Compound Sums, Their Distributions, and Actuarial Pricing

Ang Li, *The University of Western Ontario*

Supervisor: Ren, Jiandong, *The University of Western Ontario*
Joint Supervisor: Zitikis, Ricardas, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

Abstract

Compound risk models are widely used in insurance companies to mathematically describe their aggregate amount of losses during a certain time period. However, the evaluation of the distribution of compound random variables and the computation of the relevant risk measures are non-trivial. Therefore, the main purpose of this thesis is to study the bounds and simulation methods for both univariate and multivariate compound distributions. The premium setting principles related to dependent multivariate compound distributions are studied. .

In the first part of this thesis, we consider the upper and lower bounds of the tail of bivariate compound distributions. Our results extend those in the literature (e.g. Willmot and Lin (1994) and Willmot et al. (2001)) for univariate compound distributions. First, we derive the exponential upper bounds when the claim size distribution is light-tailed with a finite moment generating function. Second, we present generalized upper and lower bounds when the claim size distribution is heavy-tailed without a finite moment generating function. Numerical examples are provided to illustrate the tightness of these bounds.

In the second part of the thesis, we develop several novel variance reduction techniques for simulating tail probability and mean excess loss of the univariate and bivariate compound models. These techniques stem from possible combinations of existing commonly used variance reduction techniques. Their performances are evaluated in detail.

In the third part of the thesis, we investigate the premium setting principles when the claim frequencies and claim severities in multiple collective risk models are correlated via a background risk. We develop a novel methodology of premium setting and numerically illustrate how model parameters influence the premium level. Two empirical methods and a parametric fitting method are provided for pricing and corresponding performance assessments are presented.

Keywords: compound risk model, tail probability, upper bound, simulation, variance reduction method, mean excess loss, moment transform, background risk.

Summary for Lay Audience

Compound risk models are widely used in insurance companies to mathematically describe their aggregate amount of losses during a certain time period. The tail probability and tail moments of the aggregate losses are important risk measures of the insurer's operation. Therefore, their accurate evaluation is essential in premium setting and risk management of insurance companies. However, the calculation of the tail probabilities and the tail moments of the compound random variables are non-trivial because compound variables usually do not have an explicit probability distribution function - even for the one-dimensional case.

Literature on evaluation of the tail probability and tail moments for univariate compound risk models is extensive. However, results for multivariate compound risk models with dependence are much less. Therefore, the main purpose of this thesis is to develop a methodology to study the tail probability and tail moments of multivariate compound risk models.

In particular, in the first part of the thesis, we derive theoretical upper bounds for the tail probability of bivariate compound distributions. The results provide an analytical way to characterize tail behavior of the insurance companies' aggregate losses in two lines of businesses. Our result generalizes those for univariate models in the literature.

In the second part of the thesis, we develop several novel techniques to efficiently simulate tail probability and mean excess loss of the univariate and bivariate compound models. The methodologies are developed particularly for compound variables. We show that the our proposed simulation method is much more efficient than the simple crude simulation methods; they are essentially based on combining existing variance-reduction methods.

In the third part of the thesis, we study a multivariate compound risk model where claim frequencies and claim severities are correlated via a background risk. We introduce a new premium setting methodology and provide both non-parametric and parametric methods for parameter estimation and apply them in premium setting.

Acknowledgements

My sincere thanks go to:

- My supervisors, Dr. Jiandong Ren and Dr. Ricardas Zitikis.
- My family members and friends, for their support over the years.
- All faculty members and staffs in the Department of Statistical and Actuarial Sciences.

Finally, I would like to thank the examination committee members for their valuable remarks and considerable recommendations.

To myself.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

In the collective risk model, the aggregate loss of an insurance company during a given time period, is modeled by a compound random variable

$$S = \sum_{i=1}^{M} X_i \tag{1.1}$$

where M denotes the claim number in the given time period, and $X_1, \ldots, X_M$ denote the claim severities. The X_i's are usually assumed to be independent and identically distributed (i.i.d.) and they are independent of M. The classical collective risk model and some of its variants are introduced in popular actuarial textbooks such as Klugman et al. (2012) in great details.

An evaluation of the tail probability and the tail moments is not easy, even when the distribution of M and X_i are known. One usually has to resort to recursive formulas, such as these proposed in Panjer (1981) for the case when the distribution of M belongs to the $(a, b, 0)$ class. There is extensive literature on further developments related to the Panjer's recursive formula. For details, one is referred to the comprehensive book by Sundt and Vernic (2009).

Transform based techniques, such as fast Fourier transform (FFT), are also widely used in calculating the distribution of aggregated claims. For an introduction, one can refer to Robertson (1992) or Wang (1998). Embrechts and Frei (2009) provided an excellent comparison of the recursive and FFT methods.

Simulation methods are flexible and can be handy in estimation of the tail probability/moments of compound distributions. However, they are subject to sampling errors.

Instead of deriving the exact expressions for the distribution of compound losses, there is extensive research about upper and lower bounds for the distributions. In the most general case, exponential Chebyshev's inequality provides a rough upper bound $\mathbf{P}[S > x] \leq e^{-tx}\mathbf{E}[e^{tS}]$, as long as $\mathbf{E}[e^{tS}] < \infty$ for some $t > 0$. Willmot and Lin (1994) showed that the tail probability of S has an exponential upper bound when the X_i's are light-tailed and the failure rate of M is always less than 1. Later, Willmot et al. (2001) extended their previous results to S with heavy-tailed severities and obtained general upper and lower bound by applying a class of distributions, which we will introduce in detail in the body of the thesis.

Recently, research began to pay more attention to multivariate risk models because insurers typically face multiple sources of claims. Sundt (1999) studied the joint distribution of the aggregate losses:

$$(S_1, S_2) = \sum_{i=1}^{M} (X_{1,i}, X_{2,i}), \tag{1.2}$$

where each claim generates a two-dimensional random vector representing two types of dependent losses.

Another type of model for two lines of insurance businesses is

$$(S_1, S_2) = \Big(\sum_{i=1}^{M} X_i, \sum_{j=1}^{N} Y_j \Big). \tag{1.3}$$

In order to model the interdependence between risks S_1 and S_2 from different sources, some assume that counting variables (M, N) are dependent but all the claim sizes X_i's and Y_j's are mutually independent non-negative variables and are independent of (M, N). For example, this kind of model is commonly-used in modeling auto insurance, where M and N could represent the numbers of claims from bodily injury and property damages, respectively. Concerning about such kind of bivariate aggregate claims random variables, Hesselager (1996) introduced and derived recursive formulas for the joint distributions of (S_1, S_2) when the X_i's and Y_j's are integer-valued variables.

For a general introduction to the aggregation of dependent risk portfolios, one is referred to Wang (1998), which discussed several commonly used correlated models such as multivariate Poisson mixture model and common shock model. Specifically, the bivariate Poisson mixture model assumes frequencies $M \sim \text{Pois}(\lambda_1 \Theta)$ and $N \sim \text{Pois}(\lambda_2 \Theta)$ with common mixing parameter Θ, whereas the common shock model constructs frequencies as $M = M_0 + Z$ and $N = N_0 + Z$ with independent M_0, N_0. The latter model was explored in detail in Meyers (2007). Instead of using recursions, Jin and Ren (2010) extended the tilting method associated with FFT method (Grübel and Hermesmeier (1999)) to three specific aggregate claim models introduced in Hesselager (1996).

There are many variants of models for dependent losses. We summarize some in the following. Boudreault et al. (2006) considered a dependence structure among the interclaim time and subsequent claim size, which is an extension to the classic compound Poisson risk model. In addition, Cossette et al. (2008) constructed a dependence structure between the claim amounts and interclaim time via a generalized Farlie-Gumbel-Morgenstern copula. Kousky and Cooke (2009) assume dependence among claim severities from natural disasters, where normal copula and Gumbel copula are used. Martel-Escobar et al. (2012) modeled the dependence between parameters of primary distribution and secondary distribution by a Farlie-Gumbel-Morgenstern family. Czado et al. (2012) considered a Gaussian copula to allow the dependence between the average claim size and the number of claims. Such copula-based regression models were extended to more general copula families such as Clayton, Gumbel and Frank by Krämer et al. (2013). Shi et al. (2015) explored methods that allow for the correlation among frequency and severity components under a hurdle modeling framework. Sarabia et al. (2016) worked on aggregated risks with claim amounts following multivariate dependent Pareto distributions in the individual risk model. They also explored several relevant collective risk models with Poisson, negative binomial and logarithmic distribution as primary distribution. Cossette et al. (2018)

constructed a dependence structure among claim severities by Archimedean copula in the individual risk model and investigated such a model with regard to aggregation, capital allocation and ruin problems.

In Chapter 2, we study the upper and lower bounds for the tail probability of the bivariate compound distributions defined in (1.3). Our results generalize those in Willmot and Lin (1994) and Willmot et al. (2001). In particular, exponential upper bounds are derived when the claim size distribution is light-tailed with a finite moment generating function. Generalized upper and lower bounds are presented when the claim size distribution is heavy-tailed without a finite moment generating function.

In Chapter 3, we develop several novel variance reduction techniques for simulating tail probability and mean excess loss of the classical univariate collective risk model S. These techniques are then extended to simulating tail probability and mean excess loss of the bivariate compound model (S_1, S_2). We compare their performance with existing variance reduction techniques which are commonly used. The results show that our new simulation methods are highly effective.

In Chapter 4, we study premium pricing for multiple business lines when claim frequencies and claim severities are dependent through a background risk. We illustrate how different parameters influence the premiums. Based on whether the background risk levels can be directly observed, two non-parametric methods are given to estimate the premiums empirically. We also study a parametric model and provide a parameter fitting methodology.

Chapter 2

Bounds for the tail of bivariate compound distributions

2.1 Introduction and literature review

In classical risk theory, the aggregate claims of an insurance company during a time period are modeled by the compound random variable

$$S = \sum_{i=1}^{M} X_i, \tag{2.1}$$

where M is the claim number variable with probability function $p_m = \mathbf{P}[M = m]$. $X_1, X_2, \ldots$ are non-negative independent identically distributed claim size random variables independent of M with a common distribution function (d.f.) $F(x) = \mathbf{P}[X_i \leq x]$, $x \geq 0$. Let

$$P(z) = \sum_{m=0}^{\infty} p_m z^m, \quad |z| < z_0, \tag{2.2}$$

be the probability generating function (p.g.f) of M, where z_0 represents the radius of convergence of $P(z)$. Further, we define the survival function of M by

$$a_m = \sum_{i=m+1}^{\infty} p_i, \quad m = 0, 1, 2, \ldots, \tag{2.3}$$

which is central to our discussions and the corresponding generating function (e.g. Feller (1968))

$$A(z) = \sum_{m=0}^{\infty} a_m z^m = \frac{1 - P(z)}{1 - z}, \quad |z| < z_0. \tag{2.4}$$

An important question in actuarial science concerns the tail probability of S, defined by

$$\psi(x) = \mathbf{P}[S > x] = \sum_{m=1}^{\infty} p_m \overline{F}^{*m}(x), \quad x \geq 0, \tag{2.5}$$

where $\overline{F}^{*m}(x) = 1 - F^{*m}(x)$ with $F^{*m}(x)$ being the d.f. of $X_1 + X_2 + \cdots + X_m$.

Several methods have been developed to determine the tail probability $\psi(x)$. From Feller (1971), p. 447, one has for $t > 0$

$$\tilde{\psi}(t) = \int_0^\infty e^{-tx}\psi(x)dx = \frac{1 - \mathbf{E}[e^{-tS}]}{t} = \frac{1 - P[\tilde{f}(t)]}{t}, \tag{2.6}$$

where

$$\tilde{f}(t) = \mathbf{E}[e^{-tX_i}] = \int_0^\infty e^{-tx}\mathrm{d}F(x), \tag{2.7}$$

is the Laplace transform of the claim size distribution. Bryan (1999) provided an analytical inversion formula for the Laplace transform, but the paper also mentioned that such a means is not very practical to implement. Wang (1998) mentioned an efficient method to approximate $\psi(x)$ numerically by fast Fourier transform. On the other hand, the tail probability can be calculated using the Panjer's recursive method (e.g. Dhaene et al. (1999)). For a book-long analysis of the Panjer's recursive method, one is referred to Sundt and Vernic (2009).

However, the methods mentioned above may not provide simple analytic forms for $\psi(x)$. Consequently, simple bounds and approximations are useful. In the most general case, a related inequality sometimes known as the exponential Chebyshev's inequality states that

$$\mathbf{P}[S > x] = \mathbf{P}[e^{tS} > e^{tx}] \le e^{-tx}\mathbf{E}[e^{tS}], \quad t > 0, \tag{2.8}$$

as long as $\mathbf{E}[e^{tS}] < \infty$, which is a simple extension to classical Chebyshev's inequality. In the special case when $p_m = (1-q)q^m$, $\psi(x)$ may be interpreted as a ruin probability in a ruin process setting, thus the classical Lundberg inequality states that

$$\psi(x) \le e^{-\kappa x}, \quad x \ge 0, \tag{2.9}$$

where $\kappa > 0$ satisfies $\tilde{f}(-\kappa) = q^{-1}$, seeing Gerber (1979) for detail. To apply this inequality, one has to assume that the claim size distribution has a finite moment generating function.

Other Lundberg type bound for the tail of compound distribution had also been studied in the literature. Willmot and Lin (1994) showed that

Proposition 2.1.1. *If there exists a $\phi \in (0,1)$, such that*

$$a_{m+1} \le \phi a_m, \quad m = 0, 1, 2, \ldots, \tag{2.10}$$

then

$$\psi(x) \le \frac{a_0}{\phi}e^{-\kappa x}, \quad x \ge 0, \tag{2.11}$$

where $\kappa > 0$ satisfies

$$\phi^{-1} = \int_0^\infty e^{\kappa t}\mathrm{d}F(t). \tag{2.12}$$

Later, this classical Lundberg inequality was proved by Gerber (1994) using martingale theory. In addition, Willmot and Lin (1994) also discussed different choices of ϕ depending on the failure rate characterizations and classes of counting distributions. Specifically, let

$$h_m = \frac{p_m}{a_m} \tag{2.13}$$

be the discrete failure rate, then one has

$$\frac{a_{m+1}}{a_m} = \frac{1}{1 + h_{m+1}}. \tag{2.14}$$

A further generalization of (2.11) was proved in Willmot (1994) by applying the reliability classification of distributions, introduced in Barlow and Proschan (1975). To introduce the idea, we need the following definition:

Definition Assuming that $B(x)$ is the d.f. of a non-negative random variable and letting $\overline{B}(x) = 1 - B(x)$, then $B(x)$ is new worse than used (NWU) if

$$\overline{B}(x)\overline{B}(y) \leq \overline{B}(x + y) \tag{2.15}$$

for $x \geq 0$ and $y \geq 0$. On the contrary, $B(x)$ belongs to new better than used (NBU) distributions if

$$\overline{B}(x)\overline{B}(y) \geq \overline{B}(x + y). \tag{2.16}$$

Among the NWU class, the subclass consisting of absolutely continuous distributions with decreasing failure rate (DFR) are widely known and used. In contrast, continuous distributions with increasing failure rate (IFR) are important cases of NBU class. For example, an exponential distribution is both DFR and IFR, and a Gamma(α, β) distribution is DFR (IFR) if $\alpha \leq (\geq) 1$.

Lin (1996) derived upper and lower bounds for various compound distributions in terms of (NWU) and (NBU) distributions, respectively. By using properties of the claim size distribution, Willmot and Lin (1997) derived simpler bounds than those of Lin (1996) and gave applications in various situations. After referencing Cai and Wu (1997) which proved Lin's result inductively, Willmot et al. (2001) reconstructed their previous results and obtained a computable upper bound for aggregate claims with heavy tail severities:

Proposition 2.1.2. *If $\phi \in (0, 1)$ satisfies (2.10), and $B(x)$ is a NWU d.f. satisfying*

$$\int_0^\infty \{\overline{B}(x)\}^{-1} \mathrm{d}F(x) = \frac{1}{\phi}, \tag{2.17}$$

then

$$\psi(x) \leq \frac{a_0}{\phi}\overline{B}(x), \quad x \geq 0. \tag{2.18}$$

The results above was applied to compound geometric distributions, which play an important role in insurance risk theory. See Willmot et al. (2001) for detail.

In this chapter, we extend the above results and derive bounds for bivariate aggregate claim random variables

$$(S_1, S_2) = (\sum_{i=1}^{M} X_i, \sum_{j=1}^{N} Y_j), \tag{2.19}$$

where claim sizes X_i's, Y_j's are mutually independent and are independent of claim frequencies (M, N). The tail probability of the bivariate compound distributions is defined by

$$\overline{H}(x, y) \triangleq \mathbf{P}[S_1 > x, S_2 > y] = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} p_{m,n}\overline{F}^{*m}(x)\overline{G}^{*n}(y), \quad x, y \geq 0, \tag{2.20}$$

where $p_{m,n} = \mathbf{P}[M = m, N = n]$ for $m, n = 0, 1, 2, \ldots$, with F and G being d.f. of X_i's and Y_j's respectively. Similar to the quantities a_m in the univariate case, we define

$$a_{m,n} = \sum_{i=m+1}^{\infty} \sum_{j=n+1}^{\infty} p_{i,j}, \quad m, n = 0, 1, 2, \ldots. \tag{2.21}$$

The remainder of this chapter is organized as follows. In Section 2.2, two lemmas are given on bivariate tail probability $\overline{H}(x, y)$, which are important in the derivation of the subsequent theorems. In Section 2.3, we derive the exponential upper bounds for bivariate compound distributions when the claim size distribution is light-tailed with a finite moment generating function. We evaluate the tightness of our exponential bounds analytically by comparing them to the Lundberg bounds in Willmot and Lin (1994) for univariate case. A bivariate geometric aggregated example is given to verify our analytic results numerically. Furthermore, we relax the constraints to make the upper bound eligible for more commonly used bivariate claim frequency distributions. Section 2.4 presents upper and lower bounds which further generalize those in Section 2.3 for a wider range of claim size distributions even without a finite moment generating function. Several corollaries are given for the purpose of computation and application. We also provide an example to illustrate how our general upper bounds can be used when claim size moments are known. Section 2.5 concludes this chapter.

2.2 Two lemmas

Before we derive our main results, we introduce two Lemmas.

