Applications of claim investigation in insurance surplus and claims models

by

Mirabelle Huynh

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Dr. Jiandong Ren
Associate Professor, Department of Statistical and Actuarial Sciences
University of Western Ontario

Supervisor(s): Dr. David Landriault
Professor, Department of Statistics and Actuarial Science
University of Waterloo
Dr. Gordon E. Willmot
Professor, Department of Statistics and Actuarial Science
University of Waterloo

Internal Member: Dr. Steve Drekic
Professor, Department of Statistics and Actuarial Science
University of Waterloo
Dr. Bin Li
Assistant Professor, Department of Statistics and Actuarial Science
University of Waterloo

Internal-External Member: Dr. Qi-Ming He
Professor, Department of Management Sciences
University of Waterloo

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Claim investigation is a fundamental part of an insurer's business. Queues form as claims accumulate and claims are investigated according to some queueing mechanism. The natural existence of queues in this context prompts the inclusion of a queue-based investigation mechanism to model features like congestion inherent in the claims handling process and further to assess their overall impact on an insurer's risk management program.

This thesis explicitly models a queue-based claim investigation mechanism (CIM) in two classical models for insurance risk, namely, insurer surplus models (or risk models) and aggregate claim models (or loss models).

Incorporating a queue-based CIM into surplus and aggregate claims models provides an additional degree of realism and as a result, can help insurers better characterize and manage risk. In surplus analysis, more accurate measures for ruin-related quantities of interest such as those relating to the time to ruin and the deficit at ruin can be developed. In aggregate claims models, more granular models of the claims handling process (e.g., by decomposing claims into those that are settled and those that have been reported but not yet settled) can help insurers target the source of inefficiencies in their processing systems and later mitigate their financial impact on the insurer.

As a starting point, Chapter 2 proposes a simple CIM consisting of one server and no waiting places and superimposes this CIM onto the classical compound Poisson surplus process. An exponentially distributed investigation time is considered and then generalized to a combination of n exponentials. Standard techniques of conditioning on the first claim are used to derive a defective renewal equation (DRE) for the Gerber-Shiu discounted penalty function (or simply, the Gerber-Shiu function) $m(u)$ and probabilistic interpretations for the DRE components are provided. The Gerber-Shiu function, introduced in Gerber and Shiu [1998], is a valuable analytical tool, serving as a unified means of risk analysis as it generates

various ruin-related quantities of interest.

Chapter 3 extends and generalizes the analysis in Chapter 2 by proposing a more complex CIM consisting of a single queue with n investigation units and a finite capacity of m claims. More precisely, we consider CIMs which admit a (spectrally negative) Markov Additive Process (MAP) formulation for the insurer's surplus and the analysis will heavily rely and benefit from recent developments in the fluctuation theory of MAPs. MAP formulations for four possible CIM generalizations are more specifically analyzed.

Chapter 4 superimposes the more general CIM from Chapter 3 onto the aggregate claims process to obtain an aggregate payment process. It is shown that this aggregate payment process has a Markovian Arrival Process formulation that is preserved under considerable generalizations to the CIM. A distributional analysis of the future payments due to reported but not settled claims ("RBNS payments") is then performed under various assumptions.

Throughout the thesis, numerical analyses are used to illustrate the impact of variations in the CIM on the ruin probability (Chapters 2 and 3) and on the Value-at-Risk (VaR) and Tail-Value-at-Risk ($TVaR$) of RBNS payments (Chapter 4).

Concluding remarks and avenues for further research are found in Chapter 5.

Acknowledgements

I would like to express my gratitude to my supervisors, Professors David Landriault and Gordon E. Willmot. Over these last few years, they have served as outstanding mentors who have provided me with invaluable guidance, encouragement, and support both in my research and in my professional life. I am deeply appreciative of the kindness, patience, and generosity they showed when sharing their knowledge and insights, and when providing me with guidance. I am truly grateful for their support and feel very fortunate to have been their student.

Thank you to Professors Steve Drekic, Bin Li, Jiandong Ren, and Qi-Ming He for their valuable comments and suggestions that improved this thesis.

Many thanks to the faculty and staff in the Department of Statistics and Actuarial Science who helped make this work possible and made my PhD years enjoyable.

I wish to thank all my graduate student colleagues who created such a friendly and supportive community. I feel very grateful to have made so many valuable friendships.

Finally, thank you to my friends and family – especially my partner Avery Au and my great friends Angela Smukavich and Nathalie Moon for their care, encouragement, and support.

I gratefully acknowledge financial support from the James C. Hickman Scholar program of the Society of Actuaries and from the Canada Graduate Scholarship of the Natural Sciences and Engineering Research Council of Canada.

Dedication

To my parents and my sister.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis proposes a queueing-theoretic approach to model an insurer's claims investigation process. Two foundational areas of risk theory are addressed, surplus models and aggregate claims models, each constituting a distinct approach to analyzing and characterizing insurer risk.

Aggregate claims models (or loss models) analyze the properties of (total) losses over a finite-time horizon and do not consider the balance of premiums against losses as they are incurred over time. In contrast, surplus models (or risk models) offer a more dynamic point of view; they analyze the stochastic behaviour of an insurer's surplus as premiums are received and claims are paid out over time. The event of "ruin" is the event involving a drop in surplus below any level of capital the insurer wishes or is required to maintain.

Contemporary loss and surplus models typically proceed on the assumptions that every claim is immediately paid and the amount paid always equals the amount claimed (or, stated more implicitly, that any significant effect relating to claim investigation/delays is embedded in the model parameters). In doing so, contemporary models may sacrifice a significant amount of realism since claim investigation mechanisms (and the resulting payment adjustment and delays) are a foundational aspect of the insurance business.

Claim investigations help protect insurers against the harms of moral hazard relating to non-compliant, inflated or fraudulent claims. However, claim investigations are also a significant cost source for insurers who must devote resources to design and run effective investigation processes. Where claim investigation is inefficient or ineffective, insurers face costs related to poor customer service (due to processing delays), loss of profits (due to insufficent claims adjusting), or poor human resource management.

Incorporating a queue-based claim investigation mechanism into surplus and aggregate claims models will provide an additional degree of realism and as a result, can help insurers better characterize and manage risk. For example, higher levels of resolution could be achieved in aggregate claims models (e.g., by decomposing claims into those that are settled and those that have been reported but not yet settled) and more accurate measures for ruin-related quantities of interest such as those relating to the time to ruin and the deficit at ruin.

The remainder of this chapter provides a brief literature review of surplus and aggregate claims analysis, defines surplus and aggregate claims models that are fundamental to the thesis, and introduces mathematical tools that will be utilized. An outline of the thesis concludes the chapter.

1.1 Background

1.1.1 Insurer surplus analysis

The classical compound Poisson risk process (or Cramér-Lundberg model) was first introduced in the pioneering work of Lundberg [1903] to provide a theoretical basis for the risk management of an insurance company (see, e.g., Cramér [1969] and references therein). The Cramér-Lundberg process $\underline{U} = \{U_t\}_{t\geq 0}$ which models the insurer's surplus is defined by

$$U_t = u + ct - S_t, \tag{1.1}$$

where $u\,(u \geq 0)$ is the initial surplus level and $c\,(c > 0)$ is the level premium rate per unit time. The aggregate claim process $\underline{S} = \{S_t\}_{t\geq 0}$ is assumed to be a compound Poisson process, i.e., the aggregate claim at time t is given by

$$S_t = \begin{cases} \displaystyle\sum_{i=1}^{N_t} X_i, & N_t > 0, \\ 0, & N_t = 0, \end{cases} \tag{1.2}$$

where $\underline{N} = \{N_t\}_{t\geq 0}$ is the claim number process which is assumed to follow a homogeneous Poisson process that is independent of all else in the model. The rv's $\{X_i\}_{i=1}^{\infty}$ are assumed to be i.i.d. (also independent of all else in the model) and X_i denotes the amount of the ith claim.

To provide sound risk management, insurers must adequately understand the risk of their financial obligations. This entails understanding worst case scenarios; as such, risk theory is generally focused on understanding "ruin" which is defined as the event in which the insurer's surplus process becomes negative (see, e.g., Asmussen and Albrecher [2010]). Of particular interest is the time to ruin T defined as $T = \inf\{t \geq 0 : U_t < 0\}$ with $T = \infty$ if $U_t > 0$ for all $t \geq 0$.

A risk measure of great interest is the probability of ruin which can consider a finite or infinite time-horizon. The infinite-ruin probability is defined by $\psi(u) = P(T < \infty | U_0 = u)$ and the finite-time ruin probability is given by $\psi(u,t) = P(T \leq t | U_0 = u)$. Both measures can provide insight into the insurer's vulnerability to insolvency.

While infinite-time ruin probabilities have less practical interpretations, they are far more tractable than their finite-time counterparts and can still provide useful insights into an insurer's risk profile. Recent work on the topic of finite-time ruin can be found in, e.g., Dickson and Willmot [2005], Borovkov and Dickson [2008], and Landriault et al. [2011]. Since the Laplace transform of the time to ruin is far more tractable than its distribution function, numerical inversion techniques can be used to compute finite-time ruin probabilities (see Section 1.3 for possible techniques).

Two ruin-related quantities of interest include: the deficit (or shortfall) at ruin $|U_T|$ which corresponds to the minimum capital injection required to bring the insurer back to solvency, and the surplus immediately before ruin U_{T^-}. Note that the claim causing ruin is of size $U_{T^-} + |U_T|$.

A valuable analytical tool to understand the event of ruin is the Gerber-Shiu discounted penalty function (or simply, the Gerber-Shiu function) which is defined as

$$m(u) \equiv E\left[e^{-\delta T} w\left(U_{T^-}, |U_T|\right) 1_{\{T<\infty\}} \left| U_0 = u\right.\right], \tag{1.3}$$

for $\delta \geq 0$ and the so-called "penalty function" $w(x,y)$ for $x, y \geq 0$ satisfies mild integrability conditions (namely, we silently assume that the expectation exists). The Gerber-Shiu function generates various ruin-related quantities of interest and thus acts as a unified means of risk analysis. A considerable amount of risk theory research has been devoted to the function's study after its introduction in the seminal paper of Gerber and Shiu [1998].

In Gerber and Shiu [1998], the function was studied under the classical compound Poisson model (1.1). A comprehensive treatment of the classical surplus process (2.1) can be found in e.g., Rolski et al. [1999, Chapter 1.6] and Asmussen and Albrecher [2010, Chapter 1.1].

The classical surplus process is generalized to the ordinary Sparre Andersen model by relaxing the assumption of a Poisson arrival process and instead assuming claims arrive according to a renewal process (i.e., interclaim times remain i.i.d. but instead follow an arbitrary density). Gerber-Shiu analysis in the ordinary Sparre Andersen model can be found in, e.g., Gerber and Shiu [2005], Li and Garrido [2004], Li and Garrido [2005].

By further introducing a dependence structure between claim sizes and their incurral times, one obtains the dependent Sparre Andersen model. Studies of the Gerber-Shiu function in the dependent Sparre Andersen model can be found in, e.g., Boudreault et al. [2006], Cheung et al. [2010], and Landriault et al. [2014a].

Risk models with Markovian claim arrivals (see, e.g., Ahn and Badescu [2007], Badescu et al. [2005], and Cheung and Landriault [2009a]) have a special role in the present work. A more detailed discussion of studies relating to claim investigation is deferred to Chapter 2 and Chapter 3.

In Chapter 2 and Chapter 3, a queue-based claim investigation mechanism is proposed and superimposed onto the compound Poisson risk process (1.1). In Chapter 2, we are interested in the Gerber-Shiu function (1.3) where the deficit at ruin $|U_T|$ is generalized to $|U_T| + \eta V_T$, where $\eta \in [0, 1]$ and V_t is the total (future) payment amount due to claims under investigation at time t (if any). Hence, we assume that the insurer remains liable for a fixed portion η of the total payment amount of claims undergoing investigation when U drops below 0. The work in Chapter 3 extends Chapter 2 with a more generalized claim investigation mechanism and the joint Laplace transform of the time and generalized deficit at ruin is of main interest.

1.1.2 Aggregate claims analysis

The aggregate loss is a mathematical representation of the total payments made on all claims that occur in a fixed time period on a set of insurance contracts. There are two main ways to aggregate payments: add individual payment amounts as they become due (collective risk model) or add total payment amounts arising from each policy (individual risk model). In the collective risk model, the frequency and severity of losses are modelled separately which can lead to greater flexibility (see, e.g., Klugman et al. [2013]).

Under the collective risk model, the aggregate claims process $\{S_t\}_{t\geq 0}$ is defined using the compound sum representation given by

$$S_t = \begin{cases} \sum_{i=1}^{N_t} X_i, & N_t > 0, \\ 0, & N_t = 0, \end{cases} \tag{1.4}$$

where the claim sizes $\{X_i\}_{i=1}^{\infty}$ are assumed to form a sequence of i.i.d. rv's and the claim number process $\{N_t\}_{t\geq 0}$ is a counting process which is independent of the X_i's. Analyzing the distribution of S_t which is difficult in general has been a central problem in aggregate claims analysis. For a comprehensive treatment of the aggregate claims model given by (1.4), readers are referred to Sundt and Vernic [2009] and [Klugman et al., 2013, Section 7.2, Chapter 9].

By relaxing the independence assumption between the claim frequency and severity, a time dependent claim model is obtained. Time dependent models enable the modelling of realistic features such as interest rates, inflation, and time delays due to claim reporting and processing (see, e.g., Willmot [1989], Jang [2004], Asimit and Badescu [2010], Li et al. [2010], and Xu [2016]).

Mathematical models of aggregate claims that incorporate time lags can be broadly categorized into deterministic methods and stochastic models. Deterministic formulations (e.g., the well-known chain ladder method) are unable to capture the stochastic nature of the claim incurral and reporting processes. They are also unable to model realistic phenomena such as increased congestion in the insurer's claims handling mechanism (and the e.g., associated increase in processing time or rise in the number of claims being paid without investigation).

In Chapter 4, the queue-based claim investigation mechanism proposed in Chapter 3 is superimposed onto the compound model for aggregate claims given by (1.4) where the claim arrival process $\{N_t\}_{t\geq 0}$ is a Poisson process. Chapter 4 develops a stochastic model for claim liabilities; in this way, the randomness inherent in claim liabilities is captured and features such as congestion can be modelled. Models with Markovian claim arrivals such as Ren [2008], Ren [2016], and Kim and Kim [2007], have a special role in the present work.

A number of studies have been devoted to the study of discounted aggregate claims (see, e.g., Kim and Kim [2007] and Ren [2008]). The present work considers the special case where there is no discounting and capitalizes on the tractability of such models. We refer readers to Chapter 4 for further discussion and references related to the analysis of aggregate claims in the context of claims handling.

1.2 Risk models

1.2.1 Spectrally negative Lévy process

There has been considerable research interest in modelling the risk of an insurer's surplus process using the class of spectrally negative Lévy processes which contain the classical compound Poisson process as a special case. The downward jumps and diffusion component of such processes are well-suited to model the main features of an insurer's surplus process: payment of claims and variability from premium or investment income.

Define a strong Markov process $X = \{X_t;\, t \geq 0\}$ with càdlàg paths on a filtered probability space $(\Omega,\, \mathcal{F},\, \mathbb{F},\, \mathbb{P})$. We say X is a *Lévy process* if $P(X_0 = 0) = 1$, the increment $X_t - X_s$ is independent of $\mathcal{F}_s$ and $X_t - X_s$ has the same distribution as X_{t-s} for each $0 \leq s \leq t$. Thus, a Lévy process X is an adapted stochastic process starting from 0 and has stationary and independent increments. We also have that the distribution of X is infinitely divisible. Such models are extensively studied in, e.g., Bertoin [1998] and Kyprianou [2006].

There exists a function Ψ such

$$E\left[e^{isX_t}\right] = e^{-t\Psi(s)},$$

for $t \geq 0$ and $s \in \mathbb{R}$. The function Ψ is called the characteristic function and uniquely determines the distribution of a Lévy process. By the Lévy-Khintchine formula (see, e.g., Kyprianou [2006]), the form of Ψ is given by

$$\Psi(s) = ias + \frac{1}{2}\sigma^2 s^2 + \int_{\mathbb{R}} \left(1 - e^{isx} + isx1_{\{|x|<1\}}\right) \Pi(dx),$$

where $s, a \in \mathbb{R}$, $\sigma \geq 0$ and Π is a measure, called the Lévy measure, is concentrated on $\mathbb{R}\backslash\{0\}$ such that $\int_{\mathbb{R}} (x^2 \wedge 1)\, \Pi(dx) < \infty$.

Using the Lévy-Itô decomposition (see, e.g. Kyprianou [2006]), the Lévy process X can be represented as the independent sum

$$X = X^{(1)} + X^{(2)} + X^{(3)},$$

where $X^{(1)}$ is a linear Brownian motion with drift $-a$ and volatility σ, $X^{(2)}$ is a compound Poisson process with Poisson intensity rate $\Pi(\mathbb{R}/(-1,1))$ and i.i.d. jumps distributed as $\Pi(dx)/\Pi(\mathbb{R}/(-1,1))$, and $X^{(3)}$ determined by Π is a square integrable martingale with an almost surely countable number of jumps which are of magnitude less than 1 on each finite time interval.

The subclass of Lévy processes called the *spectrally negative Lévy process* has the restriction $\Pi(0,\infty) = 0$ such that only downward jumps are possible. In what follows, we assume

X belongs to the subclass of spectrally negative Lévy processes. A spectrally negative Lévy process is uniquely characterized by its Laplace exponent which is a function defined as

$$\psi(z) = \frac{1}{t} \log E\left[e^{zX_t}\right] = -\Psi\left(-iz\right), \tag{1.5}$$

for $z \geq 0$. When $\Pi \subseteq (-\infty, 0)$, the Laplace exponent becomes

$$\psi(z) = -az + \frac{1}{2}\sigma^2 z^2 + \int_{(-\infty,0)} \left(e^{zx} - 1 - zx1_{\{x>-1\}}\right) \Pi(dx), \tag{1.6}$$

which is infinitely differentiable, strictly convex, and $\lim_{z\to\infty} \psi(z) = \infty$. A special case which will become important in Chapter 3 is the perturbed classical Poisson risk process which is obtained by choosing the Laplace exponent

$$\psi(z) = cz + \frac{1}{2}\sigma^2 z^2 + \lambda \left(\int_0^\infty e^{-zy} P(dy) - 1 \right),$$

where $c > 0$ is the premium rate, $\sigma > 0$ is the volatility of the diffusion component, $\lambda > 0$ is the Poisson arrival rate, and P is the df of the i.i.d. jumps.

The definition of a scale function, a fundamental quantity in the study of exit problems, is presented next. For $q \geq 0$, the q-scale function $W^q(\cdot) : \mathbb{R} \mapsto [0, \infty)$ is characterized on $[0, \infty)$ as a strictly increasing and continuous function with Laplace transform

$$\int_0^\infty e^{-zx} W^q(x) dx = \frac{1}{\psi(z) - q}, \tag{1.7}$$

for $z > \Phi^q$ where Φ^q is the largest real solution to $\psi(z) = q$. The existence of scale functions is shown in Kuznetsov et al. [2012]. Also important is the second scale function Z^q defined as

$$Z^q(x) = 1 + q \int_0^x W^q(y) dy,$$

for $x \in \mathbb{R}$.

We define the first passage times of X

$$T_x^{+(-)} = \inf\{t \geq 0 : X_t > (<) x\},$$

for $x \in \mathbb{R}$. Well-known exit results which can be found in [Kyprianou, 2006, Section 8.2] are presented in the following theorem where we write $E_u[\cdot]$ for the conditional expectation $E[\cdot|X_0 = u]$.

Theorem 1. *For $q \geq 0$, the one-sided exit results are*

$$E_u\left[e^{-qT_0^-}1_{\{T_0^-<\infty\}}\right] = Z^q(u) - \frac{q}{\Phi^q}W^q(u),$$

for any $u \geq 0$ and

$$E_u\left[e^{-qT_x^+}1_{\{T_x^+<\infty\}}\right] = e^{-\Phi^q(x-u)},$$

for $0 \leq u \leq x$. The two-sided exist results are

$$E_u\left[e^{-qT_x^+}1_{\{T_x^+<T_0^-\}}\right] = \frac{W^q(u)}{W^q(x)},$$

and

$$E_u\left[e^{-qT_0^-}1_{\{T_0^-<T_x^+\}}\right] = Z^q(u) - Z^q(x)\frac{W^q(u)}{W^q(x)},$$

for $0 \leq u \leq x$.

1.2.2 Spectrally negative Markov additive process

The spectrally negative Markov additive process (MAP) studied in, e.g., Kyprianou and Palmowski [2008] and Ivanovs and Palmowski [2012], is an extension of the spectrally negative Lévy process discussed in the last section.

Define a process $X = \{X_t; t \geq 0\}$ with càdlàg paths on a filtered probability space (Ω, $\mathcal{F}$, $\mathbb{F}$, $\mathbb{P}$) and an irreducible continuous-time Markov process (CTMC) $J = \{J_t;\, t \geq 0\}$ with finite state space $E = \{1, 2, \ldots, n\}$ and infinitesimal generator $\mathbf{Q}$. The bivariate process (X, J) is a MAP if given $\{J_t = i\}$, the pair $(X_{t+s} - X_t, J_{t+s})$ is independent of (X_s, J_s) for all $0 \le s \le t$ and $i \in E$ and has the same law as $(X_s - X_0, J_s)$ given $\{J_0 = i\}$. Let us use $\mathbb{P}_{u,i}$ with $u \in \mathbb{R}$, $i \in E$ to denote the law of (X, J) given $\{X_0 = u, J_0 = i\}$.

When $\{J_t = i\}$, the process X, commonly referred to as the additive component, evolves as some spectrally negative Lévy process X^i and the processes $X^1, X^2, \dots, X^n$ are assumed to be independent. In addition, a transition of J from i to $j \neq i$ triggers a downward jump of X whose (absolute) size is distributed as $P_{i,j} \ge 0$ for $i, j \in E$. Note that only spectrally negative MAPs are considered because X is assumed to have downward jumps only. Also, the special case of a spectrally negative Lévy process is recovered when $n = 1$.

For $z \ge 0$, the Laplace exponent of the spectrally negative Lévy process X^i is given by

$$\psi_i(z) = -a_i z + \frac{1}{2}\sigma_i^2 z^2 + \int_{(-\infty,0)} \left(e^{zx} - 1 - zx1_{\{x>-1\}}\right) \Pi_i(dx),$$

where $E\left[e^{zX_t^i}\right] = e^{\psi_i(z)t}$. Furthermore, the Laplace exponent of a Lévy process given in (1.6) is generalized to an $n \times n$ matrix $\mathbf{F}^q(z)$ called the matrix exponent where

$$E\left[e^{-qt+zX_t}; J_t\right] = e^{\mathbf{F}^q(z)t},$$

and

$$\mathbf{F}^q(z) = \operatorname{diag}\{\psi_i(z)\}_{i=1}^n + \mathbf{Q} \circ \mathbf{G}(z) - q\mathbf{I}, \tag{1.8}$$

where $\mathbf{I}$ is the identify matrix, $(G(z))_{i,j} = E\left[e^{-zP_{i,j}}\right]$ for $i \neq j$ and $(G(z))_{ii} = 1$ for $i, j \in E$. The notation $\mathbf{A} \circ \mathbf{B} = (a_{ij}b_{ij})$ stands for the entry-wise (Hadamard) matrix product.

Results related to the scale matrix and exit problems for X are now given. For $q \ge 0$, the q-scale matrix $\mathbf{W}^q(x)$ is characterized by its Laplace transform,

$$\int_0^\infty e^{-zx}\mathbf{W}^q(x)dx = \mathbf{F}^q(z)^{-1},$$

for $z > \eta = max\{\mathrm{Re}(s) : s \in \mathbb{C},\ \det(F(s)) = 0\}$ and can be seen as the matrix analogue of the scale function given by (1.7). It was shown in Ivanovs and Palmowski [2012] that there exists a unique continuous function $\mathbf{W}^q : [0,\infty) \to \mathbb{R}^{n\times n}$ such that $\mathbf{W}^q(x)$ is invertible for all $x > 0$. Also, the second q-scale matrix is defined as

$$\mathbf{Z}^q(s,u) = e^{su}\left(\mathbf{I} - \int_0^u e^{-sy}\mathbf{W}^q(y)dy\mathbf{F}^q(s)\right).$$

We define the first passage times of X as

$$T_x^{X,+(-)} = \inf\{t \geq 0 : X_t > (<)x\},$$

for $x \in \mathbb{R}$. In [Ivanovs and Palmowski, 2012, Theorem 1], a second representation of the q-scale matrix is given by

$$\mathbf{W}^q(x) = e^{-\mathbf{D}^q x}\mathbf{L}^q(x),$$

where $\mathbf{D}^q$ is a transitional rate matrix of the Markov chain associated with the first passage and $\mathbf{L}^q(x)$ is a matrix associated with the expected occupation times at 0 up to the first passage time over x. More specifically, $\mathbf{D}^q$ satisfies $E\left[e^{-qT_x^{X,+}}; J_{T_x^{X,+}}\right] = e^{\mathbf{D}^q x}$, and $\left(\mathbf{L}^q(x)\right)_{i,j} = E\left[L^q(j,T_x^{X,+})|J_0 = i\right]$ where $L^q(j,t)$ is the limit in $L^2(\mathbb{P})$ of

$$L_\epsilon^q(j,t) = \frac{1}{2\epsilon}\int_0^t e^{-qs}1_{\{|X_s|<\epsilon,\, J_s=j\}}ds,$$

as $\epsilon \downarrow 0$. For a discussion concerning $\mathbf{D}^q$ and its identification through a certain matrix integral equation, see, e.g., Ivanovs and Palmowski [2012] and references therein. Define $\mathbf{L}^q = \lim_{x\to\infty}\mathbf{L}^q(x)$ and it is proven in [Ivanovs and Palmowski, 2012, Lemma 10] that $\mathbf{L}^q$ has finite entries and is invertible unless $\mathbf{Q1} = 0$ (i.e., $q = 0$) and the asymptotic drift $\lim_{t\to\infty} X_t/t = 0$, $\mathbb{P}_{0,i}$-a.s. for all $i \in E$.

The two-sided exit results for X are presented in the following theorem and can be found in [Ivanovs and Palmowski, 2012, Theorem 1 and Corollary 3]. We write $E_u\left[\,\cdot\,; J_T\right]$ to indicate a matrix with (i,j)th element corresponding to $E\left[\,\cdot\, 1_{\{J_T=j\}}\middle|\, U_0 = u, J_0 = i\right]$.

Theorem 2. *For $0 \leq u \leq x$,*

$$E_u\left[e^{-qT_x^{X,+}}1_{\{T_0^{X,-}>T_x^{X,+}\}}; J_{T_x^{X,+}}\right] = \boldsymbol{W}^q(u)\,\boldsymbol{W}^q(x)^{-1},$$

and

$$E_u\left[e^{-qT_0^{X,-}-s\left|U_{T_0^{X,-}}\right|}1_{\{T_0^{X,-}<T_x^{X,+}\}};J_{T_0^{X,-}}\right]=\boldsymbol{Z}^q(s,u)-\boldsymbol{W}^q(u)\,\boldsymbol{W}^q(x)^{-1}\boldsymbol{Z}^q(s,x).$$

More results will be provided as needed in subsequent chapters.

Remark 1. We consider a subclass of the spectrally negative MAPs called the *perturbed Markovian arrival risk process* (MArP) which will become important in Chapter 3. Under this subclass, the bivariate process (X,J) is identical to the general model defined in Section 1.2.2 except that when $\{J_t=i\}$, the process X evolves as $X^i=\{X_t^i;\,t\geq 0\}$, a compound Poisson process with diffusion (rather than the more general spectrally negative Lévy process). More specifically, we define

$$X_t^i=c_it+\sigma_iB_t^i-S_t^i,$$

where $c_i>0$, $\sigma_i>0$, $\{B_t^i;\,t\geq 0\}$ is an independent standard Brownian motion, and S_t^i is a compound Poisson process with Poisson intensity rate $\lambda_i>0$ and i.i.d. jumps distributed as P_i (where $P_i>0$ without loss of generality). The Laplace exponent of X^i is given by

$$\psi_i(z)=c_iz+\frac{\sigma_i^2}{2}z^2+\lambda_i\left(\int_0^\infty e^{-zy}G_{ii}(dy)-1\right),$$

and from (1.8), the matrix exponent is given by

$$\mathbf{F}^q(z)=\mathbf{C}z+\mathbf{\Sigma}z^2-\mathbf{\Lambda}+\mathbf{\Lambda}\circ\mathbf{P}(z)+\mathbf{Q}\circ\mathbf{G}(z)-q\mathbf{I},$$

where $\mathbf{C}=\text{diag}\{c_i\}_{i=1}^n$, $\mathbf{\Sigma}=\text{diag}\{\frac{\sigma_i^2}{2}\}_{i=1}^n$, $\mathbf{\Lambda}=\text{diag}\{\lambda_i\}_{i=1}^n$ and $\mathbf{P}(z)=\text{diag}\{E\left[e^{-zP_i}\right]\}_{i=1}^n$.

Generalized two-sided exit results for this subclass of spectrally negative MAPs can be found in Landriault et al. [2017]. In particular, when $det\left(\mathbf{F}^q(\rho_i)\right)=0$ for $\rho_i>0$ and $\rho_i\neq\rho_j$ for $i\neq j$ $(i,j\in E)$, it can be shown (see, e.g., Li [2015]) that $\mathbf{D}^q$ in (3.5) is explicitly given by

$$\mathbf{D}^q=-\mathbf{\Theta}\text{diag}\{\rho_i\}_{i=1}^n\mathbf{\Theta}^{-1},$$

where $\mathbf{\Theta}=(\boldsymbol{\theta_1},\ldots,\boldsymbol{\theta_n})$ and $\boldsymbol{\theta_i}$ denotes the right-eigenvector associated to the eigenvalue 0 of $\mathbf{F}^q(\rho_i)$, i.e. $\mathbf{F}^q(\rho_i)\mathbf{\Theta}_i=\mathbf{0}$.

1.3 Numerical Laplace inversion techniques

There are many algorithms for the numerical inversion of Laplace transforms (see, e.g., Abate and Valkó [2004] and references therein). In the following, we present three algorithms, namely, the Gaver-Stehfest (GS) method, the Gaver-Wynn-Rho (GWR) algorithm, and the Fourier-cosine (COS) series expansion.

We outline how each of the three algorithms can be used to numerically invert a LT given by $\widetilde{f}(s) = \int_0^\infty e^{-st} f(t)dt$ to obtain an approximant for $f(t)$.

1.3.1 Gaver-Stehfest method

The Gaver-Stehfest (GS) method is one of the most popular inversion algorithms. It is fast and usually gives good results, especially for smooth functions (see, e.g., [Usábel, 1999, Section 4]). The method involves the so-called Gaver functionals which are given by

$$f_k(t) = zk\binom{2k}{k}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\widetilde{f}\left((k+j)\,z\right), \tag{1.9}$$

where $z = ln(2)/t$. The Gaver functionals were developed in Gaver Jr [1966] and are useful in the numerical inversion of Laplace transforms. Since the sequence is logarithmically convergent, an acceleration method is needed (see, e.g. Valkó and Abate [2004]). The GS method utilizes Salzer summation to accelerate convergence, as proposed by Stehfest [1970] and the approximant for $f(t)$ is given by

$$f(t, N_1) = \sum_{k=1}^{N_1} w_k f_k(t),$$

where

$$w_k = (-1)^{k+N_1}\frac{k^{N_1}}{N_1!}\binom{N_1}{k},$$

i.e., the approximant for $f(t)$ involves calculating $f_1(t), \ldots, f_{N_1}(t)$ and taking a weighted sum using weights w_k (see, e.g. Valkó and Abate [2004]). Another expression for the approximant that is commonly used is given by

$$f(t, M) = z \sum_{k=1}^{M} b_k \widetilde{f}(kz), \tag{1.10}$$

where

$$b_k = (-1)^{k+\frac{M}{2}} \sum_{i=\lfloor (k+1)/2 \rfloor}^{min\left(k, \frac{M}{2}\right)} \frac{i^{\frac{M}{2}} (2i)!}{\left(\frac{M}{2} - i\right)! i! (i-1)! (k-i)! (2i-k)!}, \tag{1.11}$$

and $M = 2N_1$, i.e., the approximant for $f(t)$ involves evaluating the LT $\widetilde{f}(s)$ at $M = 2N_1$ values for the Laplace argument and taking a weighted sum using weights b_k (see, e.g., Usábel [1999]). Note that the weights b_k depend only on M and can be easily calculated.

The accuracy of the GS method increases with N_1 (or M) up to a point. The method becomes unstable for greater values of N_1 (without a corresponding increase in arithmetic precision) since (1.10) and (1.11) are prone to rounding-error.

1.3.2 Gaver-Wynn-Rho algorithm

Like the GS method discussed above, the Gaver-Wynn-Rho (GWR) algorithm also involves the Gaver functionals given by (1.9) but instead uses the Wynn rho algorithm to accelerate convergence. The Wynn rho algorithm is given by the following recursive algorithm

$$\rho_{-1}^{(n)} = 0, \qquad \rho_0^{(n)} = f_n(t), \qquad n \geq 0, \tag{1.12}$$

$$\rho_k^{(n)} = \rho_{k-2}^{(n+1)} + \frac{k}{\rho_{k-1}^{(n+1)} - \rho_{k-1}^{(n)}}, \qquad k \geq 1, \tag{1.13}$$

and under the GWR method, the approximant to $f(t)$ is obtained as

$$f(t, N_2) = \rho_{N_2}^{(0)}, \tag{1.14}$$

for even positive integer N_2 (see, e.g., Abate and Valkó [2004]). This method requires the computation of $f_1(t), \ldots, f_{N_2}(t)$ and it is clear from (1.9) that this involves evaluating the LT $\widetilde{f}(s)$ at $2N_2 + 1$ values for the Laplace arguments.

The accuracy of the approximant increases with N_2 up to a point. Thereafter, rounding error accumulation in (1.9) and (1.13) causes a decline in accuracy.

1.3.3 Fourier-cosine series expansion

Another method for numerical inversion of Laplace transforms is known as the COS method which is based on the Fourier-cosine series expansion (see, e.g., Fang and Oosterlee [2008], Zhang [2017], and references therein). This method can be easily used if the corresponding Fourier transform is available. For $\omega \in \mathbb{R}$, the Fourier transform is defined as

$$\begin{aligned} \hat{f}(\omega) &= \int_0^\infty e^{i\omega t} f(t) dt \\ &= \widetilde{f}(-i\omega). \end{aligned}$$

Under the COS method, for sufficiently large a ($a > 0$), the approximant for $f(t)$ is given by

$$f(t, N_3) = {\sum_{k=0}^{N_3-1}}' c_k \, cos\left(\frac{k\pi t}{a}\right), \tag{1.15}$$

where $\sum'$ indicates that the first term in the summation is weighted by one-half, and

$$c_k = \frac{2}{a} Re\left(\hat{f}\left(\frac{k\pi}{a}\right)\right)$$

where $Re(\cdot)$ denotes taking the real part of the argument.

In Zhang [2017], an upper bound for the approximation error

$$\left(\int_0^A \left(f(t) - f(t, N_3)\right)^2 dt\right)^{\frac{1}{2}},$$

where $0 < A \leq a$, was found and is minimized by setting parameter a to

$$a^* = O\left(K^{\frac{1}{2}}\right).$$

An approximant for $F(t) \equiv \int_0^t f(s)ds$ can be found by integrating (1.15) for a given surplus level and is given by

$$F(t, N_3) = \sum_{k=0}^{N_3-1}{}' c_k \, sinc\left(\frac{k\pi t}{a}\right) t, \tag{1.16}$$

where

$$sinc(x) = \begin{cases} \frac{sin(x)}{x}, & x \neq 0, \\ 1, & x = 0. \end{cases}$$

(see [Zhang, 2017, Remark 2]). Equation (1.16) is used later in the thesis to find the finite-time ruin probability of a surplus process.

It is clear from (1.15) (and (1.16)) that this method involves evaluating the LT $\widetilde{f}(s)$ at N_3 values for the Laplace argument.

The accuracy of the approximants (1.15) (and (1.16)) increases with N_3 up to a point but becomes unstable for large values of N_3 due to the accumulation of rounding error.

1.4 Mathematical preliminaries

1.4.1 Dickson-Hipp and Laplace transforms

The present work relies considerably on the Dickson-Hipp transform (see, e.g., Dickson and Hipp [2001], Li and Garrido [2004]). Let f be an integrable real-valued function and $r \in \mathbb{C}$ with $\Re(r) \geq 0$. The Dickson-Hipp transform of f is defined as

$$T_r f(x) \equiv \int_x^\infty e^{-r(u-x)} f(u)du, \quad x \geq 0. \tag{1.17}$$

We note that the Laplace transform is a special case of the Dickson-Hipp transform which is recovered from (1.17) by setting $x = 0$, i.e.,

$$\begin{aligned}\tilde{f}(r) &\equiv \int_0^\infty e^{-ru} f(u) du, \\ &= T_r f(0).\end{aligned} \tag{1.18}$$

Properties of the Dickson-Hipp transform can be found in Li and Garrido [2004]. In particular, the present work makes use of a property related to repeated application of the Dickson-Hipp transform which is given next.

Property 1.4.1. If $r_1, r_2, \ldots, r_k$ are distinct complex numbers with non-negative real part, then repeated application of the Dickson-Hipp transform to f is equal to

$$T_{r_k} \cdots T_{r_2} T_{r_1} f(x) = (-1)^{k-1} \sum_{l=1}^{k} \frac{T_{r_l} f(x)}{\tau_k'(r_l)}, \quad x \geq 0, \tag{1.19}$$

where $\tau_k(r) = \prod_{l=1}^{k} (r - r_l)$. In addition, for $s \in \mathbb{C}$, the corresponding Laplace transform is

$$T_s T_{r_k} \cdots T_{r_2} T_{r_1} f(0) = (-1)^k \left[\frac{\tilde{f}(s)}{\tau_k(s)} - \sum_{l=1}^{k} \frac{\tilde{f}(r_l)}{(s - r_l)\tau_k'(r_l)} \right].$$

Note that when $k = 2$, we have that

$$T_{r_2} T_{r_2} f(x) = \frac{T_{r_1} f(x) - T_{r_2} f(x)}{r_2 - r_1}.$$

1.4.2 Defective renewal equations

Defective renewal equations (see, e.g., Feller [1971]) are employed in risk theory to arrive at analytical results for the Gerber-Shiu function. In this section, we present well-known results for the general solution of defective renewal equations, namely, a closed-form expression for the general solution, exponential bounds, and the Cramér-Lundberg asymptotic result.

