Institute
and Faculty
of Actuaries

SESSIONAL PAPER

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

Yang Miao[1]*, Kristina P. Sendova[1], Bruce L. Jones[1] and Zhong Li[2]

[1]Department of Statistical and Actuarial Sciences, Western University, London, Ontario, Canada
[2]School of Insurance and Economics, University of International Business and Economics, Beijing, China
*Correspondence to: Yang Miao, Department of Statistical and Actuarial Sciences, Western University, London, Ontario, Canada. E-mail: ymiao42@uwo.ca

Abstract

In this paper, we explore potential surplus modelling improvements by investigating how well the available models describe an insurance risk process. To this end, we obtain and analyse 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 utilises a non-homogeneous Poisson process for premium counts and a Cox process for claim counts that reflect the specific features of the data.

Keywords: Cox process; Insurance; Seasonality; Surplus process

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 & Shiu (1998), and Lin & Willmot (2000).

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 Sparre Anderson (1957) 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 & Schmidli

(1994) studied a risk model where borrowing, investment and inflation were incorporated. Yang & Zhang (2001) proposed modelling the insurer's surplus using a spectrally negative Lévy process. Lin et al. (2003) introduced an absorbing barrier to model a dividend strategy. Subsequently, Lin & Pavlova (2006) extended this model to a threshold dividend strategy. Albrecher & Boxma (2004) and Boudreault et al. (2006) explored dependence structures linking inter-claim times and claim amounts. Dassios & 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 paper, we offer some data-driven extensions of the classical model. We study two data sets containing premium payments and claim payments. 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. 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 analyses 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 (2002), Labbé et al. (2011), Zhao & Yin (2012) and several others. After analysing our data, we propose to generalise the aggregate premium process by replacing the homogeneous Poisson premium-counting process by a non-homogeneous Poisson process.

Secondly, we find 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 & 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.

Lastly, 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. (2020) and Avanzi et al. (2021). 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 to provide 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. 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 paper is structured as follows. The data set is described briefly in Section 2. In Section 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 we find. In Section 4, we obtain some ruin theory results using simulation. We also compare the proposed model with existing models. Conclusions are drawn in Section 5.

2. Data

We obtained two data sets for the purpose of this paper. 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 paper. For this reason, the first data set is used extensively in this paper. 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 paper. Instead, it is used to validate the assumptions we use in the modelling process. This data set is described in Section 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:* uniquely identifies the vehicle.
- *Premium:* the single premium for the policy. The premium is paid in a lump sum.
- *Premium date:* the date when the premium was collected.
- *Accident date:* 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:* the date when the claim was paid out.
- *Claim amount:* the 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 paper, we may also build models for premium sizes and claim severities from recorded premium amounts and claim sizes. From the ruin theory perspective, this should be sufficient to fit a risk model.

A brief summary of the first data set is given in Table 1.

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.

3. Methodology

The second data set provides much more detail 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.

Table 1. Summary of the first data set

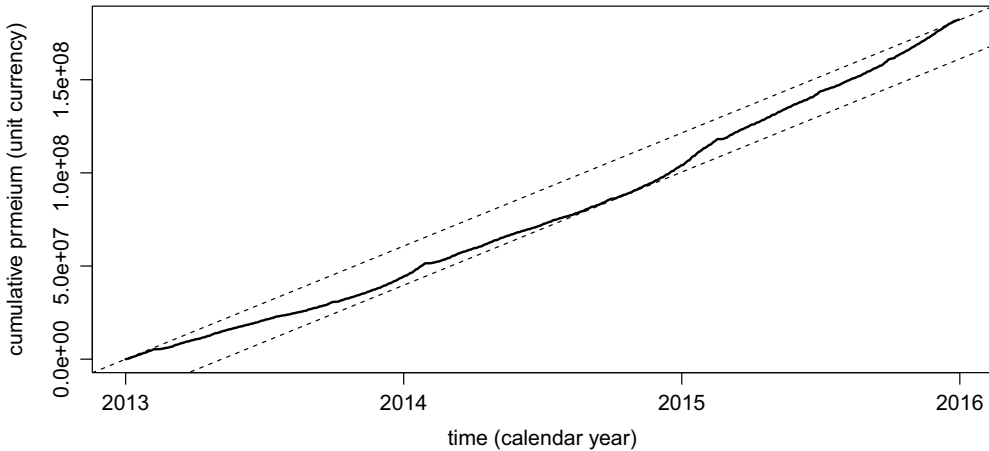| Year	Number of records	Number of claims
2013	13,180	4,442
2014	17,632	5,527
2015	23,406	4,792

Figure 1. Cumulative premium with dashed auxiliary lines to show the convexity of the plot.