Define

$$\begin{aligned}\overline{H}_{k,l}(x, y) = \sum_{m=0}^{k} \sum_{n=0}^{l} a_{m,n} \{ &\overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) + \overline{F}^{*m}(x)\overline{G}^{*n}(y) \\ &- \overline{F}^{*m}(x)\overline{G}^{*(n+1)}(y) - \overline{F}^{*(m+1)}(x)\overline{G}^{*n}(y)\}.\end{aligned} \tag{2.22}$$

Then we have the following lemma:

Lemma 2.2.1.

$$\lim_{k,l\to\infty} \overline{H}_{k,l}(x, y) = \overline{H}(x, y).$$

Proof. Since

$$a_{m,n} = a_{m+1,n+1} + \sum_{i=m+2}^{\infty} p_{i,n+1} + \sum_{j=n+2}^{\infty} p_{m+1,j} + p_{m+1,n+1}, \tag{2.23}$$

$$\sum_{i=m+2}^{\infty} p_{i,n+1} = a_{m+1,n} - a_{m+1,n+1} \tag{2.24}$$

and

$$\sum_{j=n+2}^{\infty} p_{m+1,j} = a_{m,n+1} - a_{m+1,n+1}, \tag{2.25}$$

we can directly obtain that

$$a_{m,n} = a_{m+1,n} + a_{m,n+1} + p_{m+1,n+1} - a_{m+1,n+1}. \tag{2.26}$$

Then we rearrange $\overline{H}_{k,l}(x.y)$ as

$$\begin{aligned}
\overline{H}_{k,l}(x,y) &= \sum_{m=0}^{k}\sum_{n=0}^{l} p_{m+1,n+1}\overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) \\
&\quad + \sum_{m=0}^{k}\sum_{n=0}^{l} a_{m,n}\overline{F}^{*m}(x)\overline{G}^{*n}(y) - \sum_{m=0}^{k}\sum_{n=0}^{l} a_{m+1,n+1}\overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) \\
&\quad + \sum_{m=0}^{k}\sum_{n=0}^{l} a_{m+1,n}\overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) - \sum_{m=0}^{k}\sum_{n=0}^{l} a_{m,n}\overline{F}^{*m}(x)\overline{G}^{*(n+1)}(y) \\
&\quad + \sum_{m=0}^{k}\sum_{n=0}^{l} a_{m,n+1}\overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) - \sum_{m=0}^{k}\sum_{n=0}^{l} a_{m,n}\overline{F}^{*(m+1)}(x)\overline{G}^{*n}(y) \\
&= \sum_{m=0}^{k}\sum_{n=0}^{l} p_{m+1,n+1}\overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) \\
&\quad + \sum_{m=1}^{k}\sum_{n=1}^{l} a_{m,n}\overline{F}^{*m}(x)\overline{G}^{*n}(y) - \sum_{m=1}^{k+1}\sum_{n=1}^{l+1} a_{m,n}\overline{F}^{*m}(x)\overline{G}^{*n}(y) \\
&\quad + \sum_{m=1}^{k+1}\sum_{n=0}^{l} a_{m,n}\overline{F}^{*m}(x)\overline{G}^{*(n+1)}(y) - \sum_{m=1}^{k}\sum_{n=0}^{l} a_{m,n}\overline{F}^{*m}(x)\overline{G}^{*(n+1)}(y) \\
&\quad + \sum_{m=0}^{k}\sum_{n=1}^{l+1} a_{m,n}\overline{F}^{*(m+1)}(x)\overline{G}^{*n}(y) - \sum_{m=0}^{k}\sum_{n=1}^{l} a_{m,n}\overline{F}^{*(m+1)}(x)\overline{G}^{*n}(y) \\
&= \sum_{m=0}^{k}\sum_{n=0}^{l} p_{m+1,n+1}\overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) + a_{k+1,l+1}\overline{F}^{*(k+1)}(x)\overline{G}^{*(l+1)}(y) \\
&\quad - \sum_{m=1}^{k+1} a_{m,l+1}\overline{F}^{*m}(x)\overline{G}^{*(l+1)}(y) - \sum_{n=1}^{l+1} a_{k+1,n}\overline{F}^{*(k+1)}(x)\overline{G}^{*n}(y) \\
&\quad + \sum_{m=0}^{k} a_{m,l+1}\overline{F}^{*(m+1)}(x)\overline{G}^{*(l+1)}(y) + \sum_{n=0}^{l} a_{k+1,n}\overline{F}^{*(k+1)}(x)\overline{G}^{*(n+1)}(y) \\
&= \sum_{m=0}^{k}\sum_{n=0}^{l} p_{m+1,n+1}\overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) + a_{k+1,l+1}\overline{F}^{*(k+1)}(x)\overline{G}^{*(l+1)}(y) \\
&\quad + \sum_{m=0}^{k}\sum_{j=l+2}^{\infty} p_{m+1,j}\overline{F}^{*(m+1)}(x)\overline{G}^{*(l+1)}(y) + \sum_{n=0}^{l}\sum_{i=k+2}^{\infty} p_{i,n+1}\overline{F}^{*(k+1)}(x)\overline{G}^{*(n+1)}(y).
\end{aligned} \tag{2.27}$$

The last three terms tend to 0 when $k, l \to \infty$. This ends our proof.

Remark 2.2.1. *The four terms in equation (2.27) represent, respectively:*

- *the probability of* $(\sum_{i=1}^{M} X_i > x, \sum_{j=1}^{N} Y_j > y)$, $1 \le N \le k+1$ *and* $1 \le M \le l+1$;
- *the probability of* $(\sum_{i=1}^{k+1} X_i > x, \sum_{j=1}^{l+1} Y_j > y)$, $N > k+1$ *and* $M > l+1$;
- *the probability of* $(\sum_{i=1}^{M} X_i > x, \sum_{j=1}^{l+1} Y_j > y)$, $1 \le N \le k+1$ *and* $M > l+1$;
- *the probability of* $(\sum_{i=1}^{k+1} X_i > x, \sum_{j=1}^{N} Y_j > y)$, $N > k+1$ *and* $1 \le M \le l+1$.

Hence, we could interpret $\overline{H}_{k,l}(x,y)$ *as*

$$\overline{H}_{k,l}(x,y) = \mathbf{P}\Big[\sum_{i=1}^{\min\{M,k+1\}} X_i > x, \sum_{j=1}^{\min\{N,l+1\}} Y_j > y\Big]. \tag{2.28}$$

Letting $k, l \to \infty$ *leads to Lemma 2.2.1.*

Lemma 2.2.2. *If there exists* $\phi_1 \in (0,1)$ *such that*

$$a_{m+1,n} \le \phi_1 a_{m,n}; \quad m,n = 0,1,2,\ldots, \tag{2.29}$$

then for any $k, l \ge 0$,

$$\overline{H}_{k+1,l}(x,y) \le \overline{H}_{0,l}(x,y) + \phi_1 \int_0^x \overline{H}_{k,l}(x-t,y)\mathrm{d}F(t). \tag{2.30}$$

Proof. From the law of total probability

$$\overline{F}^{*(m+1)}(x) = \overline{F}(x) + \int_0^x \overline{F}^{*m}(x-t)\mathrm{d}F(t), \tag{2.31}$$

we can obtain that

$$\begin{aligned}
&\int_0^x \overline{H}_{k,l}(x-t,y)\mathrm{d}F(t) \\
&= \sum_{m=0}^{k}\sum_{n=0}^{l} a_{m,n}\{ (\overline{F}^{*(m+2)}(x) - \overline{F}(x))\overline{G}^{*(n+1)}(y) + (\overline{F}^{*(m+1)}(x) - \overline{F}(x))\overline{G}^{*n}(y) \\
&\qquad - (\overline{F}^{*(m+1)}(x) - \overline{F}(x))\overline{G}^{*(n+1)}(y) - (\overline{F}^{*(m+2)}(x) - \overline{F}(x))\overline{G}^{*n}(y)\} \\
&= \sum_{m=0}^{k}\sum_{n=0}^{l} a_{m,n}\{ \overline{F}^{*(m+2)}(x)\overline{G}^{*(n+1)}(y) + \overline{F}^{*(m+1)}(x)\overline{G}^{*n}(y) \\
&\qquad - \overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) - \overline{F}^{*(m+2)}(x)\overline{G}^{*n}(y)\}.
\end{aligned} \tag{2.32}$$

Then we have

$$\begin{aligned}
\overline{H}_{k+1,l}(x,y) &= \sum_{m=0}^{k+1}\sum_{n=0}^{l} a_{m,n}\{\ \overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) + \overline{F}^{*m}(x)\overline{G}^{*n}(y) \\
&\qquad - \overline{F}^{*m}(x)\overline{G}^{*(n+1)}(y) - \overline{F}^{*(m+1)}(x)\overline{G}^{*n}(y)\} \\
&= \sum_{n=0}^{l} a_{0,n}\overline{F}(x)(\ \overline{G}^{*(n+1)}(y) - \overline{G}^{*n}(y)) \\
&+ \sum_{m=1}^{k+1}\sum_{n=0}^{l} a_{m,n}\{\ \overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) + \overline{F}^{*m}(x)\overline{G}^{*n}(y) \\
&\qquad - \overline{F}^{*m}(x)\overline{G}^{*(n+1)}(y) - \overline{F}^{*(m+1)}(x)\overline{G}^{*n}(y)\} \\
&= \sum_{n=0}^{l} a_{0,n}\overline{F}(x)(\ \overline{G}^{*(n+1)}(y) - \overline{G}^{*n}(y)) \\
&+ \sum_{m=1}^{k+1}\sum_{n=0}^{l} a_{m,n}\{\ \overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) + \overline{F}^{*m}(x)\overline{G}^{*n}(y) \\
&\qquad - \overline{F}^{*m}(x)\overline{G}^{*(n+1)}(y) - \overline{F}^{*(m+1)}(x)\overline{G}^{*n}(y)\} \\
&= \sum_{n=0}^{l} a_{0,n}\overline{F}(x)(\ \overline{G}^{*(n+1)}(y) - \overline{G}^{*n}(y)) \\
&\quad + \sum_{m=0}^{k}\sum_{n=0}^{l} \phi_1 a_{m,n}\{\overline{F}^{*(m+2)}(x)\overline{G}^{*(n+1)}(y) + \overline{F}^{*(m+1)}(x)\overline{G}^{*n}(y) \\
&\qquad - \overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) - \overline{F}^{*(m+2)}(x)\overline{G}^{*n}(y)\} \\
&= \sum_{n=0}^{l} a_{0,n}\overline{F}(x)(\ \overline{G}^{*(n+1)}(y) - \overline{G}^{*n}(y)) + \phi_1\int_0^x \overline{H}_{k,l}(x-t,y)\mathrm{d}F(t). \qquad (2.33)
\end{aligned}$$

By noting that

$$\overline{H}_{0,l}(x,y) = \sum_{j=0}^{l} a_{0,j}\overline{F}(x)(\overline{G}^{*(j+1)}(y) - \overline{G}^{*j}(y)), \qquad (2.34)$$

we have the desired result.

Symmetrically, we can directly write the following corollary:

Corollary 2.2.3. *If there exists $\phi_2 \in (0,1)$ such that*

$$a_{m,n+1} \le \phi_2 a_{m,n}; \quad m,n = 0,1,2,\ldots, \qquad (2.35)$$

then

$$\overline{H}_{k,l+1}(x,y) \le \overline{H}_{k,0}(x,y) + \phi_2\int_0^y \overline{H}(x,y-s)\mathrm{d}G(s). \qquad (2.36)$$

2.3 Exponential upper bounds for the tail of bivariate compound distributions

2.3.1 Exponential upper bound I

Theorem 2.3.1. *Suppose there exists* $\phi_1 \in (0,1)$ *such that*

$$a_{m+1,n} \leq \phi_1 a_{m,n}; \quad m,n = 0,1,2,\ldots, \tag{2.37}$$

then

$$\overline{H}(x,y) \leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x}, \quad x,y \geq 0 \tag{2.38}$$

where $\kappa_1 > 0$ *satisfies*

$$\phi_1^{-1} = \int_0^\infty e^{\kappa_1 x} \mathrm{d}F(x) = \tilde{f}(-\kappa_1). \tag{2.39}$$

Proof. We prove by induction that

$$\overline{H}_{k,l}(x,y) \leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x}, \quad x,y \geq 0. \tag{2.40}$$

Note that for $k = 0$,

$$\begin{aligned}
\overline{H}_{0,l}(x,y) &= \sum_{j=0}^{l} a_{0,j} \overline{F}(x)(\overline{G}^{*(j+1)}(y) - \overline{G}^{*j}(y)) \\
&= \overline{F}(x)\{\sum_{j=1}^{l}(a_{0,j-1} - a_{0,j})\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\} \\
&= \overline{F}(x)\{\sum_{i=1}^{\infty}\sum_{j=1}^{l} p_{i,j}\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\} \\
&\leq a_{0,0}\overline{F}(x)\overline{G}^{*(l+1)}(y) \\
&\leq a_{0,0}\overline{F}(x) \\
&= a_{0,0}\int_x^\infty \mathrm{d}F(t) \\
&\leq a_{0,0}\int_x^\infty e^{-\kappa_1(x-t)}\mathrm{d}F(t) \\
&\leq a_{0,0}e^{-\kappa_1 x}\int_0^\infty e^{\kappa_1 t}\mathrm{d}F(t) \\
&= \frac{a_{0,0}}{\phi_1}e^{-\kappa_1 x},
\end{aligned} \tag{2.41}$$

which shows that the result holds for $k = 0$; now if we suppose that inequality (2.40) holds for

$k = n$, then using Lemma 2.2.2

$$\begin{aligned}\overline{H}_{n+1,l}(x,y) &\le a_{0,0}\overline{F}(x) + \phi_1 \int_0^x \overline{H}_{n,l}(x-t,y)\mathrm{d}F(t) \\ &\le a_{0,0}\int_x^\infty e^{-\kappa_1(x-t)}\mathrm{d}F(t) + \phi_1\int_0^x \phi_1^{-1}a_{0,0}e^{-\kappa_1(x-t)}\mathrm{d}F(t) \\ &= a_{0,0}e^{-\kappa_1 x}\int_0^\infty e^{\kappa_1 t}\mathrm{d}F(t) \\ &= \frac{a_{0,0}}{\phi_1}e^{-\kappa_1 x}. \end{aligned} \tag{2.42}$$

Thus, $\overline{H}_{k,l}(x,y) \le \phi_1^{-1}a_{0,0}e^{-\kappa_1 x}$ holds for all $x, y \ge 0$ and $k, l = 0, 1, 2, \dots$. It follows directly that $\overline{H}(x,y) = \lim_{k,l\to\infty} \overline{H}_{k,l}(x,y) \le \phi_1^{-1}a_{0,0}e^{-\kappa_1 x}$.

The upper bound in (2.38) still holds under weaker constraints, which is a corollary of Proposition 2.1.1:

Corollary 2.3.2. *Suppose there exists* $\phi_1 \in (0,1)$ *such that*

$$a_{m+1,0} \le \phi_1 a_{m,0}; \quad m = 0, 1, 2, \dots, \tag{2.43}$$

then

$$\overline{H}(x,y) \le \frac{a_{0,0}}{\phi_1}e^{-\kappa_1 x}, \quad x, y \ge 0 \tag{2.44}$$

where $\kappa_1 > 0$ *satisfies*

$$\phi_1^{-1} = \int_0^\infty e^{\kappa_1 x}\mathrm{d}F(x) \tag{2.45}$$

Proof.

$$\begin{aligned}\overline{H}(x,y) &= \sum_{m=1}^\infty\sum_{n=1}^\infty p_{m,n}\overline{F}^{*m}(x)\overline{G}^{*n}(y) \\ &\le \sum_{m=1}^\infty\sum_{n=1}^\infty p_{m,n}\overline{F}^{*m}(x) \\ &= \sum_{m=1}^\infty p_m^*\overline{F}^{*m}(x), \end{aligned} \tag{2.46}$$

where

$$p_m^* = p_m - p_{m,0} = \sum_{n=1}^\infty p_{m,n} \quad m = 1, 2, \dots. \tag{2.47}$$

Then we consider a new distribution with probability mass function p_m^* for $m = 1, 2, \dots$ and $p_0^* = 1 - \sum_{m=1}^\infty p_m^*$. We further define a_m^* by

$$a_m^* = \sum_{i=m+1}^\infty p_m^* = \sum_{i=m+1}^\infty\sum_{j=1}^\infty p_{i,j} = a_{m,0}, \quad m = 0, 1, 2, \dots. \tag{2.48}$$

The constraint $a_{m+1,0} \leq \phi_1 a_{m,0}$ is equivalent to $a^*_{m+1} \leq \phi_1 a^*_m$ for $m = 0, 1, 2, \ldots$, therefore by Proposition 2.1.1 we have

$$\overline{H}(x,y) \leq \sum_{m=1}^{\infty} p^*_m \overline{F}^{*m}(x) \leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x}, \quad x, y \geq 0. \tag{2.49}$$

2.3.2 Exponential upper bound II

Under the same assumptions as in Theorem 2.3.1, we can further refine (2.38) and obtain that:

Theorem 2.3.3. *Under assumptions (2.37) and (2.39), an upper bound for $\overline{H}(x,y)$ is given by*

$$\overline{H}(x,y) \leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x} \overline{G}(y) + \frac{a_{0,1}}{\phi_1} e^{-\kappa_1 x} G(y), \quad x, y \geq 0. \tag{2.50}$$

Proof. Firstly, we prove by induction that

$$\overline{H}_{k,l}(x,y) \leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x} \overline{G}(y) + \frac{a_{0,1}}{\phi_1} e^{-\kappa_1 x} \int_0^y \overline{G}^{*l}(y-s) \mathrm{d}G(s). \tag{2.51}$$

When $k = 0$,

$$\begin{aligned}
\overline{H}_{0,l}(x,y) &= \sum_{j=0}^{l} a_{0,j} \overline{F}(x) (\overline{G}^{*(j+1)}(y) - \overline{G}^{*j}(y)) \\
&= a_{0,0} \overline{F}(x) \overline{G}(y) - a_{0,1} \overline{F}(x) \overline{G}(y) + \overline{F}(x) \{ \sum_{i=1}^{\infty} \sum_{j=2}^{l} p_{i,j} \overline{G}^{*j}(y) + a_{0,l} \overline{G}^{*(l+1)}(y) \} \\
&\leq a_{0,0} \overline{F}(x) \overline{G}(y) + a_{0,1} \overline{F}(x) \{ \overline{G}^{*(l+1)}(y) - \overline{G}(y) \} \\
&= a_{0,0} \overline{F}(x) \overline{G}(y) + a_{0,1} \overline{F}(x) \int_0^y \overline{G}^{*l}(y-s) \mathrm{d}G(s) \\
&\leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x} \overline{G}(y) + \frac{a_{0,1}}{\phi_1} e^{-\kappa_1 x} \int_0^y \overline{G}^{*l}(y-s) \mathrm{d}G(s).
\end{aligned} \tag{2.52}$$

Then we suppose that (2.51) holds for k. For $k+1$, we have

$$\begin{aligned}
\overline{H}_{k+1,l}(x,y) &\leq \sum_{j=0}^{l} a_{0,j} \overline{F}(x) (\overline{G}^{*(j+1)}(y) - \overline{G}^{*j}(y)) + \phi_1 \int_0^x \overline{H}_{k,l}(x-t, y) \mathrm{d}F(t) \\
&\leq a_{0,0} \overline{F}(x) \overline{G}(y) + a_{0,1} \overline{F}(x) \int_0^y \overline{G}^{*l}(y-s) \mathrm{d}G(s) + \phi_1 \int_0^x \overline{H}_{k,l}(x-t, y) \mathrm{d}F(t) \\
&\leq a_{0,0} \overline{G}(y) \int_x^{\infty} e^{-\kappa_1 (x-t)} \mathrm{d}F(t) + a_{0,1} \int_0^y \overline{G}^{*l}(y-s) \mathrm{d}G(s) \int_x^{\infty} e^{-\kappa_1 (x-t)} \mathrm{d}F(t) \\
&\quad + a_{0,0} \overline{G}(y) \int_0^x e^{-\kappa_1 (x-t)} \mathrm{d}F(t) + a_{0,1} \int_0^y \overline{G}^{*l}(y-s) \mathrm{d}G(s) \int_0^x e^{-\kappa_1 (x-t)} \mathrm{d}F(t) \\
&= a_{0,0} e^{-\kappa_1 x} \overline{G}(y) \int_0^{\infty} e^{\kappa_1 t} \mathrm{d}F(t) + a_{0,1} \int_0^y e^{-\kappa_1 x} \overline{G}^{*l}(y-s) \mathrm{d}G(s) \int_0^{\infty} e^{\kappa_1 t} \mathrm{d}F(t) \\
&= \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x} \overline{G}(y) + \frac{a_{0,1}}{\phi_1} e^{-\kappa_1 x} \int_0^y \overline{G}^{*l}(y-s) \mathrm{d}G(s).
\end{aligned} \tag{2.53}$$