Definition 1 (Defective renewal equation)**.** For $F(y) = 1 - \overline{F}(y)$ a distribution function for $y \geq 0$ with $F(0) = 0$, a function $m(x)$ satisfies a defective renewal equation if

$$m(x) = \phi \int_0^x m(x-y)dF(y) + v(x), \qquad x \geq 0, \tag{1.20}$$

where $\phi \in (0,1)$. We assume $v(x) \geq 0$ is a locally bounded function, i.e., $v(x) < \infty$ for $x < \infty$.

An associated compound geometric tail $\overline{G}(y) = 1 - G(y)$ is defined as

$$\overline{G}(y) = \sum_{n=1}^{\infty} (1-\phi)\phi^n \overline{F}^{*n}(y), \qquad y \geq 0,$$

where $\overline{F}^{*n}(y)$ is the tail distribution of the n-fold convolution of F with itself. Note that $G(y)$ has a mass point of $(1-\phi)$ at $y = 0$.

The general solution to (1.20) (see, e.g., Resnick [1992, Section 3.5]) is presented in the following proposition.

Proposition 3. The solution to the defective renewal equation (1.20) is given by

$$m(x) = v(x) + \frac{1}{1-\phi} \int_0^x v(x-y)dG(y), \qquad x \geq 0.$$

Now suppose there exists a constant $R > 0$ satisfying

$$\int_0^\infty e^{Ry} dF(y) = \frac{1}{\phi}, \tag{1.21}$$

then we have the following two propositions which provide exponential bounds (see, e.g., Willmot et al. [2001]) as well as an asymptotic result (see, e.g., Resnick [1992, Section 3.11]) for $m(x)$.

Proposition 4 (Exponential bounds)**.** If $m(x)$ satisfies (1.20) and (1.21) holds, then we have

$$C_L e^{-Rx} \leq m(x) \leq C_U e^{-Rx}, \qquad x \geq 0,$$

where $C_L = \inf_{z \geq 0} \alpha(z)$, $C_L = \sup_{z \geq 0} \alpha(z)$, and

$$\alpha(z) = \frac{e^{Rz} v(z)}{\phi \int_z^\infty e^{Ry} dF(y)}.$$

In the following proposition, note that we use $a(x) \sim b(x)$ to mean $\lim_{x \to \infty} a(x)/b(x) = 1$.

Proposition 5 (Cramér-Lundberg asymptotic result)**.** If $m(x)$ satisfies (1.20) and (1.21) holds, then we have

$$m(x) \sim Ce^{-Rx}, \qquad x \to \infty,$$

where

$$C = \frac{\int_0^\infty e^{Ry} v(y) dy}{\phi \int_0^\infty y e^{Ry} dF(y)},$$

provided $e^{Rx} v(x)$ is directly Riemann integrable.

1.4.3 Rouché's theorem and a modification to Rouché's theorem

The present work employs Rouché's theorem to show that there exists some number of roots to Lundberg's fundamental equation (see, e.g., Titchmarsh [1939]). This allows us to solve for unknown constants expressed in terms of the roots. Rouché's theorem is stated below.

Theorem 6 (Rouché's theorem)**.** *If $f(z)$ and $g(z)$ are analytic inside and on a closed contour D and $|g(z)| < |f(z)|$ on D, then $g(z)$ and $g(z) + f(z)$ have the same number of zeros inside D.*

In some instances, we can succeed in showing $|g(z)| < |f(z)|$ on all points of some closed contour except one (e.g., at $z = 0$). In such cases, the following modification of Rouché's theorem can be used (see, e.g. Klimenok [2001]).

Theorem 7 (A modification of Rouché's theorem). *Let the functions $f(z)$ and $\phi(z)$ be analytic in the open disk $|z| < 1$ and continuous on the boundary $|z| = 1$ and the following relations hold:*

$$|f(z)|_{|z|=1, z\neq 1} > |\phi(z)|_{|z|=1, z\neq 1},$$

$$f(1) = -\phi(1) \neq 0.$$

Let also the functions $f(z)$ and $\phi(z)$ have the derivatives at the point $z = 1$ and the following inequality holds

$$\frac{f'(1) + \phi'(1)}{f(1)} > 0.$$

Then the numbers $N_{f+\phi}$ and N_f of zeros of the functions $f(z)+\phi(z)$ and $f(z)$ in the domain $|z| < 1$ are related as follows

$$N_{f+\phi} = N_f - 1.$$

1.5 Outline of the thesis

In Chapter 2, a simple queue-based claim investigation mechanism (CIM) is proposed and superimposed onto the classical compound Poisson risk model. The standard technique of conditioning on the first claim is utilized to derive a defective renewal equation (DRE) satisfied by the Gerber-Shiu function. Numerical examples illustrate the impact of claim investigation on the ruin probability.

Chapter 3 extends the work in Chapter 2 by proposing a more realistic CIM. A (spectrally negative) Markov Additive Process (MAP) formulation is developed for the insurer's surplus. Four possible CIM generalizations are discussed and their MAP formulations are provided.

Numerical examples are used throughout the chapter to illustrate the impact of various CIMs on the ruin probability.

In Chapter 4, the proposed CIM from Chapter 3 is superimposed onto the aggregate claims process to obtain the aggregate payment process. It is shown that this aggregate payment process has a Markovian Arrival Process (MAP) formulation. A distributional analysis of the future payments due to RBNS claims ("RBNS payments") is then performed under some assumptions for the claim size density and CIM. Numerical examples are used to assess the impact of various CIMs on the VaR and $TVaR$ of the RBNS payments.

In conclusion, Chapter 5 provides final remarks and a discussion of potential avenues for future research.

Chapter 2

A risk model with a simple claim investigation mechanism

2.1 Introduction

Contemporary insurer surplus models do not typically consider claim investigation practices in an explicit manner. Rather, features related to claim investigation have to be a priori embedded in the defined risk model of interest by adjusting the model's parameters. In this chapter, we directly model claim investigation practices by considering a particular queue-based claim investigation mechanism. Investigation practices and strategies are developed to determine the extent of liability and identify ineligible or inflated claims which are crucial components of a sound insurance practice. For example, Juri [2002] discusses a risk process where claims are sums of dependent random variables. Such processes allow for the modelling of (allocated) loss adjustment expenses generated by claim investigations. As a result of investigation practices, claim payments are often modified to reflect investigation findings, in addition to the natural delay accompanying the investigation process. Claim payments may also be delayed by queueing times which we discuss next.

An insurer's investigation strategy is constrained by the number of investigators, time per investigation, and volume of claims, among others. Queues form as claims accumulate and claims are served according to some queueing discipline. The natural existence of queues in this context prompts the inclusion of a queue-based investigation mechanism in surplus modelling. Queueing mechanisms have been an intensive area of research for many decades. A seminal reference in the context of a single server is Cohen [1982]. While there are well-known connections between ruin and queueing problems (e.g., Asmussen and Albrecher [2010] or Sigman [2006]), there has been little study of queues in surplus models. Thus, the present model is a first step in a longer inquiry on the topic, furthering the strong ties between the two research disciplines.

A queue-based investigation mechanism will help to improve the realism of an insurer's cash flow dynamics. Many analogous modifications to improve realism have been made in the ruin theory literature such as the inclusion of dividend payments (e.g., Lin et al. [2003]) and tax payments (e.g., Albrecher and Hipp [2007]). More closely related to claim investigations, claim settlement delays and time dependent payments have also been considered in a ruin context, e.g., claims inflation (e.g., Taylor [1979]), interest rates (e.g., Cai and Dickson [2002]), and IBNR (e.g., Trufin et al. [2011] and references therein). Such features have also been discussed in an aggregate claim context under various assumptions for the number of claims process (e.g., Landriault et al. [2014b] for the nonhomogeneous birth process case, as well as references therein). In this chapter, we use a different approach to model claim settlement delay from the widely studied Chain-Ladder method and its variants (e.g., Hossack et al. [1999, Chapter 10]) where a major concern is the time until payment. The present model is intended for modelling short-term claim liabilities and moreover, its aim is to involve some queueing features (such as congestion) in an insurer's surplus analysis.

The present chapter consists of 4 sections. Section 2.2 provides a description of the proposed queue-based claim investigation mechanism, a mathematical definition of the surplus process and the definition of a generalized Gerber-Shiu function which will be the main subject matter of this chapter. Section 2.3 derives a defective renewal equation (DRE) for this Gerber-Shiu function assuming the investigation time is distributed as a combination (or generalized mixture) of exponentials. Probabilistic interpretations for the DRE components are also provided. In Section 2.4, numerical examples are presented to illustrate the impact of claim investigation strategies on the ruin probability.

2.2 Claim investigation surplus process

For completeness, we first recall the definition of the Cramér-Lundberg surplus process $\underline{U}^* = \{U_t^*\}_{t\geq 0}$, where

$$U_t^* = u + ct - S_t. \tag{2.1}$$

We note that $u\,(u \geq 0)$ is the initial surplus level and $c\,(c > 0)$ is the level premium rate per unit time. The aggregate claim process $\underline{S} = \{S_t\}_{t\geq 0}$ is assumed to be a compound Poisson process, i.e., has Poisson arrivals at rate $\lambda > 0$ and the claim sizes $\{X_1, X_2, \ldots\}$ are assumed to be i.i.d. with common density p. The time to ruin T^* is defined as $T^* = \inf\{t \geq 0 : U_t^* < 0\}$ with $T^* = \infty$ if $U_t^* > 0$ for all $t \geq 0$. A comprehensive treatment of the classical surplus process (2.1) can be found in e.g., Rolski et al. [1999, Chapter 1.6] and Asmussen and Albrecher [2010, Chapter 1.1].

In what follows, we propose to superimpose the following queueing-based claim investigation mechanism onto the surplus process (2.1):

1. The claim investigation mechanism consists of a single investigation unit that allows for at most one claim to be investigated at any given time. Hence, the investigator will

alternate between idle (no claim currently investigated) and busy periods (a claim is under investigation).

2. When the investigator is **idle**, a claim of size $x \geq 0$ will

   - avoid the investigation process with probability $q_0(x) \geq 0$ and be paid immediately in its entirety;
   - undergo investigation with probability $q_1(x) = 1 - q_0(x) \geq 0$ to assess the accuracy/validity of the amount claimed x. The investigation time is assumed to have density h which we refer to as the **investigation time density** with mean μ. Following an investigation period, the amount claimed x will result in a payment w with distribution function (df) $L_x(w)$ $(w \geq 0)$, independent of any other rv's in the risk model.

   The decision to investigate a claim of size x (when the investigator is idle) is independent of anything else in the risk model. Also, investigation times are assumed to be mutually independent and independent of any other rv's in the risk model.

3. When the investigator is **busy**, all claimed amounts will immediately be paid in full. This assumption is consistent with the concept of a single server queue with *balking* (e.g., Gross et al. [2011, Section 1.2.1]) or *finite storage* in queueing theory (see, e.g., Kleinrock [1975, Section 3.6]).

The queueing-based claim investigation mechanism can be represented diagramatically as shown in Figure 2.1.

An example for the df $L_x(w)$ is now given and will later be considered in the numerical examples of Section 2.4. Suppose a claim of size $x \geq 0$ selected for investigation is either determined fraudulent and denied with probability $\varsigma(x)$ or paid in its entirety with probability

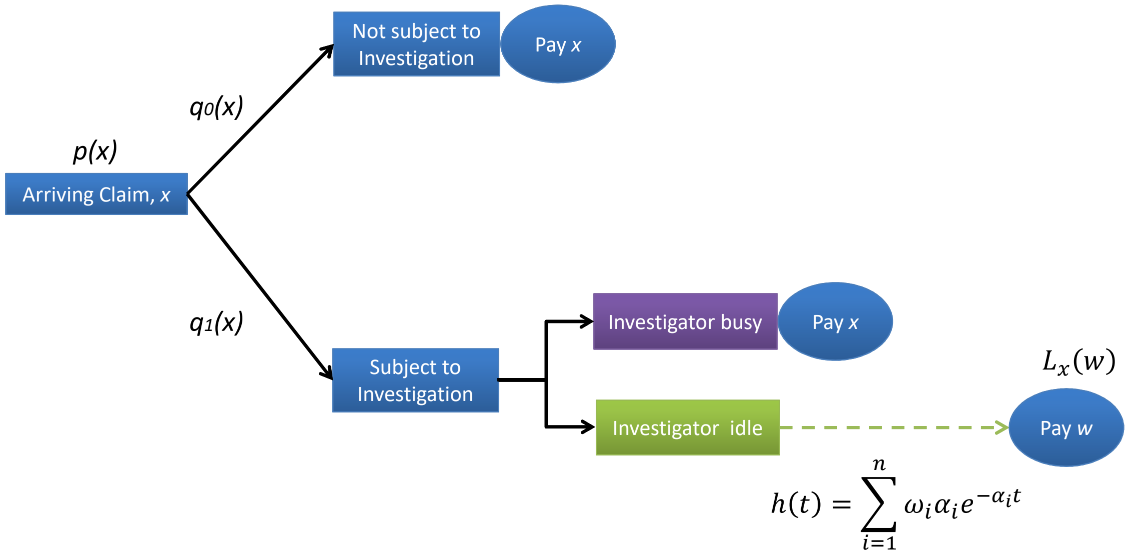

Figure 2.1: Diagramatic representation of the simple queue-based claim investigation mechanism

$1-\varsigma(x)$ (where the decision is made independently of any other rv's in the risk model), then the resulting $L_x(w)$ is of the form

$$L_x(w) = \begin{cases} \varsigma(x), & 0 \leq w < x, \\ 1, & w \geq x. \end{cases} \tag{2.2}$$

Remark 2. Throughout this chapter, the term "claim" refers to the loss amount declared by the policyholder (after appropriately applied coverage modifications, e.g. deductible, policy limit, etc.), while the term "payment" refers to the amount actually paid to the policyholder by the insurer.

We propose to refer to the resulting surplus process as a queueing-based claim investigation (QBCI) surplus process. A formal mathematical definition of this process is now provided. We first decompose the aggregate claim process $\underline{S}$ into two independent compound Poisson processes $\underline{S}^0$ and $\underline{S}^1$, where $\underline{S}^i$ has Poisson arrivals with rate $\lambda_i = \lambda \int_0^\infty p\left(y\right) q_i\left(y\right) dy$, and independent and identically distributed (i.i.d.) positive jumps with density

$$p_i(x) = \frac{p(x)\, q_i(x)}{\int_0^\infty p(y)\, q_i(y)\, dy}, \qquad x > 0.$$

We refer the reader to, e.g., Karlin and Taylor [1981, Chapter 16] for this well-known decomposition.

Given that the investigation unit alternates between idle and busy periods (and we silently assume throughout that the investigator is idle at time 0), we define $\underline{M} = \{M_t\}_{t\geq 0}$ with $M_t = \sup\{k = 1, 2, \ldots : W_1 + \ldots + W_k \leq t\}$ (where $M_t = 0$ for $t < W_1$) to be an alternating renewal process where the inter-arrival times $\{W_j\}_{j=1}^{\infty}$ are mutually independent (and also independent of any other rv's in the risk model). We assume that for j odd, the W_j's are exponentially distributed with mean $1/\lambda_1$, while for j even, W_j is distributed as a combination (or generalized mixture) of exponentials with density

$$h(t) = \sum_{i=1}^{n} \omega_i \alpha_i e^{-\alpha_i t}, \qquad t > 0, \tag{2.3}$$

where $\omega_i \in (-\infty, \infty)$ for $i = 1, 2, \ldots, n$ with $\sum_{i=1}^{n} \omega_i = 1$ and $\alpha_i > 0$ for $i = 1, 2, \ldots, n$. We have $\mu = \sum_{i=1}^{n} \omega_i/\alpha_i$ and without loss of generality, we assume that the α_i's are distinct. Note that the class of combinations of exponentials is dense in the set of continuous distributions defined on $(0, \infty)$ (see Dufresne [2007]).

For convenience, let $\epsilon_t = \int_0^t 1_{\{M_s \text{ is odd}\}} ds$ be the total time that the investigator is busy in $[0, t]$. Hence, the QBCI surplus process $\underline{U} = \{U_t\}_{t\geq 0}$ is defined as

$$U_t = u + ct - Z_t, \tag{2.4}$$

where

$$Z_t = S_t^0 + S_{\epsilon_t}^1 + J_t. \tag{2.5}$$

The process $\underline{J} = \{J_t\}_{t\geq 0}$ in (2.5) is further defined as

$$J_t = \begin{cases} \sum_{j=1}^{\lfloor \frac{M_t}{2} \rfloor} Y_j, & M_t = 2, 3, \ldots, \\ 0, & M_t = 0, 1, \end{cases}$$

where $\lfloor x \rfloor$ denotes the integer part of x, and $\{Y_j\}_{j=1}^{\infty}$ are independent rv's (also independent of any other rv's) with common df

$$K(w) = \int_0^{\infty} L_x(w)\, p_1(x)\, dx, \tag{2.6}$$

and Laplace transform (LT)

$$\widetilde{k}(s) = \int_{[0,\infty)} e^{-sw} K(dw).$$

Note that J_t corresponds to the aggregate payment of all claims that have undergone a complete investigation by time t. It is not difficult to see that $\underline{U}$ reduces to the classical compound Poisson risk process (2.1) when $q_0(x) = 1$ for all $x > 0$.

Since the seminal contribution of Gerber and Shiu [1998], the Gerber-Shiu discounted penalty function has been extensively studied in increasingly complex models and serves as a useful tool to unify the study of risk in surplus processes (see Albrecher et al. [2010]). For the surplus process (2.4), we propose to use $|U_T| + \eta V_T$ as the deficit at ruin (rather than the usual $|U_T|$), where $\eta \in [0, 1]$ and V_t is the (future) payment amount of the claim under investigation at time t (if any) given by

$$V_t = \begin{cases} Y_{\frac{M_t+1}{2}}, & M_t \text{ is odd}, \\ 0, & \text{otherwise}. \end{cases} \tag{2.7}$$

Hence, we assume that if a claim is undergoing investigation when U drops below 0, the insurer remains liable for a fixed portion η of the payment amount. Thus, for $\delta \geq 0$, the Gerber-Shiu analytic tool of interest is defined as

$$m(u) \equiv E\left[e^{-\delta T} w\left(U_{T^-}, |U_T| + \eta V_T\right) 1_{\{T<\infty\}} \,|U_0 = u\right], \tag{2.8}$$

where $T = \inf\{t \geq 0 : U_t < 0\}$ is the time of ruin with $T = \infty$ if $U_t \geq 0$ for all $t \geq 0$ and $w(x, y)$ for $x, y \geq 0$ satisfies mild integrability conditions (namely, we assume that the expectation in (2.8) exists). We therefore focus our risk analysis on the study of $m(u)$ in the next section.

Remark 3. Equation (2.7) indicates that the future payment amount of a claim under investigation has df K and is therefore determined with partial knowledge of the amount that had been claimed. Only that the claim was selected for investigation is known but not the precise amount claimed. Hence, we are assuming that the future payment amount of a claim under investigation is assessed by an external (independent) agent who does not have access to the precise amount claimed. In the event that precise knowledge of the claim amount, say x, is known, then the future payment amount would instead have df L_x.

To conclude this section, we comment on the QBCI surplus process $\underline{U}$ in the case when the investigation time density h is of phase-type form. In particular, in this context, a Markov-additive process (MAP) formulation is possible. To be specific, for a phase-type investigation time density with n states (excluding its absorbing state), we introduce an auxiliary process $\underline{\Theta} = \{\Theta_t\}_{t\geq 0}$, where $\Theta_t = 0$ if no claim is under investigation at time t and $\Theta_t = i$ ($i = 1, 2, \ldots, n$) when a claim is under investigation and the "state" of the phase-type investigation process at time t is i. Under the above assumptions, $\underline{\Theta}$ is a Markov process and more generally, the bivariate process $(\underline{U}, \underline{\Theta})$ can be shown to be a spectrally negative MAP (see Ivanovs and Palmowski [2012] and references therein for more detail on MAPs). This MAP formulation allows for $m(u)$ to be analyzed using standard MAP machinery. However, additional analysis would still be required to arrive at results of a similar nature to those of Section 2.3. Furthermore, the present approach utilizing the dense class of combinations of exponentials (2.3) can handle various distributions of both non-phase-type (e.g., the logbeta distribution of Section 2.4) and phase-type form (e.g., mixtures of exponentials).

2.3 Gerber-Shiu analysis

In this section, we derive a defective renewal equation (DRE) satisfied by the Gerber-Shiu function (2.8). A discussion of the components of the DRE follows. Note that the existence of this DRE representation for m is not immediate given the superimposition of a queueing mechanism.

2.3.1 A defective renewal equation

To first derive an integral equation satisfied by $m(u)$, we condition on the time (t) and amount (x) of the first claim and whether or not it is investigated. We note that the surplus immediately before payment of the claim is $u + ct$. If the first claim is:

(a) not selected for investigation with probability $q_0(x)$, then we further condition on whether or not the claim causes ruin. Suppose we have that:

   (i) $x > u + ct$, then the claim causes ruin. Therefore, the surplus immediately prior to ruin is $U_{T^-} = u + ct$ and the deficit at ruin $|U_T| + \eta V_T = |U_T| = x - u - ct$. Note that $V_T = 0$ since there are no claims under investigation at the time or ruin. Hence, the conditional Gerber-Shiu function in this case is $e^{-\delta t} w(u+ct, x-u-ct)$.

   (ii) $x \leq u + ct$, then the claim does not cause ruin and the surplus is reduced to $u + ct - x \geq 0$. The surplus process renews with initial surplus $u + ct - x$ and the conditional Gerber-Shiu function in this case is $e^{-\delta t} m(u + ct - x)$.

(b) selected for investigation with probability $q_1(x)$, then the claim is assumed to begin investigation and the conditional Gerber-Shiu function is

$$e^{-\delta t} E\left[e^{-\delta(T-W_1)} w\left(U_{T^-}, |U_T| + \eta V_T\right) 1_{\{T<\infty\}} \middle| U_{W_1} = u + ct, W_1 < T, X_1 = x \right],$$

where we point out the above expectation corresponds to the Gerber-Shiu counterpart to (2.8) if the QBCI surplus process is assumed to begin investigation at time 0 for a claim of size x.

Given that the time of the first claim is exponentially distributed with rate λ and the amount of the first claim has density function $p(x)$, it follows from the law of total probability that

$$m(u) = \int_0^\infty \lambda e^{-\lambda t} e^{-\delta t} \times \left(\int_0^{u+ct} m(u+ct-x) q_0(x) p(x) dx + \int_{u+ct}^\infty w(u+ct, x-u-ct) q_0(x) p(x) dx dt + \int_0^\infty E\left[e^{-\delta(T-W_1)} w\left(U_{T^-}, |U_T| + \eta V_T\right) 1_{\{T<\infty\}} \middle| U_{W_1} = u+ct, W_1 < T, X_1 = x \right] q_1(x) p(x) dx \right) dt.$$

We can more succinctly write the above integral equation using

$$m(u) = \int_0^\infty e^{-(\lambda+\delta)t} \left(\lambda_0 \sigma(u+ct) + \lambda_1 \phi(u+ct)\right) dt, \tag{2.9}$$

by defining

$$\sigma(z) = \int_0^z m(z-x) p_0(x) dx + \beta(z), \tag{2.10}$$

where

$$\beta(z) = \int_z^\infty w(z, x-z) p_0(x) dx,$$

and

$$\begin{aligned} \phi(u) &= \int_0^\infty E\left[e^{-\delta(T-W_1)} w\left(U_{T^-}, |U_T| + \eta V_T\right) 1_{\{T<\infty\}} |U_{W_1} = u, W_1 < T, X_1 = x \right] p_1(x) dx \\ &= E\left[e^{-\delta(T-W_1)} w\left(U_{T^-}, |U_T| + \eta V_T\right) 1_{\{T<\infty\}} |U_{W_1} = u, W_1 < T \right]. \end{aligned}$$

Using the Dickson-Hipp operator T_d defined as $T_d f(x) \equiv \int_x^\infty e^{-d(y-x)} f(y) dy$ for $\mathrm{Re}\,(d) \geq 0$, (2.9) can be written as

$$m(u) = \frac{1}{c} T_{\frac{\lambda+\delta}{c}} \left(\lambda_0 \sigma(u) + \lambda_1 \phi(u)\right). \tag{2.11}$$

From Property 2 related to repeated application of the Dickson-Hipp operator (see Li and Garrido [2004, p. 393]), the LT of (2.11) is

$$\begin{aligned} \widetilde{m}(s) &\equiv \int_0^\infty e^{-su} m(u) du \\ &= \frac{1}{c} T_{s, \frac{\lambda+\delta}{c}} \left(\lambda_0 \sigma(0) + \lambda_1 \phi(0)\right) \\ &= \frac{K_1 - \lambda_0 \widetilde{\sigma}(s) - \lambda_1 \widetilde{\phi}(s)}{cs - \lambda - \delta}, \end{aligned} \tag{2.12}$$

where $K_1 = cm(0)$ and $T_{d_1,d_2,\ldots,d_n} f(x) \equiv T_{d_1} T_{d_2} \cdots T_{d_n} f(x)$ for $n = 1, 2, \ldots$. In what follows, we normally add the tilde symbol $\sim$ over a quantity to denote its corresponding Laplace transform.

From (2.12), we observe that an explicit expression for the LT $\widetilde{\phi}$ is needed. When a claim is being handled for investigation, one can condition on whether ruin or the undergoing investigation finishes first. This results in

$$\begin{aligned} \phi(z) &= \mathrm{E}\left[e^{-\delta(T-W_1)} w(U_{T^-}, |U_T| + \eta V_T) 1_{\{W_1+W_2>T\}} \middle| U_{W_1} = z, W_1 < T \right] \\ &\quad + \int_0^\infty \mathrm{E}\left[e^{-\delta W_2} 1_{\left\{U_{(W_1+W_2)^-} \in dl,\, W_1+W_2<T\right\}} \middle| U_{W_1} = z, W_1 < T \right] \mathrm{E}\left[m\left(l - Y\right) \right], \end{aligned} \tag{2.13}$$

where Y has df K defined in (2.6), and

$$\mathrm{E}\left[m\left(l - Y\right) \right] = \int_{[0,l]} m(l-y) K(dy) + \int_{(l,\infty)} w(l, y-l) K(dy).$$

Equation (2.13) can be further characterized by making connections with existing results in the Cramér-Lundberg surplus process. For the second term on the RHS of (2.13), using Kyprianou [2006, Corollary 8.8], it follows that

$$\begin{aligned} &\mathrm{E}\left[e^{-\delta W_2} 1_{\left\{U_{(W_1+W_2)^-} \in dl,\, W_1+W_2<T\right\}} \middle| U_{W_1} = z, W_1 < T \right] \\ &= \sum_{i=1}^n \omega_i \alpha_i \left(e^{-\rho_i l} v_{\alpha_i+\delta}(z) - v_{\alpha_i+\delta}(z-l) 1_{\{z>l\}} \right) dl, \end{aligned} \tag{2.14}$$

where the scale function $v_{\alpha_i+\delta}$ is defined through its LT

$$\widetilde{v}_{\alpha_i+\delta}(s) = \frac{1}{cs + \lambda \widetilde{p}(s) - (\lambda + \delta + \alpha_i)}, \qquad s > \rho_i, \tag{2.15}$$

and $\rho_i = \rho(\alpha_i + \delta)$ $(i = 1, 2, \ldots, n)$ is the real and strictly positive root of $\widetilde{v}_{\alpha_i+\delta}^{-1}(s)$. By analytic continuation, we assume that the functional form of (2.15) holds for $s < \rho_i$ (obviously, $\widetilde{v}_{\alpha_i+\delta}(s)$ is not a LT on this domain). From Gerber and Shiu [1998, Equation (2.52)], we have

$$\widetilde{v}_{\alpha_i+\delta}(s) = \frac{\widetilde{g}_{\alpha_i+\delta}\left(s\right)}{s - \rho_i}, \tag{2.16}$$

where

$$\widetilde{g}_{\alpha_i+\delta}(s) \;=\; \frac{1}{c-\lambda T_{s,\rho_i}p(0)} = \frac{1}{c} + \int_0^\infty e^{-sx} g^+_{\alpha_i+\delta}(x)dx,$$

with

$$g^+_{\alpha_i+\delta}(x) = c^{-1}\sum_{j=1}^{\infty}\left(\frac{\lambda}{c}\right)^j \left(T_{\rho_i}p\right)^{*j}(x),$$

and $(T_{\rho_i}p)^{*j}$ is the j-fold convolution of $T_{\rho_i}p$ with itself.

Now, for the first term on the RHS of (2.13), given that V_T defined in (2.7) has df K and the payment amounts were assumed to be independent of any other rv's in the risk model, it follows that

$$\begin{aligned}\gamma(z) \;&:=\; \mathrm{E}\left[e^{-\delta(T-W_1)}w(U_{T^-},|U_T|+\eta V_T)1_{\{W_1+W_2>T\}}\middle| U_{W_1}=z, W_1<T\right] \\ &=\; \int_{[0,\infty)} \mathrm{E}\left[e^{-\delta(T-W_1)}w(U_{T^-},|U_T|+\eta x)1_{\{W_1+W_2>T\}}|U_{W_1}=z, W_1<T\right]K(dx).\end{aligned}$$

Furthermore, given that $U_{W_1}=z$ and $T>W_1$, the triplet $(T-W_1, U_{T^-}, |U_T|)$ is distributed as $(T^*, U^*_{T^*-}, |U^*_{T^*}|)$ with $U^*_0=z$ when $W_1+W_2>T$. Together with (2.3), one finds that

$$\begin{aligned}\gamma(z) \;&=\; \int_{[0,\infty)} \mathrm{E}\left[e^{-\delta T^*}w(U^*_{T^*-},|U^*_{T^*}|+\eta x)1_{\{W_2>T^*\}}|U^*_0=z\right]K(dx) \\ &=\; \sum_{i=1}^{n}\omega_i\int_{[0,\infty)} \mathrm{E}\left[e^{-(\alpha_i+\delta)T^*}w(U^*_{T^*-},|U^*_{T^*}|+\eta x)1_{\{T^*<\infty\}}|U^*_0=z\right]K(dx), \quad (2.17)\end{aligned}$$

which corresponds to a summation of Gerber-Shiu-type functions in the Cramér-Lundberg surplus process $\underline{U}^*$ (e.g., Gerber and Shiu [1998]). Thus,

$$\gamma(z) = \sum_{i=1}^{n}\omega_i\int_{[0,\infty)}\int_0^\infty\int_0^\infty w(a,y+\eta x)h^*_{\alpha_i+\delta}(a,y|z)dadyK(dx),$$

where

$$h^*_{\alpha_i+\delta}(a,y|z) = \frac{\lambda}{c}\left(e^{-\rho_i(a-z)}1_{\{a>z\}} + \int_0^{min\{a,z\}} e^{-\rho_i(a-t)}cg^+_{\alpha_i+\delta}(z-t)dt\right)p(a+y),$$

is known as the discounted density of the surplus immediately prior to ruin and the deficit at ruin (e.g., Landriault and Willmot [2009]).

Substituting (2.17) and (2.14) into (2.13) and taking LT on both sides results in

$$\widetilde{\phi}(s) = \widetilde{\gamma}(s) + \sum_{i=1}^{n} \omega_i \alpha_i \widetilde{v}_{\alpha_i+\delta}(s) \left(\widetilde{m}(\rho_i)\widetilde{k}(\rho_i) - \widetilde{m}(s)\widetilde{k}(s) \right) + \widetilde{b}(s), \tag{2.18}$$

where $\widetilde{b}(s)$ is the LT of

$$b(z) = \sum_{i=1}^{n} \omega_i \alpha_i \int_0^{\infty} \left(e^{-\rho_i l} v_{\alpha_i+\delta}(z) - v_{\alpha_i+\delta}(z-l) 1_{\{z>l\}} \right) \int_{(l,\infty)} w(l, y-l) K(dy)\, dl, \tag{2.19}$$

for $z \geq 0$. Substituting (2.10) and (2.18) into (2.12) yields

$$\left(cs - \lambda - \delta + \lambda_0 \widetilde{p}_0(s) - \lambda_1 \sum_{i=1}^{n} \omega_i \alpha_i \widetilde{v}_{\alpha_i+\delta}(s) \widetilde{k}(s) \right) \widetilde{m}(s) = K_1 - \widetilde{\tau}(s) - \sum_{i=1}^{n} K_{2,i} \widetilde{v}_{\alpha_i+\delta}(s), \tag{2.20}$$

where $K_{2,i} = \lambda_1 \omega_i \alpha_i \widetilde{m}(\rho_i) \widetilde{k}(\rho_i)$ $(i = 1, 2, \ldots, n)$, and

$$\widetilde{\tau}(s) = \lambda_0 \widetilde{\beta}(s) + \lambda_1 \left(\widetilde{\gamma}(s) + \widetilde{b}(s) \right). \tag{2.21}$$

Using (2.16), we multiply both sides of (2.20) by $\prod_{k=1}^{n} (\rho_k - s)$ which leads to

$$L(s)\widetilde{m}(s) \quad = \quad K_1 \prod_{k=1}^{n} (\rho_k - s) - \widetilde{\chi}(s), \tag{2.22}$$

where

$$L(s) \quad = \quad \prod_{k=1}^{n} (\rho_k - s)\,(cs - \lambda - \delta) + \widetilde{\zeta}(s), \tag{2.23}$$

with

$$\widetilde{\zeta}(s) = \lambda_0 \left\{ \prod_{k=1}^{n} (\rho_k - s) \right\} \widetilde{p}_0(s) + \lambda_1 \sum_{i=1}^{n} \omega_i \alpha_i \left\{ \prod_{k=1, k\neq i}^{n} (\rho_k - s) \right\} \widetilde{g}_{\alpha_i+\delta}(s) \widetilde{k}(s), \tag{2.24}$$

and

$$\widetilde{\chi}(s) = \left\{ \prod_{k=1}^{n} (\rho_k - s) \right\} \widetilde{\tau}(s) - \sum_{i=1}^{n} K_{2,i} \left\{ \prod_{k=1, k\neq i}^{n} (\rho_k - s) \right\} \widetilde{g}_{\alpha_i+\delta}(s). \tag{2.25}$$

Clearly from (2.22), $\widetilde{m}(s)$ depends on the unknown constants K_1 and $K_{2,i}$ $(i = 1, 2, \ldots, n)$ whose numerical values can be obtained with the help of Proposition 1. This culminates in the result of Theorem 8 where it is shown that m satisfies a DRE which provides several well-known analytical results for its solution (e.g., Resnick [1992, Section 3.5]).

Proposition 1. *For $\delta > 0$, $L(s)$ has exactly $n+1$ zeros, say $r_j = r_j(\delta)$ $(j = 1, 2, \ldots, n+1)$, with positive real part.*

Proof: As in other applications in ruin theory, we prove this result via Rouché's theorem. Let

$$L(s) = \prod_{k=1}^{n} (\rho_k - s)\,(h_1(s) - h_2(s))\,,$$

where

$$h_1(s) = cs - \lambda - \delta + \lambda_0 \widetilde{p}_0(s),$$

and

$$h_2(s) = \lambda_1 \left(\sum_{i=1}^{n} \omega_i \alpha_i \widetilde{v}_{\alpha_i+\delta}(s) \right) \widetilde{k}(s).$$

We define a contour C_r which consists of the semi-circle of radius r centered at 0 with positive real numbers which is closed by the connecting part of the imaginary axis. Given that $|h_1(s)|$ behaves asymptotically as a linear function in r, it is clear that $|h_1(s)| > |h_2(s)|$ on the semi-circle of C_r for r sufficiently large.

Using (2.15), we have

$$h_2(s) = -\lambda_1 \left(\sum_{i=1}^{n} \omega_i \frac{\alpha_i}{\alpha_i + \lambda + \delta - cs - \lambda \widetilde{p}(s)} \right) \widetilde{k}(s).$$

Then for any point on the imaginary axis

$$|h_2(s)| \;\leq\; \lambda_1 \left| \widetilde{h}\left(\lambda + \delta - cs - \lambda \widetilde{p}(s)\right) \right|,$$

where we recall that $\widetilde{h}$ is the LT of the investigation time density (2.3). Furthermore, since $\mathrm{Re}\{\lambda + \delta - cs - \lambda \widetilde{p}(s)\} \geq \delta$, we have that $|h_2(s)| < \lambda_1$. We also have

$$\begin{aligned} |h_1(s)| &\geq |\lambda + \delta - cs| - \lambda_0 \left|\widetilde{p}_0(s)\right| \\ &\geq \lambda + \delta - \lambda_0 \\ &= \lambda_1 + \delta \\ &> |h_2(s)|\,. \end{aligned}$$

Clearly, $\prod_{k=1}^{n}(\rho_k - s)\,h_1(s)$ and $\prod_{k=1}^{n}(\rho_k - s)\,h_2(s)$ are analytic inside and on C_r. Therefore, we can apply Rouché's theorem. Given it is well known that $h_1(s)$ has only one strictly positive and real zero (e.g., Gerber and Shiu [1998, Figure 2]), it follows that $\prod_{k=1}^{n}(\rho_k - s)\,h_1(s)$ and $L(s) = \prod_{k=1}^{n}(\rho_k - s)\,(h_1(s) - h_2(s))$ have the same number of zeros with strictly positive real part, namely $n+1$. $\square$

For $\delta = 0$, it can be shown that there also exist $n+1$ non-negative solutions to $L(s) = 0$ under an additional condition related to process drift. Indeed, from random walk theory (e.g., Rolski et al. [1999, Theorem 6.3.1]), this condition is given by $\mathrm{E}\left[U_{W_1+W_2} - u\right] > 0$, or equivalently

$$c > \frac{\frac{\lambda_0 \mathrm{E}\left[X^0\right]}{\lambda_1} + \lambda\mu\mathrm{E}\left[X\right] + \mathrm{E}\left[Y\right]}{\frac{1}{\lambda_1} + \mu}, \tag{2.26}$$

where X^0, X, and Y have LT $\widetilde{p}_0$, $\widetilde{p}$, and $\widetilde{k}$, respectively.

Remark 4. From a simple analysis of the function $L(s)$, we deduce that if the investigation time density (2.3) is of a mixture form (i.e. $\omega_i > 0$ $(i = 1, 2, \ldots, n)$) then the zeros $\{r_j\}_{j=1}^{n+1}$ are real and interweave with the values $\{\rho_i\}_{i=1}^{n}$, i.e. $r_1 < \rho_1 < r_2 < \rho_2 < \cdots < \rho_{n-1} < r_n < \rho_n < r_{n+1}$. In general, the ordering is unclear and we henceforth assume that the r_j's are distinct and $r_j \neq \rho_i$ for any $j = 1, 2, \ldots, n+1$ and $i = 1, 2, \ldots, n$.

