is at the core of much actuarial research. The study of this subject dates back to the early 1900s by the work of Lundberg (1903). To model the cashflow of an insurer, it suffices to consider the premium income and the claim payment. In the most basic setting, the investment is not considered, the rate at which the insurer collects premium is assumed to be constant, and the claim arrivals are modelled by *homogeneous Poisson process*. This model is known as the *classical risk model*, also known as the *compound-Poisson risk model*. Under this model, the surplus process of an insurer is given by

$$U(t) = u + ct - \sum_{k=1}^{N(t)} Y_k, \tag{3.1}$$

where u is the initial surplus of the insurer, c is the continuous premium rate, $N(t)$ is a homogeneous Poisson process which counts the number of claim arrivals up to time t, and Y_k is the amount of the kth claim. The mathematical properties of homogeneous Poisson process make the analysis of this model tractable. Consequently, many theoretical results under this model have been obtained using probabilistic arguments. An important advancement in this filed is made by Gerber and Shiu (1998), where the analysis of various quantities in interest, such as the probability of ruin, the time of ruin, the deficit at ruin, etc., are summarized systematically by the *discounted penalty function*, as known as the *Gerber-Shiu function*,

$$m(u) = \mathbb{E}\big[e^{-\delta T} w(U(T-), |U(T)|)\mathbb{1}(T < \infty)|U(0) = u\big],$$

where $T = \inf\{t : U(t) < 0\}$ is the time of ruin, the penalty function w is a bi-variate function of the surplus just before ruin and the deficit at ruin. By setting different values of w and δ, the Gerber-Shiu function reduces to different quantities in interest. See, for example, Asmussen and Albrecher (2010) Chapter XII for a summary.

The classical model (3.1) is mathematically convenient. However, the assumptions of this model are rather restrictive and unrealistic. Over the years, much research was dedicated to relaxing some assumptions of the classic model. An early attempt is Andersen (1957) where the author replaced the homogeneous Poisson process by a renewal process. Although a renewal process is a generalization of homogeneous Poisson process, it is still incapable of capturing some of the patterns, such as time-inhomogeneity, that one would expect from an insurance portfolio as some researchers have pointed out. One extension in this direction is to consider a periodic environment. Asmussen and Rolski (1994) considered a risk model where the surplus process is periodic, and Morales (2004) demonstrated how to obtain the probability of ruin of this periodic model by simulation. In recent years, the use of Cox process has attracted research interest. Different models with different Cox processes have been proposed. Noted choices include shot-noise Cox process that features random jumps in its intensity function, see, for example, Dassios and Jang (2003), Albrecher and Asmussen (2006), Dassios et al. (2015), Avanzi et al. (2021b), among many others. Another choice is Markov-modulated Poisson process, see, for example, Asmussen (1989), Lu and Li (2005), Avanzi et al. (2021a). Non-stationary Cox process has also been considered in Albrecher et al. (2021). These are all improvements on the claim modelling.

The premium income, on the other hand, is often assumed to have the same form as the classical model (3.1). Existing modification includes using a periodic (but deterministic) function, such as Asmussen and Rolski (1994). Replacing the deterministic premium income by a stochastic process is another way to modify the classical model. This model is referred to as the *stochastic premium model* in this chapter. Some theoretical results are known for this model. Boikov (2003) derived an integral equation satisfied by the ultimate survival probability and an integro-differential equation satisfied by the finite-time survival probability under this model. Closed-form expressions of the probability of ruin for two special cases are also derived. Temnov (2004) further derived limiting behaviour of the model and presented more examples. Labbé and Sendova (2009) studied the expected discounted penalty function under this model where both a renewal equation and an integral equation are established. Albrecher et al. (2010) provided an alternative approach to establish certain results concerning the expected discounted penalty function by first deriving closed-form expressions for exponential claim sizes and subsequently reformulating this results to apply to general cases. Vidmar (2018) considered stochastic dependence between premium and the claim arrivals.

In this chapter, we present a simulation-based method to obtain the probability of ruin under the stochastic premium model. As discussed in Asmussen and Binswanger (1997), this is a rare event simulation problem with infinite time horizon. One needs to apply certain simulation technique to conduct the simulation. See Asmussen and Albrecher (2010) Chapter XV for a review of available simulation techniques. In this chapter, we consider the drop of surplus upon each claim arrival. Taking advantage of the fact that this random variable may be expressed as a compound-geometric distribution, we derive a simulator based on importance sampling. With certain simple distributions, one may identify the exact distribution of the sampling random variable using the importance sampling. We give two examples of such cases to demonstrate how this simulation works. For more general cases with other premium size distributions, it may be challenging to identify the distribution of the sampling random variable from its moment generating function. Various approximation techniques are available for these cases. In this chapter, we use the fast Fourier transform to obtain an approximation. Examples concerning this scenario are also given. The simulated results are compared with known results. We also comment on the application of this model in practice.

The chapiter is structured as follows. Some useful mathematical tools and their properties that are used in this chapter are reviewed in Section 3.2. In Section 3.3, we formally introduce the risk model with stochastic premium income, and briefly summarize known results in the existing literature. A simulator based on importance sampling is derived in Section 3.4. An approximation method based on fast Fourier transform is also presented in this section. Examples are given to demonstrate the simulation approach. Conclusions are drawn in Section 3.5.

3.2 Preliminaries

In this section, we present some useful mathematical results that will be used later in the paper.

3.2.1 Moment generating function and cumulant generating function

One convenient way to represent a random variable is through its generating functions. Let X be a random variable. Its *moment generating function* M_X and its *cumulant generating function* Γ_X are defined by

$$M_X(t) = \mathbb{E}\big[e^{tX}\big], \tag{3.2}$$
$$\Gamma_X(t) = \log M_X(t).$$

These two functions are not defined for all distributions. For example, they are not defined for Cauchy distribution. Notice that it is possible that the expectation (3.2) diverts to infinity for some $t \in \mathbb{R}$, i.e., the generating functions are only defined on a subset of $\mathbb{R}$. We have the following definition.

Definition 3.1. *The domain of a function f, denoted by $\mathcal{D}(f)$, is the subset of $\mathbb{R}$ where the value of the function is finite, i.e.,*

$$\mathcal{D}(f) = \{x \in \mathbb{R} : |f(x)| < \infty\}.$$

Moment generating functions and cumulant generating functions are convenient because they provide a systematic way to calculate the moments of a random variable. If there exists $\epsilon > 0$ such that $(-\epsilon, \epsilon) \subseteq \mathcal{D}(\Gamma_X)$, then we have the following convenient results

$$\Gamma_X(0) = \log \mathbb{E}[\exp(0 \cdot X)] = 0,$$
$$\Gamma'_X(\theta) = \frac{M'_X(\theta)}{M_X(\theta)} = \frac{\mathbb{E}[X \exp(\theta X)]}{\exp(\Gamma_X(\theta))} = \mathbb{E}[X \exp(\theta X - \Gamma_X(\theta))] \quad \text{for all } \theta \in \mathcal{D}(\Gamma_X), \tag{3.3}$$
$$\Gamma'_X(0) = \mathbb{E}[X].$$

3.2.2 Importance sampling and Esscher transform

Importance sampling is an important simulation technique. It is often used to achieve variance reduction in rare event simulations.

Definition 3.2 (Importance sampling)**.** *Suppose X is a $\mathbb{R}$-valued random variable defined on a measurable space $(\Omega, \mathcal{F})$. Let $\mathbb{P}$ and $\mathbb{Q}$ be two measures on the same measurable space. Let $h : \mathbb{R} \to \mathbb{R}$ be a measurable function. Then we have*

$$\mathbb{E}_{\mathbb{P}}[h(X)] = \int_{\Omega} h(X(\omega))\mathrm{d}\mathbb{P}(\omega) = \int_{\Omega} h(X(\omega)) \cdot \frac{\mathrm{d}\mathbb{P}(\omega)}{\mathrm{d}\mathbb{Q}(\omega)} \cdot \mathrm{d}\mathbb{Q}(\omega) = \mathbb{E}_{\mathbb{Q}}[h(X) \cdot \mathcal{L}] \tag{3.4}$$

where $\mathcal{L}(\omega) = \mathrm{d}\mathbb{P}(\omega)/\mathrm{d}\mathbb{Q}(\omega)$ is the Radon-Nykodym derivative and is called the likelihood ratio.

If the random variable in Definition 3.2 has probability density function f under measure $\mathbb{P}$ and has probability density function g under measure $\mathbb{Q}$, then the likelihood ratio is f/g, i.e., equation (3.4) becomes

$$\mathbb{E}_{\mathbb{P}}[h(X)] = \int h(x) \cdot f(x)\mathrm{d}x = \int h(x) \cdot \frac{f(x)}{g(x)} \cdot g(x)\mathrm{d}x = \mathbb{E}\left[h(\tilde{X}) \cdot \frac{f(\tilde{X})}{g(\tilde{X})}\right] \tag{3.5}$$

where $\tilde{X}$ has density function g.

A major benefit of importance sampling is variance reduction. Suppose we are interested in estimating $h(X)$ by simulation. We may simulate X under probability measure $\mathbb{P}$, calculate $h(X)$, repeat this procedure n times and calculate the average of these n different realizations. To reduce the variance of the estimator, we simply increase the number of simulations n. This approach is called the *crude Monte Carlo method* or the *naïve Monte Carlo method*. Under certain circumstances, this method is not efficient. These situations usually concern simulating some rare events with very small probabilities. A number of variance reduction techniques exist for rare event simulation, see Bucklew (2013) for more details. When using importance sampling, we simulate X under probability measure $\mathbb{Q}$. The new estimator is $h(X) \cdot \mathcal{L}$. We may choose an appropriate

$\mathbb{Q}$ such that the variance of the estimator is smaller than that of the crude estimator.

The key to importance sampling is to specify a new probability measure under which the random variable X has a new distribution. If the original distribution is *absolutely continuous*, i.e., the density function exists, then it suffices to specify the new density function under the new measure. The importance sampling for this case is given by equation (3.5). We next introduce a method to define a new probability measure.

Definition 3.3 (Esscher transform)**.** *Let X be a $\mathbb{R}$-valued random variable defined on a measure space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose X has moment generating function M_X and cumulant generating function Γ_X. The Esscher transform of the measure $\mathbb{P}$, denote by $\mathbb{P}_\theta$, is given by the likelihood ratio*

$$\frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{P}}(x) = \frac{\exp(\theta x)}{M_X(\theta)} = \exp(\theta x - \Gamma_X(\theta)), \tag{3.6}$$

where θ is the parameter of the Esscher transform.

In this chapter, we denote a transformed random variable by tilde. It is easy to show that the moment generating function of the transformed distribution, denoted by $\tilde{M}_X$, is

$$\tilde{M}_X(t) = \int_\Omega e^{tX} \mathrm{d}\mathbb{P}_\theta = \int_\Omega e^{tX} \frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{P}} \mathrm{d}\mathbb{P} = \frac{\int_\Omega e^{(\theta+t)X} \mathrm{d}\mathbb{P}}{e^{\Gamma_X(\theta)}} = \frac{M_X(\theta+t)}{M_X(\theta)}. \tag{3.7}$$

Furthermore, if a random variable X has density function f, then the density function of the transformed distribution, denoted by $\tilde{f}$, is given by

$$\tilde{f}_X(x) = \frac{f(x)\exp(\theta x)}{M_X(\theta)}.$$

Some common distributions are invariant under Esscher transform. For example, if X is exponentially distributed with parameter λ, then its Esscher transform with parameter $\theta < \lambda$ is also an exponential distribution with parameter $\tilde{\lambda} = \lambda - \theta$. If X is geometrically distributed with parameter p, then its Esscher transform with parameter θ is also a geometric distribution with parameter $\tilde{p} = 1 - (1-p)\exp(\theta)$. For more details on the Esscher transform and its applications, see, for example, Asmussen and Glynn (2007).

Esscher transform is a convenient choice for importance sampling. If the likelihood ratio of the importance sampling is defined by equation (3.6), then we have

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_\theta}[X] &= \int_\Omega X \mathrm{d}\mathbb{P}_\theta = \int_\Omega X \cdot \frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{P}} \cdot \mathrm{d}\mathbb{P} \\ &= \int_\Omega X \cdot \exp(\theta X - \Gamma_X(\theta))\mathrm{d}\mathbb{P} \\ &= \mathbb{E}_{\mathbb{P}}[X \exp(\theta X - \Gamma_X(\theta))]. \end{aligned} \tag{3.8}$$

By equation (3.3) and equation (3.8), we have

$$\Gamma'_X(\theta) = \mathbb{E}_{\mathbb{P}_\theta}[X]. \tag{3.9}$$

Equation (3.9) establishes a relationship between the expected value under the new probability measure and the cumulant generating function under the original probability measure. This relationship is utilized to derive the estimator of the probability of ultimate ruin in Section 3.4.

Theorem 3.4 extends equation (3.8) to a random number of random variables.

Theorem 3.4. *If a random variable τ is a stopping time, i.e., the event $\{\tau < n\}$ is measurable with respect to the algebra generated by $\{X_1, X_2, \cdots, X_n\}$, then*

$$\mathbb{E}[f(X_1, X_2, \cdots, X_\tau)] = \mathbb{E}_{\mathbb{P}_\theta}\left[f(X_1, X_2, \cdots, X_\tau) \exp\left(-\theta \sum_{i=1}^{\tau} X_i + \tau\Gamma(\theta)\right)\right].$$

3.2.3 Fourier transform and fast Fourier transform

Fourier transform is a useful tool that is used extensively in science and engineering. It studies how a function may be approximated by a sum of trigonometric functions.

Definition 3.5 (Fourier transform)**.** *Let $f : \mathbb{R} \to \mathbb{C}$ be an integrable function, the Fourier transform of f, denoted by $\mathfrak{F}f$, is given by*

$$\mathfrak{F}f(\omega) = \int_{-\infty}^{\infty} f(x)e^{i\omega x}\mathrm{d}x,$$

where $\omega \in \mathbb{R}$. The inverse Fourier transform of a function g, denoted by $\mathfrak{F}^{-1}g$ is defined by

$$\mathfrak{F}^{-1}g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(\omega)e^{-i\omega x}\mathrm{d}\omega.$$

Remark 3.6. There are different definitions of Fourier transform and its inverse. Generically, Fourier transform can be defined as

$$\mathfrak{F}f(\omega) = \frac{1}{A}\int_{-\infty}^{\infty} f(x)e^{iB\omega x}\mathrm{d}x,$$

where A and B are two constants. For possible combinations of these two constants and their properties, see Osgood (2019) Chapter 2.4. We choose $A = B = 1$ because under this definition the Fourier transform of a probability function is the characteristic function of the distribution. □

The *discrete Fourier transform* is performed in many practical applications. Suppose instead of a continuous function f, we have $n < \infty$ equally-spaced samples, i.e.

$$f_0 = f(x_0), \quad f_1 = f(x_1), \quad f_2 = f(x_2), \quad \cdots, \quad f_{n-1} = f(x_{n-1}).$$

where $x_k - x_{k-1} = x_j - x_{j-1}$ for all $k, j \in \mathbb{N}^+$, $k, j \leq n-1$.

Definition 3.7 (Discrete Fourier transform)**.** *For a finite sequence $(f_0, f_1, \cdots, f_{n-1})$, the discrete Fourier transform and its inverse are defined as*

$$(\mathfrak{F}f)_k = \sum_{j=0}^{n-1} f_j e^{-i2\pi jk/n},$$

$$f_k = \frac{1}{n}\sum_{j=0}^{n-1}(\mathfrak{F}\mathfrak{f})_j e^{i2\pi jk/n}.$$

Discrete Fourier transform, which operates on a finite discrete sequence, may be viewed as a discretization of Fourier transform, which operates on a continuous integrable function. As the number of samples increases, the discrete Fourier transform converges to the Fourier transform. Discrete Fourier transform is of great practical value because it can be implemented through a highly efficient algorithm, called *fast Fourier transform*. A direct implementation of discrete Fourier transform has computational complexity of $O(n^2)$, while an implementation using fast Fourier transform has computational complexity of $O(n \log n)$. For more details, see for example Cooley and Tukey (1965), Brigham (1988), and Osgood (2019) Chapter 7.