Thus by induction, we obtain that

$$\overline{H}_{k,l}(x,y) \leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x}\overline{G}(y) + \frac{a_{0,1}}{\phi_1} e^{-\kappa_1 x} \int_0^y \overline{G}^{*l}(y-s)\mathrm{d}G(s). \tag{2.54}$$

It then follows directly that

$$\overline{H}(x,y) = \lim_{k,l\to\infty} \overline{H}_{k,l}(x,y) \leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x}\overline{G}(y) + \frac{a_{0,1}}{\phi_1} e^{-\kappa_1 x} G(y). \tag{2.55}$$

2.3.3 Exponential upper bound III

Our next result shows that when a reasonable condition is assumed for the distribution of N, the upper bound derived above can be tightened. We further assume that

$$b_n = \sum_{j=n+1}^{\infty} q_j, \quad n = 0, 1, 2, \ldots, \tag{2.56}$$

with $q_j = \Pr\{N = j\}$, whereas p_m, a_m keep the same assumptions as those in Section 2.1. If we further assume that $b_{n+1} \leq \phi_2 b_n$ in addition to the assumptions in Theorem 2.3.1, we can obtain the following result:

Theorem 2.3.4. *Suppose there exists $\phi_1, \phi_2 \in (0,1)$ such that*

$$a_{m+1,n} \leq \phi_1 a_{m,n}, \quad m, n = 0, 1, 2, \ldots, \tag{2.57}$$

and

$$b_{n+1} \leq \phi_2 b_n, \quad n = 0, 1, 2, \ldots, \tag{2.58}$$

then

$$\overline{H}(x,y) \leq \frac{b_0}{\phi_1\phi_2} e^{-\kappa_1 x - \kappa_2 y}, \quad x, y \geq 0 \tag{2.59}$$

where $\kappa_1, \kappa_2 > 0$ satisfy

$$\phi_1^{-1} = \int_0^\infty e^{\kappa_1 x}\mathrm{d}F(x) \tag{2.60}$$

and

$$\phi_2^{-1} = \int_0^\infty e^{\kappa_2 y}\mathrm{d}G(y). \tag{2.61}$$

Proof. Firstly, we prove by induction that

$$\overline{H}_{k,l}(x,y) \leq \phi_1^{-1} e^{-\kappa_1 x}\{\sum_{j=1}^{l} q_j \overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\}. \tag{2.62}$$

Note that for $k = 0$,

$$\begin{aligned}
\overline{H}_{0,l}(x,y) &= \overline{F}(x)\{\sum_{i=1}^{\infty}\sum_{j=1}^{l} p_{i,j}\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\} \\
&\le \overline{F}(x)\{\sum_{i=0}^{\infty}\sum_{j=1}^{l} p_{i,j}\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\} \\
&= \overline{F}(x)\{\sum_{j=1}^{l} q_j\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\} \\
&\le \phi_1^{-1}e^{-\kappa_1 x}\{\sum_{j=1}^{l} q_j\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\} \qquad (2.63)
\end{aligned}$$

Now suppose that (2.62) holds for k, then for $k+1$,

$$\begin{aligned}
\overline{H}_{k+1,l}(x,y) &\le \overline{H}_{0,l}(x,y) + \phi_1\int_0^x \overline{H}_{k,l}(x-t,y)dF(t) \\
&\le \Big(\int_x^{\infty} e^{-\kappa_1(x-t)}dF(t) + \phi_1\int_0^x \phi_1^{-1}e^{-\kappa_1(x-t)}dF(t)\Big)\{\sum_{j=1}^{l} q_j\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\} \\
&= \phi_1^{-1}e^{-\kappa_1 x}\{\sum_{j=1}^{l} q_j\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\}. \qquad (2.64)
\end{aligned}$$

Therefore,

$$\overline{H}(x,y) = \lim_{k,l\to\infty}\overline{H}_{k,l}(x,y) \le \phi_1^{-1}e^{-\kappa_1 x}\sum_{j=1}^{\infty} q_j\overline{G}^{*j}(y). \qquad (2.65)$$

Since we further assume that $b_{n+1} \le \phi_2 b_n$, the upper bounds (2.11) in univariate case can be applied to (2.65) directly, which finally results in

$$\overline{H}(x,y) \le \frac{b_0}{\phi_1\phi_2}e^{-\kappa_1 x-\kappa_2 y}. \qquad (2.66)$$

2.3.4 Evaluating exponential bounds I-III analytically

In the bivariate model (S_1, S_2), if $a_{m+1,n} \le \phi_1 a_{m,n}, m, n = 0, 1, \ldots$, we can obtain the exponential bound I for the tail probability $\mathbf{P}[S_1 > x, S_2 > y]$, that is,

$$\overline{H}(x,y) \le \frac{a_{0,0}}{\phi_1}e^{-\kappa_1 x}, \quad x, y \ge 0. \qquad (2.67)$$

Now we assume that $\psi(x) = \mathbf{P}[S_1 > x]$, the tail probability for marginal variable S_1, and that

$$a_m = \sum_{i=m+1}^{\infty} p_i = \sum_{i=m+1}^{\infty}\sum_{j=0}^{\infty} p_{i,j}, \quad m = 0, 1, 2, \ldots, \qquad (2.68)$$

which represents marginal $\mathbf{P}[M > m]$ for bivariate frequencies (M, N).

In this case, if $a_{m+1} \leq \phi_1 a_m$, $m = 0, 1, 2, \ldots$ is also satisfied, from Proposition 2.1.1 a marginal exponential bound can be obtained:

$$\psi(x) \leq \frac{a_0}{\phi_1} e^{-\kappa_1 x}, \quad x \geq 0. \tag{2.69}$$

It is of interest to consider the differences and connections between (2.67) and (2.69). Therefore, we give several clarifications:

- Obviously, $\overline{H}(x, y) \leq \overline{H}(x, 0) \leq \psi(x)$ from the aspect of definition;
- $a_{m+1,n} \leq \phi_1 a_{m,n}, m, n = 0, 1, \ldots$ is neither the sufficient condition nor the necessary condition for $a_{m+1} \leq \phi_1 a_m, m = 0, 1, \ldots$;
- $a_{m+1,n} \leq \phi_1 a_{m,n}, m, n = 0, 1, \ldots$ is a sufficient condition for $a_{m+1} \leq \phi_1 a_m, m = 0, 1, \ldots$ when N is zero-truncated;
- $a_{m+1,n} \leq \phi_1 a_{m,n}$ and $a_{m+1} \leq \phi_1 a_m$ can hold at the same time;
- $a_{0,0} \leq a_0$, hence $\frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x} \leq \frac{a_0}{\phi_1} e^{-\kappa_1 x}$.

Note 2.3.1. *Especially, in the case of independent frequencies (M, N), we have*

$$a_{m,n} = \sum_{i=m+1}^{\infty} \sum_{j=n+1}^{\infty} p_{i,j} = \sum_{i=m+1}^{\infty} \sum_{j=n+1}^{\infty} p_i q_j = a_m b_n. \tag{2.70}$$

The assumption $a_{m+1} \leq \phi_1 a_m$ is obviously equivalent to $a_{m+1,n} \leq \phi_1 a_{m,n}$ because we can multiply on both sides by b_n: $a_{m+1} b_n \leq \phi_1 a_m b_n$. Then we could connect the two upper bounds listed above through

$$\begin{aligned} \overline{H}(x, y) &= \mathbf{P}[S_1 > x, S_2 > y] \\ &\leq \mathbf{P}[S_1 > x, S_2 > 0] \\ &\leq \mathbf{P}[S_1 > x] \cdot \mathbf{P}[S_2 > 0] \\ &= \psi(x) \cdot \sum_{j=1}^{\infty} q_j \overline{G}^{*j}(0) \\ &\leq \frac{a_0}{\phi_1} e^{-\kappa_1 x} \cdot b_0 \\ &= \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x}. \end{aligned} \tag{2.71}$$

However, this is restricted by the independence assumption between claim frequencies M and N because without this assumption $\mathbf{P}[S_1 > x, S_2 > 0]$ might be greater than $\mathbf{P}[S_1 > x] \cdot \mathbf{P}[S_2 > 0]$. So our results provide an upper bound for more general conditions when M and N are dependent.

Generally in the case of dependent (M, N), we compare $a_{m+1} \leq \phi_1 a_m$ to $a_{m+1,n} \leq \phi_1 a_{m,n}$, where the constraints actually change from

$$\sum_{i=m+2}^{\infty} \sum_{j=0}^{\infty} p_{i,j} \leq \phi_1 \sum_{i=m+1}^{\infty} \sum_{j=0}^{\infty} p_{i,j}, \quad m = 0, 1, 2, \ldots, \tag{2.72}$$

to

$$\sum_{i=m+2}^{\infty} \sum_{j=n+1}^{\infty} p_{i,j} \leq \phi_1 \sum_{i=m+1}^{\infty} \sum_{j=n+1}^{\infty} p_{i,j}, \quad m, n = 0, 1, 2, \ldots. \tag{2.73}$$

The latter constraint is much stronger, but the improvement of (2.67) over (2.69) is limited. We only decrease the coefficient $a_0 = \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} p_{i,j}$ to $a_{0,0} = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{i,j}$. Obviously, the exponential bound I works well when $\mathbf{P}[M \neq 0, N = 0]$ is large.

Later, keeping the same assumptions, we refine our upper bound to be

$$\overline{H}(x, y) \leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x} \overline{G}(y) + \frac{a_{0,1}}{\phi_1} e^{-\kappa_1 x} G(y), \tag{2.74}$$

which does improve the exponential bound I by involving y. We can see easily that as y increases, the exponential bound II approaches $a_{0,1}\phi_1^{-1} e^{-\kappa x}$ away from $a_{0,0}\phi_1^{-1} e^{-\kappa x}$, narrowing the upper bounds. However, the improvement is still limited. This is due to

$$\frac{a_{0,1}}{\phi_1} e^{-\kappa_1 x} \leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x} \overline{G}(y) + \frac{a_{0,1}}{\phi_1} e^{-\kappa_1 x} G(y) \leq \frac{a_{0,0}}{\phi_1} e^{-\kappa_1 x}. \tag{2.75}$$

Even though y approaches positive infinity, the exponential bound II at best reaches $\overline{H}(x, y) \leq a_{0,1}\phi_1^{-1} e^{-\kappa x}$.

After adding another constraints $b_{n+1} \leq \phi_2 b_n$, $n = 0, 1, 2, \ldots$, the exponential III states that

$$\overline{H}(x, y) \leq \frac{b_0}{\phi_1 \phi_2} e^{-\kappa_1 x - \kappa_2 y}. \tag{2.76}$$

This is a better result as it decreases rapidly with respect to both x and y, which guarantees the tightness of the upper bound when x or y is large. A numerical example is given in the following section.

2.3.5 Numerical experiments

Here, we introduce the bivariate discrete Gumbel distribution, which is a discretized version of Gumbel's bivariate exponential distribution. Now suppose (M_0, N_0) follows a bivariate exponential distribution with survival function $S_{M_0,N_0}(x, y) = \mathbf{P}[M_0 > x, N_0 > y]$ defined as

$$S_{M_0,N_0}(x, y) = e^{-(\lambda_1 x + \lambda_2 y + \theta x y)}, \quad x, y, \lambda_1, \lambda_2 > 0, 0 \leq \theta \leq \lambda_1 \lambda_2. \tag{2.77}$$

The marginal distributions of M and N follow exponential distributions with rate λ_1 and λ_2, respectively.

In actuarial sciences, we deal with discrete claim frequencies. Thus we define (M, N), a group of discrete claim frequencies with similar survival function as the Gumbel's bivariate exponential distributions previously mentioned, that is,

$$S_{M,N}(m, n) = \mathbf{P}[M > m, N > n] = S_{M_0,N_0}(m + 1, n + 1), \quad m, n = 0, 1, 2, \ldots. \tag{2.78}$$

In another word, the joint probability mass function of (M, N) is defined as

$$p_{i,j} = \mathbf{P}[M = i, N = j] = \mathbf{P}[M_0 \in [i, i+1), N_0 \in [j, j+1)], \quad i, j = 0, 1, 2, \ldots. \tag{2.79}$$

The probability mass functions of marginal distributions M and N are

$$p_i = \mathbf{P}[M = i] = \mathbf{P}[M_0 \in [i, i+1)] = e^{-\lambda_1 i} - e^{-\lambda_1 (i+1)}, \quad i = 0, 1, 2, \ldots, \tag{2.80}$$

and

$$q_j = \mathbf{P}[N = i] = \mathbf{P}[N_0 \in [j, j+1)] = e^{-\lambda_2 j} - e^{-\lambda_2 (j+1)}, \quad j = 0, 1, 2, \ldots. \tag{2.81}$$

This discretized bivariate exponential distribution is a perfect example for our theory since we have

$$a_{m,n} = e^{-(\lambda_1(m+1)+\lambda_2(n+1)+\theta(m+1)(n+1))}, \quad m, n = 0, 1, 2, \ldots, \tag{2.82}$$

and

$$a_m = e^{-\lambda_1(m+1)} \quad m = 0, 1, 2, \ldots, \tag{2.83}$$

$$b_n = e^{-\lambda_2(n+1)} \quad n = 0, 1, 2, \ldots. \tag{2.84}$$

Then we can obtain

$$\frac{a_{m+1}}{a_m} = e^{-\lambda_1}, \quad \frac{b_{n+1}}{b_n} = e^{-\lambda_2}, \quad m, n = 0, 1, 2, \ldots \tag{2.85}$$

and

$$\frac{a_{m+1,n}}{a_{m,n}} = e^{-\lambda_1-\theta(n+1)}, \quad \frac{a_{m,n+1}}{a_{m,n}} = e^{-\lambda_2-\theta(m+1)}, \quad m, n = 0, 1, 2, \ldots. \tag{2.86}$$

Hence we are able to define $\phi_1 = e^{-\lambda_1}$ and $\phi_2 = e^{-\lambda_2}$ such that

$$a_{m+1} \leq \phi_1 a_m; \quad b_{n+1} \leq \phi_2 b_n; \quad a_{m+1,n} \leq \phi_1 a_{m,n}; \quad a_{m,n+1} \leq \phi_2 a_{m,n}, \quad m, n = 0, 1, 2, \ldots. \tag{2.87}$$

We assume that claim severities follow independent Gamma distributions $X_i \sim \text{Ga}(\alpha_1, \beta_1)$, $i = 1, 2, \ldots$, and $Y_j \sim \text{Ga}(\alpha_2, \beta_2)$, $j = 1, 2, \ldots$. Then κ_1 can be obtained from

$$\mathbf{E}[e^{\kappa_1 X_i}] = \left(1 - \frac{\kappa_1}{\beta_1}\right)^{-\alpha_1} = \frac{1}{\phi_1}, \tag{2.88}$$

$$\kappa_1 = \beta_1(1 - \phi_1^{\frac{1}{\alpha_1}}). \tag{2.89}$$

Similarly,

$$\kappa_2 = \beta_2(1 - \phi_2^{\frac{1}{\alpha_2}}). \tag{2.90}$$

In our numerical example, we set $\lambda_1 = 0.05$, $\lambda_2 = 0.08$, $\theta = 0.002$, $\alpha_1 = 20$, $\beta_1 = 0.5$, $\alpha_2 = 30$, $\beta_2 = 0.6$. The simulated survival probabilities and upper bounds are showed in Table 2.1.

At the beginning, assuming that we only know about one side $a_{m+1} \leq \phi_1 a_m$ and $a_{m+1,n} \leq \phi_1 a_{m,n}$, we obtain the first four rows in the the following table. Then, adding $b_{n+1} \leq \phi_2 b_n$, we can further obtain the fifth and sixth rows. Finally, we add $b_{m,n+1} \leq \phi_2 b_{m,n}$ and obtain the last row. The formulae in the second and fifth rows are the upper bounds for $\mathbf{P}[S_1 > x]$, $\mathbf{P}[S_2 > y]$ obtained from Proposition 2.1.1, respectively. The third row corresponds to exponential upper bound I. The fourth row corresponds to exponential upper bound II and the last two rows corresponds to exponential upper bound III. Obviously, our exponential upper bounds III outperform the Fréchet upper bound $\min\{\mathbf{P}[S_1 > x], \mathbf{P}[S_2 > y]\}$ for $\mathbf{P}[S_1 > x, S_2 > y]$.

Probability	$\overline{H}(500,100)$	$\overline{H}(1000,100)$	$\overline{H}(2000,100)$	$\overline{H}(500,500)$	$\overline{H}(1000,1000)$
Simulation(10000 times)	0.4053000	0.2086000	0.05240000	0.1773000	0.02320000
$\frac{a_0}{\phi_1}e^{-\kappa_1 x}$	0.5356794	0.2869524	0.08234170	0.5356794	0.28695244
$\frac{a_{0,0}}{\phi_1}e^{-\kappa_1 x}$	0.4935064	0.2643612	0.07585910	0.4935064	0.26436123
$\frac{a_{0,0}\overline{G}(y)}{\phi_1}e^{-\kappa_1 x}+\frac{a_{0,1}G(y)}{\phi_1}e^{-\kappa_1 x}$	0.4546539	0.2435487	0.06988690	0.4546536	0.24354859
$\frac{b_0}{\phi_2}e^{-\kappa_2 y}$	0.8523254	0.8523254	0.85232544	0.4498081	0.20232731
$\frac{b_0}{\phi_1\phi_2}e^{-\kappa_1 x-\kappa_2 y}$	0.4799822	0.2571166	0.07378023	0.2533068	0.06103503
$\frac{a_0}{\phi_1\phi_2}e^{-\kappa_1 x-\kappa_2 y}$	0.4945998	0.2649470	0.07602717	0.2610212	0.06289382

Table 2.1: Comparison of exponential upper bounds

2.3.6 Exponential upper bound IV: a generalization of the result

Theorem 2.3.5. *Suppose there exists numbers $\phi \in (0,1)$ such that*

$$a_{m+1,n+1} \le \phi a_{m,n}, \quad m,n = 0,1,2,\dots. \tag{2.91}$$

If there exists $(\phi_1,\phi_2) \in (0,1)^2$ such that

$$a_{m+1} \le \phi_1 a_m, \quad m = 0,1,\dots, \tag{2.92}$$

$$b_{n+1} \le \phi_2 b_n, \quad n = 0,1,\dots, \tag{2.93}$$

and

$$\phi_1 * \phi_2 = \phi, \tag{2.94}$$

we have

$$\overline{H}(x,y) \le \max\left\{\frac{a_0}{\phi_1\phi}, \frac{b_0}{\phi_2\phi}\right\} e^{-\kappa_1 x-\kappa_2 y}, \quad x,y \ge 0, \tag{2.95}$$

where $\kappa_1,\kappa_2 > 0$ satisfy

$$\phi_1^{-1} = \int_0^\infty e^{\kappa_1 x}\mathrm{d}F(x), \tag{2.96}$$

and

$$\phi_2^{-1} = \int_0^\infty e^{\kappa_2 y}\mathrm{d}G(y). \tag{2.97}$$

Proof. We prove by the result induction.