In the following theorem, we show that m satisfies a DRE. For simplicity, we henceforth assume that $\widetilde{k}$ is of the form

$$\widetilde{k}(s) = K(0) + \int_0^\infty e^{-sx} k^+(x)dx,$$

i.e. the df K has density k^+ on $(0, \infty)$. The analysis can easily be extended to the more general case. For convenience, we define $(g_{\alpha_i+\delta} * k)^+(y)$ as having LT $\widetilde{g}_{\alpha_i+\delta}(s)\widetilde{k}(s) - K(0)/c$, i.e.

$$(g_{\alpha_i+\delta} * k)^+(y) = g^+_{\alpha_i+\delta}(y)K(0) + \frac{k^+(y)}{c} + \int_0^y g^+_{\alpha_i+\delta}(y-x)k^+(x)dx, \qquad y > 0.$$

Theorem 8. *For $\delta > 0$ or for $\delta = 0$ and (2.26) holds, the Gerber-Shiu function $m(u)$ satisfies the following defective renewal equation*

$$m(u) = \int_0^u m(u-y)a(y)dy + \xi(u), \tag{2.27}$$

where

$$a(y) = \sum_{l=1}^{n+1} \frac{D_l}{c} \left(\lambda_0 T_{r_l} p_0(y) + \lambda_1 \sum_{i=1}^{n} \omega_i \alpha_i \frac{T_{r_l}(g_{\alpha_i+\delta} * k)^+(y)}{\rho_i - r_l} \right), \tag{2.28}$$

$$\xi(u) = \sum_{l=1}^{n+1} \frac{D_l}{c} \left(T_{r_l}\tau(u) - \sum_{i=1}^{n} K_{2,i} \frac{T_{r_l} g^+_{\alpha_i+\delta}(u)}{\rho_i - r_l} \right), \tag{2.29}$$

with

$$D_l = \frac{\prod_{j=1}^{n}(r_l - \rho_j)}{\prod_{j=1, j\neq l}^{n+1}(r_l - r_j)}, \tag{2.30}$$

and the constants $K_{2,i}$ are as given in (2.31).

Proof: From Proposition 1, one concludes that the RHS of Equation (2.20) also has zeros $\{r_j\}_{j=1}^{n+1}$ and thus, K_1 and $\{K_{2,i}\}_{i=1}^{n}$ satisfy the following system of linear equations:

$$K_1 - \widetilde{\tau}(r_j) - \sum_{i=1}^{n} K_{2,i}\widetilde{v}_{\alpha_i+\delta}(r_j) = 0,$$

for $j = 1, 2, \ldots, n+1$. In matrix form, the constants $\overline{\boldsymbol{K}}^T = (K_1, K_{2,1}, K_{2,2}, \ldots, K_{2,n})$ are given by

$$\overline{\boldsymbol{K}} = \boldsymbol{A}^{-1}\overline{\boldsymbol{\tau}}, \tag{2.31}$$

where $\overline{\boldsymbol{\tau}}^T = (\widetilde{\tau}(r_1), \widetilde{\tau}(r_2), \ldots, \widetilde{\tau}(r_{n+1}))$ and

$$\boldsymbol{A} = \begin{bmatrix} 1 & -\widetilde{v}_{\alpha_1+\delta}(r_1) & -\widetilde{v}_{\alpha_2+\delta}(r_1) & \cdots & -\widetilde{v}_{\alpha_n+\delta}(r_1) \\ 1 & -\widetilde{v}_{\alpha_1+\delta}(r_2) & -\widetilde{v}_{\alpha_2+\delta}(r_2) & \cdots & -\widetilde{v}_{\alpha_n+\delta}(r_2) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & -\widetilde{v}_{\alpha_1+\delta}(r_{n+1}) & -\widetilde{v}_{\alpha_2+\delta}(r_{n+1}) & \cdots & -\widetilde{v}_{\alpha_n+\delta}(r_{n+1}) \end{bmatrix},$$

is assumed to be invertible.

Removing the zeros r_j for $j = 1, 2, \ldots, n+1$ on both sides of (2.22), we obtain

$$\frac{L(s)}{\prod_{j=1}^{n+1}(s-r_j)}\widetilde{m}(s) = \frac{K_1\prod_{k=1}^{n}(\rho_k - s) - \widetilde{\chi}(s)}{\prod_{j=1}^{n+1}(s-r_j)}. \tag{2.32}$$

Using (2.23), partial fractions and Property 6 of the Dickson-Hipp operator (see Li and Garrido [2004, p. 394]), it is not difficult to show that

$$\frac{L(s)}{\prod_{j=1}^{n+1}(s-r_j)} = (-1)^n\left(c - T_{s,r_1,r_2,\ldots,r_{n+1}}\zeta(0)\right). \tag{2.33}$$

Again using partial fractions and Property 6 of the Dickson-Hipp operator, it can be shown that for $\widetilde{f}_1$ a LT and $\widetilde{f}_2(s) = \prod_{k=1}^{n}(\rho_k - s)\,\widetilde{f}_1(s)$, we have

$$T_{s,r_1,r_2,\ldots,r_{n+1}}f_2(0) = \sum_{l=1}^{n+1} D_l\, T_{s,r_l}f_1(0). \tag{2.34}$$

Similarly if $\widetilde{f}_{1,i}$ is a LT and $\widetilde{f}_{2,i}(s) = \prod_{k=1,k\neq i}^{n}(\rho_k - s)\,\widetilde{f}_{1,i}(s)$, we have

$$T_{s,r_1,r_2,\ldots,r_{n+1}}f_{2,i}(0) = \sum_{l=1}^{n+1} D_l\,\frac{T_{s,r_l}f_{1,i}(0)}{\rho_i - r_l}. \tag{2.35}$$

Using Equations (2.34), (2.35) and (2.24), it follows that

$$\begin{aligned} T_{s,r_1,r_2,\ldots,r_{n+1}}\zeta(0) &= \sum_{l=1}^{n+1} D_l\left(\lambda_0 T_{s,r_l}p_0(0) + \lambda_1\sum_{i=1}^{n}\omega_i\alpha_i\frac{T_{s,r_l}(g_{\alpha_i+\delta} * k)^+(0)}{\rho_i - r_l}\right) \\ &= c\,\widetilde{a}(s), \end{aligned}$$

which corresponds to the LT of (2.28) multiplied by c. Similarly, we conclude that

$$\frac{K_1\prod_{k=1}^{n}(\rho_k - s) - \widetilde{\chi}(s)}{\prod_{j=1}^{n+1}(s-r_j)} = (-1)^n T_{s,r_1,r_2,\ldots,r_{n+1}}\chi(0). \tag{2.36}$$

Using Equations (2.25), (2.34) and (2.35) again, we obtain $T_{s,r_1,r_2,\ldots,r_{n+1}}\chi(0) = c\widetilde{\xi}(s)$ given by the LT of (2.29).

Substituting (2.33) and (2.36) into (2.32) yields

$$\widetilde{m}(s) = \widetilde{m}(s)\widetilde{a}\,(s) + \widetilde{\xi}(s),$$

whose LT inversion results in (2.27). It remains to show that $a(y)$ for $y \geq 0$ is a defective density when either $\delta > 0$ or $\delta = 0$ and (2.26) holds. This is proven in Remark 5 of Section 2.3.2. □

We point out that the numerical examples of Section 2.4 focus on the ruin probability $\psi(u)$, a special case of (2.8) with $w(x, y) = 1$ and $\delta = 0$. No particular simplification arises to the DRE representation (2.27) for the special case $\psi(u)$.

The following corollary provides the special case of Theorem 8 when the investigation time density is exponentially distributed.

Corollary 1. *For exponential investigation times with mean $1/\alpha_1$, the Gerber-Shiu function $m(u)$ satisfies the following DRE (2.27) with*

$$a(y) = \frac{\lambda_0}{c}\left(\frac{\rho_1 - r_1}{r_2 - r_1}T_{r_1}p_0\,(y) + \frac{r_2 - \rho_1}{r_2 - r_1}T_{r_2}p_0\,(y)\right) + \frac{\lambda_1\alpha_1}{c}T_{r_1,r_2}(g_{\alpha_1+\delta} * k)^+(y),$$

and

$$\xi(u) = \frac{1}{c}\left(\frac{\rho_1 - r_1}{r_2 - r_1}T_{r_1}\tau(u) + \frac{r_2 - \rho_1}{r_2 - r_1}T_{r_2}\tau(u) - K_{2,1}T_{r_1,r_2}g^+_{\alpha_1+\delta}(u)\right),$$

where $K_{2,1} = -T_{r_1,r_2}\tau(0)/T_{r_1,r_2}v_{\alpha_1+\delta}(0)$ and τ can be simply obtained from (2.21), (2.19), and (2.17) with $n = 1$.

2.3.2 Defective density $a(y)$

This section provides a probabilistic interpretation for $a(y)$ as the discounted (defective) density of the first drop in surplus when no claim is being investigated.

Theorem 9. *For $\delta > 0$ or for $\delta = 0$ and (2.26) holds, the function $a(y)$ in the DRE (2.27) can be expressed as*

$$a(y) = \frac{d}{dy} E\left[e^{-\delta T^\dagger} 1_{\{|U_{T^\dagger}| \leq y, T^\dagger < \infty\}} \middle| U_0 = 0 \right],$$

where $T^\dagger = inf\{t \geq 0 : \ U_t < 0$ and M_t is even$\}$ with $T^\dagger = \infty$ if $U_t \geq 0$ whenever M_t is even for all $t \geq 0$.

Proof: From the uniqueness of Laplace transforms, it is equivalent to show $\widetilde{a}\,(z) = m^\dagger(0)$ where

$$m^\dagger\,(u) \equiv E\left[e^{-\delta T^\dagger - z|U_{T^\dagger}|} 1_{\{T^\dagger < \infty\}} \middle| U_0 = u \right].$$

The following proof employs an analysis similar to that of Section 2.3.1. As a result, only the crucial steps of the proof will be sketched. Similar to Equation (2.12), we have

$$\widetilde{m}^\dagger(s) = \frac{cm^\dagger(0) - \lambda_0 \widetilde{\sigma}^\dagger(s) - \lambda_1 \widetilde{\phi}^\dagger(s)}{cs - \lambda - \delta}, \tag{2.37}$$

where

$$\widetilde{\sigma}^\dagger(s) = \widetilde{m}^\dagger(s)\widetilde{p}_0(s) + T_{s,z}p_0(0), \tag{2.38}$$

and

$$\phi^\dagger(u) = E\left[e^{-\delta\left(T^\dagger - W_1\right) - z|U_{T^\dagger}|} 1_{\{T^\dagger < \infty\}} \middle| U_{W_1} = u, W_1 < T^\dagger \right].$$

To obtain an expression for $m^\dagger(0)$, we shall further analyze the LT $\widetilde{\phi}^\dagger$. Conditioning on the surplus level at the end of the current investigation period (but not including the investigated claim payment), it follows that

$$\begin{aligned}\phi^{\dagger}(u) &= \int_{-\infty}^{\infty} \mathrm{E}\left[e^{-\delta W_2} 1_{\left\{U_{(W_1+W_2)^-}\in dl\right\}} \middle| U_{W_1}=u, W_1<T^{\dagger}\right] \mathrm{E}\left[m^{\dagger}\left(l-Y\right)\right] \\ &= \int_{-\infty}^{\infty} \mathrm{E}\left[e^{-\delta W_2} 1_{\left\{U^*_{W_2^-}\in dl\right\}} \middle| U_0^*=u\right] \mathrm{E}\left[m^{\dagger}\left(l-Y\right)\right], \end{aligned} \tag{2.39}$$

where

$$\mathrm{E}\left[m^{\dagger}\left(l-Y\right)\right] = \begin{cases} m^{\dagger}(l)K(0)+\int_0^l m^{\dagger}(l-y)k^{+}(y)dy + T_z k^{+}(l), & l\geq 0, \\ e^{zl}\widetilde{k}(z), & l<0. \end{cases} \tag{2.40}$$

From Kyprianou [2006, Corollary 8.9], one deduces that

$$\mathrm{E}\left[e^{-\delta W_2} 1_{\left\{U^*_{W_2^-}\in dl\right\}} \middle| U_0^*=0\right] \cong \sum_{i=1}^{n}\omega_i\alpha_i\left(\widetilde{g}_{\alpha_i+\delta}(\rho_i)e^{-\rho_i l} - v_{\alpha_i+\delta}\left(-l\right)1_{\{-l>0\}}\right)dl,$$

which implies that (2.39) becomes

$$\begin{aligned}\phi^{\dagger}(u) &= \sum_{i=1}^{n}\omega_i\alpha_i\int_0^{\infty}\left(\widetilde{g}_{\alpha_i+\delta}(\rho_i)e^{-\rho_i(l-u)} - v_{\alpha_i+\delta}\left(u-l\right)1_{\{l<u\}}\right)\mathrm{E}\left[m^{\dagger}\left(l-Y\right)\right]dl \\ &+ \sum_{i=1}^{n}\omega_i\alpha_i\int_u^{\infty}\left(\widetilde{g}_{\alpha_i+\delta}(\rho_i)e^{\rho_i l} - v_{\alpha_i+\delta}\left(l\right)\right)\mathrm{E}\left[m^{\dagger}\left(u-l-Y\right)\right]dl. \end{aligned} \tag{2.41}$$

Substituting (2.40), (2.41) and (2.38) into Equation (2.37) followed by algebraic manipulations, one obtains

$$L(s)\widetilde{m}^{\dagger}(s) = \prod_{k=1}^{n}(s-\rho_k)\left(cm^{\dagger}(0)-\sum_{i=1}^{n}\frac{C_i}{s-\rho_i}-\widetilde{\tau}^{\dagger}(s)\right), \tag{2.42}$$

where $\{C_i\}_{i=1}^n$ are some constants and

$$\widetilde{\tau}^{\dagger}(s) = \lambda_0 T_{s,z}p_0(0) - \lambda_1\sum_{i=1}^{n}\omega_i\alpha_i\frac{T_{s,z}(g_{\alpha_i+\delta}*k)^{+}(0)}{s-\rho_i}. \tag{2.43}$$

From (2.42), the zeros $\{r_l\}_{l=1}^{n+1}$ of $L(s)$ must also be zeros of $cm^{\dagger}(0)-\sum_{i=1}^{n}\frac{C_i}{s-\rho_i}-\widetilde{\tau}^{\dagger}(s)$ which implies that

$$cm^{\dagger}(0)-\sum_{i=1}^{n}\frac{C_i}{r_l-\rho_i}-\widetilde{\tau}^{\dagger}(r_l)=0, \tag{2.44}$$

for $l=1,2,\ldots,n+1$. Multiplying (2.44) by D_l defined by (2.30) and summing over $l = 1,2,\ldots,n+1$ yield

$$cm^{\dagger}(0)\sum_{l=1}^{n+1} D_l - \sum_{i=1}^{n} C_i \left(\sum_{l=1}^{n+1} \frac{D_l}{r_l-\rho_i}\right) - \sum_{l=1}^{n+1} D_l \widetilde{\tau}^{\dagger}(r_l) = 0. \tag{2.45}$$

We now show that

$$\sum_{l=1}^{n+1} D_l = 1, \tag{2.46}$$

and

$$\sum_{l=1}^{n+1} \frac{D_l}{r_l-\rho_i} = 0, \tag{2.47}$$

for $i=1,2,\ldots,n$. From an application of Lagrange polynomials on $f(x)=\prod_{j=1}^{n}(x-\rho_j)$, we have

$$\begin{aligned} f(x) &= \sum_{l=1}^{n+1} f(r_l) \prod_{j=1, j\neq l}^{n+1} \frac{x-r_j}{r_l-r_j} \\ &= \sum_{l=1}^{n+1} D_l \prod_{j=1, j\neq l}^{n+1} (x-r_j)\,. \end{aligned} \tag{2.48}$$

Equating coefficients of x^n on both sides of (2.48), we obtain $\sum_{l=1}^{n+1} D_l = 1$. Also from (2.48),

$$f(x) = \left\{\prod_{j=1}^{n+1}(x-r_j)\right\} \sum_{l=1}^{n+1} \frac{D_l}{x-r_l}. \tag{2.49}$$

Evaluating (2.49) at $x=\rho_i$, we conclude that $\sum_{l=1}^{n+1} \frac{D_l}{\rho_i - r_l} = 0$ for $i=1,2,\ldots,n$, since we assume $r_j \neq \rho_i$ for any $i=1,2,\ldots,n$ and $j=1,2,\ldots,n+1$.

Finally, using (2.45), (2.46) and (2.47), we obtain

$$cm^{\dagger}(0) = \sum_{l=1}^{n+1} D_l \widetilde{\tau}^{\dagger}(r_l) \tag{2.50}$$

Substituting (2.43) into (2.50), it is easy to show $m^{\dagger}(0)=\widetilde{a}(z)$ given by the LT of (2.28), as required. □

Remark 5. From Theorem 9, it is clear that $\{a(y)\}_{y\geq 0}$ is a defective density when $\delta > 0$. For $\delta = 0$ and (2.26) holds, we note that a remains a defective density from the theory on random walks (e.g., Rolski et al. [1999, Theorem 6.3.1]).

We now turn our attention to the function $\xi(u)$, another term appearing in the DRE (2.27).

2.3.3 Comment on the function $\xi(u)$

Define

$$\zeta = \inf\{t \geq 0 : U_t < U_0 \text{ and } M_t \text{ is even}\},$$

to be the first time the surplus process is below its initial surplus level when no claim is being investigated. We have $\zeta = T^{\dagger}$ almost surely when $U_0 = 0$. Since a was shown to be the discounted (defective) ladder height density in the QBCI surplus process, it follows that

$$a(y) = \frac{d}{dy}\mathrm{E}\left[e^{-\delta\zeta}1_{\{U_0 - U_\zeta \leq y\}}|U_0 = u\right].$$

When $u > 0$, we propose to decompose $a(y)$ into two terms, namely $a^{NR}(y;u)$ and $a^{R}(y;u)$, which accounts for whether (or not) the process U makes a first passage below level 0 prior to ζ. Thus, we define

$$a^{NR}(y;u) = \frac{d}{dy}\mathrm{E}\left[e^{-\delta\zeta}1_{\{U_0 - U_\zeta \leq y, \zeta < T\}}|U_0 = u\right], \qquad 0 < y < u, \tag{2.51}$$

and

$$a^{R}(y;u) = \frac{d}{dy}\mathrm{E}\left[e^{-\delta\zeta}1_{\{U_0 - U_\zeta \leq y, \zeta \geq T\}}|U_0 = u\right], \qquad y > 0,$$

where

$$a(y) = a^{NR}(y;u) + a^{R}(y;u), \tag{2.52}$$

for $y > 0$.

Another way to obtain a renewal-like equation for the Gerber-Shiu function $m(u)$ is to condition on events related to whether $\{T \leq \zeta\}$ or $\{T > \zeta\}$. Indeed, using Equation (2.51), we have

$$m(u) = \int_0^u m(u-y)a^{NR}(y;u)dy + \mathrm{E}\left[e^{-\delta T}w(U_{T^-}, |U_T| + \eta V_T)1_{\{T\leq\zeta\}}|U_0 = u\right]. \tag{2.53}$$

Substituting (2.52) into (2.53) and subsequently comparing it to (2.27), it is clear that

$$\xi(u) = \mathrm{E}\left[e^{-\delta T}w(U_{T^-}, |U_T| + \eta V_T)1_{\{T\leq\zeta\}}|U_0 = u\right] - \int_0^u m(u-y)a^R(y;u)dy.$$

In general, it may be difficult to infer on whether $\xi(u)$ is positive for all $u \geq 0$.

However, in the special case $w(x,y) \equiv 1$, we have

$$\begin{aligned} \mathrm{E}\left[e^{-\delta T}1_{\{T\leq\zeta\}}|U_0 = u\right] &\geq \mathrm{E}\left[e^{-\delta T}1_{\left\{T<\zeta, U_\zeta\geq 0\right\}}|U_0 = u\right] \\ &\geq \int_0^u a^R(y;u)dy. \end{aligned}$$

It follows that

$$\xi(u) \geq \int_0^u a^R(y;u)\left(1 - m(u-y)\right)dy,$$

which is non-negative given that $m\,(u)$ stands for the LT of the time to ruin in this case, and is therefore bounded by 1.

2.4 Numerical examples

We now measure the impact of variations in the claim investigation mechanism on the ruin probability $\psi(u)$. Our objective is to confirm some intuitive risk management features of the model and more importantly, quantify their impact from a risk management standpoint.

As a baseline case, we assume claims arrive according to a Poisson process with rate $\lambda = 5$, and claim sizes are distributed as a mixture of an Erlang-2 and an Erlang-5 distribution with

LT

$$\widetilde{p}(s) = 0.8\left(\frac{0.6}{0.6+s}\right)^2 + 0.2\left(\frac{2}{2+s}\right)^5, \qquad s \geq 0. \tag{2.54}$$

A premium rate of $c = 18$ is assumed. Investigation times are assumed to follow a logbeta distribution, i.e. $W_2 = -\ln Y$, where Y is a beta rv with density

$$f_Y(y) = \frac{\Gamma(n+\alpha+1)}{\Gamma(\alpha+1)(n-1)!} y^{\alpha}(1-y)^{n-1}, \qquad 0 < y < 1, \tag{2.55}$$

with $\alpha = -0.7615$ and $n = 2$. It can be shown (e.g., Dufresne [2007] and Klugman et al. [2013, Example 2.2]) that W_2 has density (2.3) where $n = 2$, $\omega_1 = 1.2385$, $\alpha_1 = 0.2385$, $\omega_2 = -0.2385$, and $\alpha_2 = 1.2385$. The mean of W_2 can be found to be $\mu = 5$. Unless stated otherwise, the next four examples in this section will all use these baseline case assumptions.

We also assume $L_x(w)$ is as given in (2.2) where $\varsigma(x) = \varsigma = 0.15$ in Sections 2.4.1–2.4.3 while in Section 2.4.4, we more realistically assume that the probability $\varsigma(x)$ that a claim subject to investigation is determined to be fraudulent increases with its claim size.

Before continuing, note that the effect of catching fraudulent claims cannot be equivalently modelled by simply thinning the arrival rate of claims subject to investigation λ_1 using $1 - \varsigma(x)$ (e.g., Ross [2007, p. 310–311]). The inability to utilize thinning in our setting is due to the balking feature of the assumed queueing mechanism.

To clarify, note that the arrival rate of claims that actually undergo investigation is λ_1 only when the investigator is idle. When the investigator is busy, the arrival rate of claims that undergo investigation is equal to zero because claims arriving when the investigator is busy are immediately paid in full, i.e., there is balking.

The effect of catching fraudulent claims could be modelled by thinning as noted above if the claim investigation mechanism employs an infinite number of investigators. In this case, the mechanism does not experience congestion (since there are always available investigators to investigate incoming claims) and there is no balking. As a result, at any given time, λ_1 is the arrival rate of claims that actually undergo investigation.

2.4.1 Effect of $q_1(x)$ and cost of implementing a claim investigation mechanism

In this first example, we assume the probability a claim is subject to investigation is constant, i.e. $q_1(x) \equiv q_1$. To illustrate q_1's impact on an insurer's surplus, ruin probabilities for various q_1 are provided in Table 2.1 and plotted in Figure 2.2. We note that when $q_1 = 0$, the classical compound Poisson model (without investigation) is recovered.

Table 2.1: Ruin probabilities for varying q_1

$q_1 \setminus u$	0	15	30	45	60	75
0	0.8796	0.4083	0.1871	0.0857	0.0393	0.0180
0.05	0.8604	0.3719	0.1647	0.0738	0.0332	0.0149
0.5	0.7893	0.3362	0.1499	0.0668	0.0298	0.0133
1	0.7438	0.3336	0.1487	0.0662	0.0295	0.0131

Table 2.1 and Figure 2.2 show that all else being equal, the ruin probability decreases as q_1 increases. This suggests that introducing a claim investigation mechanism (and furthermore increasing the probability a claim is subjected to investigation) leads to lower ruin probabilities. This confirms our intuition since an investigation mechanism has the effect of causing some claim payments to be both delayed by an investigation and potentially denied. As evidenced in Table 2.1, this payment delay/denial effect is naturally less pronounced as the initial surplus level increases.

A somewhat more realistic example consists in taking into account a certain cost associated with the claim investigation activities. Here, we assume that any costs generated by

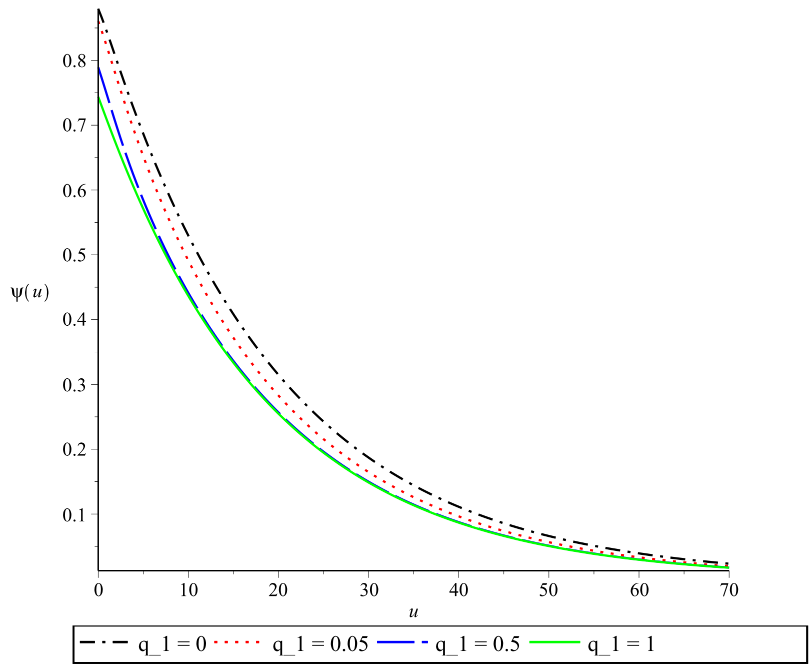

Figure 2.2: Ruin probabilities for varying q_1

the claim investigation mechanism (e.g. investigator's salary, administrative systems, etc.) can be modelled as a continuous fixed rate, c_{cost}, that effectively reduces the (net) rate of premium collected to $c - c_{cost}$. To be a sensible risk management strategy, the cost c_{cost} associated to the investigation process is assumed to not lower the drift of the surplus process (assuming no investigation), that is,

$$c_{cost} \quad < \quad \frac{\varsigma \mathrm{E}[X]}{\frac{1}{\lambda_1} + \mu}.$$

Now consider fixing $q_1 = 0.2$ and letting $c = 18 - \varsigma \mathrm{E}[X] \left(\frac{1}{\lambda_1} + \mu \right)^{-1} = 17.9208$ such that the drift is equal to that of a process with $q_1 = 0$ (no investigation) and $c = 18$. Ruin probabilities for both processes are given in Table 2.2. We observe that while the drifts of both processes are equal, the lower ruin probability associated with claim investigations can be attributed to the payment delay/denial effect.

Table 2.2: Ruin probabilities for two processes with the same drift

$(q_1, c) \setminus u$	0	15	30	45	60	75
$(0, 18)$	0.8796	0.4083	0.1871	0.0857	0.0393	0.0180
$(0.2, 17.9208)$	0.8334	0.3571	0.1626	0.0745	0.0341	0.0156

2.4.2 Effect of mean investigation time

In this example, we fix $q_1(x) = q_1 = 0.1$ and vary the mean investigation time μ by adjusting in (2.55) the value of α while fixing $n = 2$. We observe different effects on the ruin probability depending on the value of ς. To illustrate, we consider $\varsigma = 0.15$ and $\varsigma = 0.5$. First, we let $\varsigma = 0.15$ and the ruin probabilities can be found in Table 2.3 for varying μ.

Table 2.3: Ruin probabilities for varying μ when $\varsigma = 0.15$

$\mu \setminus u$	0	15	30	45	60	75
0.5	0.8554	0.3648	0.1562	0.0669	0.0287	0.0123
1	0.8521	0.3622	0.1567	0.0680	0.0295	0.0128
2	0.8495	0.3597	0.1573	0.0692	0.0305	0.0134
5	0.8475	0.3575	0.1580	0.0706	0.0317	0.0142
∞	0.8462	0.3558	0.1589	0.0723	0.0331	0.0152

Table 2.3 shows that when u is small (e.g. $u = 0, 15$), the ruin probabilities decrease with increasing mean investigation time and the opposite effect is observed for larger u (e.g. $u \geq 30$). Now consider the ruin probabilities for varying μ when $\varsigma = 0.5$ in Table 2.4. Table 2.4 shows the ruin probability increases with increasing mean investigation time for all u. The behaviour exhibited in Tables 2.3 and 2.4 can be explained by two counter effects resulting from longer mean investigation time. Longer investigation times will further delay claim

payments which will decrease the ruin probability. On the other hand, longer investigation times will decrease additional opportunities for the investigator to identify fraudulent claims which will increase the ruin probability.

Table 2.4: Ruin probabilities with varying μ when $\varsigma = 0.5$

$\mu \setminus u$	0	15	30	45	60	75
0.5	0.8342	0.3051	0.1115	0.0408	0.0149	0.0055
1	0.8356	0.3138	0.1192	0.0455	0.0173	0.0066
2	0.8380	0.3246	0.1291	0.0517	0.0207	0.0083
5	0.8415	0.3383	0.1419	0.0602	0.0256	0.0109
∞	0.8462	0.3558	0.1589	0.0723	0.0331	0.0152

When $\varsigma = 0.15$, the first effect dominates for small u while the second effect takes over for larger u. When $\varsigma = 0.5$, fraudulent claims are more prevalent so the second effect dominates for all u and the ruin probability increases with the mean investigation time for all u.

This suggests that in our proposed model, when the probability of fraudulent claims is relatively high, it is advantageous for the insurer to implement efficient investigation processes with shorter investigation times. However, this also suggests that when both the initial surplus and the probability of fraudulent claims are low, it is advantageous for the insurer to have longer investigation times. From a practical standpoint, competitive pressure would disincentivize insurers from artificially implementing inefficient investigation processes to, e.g., lower the solvency risk.

2.4.3 Three investigation strategies

In this example, we consider three different strategies to determine whether or not a claim will be subject to investigation. Unlike the previous two examples, we allow $q_1(x)$ to vary in x by assuming that

$$q_1(x) = \theta + (1-\theta)\left(1 - e^{-\kappa x}\right),$$

where θ and κ are chosen such that the mean probability of subjecting a random claim to investigation is set to be 0.1. Thus, on average, a percentage θ ($\theta \leq 0.1$) of all claims are automatically subjected to investigation; the other $(1-\theta)$ are subjected based on the claim amount with a probability of $1 - e^{-\kappa x}$ for a claim of size x. We consider the following three strategies with their respective ruin probabilities in Table 2.5:

$$\begin{array}{lll} \text{Strategy 1:} & \theta = 0.1 & (\kappa = 0) \\ \text{Strategy 2:} & \theta = 0.05 & (\kappa = 0.0173) \\ \text{Strategy 3:} & \theta = 0 & (\kappa = 0.0341) \end{array}$$

Table 2.5: Ruin probabilities with varying investigation strategies

$\theta \setminus u$	0	15	30	45	60	75
0.1	0.8475	0.3575	0.1580	0.0706	0.0317	0.0142
0.05	0.8408	0.3448	0.1513	0.0674	0.0300	0.0134
0	0.8343	0.3335	0.1456	0.0646	0.0287	0.0127

Table 2.5 shows that the ruin probability decreases with decreasing θ. This behaviour suggests that when determining whether or not a claim should be subject to investigation,

the strategy that minimizes the ruin probability is one that is completely based on the claim amount. This again agrees with intuition.

2.4.4 Probability of fraudulent claims increasing with claim size

In this example, we again fix $q_1(x) = q_1 = 0.1$. Also, we assume that a claim of size x subject to investigation will result in a payment with df (2.2). We compare the strategy of Section 2.4.2 where $\varsigma = 0.15$ (and $\mu = 5$) with $\varsigma(x) = (1 - e^{-\varphi x})/2$ with $\varphi = 0.1227$. For comparative purposes, both strategies have the same expected conditional probability that a claim is determined to be fraudulent, i.e. $\int_0^\infty \varsigma(x) p_1(x) dx = 0.15$. Ruin probabilities are provided in Table 2.6.

Table 2.6: Ruin probabilities for constant and non-constant $\varsigma(x)$

$\varsigma(x) \setminus u$	0	15	30	45	60	75
0.15	0.8475	0.3575	0.1580	0.0706	0.0317	0.0142
$(1 - e^{-\varphi x})/2$	0.8465	0.3540	0.1550	0.0687	0.0305	0.0136

When the probability of catching fraudulent claims increases with its size, the investigator is more likely to entirely deny larger claims which leads to lower ruin probabilities. This effect is confirmed in Table 2.6 where ruin probabilities when $\varsigma(x) = (1 - e^{-\varphi x})/2$ are lower than those when $\varsigma(x)$ is constant.

Chapter 3

A risk model with a more advanced claim investigation mechanism with MAP formulation

3.1 Introduction

This chapter extends the surplus analysis conducted in Chapter 2 by proposing a generalized, more realistic queue-based claim investigation mechanism (CIM). The proposed CIM consists of n ($n \in \{1, 2, \ldots\}$) investigation units and m ($m \in \{0, 1, 2, \ldots\}$) waiting places. By contrast, the CIM studied in Chapter 2 consisted of one investigation unit ($n = 1$) and zero waiting places ($m = 0$). Note that Chapter 2's CIM, with some restrictions, can be shown to be a special case of the generalized CIM proposed in this chapter.

Analyzing this generalized model using the standard techniques of conditioning on the first claim (as used in Chapter 2) is prohibitive because the resulting risk process no longer has i.i.d. interclaim times. An alternative way to analyze the model involves showing the insurer's surplus can be formulated using a (spectrally negative) Markov Additive Process (MAP) (see, e.g., Ivanovs and Palmowski [2012]).

To develop the MAP formulation, we introduce a background process that counts the number of claims in the CIM at any given time. Under the CIM assumptions, the background process behaves as a Markov process and the insurer's surplus can be represented as a MAP. More specifically, the insurer's surplus can be formulated as a special case of the MAP, namely, the perturbed Markovian Arrival Process (see, e.g., Neuts [1979]). By using a MAP formulation for the model, standard MAP machinery can be used to analyze various risk management quantities of interest, such as, the joint Laplace transform (LT) of the time and the deficit at ruin.

In addition, the flexibility of the MAP allows for a large number of possible generalizations to the CIM while preserving the surplus' MAP formulation. Specifically, the MAP formulation can accommodate any CIM modelled after a queueing system (see, e.g., Kleinrock [1975]) having an embedded Markov process and Markov-modulated Poisson arrivals.

This chapter discusses four possible CIM generalizations in detail and provides a MAP formulation for each. These more general CIMs supply an additional degree of realism to the mathematical model by enabling the model to better characterize the levels of congestion faced by insurers in their claims handling activities, i.e., the volume of claims waiting or undergoing investigation and their associated claim size. Each of the four generalizations supplies an additional degree of realism as follows:

1. We relax the assumption of exponential investigation time to consider phase-type distributed investigation times. In practice, investigation times may not follow an exponential distribution. The class of phase-type distributions is a very general distribution because it is dense in the sense of weak convergence for all distributions with positive support (see, e.g., [Asmussen and Albrecher, 2010, Chapter IX]).

2. We relax the assumption that a claim's size and investigation time are independent. In practice, we would expect that larger (smaller) claims experience longer (shorter) investigation times on average.

3. We relax the assumption that all claims entering a CIM are assumed to undergo only one stage of investigation before becoming settled. In practice, the claims handling process usually involves multiple stages, e.g., an eligibility specialist may first determine a claim's eligibility for coverage before it gets passed to investigation where the payment amount is determined, and a final stage of payment processing may be required for claims that pass investigation.

4. We relax the assumption that claims occur according to a Poisson process and consider instead a Markov-modulated Poisson claim arrival process. In practice, the claim arrival rates that insurers experience can often depend on the state of the economy (see, e.g., Wells et al. [2009]). For example, group disability insurers often experience higher claim arrival rates in weak economic times. One explanation for this effect is that, during a recession, employees who had health impairments face a lower opportunity cost when applying for benefits when they become unemployed (Maestas et al. [2015]).

Note that using our approach, any of these generalizations can be combined so long as the resulting CIM can be modelled after a queueing system having an embedded Markov process and Markov-modulated Poisson arrivals. By better characterizing the congestion inherent in the insurer's business, these more realistic CIMs can allow for more accurate estimates of ruin-related quantities. They can also better model and predict the impact of alternative CIM process designs.

There have been a number of studies of insurer surplus models having a Markovian arrival process representation (see, e.g., Asmussen [1989], Badescu et al. [2005], Ahn and Badescu [2007], and Cheung and Feng [2013]). Special cases of these risk models have also been studied in, e.g., Albrecher and Boxma [2005] and Lu and Tsai [2007]. As well, work in the context of a dividend barrier strategy can be found in, e.g., Li and Lu [2007] and Cheung and Landriault [2009b]. The model proposed in this chapter is a special case of that found in Cheung and Landriault [2009b] though without the application of dividends. Instead, we introduce a particular claim settlement application to their model and from a numerical standpoint, examine the impact of variations in an insurer's claims handling practices on their probability of ruin.

More recently, in Ahn et al. [in press], a claim developmental process is superimposed onto the insurer's surplus process to incorporate the liability due to Incurred But Not Reported (IBNR) and Reported But Not Settled (RBNS) claims. Their model's representation as a Markovian risk process is discussed. Furthermore, connections to fluid flow models are made for when payments are phase-type distributed and the Gerber-Shiu function and joint moments involving the ruin time and aggregate payments with and without claim settlement. Their numerical illustrations demonstrate the generality of their model. Furthermore, a numerical example involving a real dataset is presented where the ruin probability is computed following calibration of the model, and a discussion of some difficulties involved is also included.

Like in Ahn et al. [in press], we superimpose a claim developmental process onto the insurer's surplus process. However, we do not consider IBNR and focus only on the impact of RBNS claims on the ruin probability. In other words, we are concerned only with the claims handling process that takes place as soon as notification of a claim has reached the insurer. Our study is distinct from Ahn et al. [in press] in five major respects:

1. We focus on measuring the effect on the ruin probability of variations in the CIM at an operational level. For example, modelling a CIM that possesses stages (such as investigation, payment processing, etc.) or allowing for a dependence structure

between claim size and investigation time.

2. We consider the interplay of costs and benefits generated by variations in the CIM, such as the trade-off between the cost and benefit of hiring additional investigators.

3. We model the effect of diffusion which aims to model the variability from premium, investment income, or other factors perturbing the insurer's surplus process.

4. We assume that the insurer remains liable for a fixed portion of the total payment amount of claims undergoing investigation at the time of ruin.

5. We also compare three LT inversion methods to calculate the finite-time ruin probability in our model and find that the Gaver-Stehfest method (see, e.g., Abate and Valkó [2004] and references therein) is very efficient.

The present model is a special case of the more general (spectrally negative) Markov Additive process and their results to some exit problems can be found in Ivanovs and Palmowski [2012]. Also, for a generalization of the Markovian arrival model where waiting times between two successive events have an arbitrary distribution, see Cheung and Landriault [2010].