In risk theory, Fourier transform is often used to evaluate the probability function of a compound distribution of the form $S = \sum_{i=1}^{N} X_i$, where N is a non-negative integer-valued random variable, X_i's are independent and identically distributed positive random variables. To obtain the exact density function or mass function of the compound distribution, one needs to evaluate the self-convolution of X_i, which is often cumbersome if possible at all. Some approximation methods exist. Examples of these approximations include normal approximation, gamma approximation, and Esscher approximation. For a review on this topic, see Hardy (2006). Under certain conditions, exact numerical methods also exist, notably the Panjer recursion, see Klugman et al. (2012) Chapter 9.6 and Sundt and Jewell (1981).

Using discrete Fourier transform, one may obtain an approximation of the probability mass function of the compound distribution $S = \sum_{i=1}^{N} X_i$. We assume that the compounded random variable X_i has mass at equally-spaced points $x_0, x_1, \cdots, x_{n-1}$ with probability $q_0, q_1, \cdots, q_{n-1}$. Notice that we assume that the support of the distribution is discrete. If the distribution of the compounded random variable X_i is continuous, then it is discretized first. Under this assumption, the compound random variable is discrete. For all $k = 0, 1, 2, \cdots, n-1$, denote

$$
\begin{aligned}
q_k &= \mathbb{P}\{X_1 = x_k\}, \\
s_k &= \mathbb{P}\{S = x_k\}.
\end{aligned}
$$

Let $(\mathfrak{F}q)_k$ and $(\mathfrak{F}s)_k$ be the kth expression of the discrete Fourier transform of the sequence $\{q\}$ and the sequence $\{s\}$ respectively. These two sequences are approximations of the characteristic functions of X_1 and S. To obtain an approximation of the probability mass function of S, one first obtains the sequence $\mathfrak{F}q$, then calculate

$$(\mathfrak{F}s)_k = P_N[(\mathfrak{F}q)_k],$$

where P_N is the probability generating function of N. Finally, the mass function of S is

$$s_k = \mathfrak{F}^{-1}(\mathfrak{F}s)_k.$$

For more details of this algorithm, see, for example, Albrecher et al. (2017) Chapter 6.4 and Asmussen and Glynn (2007) Chapter 2e. We also note that this algorithm yields an approximation of the actual mass probability. The error introduced by the approximation is studied in Grübel and Hermesmeier (1999), where a procedure to reduce the approximation error is also presented.

3.3 A risk model with stochastic premium income

One way to extend the classical model (3.1) is to replace the deterministic premium income by a stochastic process. The surplus process is defined by

$$U(t) = u + \sum_{k=1}^{H(t)} X_k - \sum_{k=1}^{N(t)} Y_k, \tag{3.10}$$

where $u \geq 0$ is the initial surplus, the premium counting process $H(t)$ is a homogeneous Poisson process with rate μ, the claim counting process $N(t)$ is a homogeneous Poisson process with rate λ, $X_k > 0$ is the amount of the kth premium arrival, and $Y_k > 0$ is the size of the kth claim arrival. We assume that $H(t)$, $N(t)$, X_k's, and Y_k's are mutually independent. We make the following assumptions on the surplus process.

Assumption 3.1. The relative security loading is positive, i.e., $\mu\mathbb{E}[X_1] > \lambda\mathbb{E}[Y_1]$.

Assumption 3.2. The moment generating function of X_1 and Y_1, denoted by $M_X(t)$ and $M_Y(t)$ respectively, exist. Furthermore, for all $\gamma \in \mathbb{R}$, there is a $t_{X,\gamma} \in \mathcal{D}(M_X)$ and a $t_{Y,\gamma} \in \mathcal{D}(M_Y)$ such that $M_X(t_{X,\gamma}) > \gamma$ and $M_Y(t_{Y,\gamma}) > \gamma$.

Since the classical model has been thoroughly studied, it often serves as the benchmark model for comparison. We are interested in how much *additional* risk the stochastic premium model (3.10) can capture compared to the classical model (3.1). To this end, we need to change only the premium process of these two models while holding other factors unchanged. We call the classical model (3.1) and the stochastic premium model (3.10) *comparable* if $c = \mu\mathbb{E}[X_1]$ while N and Y_i's are the same in these two models.

Some ruin theory results have been derived for the stochastic premium model using both probabilistic arguments and through Gerber-Shiu function. The next theorem is a collection of known results pertinent to the contents of this chapter.

Theorem 3.8. *For the stochastic premium model (3.10), define the time of ruin*

$$T := \inf\{t \geq 0 : U(t) < 0 | U(0) = u\}.$$

The probability of ruin is $\psi(u) = \mathbb{P}\{T < \infty | U(0) = u\}$. *Let* $\varphi(u) = 1 - \psi(u)$ *be the survival probability. Then*

1. *(Boikov (2003)) The survival probability satisfies*

$$(\lambda + \mu)\varphi(u) = \lambda \int_0^u \varphi(u - x)\mathrm{d}F(x) + \mu \int_0^\infty \varphi(u + x)\mathrm{d}G(x),$$

   where F *is the cumulative distribution function of the claim size random variable* Y_1, *and* G *is the cumulative distribution function of the premium amount random variable* X_1.

2. *(Boikov (2003)) Let* R^* *be a positive root of*

$$\mu[\mathbb{E}e^{-RX_1} - 1] + \lambda[\mathbb{E}e^{RY_1} - 1] = 0, \tag{3.11}$$

   then by a martingale argument we have

$$\psi(u) \leq e^{-R^* u}.$$

3. *(Temnov (2004)) Let $\psi_{CL}(u)$ be the probability of ruin under the classical model (3.1) with $c = \mu\mathbb{E}[X_1]$, then $\psi(u) > \psi_{CL}(u)$ for all $u \geq 0$.*

4. *(Temnov (2004)) The classical model may be expressed as a limiting case of the stochastic premium model. The ruin probability satisfies*

$$\lim_{\substack{\mu\to\infty \\ \mu\mathbb{E}[X_1]=c}} \psi(u) = \psi_{CL}(u).$$

for all $u \geq 0$. □

For the proof of these results, the reader is referred to the original papers.

3.4 A simulation approach

In this section, we derive a simulation approach to obtain the probability of ruin. We first obtain an expression of the probability of ruin using a change of measure argument. The derivation is similar to Pham (2007), Asmussen and Albrecher (2010), and Grandell (2012).

In the classical model, the surplus increases linearly between two claims and drops at the arrival of a claim. In the stochastic premium model, the surplus has an upward jump at the arrival of a premium and has a downward drop at the arrival of a claim. Figure 3.1 gives a sample trajectory of the surplus process under the stochastic premium model.

Let E_k be the time between the $k-1$st claim arrival and the kth claim arrival (let the 0th claim arrival be 0). Since $N(t)$ is a homogeneous Poisson process with rate λ, we know that E_k's have independent and identical exponential distribution with parameter λ. Let Z_k be the kth *drop* of the surplus process, i.e.

$$Z_k = U\left(\sum_{j=1}^{k-1} E_j\right) - U\left(\sum_{j=1}^{k} E_j\right) = Y_k - \sum_{i=1}^{\xi_k} X_i, \tag{3.12}$$

where ξ_k is the number of premium arrivals between the $k-1$st claim and the kth claim

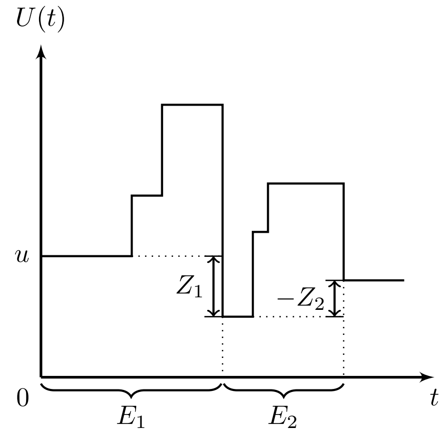


Figure 3.1: A sample trajectory of the surplus process

which may be expressed as

$$\xi_k = H\left(\sum_{j=1}^{k} E_j\right) - H\left(\sum_{j=1}^{k-1} E_j\right).$$

Since the premium arrival H follows a homogeneous Poisson process with intensity μ, we know that ξ_k follows a Poisson distribution given E_k,

$$\xi_k | E_k \sim Poisson(\mu \cdot E_k).$$

Using moment generating function, it is easy to prove the following result.

Lemma 3.9. ξ_k *is geometrically distributed with success probability* $\lambda/(\lambda+\mu)$.

Proof. The moment generating function of ξ_k, denoted by $M_\xi(t)$, may be obtained by using the law of total expectation,

$$\begin{aligned} M_\xi(t) &= \mathbb{E}\big[e^{t\xi_k}\big] = \mathbb{E}\big[\mathbb{E}\big[e^{t\xi_k}|E_k\big]\big] = \mathbb{E}\big[e^{\mu(\exp(t)-1)\cdot E_k}\big] \\ &= M_E(\mu(e^t-1)) = \frac{\lambda}{\lambda-\mu(e^t-1)} \\ &= \frac{\lambda/(\lambda+\mu)}{1-[\mu/(\lambda+\mu)]e^t}. \end{aligned} \tag{3.13}$$

Equation (3.13) is the moment generating function of a geometric distribution with success probability $\lambda/(\lambda+\mu)$. By the uniqueness of the moment generating function, we establish the lemma. □

Lemma 3.9 indicates that the total premium amount between two claim arrivals in equation (3.12) has a compound geometric distribution. This property is convenient since a number of methods exist for estimating the density function of this compound distribution.

As in the classical model, ruin can only happen when there is a claim arrival. Consequently, it suffices to consider the embedded discrete-time surplus process defined at each claim arrival, namely,

$$U_n = u - \sum_{k=1}^{n} Z_k, \tag{3.14}$$

where U_n is the surplus at the nth claim arrival, Z_k is the drop of surplus defined in equation (3.12). The discrete time of ruin, defined as the number of claim arrivals at time of ruin, is then

$$\tau := \inf\{k \in \mathbb{N}^+ : U_k < 0\} = \inf\{k \in \mathbb{N}^+ : \sum_{j=1}^{k} Z_j > u\}. \tag{3.15}$$

A Lundberg-type upper bound for the probability of ruin is given in Theorem 3.10

Theorem 3.10. *Suppose that a stochastic premium model satisfies assumption 3.1 and assumption 3.2, then the probability of ultimate ruin satisfies*

$$\psi(u) \le e^{-R^* u},$$

where R^* is the positive root of the equation

$$\Gamma_Z(R) = \Gamma_Y(R) + \log P_\xi(M_X(-R)) = 0, \tag{3.16}$$

where Γ_Z and Γ_Y are the cumulant generating function of Z_1 and Y_1 respectively, P_ξ is the probability generating function of ξ_1, and M_X is the moment generating function of X_1.

Proof. We first prove that equation (3.16) has a root in $\mathcal{D}(\Gamma_Z) \cap (0, \infty)$. Assumption 3.1 indicates that

$$\mathbb{E}[Z_1] = \mathbb{E}[Y_1] - \mathbb{E}[\xi_1] \cdot \mathbb{E}[X_1] = \mathbb{E}[Y_1] - \frac{1 - \lambda/(\lambda+\mu)}{\lambda/(\lambda+\mu)} \cdot \mathbb{E}[X_1] = \frac{\lambda\mathbb{E}[Y_1] - \mu\mathbb{E}[X_1]}{\lambda} < 0.$$

By definition of cumulant generating function, we immediately have

$$\Gamma_Z(0) = 0,$$
$$\Gamma'_Z(0) = \mathbb{E}[Z_1] < 0,$$

therefore there exists a $\epsilon_1 > 0$, such that $\Gamma(\epsilon_1) < 0$. By assumption (3.2), $\Gamma_Z(R)$ goes to infinity when R approaches the endpoint of its domain. Consequently, there exists a $\epsilon_2 > 0$, such that $\Gamma_Z(\epsilon_2) > 0$. By the continuity of Γ_Z, we may conclude that there exists a positive root of equation (3.16) in the interval $(\epsilon_1, \epsilon_2) \subseteq \mathcal{D}(\Gamma_Z) \cap (0, \infty)$. Figure 3.2 gives an illustration of equation (3.16). By the convexity of the cumulant generating function, it is easy to show that

$$\Gamma'_Z(R^*) > 0.$$

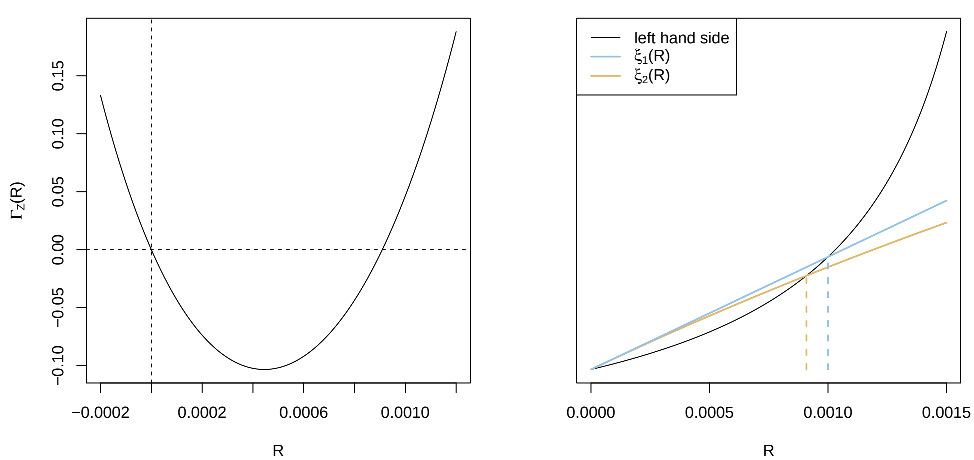


Figure 3.2: left: an example of equation (3.16); right: a comparison of adjustment coefficients under the classical model and under the stochastic premium model

Consider a new probability measure defined by Esscher transform with parameter R^*.

By equation (3.9), we have

$$\mathbb{E}_{R^*}[Z] = \Gamma'_Z(R^*) > 0, \tag{3.17}$$

i.e., the random walk $\sum_{j=1}^{k} Z_j$ in equation (3.15) has a positive drift after the Esscher transform, and consequently for any positive u, the event $\{\tau < \infty\}$ is certain. By Theorem 3.4, the probability of ultimate ruin under the original probability measure $\mathbb{P}$ may be expressed as

$$\begin{aligned}\psi(u) &= \mathbb{P}(\tau < \infty) = \mathbb{E}[\mathbb{1}_{\tau<\infty}] \\ &= \mathbb{E}_{R^*}\left[\mathbb{1}_{\tau<\infty} \cdot \exp\left(-R^* \sum_{j=1}^{\tau} Z_j + \tau\Gamma_Z(R^*)\right)\right] = \mathbb{E}_{R^*}\left[\exp\left(-R^* \sum_{j=1}^{\tau} Z_j\right)\right].\end{aligned} \tag{3.18}$$

By the definition of time of ruin (3.15), we know that

$$\sum_{j=1}^{\tau} Z_j > u \geq 0,$$

and substituting the summation in equation (3.18) by u, we have

$$\psi(u) \leq \mathbb{E}_{R^*}[\exp(-R^* \cdot u)] = e^{-R^* \cdot u}.$$

□

Remark 3.11. Equation (3.16) is equivalent to equation (3.11). Using Lemma 3.9, equation (3.16) may be rewritten as

$$\log M_Y(R) + \log\left(\frac{\lambda/(\lambda+\mu)}{1 - [\mu/(\lambda+\mu)]M_X(-R)}\right) = 0.$$

Exponentiating the equation yields

$$M_Y(R) = \frac{1 - [\mu/(\lambda+\mu)]M_X(-R)}{\lambda/(\lambda+\mu)} = \frac{\lambda + \mu - \mu M_X(-R)}{\lambda},$$

which yields equation (3.11) after rearranging. □

Corollary 3.12. *The upper bound in Theorem 3.10 is greater than or equal to the Lund-*

berg upper bound for the comparable classical model.