Firstly, from (2.63), we have

$$\begin{aligned}
\overline{H}_{0,l}(x,y) &\le \phi_1^{-1} e^{-\kappa_1 x}\Big\{\sum_{j=1}^{l} q_j \overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\Big\} \\
&\le \phi_1^{-1} e^{-\kappa_1 x}\Big\{\sum_{j=1}^{l} q_j \overline{G}^{*j}(y) + b_l\overline{G}^{*(l+1)}(y)\Big\} \\
&\le \phi_1^{-1} e^{-\kappa_1 x} \sum_{j=1}^{\infty} q_j \overline{G}^{*j}(y) \\
&\le \frac{b_0}{\phi_1\phi_2} e^{-\kappa_1 x - \kappa_2 y} \\
&\le \frac{b_0}{\phi_2\phi} e^{-\kappa_1 x - \kappa_2 y},
\end{aligned} \tag{2.98}$$

since $\phi \le \phi_1$. Similarly, we also have

$$\overline{H}_{k,0}(x,y) \le \frac{a_0}{\phi_1\phi} e^{-\kappa_1 x - \kappa_2 y}. \tag{2.99}$$

Now suppose that inequality holds for k, l, we have

$$\begin{aligned}
\overline{H}_{k+1,l+1}(x,y) &= \sum_{m=0}^{k+1}\sum_{n=0}^{l+1} a_{m,n}\{\overline{F}^{*(m+1)}(x)\overline{G}^{*(n+1)}(y) + \overline{F}^{*m}(x)\overline{G}^{*n}(y) \\
&\qquad - \overline{F}^{*m}(x)\overline{G}^{*(n+1)}(y) - \overline{F}^{*(m+1)}(x)\overline{G}^{*n}(y)\} \\
&= a_{0,0}\overline{F}(x)\overline{G}(y) + \phi\int_0^x\int_0^y \overline{H}_{k,l}(x-z,y-w)\mathrm{d}F(z)\mathrm{d}G(w) \\
&\quad + \sum_{n=1}^{l+1} a_{0,n}\overline{F}(x)\{\overline{G}^{*(n+1)}(y) - \overline{G}^{*(n)}(y)\} + \sum_{m=1}^{k+1} a_{m,0}\overline{G}(y)\{\overline{F}^{*(m+1)}(x) - \overline{F}^{*(m)}(x)\} \\
&= a_{0,0}\overline{F}(x)\overline{G}(y) + \phi\int_0^x\int_0^y \overline{H}_{k,l}(x-z,y-w)\mathrm{d}F(z)\mathrm{d}G(w) \\
&\quad + \overline{F}(x)\{\sum_{n=1}^{l}(a_{0,n}-a_{0,n+1})\overline{G}^{*(n+1)}(y) + a_{0,l+1}\overline{G}^{*(l+2)}(y) - a_{0,1}\overline{G}(y)\} \\
&\quad + \overline{G}(y)\{\sum_{m=1}^{k}(a_{m,0}-a_{m+1,0})\overline{F}^{*(m+1)}(x) + a_{k+1,0}\overline{F}^{*(k+2)}(x) - a_{1,0}\overline{F}(x)\} \\
&= a_{0,0}\overline{F}(x)\overline{G}(y) + \phi\int_0^x\int_0^y \overline{H}_{k,l}(x-z,y-w)\mathrm{d}F(z)\mathrm{d}G(w) \\
&\quad + \overline{F}(x)\{\sum_{n=1}^{l}\sum_{i=1}^{\infty} p_{i,n+1}\overline{G}^{*(n+1)}(y) + a_{0,l+1}\overline{G}^{*(l+2)}(y) - a_{0,1}\overline{G}(y)\} \\
&\quad + \overline{G}(y)\{\sum_{m=1}^{k}\sum_{j=1}^{\infty} p_{m+1,j}\overline{F}^{*(m+1)}(x) + a_{k+1,0}\overline{F}^{*(k+2)}(x) - a_{1,0}\overline{F}(x)\}
\end{aligned}$$

$$
\begin{aligned}
&= a_{0,0}\overline{F}(x)\overline{G}(y) + \phi\int_0^x\int_0^y \overline{H}_{k,l}(x-z, y-w)\mathrm{d}F(z)\mathrm{d}G(w)\\
&\quad + \overline{F}(x)\int_0^y\{\sum_{n=1}^{l}\sum_{i=1}^{\infty} p_{i,n+1}\overline{G}^{*n}(y-w) + a_{0,l+1}\overline{G}^{*(l+1)}(y-w)\}\mathrm{d}G(w)\\
&\quad + \overline{G}(y)\int_0^x\{\sum_{m=1}^{k}\sum_{j=1}^{\infty} p_{m+1,j}\overline{F}^{*m}(x-z) + a_{k+1,0}\overline{F}^{*(k+1)}(x-z)\}\mathrm{d}F(z)\\
&\le a_{0,0}\overline{F}(x)\overline{G}(y) + \phi\int_0^x\int_0^y \overline{H}_{k,l}(x-z, y-w)\mathrm{d}F(z)\mathrm{d}G(w)\\
&\quad + \overline{F}(x)\int_0^y\{\sum_{n=1}^{l}\sum_{i=1}^{\infty} p_{i,n+1}\overline{G}^{*(n+1)}(y-w) + a_{0,l+1}\overline{G}^{*(l+2)}(y-w)\}\mathrm{d}G(w)\\
&\quad + \overline{G}(y)\int_0^x\{\sum_{m=1}^{k}\sum_{j=1}^{\infty} p_{m+1,j}\overline{F}^{*(m+1)}(x-z) + a_{k+1,0}\overline{F}^{*(k+2)}(x-z)\}\mathrm{d}F(z)\\
&\le a_{0,0}\overline{F}(x)\overline{G}(y) + \phi\int_0^x\int_0^y \overline{H}_{k,l}(x-z, y-w)\mathrm{d}F(z)\mathrm{d}G(w)\\
&\quad + \overline{F}(x)\int_0^y\{\sum_{n=1}^{l} q_{n+1}\overline{G}^{*(n+1)}(y-w) + b_{l+1}\overline{G}^{*(l+2)}(y-w)\}\mathrm{d}G(w)\\
&\quad + \overline{G}(y)\int_0^x\{\sum_{m=1}^{k} p_{m+1}\overline{F}^{*(m+1)}(x-z) + a_{k+1}\overline{F}^{*(k+2)}(x-z)\}\mathrm{d}F(z)\\
&\le a_{0,0}\overline{F}(x)\overline{G}(y) + \phi\int_0^x\int_0^y \overline{H}_{k,l}(x-z, y-w)\mathrm{d}F(z)\mathrm{d}G(w)\\
&\quad + \overline{F}(x)\int_0^y\{\sum_{n=1}^{\infty} q_n\overline{G}^{*n}(y-w)\}\mathrm{d}G(w) + \overline{G}(y)\int_0^x\{\sum_{m=1}^{\infty} p_m\overline{F}^{*m}(x-z)\}\mathrm{d}F(z)\\
&\le a_{0,0}\int_x^\infty e^{-\kappa_1(x-z)}\mathrm{d}F(z)\int_y^\infty e^{-\kappa_2(y-w)}\mathrm{d}G(w)\\
&\quad + \max\Big\{\frac{a_0}{\phi_1}, \frac{b_0}{\phi_2}\Big\}\int_0^x e^{-\kappa_1(x-z)}\mathrm{d}F(z)\int_0^y e^{-\kappa_2(y-w)}\mathrm{d}G(w)\\
&\quad + \frac{b_0}{\phi_2}\int_x^\infty e^{-\kappa_1(x-z)}\mathrm{d}F(z)\int_0^y e^{-\kappa_2(y-w)}\mathrm{d}G(w)\\
&\quad + \frac{a_0}{\phi_1}\int_0^x e^{-\kappa_1(x-z)}\mathrm{d}F(z)\int_y^\infty e^{-\kappa_2(y-w)}\mathrm{d}G(w)\\
&\le \max\Big\{\frac{a_0}{\phi_1\phi}, \frac{b_0}{\phi_2\phi}\Big\}e^{-\kappa_1 x-\kappa_2 y},
\end{aligned}
\tag{2.100}
$$

since $a_{0,0} \le \min\{a_0, b_0\}$, where we prove by induction that

$$
\overline{H}_{k,l}(x,y) \le \max\Big\{\frac{a_0}{\phi_1\phi}, \frac{b_0}{\phi_2\phi}\Big\}e^{-\kappa_1 x-\kappa_2 y}. \tag{2.101}
$$

Letting $k, l \to \infty$ finished our proof.

Note 2.3.2. *If ϕ^*, ϕ_1^* and ϕ_2^* are the smallest values in $(0,1)$ such that*

$$a_{m+1,n+1} \leq \phi^* a_{m,n}, \quad m,n = 0,1,2,\dots, \tag{2.102}$$

$$a_{m+1} \leq \phi_1^* a_m, \quad m = 0,1,\dots, \tag{2.103}$$

$$b_{n+1} \leq \phi_2^* a_n, \quad n = 0,1,\dots, \tag{2.104}$$

are satisfied, we should choose $\phi \in (\phi^*, 1)$, $\phi_1 \in (\phi_1^*, 1)$ and $\phi_2 \in (\phi_2^*, 1)$. The following cases ensure:

- *If $\phi^* \leq \phi_1^* * \phi_2^*$, we may choose $\phi = \phi_1^* * \phi_2^*$, $\phi_1 = \phi_1^*$, $\phi_2 = \phi_2^*$;*
- *If $\phi_1^* * \phi_2^* \leq \phi^* < \max\{\phi_1^*, \phi_2^*\}$, we may choose $\phi = \phi^*$, $\phi_1 = \phi_1^*$; $\phi_2 = \phi^*/\phi_1^* \geq \phi_2^*$.*
- *If $\phi^* \geq \max\{\phi_1^*, \phi_2^*\}$, it is impossible to find valid (ϕ, ϕ_1, ϕ_2) for Theorem 2.3.5 .*

Note 2.3.3. *We have checked numerically that two commonly used bivariate negative binomial distributions satisfy the constraints in Theorem 2.3.5 :*

- *The bivariate Poisson distribution with mixing gamma parameter. That is, $M \sim Poisson(\lambda_1\Lambda)$, $N \sim Poisson(\lambda_2\Lambda)$ with $\Lambda \sim Ga(\alpha,\beta)$;*
- *The bivariate negative binomial distribution with common shock. That is, $M = N_0 + N_1$, $N = N_0 + N_2$, where $N_j \sim NB(r_j, p)$, $j = 0,1,2$, and are independent of each other.*

It has been proved that a_{m+1}/a_m of a negative binomial distribution is decreasing with respect to m. Numerically, for both bivariate distributions mentioned above, $a_{m+1,n+1}/a_{m,n}$ is also decreasing with respect to either m or n. In detail, this is caused by

- *$a_{m+1,n}/a_{m,n}$ decreasing with respect to m,*
- *$a_{m,n+1}/a_{m,n}$ decreasing with respect to n.*

2.4 General upper and lower bounds

In this section, we provide general upper and lower bounds for the tail of the bivariate compound distributions. These results generalize those given in Willmot et al. (2001) for the univariate case. We also introduce several classes of distributions mentioned in this chapter. The new worse (better) than used in expectation or NWUE (NBUE) is a superclass of NWU (NBU) class defined as follows:

Definition The d.f. $F(x)$ is NWUE (NBUE) if the mean residual lifetime $r(t)$ satisfies

$$r(x) \geq (\leq) r(0), \tag{2.105}$$

where

$$r(x) = \int_0^\infty \frac{\overline{F}(x+t)}{\overline{F}(x)} \mathrm{d}t. \tag{2.106}$$

That is, the mean residual lifetime is never larger than the expected lifetime.

By integration by part, we have

$$\int_x^\infty t \mathrm{d}F(x) = \overline{F}(x)\{x + r(x)\}, \tag{2.107}$$

which directly results in the following proposition.

Proposition 2.4.1. *If the d.f. $F(x)$ of variable X is NWUE, then for $x \geq 0$,*

$$\overline{F}(x) = \frac{\int_x^\infty t \mathrm{d}F(t)}{x + r(x)} \leq \frac{\mathbf{E}[X]}{x + r(x)} \leq \frac{\mathbf{E}[X]}{x + \mathbf{E}[X]}. \tag{2.108}$$

Another subclass of NWUE (NBUE) is the new worse (better) than used in convex ordering or NWUC (NBUC) class defined by

Definition The d.f. $F(x)$ is NWUC (NBUC) if

$$\overline{F}_e(t+x) \geq (\leq) \overline{F}_e(x)\overline{F}(t) \tag{2.109}$$

for all $x \geq 0, t \geq 0$, where

$$F_e(x) = 1 - \overline{F}_e(x) = \frac{\int_0^x \overline{F}(t)\mathrm{d}t}{\int_0^\infty \overline{F}(t)\mathrm{d}t}, \tag{2.110}$$

the equilibrium d.f. of $F(x)$.

2.4.1 General upper bound I

Theorem 2.4.2. *Suppose that $\phi_1 \in (0,1)$ satisfies the dominance condition that*

$$a_{m+1,n} \leq \phi_1 a_{m,n}; \quad m, n = 0, 1, 2, \ldots \tag{2.111}$$

and $B_1(t)$ is a d.f. satisfying the generalized adjustment equation

$$\int_0^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) = \frac{1}{\phi_1}. \tag{2.112}$$

If $V_1(x)$ is a d.f. satisfying

$$\overline{V}_1(x)\overline{B}_1(t) \le \overline{V}_1(x+t), \quad x \ge 0, t \ge 0, \tag{2.113}$$

then

$$\overline{H}(x,y) \le \frac{a_{0,0}}{\phi_1 \overline{V}_1(0)} c_1(x), \quad x, y \ge 0, \tag{2.114}$$

where $c_1(x)$ is defined by

$$\frac{1}{c_1(x)} = \inf_{0\le z\le x, \overline{F}(z)>0} c_1(x,z), \quad x \ge 0, \tag{2.115}$$

with $c_1(x,z)$ given by

$$c_1(x,z) = \frac{\int_z^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t)}{\overline{V}_1(x-z)\overline{F}(z)}, \quad 0 \le z \le x, \quad \overline{F}(z) > 0. \tag{2.116}$$

Proof. We prove by induction on k that for $0 \le z \le x$,

$$\overline{H}_{k,l}(z,y) \le \frac{a_{0,0}}{\phi_1 \overline{V}_1(x-z)} c_1(x), \quad x, y \ge 0. \tag{2.117}$$

Intuitively, $c_1(x)$ is a non-decreasing function which satisfies

$$\overline{F}(z) \le \frac{c_1(x)}{\overline{V}_1(x-z)} \int_z^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t), \quad 0 \le z \le x. \tag{2.118}$$

For any l,

$$\begin{aligned}
\overline{H}_{0,l}(z,y) &\le a_{0,0}\overline{F}(z) \\
&\le \frac{a_{0,0}c_1(x)}{\overline{V}_1(x-z)} \int_z^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) \\
&\le \frac{a_{0,0}c_1(x)}{\overline{V}_1(x-z)} \int_0^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) \\
&= \frac{a_{0,0}}{\phi_1 \overline{V}_1(x-z)} c_1(x),
\end{aligned} \tag{2.119}$$

which shows that the result holds for $k = 0$; now, we suppose the result holds for k, then

$$\begin{aligned}
\overline{H}_{k+1,l}(z,y) &\le a_{0,0}\overline{F}(z) + \phi_1 \int_0^z \overline{H}_{k,l}(z-t,y)\mathrm{d}F(t) \\
&\le \frac{a_{0,0}c_1(x)}{\overline{V}_1(x-z)} \int_z^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) + a_{0,0}c_1(x) \int_0^z \{\overline{V}_1(x+t-z)\}^{-1}\mathrm{d}F(t) \\
&\le \frac{a_{0,0}c_1(x)}{\overline{V}_1(x-z)} \int_0^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) \\
&= \frac{a_{0,0}}{\phi_1 \overline{V}_1(x-z)} c_1(x).
\end{aligned} \tag{2.120}$$

Now we can obtain that

$$\overline{H}(z,y) = \lim_{k,l\to\infty} \overline{H}_{k,l}(z,y) \le \frac{a_{0,0}}{\phi_1 \overline{V}_1(x-z)} c_1(x). \tag{2.121}$$

Letting $z = x$, we further obtain that

$$\overline{H}(x,y) \le \frac{a_{0,0}}{\phi_1 \overline{V}_1(0)} c_1(x). \tag{2.122}$$

Note that Theorem 2.4.2 still holds under weaker constraints:

Corollary 2.4.3. *Suppose $\phi_1 \in (0,1)$ satisfies a weaker condition that*

$$a_{m+1,0} \le \phi_1 a_{m,0}; \quad m = 0,1,2,\ldots \tag{2.123}$$

and $B_1(t)$ is a d.f. satisfying the generalized adjustment equation (2.112). If $V_1(x)$ is a d.f. satisfying (2.113), then there still holds that

$$\overline{H}(x,y) \le \frac{a_{0,0}}{\phi_1 \overline{V}_1(0)} c_1(x), \quad x,y \ge 0, \tag{2.124}$$

where $c_1(x)$ is defined by (2.115).

Proof. The proof is similar to that of Corollary 2.3.2, so for simplicity we continue from (2.46) and succeed the definitions for $p_m^*, a_m^*, m = 0,1,2,\ldots$.

In Willmot et al. (2001), it is stated in Theorem 4.2.1 that if

$$a_{m+1}^* \le \phi_1 a_m^*; \quad m = 0,1,2,\ldots, \tag{2.125}$$

we have

$$\sum_{m=1}^{\infty} p_m^* \overline{F}^{*m}(z) \le \frac{a_0^*}{\phi_1} \frac{c_1(x)}{\overline{V}_1(x-z)}; \quad 0 \le z \le x. \tag{2.126}$$

Since $a_{m+1,0} \le \phi_1 a_{m,0}$ is equivalent to $a_{m+1}^* \le \phi_1 a_m^*$ for $m = 0,1,2,\ldots$, it follows

$$\overline{H}(z,y) \le \sum_{m=1}^{\infty} p_m^* \overline{F}^{*m}(z) \le \frac{a_{0,0}}{\phi_1 \overline{V}_1(x-z)} c_1(x); \quad 0 \le z \le x. \tag{2.127}$$

Letting $z = x$ leads to (2.124).