The chapter is organized as follows. Section 3.2 presents the model and its MAP formulation. Section 3.3 gives an expression for the joint LT of the time and generalized deficit at ruin. Numerical examples are provided in Section 3.4 to assess the cost and benefit to the insurer of additional investigators as well as the effect of diffusion on the ruin probability. To conclude, a detailed discussion of generalizations to the MAP formulation for four possible CIM generalizations can be found in Section 3.5. Numerical examples are also included throughout Section 3.5 to measure the impact of variations in the CIM on the ruin probability.

3.2 Model presentation

Recall the definition of the Cramér-Lundberg surplus process with diffusion perturbation $\underline{U}^* = \{U_t^*\}_{t\geq 0}$ given by

$$U_t^* = u + ct + \sigma B_t - S_t, \tag{3.1}$$

where $u\,(u \geq 0)$ is the initial surplus level, $c\,(c > 0)$ is the level premium rate per unit time, $\{B_t\}_{t\geq 0}$ is a standard Brownian motion, and the aggregate claim process $\{S_t\}_{t\geq 0}$ is a compound Poisson process with Poisson arrivals at rate $\lambda\,(\lambda > 0)$ and positive jumps with density p, independent of $\{B_t\}_{t\geq 0}$.

In what follows, we propose a queueing-based claim investigation mechanism (CIM), superimpose it onto the surplus process (3.1) and provide the Markov Additive Process (MAP) formulation for the resulting surplus process.

3.2.1 Proposed queueing-based claim investigation mechanism

We propose a claim investigation mechanism (CIM) that consists of n investigation units where each investigation unit investigates one claim at a time. Thus, the CIM allows for at most n claims to be investigated at any given time. The CIM also permits at most m claims to wait for investigation at any given time. Details on how claims are selected to enter the CIM for investigation, (possibly) wait in the queue, and depart from the CIM are provided next.

We define the *number of claims in the CIM* as the sum of the number of claims either under investigation or waiting in the queue. When there are no claims in the CIM, we say the CIM is *empty* while the CIM is said to be *at full capacity* when there are $n+m$ claims in the CIM. We also say an investigation unit is *busy* (*idle*) if it is (not) investigating a claim.

Next, we separately consider two cases: the CIM is not at full capacity and the CIM is at full capacity. First, when the CIM is not at full capacity (i.e., when there are less than $n + m$ claims in the CIM), a claim arriving of size $x > 0$ will:

1. with probability $q_0(x) \geq 0$, not be selected for investigation, avoid the CIM and be paid immediately in its entirety; or
2. with probability $q_1(x) = 1 - q_0(x) \geq 0$, be selected to enter the CIM for investigation and:
   (a) if at least one investigation unit is idle (i.e., when there are less than n claims in the CIM), the claim will immediately undergo investigation of the amount claimed x, or
   (b) if all n investigation units are busy (i.e., when there are n or more claims in the CIM), the claim will enter the queue to wait for investigation. Claims waiting for investigation are assumed to form a single queue.

The decision to investigate a claim of size x depends only on its size x and is independent of all else in the risk model.

All claim investigation times are i.i.d. exponential rv's with mean $1/\alpha$ and are assumed to be independent of all else in the risk model. Following the completion of investigation by an investigation unit, the claim amount under investigation, say x, will result in a payment w with df $L_x(w)$ ($w \geq 0$), independent of all else in the risk model, and the claim is said to depart from the CIM. Furthermore, if there is at least one claim waiting in the queue, the investigation unit will immediately begin investigation on the claim at the front of the queue. This is consistent with the first come first serve discipline (FCFS) in queueing (see, e.g., [Kleinrock, 1975, Section 1.2]).

Lastly, we consider the case where the CIM is at full capacity (i.e., there are $n+m$ claims in the CIM). In this case, all arriving claims will be immediately paid in full. This assumption is consistent with the concept of a single server queue with *balking* (e.g., Gross et al. [2011, Section 1.2.1]) or *finite storage* in queueing theory (see, e.g., Kleinrock [1975, Section 3.6])

The queueing-based claim investigation mechanism can be represented diagramatically as shown in Figure 3.1.

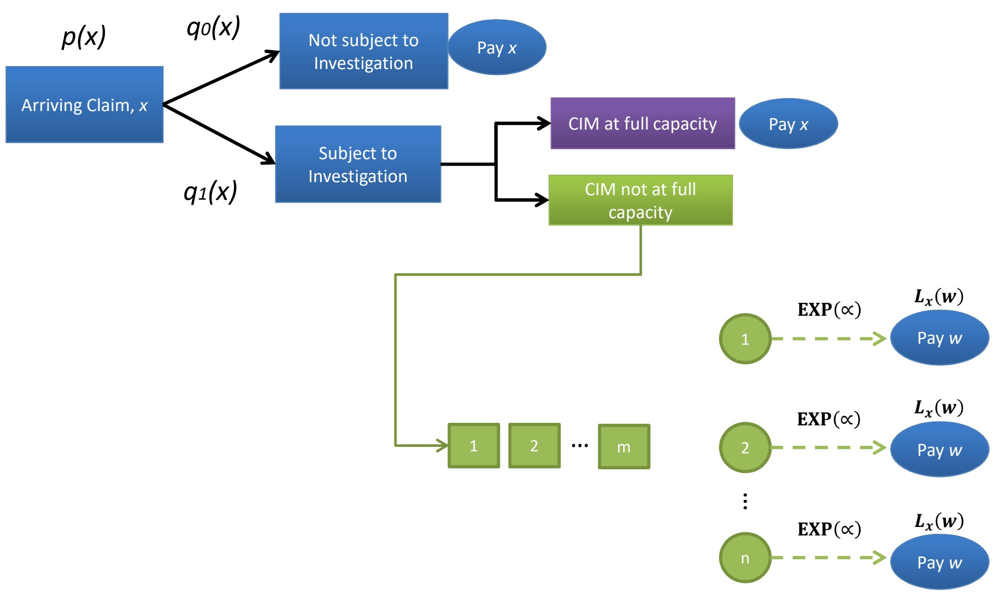

Figure 3.1: Diagramatic representation of the more advanced queue-based claim investigation mechanism

We continue to use the term “claim” to refer to the loss amount declared by the policyholder (after appropriately applied coverage modifications, e.g. deductible, policy limit, etc.), and the term “payment” to refer to the amount actually paid to the policyholder by the insurer.

Note that if we let there be one investigation unit and zero waiting places (i.e., $n = 1$ and $m = 0$), then we obtain the CIM from Chapter 2 if the investigation times are assumed to be exponential (i.e., if the investigation time density (2.3) is given by $h(t) = \alpha e^{-\alpha t}$ for $t > 0$) and independent of the claim size under investigation.

3.2.2 Markov additive process formulation

Superimposing the proposed CIM described in the last section onto the surplus process (3.1) results in a surplus process that can be formulated as a Markov additive process (MAP). Recall that a technical background on MAPs was provided in Section 1.2.2. More specifically, the resulting surplus process can be formulated as a well-known special case of the MAP called the perturbed Markovian arrival risk process (MArP) and was defined in Remark 1. In this section, we present this MArP formulation and determine the matrix exponent characterizing the surplus process.

As was done in Chapter 2, we define

$$\lambda_l = \lambda \int_0^\infty p(y) q_l(y) dy,$$

and

$$p_l(x) = \frac{p(x) q_l(x)}{\int_0^\infty p(y) q_l(y) dy},$$

for $l = 0, 1$, and $x > 0$. We have that λ_1 (λ_0) corresponds to the arrival rate of claims that are selected to enter the CIM (avoid the CIM) and $p_1(x)$ ($p_0(x)$) is the claim size density of such claims. We also define

$$K(w) = \int_0^\infty L_x(w) p_1(x) dx,$$

corresponding to the size of a payment for a claim undergoing the investigation process.

For completeness, we recall the definition of a MArP process. Define a process $U = \{U_t\}_{t \geq 0}$ and an irreducible continuous-time Markov process $J = \{J_t\}_{t \geq 0}$ with finite state space E and infinitesimal generator $\mathbf{Q}$. When $\{J_t = i\}$, the process U, commonly referred to as the additive component, evolves as some perturbed compound Poisson risk process U^i and the processes $U^1, U^2, \ldots, U^n$ are assumed to be independent. In addition, a transition of J from i to $j \neq i$ triggers a downward jump of U whose (absolute) size is denoted by rv $P_{i,j} \geq 0$ for $i, j \in E$. For $i, j \in E$, the matrix $\mathbf{G}(z)$ is defined as having (i, j)th element equal to $E\left[e^{-zP_{i,j}}\right]$ when $i \neq j$ and diagonal elements equal to one.

From the MArP definition, we note that there are two types of downward jumps in U: those triggered by state transitions in J and those occurring when J remains in some state (via a compound Poisson process). Correspondingly, note that there are two types of claim payments in our surplus process: payments from claims exiting the CIM after their investigation and payments from claims that do not enter the CIM (due to claims that avoid being investigated or arrive when the CIM is at capacity). In the following, we model our surplus process using U with:

1. payments from claims that exit the CIM modelled by jumps triggered by state transitions in J, and
2. payments from claims that do not enter the CIM modelled by jumps from a compound Poisson risk process when J remains in some state.

We let J_t correspond to the number of claims in the CIM at time t by specifying the infinitesimal generator $\mathbf{Q}$ as follows:

1. When $0 \leq i < n + m$ and $j = i + 1$, we let $(\mathbf{Q})_{i,j} = \lambda_1$. That is, when the CIM is not at full capacity, a transition in J from state i to state $i + 1$ corresponds to a claim entering the CIM which occurs at rate λ_1. Note that these transitions do not trigger any payment, i.e., $P_{i,j} = 0$.

2. When $0 < i \leq n+m$ and $j = i-1$, we let $(\mathbf{Q})_{i,j} = min(i,n)\alpha$. That is, when the CIM is not empty, a transition in J from state i to state $i-1$ corresponds to a claim exiting the CIM which occurs at rate $i\alpha$ if $i\,(i < n)$ investigation units are busy and rate $n\alpha$ if all investigation units are busy. Note that these transitions trigger a payment $P_{i,j}$ having df K.

3. Since there are no other possible transitions for J, all remaining off-diagonal elements of $\mathbf{Q}$ are equal to zero. Also, the diagonal elements of $\mathbf{Q}$ are negative such that the sum of the elements on each row of $\mathbf{Q}$ is zero.

Thus, $\mathbf{Q}$ is given by

$$\mathbf{Q} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & & & \\ \alpha & -(\lambda_1+\alpha) & \ddots & \ddots & & & \\ 0 & 2\alpha & -(\lambda_1+2\alpha) & & & & \\ \vdots & \ddots & \ddots & \ddots & & & \\ & & & n\alpha & -(\lambda_1+n\alpha) & & \\ & & & & \ddots & \ddots & \\ & & & & & n\alpha & -n\alpha \end{bmatrix}. \tag{3.2}$$

From the infinitesimal generator specified above, it is easy to see that J is irreducible. Also, we have that the matrix $\mathbf{G}(z)$ has all elements equal to 1 except $(\mathbf{G}(z))_{i,j} = \widetilde{k}(z)$ when $0 < i \leq n+m$ and $j = i-1$, i.e.,

$$\mathbf{G}(z) = \begin{bmatrix} 1 & 1 & \cdots & & \\ \widetilde{k}(z) & \ddots & \ddots & & \\ 1 & \ddots & & & \\ \vdots & & & & \\ & & & \widetilde{k}(z) & 1 \end{bmatrix}. \tag{3.3}$$

Recall that when $\{J_t = i\}$, the process U evolves as a perturbed compound Poisson risk process U^i for $i \in E$. We define

$$U_t^i = ct + \sigma B_t - S_t^i,$$

where $\{S_t^i\}_{t \geq 0}$ is a compound Poisson process. When $0 \leq i < n+m$, the CIM is not at full capacity and the compound Poisson process S_t^i has intensity rate λ_0 and i.i.d. jumps with density p_0. This corresponds to claims that are not selected for investigation and are immediately paid in their entirety. When $i = n+m$, the CIM is at full capacity and the compound Poisson process S_t^{n+m} has intensity rate λ and i.i.d. jumps with density p. This corresponds to claims that are paid immediately in full since they arrived to find the CIM at full capacity. It is assumed that the processes $U^1, U^2, \ldots, U^{n+m}$ are independent.

Let $\psi_i(z)$ be the Laplace exponent of U^i defined as

$$E\left[e^{zU_t^i}\right] = e^{\psi_i(z)t},$$

where when $0 \leq i < n+m$,

$$\psi_i(z) = cz + \frac{\sigma^2}{2}z^2 + \lambda_0\left(\widetilde{p}_0(z) - 1\right),$$

and when $i = n+m$,

$$\psi_{n+m}(z) = cz + \frac{\sigma^2}{2}z^2 + \lambda\left(\widetilde{p}(z) - 1\right).$$

Also, define $\mathbf{F}^\delta(z)$ to be the U-matrix analogue of the Laplace exponent called the matrix exponent such that

$$E\left[e^{-\delta t + zU_t}; J_t\right] = e^{\mathbf{F}^\delta(z)t}.$$

Proposition 10. The matrix exponent characterizes our surplus process U (see, e.g., Ivanovs and Palmowski [2012]) and is given by

$$\mathbf{F}^\delta(z) = \left(cz + \frac{\sigma^2}{2}z^2 - \delta\right)\mathbf{I}z - \mathbf{\Lambda} + \mathbf{\Lambda} \circ \mathbf{P}(z) + \mathbf{Q} \circ \mathbf{G}(z), \tag{3.4}$$

where $\mathbf{Q}$ and $\mathbf{G}(z)$ are given by (3.2) and (3.3), respectively. Also, $\mathbf{\Lambda} = diag\{\lambda_0, \ldots, \lambda_0, \lambda\}$ is a diagonal matrix with all diagonal entries equal to λ_0 except the last which is λ and $\mathbf{P}(z) = diag\{\widetilde{p}_0(z), \ldots, \widetilde{p}_0(z), \widetilde{p}(z)\}$ is a diagonal matrix with all diagonal entries equal to $\widetilde{p}_0(z)$ except the last which is given by $\widetilde{p}(z)$.

Thus, we conclude that the surplus process (3.1) with superimposed CIM can be formulated by a MArP via the bivariate process (U, J) specified above with matrix exponent given by (3.4).

Remark 6. Consider a special case of the CIM from Chapter 2 where investigation times are exponential (i.e., the density (2.3) is restricted to $h(t) = \alpha e^{-\alpha t}$ for $t > 0$). Note that this special case has one investigation unit and zero waiting places. The surplus process for this special case can be formulated by a MArP as described in this section with $n = 1$ and $m = 0$ having matrix exponent given by (3.4) with $\mathbf{Q}$ and $\mathbf{G}(z)$ respectively replaced by

$$\mathbf{Q} = \begin{bmatrix} -\lambda_1 & \lambda_1 \\ \alpha & -\alpha \end{bmatrix},$$

and

$$\mathbf{G}(z) = \begin{bmatrix} 1 & 1 \\ \widetilde{k}(s) & 1 \end{bmatrix}.$$

Given the MArP representation of the surplus process U, we recall some results on scale functions for MAPs which will be helpful in the characterization of first passage quantities for U. For $\delta \geq 0$, the δ-scale matrix $\mathbf{W}^\delta(x)$ is defined through its LT

$$\int_0^\infty e^{-zx} \mathbf{W}^\delta(x) dx = \mathbf{F}^\delta(z)^{-1}.$$

The δ-scale matrix $\mathbf{W}^\delta(x)$ is closely related to exit problems for MAPs. Its existence is shown in Kyprianou and Palmowski [2008] and a detailed treatment of MAPs and their exit results can be found in Ivanovs and Palmowski [2012]. Also, the second δ-scale matrix is defined as

$$\mathbf{Z}^{\delta}(z,u) = e^{zu}\left(\mathbf{I} - \int_0^u e^{-zy}\mathbf{W}^{\delta}(y)dy\mathbf{F}^{\delta}(z)\right).$$

A second representation of the δ-scale matrix given by [Ivanovs and Palmowski, 2012, Theorem 1] is introduced in Section 1.2.2 and for completeness, we recall it here. For $x \in R$, let $T_x^{U,+(-)} = inf\{t \geq 0 : U_t > (<)\, x\}$ be the first passage time of U. The δ-scale matrix can be written as

$$\mathbf{W}^{\delta}(x) = e^{-\mathbf{D}^{\delta}x}\mathbf{L}^{\delta}(x),$$

where $\mathbf{D}^{\delta}$ is a transitional rate matrix of the Markov chain associated with the first passage and $\mathbf{L}^{\delta}(x)$ is a matrix associated with the expected occupation times at 0 up to the first passage time over x. More specifically, $\mathbf{D}^{\delta}$ satisfies $E\left[e^{-\delta T_x^{X,+}}; J_{T_x^{X,+}}\right] = e^{\mathbf{D}^{\delta}x}$, and $\left(\mathbf{L}^{\delta}(x)\right)_{i,j} = E\left[L^{\delta}(j, T_x^{X,+})|J_0 = i\right]$, where $L^{\delta}(j,t)$ is the limit in $L^2(\mathbb{P})$ of

$$L_{\epsilon}^{\delta}(j,t) = \frac{1}{2\epsilon}\int_0^t e^{-\delta s}1_{\{|X_s|<\epsilon,\, J_s=j\}}ds,$$

as $\epsilon \downarrow 0$. Define $\mathbf{L}^{\delta} = \lim_{x\to\infty}\mathbf{L}^{\delta}(x)$ and it is proven in [Ivanovs and Palmowski, 2012, Lemma 10] that $\mathbf{L}^{\delta}$ has finite entries and is invertible unless $\mathbf{Q1} = 0$ (i.e., $\delta = 0$) and the asymptotic drift $\lim_{t\to\infty} U_t/t = 0$, $\mathbb{P}_{0,i}$-a.s. for all $i \in E$.

In the following section, we consider the joint LT of the time and deficit at ruin by making connections with existing MAP results.

3.3 Joint Laplace transform of the time and deficit at ruin

In this section, we propose a generalized definition for the deficit at ruin (consistent with what was done in Chapter 2) and make connections with existing MAP results to obtain an expression for the joint LT of the time and deficit at ruin.

For notational simplicity, the time of ruin $T_0^{U,-}$ is denoted by T where $T = \infty$ if $U_t \geq 0$ for all $t \geq 0$. We propose to use $|U_T| + \eta V_T$ as the deficit at ruin (rather than the usual $|U_T|$), where $\eta \in [0, 1]$ and V_t is the total (future) payment amount of all claims in the CIM (either waiting or undergoing investigation) at time t (if any). In the proposed CIM, V_t is given by

$$V_t = \sum_{i=1}^{J_t} Y_i \tag{3.5}$$

where $\{Y_i\}_{i=1}^{\infty}$ forms a sequence of i.i.d. rv's with common df K and is independent of all else in the model. Recall that J_t denotes the number of claims in the CIM at time t. We adopt the convention that the empty sum is 0. Hence, we assume that if claims are in the CIM when the surplus process U drops below 0, the insurer remains liable for a fixed portion η of their payment amounts. Note that this is consistent with the deficit at ruin as defined in Chapter 2.

In what follows, we write $E_u\left[\,\cdot\,; J_T\right]$ to indicate a matrix with (i, j)th element corresponding to

$$E\left[\,\cdot\,1_{\{J_T=j\}}\middle|\, U_0 = u, J_0 = i\right].$$

For $\delta, z \geq 0$, we let

$$\mathbf{m}_{\delta,z}(u) \equiv E_u\left[e^{-\delta T - z(|U_T| + \eta V_T)} 1_{\{T<\infty\}};\, J_T\right], \tag{3.6}$$

denote the joint matrix LT of the time and deficit at ruin matrix. Note that (3.6) is a special case of the more general Gerber-Shiu function (2.8) studied in Chapter 2.

Proposition 11. *The $(i, j)th$ element of $\mathbf{m}_{\delta,z}(u)$ is given by*

$$\left(\mathbf{m}_{\delta,z}(u)\right)_{i,j} \;=\; \left(\boldsymbol{\gamma}_{\delta,z}(u)\right)_{i,j}\left(\widetilde{k}(z\eta)\right)^j. \tag{3.7}$$

where from [Ivanovs and Palmowski, 2012, Corollary 4]

$$\begin{aligned}\boldsymbol{\gamma}_{\delta,z}(u) &:= E_u\left[e^{-\delta T - z|U_T|} 1_{\{T<\infty\}}; J_T\right] \\ &= \mathbf{Z}^{\delta}(z, u) - \mathbf{W}^{\delta}(u)\left(\mathbf{L}^{\delta}\right)^{-1}\left(z\mathbb{1}_{n+m} + \mathbf{D}^{\delta}\right)^{-1}\mathbf{L}^{\delta}\,\mathbf{F}^{\delta}(z).\end{aligned} \tag{3.8}$$

Note that (3.8) holds assuming it is not true that $\delta = 0$ and the asymptotic drift $\lim_{t\to\infty} U_t/t = 0$.

The proof of Proposition 11 involves conditioning on J_T and using (3.5).

Remark 7. We assume that the surplus process U has a positive drift by letting

$$c > \sum_{i=0}^{n+m} \pi_i \sum_{j=0}^{n+m} d_{i,j} E\left[P_{i,j}\right], \tag{3.9}$$

where $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_{n+m})$ is the stationary distribution of the background process J and can be found by solving $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$ and $\boldsymbol{\pi}\mathbb{1}_{n+m+1} = 1$. Also, $d_{i,j}$ is the rate of transition of J from state i to state j (with possibly $i = j$) with an accompanying claim. Note that π_i can be interpreted as the long-run probability of finding i claims in the CIM. Also, $\sum_{j=0}^{n+m} d_{i,j} E\left[P_{i,j}\right]$ corresponds to the infinitesimal rate of payment expense given there are i claims in the CIM. Thus, (3.9) requires that the premium rate exceeds the long-run payment expense per unit time and is known as the *positive security loading condition*.

Under the CIM we have proposed, we obtain

$$d_{i,j} = \begin{cases} min(i,n)\alpha, & 0 < i \le n+m \text{ and } j = i-1, \\ \lambda_0, & 0 \le i < n+m \text{ and } j = i, \\ \lambda, & i = j = n+m, \\ 0, & \text{otherwise}, \end{cases} \tag{3.10}$$

and

$$E\left[P_{i,j}\right] = \begin{cases} E\left[Y\right], & 0 < i \le n+m \text{ and } j = i-1, \\ E\left[X^0\right], & 0 \le i < n+m \text{ and } j = i, \\ E\left[X\right], & i = j = n+m, \\ 0, & \text{otherwise}, \end{cases} \tag{3.11}$$

by considering all possible transitions in J with an accompanying claim, namely, transitions corresponding to payment of claims that:

1. Have completed investigation,

2. Are not selected for investigation, or

3. Arrive when the CIM is at full capacity.

Substituting (3.10) and (3.11) into (3.9) yields

$$c > \lambda E\left[X\right] - \Big(\left(1 - \pi_{n+m}\right) \lambda_1 E\left[X^1\right] - \alpha E\left[B\right] E\left[Y\right] \Big), \tag{3.12}$$

where $E[B] = \sum_{j=0}^{n+m} min(j,n)\pi_j$ is the long-run mean number of busy investigators and X, X^1, and Y have LT $\widetilde{p}$, $\widetilde{p}_1$, and $\widetilde{k}$, respectively. We point out that $\lambda E\left[X\right]$ corresponds to the long-run rate of payment expenses if there is no investigation (i.e., when $q_1(x) = 0$). Also, the long-run rate of savings resulting from the CIM is given by

$$\textit{long-run rate of savings} = \left(1 - \pi_{n+m}\right) \lambda_1 E\left[X^1\right] - \alpha E\left[B\right] E\left[Y\right], \tag{3.13}$$

which corresponds to the long-run rate of claim expenses entering the CIM minus the long-run rate of payment expenses exiting the CIM. Thus, the right-hand side of (3.12) can be interpreted as the long-run rate of claim expenses net of the long-run rate of savings resulting from the CIM. Note that (3.13) is positive iff $E\left[X^1\right] > E\left[Y\right]$ because $\left(1 - \pi_{n+m}\right) \lambda_1 = \alpha E\left[B\right]$ (see, e.g., Little's law in [Kleinrock, 1975, Section 2.1]). Thus, introducing a CIM in the insurer's procedures will lower the drift of the process provided $E\left[X^1\right]$, the mean claim size under investigation, is greater than $E\left[Y\right]$, the mean payment following investigation.

It is somewhat more realistic to take into account a certain cost generated by the CIM. This can be modelled as a continuous fixed rate c_{cost} that effectively reduces the rate of premium collected. Thus, c in (3.12) can be viewed as the premium rate net of c_{cost}. In the numerical examples to follow, we will examine the impact of c_{cost} on the optimal number of investigators.

Inversion of $\mathbf{m}_{\delta,z}(u)$ in both δ and z results in the joint density of the time and deficit at ruin. Furthermore, the density of the time to ruin and the finite-time ruin probability (see, e.g., Dickson and Willmot [2005] and references therein) can be obtained by inversion in δ of $\mathbf{m}_{\delta,0}(u)$ and $\mathbf{m}_{\delta,0}(u)/\delta$, respectively. The dependence of (3.7) on z and particularly on δ is not straightforward and as such analytical Laplace inversion is in general a very challenging task. Numerical Laplace inversion techniques are discussed in Section 1.3.

We remark that the proposed CIM assumes claims wait in a single queue for homogeneous investigators with exponential investigation times. The flexibility of the model's MArP formulation allows for considerable generalizations to the model while preserving the MArP formulation which is the topic of Section 3.5.

3.4 Numerical examples

3.4.1 Baseline case

As a baseline case for the numerical examples to follow, we assume claims arrive according to a Poisson process with rate $\lambda = 5$ and claim sizes have LT given by (2.54). A (gross) premium rate of 18 is assumed. The probability a claim is selected for investigation is constant, i.e. $q_1(x) \equiv q_1 = 0.5$, and investigation times are assumed to follow an exponential distribution with mean $1/\alpha = 5$. We also assume $L_x(w)$ is as given in (2.2) where $\varsigma(x) = \varsigma = 0.25$ so that it is assumed that 25% of the claims are fraudulent. Furthermore, the surplus process is assumed to have no diffusion (i.e. $\sigma^2 = 0$). Unless stated otherwise, the remaining examples in this chapter will all use these baseline case assumptions. The matrix exponent characterizing our surplus process U is obtained by using Proposition 10 with $m = 0$.

The baseline case considered here is equivalent to that considered in Section 2.4 except that here, we have n investigators, investigation times are assumed to be exponentially distributed and $\varsigma(x) = \varsigma = 0.25$ rather than 0.15. We consider a CIM with a higher ς than in Section 2.4 to better illustrate the impact of the CIM in the numerical examples to follow. Also, recall that the baseline case from Section 2.4 has investigation times distributed as a combination of exponentials and hence its surplus process cannot be formulated as a MAP.

3.4.2 Performance of three numerical Laplace transform inversion algorithms

In this section, we evaluate the performance of the three numerical Laplace inversion algorithms introduced in Section 1.3, namely, the Gaver Stehfest (GS) method, Gaver-Wynn-Rho (GWR) algorithm, and Fourier-cosine series (COS) expansion. These techniques may be used to invert $\mathbf{m}_{\delta,0}(u)$ and $\mathbf{m}_{\delta,0}(u)/\delta$ in δ to obtain the density of the time to ruin and finite-time ruin probability, respectively. Recall that $\mathbf{m}_{\delta,0}(u)$ is the LT of the time to ruin and is obtained using Proposition 11 with $z = 0$. Note that the dependence of $\mathbf{m}_{\delta,0}(u)$ on δ is not straightforward and as such analytical Laplace inversion is in general a very challenging task. The reader is invited to consult Shi [2013] for more details on this topic.

We compare the performance of the three inversion techniques by using each technique to calculate the finite-time ruin probability $\psi(u,t) = P\left(T \leq t \middle| U_0 = u\right)$ for a CIM that follows the baseline case outlined in Section 3.4.1 with $n = 2$. We use the GS and GWR methods to approximate $\psi(u,t)$ by numerically inverting $\boldsymbol{m}_{\delta,0}(u)/\delta$. For the COS method, we instead approximate $\psi(u,t)$ by letting $\boldsymbol{m}_{\delta,0}(u)$ be the LT of interest and using (1.16).

For the GS method, we denote the approximant (1.10) using $\psi^{GS}(u,t)$ and let $N_1 = 13$. For the GWR method, we denote the approximant (1.14) using $\psi^{GWR}(u,t)$ and let $N_2 = 14$. For the COS method, we denote the approximant (1.16) using $\psi^{COS}(u,t)$ and let $N_3 = 30$.

Note that all three methods involve evaluating the LT at a comparable number of values for the Laplace argument (i.e., $2N_1 \approx 2N_2 + 1 \approx N_3$).

All calculations were performed in MATLAB using 100 significant digits and all three methods have comparable runtimes of approximately 40 seconds. Approximations to the finite-time ruin probabilities $\psi^{GS}(u,t)$ using the GS method, rounded to the 4th decimal place can be found in Table 3.1 for various u and t. Table 3.2 shows the magnitude of the difference $|\psi^{GS}(u,t) - \psi^{GWR}(u,t)|$ expressed in negative powers of 10.

Table 3.1: Approximation for finite-time ruin probabilities $\psi^{GS}(u,t)$

$u \setminus t$	1	5	10	20	40
0	0.4965	0.6449	0.6824	0.7036	0.7127
15	0.0201	0.1338	0.1931	0.2342	0.2537
30	0.0006	0.0261	0.0567	0.0859	0.1026
45	0.0000	0.0041	0.0147	0.0299	0.0410
60	0.0000	0.0005	0.0033	0.0098	0.0161
75	0.0000	0.0001	0.0007	0.0030	0.0062

From Table 3.2, we observe that the difference between the GS and GWR approximations have magnitude orders between 9 and 15.

To evaluate the performance of the COS method, consider Table 3.3 which shows the magnitude of the difference $|\psi^{GS}(u,t) - \psi^{COS}(u,t)|$ expressed in negative powers of 10.

From Table 3.3, we observe that the difference between the GS and COS approximations have magnitude orders between 2 and 6 which are less than those found in Table 3.2. Also, we find that the efficiency of the COS method tends to increase with u.

Table 3.2: Magnitude order of $|\psi^{GS}(u,t) - \psi^{GWR}(u,t)|$ in negative powers of 10

$u \setminus t$	1	5	10	20	40
0	9	9	10	10	9
15	11	10	10	10	11
30	11	10	10	10	11
45	12	10	10	10	10
60	14	11	10	10	10
75	15	11	11	10	10

Table 3.3: Magnitude order of $|\psi^{GS}(u,t) - \psi^{COS}(u,t)|$ in negative powers of 10

$u \setminus t$	1	5	10	20	40
0	2	2	3	3	4
15	3	4	4	4	3
30	4	5	4	4	3
45	6	5	4	4	4
60	6	4	5	4	4
75	6	5	5	4	4

If we instead calculate the approximant $\psi^{COS}(u,t)$ using $N_3 = 8000$, the runtime of the COS method increases from 40 seconds to approximately 5 hours and the difference between the GS and COS approximations are found to have magnitude orders between between 4 and 15. While these values are closer to those found in Table 3.2, it comes at the cost of much slower computation time.

We conclude that both the GS method and GWR algorithm perform well while the COS method is far less efficient. In the following numerical examples, we use the GS method when performing numerical LT inversions because the method is well-known and as shown above, is found to perform very efficiently.

3.4.3 Cost and benefit of adding investigators

In this example, we assume a cost per investigator that can be modelled as a continuous fixed rate c_{cost}. This cost can be viewed as expenses generated by a claim investigator (e.g., salary, office equipment, etc.). For n investigators, this cost effectively leads to a lowered (net) rate of premium collected equal to $c - nc_{cost}$. In the following, we vary the number of investigators n and observe the different impact on the ruin probability depending on the value of c_{cost}. We assume a CIM that follows the baseline case outlined in Section 3.4.1.

When investigation has no cost ($c_{cost} = 0$)

To begin, we consider the scenario where investigation has no cost (i.e., $c_{cost} = 0$). The ruin probabilities for this process can be found in Table 3.4 for varying n.

Table 3.4: Ruin probabilities for varying n when $c_{cost} = 0$

$n \setminus u$	0	15	30	45	60	75
0	0.8796	0.4083	0.1871	0.0857	0.0393	0.0180
1	0.7862	0.3275	0.1432	0.0627	0.0275	0.0120
2	0.7153	0.2597	0.1083	0.0454	0.0190	0.0080
3	0.6595	0.2042	0.0810	0.0325	0.0130	0.0052
4	0.6152	0.1596	0.0600	0.0231	0.0089	0.0034

From Table 3.4, we observe the ruin probability decreasing with increasing n for all u. As more investigators are added, the CIM has greater capacity to delay and deny claims. When there are no offsetting costs, this has the effect of reducing the ruin probability and this effect is confirmed by Table 3.4.

When investigation has non-zero cost ($c_{cost} = 0.35$)

Now let $c_{cost} = 0.35$. The ruin probabilities for this process can be found in Table 3.5 for varying n.

Table 3.5: Ruin probabilities for varying n and $c_{cost} = 0.35$

$n \setminus u$	0	15	30	45	60	75
0	0.8796	0.4083	0.1871	0.0857	0.0393	0.0180
1	0.8130	0.3819	0.1870	0.0916	0.0449	0.0220
2	0.7797	0.3700	0.1937	0.1017	0.0534	0.0280
3	0.7655	0.3713	0.2080	0.1172	0.0660	0.0372
4	0.7646	0.3856	0.2318	0.1403	0.0850	0.0515

Table 3.5 shows that when $c_{cost} = 0.35$, it is not true that the ruin probability decreases with n for all u (as was the case when $c_{cost} = 0$). For example, when $u \in \{15, 30\}$, the ruin probability decreases with n up to some point, and thereafter, the ruin probability begins to increase with n. When $u \in \{45, 60, 75\}$, the ruin probability begins to increase with n as soon as one investigator is added. When $u = 0$, the ruin probability decreases with n up to the addition of four investigators; with further analysis we find that the ruin probability begins to increase with n when at least five investigators have been added. We shade in gray the portion of Table 3.5 where ruin probabilities begin to increase with n for a given u.

The behaviour observed in Table 3.5 illustrates the tradeoff between the cost and benefit of adding investigators to the CIM. Recall that when an investigator is added, the CIM's delay and deny capacity is increased. Loosely speaking, if the marginal increase in the delay and deny capacity exceeds (falls below) the added cost of the investigator, the resulting ruin probability will decrease (increase).

We observe that the ruin probability decreases with n up to some point, and thereafter, the ruin probability begins to increase with n. Note that this point occurs immediately when $u \in \{45, 60, 75\}$. This suggests that the CIM experiences diminishing marginal increases in its delay and deny capacity with each added investigator such that the rising costs eventually causes the ruin probability to increase with n. Note that as one continues to add investigators to the CIM, the ruin probability will eventually equal 1 because the loading condition given by (3.12) will no longer be satisfied.

We refer to the number of investigators that minimizes the ruin probability n^* as the optimal number of investigators. For example, the optimal number of investigators when $u = 15$ and $u = 30$ is equal to $n^* = 2$ and $n^* = 1$, respectively. From Table 3.5, we observe that the optimal number of investigators is non-increasing with u. This is explained by considering an insurer with small u. When u is small, an insurer is particularly vulnerable to the event of ruin caused by an early claim. Thus, an increase in the delay and deny capacity of this insurer's CIM is especially valuable and results in a higher optimal number of investigators. The opposite is true when u is large. Note that analytical analysis to determine the optimal number of investigators appears to be a challenging problem.

In sum, when c_{cost} is positive, there is a tradeoff between the cost and benefit of adding investigators. Adding investigators can lower an insurer's ruin probability but only up to a point and thereafter, the ruin probability will begin to increase. Also, the optimal number of investigators is non-increasing with the initial surplus u.

Increasing c_{cost}

We now increase c_{cost} from 0.35 to 0.45 and consider its effect on the optimal number of investigators n^*. The ruin probabilities for a process with $c_{cost} = 0.45$ can be found in Table 3.6 for varying n.

Table 3.6: Ruin probabilities for varying n and $c_{cost} = 0.45$

$n \setminus u$	0	15	30	45	60	75
0	0.8796	0.4083	0.1871	0.0857	0.0393	0.0180
1	0.8210	0.3994	0.2021	0.1024	0.0518	0.0263
2	0.8001	0.4111	0.2303	0.1293	0.0726	0.0408
3	0.8023	0.4450	0.2770	0.1730	0.1081	0.0676
4	0.8226	0.5063	0.3517	0.2455	0.1713	0.1196

As c_{cost} increases, we expect that it becomes less advantageous to hire more investigators. This effect is confirmed by comparing Table 3.6 with Table 3.5. For example, when $u = 15$ and if $c_{cost} = 0.35$, it is optimal to hire two investigators but if $c_{cost} = 0.45$, then it is optimal to hire one investigator. Also, when $u = 30$ and if $c_{cost} = 0.35$, we find that it is optimal to hire one investigator but if $c_{cost} = 0.45$, it is optimal not to hire any investigators.

Finite-time ruin probability where there is non-zero cost ($c_{cost} = 0.325$)

Up to this point we have been considering the impact of varying n on the ultimate ruin probability. We now turn our attention to the finite-time ruin probability. More specifically, consider the probability of ruin before time 25 in Table 3.7 when $c_{cost} = 0.325$.

Note that Table 3.5 and Table 3.7 both consider when $c_{cost} = 0.325$ though Table 3.5

Table 3.7: Probability of ruin before time 25 for varying n and $c_{cost} = 0.325$

$n \setminus u$	0	15	30	45	60	75
0	0.8743	0.3849	0.1612	0.0646	0.0246	0.0089
1	0.8001	0.3456	0.1497	0.0618	0.0242	0.0089
2	0.7575	0.3174	0.1417	0.0601	0.0241	0.0091
3	0.7322	0.2976	0.1365	0.0594	0.0243	0.0093
4	0.7177	0.2847	0.1337	0.0596	0.0249	0.0097

provides the ultimate ruin probabilities while Table 3.7 provides the probability of ruin before time 25.

From Table 3.5 and Table 3.7, the number of investigators that minimizes the finite-time ruin probability is greater than the number that minimizes the ultimate ruin probability when $u \in \{0, 15, 30, 45, 60\}$. For example, when $u = 60$, hiring no investigators will minimize the ultimate ruin probability while hiring two investigators will minimize the finite-time ruin probability. This makes sense because adding investigators to a CIM will increase its capacity to delay claims. Thus, payments that are delayed to some point after the time horizon $t = 25$ are accounted for in the ultimate ruin probability but not the finite-time ruin probability. Thus, the finite-time ruin probability reflects a greater benefit from adding investigators which results in a higher optimal number of investigators. Note that when $u = 75$, the finite-time and ultimate ruin probability are minimized when no investigators are added. This is due to the smaller benefits from adding investigators when u is large which was explained earlier.