Proof. It suffices to show that the adjustment coefficient of the stochastic premium model is smaller than or equal to the adjustment coefficient of the comparable classical model. The Lundberg equations, expressed in terms of moment generating functions, for the classical model (3.1) and the stochastic premium model are given by

$$\text{Classical} \qquad \lambda[M_Y(R) - 1] = cR, \tag{3.19}$$

$$\text{Stochastic premium} \qquad \lambda[M_Y(R) - 1] = -\mu[M_X(-R) - 1]. \tag{3.20}$$

Denote the right hand side of these equations by $\zeta_1(R) = cR$, $\zeta_2(R) = -\mu[M_X(-R) - 1]$. At $R = 0$, we have

$$\begin{cases} \zeta_1(0) = 0 = \zeta_2(0), \\ \zeta_1'(0) = c = \mu\mathbb{E}[X] = \zeta_2'(0), \\ \zeta_1''(0) = 0 > \zeta_2''(0) = -\mu\mathbb{E}[X^2], \end{cases}$$

i.e., ζ_1 and ζ_2 have the same value and the same slope at 0, but ζ_1 is a straight line while ζ_2 is a concave function (notice the negative sign in M_X). Consequently, $\zeta_1(R) > \zeta_2(R)$ for all $R \in \mathcal{D}(\Gamma_Z) \cap (0, \infty)$. It follows that the positive root for equation (3.19) is greater than the positive root for equation (3.20). We finish the proof. Figure 3.2 illustrates equation (3.19) and (3.20). □

We showed in the proof of Theorem 3.10 that ruin is certain under the new probability measure defined by the Esscher transform with parameter R^*. Theoretically, this enables us to simulate the probability of ruin. To this end, we need to simulate the Esscher transformed random variable Z_k. Since Z_k is a linear combination of two random variables, Lemma 3.13 holds.

Lemma 3.13. *Let $X_1, X_2, \cdots, X_n$ be independent random variables. Let $a_1, a_2, \cdots, a_n$ be a sequence of real numbers. Let $X = a_1X_1 + a_2X_2 + \cdots a_nX_n$. Assume the moment* *generating function of X exists. The Esscher transform of X with parameter $\theta \in \mathcal{D}(M_X)$ has the same distribution as $a_1\tilde{X}_1 + a_2\tilde{X}_2 + \cdots + a_n\tilde{X}_n$, where $\tilde{X}_k$'s are independent random variables, and $\tilde{X}_k$ is the Esscher transform of X_k with parameter $a_k\theta$.*

Proof. By the independence of a_nX_n's, we have

$$\begin{aligned} M_X(t) &= \mathbb{E}\left[\exp\left(t \cdot \sum_{k=1}^{n} a_k X_k\right)\right] = \mathbb{E}\left[\prod_{k=1}^{n} \exp(a_k t X_k)\right] \\ &= \prod_{k=1}^{n} \mathbb{E}[\exp(a_k t X_k)] = \prod_{k=1}^{n} M_k(a_k t), \end{aligned}$$

where M_k is the moment generating function of X_k. By equation (3.7), the moment generating function of the transformed distribution is

$$\tilde{M}_X(t) = \frac{M_X(\theta + t)}{M_X(\theta)} = \prod_{k=1}^{n} \frac{M_k(a_k\theta + a_k t)}{M_k(a_k\theta)}. \tag{3.21}$$

On the other side of the equation, the moment generating function of an individual random variable X_k after the Esscher transform with parameter $a_k\theta$ is

$$\tilde{M}_k(t) = \frac{M_k(a_k\theta + t)}{M_k(a_k\theta)}.$$

Consequently, the moment generating function of $a_k\tilde{X}_k$ is

$$\mathbb{E}\left[\exp(t \cdot a_k \tilde{X}_k)\right] = \tilde{M}_k(a_k t) = \frac{M_k(a_k\theta + a_k t)}{M_k(a_k\theta)},$$

which is the kth term in equation (3.21). By the uniqueness of the moment generating function, the Esscher transform of X has the same distribution as $a_1\tilde{X}_1 + a_2\tilde{X}_2 + \cdots + a_n\tilde{X}_n$. □

Lemma 3.13 indicates that the transformed random variable Z_k may be simulated by simulating the transformed Y_k and $\sum_{i=1}^{\xi_k} X_i$ separately, and the parameter of the Esscher transform applied to Y_k is R^* as defined in equation (3.16) while the parameter for $\sum_{i=1}^{\xi_k} X_i$ is $-R^*$. This immediately allows us to conduct the simulation study for some special cases. The algorithm for these cases is

(i) Calculate the root R^* to equation (3.16).

(ii) Identify the distribution of the sampling random variables $\tilde{Z}_k = \tilde{Y}_k - \widetilde{\sum_{i=1}^{\xi_k} X_i}$ by applying Esscher transform with parameter R^* to Y_k and with parameter $-R^*$ to $\sum_{i=1}^{\xi_k} X_i$.

(iii) Simulate $\tilde{Z}_k$ until $\sum \tilde{Z}_k > u$. Return $\exp\Big(-R^* \cdot \sum \tilde{Z}_k\Big)$.

(iv) Repeat step (iii) sufficiently many times so that the relative error of the estimate of the probability of ruin is small.

(v) The probability of ruin is the average of the return values.

Example 3.14 (exponential)**.** Suppose that the premiums are exponentially distributed with mean $1/a$, and the claims are exponentially distributed with mean $1/b$. This special case has been studied in Boikov (2003) where the author derived the exact ruin probability as

$$\psi(u) = \frac{(a+b)\lambda}{(\lambda+\mu)b} e^{-R^* u}. \tag{3.22}$$

Consider the simulation approach. We simulate the transformed Z_k by simulating Y_k and $\sum_{i=1}^{\xi_k} X_i$ separately. It is well known that the Esscher transform of an exponential distribution is an exponential distribution, hence we have $\tilde{Y}_k \sim \exp(b - R^*)$. On the other hand, consider the moment generating function of the compound geometric random variable $\sum_{i=1}^{\xi_k} X_i$,

$$M(t) = \mathbb{E}\left[\exp\left(t \sum_{i=1}^{\xi_k} X_i\right)\right] = P_\xi(M_X(t)) = \frac{p}{1-(1-p)\cdot \frac{a}{a-t}} = \frac{a-t}{a-t/p}, \tag{3.23}$$

where P_ξ is the probability generating function of ξ_1, M_X is the moment generating function of X_1, and $p = \lambda/(\lambda+\mu)$ is the success probability of ξ_1. The moment generating function after the Esscher transform with parameter $-R^*$, denoted by $\tilde{M}$, is then

$$\tilde{M}(t) = \frac{M(-R^*+t)}{M(-R^*)} = \frac{(a+R^*-t)(ap+R^*)}{(ap+R^*-t)(a+R^*)} = \frac{(a+R^*)-t}{(a+R^*)\frac{(ap+R^*-t)}{ap+R^*}}$$

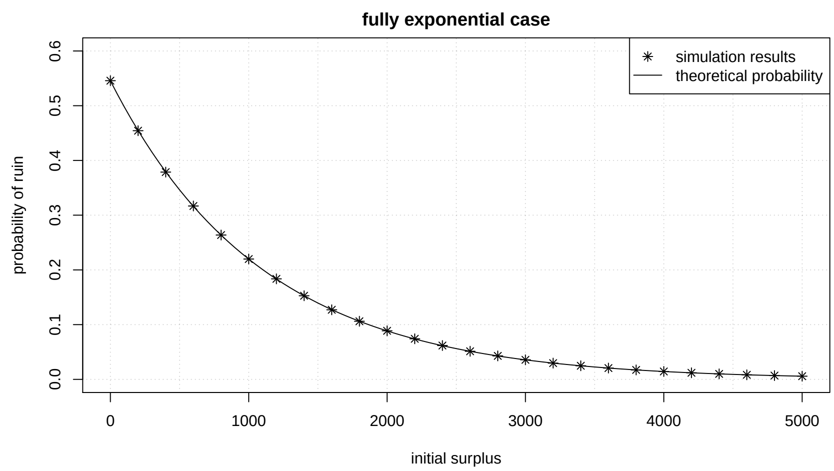


Figure 3.3: the simulated probability of ruin vs. the theoretical probability of ruin for the exponential example

$$= \frac{(a+R^*)-t}{(a+R^*)-\frac{t}{(ap+R^*)/(a+R^*)}}. \tag{3.24}$$

Observing that equation (3.24) has the same form as equation (3.23), we conclude that the Esscher transformed compound geometric-exponential distribution is still a compound geometric-exponential distribution, with

$$\tilde{\xi}_k \sim \text{geometric}\left(\tilde{p} = \frac{ap+R^*}{a+R^*}\right),$$
$$\tilde{X}_i \sim \exp(\tilde{a} = a + R^*).$$

Now all the components of the transformed surplus process may be simulated. Suppose $a = 0.01$, $b = 0.002$, $\mu = 10000$, $\lambda = 1000$. It may be calculated that the adjustment coefficient $R^* = (b\mu - a\lambda)/(\lambda + \mu)$. The theoretical probability of ruin by equation (3.22) is then

$$\psi(u) = \frac{6}{11}\exp\left(-\frac{1}{1100}u\right).$$

We conduct the simulation study by using 100,000 sample paths. The simulation results are plotted against the theoretical result in Figure 3.3. □

In the next example, we consider the case where the premium is a constant a. In some jurisdictions the premiums for obligatory liability insurance are fixed or tiered, which may be modelled by the next example. We have the following lemma about this model.

Lemma 3.15. *If the premium is a constant a, and the claim size distribution is exponential with mean $1/b$, then the probability of ultimate ruin is*

$$\psi(u) = \frac{b - R^*}{b} \exp(-R^* u) \tag{3.25}$$

where the adjustment coefficient R^* is the root of the equation

$$\frac{\lambda b}{b - R} = \lambda + \mu - \mu \exp(-aR). \tag{3.26}$$

Proof. We first derive an integral equation satisfied by the survival probability $\varphi(u) = 1 - \psi(u)$. Considering the events in a short time interval $(0, \Delta t)$ similarly to Boikov (2003), we have

$$\begin{aligned} &\varphi(u) = (1 - \lambda\Delta t)(1 - \mu\Delta t) \cdot \varphi(u) + \\ &\lambda\Delta t(1 - \mu\Delta t) \int_0^u b \exp(-bx) \cdot \varphi(u - x)\mathrm{d}x + (1 - \lambda\Delta t)\mu\Delta t \cdot \varphi(u + a) + o(\Delta t). \end{aligned} \tag{3.27}$$

Rearranging equation (3.27) and expressing terms with $(\Delta t)^2$ by $o(\Delta t)$ yields

$$(\lambda + \mu)\Delta t \varphi(u) = \lambda\Delta t \int_0^u b \exp(-bx)\varphi(u - x)\mathrm{d}x + \mu\Delta t\varphi(u + a) + o(\Delta t).$$

Dividing by Δt and letting $\Delta t \to 0$ yields

$$(\lambda + \mu)\varphi(u) = \mu\varphi(u + a) + \lambda \int_0^u b \exp(-bx)\varphi(u - x)\mathrm{d}x. \tag{3.28}$$

One may verify that

$$\varphi(u) = 1 - \frac{b - R^*}{b} \exp(-R^* u) \tag{3.29}$$

is the solution to equation (3.28) where the adjustment coefficient R^* is determined by equation (3.26). □

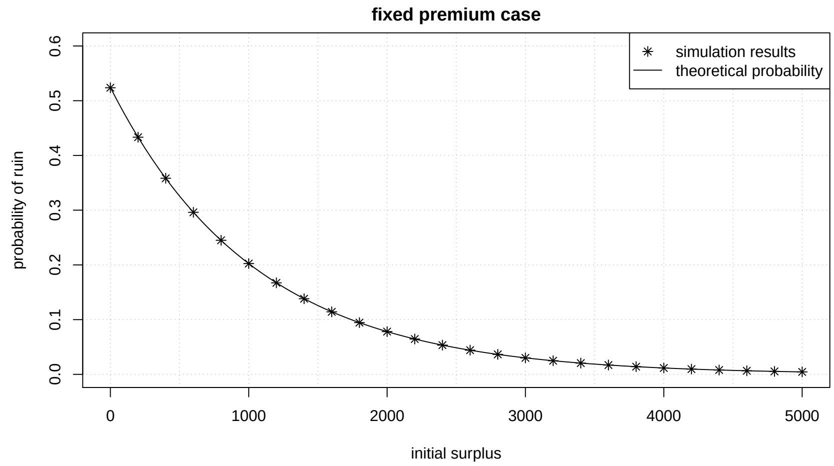


Figure 3.4: the simulated ruin probability vs. the theoretical ruin probability for the fixed premium example

Example 3.16 (fixed premium)**.** The simulation for this case is straightforward. The random variable $Z_k = Y_k - a\xi_k$ is a linear combination of random variables. By Lemma 3.13, the transformed Z_k may be simulated by $\tilde{Z}_k = \tilde{Y}_k - a\tilde{\xi}_k$ where $\tilde{Y}_k \sim \exp(b - R^*)$ and $\tilde{\xi}_k \sim \text{geometric}(\tilde{p})$ where $\tilde{p} = 1 - (1-p)\exp(-aR^*)$.

Suppose $a = 100, b = 0.002, \mu = 10000, \lambda = 1000$, it can be calculated that $R^* \approx 9.516623 \times 10^{-4}$. A comparison of the simulated ruin probability and the theoretical ruin probability is given in Figure 3.4. □

The simulations in Example 3.14 and 3.16 may be done since the exact distribution of the Esscher transformed random variables $\tilde{Z}_k$ may be identified. Specifically, the transformation of the compound random variable $\sum_{i=1}^{\xi_k} X_i$ is known to be another compound random variable. This allows us to simulate it by simulating $\tilde{\xi}_k$ and $\tilde{X}_i$'s sequentially. Notice that in this case, one does not need the probability density function of $\sum_{i=1}^{\xi_k} X_i$ to conduct the simulation. With arbitrary X_k, the Esscher transformed distribution may not be identified. In this case, only the moment generating function of the new distribution is known which is given by equation (3.7). Some techniques exist for simulating random variables from moment generating functions. For example, McLeish (2014) discusses the use of the saddlepoint approximation to simulate a random variable from its moment generating function. Various possible different approaches are also compared therein. However, this type of approximation often relies on the assumption that the distribution is close to a known distribution, such as normal distribution, and its accuracy varies in different settings. Taking advantage of the fact that the $\sum_{i=1}^{\xi_k} X_i$ is a compound geometric distribution, we next present an approximation method based on fast Fourier transform which allows us to conduct the simulation for any legible premium size distribution.

For the cases where the exact distribution of the Esscher transform of $\sum_{i=1}^{\xi_k} X_i$ can not be identified, the algorithm is:

(i) Choose a sufficiently large truncation point Π. Choose equally spaced $x_0 = 0, x_1, x_2, \cdots, x_n = \Pi$, calculate

$$q_i = \begin{cases} 0 & i = 0, \\ \mathbb{P}(x_{i-1} < X_1 \leq x_i), & i = 1, 2, \cdots, n. \end{cases}$$

(ii) Obtain a discretization of $\sum_{i=1}^{\xi_k} X_i$ by fast Fourier transform.

(a) obtain $\mathfrak{F}q$, the fast Fourier transform of q_i's.

(b) Calculate $(\mathfrak{F}f)_i = p/[1-(1-p)(\mathfrak{F}p)_i]$ as the approximation of the characteristic function.

(c) Inverse the fast Fourier transform to obtain $f_i = \mathfrak{F}^{-1}(\mathfrak{F}f)_i$. f_i's are the discretization of the compound distribution.

(iii) Calculate the positive root R^* to equation (3.16).

(iv) Simulate $\tilde{Z}_k$ until $\sum \tilde{Z}_k > u$, where $\widetilde{\sum_{i=1}^{\xi_k} X_i}$ is approximated by a discrete random variable G with $\mathbb{P}(G = x_i) = f_i \exp(-x_i R^*)$. Return $\exp(-R^* \cdot \sum \tilde{Z}_i)$.

(v) Repeat step (iv) sufficiently many times so that the relative error of the estimated probability of ruin is small.

(vi) The probability of ruin is the average of the return values.