2.4.2 Corollaries to Theorem 2.4.2

Corollary 2.4.4. *If $\phi_1 \in (0,1)$ satisfies (2.111), and $\kappa_1 > 0$ satisfies*

$$\int_0^\infty e^{\kappa_1 x} \mathrm{d}F(x) = \frac{1}{\phi_1}, \tag{2.128}$$

then

$$\overline{H}(x,y) \le \frac{a_{0,0}}{\phi_1} \alpha_1(x) e^{-\kappa_1 x}, \quad x,y \ge 0, \tag{2.129}$$

where

$$\frac{1}{\alpha_1(x)} = \inf_{0\le z\le x, \overline{F}(z)>0} \int_z^\infty e^{\kappa_1 t} \mathrm{d}F(t) \Big/ \{e^{\kappa_1 z}\overline{F}(z)\}, \quad x \ge 0. \tag{2.130}$$

Proof. Let $B_1(x) = 1 - e^{-\kappa_1 x}$, then (2.113) holds with $V_1(x) = B_1(y)$, and Theorem 2.4.2 applies. Then

$$c_1(x,z) = e^{\kappa_1 x}\frac{\int_z^\infty e^{\kappa_1 t}\mathrm{d}F(t)}{e^{\kappa_1 z}\overline{F}(z)}, \tag{2.131}$$

and

$$\frac{1}{c_1(x)} = \frac{e^{\kappa_1 x}}{\alpha_1(x)}, \tag{2.132}$$

i.e. $c_1(x) = \alpha_1(x)e^{-\kappa_1 x}$.

It is worth noting that

$$\int_z^\infty e^{\kappa_1 x}\mathrm{d}F(x) \geq \int_z^\infty e^{\kappa_1 z}\mathrm{d}F(x) = e^{\kappa_1 z}\overline{F}(z), \tag{2.133}$$

and therefore $1/\alpha_1(x) \geq 1$, i.e. $\alpha_1(x) \leq 1$. Thus, (2.114) yields

$$\overline{H}(x,y) \leq \frac{a_{0,0}}{\phi_1}e^{-\kappa_1 x}, \quad x, y \geq 0, \tag{2.134}$$

the exponential bounds we derived in the previous sections.

Corollary 2.4.5. *If $\phi_1 \in (0,1)$ satisfies (2.111), and $B_1(x)$ is a NWU d.f. satisfying* (2.112)*, then*

$$\overline{H}(x,y) \leq \frac{a_{0,0}}{\phi_1}c_1(x), \quad x, y \geq 0, \tag{2.135}$$

where $c_1(x)$ satisfies (2.115) *with*

$$c_1(x,z) = \frac{\int_0^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t)}{\overline{B}_1(x-z)\overline{F}(z)}, \quad x \geq 0, \quad 0 \leq z \leq x. \tag{2.136}$$

Proof. Since $B_1(x)$ is NWU, (2.113) holds with $V_1(x) = B_1(x)$, and Theorem 2.4.2 applies. This proof assumes $\overline{B}_1(0) = 1$.

The next result is important since it gives an upper bound that can be computed easily, and generalizes the exponential upper bounds.

Corollary 2.4.6. *Suppose that $\phi_1 \in (0,1)$ satisfies (2.111) and $B_1(x)$ is a d.f. satisfying (2.112). If $V_1(x)$ is a d.f. satisfying (2.113), then*

$$\overline{H}(x,y) \leq \frac{a_{0,0}}{\phi_1\overline{V}_1(0)}\overline{V}_1(x), \quad x, y \geq 0. \tag{2.137}$$

In particular, if $\phi_1 \in (0,1)$ satisfies (2.111) and $B_1(x)$ is a NWU d.f. satisfying (2.112), then

$$\overline{H}(x,y) \leq \frac{a_{0,0}}{\phi_1}\overline{B}_1(x), \quad x, y \geq 0. \tag{2.138}$$

Proof. Since $\overline{V}_1(x)$ is non-increasing, it follows from (2.113) that for $z \geq t$ that

$$\overline{V}_1(x-z)\overline{B}_1(t) \leq \overline{V}_1(x+t-z) \leq \overline{V}_1(x), \tag{2.139}$$

and so from (2.116),

$$c_1(x,z) \geq \frac{\int_z^\infty \{\overline{V}_1(x)\}^{-1} \mathrm{d}F(t)}{\overline{F}(z)} = \{\overline{V}_1(x)\}^{-1}. \tag{2.140}$$

Thus, from (2.115), $1/c_1(x) \geq 1/\overline{V}_1(x)$, i.e. $c_1(x) \leq \overline{V}_1(x)$. Then (2.137) follows from Theorem 2.4.2. Clearly, (2.138) is the special case where $V_1(x) = B_1(x)$ and follows from Corollary 2.4.5.

Corollary 2.4.7. *If $\phi_1 \in (0,1)$ satisfies (2.111), and $B_1(x)$ is a NWUC d.f. satisfying (2.112), then*

$$\overline{H}(x,y) \leq \frac{a_{0,0}}{\phi_1}\overline{B}_e(x), \quad x,y \geq 0, \tag{2.141}$$

with $B_e(x) = 1 - \overline{B}_e(x) = \int_0^x \overline{B}_1(t)\mathrm{d}t / \int_0^\infty \overline{B}_1(t)\mathrm{d}t$ the equilibrium d.f. of $B_1(x)$.

Proof. Since $B_1(x)$ is NWUC, (2.113) holds with $V_1(x) = B_e(x)$. The result follows from Corollary 2.4.6 and the fact that $\overline{B}_e(0) = 1$.

For the larger NWUE class we have the following.

Corollary 2.4.8. *If $\phi_1 \in (0,1)$ satisfies (2.111), and $B_1(x)$ is a NWUE d.f. satisfying (2.112), then*

$$\overline{H}(x,y) \leq \frac{a_{0,0}}{\phi_1}\frac{\int_0^\infty t\mathrm{d}B_1(t)}{x + \int_0^\infty t\mathrm{d}B_1(t)}, \quad x,y \geq 0. \tag{2.142}$$

Proof. It follows from Proposition 2.4.1 that $\overline{B}_1(x) \leq \overline{V}_1(x)$ where

$$\overline{V}_1(x) = \frac{\int_0^\infty t\mathrm{d}B_1(t)}{x + \int_0^\infty t\mathrm{d}B_1(t)}. \tag{2.143}$$

It is not difficult to see that $V_1(x)$ is a Pareto d.f. which is DFR and hence NWU. Thus, $\overline{B}_1(x) \leq \overline{V}_1(x) \leq \overline{V}_1(x+t)/\overline{V}_1(t)$ and (2.113) is satisfied with $\overline{V}_1(0) = 0$. The result follows from Corollary 2.4.6.

Remark 2.4.1. *The corollaries above can be applied to the following general upper bounds* II-IV*, since they actually specify choices of $c_1(x)/\overline{V}_1(0)$, $c_2(y)/\overline{V}_2(0)$ under different situations.*

2.4.3 General upper bound II

If we keep the same assumptions as in Theorem 2.4.2, we can further refine (2.114) similarly to what we have done for exponential upper bounds:

Theorem 2.4.9. *An upper bound for $\overline{H}(x,y)$ is given by*

$$\overline{H}(x,y) \leq \frac{a_{0,0}}{\phi_1\overline{V}_1(0)}c_1(x)\overline{G}(y) + \frac{a_{0,1}}{\phi_1\overline{V}_1(0)}c_1(x)G(y), \quad x,y \geq 0. \tag{2.144}$$

Proof. Firstly, we prove by induction on k for $0 \le z \le x$ that

$$\overline{H}_{k,l}(z,y) \le \frac{a_{0,0}c_1(x)}{\phi_1 \overline{V}_1(x-z)}\overline{G}(y) + \frac{a_{0,1}c_1(x)}{\phi_1 \overline{V}_1(x-z)} \int_0^y \overline{G}^{*l}(y-s)\mathrm{d}G(s); \quad 0 \le z \le x. \tag{2.145}$$

When $k = 0$,

$$\begin{aligned}
\overline{H}_{0,l}(z,y) &= \sum_{j=0}^{l} a_{0,j}\overline{F}(z)(\overline{G}^{*(j+1)}(y) - \overline{G}^{*j}(y)) \\
&\le a_{0,0}\overline{F}(z)\overline{G}(y) + a_{0,1}\overline{F}(z)\int_0^y \overline{G}^{*l}(y-s)\mathrm{d}G(s) \\
&\le \frac{a_{0,0}c_1(x)}{\phi_1 \overline{V}_1(x-z)}\overline{G}(y) + \frac{a_{0,1}c_1(x)}{\phi_1 \overline{V}_1(x-z)} \int_0^y \overline{G}^{*l}(y-s)\mathrm{d}G(s).
\end{aligned} \tag{2.146}$$

Then we suppose that (2.145) holds for $k = n$. For $k = n + 1$, we have

$$\begin{aligned}
\overline{H}_{n+1,l}(z,y) &\le a_{0,0}\overline{F}(z)\overline{G}(y) + a_{0,1}\overline{F}(z)\int_0^y \overline{G}^{*l}(y-s)\mathrm{d}G(s) + \phi\int_0^z \overline{H}_{n,l}(z-t,y)\mathrm{d}F(t) \\
&\le \left\{\frac{a_{0,0}c_1(x)}{\overline{V}_1(x-z)}\overline{G}(y) + \frac{a_{0,1}c_1(x)}{\overline{V}_1(x-z)}\int_0^y \overline{G}^{*l}(y-s)\mathrm{d}G(s)\right\}\int_z^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) \\
&\quad + \left\{a_{0,0}c_1(x)\overline{G}(y) + a_{0,1}c_1(x)\int_0^y \overline{G}^{*l}(y-s)\mathrm{d}G(s)\right\}\int_0^z \{\overline{V}_1(x+t-z)\}^{-1}\mathrm{d}F(t) \\
&\le \left\{\frac{a_{0,0}c_1(x)}{\overline{V}_1(x-z)}\overline{G}(y) + \frac{a_{0,1}c_1(x)}{\overline{V}_1(x-z)}\int_0^y \overline{G}^{*l}(y-s)\mathrm{d}G(s)\right\}\int_z^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) \\
&\quad + \left\{\frac{a_{0,0}c_1(x)}{\overline{V}_1(x-z)}\overline{G}(y) + \frac{a_{0,1}c_1(x)}{\overline{V}_1(x-z)}\int_0^y \overline{G}^{*l}(y-s)\mathrm{d}G(s)\right\}\int_0^z \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) \\
&= \frac{a_{0,0}c_1(x)}{\phi_1 \overline{V}_1(x-z)}\overline{G}(y) + \frac{a_{0,1}c(x)}{\phi_1 \overline{V}_1(x-z)} \int_0^y \overline{G}^{*l}(y-s)\mathrm{d}G(s).
\end{aligned} \tag{2.147}$$

From (2.145), we will have

$$\overline{H}(z,y) = \lim_{k,l\to\infty} \overline{H}_{k,l}(z,y) \le \frac{a_{0,0}}{\phi_1 \overline{V}_1(x-z)}c_1(x)\overline{G}(y) + \frac{a_{0,1}}{\phi_1 \overline{V}_1(x-z)}c_1(x)G(y), \quad 0 \le z \le x, \tag{2.148}$$

and (2.144) follows directly with $z = x$.

2.4.4 General upper bound III

If we further assume that $b_{n+1} \le \phi_2 b_n$ in addition to the assumptions in Theorem 2.3.1, we can obtain the following result:

Theorem 2.4.10. *Suppose there exists numbers $\phi_1, \phi_2 \in (0,1)$ such that*

$$a_{m+1,n} \le \phi_1 a_{m,n}, \quad m,n = 0,1,2,\ldots, \tag{2.149}$$

and

$$b_{n+1} \le \phi_2 b_n, \quad n = 0,1,2,\ldots, \tag{2.150}$$

where $B_i(t), i = 1, 2$ are d.f. satisfying the generalized adjustment equations

$$\int_0^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) = \frac{1}{\phi_1} \tag{2.151}$$

and

$$\int_0^\infty \{\overline{B}_2(t)\}^{-1}\mathrm{d}G(t) = \frac{1}{\phi_2}. \tag{2.152}$$

If $V_i(x), i = 1, 2$ are d.f. satisfying

$$\overline{V}_i(x)\overline{B}_i(t) \le \overline{V}_i(x + t), \quad x, t \ge 0, \tag{2.153}$$

then

$$\overline{H}(x, y) \le \frac{b_0 c_1(x) c_2(y)}{\phi_1\phi_2\overline{V}_1(0)\overline{V}_2(0)}, \quad x, y \ge 0, \tag{2.154}$$

where $c_1(x), c_2(y)$ satisfy

$$\frac{1}{c_1(x)} = \inf_{0\le z\le x, \overline{F}(z)>0} c_1(x, z), \quad x \ge 0, \tag{2.155}$$

$$\frac{1}{c_2(y)} = \inf_{0\le w\le y, \overline{G}(w)>0} c_2(y, w), \quad y \ge 0, \tag{2.156}$$

with $c_1(x, z), c_2(y, w)$ given by

$$c_1(x, z) = \frac{\int_z^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t)}{\overline{V}_1(x - z)\overline{F}(z)}, \quad 0 \le z \le x, \quad \overline{F}(z) > 0 \tag{2.157}$$

and

$$c_2(y, w) = \frac{\int_w^\infty \{\overline{B}_2(s)\}^{-1}\mathrm{d}G(s)}{\overline{V}_2(y - w)\overline{G}(w)}, \quad 0 \le w \le y, \quad \overline{G}(w) > 0. \tag{2.158}$$

Proof. Firstly, we prove by induction on k that for $0 \le z \le x$

$$\overline{H}_{k,l}(z, y) \le \frac{c_1(x)}{\phi_1\overline{V}_1(x - z)}\{\sum_{j=1}^{l} q_j\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\}. \tag{2.159}$$

Note that for $k = 0$,

$$\begin{aligned}
\overline{H}_{0,l}(z, y) &= \overline{F}(z)\{\sum_{i=1}^{\infty}\sum_{j=1}^{l} p_{i,j}\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\} \\
&\le \overline{F}(z)\{\sum_{i=0}^{\infty}\sum_{j=1}^{l} p_{i,j}\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\} \\
&= \overline{F}(z)\{\sum_{j=1}^{l} q_j\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\} \\
&\le \frac{c_1(x)}{\phi_1\overline{V}_1(x - z)}\{\sum_{j=1}^{l} q_j\overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\}.
\end{aligned} \tag{2.160}$$

Now suppose that (2.159) holds for $k = n$, then for $k = n + 1$, it follows that

$$\begin{aligned}\overline{H}_{n+1,l}(z,y) &\leq \overline{H}_{0,l}(z,y) + \phi_1 \int_0^z \overline{H}_{n,l}(z-t,y)\mathrm{d}F(t)\\ &\leq \Big\{\frac{c_1(x)}{\overline{V}_1(x-z)}\int_z^\infty \{\overline{B}(t)\}^{-1}\mathrm{d}F(t) + c_1(x)\int_0^z \{\overline{V}_1(x+t-z)\}^{-1}\mathrm{d}F(t)\Big\}\\ &\quad \times \{\sum_{j=1}^{l} q_j \overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\}\\ &\leq \Big\{\frac{c_1(x)}{\overline{V}_1(x-z)}\int_z^\infty \{\overline{B}(t)\}^{-1}\mathrm{d}F(t) + \frac{c_1(x)}{\overline{V}_1(x-z)}\int_0^z \{\overline{B}(t)\}^{-1}\mathrm{d}F(t)\Big\}\\ &\quad \times \{\sum_{j=1}^{l} q_j \overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\}\\ &= \frac{c_1(x)}{\phi_1 \overline{V}_1(x-z)}\{\sum_{j=1}^{l} q_j \overline{G}^{*j}(y) + a_{0,l}\overline{G}^{*(l+1)}(y)\}. \end{aligned} \tag{2.161}$$

Therefore,

$$\overline{H}(z,y) = \lim_{k,l\to\infty} \overline{H}_{k,l}(z,y) \leq \frac{c_1(x)}{\phi_1\overline{V}_1(x-z)}\sum_{j=1}^{\infty} q_j\overline{G}^{*j}(y); \quad 0 \leq z \leq x, \tag{2.162}$$

and it follows directly that

$$\overline{H}(x,y) \leq \frac{c_1(x)}{\phi_1\overline{V}_1(0)}\sum_{j=1}^{\infty} q_j\overline{G}^{*j}(y) \tag{2.163}$$

with $z = x$.

Since we further assume that $b_{n+1} \leq \phi_2 b_n$, the upper bounds (2.11) in univariate case can be applied to (2.163) directly, which finally results in

$$\overline{H}(x,y) \leq \frac{b_0 c_1(x) c_2(y)}{\phi_1\phi_2\overline{V}_1(0)\overline{V}_2(0)}. \tag{2.164}$$

2.4.5 Parametric bounds: Pareto type and numerical experiments

In Section 2.3, we already gave several results and numerical example for exponential type upper bounds, which rely on the existence of the moment generating function of the d.f. $F(x)$ and $G(y)$. For heavy-tailed severity distributions whose moment generating functions do not exist, Willmot (1994) provided an alternative to consider a moment based bound when claim size moments are known, which we can use directly.

We assume that the moments of X exist up to the order m, i.e.

$$\mathbf{E}[X^j] = \int_0^\infty x^j \mathrm{d}F(x) < \infty, \quad j \leq m. \tag{2.165}$$

Since the Pareto distributions is DFR and hence NWU, it is convenient to choose the Pareto tail $\overline{B}_1(x) = \overline{V}_1(x) = (1+\kappa_1 x)^{-m}$, where κ_1 satisfies

$$\int_0^\infty (1+\kappa_1 x)^m \mathrm{d}F(x) = \frac{1}{\phi_1}. \tag{2.166}$$

It is convenient if m is a positive integer, in which case (2.166) becomes, after a binomial expansion,

$$\sum_{j=1}^{\infty} \binom{m}{j} \mathbf{E}[X^j]\kappa_1^j = \frac{1}{\phi_1} - 1, \tag{2.167}$$

a polynomial in κ_1 of degree m. If $m = 1$ then one obtains $\kappa_1 = (1-\phi_1)/\{\phi_1 \mathbf{E}[X]\}$ whereas if $m = 2$ one obtains

$$\kappa_1 = \frac{\sqrt{\frac{\mathbf{E}[X^2]}{\phi_1} - \mathbf{Var}[X]} - \mathbf{E}[X]}{\mathbf{E}[X^2]}. \tag{2.168}$$

We obtain the following general bound.

Corollary 2.4.11. *If $\phi_1 \in (0,1)$ satisfies (2.111), and $\kappa_1 > 0$ and $m > 0$ satisfy (2.166), then*

$$\overline{H}(x,y) \le \frac{a_{0,0}}{\phi_1}(1+\kappa_1 x)^{-m}, \quad x, y \ge 0. \tag{2.169}$$

Proof. Corollary 2.4.6 applies with $\overline{B}_1(x) = (1+\kappa_1 x)^{-m}$.