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.

3.1 Characteristics of the Premium 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 1.

Two dashed parallel auxiliary lines are added to emphasise 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 analyse the premium distribution for different years. The empirical distributions of the premiums are given in Figure 2.

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.

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 ($1\% - 2\%$ annually). 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 that are illustrated in Figure 3. Since most sales happened during workdays, we see a strong weekly

Table 2. Summary of premium sizes by year

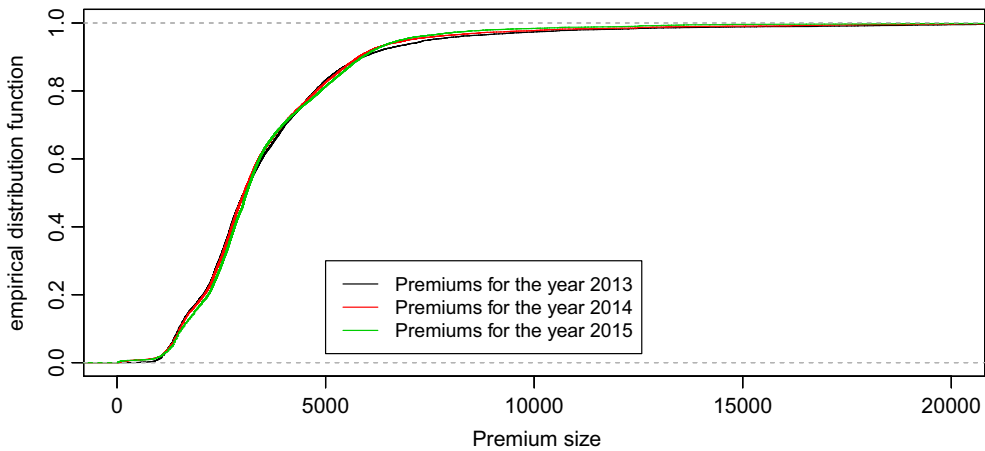| Year	1st quartile	Median	Mean	3rd quartile
2013	2,282	3,047	3,668	4,392
2014	2,290	3,039	3,593	4,363
2015	2,370	3,084	3,571	4,364

Figure 2. The empirical distribution of the premium sizes for different years.
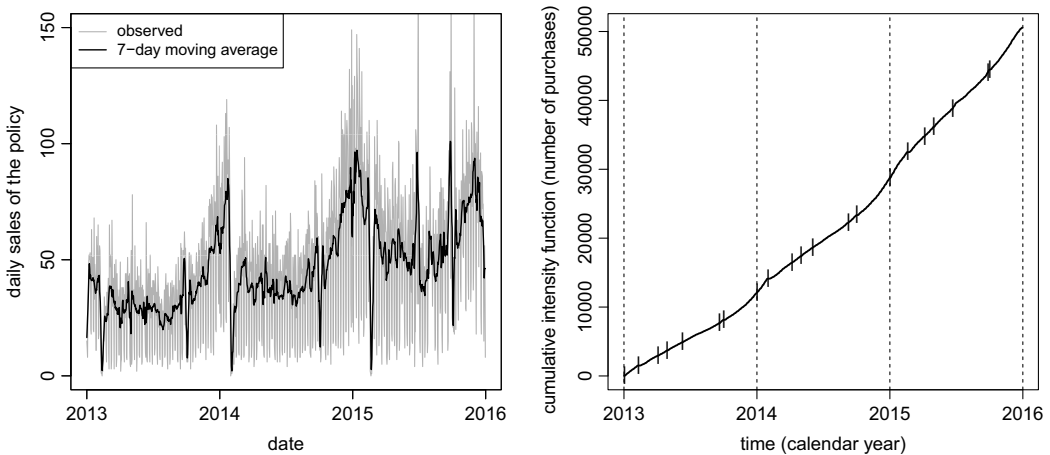

Figure 3. Left: observed daily sales of the policy with 7-day moving average. Right: observed cumulative sales of the policy, vertical short lines indicate public holidays.

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.