The approximation approach has two main sources of error. The first source of error is discretization. One may prove that the discretized random variable converges to the original random variable in distribution as the grid of the discretization gets finer, i.e., as $x_1 - x_0$ approaches 0. The second source of error is fast Fourier transform. To apply fast Fourier transform, one needs to truncate the original distribution at the chosen truncation point Π. Due to the periodic nature of Fourier transform, the density function beyond point Π is "wrapped around" and appears in $f_0, f_1, \cdots, f_n$. This error is usually termed the *aliasing error*. For demonstration, consider a compound Poisson-logarithmic distribution. It is known that this compound distribution is a negative binomial distribution. We obtain the mass function of the compound distribution using fast Fourier transform with different truncation point Π. Since Poisson distribution is in $(a, b, 0)$ class, we may also obtain the mass function using Panjer's recursion. The results are plotted in Figure 3.5. We observe that the aliasing error decreases as the truncation point Π increases. Both fast Fourier transform and Panjer's recursion can yield highly accurate approximation. However, it is known that the computation complexity of fast Fourier transform is $O(n \log n)$ while the computation complexity of Panjer's recursion is $O(n^2)$. When there are many points to evaluate, fast Fourier transform has computational advantage over Panjer's recursion. For more detailed comparison between fast Fourier transform and Panjer's recursion and an algorithm for reducing the aliasing error, see Embrechts and Frei (2009).

Example 3.17. In this example, we consider gamma claim sizes and exponential premium sizes. For the claim process, let the intensity of $N(t)$ be $\lambda = 1000$ and let $Y_k \sim \text{gamma}(\alpha, \alpha/100)$ so that $\lambda \mathbb{E}[Y_1] = 100000$; for the premium process, let the intensity of $H(t)$ be μ and let $X_k \sim \exp(\mu/200000)$ so that $\mu \mathbb{E}[X_1] = 200000$ which implies a relative security loading of 1. We vary two key parameters of the model:

- α: this parameter controls the shape of the claim size distribution. Notice that since

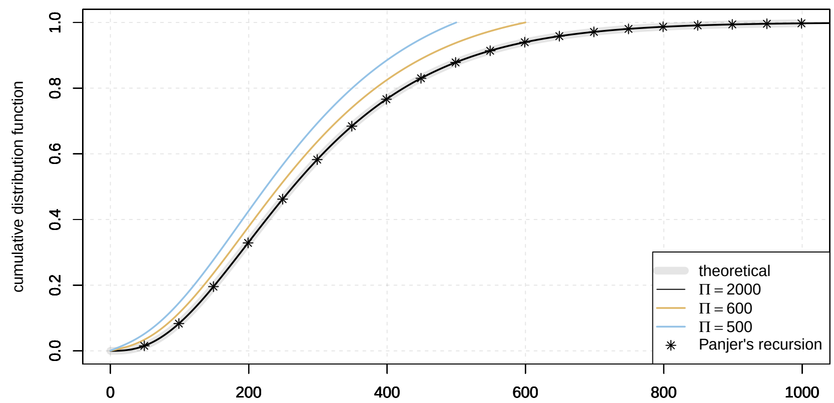


Figure 3.5: fast Fourier transform with aliasing error

the expected value of the claim size distribution is kept constant, the distribution has heavier tail with smaller α.

- μ: we let $\mu \to \infty$ to observe the limit behaviour of the model.

For each claim size distribution, we also consider the ruin probability under the classical model with the premium rate $c = 200000$. Notice that this probability is not a function with respect to μ. The simulation results with $\mu = 100, 500, 3000, 100000$ and $\alpha = 0.25, 0.75, 1, 5, 10$ are given in Figure 3.6. □

Remark 3.18. Various known behaviours of the stochastic premium model may be observed in Example 3.17. For each gamma distribution, we see that the ruin probability under the stochastic premium model is greater than the ruin probability under the classical model for all u, regardless of μ. This is result 3 in Theorem 3.8. We also observe that as μ increases, the ruin probability under the stochastic premium model converges to the ruin probability under the classical model, consistent with 4 in Theorem 3.8. □

Remark 3.19. Some cases in Example 3.17 are extreme and unlikely to happen in practice. Notice that the expected number of premium arrivals in one period of time is $\mathbb{E}[H(1)] = \mu$. For an established insurer whose business does not vary dramatically on

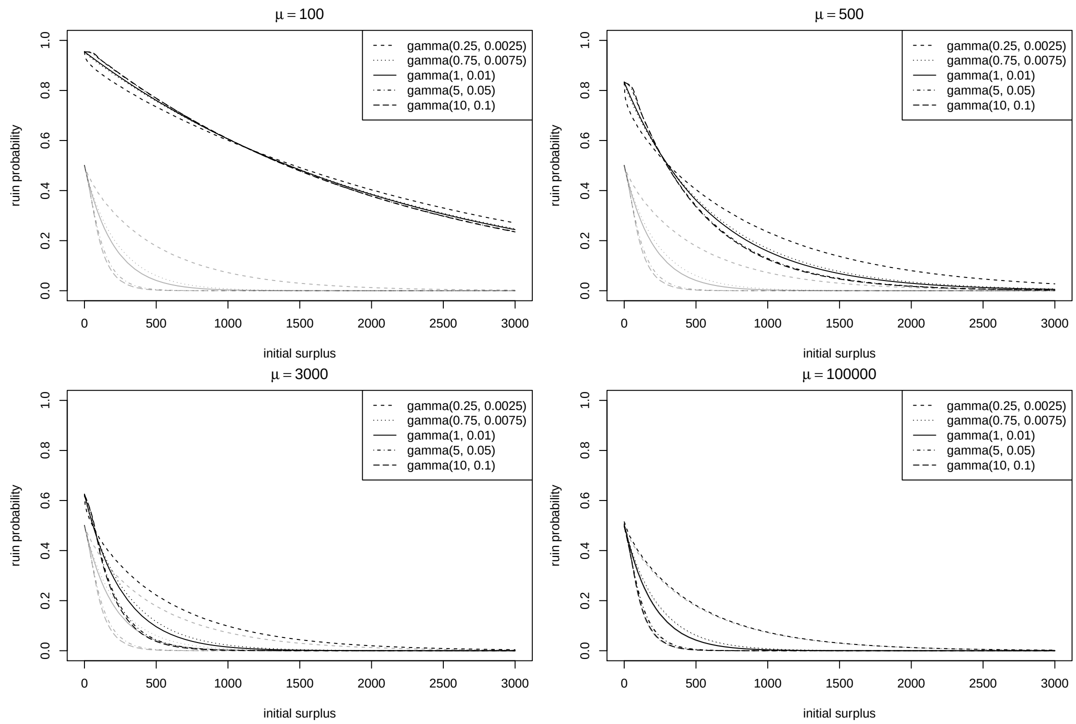


Figure 3.6: the simulated ruin probability with gamma claim sizes under the stochastic premium model (solid lines) and the classical model (dashed lines)

a year-to-year basis, μ is also approximately the portfolio size. Since all the claims are from the portfolio, the average number of claims per period of time from each policy is $\mathbb{E}[N(1)]/\mu = \lambda/\mu$, which means that the average number of claims per policy is 10 when $\mu = 100$ and is 0.01 when $\mu = 100000$. Based on some industry studies such as APCIA et al. (2020), the claim frequency per exposure varies based on physical environment and covered peril. As a result, we expect the impact of adopting the stochastic premium model to differ among different insurance portfolios. □

3.5 Conclusion

In this chapter, we derive a simulation approach to obtain the probability of ruin under the stochastic premium model. We first analyze some analytical properties of this model, specifically the drop of surplus at each claim arrival. A new probability measure is defined through Esscher transform under which ruin is certain. The simulation is done under the new probability measure, and the probability of ruin under the original probability measure is obtained. For cases where the sampling distribution is not identifiable, we take advantage of the fact that the drop of surplus may be expressed as a compound geometric distribution and use fast Fourier transform as an approximation. Some theoretical results pertinent to this model are reviewed and demonstrated through examples.

By using special cases whose theoretical probability of ruin is known, we discover that the simulation approach is fast and is capable of producing highly accurate results. The simulation approach may be readily extended to more general cases as well. We also discover that the differences between the stochastic premium model and the classical model vary based on a number of factors, such as the heaviness of tail of the distributions and the claim frequency per exposure. This means that the impact of adopting the stochastic premium model varies among different insurance portfolios and that the applicability of this model in practice is worth further investigation.

Chapter 4

On a risk model with dual seasonalities

4.1 Introduction

Modelling the cashflows of an insurance portfolio is at the core of a great amount of actuarial research. In particular, ruin theory studies the riskiness of an insurer by studying its surplus process. The customers of an insurance policy pay premiums to the insurance company in exchange for the contingent reimbursements of potential future loss. Under the *classical ruin model*, also known as the *compound-Poisson risk model*, the surplus process is modelled by

$$U(t) = u + ct - \sum_{k=1}^{N(t)} Y_k, \tag{4.1}$$

where u represents the initial surplus of the insurer, c is the continuous premium rate, i.e., the amount of premium that is collected by the insurer per unit of time. $N(t)$ is a Poisson process which counts the number of claims up to time t, and Y_i is the ith claim size, and Y_i's and $N(t)$ are mutually independent. This model dates back to the early 1900s and has been studied thoroughly afterwards. For early references, see for example

Lundberg (1903), Cramér (1955), and Gerber and Shiu (1998).

The classical model has the advantage of being mathematically tractable. The theoretical analysis is made easy due to the properties of the Poisson process. On the other hand, the assumptions of the classical model are rather restrictive and unrealistic. Over the years, many modifications have been made to the classical model to address different concerns. For example, to improve the claim counting process, Andersen (1957) proposed to use a renewal process in place of the Poisson process, Asmussen (1989) introduced a Markovian environment where the claim inter-arrival times, the claim sizes, and the premiums were influenced by an external Markovian process, Albrecher and Asmussen (2006) used a Poisson shot noise intensity process to capture the effect of sudden increases of the claim intensity.

One direction of such modification is to introduce periodic fluctuations to the surplus process. It is well known that the risk process is influenced by a number of external factors and is consequently subject to continual changes in risk propensity. The impact of such fluctuations is studied in a number of works. Berg and Haberman (1994) used a non-homogeneous Poisson process to incorporate trends in the claim process, Morales (2004) used a non-homogeneous Poisson process to address the seasonal fluctuations in the claim process, Albrecher et al. (2021) used a non-stationary Cox process to address the patterns arising in the claim process.

While the periodicity in the claim process has been introduced, the premium income is often assumed stationary. Some recent research and industrial studies point out that the premium income also has seasonal fluctuations. Several different models are able to capture this feature in the surplus process. For instances, Cramér (1955) relaxed the constant premium rate assumption by introducing an aggregate premium process to replace the deterministic premiums, Asmussen and Rolski (1994) used a deterministic function for the premium income and derived several asymptotic results.

In this chapter, we investigate how the seasonality observed in the premium income affects the probability of ruin in the presence of periodicity in the claim process. We incorporate the seasonal fluctuation in the premium income by replacing the ct component in equation (4.1) by a deterministic function. We then prove that the new model has the same probability of ruin as the model introduced in Morales (2004). Subsequently, we use the simulation algorithm presented therein to obtain the probability of ruin. After demonstrating various results by a numerical example, we examine a real-life auto insurance data set. We use the data to fit different surplus process models and obtain corresponding results. We compare the results from different models to show the potential impact of the seasonality in the claim process.

this chapter is structured as follows. We introduce the model with dual seasonalities in Section 4.2. The model specification is given in Subsection 4.2.1. The equivalent single-seasonality model is given in Subsection 4.2.2. We discuss a simulation approach in Chapter 4.3. In Subsection 4.4.1, we illustrate various results using a numerical example. An example using a real-life auto insurance data set is given in Subsection 4.4.2. Conclusions are drawn in Section 4.5.

4.2 A surplus process model with dual seasonalities

4.2.1 Model specification

We introduce the seasonal fluctuations into the surplus process. To this end, we extend the classical risk model (4.1). The new surplus process is defined by

$$U(t) = u + cP(t) - \sum_{k=1}^{N(t)} Y_k, \tag{4.2}$$

where u is the initial surplus, $P(t)$ is a periodic distortion function as described in Definition 4.1, $\{Y_k\}$ is a sequence of independent and identically distributed random variables with mean μ that represents the sizes of the claim arrivals, $N(t)$ is a *non-homogeneous Poisson process* with periodic intensity function $\lambda(t)$ which satisfies $\lambda(t) = \lambda(t+1)$ for all

$t > 0$. Let $\Lambda(t)$ denote the cumulative intensity function of $N(t)$, i.e. $\Lambda(t) = \int_0^t \lambda(s)\mathrm{d}s$. Furthermore, let $\bar{\Lambda} = \Lambda(1)$. Using the periodicity of $\lambda(t)$, it is easy to verify that

$$\Lambda(n) = n\bar{\Lambda}, \quad \text{for all } n \in \mathbb{N}. \tag{4.3}$$

Using a non-homogeneous Poisson process with periodic intensity function for the claim arrivals is a common approach to model the seasonal fluctuation, see, for example, Morales (2004) and Albrecher et al. (2021).

To model the seasonal effect in the premium income, we extend the idea of *distortion function* in the existing literature, such as Balbás et al. (2009), and define

Definition 4.1. *A periodic distortion function $P : [0, \infty) \to [0, \infty)$ is a continuous monotonically increasing function such that $P(0) = 0$, $P(1) = 1$, its derivative exists almost everywhere except on a set E with Lebesgue measure 0, and satisfies $P'(t) = P'(t+1)$ for all $t \in \mathbb{R}^+ \backslash E$.*

Table 4.1 provides three examples of periodic distortion functions and their plots.

Proposition 4.2. *For any $n \in \mathbb{N}$ and $\epsilon \in [0, 1)$, we have $P(n + \epsilon) = n + P(\epsilon)$.*

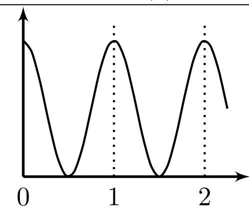
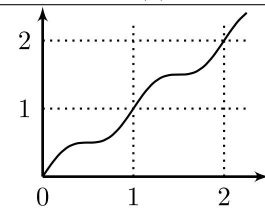
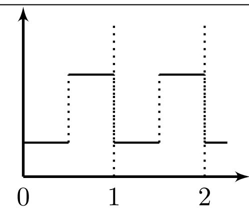
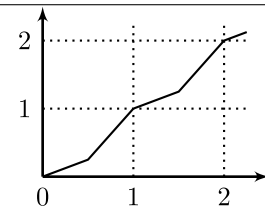
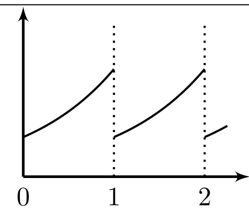
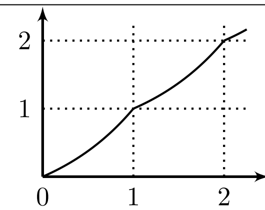
Proof. It is easy to show this result using the periodicity of the derivative of $P(t)$. We have

$$\begin{aligned} P(n+\epsilon) &= \int_0^{n+\epsilon} P'(s)\mathrm{d}s = \sum_{k=1}^{n} \int_{k-1}^{k} P'(s)\mathrm{d}s + \int_n^{n+\epsilon} P'(s)\mathrm{d}s \\ &= \sum_{k=1}^{n} \int_0^1 P'(w)\mathrm{d}w + \int_0^{\epsilon} P'(w)\mathrm{d}w = n \cdot P(1) + P(\epsilon) = n + P(\epsilon). \end{aligned}$$

□

Since P is monotonically increasing, its inverse P^{-1} exists. It is easy to show that P^{-1} is also a periodic distortion function using the inverse function theorem. As a result, Proposition 4.2 also holds for P^{-1}.

Table 4.1: Examples of periodic distortion functions and their derivatives

$P'(t)$	Plot of $P'(t)$	Plot of $P(t)$
$1 + \cos(2\pi t)$		
$0.5 + I(t - \lfloor t \rfloor \geq 0.5)$		
$(e^{t-\lfloor t \rfloor})/(e-1)$		

In essence, $P(t)$ defines the rate at which premiums are collected. One advantage of this definition is that we have $cP(t+1) - cP(t) = c$, i.e., the total amount of premium per period remains unchanged after the seasonal effect is introduced. This allows us to use the definition of the premium rate c without any modification. We define the premium rate as $c = (1 + \theta)\mu\bar{\Lambda}$, where θ and μ are defined in the same way as in the classical model (4.1).