Example 2.4.1. *We continue using the assumptions in the former numerical experiment, but replace X_i, $i = 1,2,\ldots$ with $X_i \sim Pareto(80/3, 3)$ instead, which results in the same mean. In this case, the claim size distribution has a heavy tail thus the moment generating function does not exist. For $m = 2$ and $j \le m$, $\mathbf{E}[X^j] < \infty$. Therefore, choosing κ_1 as given in (2.168), the upper bounds become*

$$\overline{H}(x,y) \le \frac{\min\{a_0, b_0\}}{\phi_1\phi_2}(1+\kappa_1 x)^{-m} e^{-\kappa_2 y} \tag{2.170}$$

with all other parameters remaining the same as those in the former example.

Probability	$\overline{H}(500,100)$	$\overline{H}(1000,100)$	$\overline{H}(2000,100)$	$\overline{H}(500,500)$	$\overline{H}(1000,1000)$
Simulation(10000 times)	0.4062000	0.2079000	0.0533000	0.1793000	0.02330000
$\frac{a_0}{\phi_1}(1+\kappa_1 x)^{-m}$	0.5781641	0.3762424	0.1956846	0.5781641	0.37624238
$\frac{a_{0,0}}{\phi_1}(1+\kappa_1 x)^{-m}$	0.5326464	0.3466216	0.1802787	0.5326464	0.34662156
$\frac{a_{0,0}\overline{G}(y)+a_{0,1}G(y)}{\phi_1}(1+\kappa_1 x)^{-m}$	0.4907125	0.3193329	0.1660858	0.4907122	0.31933272
$\frac{b_0}{\phi_2}e^{-\kappa_2 y}$	0.8523254	0.8523254	0.8523254	0.4498081	0.20232731
$\frac{b_0}{\phi_1\phi_2}(1+\kappa_1 x)^{-m}e^{-\kappa_2 y}$	0.5180496	0.3371226	0.1753383	0.2733966	0.08002707
$\frac{a_0}{\phi_1\phi_2}(1+\kappa_1 x)^{-m}e^{-\kappa_2 y}$	0.5338265	0.3473895	0.1806781	0.2817228	0.08246426

Table 2.2: Comparison of general upper bounds

2.4.6 General upper bound IV

Theorem 2.4.12. *Suppose there exists numbers $\phi \in (0,1)$ such that*

$$a_{m+1,n+1} \leq \phi a_{m,n}, \quad m,n = 0,1,2,\dots, \tag{2.171}$$

and there exists $(\phi_1, \phi_2) \in (0,1)^2$ such that

$$a_{m+1} \leq \phi_1 a_m, \quad m = 0,1,\dots, \tag{2.172}$$

$$b_{n+1} \leq \phi_2 a_n, \quad n = 0,1,\dots, \tag{2.173}$$

$$\phi_1 * \phi_2 = \phi. \tag{2.174}$$

$B_i(t)$, $i = 1,2$ are d.f. satisfying the generalized adjustment equations

$$\int_0^\infty \{\overline{B}_1(t)\}^{-1} \mathrm{d}F(t) = \frac{1}{\phi_1} \tag{2.175}$$

and

$$\int_0^\infty \{\overline{B}_2(t)\}^{-1} \mathrm{d}G(t) = \frac{1}{\phi_2}. \tag{2.176}$$

If $V_i(x)$, $i = 1,2$ are d.f. satisfying

$$\overline{V}_i(x)\overline{B}_i(t) \leq \overline{V}_i(x+t), \quad x,t \geq 0, \tag{2.177}$$

then

$$\overline{H}(x,y) \leq \max\left\{\frac{a_0}{\phi_1\phi}, \frac{b_0}{\phi_2\phi}\right\} \frac{c_1(x)c_2(y)}{\overline{V}_1(0)\overline{V}_2(0)}, \quad x,y \geq 0, \tag{2.178}$$

where $c_1(x)$, $c_2(y)$ satisfy

$$\frac{1}{c_1(x)} = \inf_{0 \leq z \leq x, \overline{F}(z) > 0} c_1(x,z), \quad x \geq 0, \tag{2.179}$$

$$\frac{1}{c_2(y)} = \inf_{0 \leq w \leq y, \overline{G}(w) > 0} c_2(y,w), \quad y \geq 0, \tag{2.180}$$

with $c_1(x,z)$, $c_2(y,w)$ given by

$$c_1(x,z) = \frac{\int_z^\infty \{\overline{B}_1(t)\}^{-1} \mathrm{d}F(t)}{\overline{V}_1(x-z)\overline{F}(z)}, \quad 0 \leq z \leq x, \quad \overline{F}(z) > 0 \tag{2.181}$$

and

$$c_2(y,w) = \frac{\int_w^\infty \{\overline{B}_2(s)\}^{-1} \mathrm{d}G(s)}{\overline{V}_2(y-w)\overline{G}(w)}, \quad 0 \leq w \leq y, \quad \overline{G}(w) > 0. \tag{2.182}$$

Proof. We prove by induction that for $0 \leq z \leq x$ and $0 \leq w \leq y$,

$$\overline{H}_{k,l}(z,w) \leq \max\left\{\frac{a_0}{\phi_1\phi}, \frac{b_0}{\phi_2\phi}\right\} \frac{c_1(x)c_2(y)}{\overline{V}_1(x-z)\overline{V}_2(y-w)}. \tag{2.183}$$

Intuitively, $c_1(x)$ is a non-decreasing function which satisfies

$$\overline{F}(z) \le \frac{c_1(x)}{\overline{V}_1(x-z)} \int_z^\infty \{\overline{B}_1(t)\}^{-1} \mathrm{d}F(t), \quad 0 \le z \le x. \tag{2.184}$$

For any l, from (2.63), we have

$$\begin{aligned}
\overline{H}_{0,l}(z,w) &\le \overline{F}(z)\{\sum_{j=1}^{l} q_j \overline{G}^{*j}(w) + a_{0,l}\overline{G}^{*(l+1)}(w)\} \\
&\le \overline{F}(z)\{\sum_{j=1}^{l} q_j \overline{G}^{*j}(w) + b_l\overline{G}^{*(l+1)}(w)\} \\
&\le \overline{F}(z)\sum_{j=1}^{\infty} q_j \overline{G}^{*j}(w) \\
&\le \frac{c_1(x)}{\overline{V}_1(x-z)} \int_z^\infty \{\overline{B}_1(t)\}^{-1} \mathrm{d}F(t) \sum_{j=1}^{\infty} q_j \overline{G}^{*j}(w), \\
&\le \frac{c_1(x)}{\phi_1 \overline{V}_1(x-z)} \sum_{j=1}^{\infty} q_j \overline{G}^{*j}(w), \\
&\le \frac{c_1(x)}{\phi_1 \overline{V}_1(x-z)} \cdot \frac{b_0 c_2(y)}{\phi_2 \overline{V}_2(y-w)}; \quad 0 \le z \le x, 0 \le w \le y,
\end{aligned} \tag{2.185}$$

where the last inequality uses the proof of Theorem 4.2.1 in Willmot and Lin (1994). Hence, for $0 \le z \le x$ and $0 \le w \le y$,

$$\overline{H}_{0,l}(z,w) \le \frac{b_0 c_1(x) c_2(y)}{\phi \overline{V}_1(x-z)\overline{V}_2(y-w)} \le \frac{b_0}{\phi_2 \phi} \frac{c_1(x) c_2(y)}{\overline{V}_1(x-z)\overline{V}_2(y-w)}, \tag{2.186}$$

and similarly, we can obtain for any k,

$$\overline{H}_{k,0}(z,w) \le \frac{a_0}{\phi_1 \phi} \frac{c_1(x) c_2(y)}{\overline{V}_1(x-z)\overline{V}_2(y-w)}. \tag{2.187}$$

Now, suppose that inequality holds for k, l. Referring to (2.100), we have

$$\begin{aligned}
\overline{H}_{k+1,l+1}(z,w) &\le a_{0,0}\overline{F}(z)\overline{G}(w) + \phi \int_0^z \int_0^w \overline{H}_{k,l}(z-t, w-s)\mathrm{d}F(t)\mathrm{d}G(s) \\
&\quad + \overline{F}(z) \int_0^w \{\sum_{n=1}^{\infty} q_n \overline{G}^{*n}(w-s)\}\mathrm{d}G(s) + \overline{G}(w) \int_0^z \{\sum_{m=1}^{\infty} p_m \overline{F}^{*m}(z-t)\}\mathrm{d}F(t) \\
&\le a_{0,0} \cdot \frac{c_1(x)}{\overline{V}_1(x-z)} \int_z^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) \cdot \frac{c_2(y)}{\overline{V}_2(y-w)} \int_w^\infty \{\overline{B}_2(s)\}^{-1}\mathrm{d}G(s) \\
&\quad + \max\left\{\frac{a_0}{\phi_1}, \frac{b_0}{\phi_2}\right\} \int_0^z \int_0^w \frac{c_1(x)c_2(y)}{\overline{V}_1(x-z+t)\overline{V}_2(y-w+s)} \mathrm{d}F(t)\mathrm{d}G(s) \\
&\quad + \frac{c_1(x)}{\overline{V}_1(x-z)} \int_z^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) \cdot \frac{b_0}{\phi_2} \int_0^w \frac{c_2(y)}{\overline{V}_2(y-w+s)}\mathrm{d}G(s)
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{c_2(y)}{\overline{V}_2(y-w)} \int_w^\infty \{\overline{B}_2(s)\}^{-1} \mathrm{d}G(s) \cdot \frac{a_0}{\phi_1} \int_0^z \frac{c_1(x)}{\overline{V}_1(x-z+t)} \mathrm{d}F(t) \\
&\le \max\Big\{\frac{a_0}{\phi_1}, \frac{b_0}{\phi_2}\Big\} \frac{c_1(x)c_2(y)}{\overline{V}_1(x-z)\overline{V}_2(y-w)} \int_0^\infty \{\overline{B}_1(t)\}^{-1} \mathrm{d}F(t) \int_0^\infty \{\overline{B}_2(s)\}^{-1} \mathrm{d}G(s) \\
&= \max\Big\{\frac{a_0}{\phi_1\phi}, \frac{b_0}{\phi_2\phi}\Big\} \frac{c_1(x)c_2(y)}{\overline{V}_1(x-z)\overline{V}_2(y-w)}, \qquad (2.188)
\end{aligned}
$$

since $a_{0,0} \le \min\{a_0, b_0\}$, where $0 \le z \le x$, $0 \le w \le y$. Letting $z = x$ and $w = y$ completes our proof.

2.4.7 General lower bound

Theorem 2.4.13. *Suppose that $\theta_1 \in (0,1)$ satisfies the dominance condition that $a_{m+1,n} \ge \theta_1 a_{m,n}; m,n = 0,1,2,\ldots,$ where $B(t)$ is a d.f. satisfying the generalized adjustment equation $\int_0^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) = 1/\theta_1$. If $W(x)$ is a d.f. satisfying $\overline{W}(x)\overline{B}(t) \ge \overline{W}(x+t), x \ge 0, t \ge 0$, then*

$$
\overline{H}(x,y) \ge \frac{a_{0,0}}{\theta_1 \overline{W}_1(0)} d_1(x)\overline{G}(y), \quad x,y \ge 0, \qquad (2.189)
$$

where $d_1(x)$ satisfies

$$
\frac{1}{d_1(x)} = \sup_{0\le z\le x, \overline{F}(z)>0} d_1(x,z), \quad x \ge 0, \qquad (2.190)
$$

with $d(x,z)$ given by

$$
d_1(x,z) = \frac{\int_z^\infty \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t)}{\overline{W}_1(x-z)\overline{F}(z)}, \quad 0 \le z \le x, \quad \overline{F}(z) > 0. \qquad (2.191)
$$

Proof. We define that

$$
A_1(z) = 1 - \overline{A}_1(z) = \theta_1 \int_0^z \{\overline{B}_1(t)\}^{-1}\mathrm{d}F(t) \qquad (2.192)
$$

the sum of k independent random variables with density function $A_1(z)$, has density function $A_k(z) = 1 - \overline{A}_k(z)$, then by the law of total probability

$$
\overline{A}_{k+1}(z) = \overline{A}(z) + \int_0^z \overline{A}_k(z-y)\mathrm{d}A_1(y). \qquad (2.193)
$$

If

$$
a_{m,n} \ge \theta_1 a_{m-1,n}, \qquad (2.194)
$$

we can prove that

$$
\overline{H}(z,y) = \lim_{k,l\to\infty} \overline{H}_{k,l}(z,y) \ge \frac{a_{0,0}\overline{G}(y)}{\theta_1 \overline{W}_1(x-z)} d_1(x) \quad 0 \le z \le x. \qquad (2.195)
$$

We also prove the above conclusion by showing that for $k = 0,1,2,\ldots,$

$$
\overline{H}_{k,l}(z,y) \ge \frac{a_{0,0}\overline{G}(y)d_1(x)}{\theta_1 \overline{W}_1(x-z)} \overline{A}_{k+1}(z) \quad 0 \le z \le x. \qquad (2.196)
$$

Intuitively, $c_1(x)$ is a non-decreasing function which satisfies

$$\overline{F}(z) \geq \frac{d_1(x)}{\overline{W}_1(x-z)} \int_z^{\infty} \{\overline{B}_1(t)\}^{-1} \mathrm{d}F(t), \quad 0 \leq z \leq x. \tag{2.197}$$

For $k = 0$,

$$\begin{aligned}
\overline{H}_{0,l}(z, y) &= \sum_{n=0}^{l} a_{0,n} \overline{F}(z)(\overline{G}^{*(n+1)}(y) - \overline{G}^{*n}(y)) \\
&= \overline{F}(z)\{ \sum_{n=1}^{l} (a_{0,n-1} - a_{0,n}) \overline{G}^{*n}(y) + a_{0,l} \overline{G}^{*(l+1)}(y)\} \\
&= \overline{F}(z)\{ \sum_{n=1}^{l} \sum_{i=1}^{\infty} p_{i,n} \overline{G}^{*n}(y) + a_{0,l} \overline{G}^{*(l+1)}(y)\} \\
&\geq a_{0,0} \overline{F}(z) \overline{G}(y) \\
&\geq \frac{a_{0,0} \overline{G}(y) d_1(x)}{\overline{W}_1(x-z)} \int_z^{\infty} \{\overline{B}_1(t)\}^{-1} \mathrm{d}F(t) \\
&= \frac{a_{0,0} \overline{G}(y) d_1(x)}{\theta_1 \overline{W}_1(x-z)} \overline{A}_1(z).
\end{aligned} \tag{2.198}$$

Suppose the conclusion holds for k, then for $k+1$, we have

$$\begin{aligned}
\overline{H}_{k+1,l}(z, y) &\geq a_{0,0} \overline{F}(z) \overline{G}(y) + \theta_1 \int_0^z \overline{H}_{k,l}(z-t, y) \mathrm{d}F(t) \\
&\geq \frac{a_{0,0} \overline{G}(y) d_1(x)}{\theta_1 \overline{W}_1(x-z)} \overline{A}_1(z) + a_{0,0} d_1(x) \overline{G}(y) \int_0^z \{\overline{W}_1(x+t-z)\}^{-1} \overline{A}_{k+1}(z-t) \mathrm{d}F(t) \\
&\geq \frac{a_{0,0} \overline{G}(y) d_1(x)}{\theta_1 \overline{W}_1(x-z)} \Big\{ \overline{A}_1(z) + \theta_1 \int_0^z \{\overline{B}_1(t)\}^{-1} \overline{A}_{k+1}(z-t) \mathrm{d}F(t) \Big\} \\
&= \frac{a_{0,0} \overline{G}(y) d_1(x)}{\theta_1 \overline{W}_1(x-z)} \overline{A}_{k+2}(z),
\end{aligned} \tag{2.199}$$

which ends the proof.

It follows from Ross et al. (1996) that

$$m(x) = \sum_{k=1}^{\infty} A_k(x) \leq \infty, \tag{2.200}$$

implying that $\lim_{k \to \infty} A_k(x) = 0$. Thus,

$$\lim_{k \to \infty} \overline{A}_k(x) = 1. \tag{2.201}$$

For $0 \leq z \leq x$,

$$\overline{H}(z, y) = \lim_{k,l \to \infty} \overline{H}_{k,l}(z, y) \geq \frac{a_{0,0} \overline{G}(y)}{\theta_1 \overline{W}_1(x-z)} d_1(x) \lim_{k \to \infty} \overline{A}_{k+1}(x) = \frac{a_{0,0} \overline{G}(y)}{\theta_1 \overline{W}_1(x-z)} d_1(x). \tag{2.202}$$

Letting $z = x$, we finally obtain that

$$\overline{H}(x, y) \geq \frac{a_{0,0}\overline{G}(y)}{\theta_1 \overline{\overline{W}}_1(0)} d_1(x). \tag{2.203}$$

2.5 Conclusion

In this chapter, we extended the results in Willmot and Lin (1994) and Willmot et al. (2001) to a bivariate aggregate claim model. We derived four different upper bounds of bivariate tail probability for both light-tailed claim severities with finite moment generating functions and heavy-tailed claim severities without finite moment generating functions. The proposed upper bounds outperform the univariate one given in Willmot and Lin (1994) and Willmot et al. (2001) when they are used to bound bivariate tail probability. Several corollaries are given for the purpose of computation and application by introducing specific classes of distributions such as NWU and NWUC. A general lower bound for bivariate tail probability which is usually of less concern in practice is also presented.

Chapter 3

Simulation methods for compound distributions

3.1 Introduction and literature review

In collective risk theory, the total amount of losses that an insurance company incurs during a time period is modeled by a compound random variable

$$S = \sum_{i=1}^{M} X_i, \tag{3.1}$$

where M is a discrete number variable representing the number of claims; $X_1, X_2, \ldots$ are non-negative independent identically distributed claim size random variables independent of M. The tail probability of S, $\mathbf{P}[S > c]$ for some specified value c, and the tail mean, defined by $\mathbf{E}[(S-c)_+]$, are important quantities because they are closely related to important risk measures, such as the value at risk (VaR) and the tail value at risk (TVaR).

Evaluations of the tail probability and the tail mean are not easy, even when the distributions of M and X_i are known. One usually has to resort to recursive formulas, such as those proposed in Panjer (1981) for the case when the distribution of M belongs to the $(a, b, 0)$ class. There is an extensive literature on further developments related to Panjer's recursive formula. For details, one is referred to the comprehensive book by Sundt and Vernic (2009).

Transform based techniques, such as fast Fourier transform (FFT), are also widely used in calculating the distribution of aggregated claims. For an introduction, one can refer to Robertson (1992) or Wang (1998). Embrechts and Frei (2009) provided an excellent comparison of the recursive and FFT methods.