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 further research. See, for example, Boikov (2002), Labbé & 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 & Leemis (2000) and Henderson (2003). Asymptotic properties of these estimators as the number of observations increases to infinity are derived in these works. Chernobai et al. (2007) demonstrate that when dealing with one realisation 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 3. In this plot, we use vertical short solid lines to indicate public holidays in the region where the insurance company operates.

Some patterns in the data are immediately noticeable in Figure 3. There is a clear yearly cycle in the sales of the policy, which means that the premium income is not uniform over the 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 want to emphasise that the goal here is to simply observe whether there is any temporal pattern in the purchasing process. The methods employed here are sufficient to illustrate the existence of seasonalities but may be improved if one is interested in fitting the proposed model to their own data. For instance, one may use more sophisticated predictive models to better understand what drives these periodic patterns in their specific data set. For the purpose of this paper, we use simpler methods because the interpretation of their results is more straightforward.

For the estimated 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 minimising 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 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 for R in our analysis (Balcilar et al. 2019). Similar packages are available for different languages, such as the *statsmodels* module for Python (Seabold & Perktold 2010). 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 growth component g_t and a cyclical component c_t,

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

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

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

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 & Prescott (1997). Ravn & Uhlig (2002) showed that the parameter λ should be adjusted
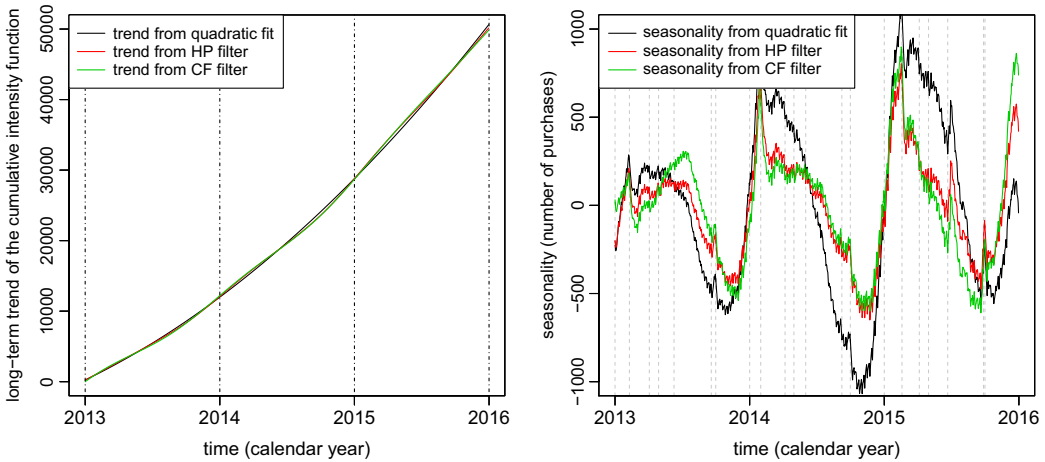
Figure 4. Left: Comparison of the long-term trend of the cumulative intensity function of insurance policy purchases captured by different algorithms. Right: comparison of the seasonalities of the cumulative intensity function of insurance policy purchases captured by different algorithms, dotted vertical lines mark the public holidays in the region where the insurance company operates.

according to the fourth power of the frequency of observations. Based on this result, we choose $\lambda = 1600 \times 90^4$ 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 analyses 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 & Fitzgerald (2003). As it is clear from Figure 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 4.

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. Although the HP filter and the CF filter are developed based on different mechanisms, they yield virtually identical seasonality components. This further confirms the existence of the seasonal fluctuations in the data set. 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 fewer policies are sold in the middle of a year compared to other times of the year.

We observe that there are peaks in the seasonality curve in Figure 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 a larger impact on consumer behaviours.
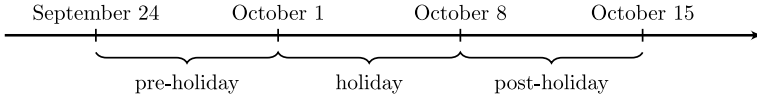
Figure 5. Illustration of three time periods around a public holiday (assume October 1 – October 7 are public holidays).

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 the public holiday. We choose 7 days to eliminate possible weekly fluctuation in purchases. Figure 5 gives an illustration of these three periods assuming that October 1–October 7 is a public holiday.

Analysing the impact of a specific event is frequently conducted across different disciplines. For count data, a generalised 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, 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 and the impact of weekends, together with the impact of public holidays. We assume

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

(1)

where $H(t)$ is the number of policies sold on 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 estimates of these parameters are given in Table 3.