4.2.2 An equivalent surplus process

Both the premium income and the claim process have seasonal fluctuations in Model (4.2). These behaviours are likely driven by different factors. For example, Copeland et al. (2011) documented that the activities of the automobile market have seasonal fluctuations, which could in turn cause fluctuations in the demand of auto insurance. On the other hand, the road conditions in different seasons may contribute to the seasonal variation of the accident rate. One may investigate the different drivers for the seasonalities in the premium income and the claim process and choose two periodic functions that best describe an insurer's experience. We may also investigate the relationship between the seasonalities in different processes by studying the evolution of the exposure, that is the number of policies that are in force at any given time. Suppose the coverage of an ℓ-year term insurance starts at time t, then this policy expires at time $t + \ell$. Let $H(t)$ be the number of policies sold up to time t, let $E(t)$ be the exposure at time t, then we have $E(t) = H(t) - H(t - \ell)$. It is apparent that if ℓ is a multiple of the period of $H(t)$, then the exposure $E(t)$ does not have any seasonal fluctuation. In other words, the seasonalities in this case are independently determined in relationship to their particular cause. One such example is auto insurances which commonly have a one-year term in many parts of the world. For other types of insurance policies, such as travel insurance or health insurance, the two seasonalities may be dependent. In this paper, we assume the exposure does not exhibit seasonal fluctuations, which allows us to separate the seasonal component of the premium income and the seasonal component of the claim process. We next show that the probability of ruin under model (4.2) is the same as the one under a risk model with constant premium rate and seasonal claim intensity.

For model (4.2), the *transformed surplus process* is defined as

$$\check{U}(t) = U(P^{-1}(t)) = u + ct - \sum_{k=1}^{\check{N}(t)} Y_k, \tag{4.4}$$

where

$$\check{N}(t) = N(P^{-1}(t)). \tag{4.5}$$

The transformed surplus process has constant premium rate c. To examine the new counting process $\check{N}$, we use the following lemma from Çinlar (2013) (Corollary 7.8 there).

Lemma 4.3. *$\{T_1, T_2, \cdots,\}$ are the arrival times of a non-homogenous Poisson process with cumulative intensity function Λ if and only if $\{\Lambda(T_1), \Lambda(T_2), \cdots\}$ are the arrival times of a homogeneous Poisson process with rate 1.*

Proposition 4.4. *$\check{N}$ is a non-homogeneous Poisson process with cumulative intensity function $\Lambda \circ P^{-1}$.*

Proof. Let T_n be the nth arrival time of N, and let $\check{T}_n$ be the nth arrival time of $\check{N}$. By equation (4.5), $P^{-1}(\check{T}_n)$ has the same distribution as T_n, i.e., $P^{-1}(\check{T}_n) \stackrel{\mathcal{D}}{=} T_n$. By Lemma 4.3, we have

$$T_n \stackrel{\mathcal{D}}{=} \Lambda^{-1}(E_1 + \cdots + E_n),$$

where E_i's are independent and identically distributed exponential random variables with rate 1. Consequently,

$$\check{T}_n \stackrel{\mathcal{D}}{=} P(T_n) \stackrel{\mathcal{D}}{=} P(\Lambda^{-1}(E_1 + \cdots + E_n)).$$

Therefore,

$$(\Lambda \circ P^{-1})(\check{T}_n) \stackrel{\mathcal{D}}{=} E_1 + \cdots E_n.$$

Finally, by Lemma 4.3, we establish that $\check{N}$ is a non-homogeneous Poisson process with cumulative intensity function $\Lambda \circ P^{-1}$. □

By employing this change-of-time technique, we are able to simplify the premium income. Given that both Λ and P^{-1} are periodic functions with period 1, it is easy to prove that for all $n \in \mathbb{N}$ and $\epsilon \in [0, 1)$,

$$(\Lambda \circ P^{-1})(n + \epsilon) = n\bar{\Lambda} + (\Lambda \circ P^{-1})(\epsilon). \tag{4.6}$$

The expected number of claims per period of the transformed surplus process remains unchanged since $(\Lambda \circ P^{-1})(n) = n\bar{\Lambda}$.

We now show that the original model (4.2) and the transformed model (4.4) are equivalent in terms of the probability of ruin. Define the time of ruin as the time when the surplus drops below 0 for the first time. Let T be the time of ruin of the original model, and let $\check{T}$ be the time of ruin of the transformed model, i.e., $T = \inf\{t : U(t) < 0\}$, and $\check{T} = \inf\{t : \check{U}(t) < 0\}$.

We have the following result. Suppose the underlying probability space is $(\Omega, \mathcal{F}, \mathbb{P})$. By fixing $\omega \in \Omega$, we fix a specific sample path, i.e., we fix the values of claim sizes $Y_1, Y_2, \cdots$ and the claim arrival times $T_1, T_2, \cdots$ for U. Also notice that the transformation of the time is deterministic. As a result, once the sample path of U is fixed, the corresponding $\check{U}$ is fixed as well.

Lemma 4.5. *For all $\omega \in \Omega$, we have $\check{T}(\omega) = P(T(\omega))$.*

Proof. By definition, we have

$$P(T(\omega)) = P(\inf\{t : U(t;\omega) < 0\}).$$

Since P is a monotonically increasing function, we have

$$\begin{aligned} P(T(\omega)) &= \inf\{P(t) : U(t;\omega) < 0\} = \inf\{z : U(P^{-1}(z);\omega) < 0\} \\ &= \inf\{z : \check{U}(z;\omega) < 0\} = \check{T}(\omega). \end{aligned}$$

□

Proposition 4.6. *The probability of ruin remains unchanged after the change of time, i.e., $\mathbb{P}\{\, T < \infty \,\} = \mathbb{P}\{\, \check{T} < \infty \,\}$.*

Proof. By Proposition 4.2, we have

$$\check{T} = P(T) = P(\lfloor T \rfloor + T - \lfloor T \rfloor) = \lfloor T \rfloor + P(T - \lfloor T \rfloor),$$

where $\lfloor x \rfloor = \max\{y \in \mathbb{Z} : y \leq x\}$. Since $T - \lfloor T \rfloor \in [0, 1)$, we have $P(T - \lfloor T \rfloor) < P(1) = 1$. Consequently, $\check{T} < \infty$ if and only if $\lfloor T \rfloor < \infty$, which is equivalent to $T < \infty$. Therefore $\mathbb{P}\{\, T < \infty \,\} = \mathbb{P}\{\, \check{T} < \infty \,\}$. □

Proposition 4.6 implies that we may study the transformed model in place of the original model (4.2). In the next Section, we review a simulation method to study the probability of ruin.

4.3 A simulation approach

In Section 4.2, we showed that we may consider the transformed surplus process $\check{U}$ instead of U. We also showed that the transformed cumulative intensity function is $\Lambda \circ P^{-1}$. One difficulty with this approach is that the transformed cumulative intensity function is complex. In the existing literature, the analytical results usually rely on the renewal property of the counting process. For surplus processes with arbitrary counting process, we usually obtain the quantity of interest by simulation.

The probability of ruin is difficult to obtain using naïve Monte-Carlo simulation. As pointed out by Asmussen and Binswanger (1997), we often face two difficulties:

- The time horizon of the simulation is infinite, making it difficult to define the stopping condition. Since ruin is not certain, the surplus for some simulated trajectories is positive for all $t > 0$, meaning the stopping condition is never met for these paths.

- The variance of the simulator might be too large such that it is difficult to obtain a reliable result. Specifically, the relative error could be unbounded, meaning the error could not be reduced by simply increasing the number of simulations.

One way to solve these problems is to apply the *importance sampling* technique, see, for example, Rubino and Tuffin (2009), Chapter 2. The idea is to increase the claim severity and the claim frequency at the same time so that the probability of ruin becomes 1. The original probability of ruin may be obtained subsequently using the *Radon-Nikodym theorem.*

There are already a number of papers using the importance sampling technique to conduct simulation studies. The following estimator is from Morales (2004). For derivation and proofs, we refer the readers to the original paper.

Lemma 4.7 (Morales (2004))**.** *Consider the surplus process*

$$U(t) = u + ct - \sum_{k=1}^{N(t)} Y_k,$$

where $N(t)$ is a non-homogeneous Poisson process with periodic intensity function $\lambda(t)$ and cumulative intensity function $\Lambda(t)$. Let $\bar{\Lambda} = \Lambda(1)$ be the expected number of arrivals of N per unit of time. Y_k's are independent and identically distributed claim sizes with common density function $f(y)$ and mean μ. The premium rate is $c = (1+\theta)\bar{\Lambda}\mu$ where θ is the relative security loading. Suppose the moment generating function of the claim-size distribution exists and is denoted by M. The probability of ruin as a function of the initial surplus u, denoted by $\psi(u)$, may be simulated by

$$\psi(u) = \mathbb{E}\left[\exp\left(-R \cdot \left[\sum_{k=1}^{\tilde{\tau}} \tilde{Y}_k - (1+\theta)\mu\eta \sum_{k=1}^{\tilde{\tau}} \tilde{E}_k\right]\right)\right], \tag{4.7}$$

where $\eta = \inf_{0 \le t < 1}\{\bar{\Lambda}t/\Lambda(t)\}$, R (if exists) is the positive root of the equation

$$M(x) = 1 + (1+\theta)\mu\eta x. \tag{4.8}$$

$\tilde{Y}_k$ is the Esscher transform of the kth claim size with parameter R, i.e., the density function of $\tilde{Y}_k$ is given by

$$\tilde{f}(y) = \frac{e^{Ry} f(y)}{M(R)}.$$

$\tilde{E}_k$'s are independent exponential random variables with parameter $1+\delta$ where $\delta = (1+\theta)\mu\eta R$. Finally, $\tilde{\tau}$ is the number of claims at the time of ruin, i.e.,

$$\tilde{\tau} = \inf\left\{n \in \mathbb{Z} : u + c\Lambda^{-1}\left(\sum_{k=1}^{n} \tilde{E}_k\right) - \sum_{k=1}^{n} \tilde{Y}_k < 0\right\}. \tag{4.9}$$

Remark 4.8. Notice that 0 is always a root of equation (4.8), however only strictly positive root can be used in the simulation. It is easy to verify that no transformation is performed with $R = 0$, and consequently ruin is not certain with this choice.

Remark 4.9. In Morales (2004), the estimator depends on a specific family of intensity functions, namely the *bell-shaped intensities*. The advantage of this family of intensity functions is that the inverse of the cumulative intensity function has a closed-form expression, and hence, it is easy to simulate the non-homogeneous Poisson process using the inversion method in equation (4.9). However, this is not a necessary condition.

We emphasize that the following condition must hold for this approach to work.

Proposition 4.10. *Let R be the positive root of equation (4.8). Ruin is certain after the transformations defined in Lemma 4.7 if and only if $M'(R) > (1+\theta)\mu$.*

Proof. Let $\tilde{T}_n$ be the time of the nth claim arrival, which is simulated by

$$\tilde{T}_n = \Lambda^{-1}(\tilde{E}_1 + \cdots + \tilde{E}_n),$$

where $\tilde{E}_k \sim \exp(1+\delta)$ for all k. Consequently, we have

$$\Lambda(\tilde{T}_n) = \tilde{E}_1 + \cdots + \tilde{E}_n \overset{\mathcal{D}}{=} \frac{E_1 + \cdots + E_n}{1+\delta}.$$

Thus,

$$(1+\delta)\Lambda(\tilde{T}_n) \overset{\mathcal{D}}{=} E_1 + \cdots + E_n,$$

where $E_k \sim \exp(1)$ for all k. By Lemma 4.3, we conclude that $\tilde{T}_n$ is the nth arrival time of a non-homogeneous Poisson process with cumulative intensity function $(1+\delta)\cdot\Lambda(t)$. It follows that the expected number of claims per period of time is

$$\mathbb{E}\Big[\tilde{N}(1)\Big] = (1+\delta)\cdot\bar{\Lambda}.$$

On the other hand, the expected value of the transformed claim-size distribution is

$$\mathbb{E}\Big[\tilde{Y}_k\Big] = \frac{M'(R)}{M(R)}.$$

Consequently, the total expected claim per period is

$$\mathbb{E}\Big[\tilde{N}(1)\Big] \cdot \mathbb{E}\Big[\tilde{Y}_1\Big] = (1+\delta) \cdot \bar{\Lambda} \cdot \frac{M'(R)}{M(R)} = \bar{\Lambda} \cdot M'(R).$$

To guarantee that ruin is certain under the transformation, we need

$$\mathbb{E}\Big[\tilde{N}(1)\Big] \cdot \mathbb{E}\Big[\tilde{Y}_1\Big] = \bar{\Lambda} \cdot M'(R) > c = (1+\theta)\bar{\Lambda}\mu,$$

or equivalently,

$$M'(R) > (1+\theta)\mu, \tag{4.10}$$

which concludes the proof. □

Remark 4.11. Proposition 4.10 implies that the simulation approach does not work for all combinations of intensity function and claim-severity distribution. This is a difference from the classical model. For the classical model with homogeneous Poisson claim arrivals, Proposition 4.10 holds as long as $\theta > 0$. See for example Klugman et al. (2012), p. 279. In practice, this condition is not highly restrictive, as the insurers usually charge sufficient amount of premium, and the seasonal effect is usually mild to moderate. We demonstrate this by examining a real-life data set in Section 4.4.2.

4.4 Two examples

We have showed that the dual-seasonality model (4.2) may be transformed into a single-seasonality model (4.4), and that the probability of ruin of the single-seasonality model may be simulated using Lemma 4.7, given that condition (4.10) holds for the transformed model. In this section, we first give a numerical example to illustrate various options. We then proceed to apply the simulation method to a real-life auto insurance data set.

4.4.1 A numerical example

We consider three different cases for the dual-seasonality model

$$U(t) = u + cP(t) - \sum_{k=1}^{N(t)} Y_k.$$

The periodic distortion function P and the Poisson intensity function $\lambda(t)$ are specified in Table 4.2. Notice that Model 1 is the classical model, Model 2 is the single-seasonality model in the existing literature, and Model 3 is the dual-seasonality model. For all three models, the expected number of claims per period is $\bar{\Lambda} = 1000$. For all three models, the claim sizes are assumed to be exponential with mean 100. The relative security loading θ is assumed to be $0.4, 0.8, 1.2$, and 1.6, and the premium rate $c = (1+\theta)\mu\bar{\Lambda}$ is calculated accordingly. We obtain the probability of ruin via simulation with initial surplus 0, 250, 500, and 1000. The introduction of s_1 and s_2 allows us to consider different locations for peak seasons by shifting the seasonal components horizontally. Figure 4.1 presents claim arrival intensities with seasonalities that peak at different times of year.

Table 4.2: Specifications of the seasonal components of the three models

	Model 1	Model 2	Model 3
$P'(t)$	1	1	$1 + 0.1\sin(2\pi(t - s_1))$
$\lambda(t)$	1000	$1000 + 50\sin(2\pi(t - s_2))$	$1000 + 50\sin(2\pi(t - s_2))$

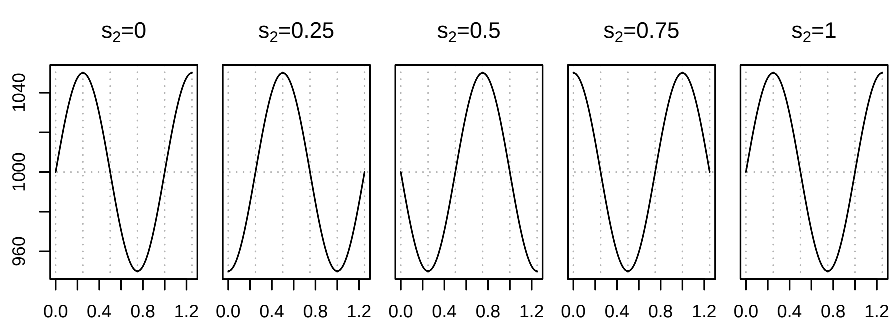


Figure 4.1: the intensity function $\lambda(t)$ with different s_2

We show the results with relative security loading 0.8 and initial surplus 0. The full simulation results may be found in Appendix A. For Model 1, since the claim sizes are exponentially distributed, it is well known that the probability of ruin is

$$\psi(u) = \frac{1}{1+\theta} \exp\left(-\frac{\theta}{(1+\theta)\mu} \cdot u \right)$$

and therefore, $\psi(0) = 0.5556$. For Model 2, the probability of ruin is simulated using Lemma 4.7. For Model 3, we first transform the surplus process according to (4.4), and then apply Lemma 4.7. The results with $u = 0$ are given in Figure 4.2.