In conclusion, the optimal number of investigators will depend on the time-horizon t considered and more specifically, it will be more advantageous to hire more investigators as t decreases.

3.4.4 Effect of diffusion

In this example, we illustrate the impact of the diffusion component on the surplus process for varying σ. Recall that the diffusion component models the variability from premium, investment income, or other factors perturbing the insurer's surplus process. We let $n = 2$ such that there are two investigators in the CIM and all other features of the CIM are assumed to satisfy the baseline case. Ruin probabilities for various σ are provided in Table 3.8.

Table 3.8: Ruin probabilities for varying σ

$\sigma \setminus u$	0	15	30	45	60	75
0	0.7153	0.2597	0.1083	0.0454	0.0190	0.0080
1	1.0000	0.2638	0.1110	0.0469	0.0198	0.0084
2	1.0000	0.2760	0.1191	0.0517	0.0224	0.0097
3	1.0000	0.2958	0.1329	0.0600	0.0271	0.0122
4	1.0000	0.3227	0.1525	0.0725	0.0344	0.0164

Table 3.8 shows that the ruin probability increases with σ. This confirms our intuition since increasing the volatility of the diffusion process will increase the variability of the surplus process which leads to a greater risk of ruin. Note that when $u = 0$, a positive σ leads to ruin probabilities equal to 1.

3.5 Generalizations to the claim investigation mechanism

The CIM proposed in Section 3.2.1 assumes claims arrive according to a Poisson process and investigation times are i.i.d. exponential rv's that are independent of all else in the model. In this section, we discuss possible generalizations to the CIM while preserving the model's useful MArP formulation. The flexibility of the MArP and in particular, the flexibility in defining and developing its background process allows for a large number of possible generalizations and a considerable amount of added realism.

Recall that in the MArP formulation presented in Section 3.2.2, the background process J is given by the Markov process corresponding to the number of claims in the CIM. By more generally viewing J as corresponding to the state process of the CIM, it is not difficult to see that the model's MArP formulation can accommodate any CIM (more general than that proposed) provided its state process can be defined as a Markov process and claims arrive according to a Markov-modulated Poisson process (MMPP). Thus, a CIM modelled after any queueing system studied in the vast number of queueing problems can be accommodated, provided the queue contains an embedded Markov process and has MMPP arrivals. Such queues include but are not limited to those with: heterogeneous servers, responsive servers, phase-type service times, bulk claim arrivals/service, server vacations, parallel queues and networks of service nodes (see, e.g., [Kleinrock, 1975, Chapter 3 and 4], [He, 2014, Chapter 4], [Sericola, 2013, Section 5.8] and references therein). Indeed, the background process J developed in Section 3.2.2 for the CIM proposed in Section 3.2.1 corresponds to the system state process of the well-known $M/M/n/n+m$ queue with customer arrival rate λ_1 and service rate α (see, e.g., [Gross et al., 2011, Section 2.5]).

In the following subsections, we generalize the proposed CIM while preserving the model's

MArP formulation. We consider four generalizations to the proposed CIM, namely, (1) phase-type investigation times, (2) investigation time dependence on claim size, (3) claim investigation networks, and (4) Markov modulated Poisson claim arrivals.

For each generalization, the key objective is to determine the matrix exponent $\mathbf{F}^{\delta}(z)$ for the resulting surplus by updating the MArP formulation of the proposed CIM from Section 3.2.2. We recall that the matrix exponent characterizes the surplus process U and plays a central role in exit results via the closely-related scale matrix.

After each CIM generalization is introduced, each subsection proceeds as follows:

1. A multivariate background process $\mathbf{J}$ is defined which serves to enlarge the state space and to keep the Markovian structure of the $(U, \mathbf{J})$ process.
2. The infinitesimal generator $\mathbf{Q}$ is identified.
3. The matrix $\mathbf{G}(z)$ is identified.
4. The matrix exponent $\mathbf{F}^{\delta}(z)$ for the resulting surplus process is obtained by outlining the required modifications to (3.4).
5. The joint LT of the time and deficit at ruin is obtained by outlining the required modifications to (3.7).

For the subsections relating to investigation time dependence on claim size and claim investigation networks, a numerical example is provided to illustrate the impact of the generalization on the ruin probability.

The following notation will be used in the remainder of this section. We use $\mathbf{I}_n$ to denote an $n \times n$ identity matrix and $\mathbb{1}_n$ to denote a column vector of size n with all entries equal to 1. We denote by $\mathbf{e}_{n,h}$ the row vector of size n with hth entry equal to 1 and all other entries equal to 0. Suppose we have two matrices $\mathbf{A} = (a_{i,j})$ of size $k \times l$ and $\mathbf{B} = (b_{i,j})$ of size $m \times n$. The Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \otimes \mathbf{B}$ and defined by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,l}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,l}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1}\mathbf{B} & a_{k,2}\mathbf{B} & \cdots & a_{k,l}\mathbf{B} \end{bmatrix}.$$

3.5.1 Phase-type investigation times

In this section, we generalize the proposed CIM described in Section 3.2.1 by assuming that investigation times are distributed according to a phase-type distribution $(\boldsymbol{\beta}, \mathbf{T})$ of order s with $\boldsymbol{\beta}\mathbb{1}_s = 1$ and we set $\boldsymbol{T}_0 = -\boldsymbol{T}\mathbb{1}_s$. We continue to assume that investigation times are i.i.d. and independent of all else in the model. All other features of the CIM are assumed to satisfy the proposed CIM of Section 3.2.1. In what follows, we superimpose this generalized CIM onto the surplus process and present the MArP formulation of the resulting surplus process. We remark that the generalized CIM is modelled after the $M/PH/n/n+m$ queue studied in queueing theory. A treatment of the $M/PH/n$ queue where infinite capacity is assumed can be found in Sericola [2013].

Note that the class of phase-type distributions is dense in the sense of weak convergence for all distributions with positive support and is therefore a very general distributional assumption. Special cases of phase-type distributions include the exponential, mixture of exponential, Erlang, and Coxian distributions. A more detailed account of phase-type distributions can be found in [Asmussen and Albrecher, 2010, Chapter IX].

1. Defining background process J

We begin by defining the generalized background process and its state space. For $j \in \{1, \ldots, s\}$, define $R_j = \{R_{j,t}\}_{t \geq 0}$ where $R_{j,t}$ corresponds to the number of claims under investigation in phase j at time t. The generalized background process is defined by the multivariate Markov process $\mathbf{J} = \{(J_t, R_{1,t}, \ldots, R_{s,t})\}_{t \geq 0}$ where $\mathbf{J}_t = (J_t, R_{1,t}, \ldots, R_{s,t})$ corresponds to not only the number of claims in the CIM but also, the number of claims under investigation in each phase at time t. Thus, the generalized state space is given by

$$E = \bigcup_{i=0}^{n+m} C_i,$$

where for $i > 0$,

$$C_i = \{(i, r_1, \ldots, r_s)\,,\, 0 \leq r_h \leq n, \text{ for all } h \in \{1, \ldots, s\} \text{ and } r_1 + \cdots + r_s = min(n, i)\},$$

and to simplify the notation in what follows, we let $C_0 = (0, 0, \ldots, 0)$, a zero vector of size $s+1$. Note that when at least one of the r_h's is non-zero, then $i > 0$. Also, the cardinality of C_i when $i > 0$ depends on i and is equal to $|C_i| = \binom{s+min(n,i)-1}{min(n,i)}$ (see, e.g., [Feller, 1971, Section 5]).

2. Identifying infinitesimal generator Q

We now determine the transition rates of $\mathbf{J}$. Let $\mathbf{r} = (r_1, \ldots, r_s)$ with $r_1 + \cdots + r_s = min(n, i)$ and $\mathbf{r}' = (r'_1, \ldots, r'_s)$ with $r'_1 + \cdots + r'_s = \min(n, i')$. For $(i, \mathbf{r}), (i', \mathbf{r}') \in E$, denote the transition rate of $\mathbf{J}$ from state $(i, \mathbf{r})$ to $(i', \mathbf{r}')$ by $q_{(i,\mathbf{r}),(i',\mathbf{r}')}$. All transition rates of $\mathbf{J}$ starting from state $(i, \mathbf{r}) \in C_i$ are given in the following:

1. **Claims entering the CIM**: If the CIM is not at capacity, claims enter the CIM at rate λ_1. Suppose an entering claim finds i claims in the CIM. If $0 \leq i < n$, then not

all servers are busy and the claim immediately begins investigation at phase h with probability β_h for $h \in \{1, \ldots, s\}$. Otherwise, if $n \leq i < n+m$, then all servers are busy and the claim joins the queue. Thus, if $0 \leq i < n$,

$$q_{(i,\mathbf{r}),\left(i+1,\mathbf{r}+\mathbf{e}_{s,h}\right)} = \lambda_1 \beta_h, \qquad \text{for all } 1 \leq h \leq s,$$

and if $n \leq i < n+m$,

$$q_{(i,\mathbf{r}),(i+1,\mathbf{r})} = \lambda_1.$$

2. **Claims departing the CIM:** Suppose the CIM is not empty and furthermore, that there are i claims in the CIM. If there are r_g claims at phase g of investigation, then a claim departs from the CIM from phase g at rate $r_g \cdot (\boldsymbol{T}_0)_g$ for $g \in \{1, \ldots, s\}$. Furthermore, if the claim makes its departure while there is at least one claim waiting in the queue (i.e., $n < i \leq n+m$), then the claim at the front of the queue begins investigation at phase h with probability β_h where $h \in \{1, \ldots, s\}$. Thus, if $0 < i \leq n$,

$$q_{(i,\mathbf{r}),(i-1,\mathbf{r}-\mathbf{e}_{s,g})} = r_g \cdot (\mathbf{T}_0)_g, \qquad \text{for all } 1 \leq g \leq s,$$

and if $n < i \leq n+m$,

$$q_{(i,\mathbf{r}),\left(i-1,\mathbf{r}-\mathbf{e}_{s,g}+\mathbf{e}_{s,h}\right)} = r_g \cdot (\mathbf{T}_0)_g \, \beta_h, \qquad \text{for all } 1 \leq g, h \leq s \text{ and } g \neq h,$$

and

$$q_{(i,\mathbf{r}),(i-1,\mathbf{r})} = \sum_{g=1}^{s} r_g \cdot (\mathbf{T}_0)_g \, \beta_g. \tag{3.14}$$

The summation over $g \in \{1, \ldots, s\}$ in (3.14) is explained by considering all the ways that $\mathbf{J}$ transitions from state $(i, \mathbf{r})$ to state $(i-1, \mathbf{r})$. That is, such a transition occurs whenever a claim departs the CIM from phase g and a waiting claim enters investigation at the same phase g, for any $g \in \{1, \ldots, s\}$.

3. **For transitions with a fixed number of claims in the CIM:** A transition in $\mathbf{J}$ with a fixed number of claims in the CIM occurs when a claim undergoing investigation transitions between phases of investigation. If there are r_g claims at phase g of

investigation, a claim transitions from phase g to phase h with rate $r_g \cdot (\mathbf{T})_{g,h}$. Thus, if $0 < i \leq n+m$,

$$q_{(i,\mathbf{r}),\left(i,\mathbf{r}-\mathbf{e}_{s,g}+\mathbf{e}_{s,h}\right)} = r_g \cdot (\mathbf{T})_{g,h}, \qquad \text{for all } 1 \leq g,h \leq s \text{ and } g \neq h.$$

No other transitions in $\mathbf{J}$ are possible.

Using the transition rates for $\mathbf{J}$ described above, we are able to determine the infinitesimal generator $\mathbf{Q}$. The diagonal elements of $\mathbf{Q}$ are found such that its rows sum to zero. According to the transition rates given above, it follows that

$$q_{(i,\mathbf{r}),(i,\mathbf{r})} = \begin{cases} -\lambda_1, & i = 0, \\ -\lambda_1 + \sum_{j=1}^{s} r_j \cdot (\mathbf{T})_{j,j}, & 0 < i < n+m, \\ \sum_{j=1}^{s} r_j \cdot (\mathbf{T})_{j,j}, & i = n+m, \end{cases}$$

for all $(i,\mathbf{r}) \in E$. The matrix $\mathbf{Q}$ has the same structure as (3.25). However, according to the transition rates described above, we do not have simple expressions for the submatrices $Q_{i,i'}$ like we do in Section 3.5.4.

3. Identifying matrix $\mathbf{G}(z)$

Next, we consider the LT of payments triggered by transitions in $\mathbf{J}$ to determine the matrix $\mathbf{G}(z)$. Let $g(z)_{(i,\mathbf{r}),(i',\mathbf{r}')}$ denote the LT of the claim payment triggered when $\mathbf{J}$ transitions from state $(i,\mathbf{r})$ to $(i',\mathbf{r}')$ where $(i,\mathbf{r})$, $(i',\mathbf{r}') \in E$. Recall that a payment is triggered only at transitions in $\mathbf{J}$ corresponding to a claim completing investigation and departing the CIM and that the payment sizes have LT $\widetilde{k}(z)$. Thus, if $1 \leq i \leq n$,

$$g(z)_{(i,\mathbf{r}),(i-1,\mathbf{r}-\mathbf{e}_{s,g})} = \widetilde{k}(z), \qquad \text{for all } 1 \leq g,h \leq s,$$

and if $n < i \leq n+m$,

$$g(z)_{(i,\mathbf{r}),\left(i-1,\mathbf{r}-\mathbf{e}_{s,g}+\mathbf{e}_{s,h}\right)} = \widetilde{k}(z), \qquad \text{for all } 1 \leq g,h \leq s \text{ and } g \neq h,$$

and

$$g(z)_{(i,\mathbf{r}),(i-1,\mathbf{r})} = \widetilde{k}(z).$$

Since all other transitions in $\mathbf{J}$ do not trigger payments (which is mathematically equivalent to assuming such transitions trigger a payment of size 0), all other entries in $\mathbf{G}(z)$ are equal to 1. The matrix $\mathbf{G}(z)$ has the same structure as (3.27). However, we do not have simple expressions for the submatrices like we do in Section 3.5.4.

4. Matrix exponent $\mathbf{F}^{\delta}(z)$

Given the MArP formulation provided thus far, we are ready to give the matrix exponent of the surplus process.

Proposition 12. For phase-time investigation times, the resulting surplus process U is characterized by the matrix exponent given by (3.4) with $\mathbf{Q}$ and $\mathbf{G}(z)$ as identified in this section. Also, $\boldsymbol{\Lambda}$ and $\mathbf{P}(z)$ are respectively replaced by

$$\boldsymbol{\Lambda} = diag\{\lambda_0,\, \ldots,\, \lambda_0, \underbrace{\lambda, \ldots, \lambda}_{|C_{n+m}|}\},$$

a diagonal matrix where all diagonal elements are equal to λ_0 except for the last $|C_{n+m}|$ diagonal elements which are equal to λ, and

$$\mathbf{P}(z) = diag\{\widetilde{p}_0(z), \ldots, \widetilde{p}_0(z), \underbrace{\widetilde{p}(z) \ldots, \widetilde{p}(z)}_{|C_{n+m}|}\},$$

a diagonal matrix where all diagonal entries are equal to $\widetilde{p}_0(z)$ except for the last $|C_{n+m}|$ diagonal elements which are equal to $\widetilde{p}(z)$.

5. Joint LT of the time and deficit at ruin

Lastly, we note that the joint LT of the time and deficit at ruin is given by (3.7) and (3.8) using the matrix exponent $\mathbf{F}^{\delta}(z)$ identified in Proposition 12. No further modifications are necessary because V_t, which we recall is defined as the total (future) payment amount of all claims in the CIM at time t (if any), is still given by (3.5).

3.5.2 Investigation time dependence on claim size

In this section, we introduce a dependence structure into the CIM proposed in Section 3.2.1. More precisely, we consider a dependence structure where investigation times are distributed as a mixture of exponentials.

We continue to assume that investigation times are i.i.d. exponential rv's except that the rate of investigation is now assumed to depend on the size of the claim under investigation. More precisely, a claim under investigation of size x will undergo investigation with duration having df $\sum_{j=1}^{d} f_j(x)\,(1 - e^{-\alpha_j t})$, for $t > 0$. That is, conditional on the claim size x, a claim under investigation is investigated for an exponential amount of time at rate $\alpha_j > 0$ (i.e., having mean $1/\alpha_j$) with probability $f_j(x) > 0$ for $j \in \{1, 2, \ldots, d\}$. For simplicity, we will say that an investigation time is $Exp(\alpha_j)$ if it is exponentially distributed with mean $1/\alpha_j$. We assume $\sum_{j=1}^{d} f_j(x) = 1$ for all $x > 0$. Investigation times are assumed to remain independent of any other rv's in the risk model and all other features of the CIM are assumed to satisfy the proposed CIM given in Section 3.2.1. For a random claim selected for investigation, we have that investigation times follow a mixture of exponential distributions where the jth exponential has rate α_j with corresponding mixing weight

$$w_j \equiv \int_0^\infty f_j(x) p_1(x) dx,$$

for $j \in \{1, \ldots, d\}$. Hence, the resulting CIM is said to behave as a $M/H_d/n/n+m$ queue studied in queueing theory (see, e.g. [Kleinrock, 1975, Section 4.7] for a treatment of the $M/H_d/1$ queue where a single server and infinite capacity is assumed). Note that if $d = 1$, we recover the proposed model from Section 3.2.1 with α replaced by α_1.

Remark 8. Note that in practice it is often the case that larger claims experience longer investigation times on average. This relationship can be modelled under our proposed dependence structure by having larger claims be more likely to have smaller rates of investigation. Supposing $\alpha_1 < \cdots < \alpha_d$, we can, for example, choose $\{f_j(x)\}_{j=1}^d$ such that for large (small) x, $f_j(x)$ is decreasing (increasing) in j.

It is well-known that the mixture of exponential distributions is included in the class of phase-type distributions (see, e.g., [Latouche and Ramaswami, 1999, Section 2.2]). Thus, the investigation times under our proposed dependence structure are phase-type $(\boldsymbol{\beta}, \mathbf{T})$ of order d with $\boldsymbol{\beta} = (w_1, \ldots, w_d)$,

$$\mathbf{T} = -diag\{\alpha_j\}_{j=1}^d,$$

and $\mathbf{T}_0 = (\alpha_1, \ldots, \alpha_d)^T$ and the resulting surplus process can be formulated as a special case of that in Section 3.5.1 where phase-type investigation times were discussed.

1. Defining background process J

We obtain the background process from Section 3.5.1 by setting $s = d$ and letting $R_{j,t}$ correspond to the number of claims having $Exp(\alpha_j)$ investigaiton time at time t.

2. Identifying infinitesimal generator Q

We obtain the infinitesimal generator $\mathbf{Q}$ from Section 3.5.1 by setting $s = d$, $(\mathbf{T}_0)_g = \alpha_g$, $(\mathbf{T})_{g,h} = 0$ when $g \neq h$, and $(\mathbf{T})_{j,j} = -\alpha_j$ for all $1 \leq g, h, j \leq d$.

3. Identifying matrix $\mathbf{G}(z)$

Next, we consider the LT of payments triggered by transitions in $\mathbf{J}$ to determine the matrix $\mathbf{G}(z)$. We define

$$\widetilde{k}_j(z) = \frac{1}{w_j} \int_0^\infty \widetilde{l}_x(z) f_j(x) p_1(x) dx,$$

for $j \in \{1, \dots, d\}$, corresponding to the LT of the payment size following investigation given the investigation time was $Exp(\alpha_j)$. Thus, similar to Section 3.5.1, if $0 < i \leq n$,

$$g(z)_{(i,\mathbf{r}),(i-1,\mathbf{r}-\mathbf{e}_{s,g})} = \widetilde{k}_g(z), \qquad \text{for all } 1 \leq g \leq s \text{ where } r_g \geq 1,$$

and if $n < i \leq n+m$,

$$g(z)_{(i,\mathbf{r}),\left(i-1,\mathbf{r}-\mathbf{e}_{s,g}+\mathbf{e}_{s,h}\right)} = \widetilde{k}_g(z), \qquad \text{for all } 1 \leq g,h \leq s \text{ and } g \neq h \text{ where } r_g \geq 1$$

and

$$g(z)_{(i,\mathbf{r}),(i-1,\mathbf{r})} = \sum_{g=1}^{d} \left(\frac{r_g \alpha_g}{\sum_{l=1}^{d} r_l \alpha_l} \right) \widetilde{k}_g(z). \tag{3.15}$$

The summation over g in (3.15) is explained by considering all possible transitions of $\mathbf{J}$ from state $(i, \mathbf{r})$ to state $(i-1, \mathbf{r})$. This transition occurs whenever a claim departs the CIM after having an $Exp(\alpha_g)$ investigation time and a waiting claim begins investigation (also) with an $Exp(\alpha_g)$ investigation time, for any $g \in \{1, \dots, d\}$. A claim departs the CIM after having an $Exp(\alpha_g)$ investigation time with probability $r_g\alpha_g / \sum_{l=1}^{d} r_l \alpha_l$ and triggers a payment with LT $\widetilde{k}_g(z)$. Thus, the LT of the payment triggered by this transition in $\mathbf{J}$ involves an expectation over all the possible rates of investigation for the departing claim.

Since all other transitions in $\mathbf{J}$ do not trigger payments (which is mathematically equivalent to assuming such transitions trigger a payment of size 0), all other entries in $\mathbf{G}(z)$ are equal to 1. The matrix $\mathbf{G}(z)$ has the same structure as (3.27). However, we do not have simple expressions for the submatrices like we do in Section 3.5.4.

4. Matrix exponent $\mathbf{F}^{\delta}(z)$

Given the MArP formulation provided thus far, we are ready to give the matrix exponent of the surplus process.

Proposition 13. Introducing the proposed dependence structure results in a surplus process U characterized by a matrix exponent given by (3.4) with $\mathbf{Q}$ and $\mathbf{G}(z)$ as identified in this section. Also, $\mathbf{\Lambda}$ and $\mathbf{P}(z)$ are respectively replaced by

$$\mathbf{\Lambda} = diag\{\lambda_0,\, \dots,\, \lambda_0, \underbrace{\lambda, \dots, \lambda}_{|C_{n+m}|}\},$$

a diagonal matrix where all diagonal elements are equal to λ_0 except for the last $|C_{n+m}|$ diagonal elements which are equal to λ, and

$$\mathbf{P}(z) = diag\{\widetilde{p}_0(z), \dots, \widetilde{p}_0(z), \underbrace{\widetilde{p}(z) \dots, \widetilde{p}(z)}_{|C_{n+m}|}\},$$

a diagonal matrix where all diagonal entries are equal to $\widetilde{p}_0(z)$ except for the last $|C_{n+m}|$ diagonal elements which are equal to $\widetilde{p}(z)$.

Remark 9. The risk model from Chapter 2 with the added restriction that the investigation time density (2.3) is of a mixture form (i.e., all mixing weights $\omega_i > 0$) is obtained by the CIM discussed in this section by setting $n = 1$ and $m = 0$. Thus, the resulting surplus process has matrix exponent given by (3.4) with $\mathbf{Q}$ replaced by

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{0,0} & \mathbf{Q}_{0,1} \\ \mathbf{Q}_{1,0} & \mathbf{Q}_{1,1} \end{bmatrix},$$

where $\mathbf{Q}_{0,0} = -\lambda_1$, $\mathbf{Q}_{0,1} = \lambda_1\,(w_1, \dots, w_d)$, $\mathbf{Q}_{1,0} = (\alpha_1, \dots, \alpha_d)^T$, and $\mathbf{Q}_{1,1} = -diag\{\alpha_j\}_{j=1}^{d}$, and the matrix $\mathbf{G}(z)$ is given by

$$\mathbf{G}(z) = \begin{bmatrix} 0 & 0 \\ \mathbf{G}_{1,0}(z) & 0 \end{bmatrix},$$

where $\mathbf{G}_{1,0}(z) = \left(\widetilde{k}_1(z), \dots, \widetilde{k}_d(z)\right)^T$.

5. Joint LT of the time and deficit at ruin

Lastly, to obtain the joint LT of the time and deficit at ruin matrix

$$\mathbf{m}_{\delta,z}(u) \equiv E_u\left[e^{-\delta T - z(|U_T| + \eta V_T)} 1_{\{T<\infty\}};\, \mathbf{J}_T\right],$$

some modifications to (3.7) are required which we outline below. Modifications are required because V_t is no longer given by (3.5) under the proposed dependence structure introduced in this section.

Recall that V_t is the total (future) payment amount of all claims in the CIM at time t (if any). Let $\mathbf{r} = (r_1, \ldots, r_d)$ with $r_1 + \cdots + r_d = min(n, i)$ and $\mathbf{r}' = (r'_1, \ldots, r'_d)$ with $r'_1 + \cdots + r'_d = \min(n, i')$. In what follows, for $(i, \mathbf{r}), (i', \mathbf{r}') \in E$, we write $(E_u\left[\,\cdot\,; J_T\right])_{(i,\mathbf{r}),(i',\mathbf{r}')}$ to denote

$$E\left[\,\cdot\, 1_{\{\mathbf{J}_T=(i',\mathbf{r}')\}}\middle|\, U_0 = u, \mathbf{J}_0 = (i, \mathbf{r})\right].$$

It follows that

$$\begin{aligned}(\mathbf{m}_{\delta,z}(u))_{(i,\mathbf{r}),(i',\mathbf{r}')} &= E\left[\, e^{-\delta T - z(|U_T|+\eta V_T)} 1_{\{\mathbf{J}_T=(i',\mathbf{r}'),\, T<\infty\}} \,|U_0 = u, \mathbf{J}_0 = (i,\mathbf{r})\right] \\ &= \left(\boldsymbol{\gamma}_{\delta,z}(u)\right)_{(i,\mathbf{r}),(i',\mathbf{r}')} E\left[e^{-z\eta V_T}|\mathbf{J}_T = (i', \mathbf{r}')\right], \end{aligned} \tag{3.16}$$

where $\boldsymbol{\gamma}_{\delta,z}(u) := E_u\left[e^{-\delta T - z|U_T|} 1_{\{T<\infty\}}; \mathbf{J}_T\right]$ is given by (3.8) using the matrix exponent $\mathbf{F}^{\delta}(z)$ identified in Proposition 13.

In the following, we consider V_T given $\mathbf{J}_T = (i', \mathbf{r}')$ where $(i', \mathbf{r}') \in E$. Note that if $\mathbf{J}_T = (0, 0, \ldots, 0)$, a zero vector of size $d + 1$, then the CIM is empty at the time of ruin and $V_T = 0$ such that $E\left[e^{-z\eta V_T}|\mathbf{J}_T = (0, 0, \ldots, 0)\right] = 1$.

Recall that a claim with an $Exp(\alpha_g)$ investigation time results in a payment amount having LT $\widetilde{k}_g(z)$ for $g \in \{1, \ldots, d\}$. Thus, for $g \in \{1, \ldots, d\}$, if there are r'_g claims being investigated for an $Exp(\alpha_g)$ time, their total payments have LT $\left(\widetilde{k}_g(z)\right)^{r'_g}$.

If $n < i' \leq n+m$, then there are $(i' - min(i', n))$ claims waiting in the queue. Recall that a claim waiting in the queue has probability w_g of having an $Exp(\alpha_g)$ investigation time for $g \in \{1, \ldots, d\}$. Thus, the total payments from claims waiting in the queue have LT given by $\left(\sum_{g=1}^{d} w_g \widetilde{k}_g(z)\right)^{i'-min(i',n)}$.

Thus, it follows that

$$E\left[e^{-z\eta V_T} | J_T = (i', \mathbf{r}')\right] = \left(\sum_{g=1}^{d} \left[\widetilde{k}_g(z\eta)\right]^{r'_g}\right) \left(\sum_{g=1}^{d} w_g \widetilde{k}_g(z\eta)\right)^{i'-min(n,i')}. \tag{3.17}$$

Thus, the joint LT of the time and deficit at ruin under the proposed dependence structure is given by substituting (3.17) into (3.16).

6. Numerical example

In this example, we compare the ruin probability of a CIM following the baseline case outlined in Section 3.4.1 with $n = 1$ before and after a dependence structure between the investigation time and claim size is introduced. Recall that the baseline case with $n = 1$ has exponential investigation times having mean $1/\alpha = 5$ that are independent of all other rv's in the model. The ruin probability for this process can be found in Table 3.9.

The dependence structure we introduce follows that discussed in this section and assumes $d = 2$, $f_1(x) = e^{-\beta x}$, $f_2(x) = 1 - e^{-\beta x}$, $1/\alpha_1 = 2$, and $1/\alpha_2 = 10.5714$. Thus, with probability $w = w_1 = \widetilde{p_1}(\beta)$, a random claim selected for investigation will have a "short" mean investigation time of $1/\alpha_1 = 2$; otherwise, with probability $1 - w$, it will have a "long" mean investigation time of $1/\alpha_2 = 10.5714$. We set $\beta = 0.1508$ such that $w = \widetilde{p_1}(\beta) = 0.65$. All remaining assumptions follow the baseline case assumptions outlined in Section 3.4.1.

Note that $f_1(x)$ ($f_2(x)$) is decreasing (increasing) in x. Thus, larger claims are more likely to have the long mean investigation time of $1/\alpha_2 = 10.5714$ and smaller claims are more likely to have the short mean investigation time of $1/\alpha_1 = 2$.

The matrix exponent for this process is given by Proposition 13 and its ruin probabilities for various u are also given in Table 3.9.

Table 3.9: Ruin probabilities with and without a dependence structure

Investigation times $\setminus\, u$	0	15	30	45	60	75
Independent	0.7862	0.3275	0.1432	0.0627	0.0275	0.0120
Dependent	0.7821	0.3190	0.1381	0.0602	0.0263	0.0115

From Table 3.9, we observe that the ruin probability is lower with the introduction of the dependence structure. Note that under both processes, the mean investigation time for a random claim selected for investigation is equal to 5. However, when investigation times are dependent on claim sizes, payment of larger claims tend to be delayed by longer investigation times - which is not the case when investigation times are independent of claim sizes. Since insurers have increased exposure to ruin from larger claims, longer delays associated with larger claims will lower the risk of ruin. This effect is confirmed in Table 3.9.

3.5.3 Claim investigation networks

Up to this point, a claim entering a CIM is assumed to undergo only one stage of investigation before it is paid. However, more realistically, the claims handling process may involve multiple stages. For example, a first stage may be dedicated to determining a claim's eligibility for coverage. An eligibility specialist may screen key documentation and determine whether the claim complies with policy terms. In subsequent stages, eligible claims may be assessed in greater detail, additional information may be collected (e.g., for disability insurance, physician and employment records are typically required), and fair payment amounts are determined. In the final stages, claims may undergo administrative processing before payments are made to policyholders.

In this section, we assume a network-based CIM consisting of n nodes where each node represents a stage in the claims handling process. At each node, we say that a claim receives *service* which might include, e.g., eligibility review, investigation, and administrative processing. We remark that the CIM proposed in Section 3.2.1 consists of one node that is assumed to encompass all stages of the claims handling process. Furthermore, the network-based CIM is assumed to satisfy the following:

1. Claims enter the CIM at node j ($j \in \{1, \dots, n\}$) with probability $\gamma_j \geq 0$ where $\sum_{j=1}^{n} \gamma_j = 1$.

2. Each node consists of a single server having i.i.d. exponential service times and a single queue where claims enter service according to a FCFS discipline. We let μ_j be the service rate at node j. All nodes are assumed to be independent.

3. Having completed service at node j, a claim will either transfer to node h with probability $p_{j,h} \geq 0$ for $h \in \{1, \dots, n\}$ or depart from the CIM with probability $p_{j,0} \geq 0$ independently of all else in the model. We assume $\sum_{h=0}^{n} p_{j,h} = 1$ for all $j \in \{1, \dots, n\}$. A claim of size x departing from the CIM at node j is assumed to trigger a (possibly zero) payment with LT $\widetilde{l}_{x,j}(z) \leq 1$. Note that a zero payment is triggered if, for example, the claim is determined to be fraudulent and is therefore denied.

4. The CIM is said to be at *full capacity* when the total number of claims (waiting in a queue or being serviced) across all n nodes is equal to $n + m$. When the CIM is at full capacity, all arriving claims will not enter the CIM and be immediately paid in entirety. Thus, each node has capacity $n + m$ with the added restriction that the total number of claims across all nodes does not exceed $n + m$ and therefore, it is possible for $n + m$ claims to be at a single node provided all other nodes are empty.

All other features of the network-based CIM are assumed to be given by the proposed CIM. From the four characteristics listed above, we remark that a claim is permitted to enter the CIM at any node, traverse among nodes where it is allowed to avoid and revisit nodes, and depart from any node to trigger a (possibly zero) payment. Also, claims do not necessary flow through nodes in the same order.

In queueing problems, a queueing network having the first three of the four characteristics listed above is referred to as an open Jackson network with single-server nodes. Queueing networks is a large area of research in queueing theory and interested readers are referred to, e.g., [Kleinrock, 1975, Section 4.8]. Note that the special case where claims always arrive at node 1 and must flow through all nodes in a single direction before departure at node n is obtained by setting $\gamma_1 = 1$, $p_{i,i+1} = 1$ for $1 \leq i \leq n-1$, and $p_{n,0} = 1$. In queueing problems, these network queues are called *series* or *tandem* queues.

Remark 10. Note that this network-based CIM can be generalized to allow for multi-server nodes while preserving the model's MArP formulation by further enlarging the state space.

In what follows, we superimpose the network-based CIM onto the surplus process which preserves the model's MArP formulation. We present the MArP formulation and determine the matrix exponent characterizing the resulting surplus process.

1. Defining background process J

We begin by defining the generalized background process and its state space. Let us define $G_j = \{G_{j,t}\}_{t\geq 0}$ such that $G_{j,t}$ corresponds to the number of claims at node j at time t for $j \in \{1, \ldots, n\}$. The generalized background process is chosen to be the multivariate Markov process $\mathbf{J} = \{(J_t, G_{1,t}, \ldots, G_{n,t})\}_{t\geq 0}$ where $\mathbf{J}_t = (J_t, G_{1,t}, \ldots, G_{n,t})$ corresponds to not only the number of claims in the CIM but also, the number of claims at each node at time t.

Thus, the generalized state space is given by

$$E = \bigcup_{i=0}^{n+m} C_i,$$

where for $i > 0$,

$$C_i = \{(i, g_1, \ldots, g_n)\,,\, 0 \le g_h \le i, \text{ for all } h \in \{1, \ldots, k\} \text{ and } g_1 + \cdots + g_n = i\},$$

and to simplify the notation in the following, we let $C_0 = (0, 0, \ldots, 0)$, a zero vector of size $n + 1$. Note that when at least one of the g_h's is non-zero, then $i > 0$. Also, the cardinality of C_i when $i > 0$ depends on i and is equal to $|C_i| = \binom{n+i-1}{i}$ (see, e.g., [Feller, 1971, Section 5]).

2. Identifying infinitesimal generator Q

We now determine the transition rates of $\mathbf{J}$. Let $\mathbf{g} = (g_1, \ldots, g_n)$ with $g_1 + \cdots + g_n = i$ and $\mathbf{g}' = (g'_1, \ldots, g'_n)$ with $g'_1 + \cdots + g'_n = i'$. For $(i, \mathbf{g}), (i', \mathbf{g}') \in E$, denote the transition rate of $\mathbf{J}$ from state $(i, \mathbf{g})$ to $(i', \mathbf{g}')$ by $q_{(i,\mathbf{g}),(i',\mathbf{g}')}$. All transition rates for $\mathbf{J}$ starting from state $(i, \mathbf{g}) \in C_i$ are given as follows:

1. **Claims entering the CIM**: Suppose the CIM is not at capacity. Claims enter the CIM at rate λ_1 and begin service at node j with probability γ_j for $j \in \{1, \ldots, n\}$. Thus, if $0 \le i < n + m$,

$$q_{(i,\mathbf{g}),(i+1,\mathbf{g}+\mathbf{e}_{n,j})} = \lambda_1 \gamma_j, \qquad \text{for all } 1 \le j \le n.$$

2. **Claims departing the CIM**: Suppose the CIM is not empty and furthermore, that there is a claim undergoing service at node j for $j \in \{1, \ldots, n\}$, i.e. $g_j \ge 1$. The claim completes service with rate μ_j and subsequently departs the CIM with probability $p_{j,0}$. Thus, if $0 < i \le n + m$,

$$q_{(i,\mathbf{g}),(i-1,\mathbf{g}-\mathbf{e}_{n,j})} = \mu_j p_{j,0}, \qquad \text{if } g_j \ge 1 \text{ for all } 1 \le j \le n.$$

3. **Transitions with fixed number of claims in the CIM**: A transition in $\mathbf{J}$ with a fixed number of claims in the CIM occurs when a claim completes service at one node and transfers to another node (or revisits the same node). Suppose there is a claim undergoing investigation at node j for $j \in \{1, \ldots, n\}$, i.e. $g_j \geq 1$, it completes service at rate μ_j and subsequently transfers to node h with probability $p_{j,h}$ for $h \in \{1, \ldots, n\}$. Thus, if $1 < i \leq n+m$,

$$q_{(i,\mathbf{g}),\left(i,\mathbf{g}-\mathbf{e}_{n,j}+\mathbf{e}_{n,h}\right)} = \mu_j p_{j,h}, \qquad \text{if } g_j \geq 1 \text{ for all } 1 \leq j, h \leq n.$$

No other transitions in $\mathbf{J}$ are possible.

Using the transition rates for $\mathbf{J}$ described above, we are able to determine the infinitesimal generator $\mathbf{Q}$. The diagonal elements of $\mathbf{Q}$ are found such that its rows sum to zero. According to the transition rates given above, it follows that

$$q_{(i,\mathbf{g}),(i,\mathbf{g})} = \begin{cases} -\lambda_1, & i = 0, \\ -\lambda_1 - \sum_{j=1}^{n} \mu_j \left(1 - p_{j,j}\right) I_{\{g_j > 0\}}, & 0 < i < n+m, \\ -\sum_{j=1}^{n} \mu_j \left(1 - p_{j,j}\right) I_{\{g_j > 0\}}, & i = n+m, \end{cases}$$

for all $(i, \mathbf{g}) \in E$. The matrix $\mathbf{Q}$ has the same structure as (3.25). However, according to the transition rates described above, we do not have simple expressions for the submatrices like we do in Section 3.5.4.

3. Identifying matrix $\mathbf{G}(z)$

Next, we consider the LT of payments triggered by transitions in $\mathbf{J}$ to determine the matrix $\mathbf{G}(z)$. Let $g(z)_{(i,\mathbf{g}),(i',\mathbf{g}')}$ denote the LT of the claim payment triggered when $\mathbf{J}$ transitions from state $(i, \mathbf{g})$ to $(i', \mathbf{g}')$ where $(i, \mathbf{g})$, $(i', \mathbf{g}') \in E$. Recall that a payment is triggered only at transitions in $\mathbf{J}$ corresponding to a claim completing investigation and departing the CIM.