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 sales 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 4. We also notice that there is a 67% decrease on weekends, which explains the pattern observed in Figure 3.

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

$$\omega = \arctan\left(\frac{\hat{\gamma}_{sin}}{\hat{\gamma}_{cos}}\right) = 0.0249 \approx 0.0079\pi,$$

(2)

Table 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)	Significance level
β_0	0.2701	1.310039	***
β_1	−1.8147	0.162894	***
β_2	−0.0872	0.916525	***
β_3	−1.1192	0.326527	***
γ_0	3.5940	36.37954	***
γ_1	0.0008	1.000787	***
γ_{cos}	0.2486	1.282246	***
γ_{sin}	0.0062	1.006248	

Table 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%	

$$\alpha = \frac{\hat{\gamma}_{cos}}{\cos(\omega)} = 0.2487.$$

Equation 2 indicates that the seasonality component is a cosine function. This is consistent with the seasonality obtained in Figure 4.

Using the estimates in Table 3, the parameter (1) becomes

$$\log \zeta(t) = 0.27 I_0(t) - 1.81 I_1(t) + -0.09 I_2(t) - 1.12 I_3(t)$$

$$+ 3.5940 + 0.0008t + 0.2487 \sin\left(\frac{2\pi}{365}(t + 89.8)\right). \tag{3}$$

This estimate is used later as the intensity function in the simulation study.

In this section, we considered different components in the purchasing process. As shown in Figures 3 and 4, and subsequent analysis, the intensity function of the Poisson process should incorporate long-term trend, seasonalities and impact of public holidays.

3.2 Characteristics of the Claim Process

We next analyse the patterns exhibited by the claim process over time. In this subsection, we study the claim process as a stand-alone process. In Subsection 3.3, we explore a possible relationship between the purchasing process and the claim process. Analysing 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 3.1, we omit the technical details and simply state the results.
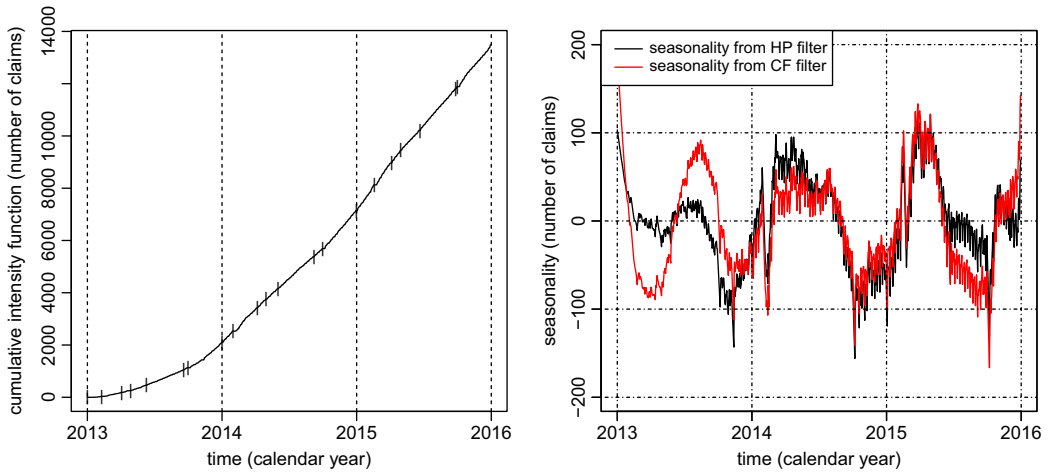
Figure 6. Left: estimated cumulative intensity function. Right: seasonal patterns of the claim process.

Table 4 summarises the claims by year and season. As shown in this table, more claims happened in winter. A Chi-square test 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.

Using the same algorithms, the estimated cumulative intensity function and the seasonality are plotted in Figure 6. The cumulative intensity function is again convex, which means that claims are paid increasingly frequently. A possible explanation is explored in Section 3.3. Also, there are more claims in winter than in summer, which is consistent with Table 4.

We notice that the seasonal pattern in the first year is different from that in the last two years. As mentioned in Section 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 is 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.

3.3 Relations between the Two Processes

In this subsection, we investigate a possible relationship between the two counting processes. We discover in Sections 3.1 and 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 year, then one would expect more claims in that period. In light of this, we explore the evolution of the *exposure* of the insurer over time. This is a natural extension to the *collective risk model* (Chapter 9 of Klugman et al. (2012)).