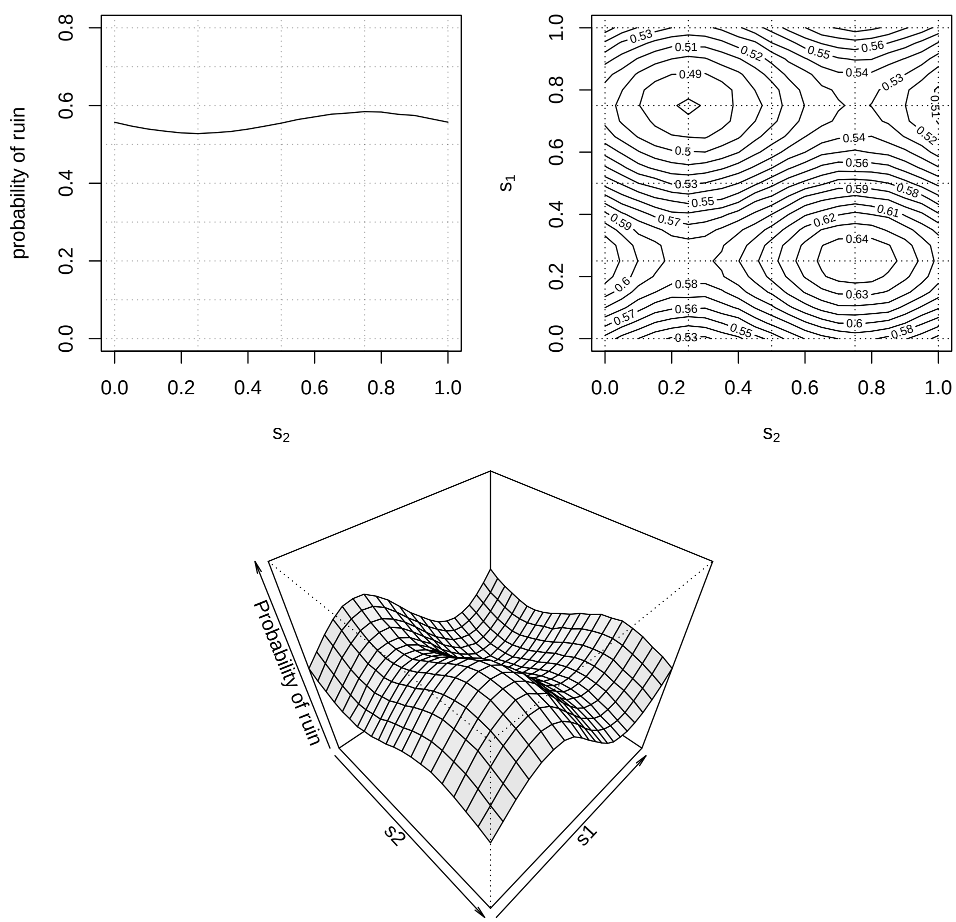


Figure 4.2: the probability of ruin for Model 2 (upper left) and Model 3 (upper right and bottom) with $u = 0$

Remark 4.12. We may verify that condition (4.10) holds for all combinations of P, λ

and relative security loading θ. □

Remark 4.13. For Model 1, the probability of ruin is 0.5556. For Model 2, the minimum probability of ruin is 0.5278 when $s_2 = 0.25$, and the maximum probability of ruin is 0.5843 when $s_2 = 0.75$. For Model 3, the minimum probability of ruin is 0.4781 when $s_1 = 0.75$ and $s_2 = 0.25$, and the maximum probability of ruin is 0.6477 when $s_1 = 0.25$ and $s_2 = 0.75$. Depending on the combination of the two seasonal components, the seasonality in the premium income may amplify or reduce the impact of the seasonality in the claim process. □

Remark 4.14. It may be observed from the simulation results that the insurer is subject to a significant risk of ruin if they do not hold sufficient initial surplus. This is true even if the insurer charges very high premium, i.e., the relative security loading is high. This result stresses the importance of holding sufficient reserve by insurance companies so that the insolvency risk is properly managed. □

We next examine another important factor that affects the probability of ruin, namely the heaviness of the tail of the claim-size distribution. We fix $s_1 = 0.25$ and $s_2 = 0.75$, i.e., the combination corresponding to the maximum probability of ruin. The claim sizes are now assumed to follow a gamma distribution with density function

$$f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, y > 0.$$

To make the results comparable, we require that $\beta = \alpha/100$ such that the expected value of the gamma distribution is $\alpha/\beta = 100$. Let $\alpha = 0.1, 0.5, 1, 2, 5$. Notice that $\alpha = 1$ is the exponential case considered in the previous simulation. We hold other parameters of the model unchanged. The probabilities of ruin under three models and their logarithms are given in Figure 4.3.

Remark 4.15. As the initial surplus increases, the probability of ruin decreases. Among all the claim-size distributions used in the simulation, those with heavier tail result in

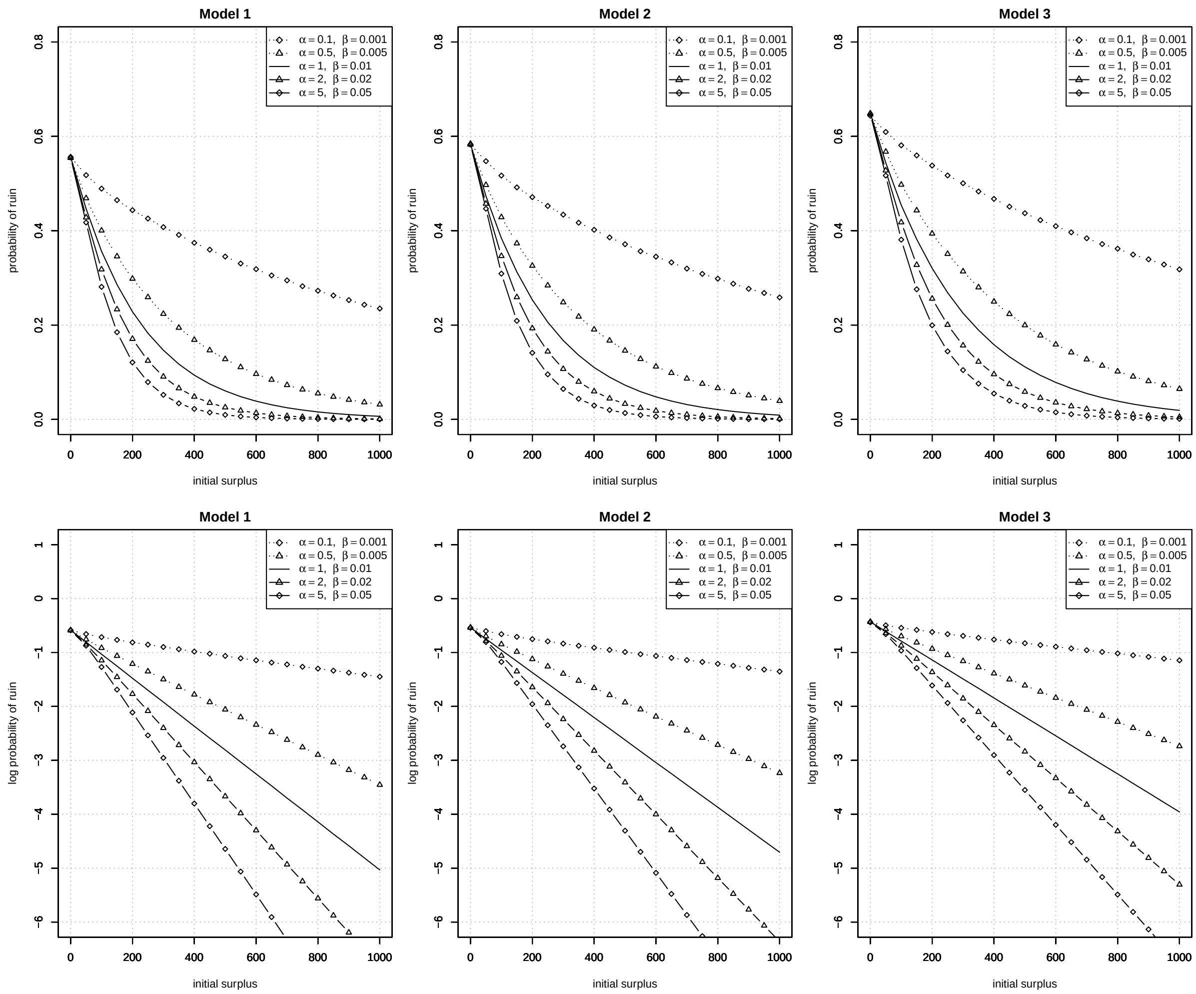


Figure 4.3: the probabilities of ruin (top) and their logarithms (bottom) with different initial surplus and claim-size distribution

lower decreasing rate of the probability of ruin, while those with lighter tail result in higher decreasing rate of the probability of ruin. Also, the probability of ruin decreases at an exponential rate with large initial surplus regardless of the claim-size distribution. This is consistent with the results presented in Asmussen and Rolski (1994). □

4.4.2 A real-life example

In this section, we apply the simulation technique to a real-life auto insurance data set that is provided by an anonymous insurer. A more detailed analysis of the data set may be found in Miao et al. (2021). The original data set contains the premium data and the claim data for the period of 2013 to 2015. The data set is extracted by the effective date of the policies, i.e., it contains only the policies that became effective after January 1, 2013. Consequently, the policies that became effective in 2012 and expired in 2013 are not included. In other words, the claim data for 2013 is not complete. For this reason we use only the data for years 2014 and 2015. Table 4.3 gives a brief summary of the data set.

Table 4.3: description of the data set

year	2014	2015
total premium	59,779,468	78,186,959
number of claim	4,847	5,976
total claim	18,370,311	24,824,995

Plots of the observed premium income and the claim arrivals are given in Figure 4.4. Notice that the 7-day moving averages are used for both premium and claim arrivals to remove the weekly pattern observed in the data set.

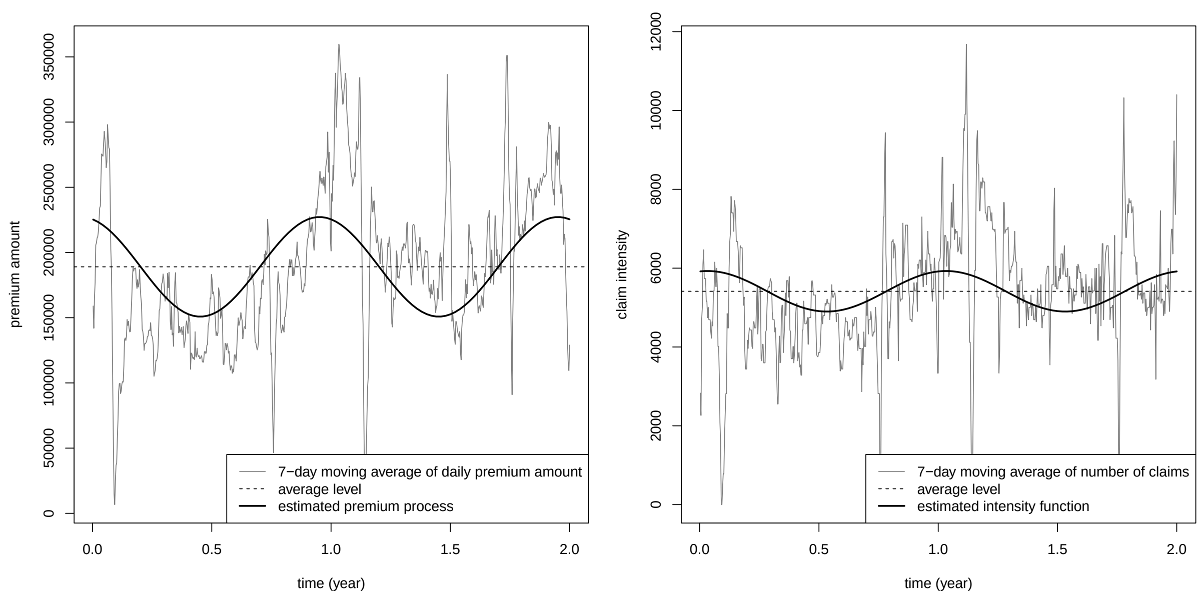


Figure 4.4: estimated premium income (left) and estimated intensity function for $N(t)$ (right)

All components of Model 4.2 may be estimated from the data set. The relative security loading may be calculated from Table 4.3. The estimated relative security loading is 2.25

for 2014 and is 2.15 for 2015. We use $\theta = 2$ in the simulation. The distortion function for the premium and the intensity function for the claim arrivals may be fitted to the data set. For the premium income component $cP(t)$, we first obtain c by calculating the average annual premium amount from the data set, we then use the method of least squares to estimate the unknown phase and the magnitude of the seasonality of the distortion function $P(t)$. Similar procedure is applied to obtain an estimate of the intensity function of the claim counting process $N(t)$. For a discussion on how to estimate the intensity function from the data, see Chernobai et al. (2007). The estimates are

$$P'(t) = 1 + 0.2\sin(2\pi(t - 0.70)), \tag{4.11}$$

$$\lambda(t) = 5414 + 515\sin(2\pi(t - 0.78)). \tag{4.12}$$

For the claim-size distribution used in the simulation, we use both the empirical distribution obtained directly from the data set, and a gamma distribution with parameters estimated by matching the first two moments. For the empirical distribution case, a procedure similar to the one introduced in Grandell (1979) is used to estimate R. We use a gamma distribution here because it is mathematically convenient due to the fact that the sampling random variables $\tilde{Y}_k$'s also follow a gamma distribution. One could fit a better parametric model for the claim size distribution to obtain a better estimate of the probability of ruin, however the two distributions we use for the simulation should suffice to illustrate how the proposed model compare with alternative models.

Three different models similar to those defined in Table 4.2 are considered. We obtain the probability of ruin with different initial surplus u. The result is given in Figure 4.5. In this case, the seasonality in the premium income reduces the probability of ruin. Among these three models, the single-seasonality model yields the highest probability of ruin, while the dual seasonal model yields the lowest probability of ruin.

We may also calculate the required amount of initial surplus such that the probability of ruin is under certain level. Table 4.4 lists the required initial surplus in thousands

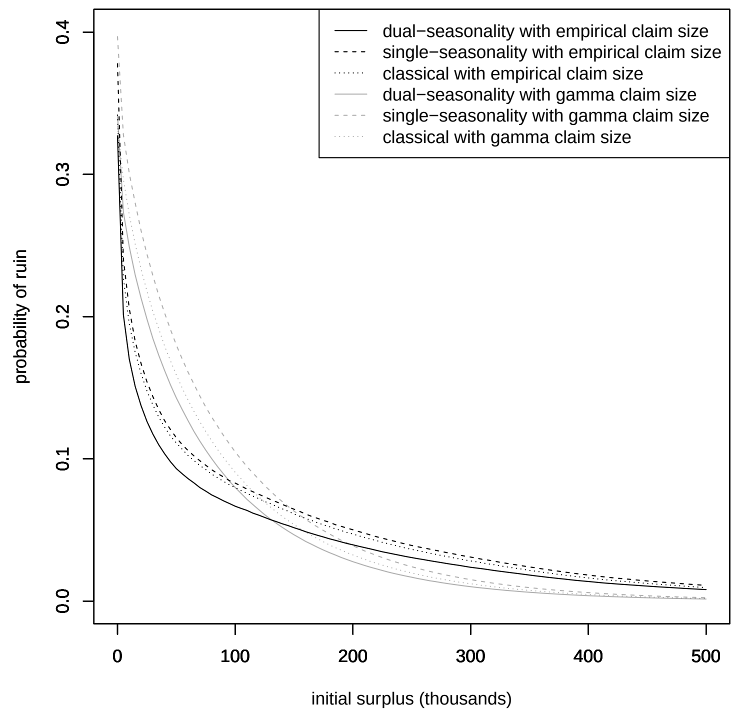


Figure 4.5: the probability of ruin for three models using different initial surplus

with which the probability of ruin is less than 0.01 under three different models with two different claim-size distributions. Among the three models, the dual-seasonality model yields the lowest required initial surplus. This is consistent with previous results where the dual-seasonality model yields the lowest probability of ruin among the three models.