Also note that the payment size triggered by a claim departing the CIM from node j has LT

$$\widetilde{k}_j(z) = \int_0^\infty \widetilde{l}_{x,j}(z) p_1(x) dx.$$

Thus, if $0 < i \le n + m$,

$$g_{(i,\mathbf{g}),(i-1,\mathbf{g}-\mathbf{e}_{n,j})}(z) = \widetilde{k}_j(z), \qquad \text{if } g_j \ge 1 \text{ for all } 1 \le j \le n,$$

Since all other transitions in $\mathbf{J}$ do not trigger payments (which is mathematically equivalent to assuming such transitions trigger a payment of size 0), all other entries in $\mathbf{G}(z)$ are equal to 1. The matrix $\mathbf{G}(z)$ has the same structure as (3.27). However, we do not have simple expressions for the submatrices like we do in Section 3.5.4.

4. Matrix exponent $\mathbf{F}^\delta(z)$

Given the MArP formulation provided thus far, we are ready to give the matrix exponent of the surplus process.

Proposition 14. For the network-based CIM proposed in this section, the resulting surplus process U is characterized by the matrix exponent given by (3.4) with $\mathbf{Q}$ and $\mathbf{G}(z)$ as identified in this section. Also, $\mathbf{\Lambda}$ and $\mathbf{P}(z)$ are respectively replaced by

$$\mathbf{\Lambda} = diag\{\lambda_0, \ldots, \lambda_0, \underbrace{\lambda, \ldots, \lambda}_{|C_{n+m}|}\},$$

a diagonal matrix where all diagonal elements are equal to λ_0 except for the last $|C_{n+m}|$ diagonal elements which are equal to λ, and

$$\mathbf{P}(z) = diag\{\widetilde{p}_0(z), \ldots, \widetilde{p}_0(z), \underbrace{\widetilde{p}(z) \ldots, \widetilde{p}(z)}_{|C_{n+m}|}\},$$

a diagonal matrix where all diagonal entries are equal to $\widetilde{p}_0(z)$ except for the last $|C_{n+m}|$ diagonal elements which are equal to $\widetilde{p}(z)$.

5. Joint LT of the time and deficit at ruin

Lastly, to obtain the joint matrix LT of the time and deficit at ruin

$$\mathbf{m}_{\delta,z}(u) \equiv E_u\left[e^{-\delta T - z(|U_T|+\eta V_T)} 1_{\{T<\infty\}};\, \mathbf{J}_T\right],$$

some modifications to (3.7) are required which we outline below. Modifications are required because V_t is no longer given by (3.5) for the network-based CIM proposed in this section.

Recall that V_t is the total (future) payment amount of all claims in the CIM at time t (if any). Let $\mathbf{g} = (g_1, \ldots, g_n)$ with $g_1 + \cdots + g_n = i$ and $\mathbf{g}' = (g_1', \ldots, g_n')$ with $g_1' + \cdots + g_n' = i'$. In what follows, for $(i, \mathbf{g}), (i', \mathbf{g}') \in E$, we write $(E_u\left[\,\cdot\,; J_T\right])_{(i,\mathbf{g}),(i',\mathbf{g}')}$ to denote

$$E\left[\,\cdot\, 1_{\{\mathbf{J}_T=(i',\mathbf{g}')\}} \middle| U_0 = u, \mathbf{J}_0 = (i, \mathbf{g})\right].$$

It follows that

$$\begin{aligned}
(\mathbf{m}_{\delta,z}(u))_{(i,\mathbf{g}),(i',\mathbf{g}')} &= E\left[\, e^{-\delta T - z(|U_T|+\eta V_T)} 1_{\{\mathbf{J}_T=(i',\mathbf{g}'),\, T<\infty\}}; |U_0 = u, \mathbf{J}_0 = (i, \mathbf{g})\right] \\
&= \left(\boldsymbol{\gamma}_{\delta,z}(u)\right)_{(i,\mathbf{g}),(i',\mathbf{g}')} E\left[e^{-z\eta V_T} | \mathbf{J}_T = (i', \mathbf{g}')\right], \qquad (3.18)
\end{aligned}$$

where $\boldsymbol{\gamma}_{\delta,z}(u) := E_u\left[e^{-\delta T - z|U_T|} 1_{\{T<\infty\}}; \mathbf{J}_T\right]$ is given by (3.8) using the matrix exponent $\mathbf{F}^\delta(z)$ identified in Proposition 14.

In the following, we consider V_T given $\mathbf{J}_T = (i', \mathbf{g}')$ where $(i', \mathbf{g}') \in E$. Note that if $\mathbf{J}_T = (0, 0, \ldots, 0)$, a zero vector of size $n+1$, then the CIM is empty at the time of ruin and $V_T = 0$ such that $E\left[e^{-z\eta V_T} | \mathbf{J}_T = (0, 0, \ldots, 0)\right] = 1$.

Recall that a claim departing the CIM from node j results in a payment amount having LT $\widetilde{k}_j(z)$ for $j \in \{1, \ldots, n\}$. Now consider a claim at node h (either being serviced or waiting in the node's queue) for $h \in \{1, \ldots, n\}$. The payment amount for this claim is determined based on the node from which it eventually departs the CIM. Suppose $\omega_{h,j}$ is the probability that a claim at node h departs the CIM from node j. Then the claim at node h results in a payment with LT $\sum_{j=1}^{n} \omega_{h,j} \widetilde{k}_j(z)$. Furthermore, if there are g'_h claims in node h, their total payments have LT $\left(\sum_{j=1}^{n} \omega_{h,j} \widetilde{k}_j(z)\right)^{g'_h}$ for $h \in \{1, \ldots, n\}$. Thus, it follows that

$$E\left[e^{-z\eta V_T} | J_T = (i', \mathbf{g}')\right] = \prod_{h=1}^{n} \left(\sum_{j=1}^{n} \omega_{h,j} \widetilde{k}_j(z\eta)\right)^{g'_h}. \tag{3.19}$$

Therefore, the joint LT of the time and deficit at ruin under the network-based CIM is given by substituting (3.19) into (3.18). Explicit expressions for $\{\omega_{h,j}\}_{h,j=1}^{n}$ are provided next.

Consider $\{N_t\}_{t\geq 0}$ where N_t represents the node at which the claim resides at time t. We have that $\{N_t\}_{t\geq 0}$ is a semi-Markov chain with transient states $\{1, \ldots, n\}$ representing the n service nodes and an absorbing state $\{0\}$ representing the claim's state after it departs the CIM. The process $\{N_t\}_{t\geq 0}$ is Markovian on the claim's service completions at each node it visits before departure. Recall that a claim at node h transfers to node j at rate $\mu_h p_{h,j}$ and departs the CIM at rate $\mu_h p_{h,0}$. It follows that $\omega_{h,j}$ is the probability that N_t is eventually absorbed into state 0 from state j given $N_0 = h$.

We remark that $\omega_{h,j}$ is independent of the time spent (waiting for or undergoing service) at each node. Thus, consider $\{N_t^*\}_{t\geq 0}$, a (fully Markovian) CTMC with the same underlying Markov chain as $\{N_t\}_{t\geq 0}$. We have that $\omega_{h,j}$ is also the probability that N_t^* is eventually absorbed into state 0 from state j given $N_0^* = h$. Note that N_t^* represents the node at which the claim resides at time t assuming there are an infinite number of servers at each node such that the claim spends no time waiting for service. The infinitesimal generator of $\{N_t^*\}_{t\geq 0}$ is given by

$$\mathbf{R} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{S}_0 & \mathbf{S} \end{bmatrix},$$

where $\mathbf{S} = [s_{h,j}]_{h,j=1}^{n}$ and $s_{h,j} = \mu_h p_{h,j}$ when $h \neq j$, and $s_{h,h} = -\mu_h(1 - p_{h,h})$ for $h, j \in \{1, \ldots, n\}$. Also,

$$\mathbf{S}_0 = -\mathbf{S}\mathbb{1}_n = (\mu_1 p_{1,0}, \ldots, \mu_n p_{n,0})^T.$$

From the general theory of CTMCs, we have that

$$P\left(N_t^* = j | N_0^* = h\right) = \left(e^{\mathbf{R}t}\right)_{h,j}.$$

Then $\left(e^{\mathbf{R}t}\right)_{h,j}(\mathbf{S}_0)_j$ is the probability that N_t^* equals j at time t and is then immediately absorbed, given $N_0^* = t$. Thus, by integrating over all possible time t, we obtain the following expression for $\omega_{h,j}$

$$\omega_{h,j} = \int_0^\infty \left(e^{\mathbf{R}t}\right)_{h,j}(\mathbf{S}_0)_j \, dt,$$

for $h, j \in \{1, \ldots, n\}$.

6. Numerical example

In this example, we compare a one-node CIM composed of one node and one waiting place (i.e., $n = 1$ and $m = 1$) with a two-node CIM composed of two nodes and no waiting places (i.e., $n = 2$ and $m = 0$). Note that both CIMs reach full capacity when two claims are in the system.

In the one-node CIM, all claims handling is assumed to occur in a single node. All other features of the process follows the baseline case outlined in Section 3.4.1. For example, service time at the single node is assumed to be exponentially distributed with mean $1/\alpha = 5$ and a claim that completes service is either determined fraudulent and denied with probability $\varsigma = 0.25$ or paid in its entirety with probability $1 - \varsigma = 0.75$. The matrix exponent for this process is given by either Proposition 14 or Proposition 10 with $n = 1$ and $m = 1$.

In the two-node CIM, claims handling is assumed to span two nodes. We let Node 1 be dedicated to claims investigation and Node 2 be dedicated to administrative payment processing. We assume that a claim entering the CIM arrives for service at Node 1 (i.e. $\gamma_1 = 1$). Claim investigation at Node 1 takes place for an exponential amount of time with mean $1/\mu_1 = 3a/5$. A claim that completes investigation at Node 1 is determined to be fraudulent and departs the CIM with no payment with probability $p_{1,0} = \varsigma = 0.25$. Otherwise, the claim is determined to be not fraudulent and will transfer to Node 2 with probability $p_{1,2} = 1 - \varsigma = 0.75$. We assume that no payment is triggered when the claim transfers to Node 2 (i.e. $\widetilde{k}_1(z) = 1$).

At Node 2, administrative payment processing takes place for an exponential amount of time with mean $1/\mu_2 = 2a/5$. For comparative purposes with the one-node system, a claim that completes processing at Node 2 will exit the CIM (i.e. $p_{2,0} = 1$) and a payment is triggered equal to the full amount claimed (i.e. $\widetilde{k}_2(z) = \widetilde{p}_1(z)$). All other features of the process follows the baseline case outlined in Section 3.4.1. The matrix exponent for the resulting surplus process is given by Proposition 14.

Let ξ_1 and ξ_2 be, respectively, the long-run mean time an arriving claim spends (either waiting for or undergoing service) in the one-node and the two-node CIM, given the claim enters the CIM. Recall that an arriving claim enters either the one-node or two-node CIM only if there are less than 2 claims in the CIM. We set parameter a such that $\xi_1 = \xi_2$.

In the following, we find ξ_1 and ξ_2. For $j = 1, 2$, let $\mathbf{J}_j$ be the background process for the j-node CIM having state space E_j and infinitesimal generator $\mathbf{Q}_j$. Note that we have

$$E_1 = \{(0,0), (1,1), (2,2)\},$$

and

$$E_2 = \{(0,0,0), (1,1,0), (1,0,1), (2,2,0), (2,0,2), (2,1,1)\}.$$

Define

$$\boldsymbol{\pi}_1 = \left(\pi_1(i,g) \right)_{(i,g)\in E_1},$$

and

$$\boldsymbol{\pi}_2 = \left(\pi_2(i,\mathbf{g}) \right)_{(i,\mathbf{g})\in E_2},$$

to be, respectively, the stationary distribution of the one-node CIM and the two-node CIM (i.e., $\boldsymbol{\pi}_j \mathbf{Q}_j = \mathbf{0}$ and $\boldsymbol{\pi}_j \mathbb{1}_{|E_j|} = 1$ for $j = 1, 2$). In what follows, for simplicity, we use $Exp(\mu)$ to refer to an exponential rv with mean $1/\mu$.

In the one-node CIM, given a claim enters the CIM, if it finds:

1. No claims in the CIM, with long-run probability $\pi_1(0,0)/(\pi_1(0,0) + \pi_1(1,1))$, it will immediately enter service for an $Exp(0.2)$ amount of time, and then exit the CIM.
2. One claim in the CIM, with long-run probability $\pi_1(1,1)/(\pi_1(0,0) + \pi_1(1,1))$, it will wait for service for an $Exp(0.2)$ amount of time, then enter service for an $Exp(0.2)$ amount of time, and then exit the CIM.

Thus, we have

$$\xi_1 = \frac{\pi_1(0,0)}{\pi_1(0,0) + \pi_1(1,1)} \frac{1}{0.2} + \frac{\pi_1(1,1)}{\pi_1(0,0) + \pi_1(1,1)} \left(\frac{1}{0.2} + \frac{1}{0.2} \right) = 9.6296. \tag{3.20}$$

To find ξ_2, we will separately consider the cases where the claim entering the CIM is fraudulent versus not fraudulent. Let $\xi_{2,fraud}$ ($\xi_{2,notfraud}$) be the long-run mean time that a fraudulent (not fraudulent) claim spends in the two-node CIM, given that it enters the CIM, such that

$$\xi_2 = \varsigma \xi_{2,fraud} + (1 - \varsigma) \xi_{2,notfraud}. \tag{3.21}$$

For simplicity, define

$$\pi_2^*(i, g_1, g_2) \equiv \frac{\pi_2(i, g_1, g_2)}{\pi_2(0,0,0) + \pi_2(1,1,0) + \pi_2(1,0,1)} \tag{3.22}$$

to be the long-run probability that an arriving claim finds the CIM in state (i, g_1, g_2), given that the claim enters the CIM.

Given a claim is fraudulent and it enters the CIM, if it finds:

1. No claims in the CIM, with long-run probability $\pi_2^*(0,0,0)$, it will immediately enter service at Node 1 for an $Exp(\mu_1)$ amount of time, and then exit the CIM.

2. One claim in the CIM undergoing service at Node 2, with long-run probability $\pi_2^*(1,0,1)$, it will immediately enter service at Node 1 for an $Exp(\mu_1)$ amount of time, and then exit the CIM.

3. One claim in the CIM undergoing service at Node 1, with long-run probability $\pi_2^*(1,1,0)$, it will wait for service at Node 1 for an $Exp(\mu_1)$ amount of time, then enter service at Node 1 for an $Exp(\mu_1)$ amount of time, and then exit the CIM.

Thus, we have

$$\xi_{2,fraud} \equiv \pi_2^*(0,0,0)\frac{1}{\mu_1} + \pi_2^*(1,0,1)\frac{1}{\mu_1} + \pi_2^*(1,1,0)\frac{2}{\mu_1}. \tag{3.23}$$

Given a claim is non-fraudulent and it enters the CIM, if it finds:

1. No claims in the CIM, with long-run probability $\pi_2^*(0,0,0)$, it will immediately enter service at Node 1 for an $Exp(\mu_1)$ amount of time, then immediately enter service at Node 2 for an $Exp(\mu_2)$ amount of time, and then exit the CIM.

2. One claim in the CIM undergoing service at Node 2, with long-run probability $\pi_2^*(1,0,1)$, it will immediately enter service at Node 1 and:

   (a) If service at Node 2 completes before service at Node 1, the claim will spend an $Exp(\mu_1)$ amount of time in Node 1, then immediately enter service at Node 2 for an $Exp(\mu_2)$ amount of time, and then exit the CIM.

   (b) If service at Node 1 completes before service at Node 2, the claim will be in service at Node 1 and wait for service at Node 2 for an $Exp(\mu_2)$ amount of time, then enter service at Node 2 for an $Exp(\mu_2)$ amount of time, and then exit the CIM.

3. One claim in the CIM undergoing service at Node 1, with long-run probability $\pi_2^*(1,1,0)$, it will wait for the claim ahead to complete service for an $Exp(\mu_1)$ amount of time, then enter Node 1 where:

   (a) If the claim ahead was fraudulent (and has exited the CIM), it will spend an $Exp(\mu_1)$ amount of time in Node 1, then immediately enter service at Node 2 for an $Exp(\mu_2)$ amount of time, and then exit the CIM.

   (b) If the claim ahead was not fraudulent (and has moved to Node 2), then we have the second scenario described above.

Thus, using (3.22), we have

$$\begin{aligned}\xi_{2,notfraud} &= \pi_2^*(0,0,0)\left(\frac{1}{\mu_1}+\frac{1}{\mu_2}\right)+\pi_2^*(1,0,1)\left(E\left[\max(Exp(\mu_1),Exp(\mu_2))\right]+\frac{1}{\mu_2}\right)\\ &+ \pi_2^*(1,1,0)\left(\frac{1}{\mu_1}+\varsigma\frac{1}{\mu_1}+(1-\varsigma)E\left[\max(Exp(\mu_1),Exp(\mu_2))\right]+\frac{1}{\mu_2}\right).\end{aligned} \tag{3.24}$$

Note that $E\left[\max(Exp(\mu_1),Exp(\mu_2))\right]=1/\mu_1+1/\mu_2-1/\left(\mu_1+\mu_2\right)$.

Setting (3.20) equal to (3.21) and using (3.23) and (3.24), we solve for $a\approx 7.0276$. The ruin probability for both the one-node and two-node CIMs are given in Table 3.10.

Table 3.10: Ruin probabilities for a network-based CIM

$(n,m)\setminus u$	0	15	30	45	60	75
(1,1)	0.7214	0.2732	0.1188	0.0519	0.0227	0.0099
(2,0)	0.7170	0.2710	0.1179	0.0515	0.0225	0.0099

Suppose the claims handling process is indeed composed of two stages (i.e., a claim is first investigated for fraudulence, then it moves onto processing if it is not fraudulent). Modelling this process using a one-node CIM would overestimate the ruin probability since it does not reflect the speedier exit of fraudulent claims from the CIM. The speedier exit of fraudulent claims means more claims can enter the CIM (less balking) and greater opportunity to catch more fraudulent claims which should lower the ruin probability. This is confirmed by the Table 3.10 where we observe the ruin probability for the two-node CIM is lower than for the one-node CIM.

3.5.4 Markov-modulated Poisson claim arrivals

The claim arrival rates that insurers experience can often depend on the state of the economy (see, e.g., Wells et al. [2009]). For example, group disability insurers often experience higher claim arrival rates in weak economic times. One explanation for this effect is that, during a recession, employees who had health impairments face a lower opportunity cost when applying for benefits when they become unemployed (Maestas et al. [2015]).

Recall that in the proposed CIM, claims arrive according to a Poisson process with rate λ. In this section, we let the claim arrival process be governed by the more general Markov modulated Poisson process (MMPP). That is, claim arrivals are assumed to be governed by a Poisson process whose rate depends on the state of $R = \{R_t\}_{t\geq 0}$ which is defined to be a continuous-time Markov process with finite state space $\{1, \ldots, v\}$ and infinitesimal generator $\mathbf{D} = (d_{l,l'})$ where $l, l' \in \{1, \ldots, v\}$. When $\{R_t = l\}$, we say the 'environment' is in state l at time t and claims are assumed to arrive according to a Poisson process with rate $\lambda^{(l)}$ ($\lambda^{(l)} > 0$). All other features of the CIM are assumed to satisfy the proposed CIM given in Section 3.2.1. Thus, if there is only one environmental state (i.e., $v = 1$), we recover the proposed CIM with λ replaced by $\lambda^{(1)}$.

Superimposing this generalized CIM onto the surplus process preserves the MArP formulation and in this section, we present the MArP formulation for the resulting surplus process and determine its matrix exponent.

1. Defining background process J

We begin by defining the generalized background process and its state space. The generalized background process is defined by the bivariate Markov process $\mathbf{J} = \{(J_t, R_t)\}_{t \geq 0}$ where $\mathbf{J}_t = (J_t, R_t)$ corresponds to not only the number of claims in the CIM but also, the state of the environment at time t. Thus, the generalized state space is given by

$$E = \bigcup_{i=0}^{n+m} C_i,$$

where $C_i = \{\,(i,1)\,,\,\ldots\,,\,(i,v)\}$ for $i \in \{0,\ldots,n+m\}$.

2. Identifying infinitesimal generator Q

We now determine the transition rates of $\mathbf{J}$. For $j = 0, 1$ and $l \in \{1, \ldots, v\}$, define

$$\lambda_j^{(l)} = \lambda^{(l)} \int_0^\infty p(y) q_j(y) dy.$$

When the CIM is not at capacity and the environment is in state l, $\lambda_1^{(l)}$ ($\lambda_0^{(l)}$) corresponds to the arrival rate of claims that enter (avoid) the CIM. When the CIM is at capacity and the environment is in state l, recall that all arriving claims (arriving at rate $\lambda^{(l)}$) avoid the CIM. Denote the transition rate of $\mathbf{J}$ from state (i,l) to (i',l') by $q_{(i,l),(i',l')}$ for (i,l), $(i',l') \in E$. All transition rates starting from state $(i,l) \in C_i$ are given below:

1. **Claims entering the CIM:** Given the CIM is not at capacity and the environment is in state l, claims enter the CIM at rate $\lambda_1^{(l)}$. Thus, if $0 \leq i < n+m$,

$$q_{(i,l),(i+1,l)} = \lambda_1^{(l)}, \qquad \text{for all } 1 \leq l \leq v.$$

2. **Claims departing the CIM:** Given the CIM is not empty and the environment is in state l, claims complete investigation and depart the CIM at rate $i\alpha$ if i $(i < n)$ investigation units are busy and depart from the CIM at rate $n\alpha$ if all investigation units are busy. Thus, if $0 < i \leq n + m$,

$$q_{(i,l),(i-1,l)} = min(n, i)\alpha, \qquad \text{for all } 1 \leq l \leq v.$$

3. **Transitions with a fixed number of claims in the CIM:** A transition with a fixed number of claims in the CIM occurs when there is a transition in the environment process R. We have that a transition in R from state l to l' occurs at rate $d_{l,l'}$. Thus, for all $0 \leq i \leq n + m$,

$$q_{(i,l),(i,l')} = d_{l,l'}, \qquad \text{for all } 1 \leq l, l' \leq v.$$

No other transitions in $\mathbf{J}$ are possible.

Using the transition rates for $\mathbf{J}$ described above, we are able to give the infinitesimal generator $\mathbf{Q}$. The diagonal elements of $\mathbf{Q}$ are found such that its rows sum to zero. According to the transition rates given above, it follows that

$$q_{(i,l),(i,l)} = \begin{cases} d_{l,l} - \lambda_1^{(l)}, & i = 0, \\ d_{l,l} - \lambda_1^{(l)} - min(n, i)\alpha, & 0 < i < n + m, \\ d_{l,l} - min(n, i)\alpha, & i = n + m, \end{cases}$$

for all $(i, l) \in E$. Let us define

$$\mathbf{\Lambda}_j = diag \left\{ \lambda_j^{(l)} \right\}_{l=1}^{v},$$

for $j = 0, 1$, and

$$\mathbf{\Lambda} = diag \left\{ \lambda^{(l)} \right\}_{l=1}^{v}.$$

The infinitesimal generator $\mathbf{Q}$ can be written and partitioned as follows

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{0,0} & \mathbf{Q}_{0,1} & 0 & 0 & 0 & \cdots \\ \mathbf{Q}_{1,0} & \mathbf{Q}_{1,1} & \mathbf{Q}_{1,2} & 0 & 0 & \cdots \\ 0 & \mathbf{Q}_{2,1} & \mathbf{Q}_{2,2} & \mathbf{Q}_{2,3} & 0 & \cdots \\ 0 & 0 & \mathbf{Q}_{3,2} & \mathbf{Q}_{3,3} & \mathbf{Q}_{3,4} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\ & & & & \mathbf{Q}_{n+m,n+m-1} & \mathbf{Q}_{n+m,n+m} \end{bmatrix}. \tag{3.25}$$

where $\mathbf{Q}_{i,i'}$ are submatrices of transition rates from states in C_i to states in $C_{i'}$. According to the transition rates described above, we have simple expressions for these submatrices. For $i = 0$, we have

$$\mathbf{Q}_{0,0} = \mathbf{D} - \mathbf{\Lambda}_1, \qquad \text{and} \qquad \mathbf{Q}_{0,1} = \mathbf{\Lambda}_1.$$

When $0 < i < n + m$, we have

$$\mathbf{Q}_{i,i-1} = min(n,i)\alpha\mathbf{I}, \qquad \mathbf{Q}_{i,i} = \mathbf{D} - min(n,i)\alpha\mathbf{I} - \mathbf{\Lambda}_1, \qquad \text{and} \qquad \mathbf{Q}_{i,i+1} = \mathbf{\Lambda}_1.$$

Finally, when $i = n + m$, we have

$$\mathbf{Q}_{n+m,n+m-1} = n\alpha\mathbf{I}, \qquad \mathbf{Q}_{n+m,n+m} = \mathbf{D} - n\alpha\mathbf{I}.$$

3. Identifying matrix $\mathbf{G}(z)$

Next, we consider the LT of payments triggered by transitions in $\mathbf{J}$ to determine the matrix $\mathbf{G}(z)$. Let $g(z)_{(i,l),(i',l')}$ denote the LT of the payment triggered when $\mathbf{J}$ transitions from state (i,l) to (i',l') where (i,l), $(i',l') \in E$. Recall that a payment is triggered only at transitions in $\mathbf{J}$ corresponding to a claim completing investigation and departing the CIM and that the payment sizes have LT $\widetilde{k}(z)$. Thus, if $0 < i \leq n + m$,

$$g(z)_{(i,l),(i-1,l)} = \widetilde{k}(z), \qquad \text{for all } 1 \leq l \leq v. \tag{3.26}$$

Since all other transitions in $\mathbf{J}$ do not trigger payments (which is mathematically equivalent to assuming such transitions trigger a payment of size 0), all other entries in $\mathbf{G}(z)$ are equal to 1. Thus, we can write and partition the matrix $\mathbf{G}(z)$ as follows

$$\mathbf{G}(z) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & \cdots \\ \mathbf{G}_{1,0}(z) & 1 & 1 & 1 & 1 & \cdots \\ 0 & \mathbf{G}_{2,1}(z) & 1 & 1 & 1 & \cdots \\ 1 & 1 & \mathbf{G}_{3,2}(z) & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\ & & & & \mathbf{G}_{n+m,n+m-1}(z) & 1 \end{bmatrix} \tag{3.27}$$

where $\mathbf{G}_{i,i-1}(z)$ are submatrices of the LT of payments triggered by transitions in $\mathbf{J}$ from states in C_i to states in C_{i-1}. According to (3.26), for all $0 < i \leq n+m$, $\mathbf{G}_{i,i-1}(z)$ is given by a $v \times v$ matrix where all diagonal elements are equal to $\widetilde{k}(z)$ and all off-diagonal elements are equal to 1.

4. Matrix exponent $\mathbf{F}^{\delta}(z)$

Next, we consider the surplus process when $\mathbf{J}$ remains in a particular state. Let the surplus process U evolve as $U^{(i,l)} = \{U_t^{(i,l)}\}_{t\geq 0}$ when $\{\mathbf{J}_t = (i,l)\}$ and denote the Laplace exponent of $U^{(i,l)}$ by $\psi_i^{(l)}(z)$ for $(i,l) \in E$. When $0 \leq i < n+m$, we have

$$\psi_i^{(l)}(z) = cz + \frac{\sigma^2}{2}z^2 + \lambda_0^{(l)}\left(\widetilde{p}_0(z) - 1\right), \qquad \text{for all } 1 \leq l \leq v,$$

and when $i = n+m$, we have

$$\psi_{n+m}^{(l)}(z) = cz + \frac{\sigma^2}{2}z^2 + \lambda^{(l)}\left(\widetilde{p}(z) - 1\right), \qquad \text{for all } 1 \leq l \leq v.$$

Given the MArP formulation provided thus far, we are ready to give the matrix exponent of the surplus process.

Proposition 15. For claim arrivals governed by a MMPP, the resulting surplus process U is characterized by the matrix exponent $\mathbf{F}^{\delta}(z)$ given by (3.4) with $\mathbf{Q}$ and $\mathbf{G}(z)$ replaced by (3.25) and (3.27), respectively. Also, $\mathbf{\Lambda}$ and $\mathbf{P}(z)$ are respectively replaced by

$$\mathbf{\Lambda} = diag\{\mathbf{\Lambda}_0, \ldots, \mathbf{\Lambda}_0, \, \mathbf{\Lambda}\},$$

a block diagonal matrix where all diagonal blocks are equal to $\mathbf{\Lambda}_0$ except for the last which is equal to $\mathbf{\Lambda}$, and

$$\mathbf{P}(z) = diag\{\widetilde{p}_0(z), \ldots, \widetilde{p}_0(z), \, \widetilde{p}(z)\} \otimes \mathbf{I},$$

a diagonal matrix where all diagonal entries are equal to $\widetilde{p}_0(z)$ except for the last v diagonal entries which are all equal to $\widetilde{p}(z)$.

5. Joint LT of the time and deficit at ruin

We note that the joint LT of the time and deficit at ruin is given by (3.7) and (3.8) using the matrix exponent $\mathbf{F}^{\delta}(z)$ identified in Proposition 15. No further modifications are necessary because V_t, which we recall is defined as the total (future) payment amount of all claims in the CIM at time t (if any), is still given by (3.5).

Chapter 4

An aggregate claims model with claims investigation

4.1 Introduction

Chapters 2 and 3 considered a queue-based claim investigation mechanism (CIM) superimposed on the insurer's surplus process. This chapter extends this inquiry by superimposing the CIM onto the insurer's aggregate claims process. An insurer's aggregate claim is a mathematical representation of the total claims received by the insurer over some time period.

There is a time lag between the time a claim is reported to an insurer and the time it is settled. An insurer experiences congestion whenever claims are reported but not immediately paid as a result of investigation/processing times or time spent waiting in queue for investigation/processing. As a result of payments not being made as soon as they are reported, insurers must estimate outstanding claim liabilities. Developing accurate estimates of claim liabilities associated with congestion is important because these estimates can affect the amount of reserves insurers are required by regulation to set aside to meet future obligations.

There are two major categories of the claim liability associated with time lags:

- Incurred But Not Reported (IBNR) which refer to incurred losses for which the insurer has not (yet) received notification of.
- Reported But Not Settled (RBNS), which refer to claims for which notification has reached the insurer but for which no payment has been made.

For IBNR claims, the time taken to report a claim corresponding to an incurred loss can be assumed to be independent of other claims in a similar IBNR state. By contrast, for RBNS claims, the time it takes to settle a reported claim does depend on the presence of other RBNS claims, i.e., claims that are currently waiting for or are under investigation.

Mathematical models of aggregate claims that incorporate time lags can be broadly categorized into deterministic methods such as the well-known chain-ladder method (see, e.g., Brown and Gottlieb [2007]), macro-level (stochastic) models, and micro-level (stochastic) models (see, e.g., Badescu et al. [2016] and references therein). The model developed in this chapter would be described as a micro-level model since it characterizes the behaviour of individual claims. Readers are referred to, e.g., Norberg [1993] and its follow-up Norberg [1999] for micro-level models of IBNR and RBNS claim liabilities where claims follow a non-homogeneous marked Poisson process.

The present work focuses on formulating a model for the RBNS claim liability. By altering the design of the claim investigation and settlement process, insurers can control the RBNS claim liability. A model for the RBNS claim liability can be of value to insurers because they could be used to measure the impact of alternative process designs. More specifically, such a model could help insurers quantify the costs/benefits of particular claims handling operations and identify alternative practices, e.g., determining a sufficient number of investigators, choosing target service rates, or changing the set-up of the claim investigation mechanism. Furthermore, insurers could use such a model to predict the future RBNS claim liability.

A RBNS liability model may be especially helpful when data relating to RBNS claim size, count, and frequency is either not available or costly to obtain (see, e.g., Willmot [1990]).

In general, a claims handling system benefits insurers when it denies ineligible claims or corrects inflated claims. It creates costs for insurers due to the need for personnel, administration and, when executed poorly, the customer dissatisfaction and risk to reputation resulting from excessive wait times or inaccurate assessments. The time delay between receiving and paying a claim reflects the efficiency of a claims processing system. Being able to process claims in shorter periods of time (on average) may also be a particularly valuable competitive advantage. More granular models of the claims handling process could help insurers target the source of inefficiencies in their processing systems.

Finally, queue-based claim investigation processes raise many practical questions for which queueing-theoretic models have the potential to provide insights. For example:

- What is the effect of changes to an insurer's investigation strategies on their risk profile?
- How does the RBNS liability evolve over time?
- What is the long-run behaviour of the RBNS claims in the CIM?
- For two different CIMs, which generates greater RBNS liability risk?

These kinds of questions motivate the development of a queueing-theoretic mathematical model of claims investigations.

Some studies have been done on claim liabilities using a queue-based model for the settlement process. Hachemeister [1980] proposed a claim reporting/handling process represented by a Markov chain. It was assumed that an incurred claim went through some staged reporting and handling process such as “IBNR”, “RBNS”, and “settled”.

Willmot [1990] studied the liability of reported and unreported claims when modelling the claim settlement process using an infinite capacity queue with Poisson claim arrivals. Queues of varying complexity were considered and the claim settlement process was assumed to have reached equilibrium. Connecting with results from Queueing Theory (see, e.g., Kleinrock [1975]), the approximate right-tail behaviour of the liability from reported claims was examined which allowed for an estimate of the amount needed to cover such liabilities with a specified probability.

Kim and Kim [2007] and Ren [2008] studied the aggregate discounted claims with a Markovian claim arrival process. The present work introduces a particular claim settlement application to this work and considers the special case where there is no discounting. Aggregate claims models that do not incorporate discounting are more tractable due to their regenerative features and the present work capitalizes on the explicit results that are available for such models. The present model is intended for claims with short time lags or when it is believed that inflation on claim severities offsets interest earned on assets backing the reserve (see, e.g., Léveillé and Garrido [2001]). If the claims being considered have longer delays or inflation is not assumed to offset interest, then the present model could be viewed as providing a conservative estimate of the claim liability.

More recently, Ren [2016] found the joint distribution and moments of the IBNR, RBNS, and settled claims in a model for the claim occurrence, reporting, and handling process assuming claims occur according to a Markovian arrival process. The model in this chapter is a special case of that found in Ren [2016] since it examines only claim liabilities arising from RBNS and settled claims, where claims arrive according to a Poisson process. However, this work conducts a more in-depth analysis of future payments due to RBNS claims.

As done in Ren [2016], this chapter superimposes a claim investigation mechanism (CIM) onto the aggregate claims process. The resulting process models the total payments actually paid to the policyholder by the insurer and is referred to as the "aggregate payment process". The CIM that is superimposed is assumed to be the queue-based CIM described in Section 3.2.1.

Under the model assumptions, the aggregate payment process is found to have a Markovian Arrival Process (MAP) formulation. Furthermore, from the discussions on CIM generalizations in Chapter 3, it follows that considerable generalizations to the CIM are possible while preserving the MAP formulation.

We define the "RBNS payments" to be the future (eventual) payments due to reported but not settled claims. By making connections with existing MAP results, the joint LT and moments of the settled and RBNS payments are obtained.

The RBNS payments is a quantity of particular interest in this chapter. Its distribution is found under some assumptions. In addition, its distribution is obtained when it is further assumed that the claim size density follows a mixed Erlang, gamma, or inverse Gaussian distribution.

In numerical examples, the Value-at-Risk (VaR) and Conditional Tail Expectation ($TVaR$) of the RBNS payments are computed for varying claim investigation practices. The VaR of a loss denoted by rv X is interpreted as a threshold value for which the probability that the loss exceeds this threshold is less than some specified probability. More precisely, at confidence level p $(0 < p < 1)$, a loss X has its VaR defined as

$$VaR_p = inf\{x \geq 0 : P\left(X > x\right) \leq 1 - p\}.$$

The $TVaR$ is a more conservative risk measure associated with the VaR. At confidence level

p $(0 < p < 1)$, a loss has its $TVaR$ defined by (see, e.g., [Klugman et al., 2013, Chapter 5])

$$\begin{aligned} TVaR_p &= \frac{\int_p^1 VaR_\phi d\phi}{1-p} \\ &= VaR_p + \frac{\overline{F}_X(VaR_p)}{1-p}\Big(E\left[X|X > VaR_p\right] - VaR_p\Big), \end{aligned}$$

where $\overline{F}_X$ is the survival function of X. Note that when the distribution of X is continuous at VaR_p, then $TVaR_p$ is given by the conditional expectation of X given that it exceeds the VaR_p, i.e.,

$$TVaR_p = E\left[X|X > VaR_p\right].$$

The VaR and $TVaR$ of RBNS payments in the long-run is also considered as these quantities can provide a measure of the overall risk due to RBNS claims. That is, we consider the risk due to RBNS claims when the CIM "settles down" to some equilibrium state at some arbitrary time in the distant future (see, e.g., [Kleinrock, 1975, Chapter 3]).

Overall, we will observe that our findings from the numerical examples in this chapter are consistent with those from Chapters 2 and 3.

The remainder of this chapter is organized as follows: Section 4.2 presents the model, the joint LT and moments of settled and RBNS payments are found in Section 4.3, Section 4.4 contains a distributional analysis of the RBNS payments, remarks on possible generalizations to the CIM are included in Section 4.5, and numerical studies are included in Section 4.6 to conclude the chapter.

4.2 Model presentation

In what follows, we propose a queue-based claim investigation mechanism (CIM) that is identical to that proposed in Chapter 3 and we refer readers to Section 3.2.1 for the CIM assumptions.

Recall the definition of the aggregate claims process $S^* = \{S_t^*\}_{t\geq 0}$ under the collective risk model (see, e.g., [Klugman et al., 2012, Chapter 9] or Buhlmann [1970]) where

$$S_t^* = \sum_{i=1}^{N_t^*} X_i, \qquad t \geq 0, \tag{4.1}$$

with the convention that $S_t^* = 0$ if $N_t^* = 0$. For the compound sum (4.1), the claim sizes $\{X_i\}_{i=1}^{\infty}$ are assumed to form a sequence of i.i.d. positive rv's and the claim number process $\{N_t^*\}_{t\geq 0}$ is a counting process with N_t^* representing the number of claims that have arrived by time t. It is generally assumed that the counting process $\{N_t^*\}_{t\geq 0}$ is independent of $\{X_i\}_{i=1}^{\infty}$.

In this section, we superimpose the proposed CIM described in Section 3.2.1 onto the aggregate claims process (4.1) which in essence consists in incorporating a settlement delay when claims are reported to the insurer. We assume the counting process $\{N_t^*\}_{t\geq 0}$ is a Poisson process and we find that the resulting payment process has payment arrivals that follow a Markovian arrival process (MAP).