Let $M(t)$ denote the non-homogeneous Poisson process for the premium arrivals and let $\mu(t)$ be the intensity function of $M(t)$. Based on the results in Section 3.2, let

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

where κ is a constant representing the initial business scale and $g(t)$ and $s(t)$ are two generic functions representing the growth component and the seasonality component, respectively.
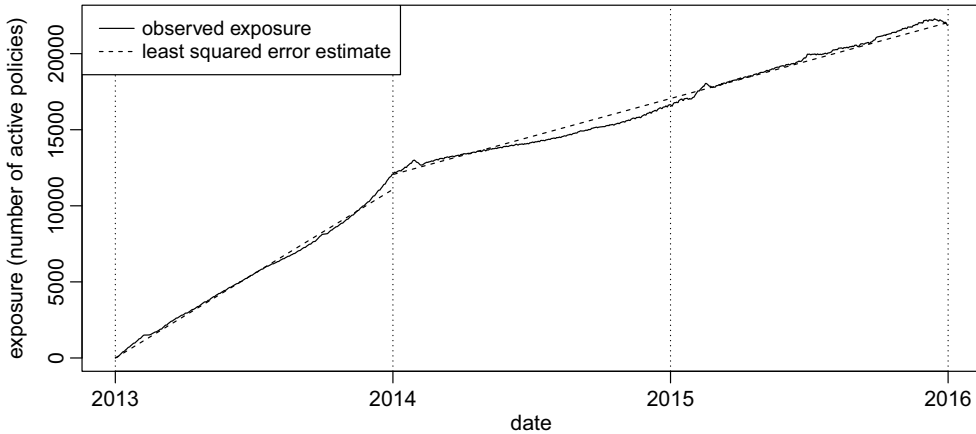
Figure 7. Estimated exposure, using policy purchasing dates as effective dates.

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

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

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

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

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[\xi(t)] = \begin{cases} \kappa + g(t) + s(t), & t \leq \ell \\ [g(t) - g(t - \ell)] + [s(t) - s(t - \ell)], & t > \ell \end{cases}, \tag{4}$$

Consider $[s(t) - s(t - \ell)]$ in Equation 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 7 is the observed exposure from the data set. A linear function is fitted to the estimated exposure using the least-squares-error estimates. There is no evidence of yearly fluctuation.

Another simplifying assumption we make is that an insurance policy becomes effective immediately upon purchase. 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 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
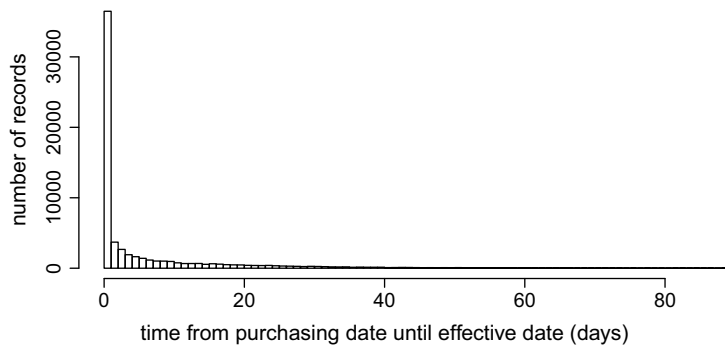
Figure 8. Histogram of the time difference between purchasing and effective date.

process will impact the claim process. One may track different drivers of seasonality for different processes and determine whether there is an interaction between them based on the specification of the portfolio.

3.4 A Cox Process Modelling the Claim Arrivals

We argue in Section 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 that 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 follow 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.