Table 4.4: required initial surplus in thousands such that the probability of ruin is lower than 0.01

Model	claim-size distribution	
	empirical distribution	gamma distribution
classical	495	325
single-seasonality	525	345
dual-seasonality	460	305

Remark 4.16. Using the parameters estimated from this particular data set, the seasonality in the premium income offsets the impact of the seasonality in the claim process. Since we are assuming a positive relative security loading, the surplus process has a positive drift. Consequently, the surplus would increase to infinity almost surely. This is true

even in the cases when ruin occurs in finite time as long as we allow the surplus process to continue evolving after it hits 0. This implies that the risk of ruin is more pronounced for small t. It is clear from Figure 4.4 that both the premium income and the claim arrivals have a high season around time 0. While a higher volume of claim arrivals is especially risky when the insurer does not have sufficient time to build an ample reserve, more premium income at this period of time provides some additional safety buffer. This clarifies why the risk of ruin is higher under Model 2 than under Model 3. This reasoning also explains the patterns in Appendix A. $\square$

4.5 Conclusion

In this chapter, we extend the classical ruin model to allow for seasonal fluctuations in the premium income. We propose a method to incorporate the seasonal component in the premium income without affecting other components of the surplus process model. We show that employing a certain change-of-time technique, the dual-seasonality model is equivalent to a single-seasonality model, which is easier to analyze. We then review and apply a simulation technique utilizing importance sampling. Some ruin theory results are demonstrated using a numerical example and a real-life auto-insurance example.

We discover that, depending on the relative position of the peak season of the premium and the peak season of the claim, the seasonality in the premium income may either increase or reduce the probability of ruin. Specifically, the probability of ruin is sensitive to the initial condition, i.e., whether one has a peak season or a low season for the premium and the claim around $t = 0$. The probability of ruin decreases with the initial surplus at a rate that is related to the specific claim-size distribution. Furthermore, the way the probability of ruin decreases is not exponential for smaller u, but it is approximately exponential as $u \to \infty$.

Appendix A

Plots of additional simulation results

Table A.1: Simulation results with relative security loading $\theta = 0.4$

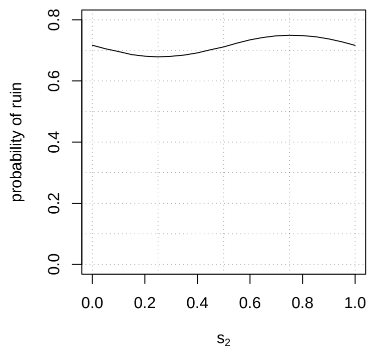
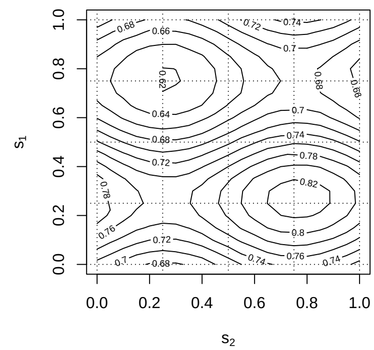
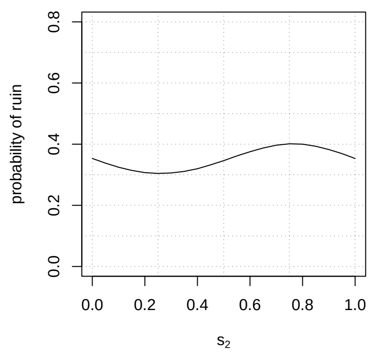
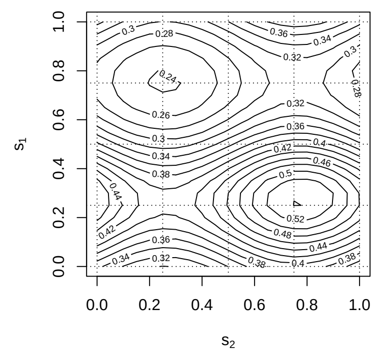
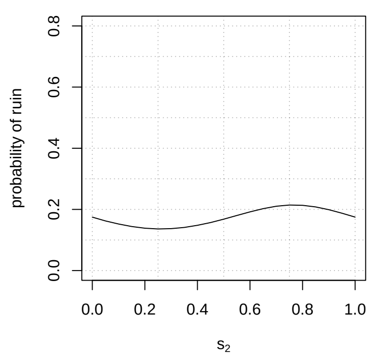
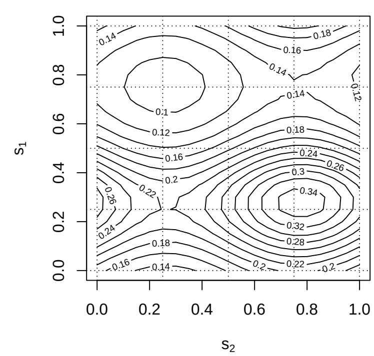
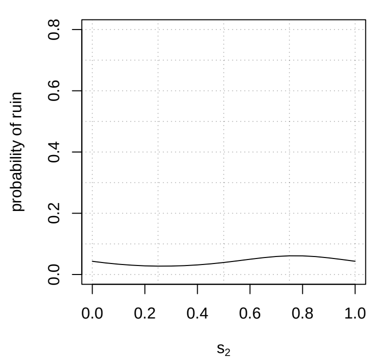
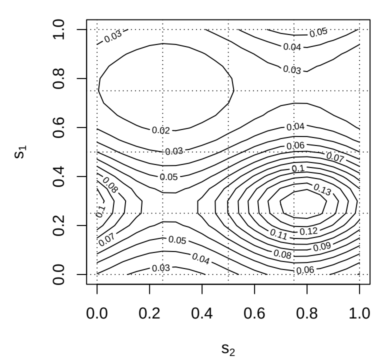
Initial surplus	Model 1	Model 2	Model 3
0	0.7143		
250	0.3497		
500	0.1712		
1000	0.0410		

Table A.2: Simulation results with relative security loading $\theta = 0.8$

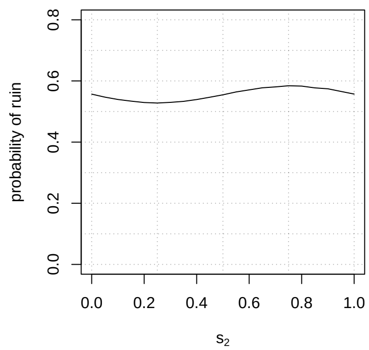
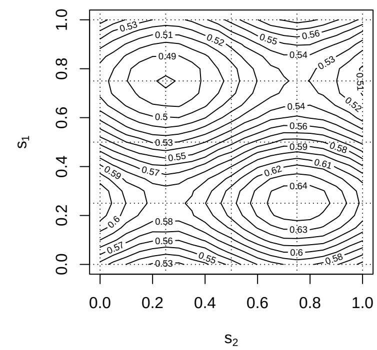
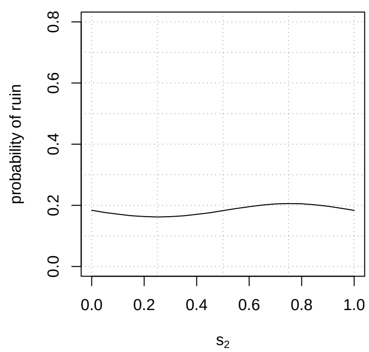
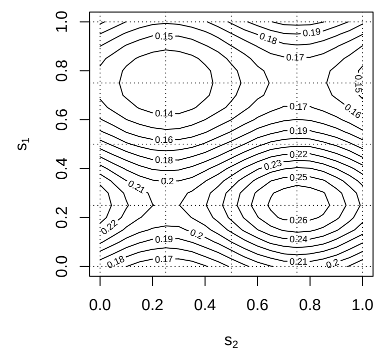
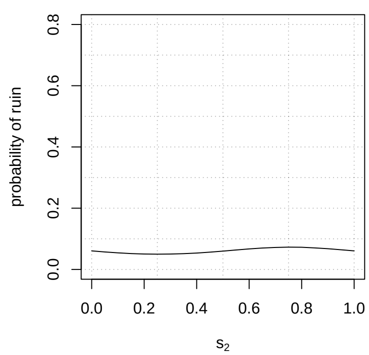
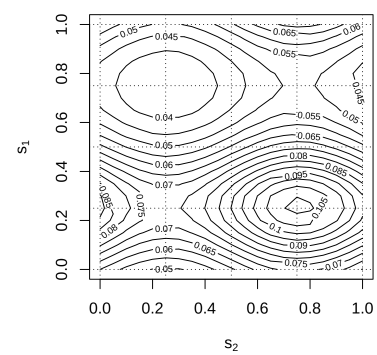
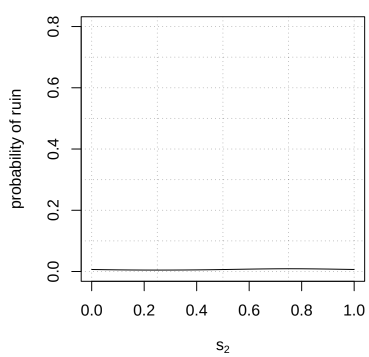
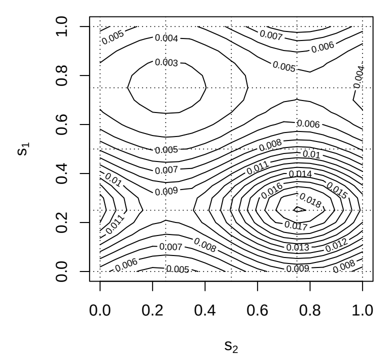
Initial surplus	Model 1	Model 2	Model 3
0	0.5556		
250	0.1829		
500	0.0602		
1000	0.0065		

Table A.3: Simulation results with relative security loading $\theta = 1.2$

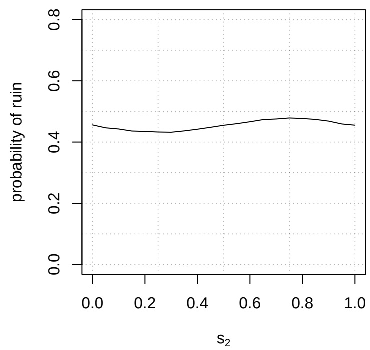
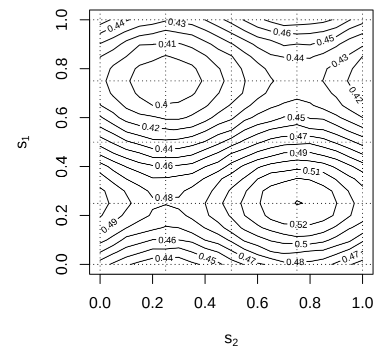
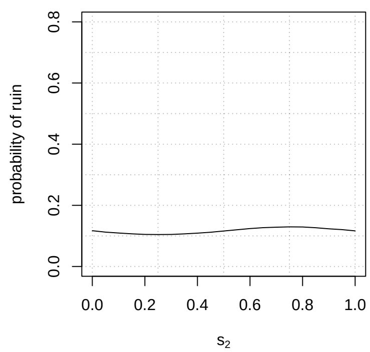
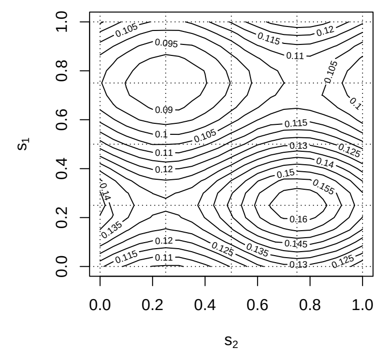
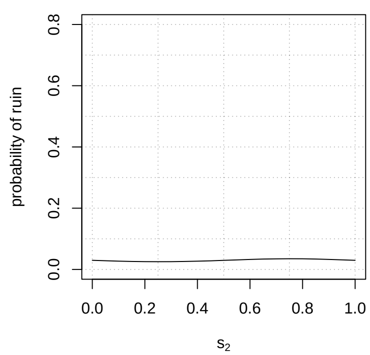
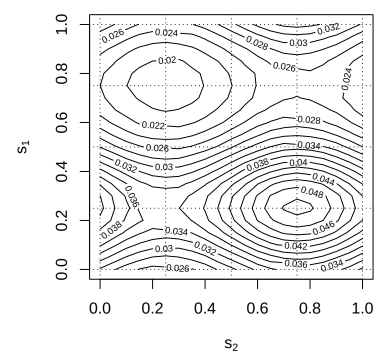
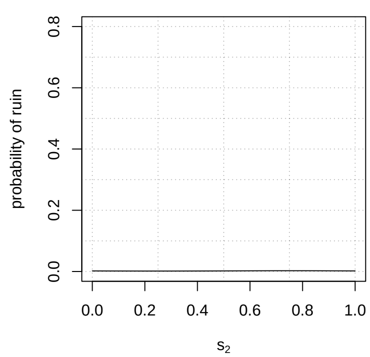
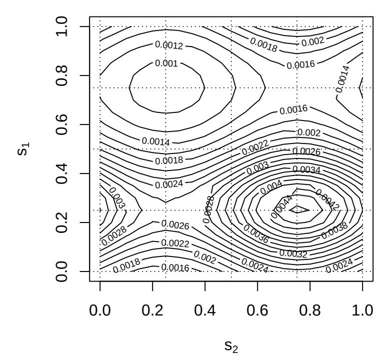
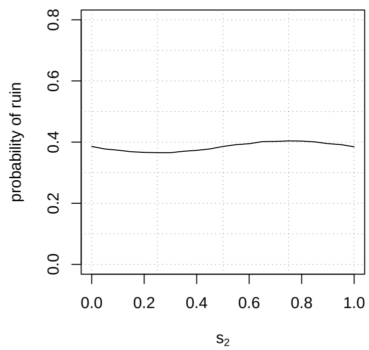
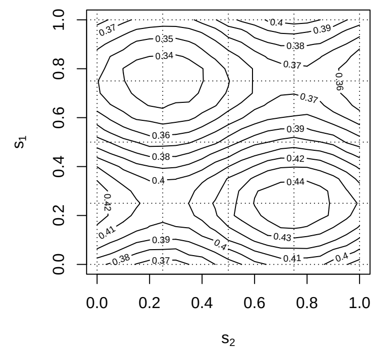
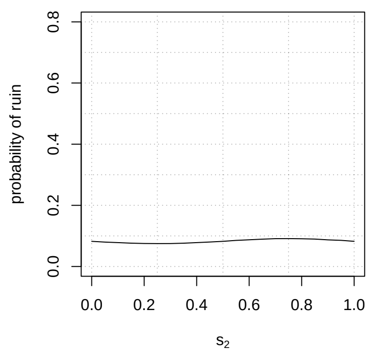
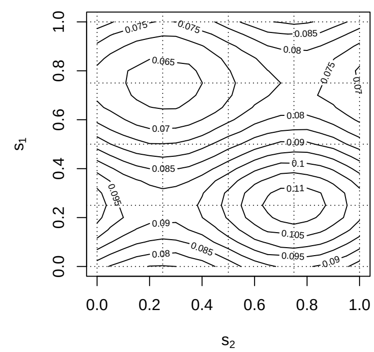
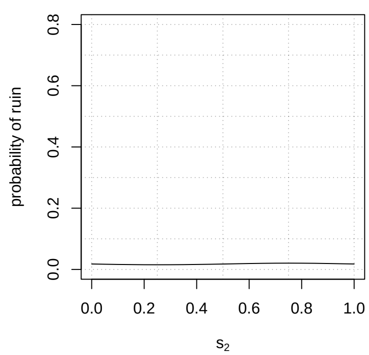
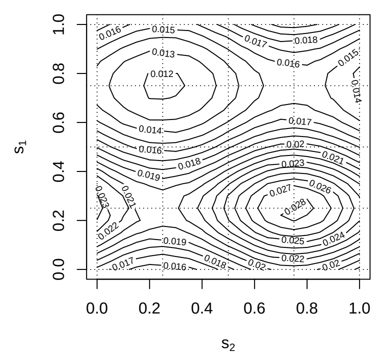
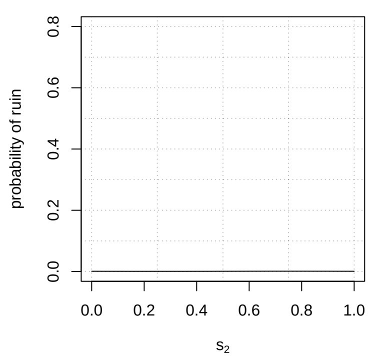
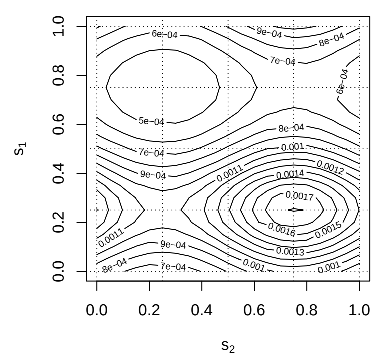
Initial surplus	Model 1	Model 2	Model 3
0	0.4545		
250	0.1162		
500	0.0297		
1000	0.0019		

Table A.4: Simulation results with relative security loading $\theta = 1.6$

Initial surplus	Model 1	Model 2	Model 3
0	0.3846		
250	0.0826		
500	0.0177		
1000	0.0008		

Chapter 5

Summary and future research

In this thesis, we have investigated a real-life insurance data set and studied some risk models inspired by the characteristics of the data set. We summarize our results as follows.