Remark 1. In this chapter, "aggregate claims" refers to the total amount declared by policyholders (after appropriately applied coverage modifications, e.g. deductible, policy limit, etc.) and "aggregate payment" refers to the total payments actually paid to the policyholders by the insurer.

Remark 2. The Markovian arrival process (MAP) discussed in this chapter is not to be confused with either the Markovian additive process or the Markovian arrival risk process (both discussed in Chapter 3), despite the fact that these three processes typically share the "MAP" abbreviation.

The following notation will be used in the remainder of this chapter. We use $\mathbf{I}_n$ to denote an $n \times n$ identity matrix and $\mathbb{1}_n$ to denote a column vector of size n with all entries equal to

1. We denote by $\mathbf{e}_{n,h}$ the row vector of size n with hth entry equal to 1 and all other entries equal to 0. Suppose we have two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimension. The Hadamard product of $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \circ \mathbf{B}$ and has (i,j)th element given by

$$(A \circ B)_{i,j} = (A)_{i,j} (B)_{i,j} .$$

For completeness, we recall the definition of an aggregate loss process with MAP arrivals (see, e.g., [Latouche and Ramaswami, 1999, pg. 78]). Define an aggregate loss process $S = \{S_t\}_{t \geq 0}$ given by

$$S_t = \sum_{i=1}^{N_t} Y_i, \qquad t \geq 0, \tag{4.2}$$

where payments of size $\{Y_i\}_{i=1}^{\infty}$ occur according to a MAP. That is, payments occur according to a MAP with background process $J = \{J_t\}_{t \geq 0}$ which is assumed to be an irreducible continuous-time Markov process with finite state space E and infinitesimal generator $\mathbf{Q}$. There are two types of transitions in J:

- *Type-1*: Transitions in J from state i to state j ($j \neq i$) without an accompanying payment. Such transitions are assumed to occur at rate $d_{0,i,j}$.
- *Type-2* : Transitions in J from state i to state j (with possibly $j = i$) with an accompanying claim denoted by rv $P_{i,j}$ having LT $\widetilde{h}_{i,j}$. Such transitions are assumed to occur at rate $d_{1,i,j}$.

We let Type-1 transitions be governed by the matrix $\mathbf{D}_0 = (d_{0,i,j})_{i,j \in E}$ and Type-2 transitions be governed by the matrix $\mathbf{D}_1 = (d_{1,i,j})_{i,j \in E}$. The diagonal elements of $\mathbf{D}_0$ are assumed to be negative such that the elements of each row of $\mathbf{Q} = \mathbf{D}_0 + \mathbf{D}_1$ sum to zero. We also define the matrix $\mathbf{H}(s) = \left(\widetilde{h}_{i,j}(s)\right)_{i,j \in E}$. It follows that the LT of the aggregate loss process S_t is given by (see, e.g., [Latouche and Ramaswami, 1999, pg. 78])

$$E\left[e^{-sS_t}; J_t\right] = e^{(\mathbf{D}_0 + \mathbf{D}_1 \circ H(s))t}, \qquad t \geq 0. \tag{4.3}$$

Note that we write $E\left[\,\cdot\,; J_t\right]$ to indicate a matrix with (i,j)th element corresponding to $E\left[\,\cdot\, 1_{\{J_t=j\}} \middle| J_0 = i\right]$.

By the definition of the aggregate loss process with MAP arrivals given by (4.2), we can have:

1. a payment when J makes a state change,
2. no payment when J makes a state change, and
3. a payment when J makes a transition but does not change state (i.e., J revisits the state it is in).

Correspondingly, by superimposing the proposed CIM onto the aggregate claims process given by (4.1), we can have:

1. a payment when the number of claims in the CIM changes,

   i.e., a payment made following a claim investigation where the number of claims in the CIM decreases by one,

2. no payment when the number of claims in the CIM changes,

   i.e., no payment made when a claim is selected for investigation and enters the CIM where the number of claims in the CIM increases by one,

3. a payment without a change in the number of claims in the CIM,

   i.e., a payment made following a claim that is not selected for investigation or arrives when the CIM is full where the number of claims in the CIM does not change.

Thus, letting J be the Markov process corresponding to the number of claims in the CIM (where $E = \{0, 1, \ldots, n+m\}$) results in an aggregate payment process for settled claims that can be modelled by S_t with LT given by (4.3).

Since the CIM assumed here follows that proposed in Chapter 3, the infinitesimal generator $\mathbf{Q}$ is given by (3.2). As was done in Chapters 2 and 3, we define

$$\lambda_l = \lambda \int_0^\infty p(y)q_l(y)dy, \tag{4.4}$$

and

$$p_l(x) = \frac{p(x)q_l(x)}{\int_0^\infty p(y)q_l(y)dy}, \tag{4.5}$$

for $l = 0, 1$, and $x > 0$. We have that λ_1 (λ_0) corresponds to the arrival rate of claims that are selected to enter the CIM (avoid the CIM) and $p_1(x)$ ($p_0(x)$) is the claim size density of such claims. We also define

$$K(w) = \int_0^\infty L_x(w)p_1(x)dx, \tag{4.6}$$

corresponding to the size of a payment for a claim undergoing the investigation process. To determine $\mathbf{D}_0$ and $\mathbf{D}_1$, we consider the transitions in J with and without an accompanying claim. Under the proposed CIM, we obtain

$$d_{1,i,j} = \begin{cases} min(i,n)\alpha, & 0 < i \leq n+m \text{ and } j = i-1, \\ \lambda_0, & 0 \leq i < n+m \text{ and } j = i, \\ \lambda, & i = j = n+m, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$\widetilde{h}_{i,j}(s) = \begin{cases} \widetilde{k}(s), & 0 < i \leq n+m \text{ and } j = i-1, \\ \widetilde{p}_0(s), & 0 \leq i < n+m \text{ and } j = i, \\ \widetilde{p}(s), & i = j = n+m, \\ 1, & \text{otherwise}, \end{cases}$$

by considering all possible transitions in J with an accompanying claim, namely, transitions corresponding to payment of claims that: have completed investigation, are not selected for investigation, or arrive when the CIM is at full capacity. Thus, $\mathbf{D}_1$ and $\mathbf{H}(s)$ are respectively given by

$$\mathbf{D}_1 = \begin{bmatrix} \lambda_0 & 0 & \cdots & & & & \\ \alpha & \lambda_0 & \ddots & \ddots & & & \\ 0 & 2\alpha & \lambda_0 & & & & \\ \vdots & \ddots & \ddots & \ddots & & & \\ & & & n\alpha & \lambda_0 & & \\ & & & & \ddots & \ddots & \\ & & & & & n\alpha & \lambda \end{bmatrix}, \tag{4.7}$$

and

$$\mathbf{H}(s) = \begin{bmatrix} \widetilde{p}_0(s) & 1 & \cdots & & & \\ \widetilde{k}(s) & \widetilde{p}_0(s) & \ddots & \ddots & & \\ 1 & \widetilde{k}(s) & \widetilde{p}_0(s) & & & \\ \vdots & \ddots & \ddots & \ddots & & \\ & & & & \widetilde{k}(s) & \widetilde{p}(s) \end{bmatrix}. \tag{4.8}$$

We also obtain

$$d_{0,i,j} = \begin{cases} \lambda_1, & 0 \le i < n+m \text{ and } j = i+1, \\ -\left(\lambda + min(i,n)\alpha\right), & 0 \le i \le n+m \text{ and } j = i, \\ 0, & \text{otherwise,} \end{cases}$$

by considering all possible transitions in J without an accompanying claim, namely, transitions corresponding to claims that enter the CIM. Also, recall that the diagonal elements of $\mathbf{D}_0$ are set such that the elements on each row of the matrix $\mathbf{Q} = \mathbf{D}_0 + \mathbf{D}_1$ sum to zero.

Thus, $\mathbf{D}_0$ is given by

$$\mathbf{D}_0 = \begin{bmatrix} -\lambda & \lambda_1 & 0 & \cdots & & & \\ 0 & -(\lambda+\alpha) & \lambda_1 & \ddots & & & \\ \vdots & \ddots & -(\lambda+2\alpha) & \ddots & & & \\ & & & \ddots & & & \\ & & & & -(\lambda+n\alpha) & \ddots & \\ & & & & & \ddots & \lambda_1 \\ & & & & & & -(\lambda+n\alpha) \end{bmatrix}. \tag{4.9}$$

For $t \geq 0$, define

$$R_t = \sum_{i=1}^{J_t} Y_i, \tag{4.10}$$

where $\{Y_i\}_{i=1}^{\infty}$ are i.i.d. rv's with LT $\widetilde{k}$. For a fixed time t, we can interpret R_t as the total future payments due to Reported But Not Settled (RBNS) claims at time t and we henceforth refer to R_t as the "RBNS payments" at time t. In the following section, for a fixed time t, we consider the joint LT of the aggregate payments for settled claims S_t and the RBNS payments R_t. Note that if a study of the total future payments due to RBNS claims involves varying time t (such as a study of joint distributions), then a more rigorous definition than that given by (4.10) is required.

4.3 Joint Laplace transform and moments of settled and RBNS payments

In this section, we present the joint LT and moments of the aggregate payments for settled claims S_t and RBNS payments R_t by making connections with existing MAP results. In the following proposition, connections to the LT of S_t (given by (4.3)) are made to obtain the joint LT of S_t and R_t.

Proposition 16. The joint LT of the settled payments S_t and RBNS payments R_t, conditional on $J_0 = i$ is given by

$$E\left[e^{-sS_t - zR_t}\middle|\, J_0 = i\right] = \mathbf{e}_{n+m+1,i} e^{(\mathbf{D}_0 + \mathbf{D}_1 \circ \mathbf{H}(s))t} \mathbf{a}(z),$$

where $\mathbf{a}(z)$ is a column-vector given by

$$\mathbf{a}(z) = \left(\widetilde{k}(z)^j\right)_{j\in E},$$

and the elements of $\mathbf{D}_0$, $\mathbf{D}_1$, and $\mathbf{H}(z)$ are given, respectively, by (4.9), (4.7), and (4.8).

Also, by making connections with results for the moments of S_t (see, e.g., Ren [2008]), we have the following proposition for the joint moments of S_t and R_t.

Proposition 17. For c, d non-negative integers, the joint moments of S_t and R_t are given by

$$E\left[S_t^c R_t^d\middle|\, J_0 = i\right] = (-1)^c \mathbf{g}_{c,i}(t)\mathbf{b}_d,$$

where $\mathbf{g}_{c,i}(t)$ is a row vector given by

$$\mathbf{g}_{c,i}(t) = \left((-1)^c E\left[S_t^c \mathbf{1}_{\{J_t = j\}}\middle|\, J_0 = i\right]\right)_{j\in E}$$

and $\mathbf{b}_d$ is a column vector given by

$$\mathbf{b}_d = \left(E\left[(jY)^d\right]\right)_{j\in E} = \left((-j)^d \left.\frac{d^d}{dz^d}\widetilde{k}(z)\right|_{z=0}\right)_{j\in E}.$$

We have $g_{0,i}(t) = \mathbf{e}_{n+m+1,i} e^{\mathbf{Q}t}$ and for $c \geq 1$, $\mathbf{g}_{c,i}(t)$ can be found by numerically solving the following system of differential equations (see Ren [2008])

$$\frac{d\mathbf{g}_{c,i}(t)}{dt} = \mathbf{g}_{c,i}(t)\mathbf{Q} + \sum_{k=0}^{c-1} \binom{c}{k} (-1)^{c-k} \mathbf{g}_{k,i}(t)\left(\mathbf{D}_1 \circ \Delta_{n-k}\right),$$

where

$$\Delta_k = \left(E\left[P_{i,j}^k\right]\right)_{i,j\in E} = (-1)^k \left.\frac{d^k}{dz^k}\mathbf{H}(z)\right|_{z=0}.$$

In what follows, we focus our attention on the RBNS payments R_t. Suppose

$$e^{\mathbf{Q}t} = \left(P\left(J_t = j | J_0 = i\right)\right)_{i,j \in E} \equiv \left(a_{i,j}(t)\right)_{i,j \in E},$$

By Proposition 16 with $s = 0$, the LT of R_t conditional on $J_0 = i$ is given by

$$E\left[e^{-zR_t} \middle| J_0 = i\right] = \sum_{j=0}^{n+m} a_{i,j}(t)\widetilde{k}(z)^j = A_{i,t}\left(\widetilde{k}(z)\right), \tag{4.11}$$

where $A_{i,t}(z) = \sum_{j=0}^{n+m} a_{i,j}(t)z^j$ is a probability generating function (pgf) for the discrete probability distribution $\{a_{i,0}(t), a_{i,1}(t), \ldots, a_{i,n+m}(t)\}$. Identifying the distribution of R_t and studying quantities such as the expectation, Value-at-Risk, and Conditional Tail Expectation of R_t serves to quantify the risk of the insurer's future liability arising from claims that have been reported but not yet settled due to claim investigation congestion. Furthermore, it allows for analysis of the impact of various CIMs on R_t.

Remark 3. We can also consider R_t when t is large to assess the risk due to reported claims that remain unpaid in the long-run due to claims investigation congestion. This analysis can provide an indicator of the overall risk of an insurer's CIM associated with RBNS claims. Let $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_{n+m})$ be the stationary distribution of the background process J which can be found by solving $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$ and $\boldsymbol{\pi}\mathbb{1}_{n+m+1} = 1$. We have that the stationary distribution $\boldsymbol{\pi}$ exists and is unique since $\mathbf{J}$ is an irreducible CTMC with finite state space (see [Ross, 2007, Chapter 6]).

Then the LT of R_t in the long-run (i.e., at some arbitrary time in the distant future) is given by

$$\lim_{t\to\infty} E\left[e^{-zR_t} \middle| J_0 = i\right] = \sum_{j=0}^{n+m} \pi_j \widetilde{k}(z)^j. \tag{4.12}$$

4.4 Distribution of the RBNS payments

In this section, we identify the distribution of RBNS payments R_t (conditional on $J_0 = i$) under some assumptions including an investigation strategy that was discussed in Chapter 2 which impacts $q_1(x)$.

Recall that following an investigation, the amount claimed x will result in a payment w ($w \geq 0$) with df $L_x(w)$. Suppose that a claim of size x ($x \geq 0$) selected for investigation is either determined fraudulent and denied with probability $\varsigma(x)$ or paid in its entirety with probability $1 - \varsigma(x)$. Then $L_x(w)$ is given by (2.2) and

$$\begin{aligned} \widetilde{l}_x(z) &= \int_{[0,\infty)} e^{-zw} L_x(dw) \\ &= \varsigma(x) + (1 - \varsigma(x))\, e^{-zx}. \end{aligned} \tag{4.13}$$

Also suppose that the probability of a fraudulent claim increases with its size. More precisely, suppose that (see Section 2.4.4)

$$\varsigma(x) = \left(1 - e^{-\gamma x}\right) \varsigma_0, \tag{4.14}$$

where $0 \leq \varsigma_0 \leq 1$ and $\gamma \geq 0$. As a result, from (4.6) and using (4.14), the payment size of a claim undergoing investigation has LT

$$\begin{aligned} \widetilde{k}(z) &= \int_0^\infty \widetilde{l}_x(z) p_1(x) dx \\ &= (1 - \widetilde{p}_1(\gamma))\, \varsigma_0 + (1 - \varsigma_0)\, \widetilde{p}_1(z) + \varsigma_0 \widetilde{p}_1(\gamma) \widetilde{p}^*_{1,\gamma}(z), \end{aligned} \tag{4.15}$$

where

$$p^*_{1,\gamma}(x) = \frac{e^{-\gamma x} p_1(x)}{\widetilde{p}_1(\gamma)},$$

is the Esscher transform density of p_1 (see, e.g., Gerber and Shiu [1994]) with LT $\widetilde{p}^*_{1,\gamma}(z) = \widetilde{p}_1(\gamma + z)/\widetilde{p}_1(\gamma)$.

It follows from (4.15), that

$$\widetilde{k}(z) \;=\; \xi + (1-\xi)\,\widetilde{h}(z), \tag{4.16}$$

where

$$\xi = (1 - \widetilde{p}_1(\gamma))\,\varsigma_0, \tag{4.17}$$

and

$$\widetilde{h}(z) \;=\; \frac{1}{1-\xi}\left((1-\varsigma_0)\,\widetilde{p}_1(z) + \varsigma_0 \widetilde{p}_1(\gamma)\widetilde{p}^{*}_{1,\gamma}(z)\right). \tag{4.18}$$

Thus,

$$\widetilde{k}(z) \;=\; P\left(\widetilde{h}(z)\right), \tag{4.19}$$

where $P(z) = \xi + (1-\xi)\,z$ is a Bernoulli pgf. That is, if $\widetilde{l}_x(w)$ and $\varsigma(x)$ are respectively given by (4.13) and (4.14), k is distributed as a compound Bernoulli rv with secondary LT given by (4.18) which is a mixture of p_1 and its Esscher transform density $p^{*}_{1,\gamma}$. Thus, from (4.11), it follows that R_t (conditional on $J_0 = i$) is distributed as a compound rv with primary pgf $A_{i,t}(z)$ and secondary LT a compound Bernoulli rv with LT (4.19).

In what follows, we further particularize $\widetilde{k}(z)$ under the following assumption on $q_1(x)$:

$$\begin{aligned} q_1(x) \;&=\; \theta + (1-\theta)(1-e^{-\kappa x}) \\ &=\; 1 - (1-\theta)e^{-\kappa x}, \end{aligned} \tag{4.20}$$

where $0 \leq \theta \leq 1$ and $\kappa \geq 0$ (see Section 2.4.3). Recall that $q_1(x)$ is the probability that a claim of size x is selected for investigation. Thus, from (4.20), on average, a percentage θ of all claims are automatically subjected to investigation and the other $(1-\theta)$ are subjected based on their claim amount with probability $1 - e^{-\kappa x}$ for a claim of size x. Recall from (4.4) and (4.5) that claims that have been selected for investigation have density

$$p_1(x) \;=\; \frac{\lambda}{\lambda_1} p(x) q_1(x).$$

If $q_1(x)$ is given by (4.20), then we have

$$\widetilde{p}_1(z) \;=\; \frac{\lambda}{\lambda_1}\left(\widetilde{p}(z) - (1-\theta)\widetilde{p}(\kappa)\widetilde{p}^*_\kappa(z)\right), \tag{4.21}$$

where

$$\frac{\lambda}{\lambda_1} = \frac{1}{1-(1-\theta)\,\widetilde{p}(\kappa)}, \tag{4.22}$$

and

$$p^*_\kappa(x) = \frac{e^{-\kappa x}p(x)}{\widetilde{p}(\kappa)},$$

is the Esscher transform density of p with LT $\widetilde{p}^*_\kappa(z) = \widetilde{p}(\kappa+z)/\widetilde{p}(\kappa)$. Thus, if $q_1(x)$ is given by (4.20), then from (4.21), p_1 is a combination of p and its Esscher transform density p^*_κ.

Now suppose $q_1(x)$ is given by (4.20) and $\varsigma(x)$ is given by (4.14), then we can show that

$$\widetilde{h}(z) = C_0\left(C_1\widetilde{p}(z) + C_2\widetilde{p}^*_\kappa(z) + C_3\widetilde{p}^*_\gamma(z) + C_4\widetilde{p}^*_{\kappa+\gamma}(z)\right), \tag{4.23}$$

where

$$\begin{aligned} C_1 &= (1-\varsigma_0), & C_2 &= -(1-\varsigma_0)(1-\theta)\,\widetilde{p}(\kappa) \\ C_3 &= \varsigma_0\widetilde{p}(\gamma), & C_4 &= -\varsigma_0(1-\theta)\widetilde{p}(\kappa+\gamma), \end{aligned} \tag{4.24}$$

and from (4.17), (4.22), and (4.24), $C_0 = \frac{\lambda}{\lambda_1(1-\xi)} = (C_1+C_2+C_3+C_4)^{-1}$. Thus, from (4.19), it follows that k is distributed as a compound Bernoulli rv with secondary LT given by (4.23) which is a combination of p and its Esscher transform densities with parameters κ, γ, and $\kappa+\gamma$.

Note that if $\gamma = \infty$ such that from (4.14), $\varsigma(x) = \varsigma_0$, then (4.23) simplifies to

$$\widetilde{h}(z) = \frac{1}{\left(1-(1-\theta)\widetilde{p}(\kappa)\right)(1-\varsigma_0)}\left((1-\varsigma_0)\widetilde{p}(z) - (1-\varsigma_0)(1-\theta)\widetilde{p}(\kappa)\widetilde{p}^*_\kappa(z)\right),$$

i.e., $\widetilde{h}$ is the LT of a combination of p and its Esscher transform density p^*_κ. Instead, if $\kappa = \infty$ such that from (4.20), $q_1(x) = 1$, then $C_2 = C_4 = 0$ and (4.23) simplifies to

$$\widetilde{h}(z) = \frac{1}{1-(1-\widetilde{p}_1(\gamma))\,\varsigma_0}\left((1-\varsigma_0)\widetilde{p}(z) + \varsigma_0\widetilde{p}(\gamma)\widetilde{p}^*_\gamma(z)\right),$$

i.e., $\widetilde{h}$ is a mixture of p and its Esscher transform density p^*_γ.

Next, we present three examples where we identify the distribution of the total RBNS payments R_t (conditional on $J_0 = i$) under various distributional assumptions for the claim size density p. We also assume $q_1(x)$ and $\varsigma(x)$ are respectively given by (4.14) and (4.20).

Example 1. Mixture of Erlangs distributed claim size

Suppose the claim size is mixed Erlang distributed with density

$$p(x) = \sum_{j=1}^{\infty} q_j e_j(x), \qquad x > 0, \tag{4.25}$$

where

$$e_j(x) = \frac{\beta\,(\beta x)^{j-1}\, e^{-\beta x}}{(j-1)!}, \qquad x > 0,$$

for $\beta > 0$ and $\mathbf{q} = \{q_1, q_2, \ldots\}$ is a discrete probability distribution with pgf $Q(z) = \sum_{j=1}^{\infty} q_j z^j$. It follows that the LT of (4.25) is given by

$$\begin{aligned}\widetilde{p}(z) &= \sum_{j=1}^{\infty} q_j \left(\frac{\beta}{\beta+z}\right)^j \\ &= Q\left(\frac{\beta}{\beta+z}\right).\end{aligned} \tag{4.26}$$

It is well-known that the mixed Erlang class of distributions is very large and is also dense in the class of positive continuous probability distributions (see, e.g., Willmot and Lin [2011] and references therein).

In [Willmot and Woo, 2018, Example 1.4.6], it is shown that for $\mu > 0$, $\widetilde{p}^*_\mu(z) = \widetilde{p}(\mu + z)/\widetilde{p}(\mu)$ can be expressed as

$$\widetilde{p}^*_\mu(z) = Q_\mu\left(\frac{\beta+\mu}{\beta+\mu+z}\right), \tag{4.27}$$

where $Q_\mu(z) = \sum_{n=1}^{\infty} q_{n,\mu} z^n$ is a pgf with

$$q_{n,\mu} = \frac{\left(\frac{\beta}{\beta+\mu}\right)^n q_n}{Q\left(\frac{\beta}{\beta+\mu}\right)}, \tag{4.28}$$

which has a discrete Esscher form. That is, the Esscher transform density of a mixture of Erlangs is also a mixture of Erlangs. We define $\mathbf{q}_\mu = \{q_{1,\mu}, q_{2,\mu}, \ldots\}$.

Consider the following algebraic identity (see, e.g., Willmot and Lin [2011])

$$\frac{\beta}{\beta+z} = \frac{\beta+\mu}{\beta+\mu+z}\left(\frac{\frac{\beta}{\beta+\mu}}{1-\left(1-\frac{\beta}{\beta+\mu}\right)\frac{\beta+\mu}{\beta+\mu+z}}\right). \tag{4.29}$$

Using (4.29), (4.26) may be expressed as (see [Willmot and Woo, 2007, pg. 103] for further details)

$$\widetilde{p}(z) = Q\left(\frac{\beta}{\beta+z}\right) = D_{\beta,\mu,\mathbf{q}}\left(\frac{\beta+\mu}{\beta+\mu+z}\right), \tag{4.30}$$

where $D_{\beta,\mu,\mathbf{q}}(z) = \sum_{n=1}^{\infty} d_{n,\beta,\mu,\mathbf{q}} z^n$ is a pgf with

$$d_{n,\beta,\mu,\mathbf{q}} = \sum_{j=1}^{n} q_j \binom{n-1}{n-j}\left(\frac{\beta}{\beta+\mu}\right)^j\left(1-\frac{\beta}{\beta+\mu}\right)^{n-j}. \tag{4.31}$$

That is, a mixture of Erlangs can be expressed as a different mixture of Erlangs with larger scale parameter.

Using (4.26) and (4.27), it follows that (4.23) is given by

$$\begin{aligned}\widetilde{h}(z) \quad &= \quad C_0\Bigg\{C_1 Q\left(\frac{\beta}{\beta+z}\right) + C_2 Q_\kappa\left(\frac{\beta+\kappa}{\beta+\kappa+z}\right) + C_3 Q_\gamma\left(\frac{\beta+\gamma}{\beta+\gamma+z}\right) \\ &+ \quad C_4 Q_{\kappa+\gamma}\left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right)\Bigg\},\end{aligned} \tag{4.32}$$

where $C_1, C_2, C_3,$ and C_4 are given by (4.24). Additionally, using (4.30) on the first three terms of (4.32), we obtain

$$\widetilde{h}(z) = \sum_{n=1}^{\infty} d_n \left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right)^n, \tag{4.33}$$

where

$$d_n = C_0\left(C_1 d_{n,\beta,\kappa+\gamma,\mathbf{q}} + C_2 d_{n,\beta+\kappa,\gamma,\mathbf{q}_\kappa} + C_3 d_{n,\beta+\gamma,\kappa,\mathbf{q}_\gamma} + C_4 q_{n,\kappa+\gamma}\right). \tag{4.34}$$

Using (4.28) and (4.31), it is easy to show that (4.34) is given by

$$d_n = \frac{C_0}{(\beta+\kappa+\gamma)^n}\left(\sum_{j=1}^{n}\binom{n-1}{n-j}q_j\beta^j r_{n-j} - \varsigma_0(1-\theta)q_n\beta^n\right), \tag{4.35}$$

where

$$\begin{aligned} r_{n-j} &= (1-\varsigma_0)\left(\kappa+\gamma\right)^{n-j} - (1-\varsigma_0)(1-\theta)\gamma^{n-j} + \varsigma_0\kappa^{n-j} \\ &= (1-\varsigma_0)\left[(\kappa+\gamma)^{n-j} - (1-\theta)\gamma^{n-j}\right] + \varsigma_0\kappa^{n-j}. \end{aligned}$$

By removing the last term from the summation in (4.35) and simplifying, we have that

$$d_n = \frac{C_0}{(\beta+\kappa+\gamma)^n}\left(\sum_{j=1}^{n-1}\binom{n-1}{n-j}q_j\beta^j r_{n-j} + q_n\beta^n\theta\right),$$

where we adopt the notational convention that the empty sum is 0. It is easy to see that $r_{n-j} > 0$ for $j < n$ and thus, $d_n > 0$ for $n \geq 1$. Therefore, (4.33) can be expressed as

$$\widetilde{h}(z) = D\left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right),$$

where $D(z) = \sum_{n=0}^{\infty} d_n z^n$ is a pgf. Also, from (4.19), $\widetilde{k}(z)$ is distributed as a mixed Erlang with LT

$$\widetilde{k}(z) = B\left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right),$$

where $B(z) = \sum_{n=0}^{\infty} b_n z^n = P(D(z))$ is a compound Bernoulli rv with secondary pgf $D(z)$. Since P is a pgf of a member of the $(a, b, 1)$ class, recursive techniques are available to easily evaluate b_n (see, e.g., [Klugman et al., 2013, Chapter 6]). Furthermore, from (4.11), we have

$$E\left[e^{-zR_t}\middle| J_0 = i\right] = G\left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right),$$

where $G(z) = \sum_{n=0}^{\infty} g_n z^n = A_{i,t}(B(z))$ is a compound rv with primary pgf $A_{i,t}(z)$ and secondary pgf $B(z)$. Thus, the RBNS payments R_t is distributed as mixed Erlang rv with mixing weights $\{g_1, g_2, \ldots\}$. Evaluation of $\{g_1, g_2, \ldots\}$ is a well understood problem in loss analysis (see, e.g., [Klugman et al., 2013, Section 3.4]). For example, if $A_{i,t}(z)$ is the pgf of a distribution from the $(a, b, 1)$ class, the mixing weights $\{g_0, g_1, \ldots\}$ are easily obtained.

Example 2. Gamma distributed claim size

For $\beta > 0$ and $\alpha > 0$, suppose the claim size density is gamma distributed with density

$$p(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \qquad x > 0, \tag{4.36}$$

and LT

$$\widetilde{p}(z) = \left(\frac{\beta}{\beta+z}\right)^{\alpha}.$$

We have that the Esscher transform density of a gamma rv is also a gamma rv with different scale parameter, i.e., for $\mu > 0$,

$$\widetilde{p}_{\mu}^{*} = \frac{\widetilde{p}(\mu+z)}{\widetilde{p}(\mu)} = \left(\frac{\beta+\mu}{\beta+\mu+z}\right)^{\alpha}.$$

Using the algebraic identity (4.29), we have

$$\begin{aligned}
\left(\frac{\beta}{\beta+z}\right)^{\alpha} &= \left(\frac{\beta+\mu}{\beta+\mu+z}\right)^{\alpha} \left(\frac{\frac{\beta}{\beta+\mu}}{1-\left(1-\frac{\beta}{\beta+\mu}\right)\frac{\beta+\mu}{\beta+\mu+z}}\right)^{\alpha} \\
&= \left(\frac{\beta+\mu}{\beta+\mu+z}\right)^{\alpha} \sum_{n=0}^{\infty} f_{n,\beta,\mu} \left(\frac{\beta+\mu}{\beta+\mu+z}\right)^{n} \\
&= \sum_{n=0}^{\infty} f_{n,\beta,\mu} \left(\frac{\beta+\mu}{\beta+\mu+z}\right)^{\alpha+n} \qquad (4.37)
\end{aligned}$$

where

$$f_{n,\beta,\mu} = \binom{n+\alpha-1}{n} \left(\frac{\beta}{\beta+\mu}\right)^{\alpha} \left(1-\frac{\beta}{\beta+\mu}\right)^{n}, \tag{4.38}$$

corresponding to the negative binomial probabilities (see, e.g., [Willmot and Lin, 2011, Example 2]).

Then using (4.37), it follows that (4.23) is given by

$$\begin{aligned}
\widetilde{h}(z) &= C_0 \Bigg\{ C_1 \left(\frac{\beta}{\beta+\kappa}\right)^{\alpha} + C_2 \left(\frac{\beta+\kappa}{\beta+\kappa+z}\right)^{\alpha} + C_3 \left(\frac{\beta+\gamma}{\beta+\gamma+z}\right)^{\alpha} \\
&+ C_4 \left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right)^{\alpha} \Bigg\}, \qquad (4.39)
\end{aligned}$$

where $C_1, C_2, C_3,$ and C_4 are given by (4.24) and using (4.37) on the first three terms of (4.39), we have

$$\widetilde{h}(z) = \sum_{n=0}^{\infty} d_n \left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right)^{\alpha+n}, \tag{4.40}$$

where

$$d_0 = C_0 \left(C_1 f_{0,\beta,\kappa+\gamma} + C_2 f_{0,\beta+\kappa,\gamma} + C_3 f_{0,\beta+\gamma,\kappa} + C_4\right) \tag{4.41}$$

and for $n \geq 1$

$$d_n = C_0 \left(C_1 f_{n,\beta,\kappa+\gamma} + C_2 f_{n,\beta+\kappa,\gamma} + C_3 f_{n,\beta+\gamma,\kappa}\right). \tag{4.42}$$

Using (4.38), it is easy to show that (4.41) is given by

$$d_0 = C_0 \left(\frac{\beta}{\beta+\kappa+\gamma}\right)^{\alpha} \theta,$$

and for $n \geq 1$, (4.42) is given by

$$d_n = C_0 \binom{n+\alpha-1}{n} \left(\frac{\beta}{\beta+\kappa+\gamma}\right)^{\alpha} \frac{(1-\varsigma_0)\left[(\kappa+\gamma)^n - (1-\theta)\gamma^n\right] + \varsigma_0 \kappa^n}{(\beta+\kappa+\gamma)^n}.$$

Since $(\kappa+\gamma)^n - (1-\theta)\gamma^n > 0$ for $n \geq 1$, we have that $d_n > 0$ for $n \geq 0$ and it follows that (4.40) is the LT of a mixture of gammas.

Thus, from (4.11) and (4.19),

$$\begin{aligned}
& E\left[e^{-zR_t}\middle|\, J_0 = i\right] \\
= & \sum_{j=0}^{n+m} a_{i,j}(t) \left(\xi + (1-\xi) \sum_{r=0}^{\infty} d_r \left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right)^{r+\alpha}\right)^j \\
= & \sum_{j=0}^{n+m} a_{i,j}(t) \sum_{l=0}^{j} \binom{j}{l} \xi^{j-l} (1-\xi)^l \left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right)^{\alpha l} \left(\sum_{r=0}^{\infty} d_r \left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right)^{r}\right)^l,
\end{aligned}$$

using binomial expansion. If we define the pgf $D_l^*(z) = \sum_{r=0}^{\infty} d_{r,l}^* z^r = D(z)^l$, then we have

$$E\left[e^{-zR_t}\middle|\, J_0 = i\right] = \sum_{r=0}^{\infty} \sum_{l=0}^{n+m} g_{r,l} \left(\frac{\beta+\kappa+\gamma}{\beta+\kappa+\gamma+z}\right)^{r+\alpha l}, \tag{4.43}$$

by changing the order of summation, where

$$g_{j,l} = \sum_{j=l}^{n+m} a_{i,j}(t) \binom{j}{l} \xi^{j-l} (1-\xi)^l d_{r,l}^*,$$

Thus, the RBNS payments R_t (conditional on $J_0 = i$) is distributed as a mixture of gammas with LT (4.43).

Example 3. Inverse Gaussian distributed claim size

Suppose the claim size follows an inverse Gaussian distribution with density

$$p(x) = \frac{\mu}{\sqrt{2\pi\sigma x^3}} e^{-\frac{(x-\mu)^2}{2\sigma x}}, \qquad x > 0, \tag{4.44}$$

where $\sigma > 0$ and $\mu > 0$. It follows that the LT of (4.44) is given by

$$\widetilde{p}(z) = e^{-\frac{\mu}{\sigma}\left(\sqrt{1+2\sigma z}-1\right)} \equiv G\left(z;\mu,\sigma\right).$$

We can show that the Esscher transform density of p is also inverse Gaussian distributed, i.e.,

$$\begin{aligned}
\widetilde{p}_\kappa^*(z) = \frac{\widetilde{p}(z+\kappa)}{\widetilde{p}(\kappa)} &= exp\left\{-\frac{\mu}{\sigma}\left(\sqrt{1+2\sigma(z+\kappa)}-1\right) + \frac{\mu}{\sigma}\left(\sqrt{1+2\sigma\kappa}-1\right)\right\} \\
&= e^{-\frac{\mu_\kappa}{\sigma_\kappa}\left(\sqrt{1+2\sigma_\kappa z}-1\right)} \\
&= G(z;\mu_\kappa,\sigma_\kappa)
\end{aligned} \tag{4.45}$$

where $\mu_\kappa = \mu/\sqrt{1+2\sigma\kappa}$ and $\sigma_\kappa = \sigma/\left(1+2\sigma\kappa\right)$.

Thus, using (4.45), it follows that (4.23) is given by

$$\widetilde{h}(z) = C_0\Big(C_1 G(z;\mu,\sigma) + C_2 G(z;\mu_\kappa,\sigma_\kappa) + C_3 G(z;\mu_\gamma,\sigma_\gamma) + C_4 G(z;\mu_{\kappa+\gamma},\sigma_{\kappa+\gamma})\Big), \tag{4.46}$$

It follows from (4.19) that $\widetilde{k}(z)$ is a compound Bernoulli LT where the secondary distribution is a combination of inverse Gaussian distributions with LT given by (4.46).

Thus, from (4.11), it follows that R_t (conditional on $J_0 = i$) is distributed as a compound rv with primary pgf $A(z)$ and secondary LT a compound Bernoulli rv with LT (4.19) where $\widetilde{h}(z)$ is given by (4.46).

Further Remarks

Up to this point, we have considered $\varsigma(x)$ and $q_1(x)$ to be respectively given by (4.14) and (4.20) which both contain an exponential tail. We remark that this section's analysis can be easily extended if $\varsigma(x)$ and $q_1(x)$ are generalized to instead contain a mixed Erlang tail. In what follows, we present a brief outline of how the analysis would be extended. Suppose

$$\varsigma(x) = \left(1 - \sum_{j=0}^{\infty} \frac{\overline{F}_j}{j!} (\gamma x)^j e^{-\gamma x}\right) \varsigma_0, \tag{4.47}$$

where $\overline{F}_j = \sum_{i=j+1}^{\infty} f_i$ and $\{f_1, f_2, \ldots\}$ is a discrete probability distribution and

$$q_1(x) \quad = \quad 1 - (1-\theta) \sum_{n=0}^{\infty} \frac{\overline{G}_n}{n!} (\kappa x)^n e^{-\kappa x}, \tag{4.48}$$

where $\overline{G}_n = \sum_{j=n+1}^{\infty} g_j$ and $\{g_1, g_2, \ldots\}$ is a discrete probability distribution. Note that (4.14) and (4.20) are respectively recovered from (4.47) and (4.48) by letting $f_1 = 1$ and $g_1 = 1$.

Define the density

$$h_{n,\kappa}(x) = \frac{(\kappa x)^n e^{-\kappa x} p(x)}{c_{n,\kappa}}, \qquad x > 0, \tag{4.49}$$

where $c_{n,\kappa}$ is easily obtained using $c_{n,\kappa} = (-\kappa)^n \widetilde{p}^{(n)}(\kappa)$. Note that when $n = 0$, $h_{0,\kappa}$ is the Esscher transform density of p with parameter κ.