Proposition 3.1.. Suppose $N_1, N_2, \ldots$ 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. 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 maximising the likelihood function in a similar way to the analysis in Section 3.1. While fitting the Cox process, the observed exposure, as shown in Figure 7, is used as the offset. Figure 9 compares the outputs of the three different models. The Cox model is
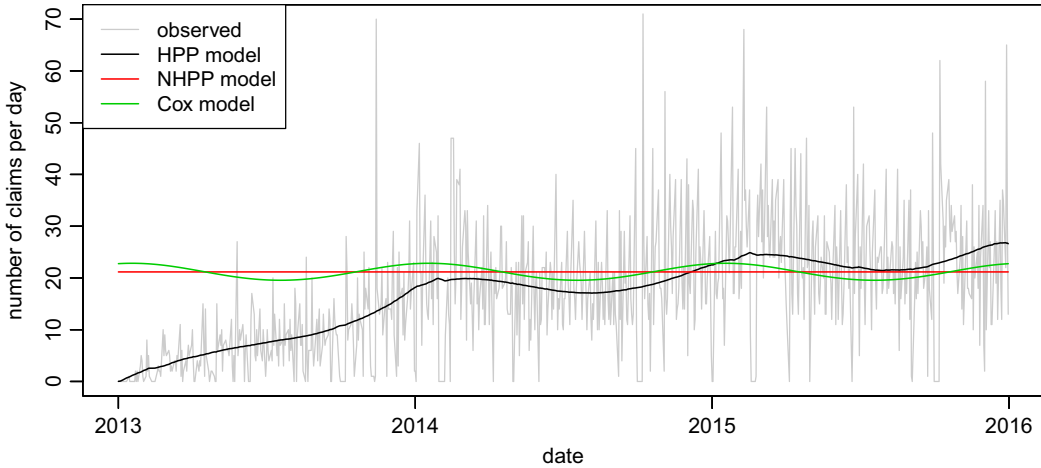
Figure 9. Comparison of three different numbers of claims. Three models are fitted to the data, then the predicted number of claims is calculated and plotted for each model.

Table 5. Summary statistics for model comparison

Statistics for 2014–2015	HPP	NHPP	Cox
$\sum e_i^2$	72,361	71,664	68,724
AIC	6,250	6,221	6,083
statistics for 2013–2015	HPP	NHPP	Cox
$\sum e_i^2$	120,737	119,788	80,018
AIC	11,205	11,154	8,123

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 summary statistics are provided in Table 5. We compare the sum of squared error and the Akaike information criteria for different models. As shown by these statistics, the Cox process yields the smallest error, and hence is the closest to the data.

Remark 1. The differences between any two models in terms of error statistics for this data set are 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 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. □

Table 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

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 to reflect the patterns exhibited in the data set as follows:

$$U(t) = u + \sum_{k=1}^{M(t)} X_k - \sum_{i=1}^{N(t)} Y_i, \quad u \geq 0, \tag{5}$$

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 $v(t)$, respectively. The dependence structure between $M(t)$ and $N(t)$ is given by

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

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 in Equation 3 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

$$\mu(t) = 365 \exp[3.5940 + 0.2487 \cdot \sin(2\pi(t + 0.246027))], \tag{6}$$

$$v(t) = \xi(t) \cdot [0.488972 + 0.074706 \cdot \sin(2\pi(t + 0.120373))], \tag{7}$$

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 the 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. The quarterly risk measures are given in Table 6.

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 6 and 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 $\nu(t)$ is the exposure at time t. We allow the sinusoidal functions to shift horizontally to capture the impact of different combinations of seasonalities. Although in the data set that we analysed, premium arrivals and claim arrivals have peak seasons around the same time of year, it is possible that the two seasonalities are overall unsynchronised. By shifting the functions representing the seasonalities along the horizontal axis, we are able to accommodate 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

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

We consider different combinations of a and b. The one-year ruin probability is given in Figure 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 10 shows that the seasonality in the premium process also has impact on the riskiness of the business.

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, CAS, PCI, and SOA (2018), examines the drivers of 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 processes to model these characteristics. For example, Lu & 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. (2020) construct a Cox process by using a subordinator and demonstrate the success of this model using Dutch fire
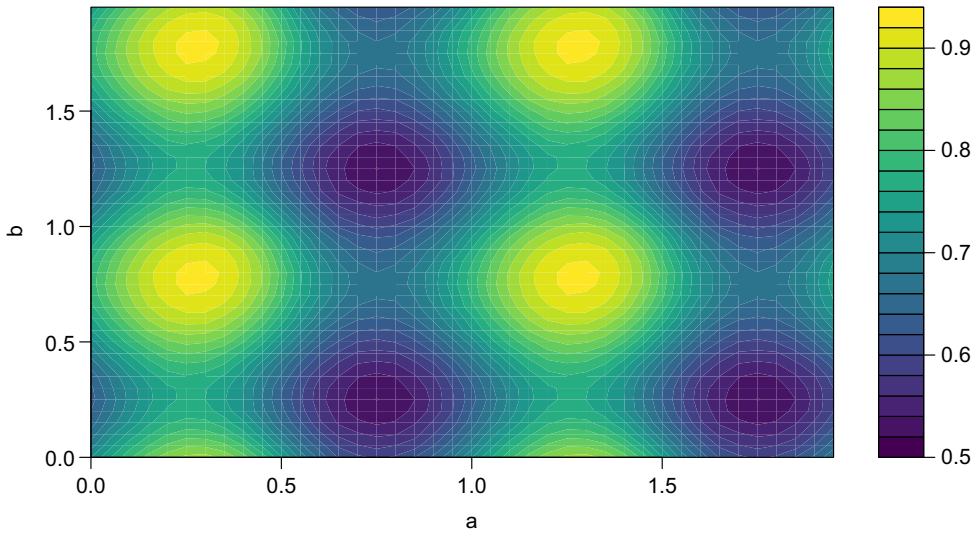
Figure 10. Contour plot of one-year ruin probability with different combinations of seasonalities: *a* represents the initial season of the premium process, *b* represents the initial season of the claim process.