In Chapter 2, we studied the time patterns of the premium and the claim processes of an insurer. We applied and compared several different statistical techniques to separate long-term trends and short-term fluctuations. We concluded that both the premium and the claim processes exhibit evident time patterns. Specifically, both processes exhibit increasing intensities with seasonal fluctuations. We proposed to model the dependency between these two processes through the exposure. Based on these findings, we proposed a risk model that utilizes both a non-homogeneous Poisson process and a Cox process as counting processes. A preliminary simulation study on the probability of finite-time ruin was conducted to demonstrate how this new model could potentially capture more risks. As this new model offers greater flexibility than the existing risk models, it may fit real-life insurance data sets better.

In Chapter 3, we studied a risk model with stochastic premium income. This model has been studied in the existing literature and various results have been obtained. We first studied the drop of surplus at each claim arrival. Exploiting the fact that this random variable may be expressed as the difference of a known random variable and a compound geometric random variable, we derived a simulation method based on importance sampling to obtain the probability of ultimate ruin under this model. For cases where the sampling distribution is not easily identifiable from its moment generating function, we proposed to use fast Fourier transform to obtain an approximation. Numerical examples were given to demonstrate the behaviour of this model. Simulated results are compared with known theoretical results. The proposed method allows one to obtain the risk measure of interest for a wider range of distributions and provides more insight on the insolvency risk of an insurance portfolio.

In Chapter 4, we studied a risk model incorporating seasonal fluctuations in both the premium and the claim processes. We replaced the constant premium rate in the classical model by a deterministic periodic function. By a change-of-time technique similar to Asmussen (1989), we proved the equivalence of the proposed model and the model proposed in Morales (2004). We analyzed the simulation approach therein and identified the condition that must be satisfied for this approach to work. We then demonstrated how to obtain the probability of ruin under this model via several numerical examples. We also fitted the model to the real-life insurance data set and compared the probability of ruin under different models. The analysis in this Chapter demonstrates how seasonalities affect the riskiness of a portfolio.

To conclude, we studied risk models driven by real-life data in this thesis. These models may better describe the evolution of an insurer's surplus since they allow for characteristics that we observe in the real-life data. Hence, insurers could gain a better understanding of the risks they are taking. Future research topics in this direction include

- In Chapter 2, we only showed the probability of finite-time ruin. With a long time horizon, this probability should be close to the probability of ultimate ruin. One could potentially simulate the latter by employing simulation techniques.

- In Chapter 3, we showed the simulated results. We observed some distinct behaviours of this model. Only some of these behaviours can be explained by known theoretical results. One could investigate from a theoretical point of view why these behaviours occur.

- In Chapter 4, we used a periodic but deterministic function to replace the constant premium rate. One could further assume that the premium income is modelled by a compound Poisson process with periodic intensity function. Both simulation methods and theoretical methods are yet to be derived for this model.

Bibliography

H. Albrecher and S. Asmussen. Ruin probabilities and aggregrate claims distributions for shot noise Cox processes. *Scandinavian Actuarial Journal*, 2006(2):86–110, 2006.

H. Albrecher and O. Boxma. A ruin model with dependence between claim sizes and claim intervals. *Insurance: Mathematics and Economics*, 35:245–254, 2004.

H. Albrecher, H. U. Gerber, and H. Yang. A direct approach to the discounted penalty function. *North American Actuarial Journal*, 14(4):420–434, 2010.

H. Albrecher, C. Constantinescu, and S. Loisel. Explicit ruin formulas for models with dependence among risks. *Insurance: Mathematics and Economics*, 48(2):265–270, 2011.

H. Albrecher, J. Beirlant, and J. L. Teugels. *Reinsurance: actuarial and statistical aspects*. John Wiley & Sons, 2017.

H. Albrecher, J. C. Araujo-Acuna, and J. Beirlant. Fitting nonstationary Cox processes: An application to fire insurance data. *North American Actuarial Journal*, 25(2):135–162, 2021.

E. S. Andersen. On the collective theory of risk in case of contagion between claims. *Bulletin of the Institute of Mathematics and its Applications*, 12(2):275–279, 1957.

APCIA, CAS, and SOA. Auto loss cost trends 2019 update, 2020. URL `https://www.soa.org/resources/research-reports/2020/auto-loss-cost/`.

B. L. Arkin and L. M. Leemis. Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process from overlapping realizations. *Management Science*, 46(7):989–998, 2000.

S. Asmussen. Risk theory in a Markovian environment. *Scandinavian Actuarial Journal*, 1989(2):69–100, 1989.

S. Asmussen and H. Albrecher. *Ruin probabilities*, volume 14. World scientific, 2010.

S. Asmussen and K. Binswanger. Simulation of ruin probabilities for subexponential claims. *ASTIN Bulletin: The Journal of the IAA*, 27(2):297–318, 1997.

S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer Science & Business Media, 2007.

S. Asmussen and T. Rolski. Risk theory in a periodic environment: the Cramer-Lundberg approximation and Lundberg's inequality. *Mathematics of Operations Research*, 19(2): 410–433, 1994.

B. Avanzi, G. Taylor, B. Wong, and A. Xian. Modelling and understanding count processes through a Markov-modulated non-homogeneous Poisson process framework. *European Journal of Operational Research*, 290(1):177–195, 2021a.

B. Avanzi, G. Taylor, B. Wong, and X. Yang. On the modelling of multivariate counts with Cox processes and dependent shot noise intensities. *Insurance: Mathematics and Economics*, 99:9–24, 2021b.

A. Balbás, J. Garrido, and S. Mayoral. Properties of distortion risk measures. *Methodology and Computing in Applied Probability*, 11(3):385, 2009.

R. E. Beard, T. Pentikäinen, and E. Pesonen. *Risk theory.* Chapman & Hall, London, 3rd edition, 1984.

M. P. Berg and S. Haberman. Trend analysis and prediction procedures for time non-homogeneous claim processes. *Insurance: Mathematics and Economics*, 14(1):19–32, 1994.

A. V. Boikov. The Cramér–Lundberg model with stochastic premium process. *Theory of Probability & Its Applications*, 47(3):489–493, 2003.

M. Boudreault, H. Cossette, D. Landriault, and E. Marceau. On a risk model with a dependence between interclaim arrivals and claim sizes. *Scandinavian Actuarial Journal*, 5:265–285, 2006.

E. O. Brigham. *The fast Fourier transform and its applications.* Prentice-Hall, Inc., 1988.

J. Bucklew. *Introduction to rare event simulation.* Springer Science & Business Media, 2013.

E. Çinlar. *Introduction to stochastic processes.* Courier Corporation, 2013.

T. Y. Chang, W. Huang, and Y. Wang. Something in the Air: Pollution and the Demand for Health Insurance. *The Review of Economic Studies*, 85(3):1609–1634, 03 2018. ISSN 0034-6527. doi: 10.1093/restud/rdy016. URL `https://dx.doi.org/10.1093/restud/rdy016`.

A. S. Chernobai, S. T. Rachev, and F. J. Fabozzi. *Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis.* Wiley Publishing, 2007.

L. J. Christiano and T. J. Fitzgerald. The band pass filter. *International Economic Review*, 44(2):435–465, 2003.

J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

A. Copeland, W. Dunn, and G. Hall. Inventories and the automobile market. *The RAND Journal of Economics*, 42(1):121–149, 2011.

H. Cramér. *Collected Works*, volume 2. Springer-Verlag, 1955.

A. Dassios and J.-W. Jang. Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. *Finance and Stochastics*, 7(1):73–95, 2003.

A. Dassios and S. Wu. Parisian ruin with exponential claims. *Working paper, LSE, London*, page 11 pages, 2008. URL `http://stats.lse.ac.uk/angelos/docs/exponentialjump.pdf`.

A. Dassios, J. Jang, and H. Zhao. A risk model with renewal shot-noise Cox process. *Insurance: Mathematics and Economics*, 65:55–65, 2015.

C. D. Daykin, T. Pentikäinen, and M. Pesonen. *Practical risk theory for actuaries*. Chapman & Hall, London, 1st edition, 1994.

D. C. Dickson. On the distribution of the surplus prior to ruin. *Insurance: Mathematics and Economics*, 11:191–2–7, 1992.

C. Dutang and A. Charpentier. *CASdatasets: Insurance Datasets*, 2020. R package version 1.0-11.

P. M. Ellis. Characterization of the annual pattern of life insurance sales. *The Journal of Risk and Insurance*, 41(4):735–738, 1974. ISSN 00224367, 15396975. URL `http://www.jstor.org/stable/251969`.

P. Embrechts and M. Frei. Panjer recursion versus FFT for compound distributions. *Mathematical Methods of Operations Research*, 69(3):497–508, 2009.

P. Embrechts and H. Schmidli. Ruin estimation for a general insurance risk model. *Advances in Applied Probability*, 26:404–422, 1994.

H. U. Gerber and E. S. Shiu. On the time value of ruin. *North American Actuarial Journal*, 2(1):48–72, 1998.

J. Grandell. Empirical bounds for ruin probabilities. *Stochastic Processes and their Applications*, 8(3):243–255, 1979.

J. Grandell. *Aspects of risk theory*. Springer Science & Business Media, 2012.

R. Grübel and R. Hermesmeier. Computation of compound distributions I: Aliasing errors and exponential tilting. *ASTIN Bulletin: The Journal of the IAA*, 29(2):197–214, 1999.

A. Guillou, S. Loisel, and G. Stupfler. Estimating the parameters of a seasonal Markov-modulated Poisson process. *Statistical Methodology*, 26:103 – 123, 2015. ISSN 1572-3127. doi: https://doi.org/10.1016/j.stamet.2015.04.003. URL `http://www.sciencedirect.com/science/article/pii/S1572312715000325`.

M. R. Hardy. Approximating the aggregate claims distribution. *Encyclopedia of Actuarial Science*, 1, 2006.

S. G. Henderson. Estimation for nonhomogeneous Poisson processes from aggregated data. *Operations Research Letters*, 31(5):375–382, 2003.

R. J. Hodrick and E. C. Prescott. Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1):1–16, 1997.

S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss models: from data to decisions*, volume 715. John Wiley & Sons, 2012.

C. Labbé and K. P. Sendova. The expected discounted penalty function under a risk model with stochastic income. *Applied Mathematics and Computation*, 215(5):1852–1867, 2009.

C. Labbé, H. S. Sendov, and K. P. Sendova. The Gerber-Shiu function and the generalized Cramér-Lundberg model. *Applied Mathematics and Computation*, 218:3035–3056, 2011.

L. M. Leemis. Nonparametric estimation of the cumulative intensity function for a non-homogeneous Poisson process. *Management Science*, 37(7):886–900, 1991.

S. X. Lin and K. P. Pavlova. The compound Poisson risk model with a threshold dividend strategy. *Insurance: Mathematics and Economics*, 28:57–80, 2006.

S. X. Lin and G. E. Willmot. The moments of the time of ruin, the surplus before ruin, and the deficit at ruin. *Insurance: Mathematics and Economics*, 27:19–44, 2000.

S. X. Lin, G. E. Willmot, and S. Drekic. The classical risk model with a constant dividend barrier: analysis of the Gerber-Shiu function. *Insurance: Mathematics and Economics*, 33:551–566, 2003.

Y. Lu and J. Garrido. Doubly periodic non-homogeneous Poisson models for hurricane data. *Statistical Methodology*, 2(1):17–35, 2005.

Y. Lu and S. Li. On the probability of ruin in a Markov-modulated risk model. *Insurance: Mathematics and Economics*, 37(3):522–532, 2005.

F. Lundberg. *I. Approximerad framstallning af sannolikhetsfunktionen: II. Aterforsakring af kollektivrisker*. 1903.

D. McLeish. Simulating random variables using moment-generating functions and the saddlepoint approximation. *Journal of Statistical Computation and Simulation*, 84(2): 324–334, 2014.

Y. Miao, W. Huang, B. L. Jones, and K. P. Sendova. Some observations on the temporal patterns in the surplus process of an insurer. *a manuscript,* pp. 38, 2021.

M. Morales. On a surplus process under a periodic environment: a simulation approach. *North American Actuarial Journal*, 8(4):76–89, 2004.

B. G. Osgood. *Lectures on the Fourier transform and its applications*, volume 33. American Mathematical Soc., 2019.

H. Pham. Some applications and methods of large deviations in finance and insurance. In *Paris-Princeton Lectures on Mathematical Finance 2004*, pages 191–244. Springer, 2007.

M. O. Ravn and H. Uhlig. On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of economics and statistics*, 84(2):371–376, 2002.

G. Rubino and B. Tuffin. *Rare event simulation using Monte Carlo methods*. John Wiley & Sons, 2009.

S. Schlegel. Ruin probabilities in perturbed risk models. *Insurance: Mathematics and Economics*, 22(1):93–104, 1998.

K. P. Sendova, C. Yang, and R. Zhang. Dividend barrier strategy: Proceed with caution. *Statistics & Probability Letters*, 137:157–164, 2018.

B. Sundt and W. S. Jewell. Further results on recursive evaluation of compound distributions. *ASTIN Bulletin: The Journal of the IAA*, 12(1):27–39, 1981.

G. Temnov. Risk process with random income. *Journal of Mathematical Sciences*, 123 (1):3780–3794, 2004.

G. Temnov. Risk models with stochastic premium and ruin probability estimation. *Journal of Mathematical Sciences*, 196(1):84–96, 2014.

M. Vidmar. Ruin under stochastic dependence between premium and claim arrivals. *Scandinavian Actuarial Journal*, 2018(6):505–513, 2018.

C. Yang, K. P. Sendova, and Z. Li. Parisian ruin with a threshold dividend strategy under the dual Lévy risk model. *Insurance: Mathematics and Economics*, 90:135–150, 2020.

H. Yang and L. Zhang. Spectrally negative Lévy processes with applications in risk theory. *Advances in Applied probability*, 33:281–291, 2001.

Y. Zhao and C. Yin. The expected discounted penalty function under a renewal risk model with stochastic income. *Applied Mathematics and Computation*, 218:6144–6154, 2012.

Curriculum Vitae

Name:	Yang Miao
Post-Secondary Education and Degrees:	Ph.D. in Statistics, Actuarial Science University of Western Ontario, 2017-2022
	M.Sc. in Statistics, Actuarial Science University of Western Ontario, 2016 - 2017
	B.Sc. in Applied Mathematics, Statistics South China University of Technology, 2012 - 2016
Honours and Awards:	Michael Bean Society of Actuaries' Graduate Scholarship 2017
	MITACS Accelerate Award *Risk aggregation beyond the normal limits* 2018
	Western Graduate Research Scholarship 2017 - 2021
Related Work Experience:	Sessional Lecturer University of Western Ontario, 2018
	Teaching Assistant University of Western Ontario, 2016 - 2021
	Research Assistant University of Western Ontario, 2016 - 2021

Publications:

Y. Miao, W. Huang, B. L. Jones, and K. P. Sendova. Some observations on the temporal patterns in the surplus process of an insurer, submitted.

Y. Miao, B. L. Jones, and K. P. Sendova. On a risk model with dual seasonalities, reviewed by *North American Actuarial Journal*, under revision.

Y. Miao, K. P. Sendova. A simulation approach to a risk model with stochastic premium income, a manuscript.