Recalling that $p_1(x) = \frac{\lambda}{\lambda_1} p(x) q_1(x)$ and assuming $q_1(x)$ is given by (4.48), it follows that

$$\widetilde{p}_1(z) \quad = \quad \frac{\lambda}{\lambda_1} \left(\widetilde{p}(z) - (1-\theta) \sum_{n=0}^{\infty} \frac{\overline{G}_n}{n!} c_{n,\kappa} \widetilde{h}_{n,\kappa}(z)\right), \tag{4.50}$$

i.e., p_1 is a combination of p and $\{h_{n,\kappa}\}_{n=0}^{\infty}$. Also, using (4.13), (4.47) and (4.50), it can be shown that

$$\begin{aligned}\widetilde{k}(z) &= \int_0^\infty \widetilde{l}_x(z)p_1(x)dx \\ &= \xi^* + (1-\xi^*)\widetilde{h}^*(z), \end{aligned} \tag{4.51}$$

where $\xi^* = \varsigma_0\left(1 - \sum_{j=0}^{\infty}\overline{F}_j\widetilde{p}_1^{(j)}(\gamma)/j!\right)$ and

$$\widetilde{h}^*(z) = \frac{1}{C_0^*}\left(C_1^*\widetilde{p}(z) + \sum_{n=0}^{\infty}C_{2,n}^*\widetilde{h}_{n,\kappa}(z) + \sum_{j=0}^{\infty}C_{3,j}^*\widetilde{h}_{j,\gamma}(z) + \sum_{j=0}^{\infty}\sum_{n=0}^{\infty}C_{4,j,n}^*\widetilde{h}_{n+j,\kappa+\gamma}(z)\right), \tag{4.52}$$

where

$$C_1^* = 1-\varsigma, \qquad C_{2,n}^* = -(1-\varsigma_0)(1-\theta)\frac{\overline{G}_n}{n!}c_{n,\kappa},$$
$$C_{3,j}^* = \varsigma_0\frac{\overline{F}_j}{j!}c_{j,\gamma}, \qquad C_{4,n,j}^* = -\varsigma_0(1-\theta)\frac{\overline{F}_j}{j!}\frac{\overline{G}_n}{n!}\left(\frac{\gamma}{\kappa+\gamma}\right)^j\left(\frac{\kappa}{\kappa+\gamma}\right)^n c_{n+j,\kappa+\gamma},$$

and

$$C_0^* = C_1^* + \sum_{n=0}^{\infty}C_{2,n}^* + \sum_{j=0}^{\infty}C_{3,j}^* + \sum_{j=0}^{\infty}\sum_{n=0}^{\infty}C_{4,j,n}^*.$$

Thus, from (4.51), it follows that k is distributed as a compound Bernoulli rv with secondary LT given by (4.52) which is a combination of p and the transforms of p given by (4.49) with various parameters. Note that if $f_1 = g_1 = 1$, we recover (4.16) and (4.23) respectively from (4.51) and (4.52).

Suppose that $p(x)$ is mixed Erlang distributed with density (4.25) (see Example 1). It would follow that $h_{n,\kappa}(x)$ given by (4.49) is a mixed Erlang density given by

$$h_{n,\kappa}(x) = \sum_{j=1}^{\infty}q_{j,n,\kappa}\frac{(\beta+\kappa)^{n+j}\,x^{n+j-1}e^{-(\beta+\kappa)x}}{(n+j-1)!},$$

where

$$q_{j,n,\kappa} = \frac{1}{c_{n,\kappa}}q_j\frac{(n+j-1)!}{(j-1)!}\left(\frac{\beta}{\beta+\kappa}\right)^j\left(\frac{\kappa}{\beta+\kappa}\right)^n,$$

and $c_{n,\kappa} = \sum_{j=1}^{\infty} q_j \frac{(n+j-1)!}{(j-1)!} \left(\frac{\beta}{\beta+\kappa}\right)^j \left(\frac{\kappa}{\beta+\kappa}\right)^n$. Thus,

$$\widetilde{h}_{n,\kappa}(z) = Q_{n,\kappa}\left(\frac{\beta+\kappa}{\beta+\kappa+z}\right),$$

where $Q_{n,\kappa}(z) = z^n \sum_{j=1}^{\infty} q_{j,n,\kappa} z^j$ is a pgf of the distribution $\{q_{j,n,\kappa}\}_{j=1}^{\infty}$ shifted to the right by n. It also follows that h^* given by (4.52) is a mixed Erlang density; the details are tedious and involve similar techniques employed in Example 1 and as a result, they are omitted.

Now suppose that p is gamma distributed with density (4.36) (see Example 2), then $h_{n,\kappa}(x)$ given by (4.52) has the gamma density

$$h_{n,\kappa}(x) = \frac{(\beta+\kappa)^{n+j}}{\Gamma(n+\alpha)} x^{n+\alpha-1} e^{-(\beta+\kappa)x}, \qquad x > 0,$$

and using similar techniques employed in Example 2, it can be shown that h^* given by (4.52) is a mixed gamma density. The details are tedious and thus omitted.

Finally, suppose p is inverse Gaussian distributed (see Example 3) with density given by (4.44), then from (4.49), we can show that

$$h_{n,\kappa}(x) = \frac{\left(\frac{a}{b}\right)^{\frac{p}{2}}}{2K_p\left(\sqrt{ab}\right)} x^{p-1} e^{-\left(ax+\frac{b}{x}\right)/2}, \qquad x > 0,$$

where $p = n - 1/2$, $a = 2k + 1/\sigma$, $b = \mu^2/\sigma$, and K_p is a modified Bessel function of the second kind, i.e., $h_{n,\kappa}(x)$ is a generalized inverse Gaussian density.

4.5 Generalizations to the claim investigation mechanism

Considerable generalizations to the CIM proposed in Section 3.2.1 are possible while preserving the Markovian arrival process formulation of the aggregate payment process presented in Section 4.2. We refer readers to Section 3.5 for a detailed discussion of such generalizations.

Generalizations to the CIM of the type discussed in Section 3.5 involve viewing $J = \{J_t\}_{t\geq 0}$ more generally as the state process of the CIM. Suppose we define the generalized state process by the CTMC $\mathbf{J} = \{\mathbf{J}_t\}_{t\geq 0}$ with generalized state space E^*. As discussed in Section 4.2, the matrices $\mathbf{D}_0$ and $\mathbf{D}_1$ are respectively found by determining the transition rates of $\mathbf{J}$ that occur with and without an accompanying claim and the infinitestimal generator of $\mathbf{J}$ is given by $\mathbf{Q} = \mathbf{D}_0 + \mathbf{D}_1$. Also, we have $\mathbf{H}(s) = \left(\widetilde{h}_{\mathbf{i},\mathbf{j}}(s)\right)_{\mathbf{i},\mathbf{j}\in E^*}$ where $\widetilde{h}_{\mathbf{i},\mathbf{j}}$ is the LT of a claim accompanying a transition in $\mathbf{J}$ from state $\mathbf{i}$ to $\mathbf{j}$.

The joint LT of the settled payments S_t and RBNS payments R_t (conditional on $\mathbf{J}_0$) under the generalized CIM is easily determined from Proposition 16 by replacing $a(z)$ with $\mathbf{a}^*(z) = \left(\widetilde{a}_{\mathbf{j}}(z)\right)_{\mathbf{j}\in E^*}$ where $\widetilde{a}_{\mathbf{j}}(z)$ denotes the LT of the RBNS payments in the CIM when $\mathbf{J} = \mathbf{j}$. Similarly, the joint moments of S_t and R_t can be found using Proposition 17 with $\mathbf{b}_d$ replaced by $\mathbf{b}^*_d = \left((-1)^d \frac{d^d}{dz^d}\widetilde{a}_{\mathbf{j}}(z)\Big|_{z=0}\right)_{\mathbf{j}\in E^*}$. Also if $\boldsymbol{\pi}^*$ is the stationary distribution of the background process $\mathbf{J}$ which can be found by solving $\boldsymbol{\pi}^*\mathbf{Q} = \mathbf{0}$ and $\boldsymbol{\pi}^*\mathbb{1}_{n+m+1} = 1$, then from (4.12), the LT of R_t in the long-run is given by

$$\lim_{t\to\infty} E\left[e^{-zR_t}\middle|\, \mathbf{J}_0 = \mathbf{i}\right] = \boldsymbol{\pi}^*\mathbf{a}^*(z). \tag{4.53}$$

4.6 Numerical Examples

This section presents numerical examples which quantify the impact of variations in the CIM on an insurer's RBNS payments with the objective of measuring the risk associated with claims congestion caused by the claims handling process. Recall that Chapter 2 and Chapter 3 also considered variations in the CIM however their impact on the ruin probability was measured. In particular, the following numerical examples compare the Value-at-Risk (VaR) and Tail Value-at-Risk ($TVaR$) of R_t for various CIMs.

The following examples each revisit scenarios considered earlier in Chapter 2 and Chapter

3. Also, all the examples except the first implicitly assume that $J_0 = 0$, i.e., the CIM is empty at time 0.

4.6.1 Varying time horizons

In this example, we consider the VaR and $TVaR$ of the $RBNS$ payments R_t at various t, conditional on the number of claims in the CIM at time 0. We make the same assumptions as in Example 2.4.1 with $q_1(x) = q_1 = 0.05$, except that the investigation time here is instead assumed to follow an exponential distribution with rate $\alpha = 0.2$. The LT of R_t is given by (4.11) when $t < \infty$ and by (4.12) when $t = \infty$ (such that the CIM is in steady-state). Table 4.1 and Table 4.2 respectively present the VaR and $TVaR$ of the RBNS payments R_t at various t given $J_0 = 0$ and $J_0 = 1$.

Table 4.1: VaR and $TVaR$ of RBNS payments for varying t when $J_0 = 0$

t	$VaR_{0.85}$	$TVaR_{0.85}$	$VaR_{0.9}$	$TVaR_{0.9}$	$VaR_{0.95}$	$TVaR_{0.95}$	$VaR_{0.995}$	$TVaR_{0.995}$
0.5	0	1.2741	0	2.8667	2.5385	4.6481	7.3604	9.3295
1	1.0771	3.5145	2.2887	4.4228	3.8085	5.8674	8.5384	10.4757
5	3.3974	5.4653	4.2492	6.3001	5.6933	7.7110	10.3096	12.2080
10	3.6094	5.6723	4.4584	6.5054	5.9010	7.9129	10.5033	12.3980
∞	3.6329	5.6953	4.4818	6.5283	5.9241	7.9354	10.5249	12.4192

From Table 4.1 and Table 4.2, we are able to see how the VaR and $TVaR$ values evolve over time t conditional on $J_0 = 0$ or $J_0 = 1$, respectively. Table 4.1 shows that when $J_0 = 0$ ($J_0 = 1$), the VaR and $TVaR$ increase (decreases) with time which corresponds to the 'ramping-up' ('emptying') of CIM operations from its empty (at capacity) starting point. We observe that as t increases, the VaR and $TVaR$ converge to the same values as the CIM

Table 4.2: VaR and $TVaR$ of RBNS payments for varying t when $J_0 = 1$

t	$VaR_{0.85}$	$TVaR_{0.85}$	$VaR_{0.9}$	$TVaR_{0.9}$	$VaR_{0.95}$	$TVaR_{0.95}$	$VaR_{0.995}$	$TVaR_{0.995}$
0.5	4.6664	6.7092	5.5109	7.5336	6.9402	8.9215	11.4728	13.3508
1	4.4958	6.5420	5.3408	7.3680	6.7730	8.7592	11.3166	13.1971
5	3.8031	5.8621	4.6507	6.6938	6.0916	8.0980	10.6810	12.5724
10	3.6515	5.7135	4.5002	6.5464	5.9425	7.9532	10.5419	12.4360
∞	3.6329	5.6953	4.4818	6.5283	5.9241	7.9354	10.5249	12.4192

reaches its steady-state whether $J_0 = 0$ or $J_0 = 1$. We note that the CIM is close to reaching steady-state by time 10 since the VaR and $TVaR$ values at $t = 10$ and $t = \infty$ are quite close.

In the following examples, we let $t = \infty$ for all calculations and consider the RBNS payments in the long-run, i.e., when the CIM has approached steady-state. Recall that risk measurements that involve the CIM in steady-state can provide an indicator of the insurer's overall risk due to RBNS claims.

4.6.2 Effect of $q_1(x)$

In this example, we assume the probability that a claim is subject to investigation is constant, i.e. $q_1(x) \equiv q_1$, and consider q_1's impact on the RBNS payments. All assumptions here follow those made in Example 2.4.1 except the investigation time is instead assumed to be exponentially distributed with rate $\alpha = 0.2$. The long-run RBNS payments has LT given by (4.12) and its VaR and $TVaR$ values are presented in Table 4.3. Note that if $q_1 = 0$, we have that $VaR = TVaR = 0$ since there is no investigation and all claims are immediately settled.

Table 4.3: VaR and $TVaR$ of RBNS payments for varying q_1

q_1	$VaR_{0.85}$	$TVaR_{0.85}$	$VaR_{0.9}$	$TVaR_{0.9}$	$VaR_{0.95}$	$TVaR_{0.95}$	$VaR_{0.995}$	$TVaR_{0.995}$
0.5	3.6329	5.6953	4.4818	6.5283	5.9241	7.9354	10.5249	12.4192
0.1	4.1598	6.2123	5.0059	7.0412	6.4429	8.4389	11.0085	12.8942
0.5	4.7016	6.7437	5.5459	7.5677	6.9746	8.9549	11.5050	13.3825
1	4.7803	6.8207	5.6243	7.6440	7.0515	9.0296	11.5769	13.4533

As q_1 increases, it becomes more likely that claims will enter the CIM (when it is not at capacity). This leads to more claims residing in the CIM in the long-run and all else being equal, greater RBNS payments for the insurer. Table 4.3 confirms our intuition since the VaR and $TVaR$ of the long-run RBNS payments increases as q_1 increases.

4.6.3 Three investigation strategies

In this example, we revisit the three investigation strategies discussed in Example 2.4.3 that are used to determine whether or not a claim will be subject to investigation. The three strategies involve allowing $q_1(x)$ to vary in x and more precisely, assume that

$$q_1(x) = \theta + (1-\theta)\left(1 - e^{-\kappa x}\right), \tag{4.54}$$

where

$$\begin{aligned} &\text{Strategy 1:} \quad \theta = 0.1 \qquad (\kappa = 0) \\ &\text{Strategy 2:} \quad \theta = 0.05 \quad\ (\kappa = 0.0173) \\ &\text{Strategy 3:} \quad \theta = 0 \qquad\ \ (\kappa = 0.0341) \end{aligned}$$

Recall that θ and κ are chosen such that the mean probability of subjecting a random claim to investigation is set to be 0.1, i.e., $\int_0^\infty q_1(x)p(x) = 0.1$. According to (4.54), on average, a percentage θ ($\theta \leq 0.1$) of all claims are automatically subjected to investigation and the other $(1-\theta)$ are subjected based on their claim amount with a probability $1 - e^{-\kappa x}$ for a claim of size x. All other assumptions follow those in Example 2.4.3 except the combination of exponential investigation time is replaced here by an exponential investigation time with the same mean, i.e. $\alpha = 0.2$. The long-run RBNS payments has LT given by (4.12) and its VaR and $TVaR$ for each of the three investigation strategies are presented in Table 4.4 at various confidence levels.

Table 4.4: VaR and $TVaR$ for varying investigation strategies

θ	$VaR_{0.85}$	$TVaR_{0.85}$	$VaR_{0.9}$	$TVaR_{0.9}$	$VaR_{0.95}$	$TVaR_{0.95}$	$VaR_{0.995}$	$TVaR_{0.995}$
0.1	3.5562	5.6203	5.0059	7.0412	6.4429	8.4389	11.0085	12.8942
0.05	5.1460	7.5216	6.1642	8.4705	7.8271	10.0327	12.8453	14.8623
0	5.9495	8.3582	6.9986	9.3155	8.6780	10.8786	13.6794	15.6819

Table 4.4 shows that the VaR and $TVaR$ increase with decreasing θ. From (4.54), as we decrease θ, we increase the degree to which the investigation strategy is based on the claim size. Thus, for smaller θ, claims that have been selected to enter the CIM will tend to be larger on average and therefore lead to larger RBNS payments in the long-run. This effect is confirmed by Table 4.4.

4.6.4 Investigation time dependence on claim size

In this example, we revisit the two scenarios discussed in the numerical example from Section 3.5.2 and compare the VaR and $TVaR$ of the RBNS payments under both scenarios. The first scenario involves a CIM with investigation times that are independent of all else in the model whereas the second scenario involves a CIM with investigation times that depend on claim sizes. The details of the second scenario's dependence structure and all other assumptions can be found in Section 3.5.2. Note that the mean investigation time for a random claim is set to be equal under both scenarios. The LT of the total long-run RBNS payments in the two-node CIM is given by (4.53) where

$$\begin{aligned}
\widetilde{a}_{(0,0,0)}(z) &= 1, \\
\widetilde{a}_{(1,1,0)}(z) &= \widetilde{k}_1(z), \\
\widetilde{a}_{(1,0,1)}(z) &= \widetilde{k}_2(z),
\end{aligned}$$

where $\widetilde{k}_1$, $\widetilde{k}_2$, and $\mathbf{Q}$ are formally defined in Section 3.5.2. The VaR and $TVaR$ for the long-run RBNS payments under both scenarios can be found in Table 4.5 for varying confidence levels.

Table 4.5: VaR and $TVaR$ with and without a dependence structure

Dependence	$VaR_{0.85}$	$TVaR_{0.85}$	$VaR_{0.9}$	$TVaR_{0.9}$	$VaR_{0.95}$	$TVaR_{0.95}$	$VaR_{0.995}$	$TVaR_{0.995}$
No	4.4404	6.4877	5.2857	7.3142	6.7187	8.7065	11.2658	13.1472
Yes	5.3224	7.5268	6.2637	8.4083	7.8040	9.8641	12.4967	14.4023

From Table 4.5, we observe that the VaR and $TVaR$ increase with the introduction of the dependence structure. This makes sense because when investigation times are dependent on claim sizes, smaller claims tend to exit faster from the CIM while larger claims tend to be delayed by longer investigation times. Thus, the in long-run, larger claims are more likely to be found in the CIM (which lead to larger long-run RBNS payments). The numerical results are therefore consistent with our intuition.

4.6.5 Claim investigation network

In this example, we revisit the two scenarios discussed in the numerical example of Section 3.5.3. The first scenario involves a one-node CIM composed of one node and one waiting place (i.e., $n = 1$ and $m = 1$) and the second scenario involves a two-node CIM composed of two nodes and no waiting places (i.e., $n = 2$ and $m = 0$). Readers are referred to the numerical example of Section 3.5.3 for the scenario assumptions. The LT of the total long-run RBNS payments in the two-node CIM is given by (4.53) where

$$\begin{aligned} \widetilde{a}_{(0,0,0)}(z) &= 1, & \widetilde{a}_{(1,1,0)}(z) &= \widetilde{k}(z) \\ \widetilde{a}_{(1,0,1)}(z) &= \widetilde{p}_1(z), & \widetilde{a}_{(2,1,1)}(z) &= \widetilde{k}(z)\widetilde{p}_1(z) \\ \widetilde{a}_{(2,2,0)}(z) &= \widetilde{k}(z)^2, & \widetilde{a}_{(2,0,2)}(z) &= \widetilde{p}_1(z)^2, \end{aligned}$$

and $\mathbf{Q}$ is as discussed in Section 3.5.3. The VaR and $TVaR$ for both the one-node and two-node CIMs are given in Table 4.6.

Table 4.6: VaR and $TVaR$ for a network-based CIM

(n, m)	$VaR_{0.85}$	$TVaR_{0.85}$	$VaR_{0.9}$	$TVaR_{0.9}$	$VaR_{0.95}$	$TVaR_{0.95}$	$VaR_{0.995}$	$TVaR_{0.995}$
(1,1)	7.8749	10.4362	8.9792	11.4572	10.7651	13.1355	16.1616	18.3263
(2,0)	8.2994	10.8470	9.3967	11.8628	11.1731	13.5336	16.5482	18.7070

Suppose the claims handling process is indeed composed of two stages as described in Section 3.5.3 (i.e., a claim is first investigated for fraudulence, then it moves onto processing if it is not fraudulent). Modelling this process using a one-node CIM would underestimate the RBNS payments since it would not reflect that claims that have completed the first stage of investigation and successfully moved onto administrative processing will be paid in full. This is confirmed by the Table 4.6 where we observe the VaR and $TVaR$ for the two-node CIM is higher than for the one-node CIM.

Chapter 5

Conclusion and Future Research

There has been little study of queues in insurer surplus and aggregate claim models. The work of this thesis may be viewed as a first steps in this line of inquiry by explicitly modelling claim investigation practices in classical models for insurance risk. In particular, a queue-based claim investigation mechanism (CIM) was proposed to model features like congestion in the claims handling process. The CIM was superimposed onto the classical insurer's surplus model as well as the aggregate claims model to measure the impact of the CIM on some common risk-based quantities of interest.

Chapter 2 proposed a simple CIM consisting of one server and no waiting places. It was assumed that the investigation times were distributed as a combination of n exponentials. Standard techniques of conditioning on the first claim were used to derive a defective renewal equation (DRE) for the Gerber-Shiu function $m(u)$ and probabilistic interpretations for the DRE components were provided. Via numerical examples, the impacts of various claim investigation strategies and CIMs on the ruin probability were measured.

Chapter 3 extended and generalized the surplus analysis conducted in Chapter 2 by proposing a more realistic queue-based claim investigation mechanism (CIM). The proposed

CIM consisted of a single queue with n investigation units and a finite capacity of m claims. More specifically, we considered CIMs which admit a (spectrally negative) Markov Additive Process (MAP) formulation of the insurer's surplus and relied on recent developments in the fluctuation theory of MAPs to perform our analysis. The flexibility of the MAP enabled generalizations to the CIM while preserving the surplus' MAP formulation. A detailed discussion of four possible CIM generalizations and their MAP formulations were provided. Numerical examples were used to assess the cost and benefit of additional investigators as well as the impact of variations in the CIM on the ruin probability.

In Chapter 4, the proposed CIM from Chapter 3 was superimposed onto the aggregate claims process to obtain the aggregate payment process. It was shown that this aggregate payment process has a Markovian Arrival Process (MAP) formulation that is preserved under considerable generalizations to the CIM. A distributional analysis of the future payments due to RBNS claims ("RBNS payments") was then performed under assumptions for $q_1(x)$ (the probability that a claim of size x is selected for investigation) and L_x (the distribution of the payment amount due to a claim of size x that had undergone investigation). The RBNS payment distribution was also found when the claim size density p was furthermore assumed to follow either a mixed Erlang, gamma, or inverse Gaussian density. Finally, numerical examples were used to assess the impact of various claim investigation practices on the VaR and $TVaR$ of the RBNS payments. Numerical findings were consistent with those from Chapters 2 and 3.

There are a number of directions in which this work may be extended which I briefly discuss below.

Chapters 2 and 3 of this thesis focused on classical quantities of interest that can be studied using the Gerber-Shiu function such as the time and deficit at ruin and the ruin probability. One way to extend this effort is to study other quantities of interest. For example, an insurer performing a cost-benefit analysis of their claims handling process could benefit from a quantity that measures the net benefit of a CIM. The costs of a CIM include investigator salaries and administration system expenses. The benefit may be defined, for example, as the difference - discovered via investigation - between an inflated claim amount and the amount actually payable to the policyholder.

Assuming costs are incurred continuously at rate κ, the following quantity

$$\Pi_t := \sum_{j=1}^{\lfloor \frac{M_t}{2} \rfloor} (X_{1,j} - Y_j) - \kappa t, \tag{5.1}$$

may be interpreted as the benefit of the CIM minus its cost, where $\lfloor M_t/2 \rfloor$ is the number of claims that have completed investigation by time t, and $X_{1,j}$ (Y_j) is the claim (payment) size of the jth claim completing investigation. Deriving a DRE for Π_t under Chapter 2's risk model by using the same approach found in the chapter appears to be a promising avenue of investigation. Further work would be needed to study this quantity in the risk model proposed in Chapters 3 and 4. Another cost associated with the CIM that can be measured is the excess payment due to inflated claims that were paid because the claim had arrived when the CIM was at full capacity and was therefore paid without investigation.

Work on optimality questions is another direction in which this work may be extended. It would be of interest to determine an optimal claim investigation mechanism that minimizes (or maximizes) some objective function such as the ruin probability or the CIM's net benefit (see above). For example, consider increasing the probability a claim (of size x) is investigated $q_1(x)$ or the number of investigators n. While this will increase the cost of investigation, it will also lead to the adjustment of a greater number of claims. Analytical studies can be performed to examine how $q_1(x)$ and/or n can be chosen to, for example, minimize the ruin probability. Numerical analysis can also be performed, e.g., some examples in Chapter 3 employed numerical analysis to determine the optimal number of investigators n that minimized the ruin probability.

Another way this work may be extended is to further improve realism by considering adaptive claim investigation mechanisms. For example, in bad economic times associated with large job losses, policyholders may be more inclined to file fraudulent or inflated claims. This results in a greater claim arrival rate coupled with an increased probability that a claim (of size x) is fraudulent $\varsigma(x)$. The insurer may respond by reinforcing their CIM, e.g., adding more investigators to their CIM and increasing the probability a claim (of size x) is investigated $q_1(x)$. The increased cost from the reinforced CIM would also need to be accounted for. Also, the insurer may wish to undo their reinforcements when economic times improve. The greater frequency of claims in bad economic times may be modelled using Markov modulated Poisson arrivals (see Section 3.5.4) though further work would be needed to model the other adaptive features. Also, insights from industry can be used to set plausible changes in model assumptions, for example, the change in $\varsigma(x)$ when economic times deteriorate. In addition, optimality questions may be explored relating to how the CIM should best be adapted in response to changes in the environment to minimize (or maximize) some objective functions.

The CIM could also be assumed to adapt in response to its level of congestion. For example, in times when the mechanism is approaching full capacity, the insurer may decide to investigate only the largest claims. For instance, suppose that when the mechanism is less than 80% full, the insurer investigates all claims over size l but when the mechanism is greater than 80% full, the insurer will focus its investigation on claims over size h ($h > l$). Here again, this direction of research would be of interest to insurers in the design of optimal CIMs which is an important component of their comprehensive set of risk management activities.

References

J. Abate and P. P. Valkó. Multi-precision Laplace transform inversion. *International Journal for Numerical Methods in Engineering*, 60(5):979–993, 2004.

S. Ahn and A. L. Badescu. On the analysis of the Gerber-Shiu discounted penalty function for risk processes with Markovian arrivals. *Insurance: Mathematics and Economics*, 41 (2):234–249, 2007.

S. Ahn, A. L. Badescu, E. C. Cheung, and J. R. Kim. An IBNR–RBNS insurance risk model with marked Poisson arrivals. *Insurance: Mathematics and Economics*, in press.

H. Albrecher and O. J. Boxma. On the discounted penalty function in a Markov-dependent risk model. *Insurance: Mathematics and Economics*, 37(3):650–672, 2005.

H. Albrecher and C. Hipp. Lundberg's risk process with tax. *Blätter der DGVFM*, 28(1): 13–28, 2007.

H. Albrecher, C. Constantinescu, and J. Garrido. Editorial for the special issue on Gerber-Shiu functions. *Insurance: Mathematics and Economics*, 46(1):1–2, 2010.

A. V. Asimit and A. L. Badescu. Extremes on the discounted aggregate claims in a time dependent risk model. *Scandinavian Actuarial Journal*, 2010(2):93–104, 2010.

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

S. Asmussen and H. Albrecher. *Ruin Probabilities.* Singapore: World Scientific, 2010.

A. Badescu, L. Breuer, A. Da Silva Soares, G. Latouche, M. A. Remiche, and D. Stanford. Risk processes analyzed as fluid queues. *Scandinavian Actuarial Journal*, 2005(2):127–141, 2005.

A. L. Badescu, X. S. Lin, and D. Tang. A marked Cox model for the number of IBNR claims: Theory. *Insurance: Mathematics and Economics*, 69:29–37, 2016.

J. Bertoin. *Lévy Processes.* Cambridge University Press, 1998.

K. A. Borovkov and D. C. Dickson. On the ruin time distribution for a Sparre Andersen process with exponential claim sizes. *Insurance: Mathematics and Economics*, 42(3): 1104–1108, 2008.

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

R. L. Brown and L. R. Gottlieb. *Introduction to Ratemaking and Loss Reserving for Property and Casualty Insurance.* Actex Publications, 2007.

H. Buhlmann. Mathematical methods in risk theory. *Berlin: Springer–Verlag*, 1970.

J. Cai and D. C. M. Dickson. On the expected discounted penalty function at ruin of a surplus process with interest. *Insurance: Mathematics and Economics*, 30(3):389–404, 2002.

E. C. Cheung and R. Feng. A unified analysis of claim costs up to ruin in a Markovian arrival risk model. *Insurance: Mathematics and Economics*, 53(1):98–109, 2013.

E. C. Cheung and D. Landriault. Analysis of a generalized penalty function in a semi-Markovian risk model. *North American Actuarial Journal*, 13(4):497–513, 2009a.

E. C. Cheung and D. Landriault. Perturbed MAP risk models with dividend barrier strategies. *Journal of Applied Probability*, 46(2):521–541, 2009b.

E. C. Cheung and D. Landriault. A generalized penalty function with the maximum surplus prior to ruin in a MAP risk model. *Insurance: Mathematics and Economics*, 46(1):127–134, 2010.

E. C. Cheung, D. Landriault, G. E. Willmot, and J. K. Woo. Structural properties of Gerber-Shiu functions in dependent Sparre Andersen models. *Insurance: Mathematics and Economics*, 46(1):117–126, 2010.

J. W. Cohen. *The Single Server Queue*. North-Holland, Amsterdam, 2 edition, 1982.

H. Cramér. Historical review of Filip Lundberg's works on risk theory. *Scandinavian Actuarial Journal*, 1969(sup3):6–12, 1969.

D. C. Dickson and C. Hipp. On the time to ruin for Erlang(2) risk processes. *Insurance: Mathematics and Economics*, 29(3):333–344, 2001.

D. C. M. Dickson and G. E. Willmot. The density of the time to ruin in the classical Poisson risk model. *Astin Bulletin*, 35(1):45–60, 2005.

D. Dufresne. Fitting combinations of exponentials to probability distributions. *Applied Stochastic Models in Business and Industry*, 23(1):23–48, 2007.

F. Fang and C. W. Oosterlee. A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848, 2008.

W. Feller. *An Introduction to Probability Theory and its Applications*, volume 2. John Wiley, New York, 2 edition, 1971.

D. P. Gaver Jr. Observing stochastic processes, and approximate transform inversion. *Operations Research*, 14(3):444–459, 1966.

H. U. Gerber and E. S. Shiu. Option pricing by Esscher transforms. *Transactions of the Society of Actuaries*, 46:99–191, 1994.

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

H. U. Gerber and E. S. Shiu. The time value of ruin in a Sparre Andersen model. *North American Actuarial Journal*, 9(2):49–69, 2005.

D. Gross, J. Shortle, J. Thompson, and C. Harris. *Fundamentals of Queueing Theory*. Wiley Series in Probability and Statistics. Wiley, 2011.

C. A. Hachemeister. A stochastic model for loss reserving. In *Proceedings. International Congress of Actuaries*, pages 185–194, 1980.

Q. M. He. *Fundamentals of Matrix–Analytic Methods*. Springer, 2014.

I. B. Hossack, J. H. Pollard, and B. Zehnwirth. *Introductory Statistics with Applications in General Insurance*. Cambridge University Press, 1999.

J. Ivanovs and Z. Palmowski. Occupation densities in solving exit problems for Markov additive processes and their reflections. *Stochastic Processes and their Applications*, 122 (9):3342–3360, 2012.

J. W. Jang. Martingale approach for moments of discounted aggregate claims. *Journal of Risk and Insurance*, 71(2):201–211, 2004.

A. Juri. Supermodular order and Lundberg exponents. *Scandinavian Actuarial Journal*, (1):17–36, 2002.

S. Karlin and H. M. Taylor. *A Second Course in Stochastic Processes*, volume 2. Gulf Professional Publishing, 1981.

B. Kim and H. S. Kim. Moments of claims in a Markovian environment. *Insurance: Mathematics and Economics*, 40(3):485–497, 2007.

L. Kleinrock. *Queueing Systems: Theory*, volume 1. Wiley Interscience, 1975.

V. Klimenok. On the modification of Rouche's theorem for the queueing theory problems. *Queueing Systems*, 38(4):431–434, 2001.

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

S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss models: Further Topics*. John Wiley & Sons, 2013.

A. Kuznetsov, A. E. Kyprianou, and V. Rivero. The theory of scale functions for spectrally negative Lévy processes. In *Lévy Matters*. Springer, 2012.

A. E. Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, 2006.

A. E. Kyprianou and Z. Palmowski. Fluctuations of spectrally negative Markov additive processes. In *Séminaire de probabilités XLI*. Springer, 2008.

D. Landriault and G. E. Willmot. On the joint distributions of the time to ruin, the surplus prior to ruin, and the deficit at ruin in the classical risk model. *North American Actuarial Journal*, 13(2):252–270, 2009.

D. Landriault, T. Shi, and G. E. Willmot. Joint densities involving the time to ruin in the Sparre Andersen risk model under exponential assumptions. *Insurance: Mathematics and Economics*, 49(3):371–379, 2011.

D. Landriault, W. Y. Lee, G. E. Willmot, and J. K. Woo. A note on deficit analysis in dependency models involving Coxian claim amounts. *Scandinavian Actuarial Journal*, 2014(5):405–423, 2014a.

D. Landriault, G. E. Willmot, and D. Xu. On the analysis of time dependent claims in a class of birth process claim count models. *Insurance: Mathematics and Economics*, 58: 168–173, 2014b.

D. Landriault, B. Li, and S. Li. Drawdown analysis for the renewal insurance risk process. *Scandinavian Actuarial Journal*, 2017(3):267–285, 2017.

G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*, volume 5. SIAM and ASA, Philadelphia, 1999.

G. Léveillé and J. Garrido. Moments of compound renewal sums with discounted claims. *Insurance: Mathematics and Economics*, 28(2):217–231, 2001.

J. Li, Q. Tang, and R. Wu. Subexponential tails of discounted aggregate claims in a time-dependent renewal risk model. *Advances in Applied Probability*, 42(4):1126–1146, 2010.

S. Li. *Adaptive policies and drawdown problems in insurance risk models*. PhD thesis, University of Waterloo, 2015.

S. Li and J. Garrido. On ruin for the Erlang(n) risk process. *Insurance: Mathematics and Economics*, 34(3):391–408, 2004.

S. Li and J. Garrido. The Gerber–Shiu function in a Sparre Andersen risk process perturbed by diffusion. *Scandinavian Actuarial Journal*, 2005(3):161–186, 2005.

S. Li and Y. Lu. Moments of the dividend payments and related problems in a Markov-modulated risk model. *North American Actuarial Journal*, 11(2):65–76, 2007.

S. X. Lin, G. E. Willmot, and S. Drekic. The classical risk model with a constant dividend barrier: analysis of the Gerber-Shiu discounted penalty function. *Insurance: Mathematics and Economics*, 33(3):551–566, 2003.

Y. Lu and C. C. L. Tsai. The expected discounted penalty at ruin for a Markov-modulated risk process perturbed by diffusion. *North American Actuarial Journal*, 11(2):136–149, 2007.

F. Lundberg. *Approximations of the probability function/Reinsurance of Collective Risks.* PhD thesis, University of Uppsala, 1903.

N. Maestas, K. J. Mullen, and A. Strand. Disability insurance and the Great Recession. *The American Economic Review*, 105(5):177–182, 2015.

M. F. Neuts. A versatile Markovian point process. *Journal of Applied Probability*, 16(4): 764–779, 1979.

R. Norberg. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin: The Journal of the IAA*, 23(1):95–115, 1993.

R. Norberg. Prediction of outstanding liabilities ii. model variations and extensions. *ASTIN Bulletin: The Journal of the IAA*, 29(1):5–25, 1999.

J. Ren. On the Laplace transform of the aggregate discounted claims with Markovian arrivals. *North American Actuarial Journal*, 12(2):198–206, 2008.

J. Ren. Analysis of insurance claim settlement process with Markovian arrival processes. *Risks*, 4(1):6, 2016.

S. I. Resnick. *Adventures in Stochastic Processes.* Birkhauser, Boston, 1992.

T. Rolski, H. Schmidli, V. Schmidt, and J. Teugels. *Stochastic Processes for Insurance and Finance.* John Wiley & Sons, 1999.

S. Ross. *Introduction to Probability Models.* Academic Press, 2007.

B. Sericola. *Markov Chains: Theory and Applications.* John Wiley & Sons, 2013.

T. Shi. *On the distribution of the time to ruin and related topics*. PhD thesis, University of Waterloo, 2013.

K. Sigman. Queueing theory. *Encyclopedia of Actuarial Science*, 3:1357–1361, 2006.

H. Stehfest. Algorithm 368: Numerical inversion of Laplace transforms. *Communications of the ACM*, 13(1):47–49, 1970.

B. Sundt and R. Vernic. *Recursions for convolutions and compound distributions with insurance applications*. Springer Science & Business Media, 2009.

G. C. Taylor. Probability of ruin under inflationary conditions or under experience rating. *Astin Bulletin*, 10(02):149–162, 1979.

E. C. Titchmarsh. *The Theory of Functions*. Oxford University Press, USA, 1939.

J. Trufin, H. Albrecher, and M. Denuit. Ruin problems under IBNR dynamics. *Applied Stochastic Models in Business and Industry*, 27(6):619–632, 2011.

M. Usábel. Calculating multivariate ruin probabilities via Gaver–Stehfest inversion technique. *Insurance: Mathematics and Economics*, 25(2):133–142, 1999.

P. P. Valkó and J. Abate. Comparison of sequence accelerators for the Gaver method of numerical Laplace transform inversion. *Computers & Mathematics with Applications*, 48 (3):629–636, 2004.

G. Wells, D. Newton, and D. Sanders. Non-life insurance claims in a recession. *Milliman White Paper*, 2009. URL http://uk.milliman.com/insight/Research/perspective/research/pdfs/Non-life-insurance-claims-in-a-recession/.

G. E. Willmot. The total claims distribution under inflationary conditions. *Scandinavian Actuarial Journal*, 1989(1):1–12, 1989.

G. E. Willmot. A queueing theoretic approach to the analysis of the claims payment process. *Transactions of the Society of Actuaries*, 42:447–497, 1990.

G. E. Willmot and X. S. Lin. Risk modelling with the mixed Erlang distribution. *Applied Stochastic Models in Business and Industry*, 27(1):2–16, 2011.

G. E. Willmot and J. K. Woo. On the class of Erlang mixtures with risk theoretic applications. *North American Actuarial Journal*, 11(2):99–115, 2007.

G. E. Willmot and J. K. Woo. *Surplus analysis of Sparre Andersen insurance risk processes*. Springer International Publishing, 2018.

G. E. Willmot, J. Cai, and X. S. Lin. Lundberg inequalities for renewal equations. *Advances in Applied Probability*, 33(3):674–689, 2001.

D. Xu. *Analysis of Time Dependent Aggregate Claims*. PhD thesis, University of Waterloo, 2016.

Z. Zhang. Approximating the density of the time to ruin via Fourier-cosine series expansion. *ASTIN Bulletin: The Journal of the IAA*, 47(1):169–198, 2017.