insurance claims. Avanzi et al. (2021) construct a Markov-modulated Poisson process to account for both known information and unobservable drivers. In this study, 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 an input in the hidden Markov chain calibration. While the model is different, similar results are obtained in this paper.

Other research projects are dedicated to improving premium modelling. The seasonality in the premium is discussed by Asmussen & 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 (2002), Labbé & 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 to be stationary, and the premium and the claims are assumed to be 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 (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
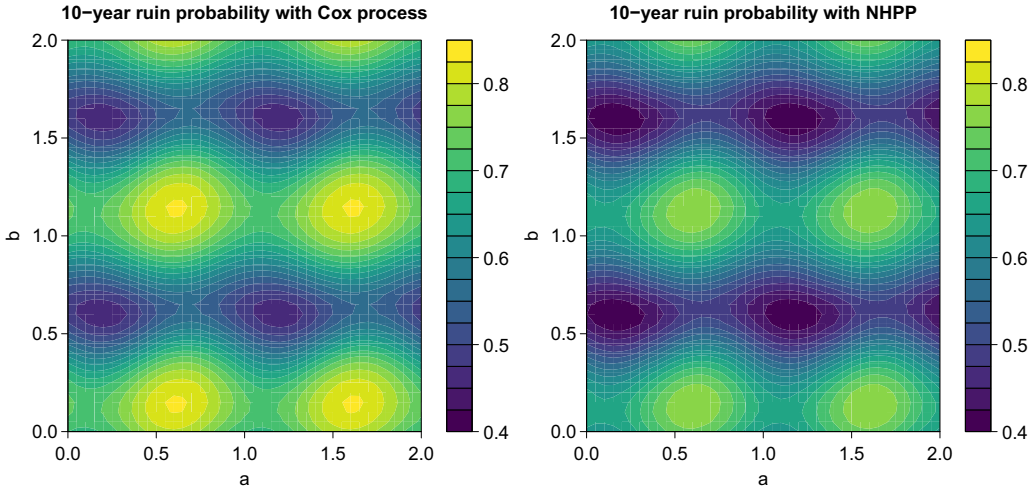
Figure 11. Contour plot of 10-year ruin probability with different combinations of seasonality of the premiums (a) and seasonality of the claims (b). Left: the claim arrivals from a Cox process Right: the claim arrivals from a NHPP.

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:

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

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

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 11. In this scenario, the model with dependence is able to capture the additional risk.

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 risk-based capital requirements, while the Life Insurance Capital Adequacy Test 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 by Institute of Risk Management (2015) 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. With a more representative model of their cash flows, insurers are better positioned to manage their assets to meet their future obligations. This could improve both the insurer's ability to withstand the risk of loss due to fluctuations in their business, and potentially the profitability.

5. Conclusion

In this paper, we study the time patterns of the premium and claim processes of an insurer. We find that both processes exhibit increasing intensities with seasonal fluctuations. Major public holidays also have an 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 utilises both a non-homogeneous Poisson process and a Cox process as counting processes.

The model suggested in this paper 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 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. By studying the time patterns of cash inflow and cash outflow, insurers are in a better position to optimally manage their assets, to achieve both higher profitability and financial security.

Due to the lack of appropriate data, the authors 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 should be examined. Theoretical results are yet to be derived under this model. Simulation techniques may be employed to obtain the results of interest.

References

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

Albrecher, H., Araujo-Acuna, J. C., & Beirlant, J. (2020). Fitting nonstationary Cox processes: an application to fire insurance data. *North American Actuarial Journal*, **25**(2), 1–28.