Simulation methods are flexible and can be handy in estimation of the tail probability/moments of compound distributions. However, they are subject to sampling errors. For example, the crude Monte Carlo method (CMC) for estimating $\theta = \mathbf{P}[S > c]$ is

$$\hat{\theta}_0 = \mathbb{I}(S > c), \tag{3.2}$$

where $\mathbb{I}(\cdot)$ is an indicator function, which takes value one if the argument is true and zero otherwise. Then $\hat{\theta}_0$ is an unbiased estimator of θ because

$$\mathbf{E}[\hat{\theta}] = \theta.$$

The variance of $\hat{\theta}_0$ is

$$\mathbf{Var}[\hat{\theta}_0] = \theta(1-\theta). \tag{3.3}$$

The coefficient of variation of $\hat{\theta}_0$ is

$$CV(\hat{\theta}_0) = \frac{\sqrt{\theta(1-\theta)}}{\theta} = \sqrt{\frac{(1-\theta)}{\theta}}. \tag{3.4}$$

When we are interested in tail probability so that θ is close to zero, the coefficient of variation will be huge, which makes the simulation method inefficient. Notice that when we conduct the simulation n times, the estimator for θ is

$$\hat{\theta}_{0,n} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}(S^{(j)} > c), \tag{3.5}$$

which has variance $\frac{1}{n}\theta(1-\theta)$ and

$$\begin{aligned} CV(\hat{\theta}_{0,n}) &= \frac{\sqrt{\theta(1-\theta)/n}}{\theta} = \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{(1-\theta)}{\theta}} \\ &\simeq \frac{1}{\sqrt{n\hat{\theta}}} = \Big(\sum_{j=1}^{n} \mathbb{I}(S^{(j)} > c)\Big)^{-1/2}. \end{aligned} \tag{3.6}$$

When θ is close to zero, $\sum_{j=1}^{n} \mathbb{I}(S^{(j)} > c)$ will be small, resulting in a large coefficient of variation.

When n is large, the distribution of $\hat{\theta}_{0,n}$ is approximately normal by the Central Limit Theorem. Then the $1-\alpha$ confidence interval of θ is given by

$$(\hat{\theta}_{0,n} - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\mathbf{Var}[\hat{\theta}_0]}, \hat{\theta}_{0,n} + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\mathbf{Var}[\hat{\theta}_0]}).$$

Therefore, the ratio of the length of the CI and the magnitude of the parameter of interest is about $2 \times \frac{z_{1-\alpha/2}}{\sqrt{n}} \times CV(\hat{\theta}_0)$. Consequently, for any estimator $\hat{\theta}$, $CV(\hat{\theta})$ is a good measure for its quality of as an estimator of θ. In fact, it can be considered as the relative error of the simulation estimator $\hat{\theta}$. For our case, the relative error of $\hat{\theta}_0$ is given in (3.4), which shows that $\hat{\theta}_0$ is not very efficient when θ is small.

Various variance reduction methods exist for improving the efficiency of CMC. Commonly used techniques include importance sampling, conditioning, stratification, and the use of control variable. For detailed introductions to variance reduction methods, one is referred for example, to Ross (2013) or Asmussen and Glynn (2007).

Clearly, simulation methods for evaluating the tail probability and tail mean of compound variables are very important for actuaries. However, quite surprisingly, the literature in this area is quite sparse. Relevant references include Peköz and Ross (2004), in which the "conditioning" approach was introduced specifically for compound variables. Glasserman et al. (2000) proposed a method for simulating tail probability (Value at Risk) of investment portfolios by combining the importance sampling and stratification method.

In this chapter, we study in detail variance reduction methods for simulating tail probability and tail mean of both univariate and bivariate compound variables. Then we propose several novel combinations of variance reduction methods specifically for compound distributions. These include the combination of importance sampling and conditioning and the combination of importance sampling and stratified sampling.

We also extend our methods to estimating the tail probability of bivariate compound variables. Particularly, we consider losses from two lines of business:

$$(S_1, S_2) = \Big(\sum_{i=1}^{M} X_i, \sum_{j=1}^{N} Y_j\Big), \tag{3.7}$$

where (M, N) is a vector of random variables representing the number of the claims or the two lines of business. M and N which can be dependent, have joint distribution $p_{m,n}$. The claim size random variables X_i, $i = 1, 2, \ldots$ and Y_j, $j = 1, 2, \ldots$ are mutually independent and are independent of the claim number M and N. We will provide several techniques for efficiently simulating the bivariate tail probability $\mathbf{P}[S_1 > c, S_2 > d]$ and bivariate tail moment $\mathbf{E}[(S_1 - c)_+ \times (S_2 - d)_+]$.

The remainder of this chapter is organized as follows. Section 3.2 reviews commonly used variance reduction methods; Section 3.3 applies them to the estimation of tail probability $P[S > c]$. Section 3.4 studies the simulation methods for the tail moment $(S - c)_+$. Sections 3.5 and 3.6 extend the results to bivariate compound variables. Section 3.7 concludes this chapter.

3.2 Review of variance reduction techniques

In this section, we briefly review several variance reduction techniques that will be used in this chapter. For a more comprehensive introduction to the topic, one is referred to Ross (2013) and Asmussen and Glynn (2007).

3.2.1 Importance sampling

Let $\mathbf{Z} = (Z_1, \ldots, Z_n)$ denote a vector of random variables having a joint density function $f(z) = f(z_1, \ldots, z_n)$ and suppose that we want to estimate

$$\theta = \mathbf{E}[g(\mathbf{Z})] = \int g(z) f(z) \mathrm{d}z, \tag{3.8}$$

where the integral is n-dimensional and over the support of z.

In importance sampling, one find a probability density function $f^*(z)$ such that $f(z) = 0$ whenever $f^*(z) = 0$. Since

$$\theta = \int \frac{g(z) f(z)}{f^*(z)} f^*(z) \mathrm{d}z = \mathbf{E}\Big[\frac{g(\mathbf{Z}^*) f(\mathbf{Z}^*)}{f^*(\mathbf{Z}^*)}\Big], \tag{3.9}$$

where $\mathbf{Z}^*$ has density $f^*(z)$,

$$\hat{\theta}_I = \frac{g(\mathbf{Z}^*) f(\mathbf{Z}^*)}{f^*(\mathbf{Z}^*)}$$

is an estimator for θ.

The importance sampling approach aims to choose an appropriate $f^*(z)$ which results in a smaller variance compared to the crude simulation in (3.8).

Since

$$\mathbf{Var}\Big[\frac{g(\boldsymbol{Z}^*)f(\boldsymbol{Z}^*)}{f^*(\boldsymbol{Z}^*)}\Big] = \mathbf{E}\Big[\Big(\frac{g(\boldsymbol{Z}^*)f(\boldsymbol{Z}^*)}{f^*(\boldsymbol{Z}^*)} - \theta\Big)^2\Big] = \int \frac{(g(z)f(z) - \theta f^*(z))^2}{f^*(z)}\mathrm{d}z, \tag{3.10}$$

in order to achieve a smaller variance, $f^*(z)$ should be chosen such that $g(z)f(z) - \theta f^*(z)$ is close to zero. That is, $f^*(z)$ is proportional to $g(z)f(z)$.

If $\boldsymbol{Z}$ has a finite moment generating function (m.g.f.)

$$M_{\boldsymbol{Z}}(\boldsymbol{t}) = \mathbf{E}[e^{\boldsymbol{Z}\cdot\boldsymbol{t}}],$$

then it is usually handy to choose $\boldsymbol{Z}^*$ to be the Esscher (exponential tilting) transform of $\boldsymbol{Z}$. That is, we let $\boldsymbol{Z}^*$ have p.d.f.

$$f^*(z) = \frac{e^{\boldsymbol{t}\cdot z}}{M_{\boldsymbol{Z}}(\boldsymbol{h})}f(z), \tag{3.11}$$

for some tilting parameter $\boldsymbol{h} \in (\boldsymbol{0}, \boldsymbol{b})$. In addition, the m.g.f. of $\boldsymbol{Z}^*$ is given by

$$M_{\boldsymbol{Z}^*}(\boldsymbol{t}) = \mathbf{E}[e^{\boldsymbol{t}\cdot\boldsymbol{Z}^*}] = \frac{\mathbf{E}[e^{(\boldsymbol{t}+\boldsymbol{h})\cdot\boldsymbol{Z}}]}{\mathbf{E}[e^{\boldsymbol{h}\cdot\boldsymbol{Z}}]} = \frac{M_{\boldsymbol{Z}}(\boldsymbol{t}+\boldsymbol{h})}{M_{\boldsymbol{Z}}(\boldsymbol{h})}. \tag{3.12}$$

Notice that if $\boldsymbol{Z}$ is a discrete random variable, we interpret f and f^* as probability mass function.

3.2.2 Variance reduction by conditioning

Suppose that Z and W are dependent random variables and we are interested in estimating

$$\theta = \mathbf{E}[Z]. \tag{3.13}$$

Suppose that $\mathbf{E}[Z \mid W]$ can be determined after W is simulated (efficiently), then it is a more efficient estimator of θ than the crude estimator Z because it is unbiased, i.e.,

$$\mathbf{E}[\mathbf{E}[Z \mid W]] = \mathbf{E}[Z] \tag{3.14}$$

and it has smaller variance than Z, because

$$\mathbf{Var}[Z] = \mathbf{Var}[\mathbf{E}[Z \mid W]] + \mathbf{E}[\mathbf{Var}[Z \mid W]] \geq \mathbf{Var}[\mathbf{E}[Z \mid W]]. \tag{3.15}$$

3.2.3 Stratified sampling

Stratified sampling resembles the conditioning method in the sense that a random variable W can help the simulation of the mean of a random variable Z. Suppose that W takes values in k strata $\mathcal{W}_1, \ldots, \mathcal{W}_k$ with probability $p_i = \mathbf{P}[W \in \mathcal{W}_i]$, then

$$\mathbf{E}[Z] = \sum_{i=1}^{k} \mathbf{E}[Z \mid W \in \mathcal{W}_i]p_i. \tag{3.16}$$

Suppose that we are planning to estimate $\mathbf{E}[Z]$ by n runs of simulation, with n_i samples from stratum i, then the stratified estimate of $\mathbf{E}[Z]$ is

$$\hat{\theta}_{S,n} = \sum_{i=1}^{k} \bar{Z}_i p_i, \tag{3.17}$$

where $\bar{Z}_i$ is the sample average of Z_i's conditional on $W = w_i$.

Let $\sigma_i = \mathbf{Var}[Z \mid W \in \mathcal{W}_i]$, then the variance of $\hat{\theta}_{S,n}$ is

$$Var(\hat{\theta}_{S,n}) = \sum_{i=1}^{k} p_i^2 \frac{\sigma_i^2}{n_i}. \tag{3.18}$$

If we choose $n_i = np_i$, $\hat{\theta}_{S,n}$ has a smaller variance than the crude simulation estimator $\hat{\theta}_{0,n} = \sum X_i/n$ because

$$\mathbf{Var}[\hat{\theta}_{S,n}] = \frac{1}{n}\sum_{i=1}^{k} p_i\sigma_i^2 = \frac{1}{n}\mathbf{E}[Var[Z \mid W]] \leq \frac{1}{n}[Var[Z]] = \mathbf{Var}[\hat{\theta}_{0,n}]. \tag{3.19}$$

It is worthwhile noting that $n_i = np_i$ is not necessarily the optimal number of simulations in stratum i. Particularly, if n_i is chosen to be proportional to $p_i\sigma_i$ (Fishman 1996), then $\mathbf{Var}[\hat{\theta}_{S,n}]$ is minimized.

3.2.4 The use of control variables

When using a control variable to reduce the variance of the estimator $\theta = \mathbf{E}[Z]$, we select a control variate W that is strongly positively or negatively correlated with Z. Then an unbiased estimator for θ is

$$\hat{\theta}_{CV} = Z + \gamma(W - \mathbf{E}[W])$$

for some constant γ. The variance of $\hat{\theta}_{CV}$ is

$$\mathbf{Var}[Z] + \gamma^2\mathbf{Var}[W] + 2\gamma\mathbf{Cov}(Z, W), \tag{3.20}$$

which is minimized when $\gamma = -\mathbf{Cov}(Z, W)/\mathbf{Var}[W]$. The minimum value is

$$\mathbf{Var}[Z](1 - \rho^2), \tag{3.21}$$

where

$$\rho = \mathbf{Corr}(Z, W) = \frac{\mathbf{Cov}(Z, W)}{\sqrt{\mathbf{Var}[Z]\mathbf{Var}[W]}}. \tag{3.22}$$

Remark 3.2.1. *There are many other simulation variance reduction methods. We only introduced a few that will be used in the later sections of this chapter. Note that these methods can be combined to reduce the simulation variance even further. We will explore a few combinations in the next section.*

3.3 Simulation of the tail probability of compound distributions

In this section, we applied the variance reduction methods introduced in the previous section to the estimation of the tail probability $\theta = \mathbf{P}[S > c]$ of the compound random variable defined in (3.1).

The crude estimator is simply $\hat{\theta}_0 = \mathbb{I}(S > c)$. If the simulation is run n times, then the estimator is given by $\hat{\theta}_{0,n}$ as defined in (3.5). The problem with the crude method is that when c is large, the sample size needs to be very large in order for $S > c$ to occur. In other words, as shown in equation (3.6), the crude estimator is not efficient.

3.3.1 Importance sampling

Let $f_S(x)$ be the probability density function (p.d.f.) of S[1]. We assume that S has a finite moment generating function (m.g.f.) $M_S(t) = \mathbf{E}[e^{tS}]$ on the interval $t \in [0, b)$, where $b > 0$.

In importance sampling, instead of sampling $\mathbb{I}(S > c)$, one samples $\mathbb{I}(S^* > c)L(S^*)$, where S^* has p.d.f. f^* and $L(S^*) = \frac{f(S^*)}{f^*(S^*)}$ is the likelihood ratio. Since we assume that S has a finite m.g.f., we may choose S^* to be the exponential tilting version (Esscher transform) of S. That is, we let S^* have p.d.f.

$$f_{S^*}(x) = \frac{e^{hx}}{M_S(h)} f_S(x), \tag{3.23}$$

for some tilting parameter h. In addition, the m.g.f. of S^* is given by

$$M_{S^*}(t) = \mathbf{E}[e^{tS^*}] = \frac{\mathbf{E}[e^{(t+h)S}]}{\mathbf{E}[e^{hS}]} = \frac{M_S(t+h)}{M_S(h)}. \tag{3.24}$$

Then, we have

$$\mathbf{E}[\mathbb{I}(S > c)] = \mathbf{E}\left[\mathbb{I}(S^* > c)\frac{M_S(h)}{e^{hS^*}}\right], \tag{3.25}$$

where S^* has p.d.f. $f_{S^*}(x)$.

Hence, the importance sampling estimator of θ is given by

$$\hat{\theta}_I = \mathbb{I}(S^* > c) M_S(h) \exp(-hS^*). \tag{3.26}$$

If the sample size is n, then we have

$$\hat{\theta}_{I,n} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}(S^{*(j)} > c) M_S(h) \exp(-hS^{*(j)}),$$

where $S^{*(j)}$ is the jth simulated value of S^*.

Note that choosing the appropriate value of h is critical in applying the importance sampling method. For our case, we would like to choose h such that $\mathbb{I}(S^* > c)$ is more likely to occur than $\mathbb{I}(S > c)$. Thus, a natural way is to select h such that $\mathbf{E}[S^*] = c$.

[1]Notice that S has a point mass at zero, where the p.d.f. is not defined. In the following, $f(0)$ and $f^*(0)$ are taken to be probability mass at zero.

Note 3.3.1. *As pointed out in Ross (2013),*

$$\hat{\theta}_I = \mathbb{I}(S^* > c) M_S(h) e^{-hS^*} \leq M_S(h) e^{-hc}. \tag{3.27}$$

Thus, one can select h to minimize the upper bound $M_S(h)e^{-hc}$. Setting its derivative to 0, we have

$$M'_S(h) - cM_S(h) = 0. \tag{3.28}$$

Thus, the optimal h should satisfy

$$c = \frac{M'_S(h)}{M_S(h)} = \mathbf{E}\left[S \frac{e^{hS}}{M_S(h)}\right] = \mathbf{E}[S^*]. \tag{3.29}$$

This verifies the intuitive selection of h.

We now study the distribution of S^*. Firstly, the m.g.f of S^* is

$$M_{S^*}(t) = \frac{M_S(t+h)}{M_S(h)} = \frac{P_M(M_X(t+h))}{P_M(M_X(h))} = \frac{P_M(M_X(h)M_{X^*}(t))}{P_M(M_X(h))}, \tag{3.30}$$

where $P_M(z) = \mathbf{E}[z^M]$ represents the probability generating function (p.g.f.) of M and X^* is Esscher transform of X with parameter h.

Denote $c(h) = \log(M_X(h))$ and assume that M^* is the Esscher transform of M with parameter $c(h)$. Then, the p.g.f. of M^* is

$$P_{M^*}(z) = \frac{\mathbf{E}[e^{c(h)M} \cdot z^M]}{\mathbf{E}[e^{c(h)M}]} = \frac{P_M(e^{c(h)} \cdot z)}{P_M(e^{c(h)})} = \frac{P_M(M_X(h)z)}{P_M(M_X(h))}. \tag{3.31}$$

Therefore by (3.30),

$$M_{S^*}(t) = P_{M^*}(M_{X^*}(t)), \tag{3.32}$$

which shows that S^* is again a compound sum variable with claim number random variable M^* and claim size random variable X^*.

We next state the result.

Proposition 3.3.1. *The Esscher transform of $S = \sum_{i=1}^{M} X_i$ with parameter h is*

$$S^* = \sum_{i=1}^{M^*} X_i^*, \tag{3.33}$$

where X^* is the Esscher transform of X with parameter h and M^* is the Esscher transform of M with parameter $c(h) = \log(M_X(h))$.

The Esscher transforms of some commonly used distributions with parameter h are listed below:

- if $M \sim \mathrm{B}(n, p)$ with the probability mass function (p.m.f.)

$$p_M(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n, \tag{3.34}$$

  then $M^* \sim \mathrm{B}(n, \frac{pe^h}{1-p+pe^h})$;

- if $M \sim \mathrm{NB}(r, p)$ with the p.m.f.

$$p_M(k) = \binom{k+r-1}{k}(1-p)^r p^k, \quad k = 0, 1, 2, \ldots, \tag{3.35}$$

  then $M^* \sim \mathrm{NB}(r, pe^h)$, $h < -\ln p$;

- if $M \sim \mathrm{Poisson}(\lambda)$ with the p.m.f.

$$p_M(k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, 2, \ldots, \tag{3.36}$$

  then $M^* \sim \mathrm{Poisson}(\lambda e^h)$;

- if $X \sim \mathrm{Ga}(\alpha, \beta)$, with the p.d.f.

$$f_X(x) = \beta^\alpha x^{\alpha-1} e^{-\beta x}/\Gamma(\alpha), \quad x > 0, \tag{3.37}$$

  then $X^* \sim \mathrm{Ga}(\alpha, \beta - h)$, $h < \beta$.

Proposition 3.3.1 suggests that S^* is easy to simulate and the importance sampling method is easily implemented.

Example 3.3.1. *Assuming that $M \sim Possion(\lambda)$ and X_i, $i = 1, 2, \ldots$ have the common distribution $X \sim Ga(\alpha, \beta)$. Then*

$$S^* = \sum_{i=1}^{M^*} X_i^*, \tag{3.38}$$

where $M^* \sim Possion\left(\lambda(\frac{\beta}{\beta-h})^\alpha\right)$ and $X_i^* \sim Ga(\alpha, \beta - h)$.