Arkin, B. L., & Leemis, L. M. (2000). Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process from overlapping realizations. *Management Science*, **460**(7), 989–998.

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

Asmussen, S., & Rolski, T. (1994). Risk theory in a periodic environment: the Cramér-Lundberg approximation and Lundberg's inequality. *Mathematics of Operations Research*, **190**(2), 410–433.

Avanzi, B., Taylor, G., Wong, B., & Xian, A. (2021). Modelling and understanding count processes through a Markov-modulated non-homogeneous Poisson process framework. *European Journal of Operational Research*, **2900**(1), 177–195.

Balcilar, M., Hodrick-Prescott, B., & Balcilar, M. M. (2019). Package 'mFilter'.

Beard, R. E., Pentikäinen, T., & Pesonen, E. (1984). *Risk Theory* (3rd ed.). Chapman & Hall, London.

Boikov, A. (2002). The Cramér-Lundberg model with stochastic premium process. *Theory of Probability and Its Applications*, **470**(3), 281–291.

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

CAS, PCI, and SOA. (2018). Auto loss cost drivers: physical damage. https://www.soa.org/globalassets/assets/files/resources/research-report/2018/auto-loss-cost-drivers.pdf.

Chang, T. Y., Huang, W., & Wang, Y. (2018). Something in the air: pollution and the demand for health insurance. *The Review of Economic Studies*, **850**(3), 1609–1634. ISSN 0034-6527. https://doi.org/10.1093/restud/rdy016.

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

Christiano, L. J., & Fitzgerald, T. J. (2003). The band pass filter. *International Economic Review*, **440**(2), 435–465.

Cramér, H. (1955). *Collected Works*, vol. **2**. Springer-Verlag.

Dassios, A., & Wu, S. (2008). Parisian ruin with exponential claims. *Working Paper, LSE, London*, 11 pages. http://stats.lse.ac.uk/angelos/docs/exponentialjump.pdf.

Daykin, C. D., Pentikäinen, T., & Pesonen, M. (1994). *Practical Risk Theory for Actuaries* (1st ed.). London: Chapman & Hall.

Dickson, D. C. (1992). On the distribution of the surplus prior to ruin. *Insurance: Mathematics and Economics*, **11**, 191–207.

Ellis, P. M. (1974). Characterization of the annual pattern of life insurance sales. *The Journal of Risk and Insurance*, **410**(4), 735–738. ISSN 00224367, 15396975. http://www.jstor.org/stable/251969.

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

Gerber, H. U., & Shiu, E. S. (1998). On the time value of ruin. *North American Actuarial Journal*, **20**(1), 48–78.

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

Henderson, S. G. (2003). Estimation for nonhomogeneous Poisson processes from aggregated data. *Operations Research Letters*, **310**(5), 375–382.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit and Banking*, **290**(1), 1–16.

Institute of Risk Management (2015). Operational risk modelling: common practices and future development. https://www.theirm.org/media/6809/irm_operational-risks_booklet_hi-res_web-2.pdf

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss Models: From Data to Decisions*, vol. **715**. John Wiley & Sons.

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

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

Leemis, L. M. (1991). Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process. *Management Science*, **370**(7), 886–900.

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

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

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

Lu, Y., & Garrido, J. (2005). Doubly periodic non-homogeneous Poisson models for hurricane data. *Statistical Methodology*, **20**(1), 17–35.

Lundberg, F. (1903). I. Approximerad framstallning af sannolikhetsfunktionen: II. Aterforsakring af kollektivrisker.

Morales, M. (2004). On a surplus process under a periodic environment. *North American Actuarial Journal*, **80**(4), 76–89.

Ravn, M. O., & Uhlig, H. (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of Economics and Statistics*, **840**(2), 371–376.

Seabold, S., & Perktold, J. (2010). Statsmodels: econometric and statistical modeling with python, in *Proceeding of the 9th Python in Science Conference*, Vol. 57, pp. 10–25080, Austin, TX.

Temnov, G. (2004). Risk process with random income. *Journal of Mathematical Sciences*, **1230**(1), 3780–3794.

Temnov, G. (2014). Risk models with stochastic premium and ruin probability estimation. *Journal of Mathematical Sciences*, **1960**(1), 84–96.

Yang, H., & Zhang, L. (2001). Spectrally negative Lévy processes with applications in risk theory. *Advances in Applied Probability*, **33**, 281–291.

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