3.3.2 Importance and stratified sampling

A method for applying stratified sampling in simulating quantities related to compound distribution was introduced in Section 9.5.3 of Ross (2013). This is done by treating the claim number M as the stratifying variable. Specifically, to simulate $\theta = g_N(X_1, \ldots, X_N)$, one can choose m such that $\mathbf{P}(N > m)$ is small and then make use of the fact that

$$\theta = \sum_{n=1}^{m} g_n(X_1, \ldots, X_n)p_n + \mathbf{E}[g_N(X_1, \ldots, X_N)|N > m](1 - \sum_{n=0}^{m} p_m).$$

This method can be applied to estimate the tail probability of S if we let $g_N(X_1, \ldots, X_N) = \mathbb{I}((X_1 + \cdots + X_N) > c)$. However, directly applying the method is not efficient because when c is large and the probability is small, one has to select m to be very large, resulting in a large number of strata. In addition, on strata with small N, the function to be evaluated $\mathbb{I}((X_1 + \cdots + X_N) > c)$ is likely to be zero.

Therefore, we propose a method that combines importance sampling and stratified sampling to simulate the tail probability of S. For this purpose, suppose that S^* is the Esscher transform of S with parameter h and that $\mathbf{E}[S^*] = c$.

Letting

$$g_m^*(x_1, \dots, x_m) = M_S(h)\mathbb{I}(\sum_{i=1}^{m} x_i > c)\exp(-h\sum_{i=1}^{m} x_i) \tag{3.39}$$

and

$$p_m^* = \mathbf{P}[M^* = m], \quad m = 1, 2, \dots,$$

then for a given value of m_l we have

$$\begin{aligned}
\theta &= \mathbf{E}[\mathbb{I}(S > c)] \\
&= \mathbf{E}\Big[\mathbb{I}(S^* > c)\frac{M_S(h)}{e^{hS^*}}\Big] \\
&= \mathbf{E}[g_{M^*}^*(X_1^*, \dots, X_{M^*}^*)] \\
&= \sum_{m=0}^{m_l} \mathbf{E}[g_{M^*}^*(X_1^*, \dots, X_{M^*}^*) \mid M^* = m]p_m^* + \mathbf{E}[g_{M^*}^*(X_1^*, \dots, X_{M^*}^*) \mid M^* > m_l]\mathbf{P}[M^* > m_l] \\
&= \sum_{m=0}^{m_l} \mathbf{E}[g_m^*(X_1^*, \dots, X_m^*)]p_m^* + \mathbf{E}[g_{M^*}^*(X_1^*, \dots, X_{M^*}^*) \mid M^* > m_l](1 - \sum_{m=0}^{m_l} p_m^*).
\end{aligned} \tag{3.40}$$

With this, we can follow the procedure outlined in Ross (2013) to simulate $\theta = \mathbf{P}[S > c]$. Concretely, we first fix a value of m_l such that $\mathbf{P}[M^* > m_l]$ is small and generate the value of M^* conditional on it exceeding m_l. Suppose that the generated value in a run is m'. Then generate m' independent random variables $X_1^*, \dots, X_{m'}^*$, the estimator from that run being

$$\mathcal{E} = \sum_{m=1}^{m_l} g_m^*(X_1^*, \dots, X_m^*)p_m^* + g_{m'}^*(X_1^*, \dots, X_{m'}^*)(1 - \sum_{m=0}^{m_l} p_m^*). \tag{3.41}$$

As pointed out in Ross (2013), since it is relatively easy to compute the functions g_m^*, we can use the data $X_1^*, \dots, X_{m'}^*$ in reverse order to obtain a second estimator, and then average the two estimators. That is, we let

$$\mathcal{E}' = \sum_{m=1}^{m_l} g_m^*(X_{m'}^*, \dots, X_{m'-m+1}^*)p_m^* + g_{m'}^*(X_{m'}^*, \dots, X_1^*)(1 - \sum_{m=0}^{m_l} p_m^*)$$

and the estimator from one run of simulation now becomes

$$\hat{\theta}_{I+S} = \frac{1}{2}(\mathcal{E} + \mathcal{E}'). \tag{3.42}$$

Remark 3.3.1. *By* (3.39)*, the second term in* (3.40) *is bounded by* $M_S(h)\mathbf{P}[M^* > m_l]$*, that is,*

$$\mathbf{E}[g_{M^*}^*(X_1^*, \dots, X_{M^*}^*) \mid M^* > m_l]\mathbf{P}[M^* > m_l] \leq M_S(h)\mathbf{P}[M^* > m_l].$$

Thus, if we select m_l *sufficiently large, so that* $M_S(h)\mathbf{P}[M^* > m_l]$ *is negligible, then the second term in* (3.41) *can be omitted.*

3.3.3 Conditioning

Peköz and Ross (2004) introduced a very effective way of simulating the tail probability of a compound distribution based on the conditioning method. Let

$$T(c) = \min\Big(m : \sum_{i=1}^{m} X_i > c\Big). \tag{3.43}$$

Then, one has that

$$S > c \Leftrightarrow M \geq T(c). \tag{3.44}$$

Therefore,

$$\begin{aligned} \mathbf{E}[\mathbb{I}(S > c)] &= \mathbf{E}[\mathbf{E}[\mathbb{I}(S > c) \mid T(c)]] \\ &= \mathbf{E}[\mathbf{E}[\mathbb{I}(M > T(c)) \mid T(c)]] \\ &= \mathbf{E}_{T(c)}[\mathbf{P}[M \geq T(c) \mid T(c)]]. \end{aligned}$$

To implement the method, we start by generating the values of X_i in sequence, and stops when the sum of generated values exceeds c. If the generated value of $T(c)$ is t_c, then we use $\mathbf{P}[M \geq t_c]$ as the estimate of $\mathbf{P}[S > c]$ from this run.

Overall, when applying conditioning method, the estimator for θ is

$$\hat{\theta}_{CD} = \mathbf{P}[M \geq T(c)]. \tag{3.45}$$

where $a_m = \mathbf{P}[M > m], m = 0, 1, 2, \ldots.$

We point out that (3.44) agrees with Proposition 1.1 of Lin (1996).

3.3.4 Combining conditioning method with control variates

As shown in Peköz and Ross (2004), the preceding conditional expectation estimator can be further improved by using a control variate. The idea is to select a control variate W that is strongly positively or negatively correlated with $\mathbf{E}[\mathbb{I}(M \geq T(c)) \mid T(c)]$. As suggested in Peköz and Ross (2004), one choice for W is

$$W = \sum_{i=1}^{T(c)} (X_i - \mathbf{E}[X]), \tag{3.46}$$

which has zero mean. W is positively corrected with $\mathbf{E}[\mathbb{I}(M \geq T(c)) \mid T(c)]$ because when $T(c)$ is large, (1) $\mathbf{E}[\mathbb{I}(M \geq T(c)) \mid T(c)]$ will be small; and (2) X_i's are likely to be small so that W will be small.

With the above, when combining the conditioning method with the control variates, the estimator for θ is

$$\hat{\theta}_{CD+CV} = \hat{\theta}_{CD} - \gamma W, \tag{3.47}$$

where γ can be chosen to be $\gamma = cov(\hat{\theta}_{CD}, W)/Var(W)$, which may be estimated using the simulated values of $\hat{\theta}_{CD}$ and W.

3.3.5 Combining importance sampling with conditioning

We show in this section that importance sampling can be combined with the conditioning method to efficiently simulate the tail probability of compound sums.

As in Section 3.3.1, let S^* and X^* be the Esscher transform of S and X with tilting parameter h, respectively. Let M^* be the Esscher transform of M with parameter $\log(M_X(h))$.

Define

$$T^*(c) = \min\Big(m : \sum_{i=1}^{m} X_i^* > c\Big), \tag{3.48}$$

and

$$S_M^* = \sum_{i=1}^{M} X_i^*. \tag{3.49}$$

We then have

$$\begin{aligned}
&\mathbf{E}[\mathbb{I}\{S > c\}]\\
&= \mathbf{E}\Big[\mathbb{I}\{S^* > c\}\frac{\mathbf{E}[e^{hS}]}{e^{hS^*}}\Big]\\
&= \mathbf{E}\Big[\mathbb{I}\{S_M^* \ge c\}\frac{\mathbf{E}[e^{hS}]}{e^{hS_M^*}}\frac{(M_X(h))^M}{\mathbf{E}[(M_X(h))^M]}\Big]\\
&= \mathbf{E}\Big[\mathbb{I}\{S_M^* \ge c\}\frac{M_X(h)^M}{e^{hS_M^*}}\Big]\\
&= \mathbf{E}\Big[\mathbb{I}\{M \ge T^*(c)\}\frac{M_X(h)^M}{e^{h\sum_{i=1}^M X_i^*}}\Big]\\
&= \mathbf{E}\Big[\mathbf{E}\Big[\mathbb{I}\{M \ge T^*(c)\}\frac{M_X(h)^M}{e^{h\sum_{i=1}^M X_i^*}} \mid T^*(c), X_1^*, \dots, X_{T^*(c)}^*\Big]\Big]\\
&= \mathbf{E}\Big[\mathbf{E}\Big[\mathbb{I}\{M \ge T^*(c)\}\frac{M_X(h)^{M-T^*(c)}}{e^{h(\sum_{i=1}^M X_i^* - \sum_{i=1}^{T^*(c)} X_i^*)}} \mid T^*(c), X_1^*, \dots, X_{T^*(c)}^*\Big]\frac{M_X(h)^{T^*(c)}}{e^{h\sum_{i=1}^{T^*(c)} X_i^*}}\Big]\\
&= \mathbf{E}\Big[\mathbf{P}[M \ge T^*(c)]\mathbf{E}\Big[\frac{M_X(h)^{M-T^*(c)}}{e^{h(\sum_{i=1}^M X_i^* - \sum_{i=1}^{T^*(c)} X_i^*)}} \mid T^*(c), M \ge T^*(c)\Big]\frac{M_X(h)^{T^*(c)}}{e^{h\sum_{i=1}^{T^*(c)} X_i^*}}\Big]\\
&= \mathbf{E}\Big[\mathbf{P}[M \ge T^*(c)]\frac{M_X(h)^{T^*(c)}}{e^{h\sum_{i=1}^{T^*(c)} X_i^*}}\Big].
\end{aligned} \tag{3.50}$$

In the last line, we use the fact that

$$\mathbf{E}\Big[\frac{M_X(h)^{M-T^*(c)}}{e^{h(\sum_{i=1}^M X_i^* - \sum_{i=1}^{T^*(c)} X_i^*)}} \mid T^*(c), M \ge T^*(c)\Big] = 1,$$

which holds since for any value of $M \ge T^*(c)$,

$$\mathbf{E}\Big[\frac{M_X(h)^{M-T^*(c)}}{e^{h(\sum_{i=1}^M X_i^* - \sum_{i=1}^{T^*(c)} X_i^*)}} \mid T^*(c)\Big] = 1.$$

To utilize this result in simulation, we start with generating the values of X_i^* in sequence, and stop when the sum of generated values exceeds c. We record the values of $T^*(c)$ and $X_1^*, \dots, X_{T^*(c)}^*$. Then the estimate of $\mathbf{P}[S > c]$ is

$$\hat{\theta}_{I+CD} = \mathbf{P}[M \geq T^*(c)] \frac{M_X(h)^{T^*(c)}}{e^{h\sum_{i=1}^{T^*(c)} X_i^*}}. \tag{3.51}$$

Remark 3.3.2. *The parameter h needs to be selected when combining the importance sampling and the conditioning methods. We suggest that a good value of h could be such that* $\mathbf{E}[S^*] = c$*, as in the importance sampling case. The reason is that the value of h that minimizes the variance of* $\mathbf{P}[M \geq T^*(c)] \frac{M_X(h)^{T^*(c)}}{e^{h\sum_{i=1}^{T^*(c)} X_i^*}}$ *is roughly the same as that minimizing the variance of* $\mathbb{I}\{S^* > c\} \frac{\mathbf{E}[e^{hS}]}{e^{hS^*}}$*. For a proof of the statement, one can refer to Appendix A.1.*

Remark 3.3.3. *A zero-mean control variate*

$$W = \sum_{i=1}^{T^*(c)} (X_i^* - \mathbf{E}[X^*]) \tag{3.52}$$

can be chosen to improve the estimator $\hat{\theta}_{I+CD}$*. This results in*

$$\hat{\theta}_{I+CD+CV} = \hat{\theta}_{I+CD} - \gamma W. \tag{3.53}$$

The numerical example in the next section will illustrate the effectiveness of this combination.

3.3.6 Numerical experiments

This section presents a numerical example to illustrate the efficiency of the different methods of simulating tail probability. We assume that $M \sim$ Poisson(20), $X_i \sim$ Ga(20, 0.5).

For each method, the tail probability was estimated using 1000 simulated samples (one round). The standard deviations were based on 100 rounds of simulations. We will follow this convention in all the numerical examples in the rest parts of the chapter.

In the technique "I+S" where stratified sampling is applied, we set $m_l = 50$. Our results are shown in Table 3.1 .

We observe that the combinations "I+S", "I+CD" and "I+CD+CV" methods performed well. The methods involving importance sampling tend to have small variance when c is large. Therefore, for simulating small tail probabilities, "I+S" and "I+CD" are recommended. The method "I+CD+CV" did not provide enough improvement over the method "I+CD", thus is not worth the extra steps.

3.4 Simulation of mean excess losses

This section introduces variance reduction methods for simulating the mean excess losses

$$\tau = \mathbf{E}[(S - c)_+].$$

c		1000	1200	1400
Analytical	mean	1.3908e-01	1.9822e-02	1.4701e-03
Technique 1	mean	1.3867e-01	1.9300e-02	1.4700e-03
(C)	sd	1.0866e-02	4.1280e-03	1.2906e-03
	sd/mean	7.8356e-02	2.1389e-01	8.7799e-01
Technique 2	mean	1.3919e-01	2.0001e-02	1.4599e-03
(I)	sd	5.5503e-03	1.0284e-03	9.2773e-05
	sd/mean	3.9877e-02	5.1420e-02	6.3549e-02
Technique 3	mean	1.3901e-01	1.9803e-02	1.4703e-03
(I+S)	sd	5.0935e-04	7.8608e-05	7.6717e-06
	sd/mean	3.6642e-03	3.9695e-03	5.2179e-03
Technique 4	mean	1.3878e-01	1.9892e-02	1.4752e-03
(CD)	sd	1.8385e-03	4.3484e-04	4.5747e-05
	sd/mean	1.3248e-02	2.1860e-02	3.1012e-02
Technique 5	mean	1.3906e-01	1.9811e-02	1.4677e-03
(CD+CV)	sd	5.1966e-04	1.7081e-04	2.0931e-05
	sd/mean	3.7371e-03	8.6218e-03	1.4261e-02
Technique 6	mean	1.3910e-01	1.9829e-02	1.4688e-03
(I+CD)	sd	6.4308e-04	1.1943e-04	9.0533e-06
	sd/mean	4.6232e-03	6.0230e-03	6.1636e-03
Technique 7	mean	1.3905e-01	1.9820e-02	1.4697e-03
(I+CD+CV)	sd	4.7882e-04	9.5443e-05	8.2514e-06
	sd/mean	3.4436e-03	4.8155e-03	5.6142e-03

Table 3.1: Comparison of the simulation methods for $\mathbf{P}[S > c]$

3.4.1 Importance sampling & importance and stratified sampling

Importance sampling method for simulating $\mathbf{E}[(S - c)_+]$ is similar to that for $\mathbf{P}[S > c]$. One simply replace $\mathbb{I}\{S^* > c\}$ by $(S^* - c)_+$ in equation (3.26), yielding

$$\hat{\tau}_I = (S^* - c)_+ M_S(h) \exp(-hS^*). \tag{3.54}$$

To combine importance and stratified sampling, one simply replace the function g_m^* in (3.39) by

$$g_m^*(x_1, \ldots, x_m) = \left(\sum_{i=1}^{m} x_i - c \right)_+ M_S(h) e^{-h \sum_{i=1}^{m} x_i}.$$

3.4.2 Conditioning

A method for simulating the mean excess losses using the conditioning method was discussed in Peköz and Ross (2004). The main idea is to define the quantity

$$A = \sum_{i=1}^{T(c)} X_i - c. \tag{3.55}$$

Then the conditional expectation estimator is constructed as

$$
\begin{aligned}
\hat{\tau}_{CD} &= \mathbf{E}[(S-c)_+ \mid T(c), A] \\
&= \sum_{i \geq T(c)} (A + (i - T(c))\mathbf{E}[X])\mathbf{P}[M = i] \\
&= (A - T(c)\mathbf{E}[X])\mathbf{P}[M \geq T(c)] + \mathbf{E}[X]\Big(\mathbf{E}[M] - \sum_{i<T(c)} i\mathbf{P}[M = i]\Big). \qquad (3.56)
\end{aligned}
$$

3.4.3 Combining conditioning method with control variates

The following two control variables can be used as control variates to enhance the performance of the conditioning method without increasing computational efforts:

$$W_1 = \sum_{i=1}^{T(c)} (X_i - \mathbf{E}[X]), \qquad (3.57)$$

and

$$W_2 = A - \mathbf{E}[A], \qquad (3.58)$$

both of which have zero mean.

Both W_1 and W_2 can easily be obtained with the sample generated for the conditioning method. $\mathbf{E}[A]$ is the numerical mean of generated A's, whereas $\mathbf{E}[X]$ can be calculated theoretically. Thus, we can construct an estimator

$$\hat{\tau}_{CD+CV} = \hat{\tau}_{CD} - \gamma_1 W_1 - \gamma_2 W_2, \qquad (3.59)$$

where γ_1, γ_2 are chosen to minimize the variance of $\hat{\tau}_{C+CV}$. This could be achieved by letting γ_1 and γ_2 take values that results from the linear regression

$$\hat{\tau}_{CD} = \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2 + \epsilon, \qquad (3.60)$$

where the values of the parameters $\hat{\tau}_{CD}$, W_1 and W_2 are generated in simulation using the conditioning method.

For general discussions on least squares or regression based methods to be used in conjunction with control variates, one is referred to Lavenberg and Welch (1981) and Davidson and MacKinnon (1992).

3.4.4 Combining importance sampling with conditioning

Similar to the simulation of a tail probability, we have

$$\begin{aligned}
\mathbf{E}[(S-c)_+] &= \mathbf{E}\Big[(S^*-c)_+\frac{\mathbf{E}[e^{hS}]}{e^{hS^*}}\Big] \\
&= \mathbf{E}\Big[(S^*-c)_+\frac{\mathbf{E}[M_X(h)^M]}{M_X(h)^{M^*}}\frac{M_X(h)^{M^*}}{e^{hS^*}}\Big] \\
&= \mathbf{E}\Big[(S_M^*-c)_+\frac{M_X(h)^M}{e^{hS_M^*}}\Big] \\
&= \mathbf{E}\Big[(S_M^*-c)_+\frac{M_X(h)^M}{e^{h\sum_{i=1}^M X_i^*}}\Big] \\
&= \mathbf{E}\Big[\mathbf{E}\Big[1\{M\geq T^*(c)\}(A^*+\sum_{i=T^*(c)+1}^{M}X_i^*)\frac{M_X(h)^{M-T^*(c)}}{e^{h\sum_{i=T^*(c)+1}^M X_i^*}}\mid T^*(c),A^*\Big]\frac{M_X(h)^{T^*(c)}}{e^{h\sum_{i=1}^{T^*(c)}X_i^*}}\Big] \\
&= \mathbf{E}\Big[(A^*+(\mathbf{E}[M-T^*(c)\mid M\geq T^*(c)])\mathbf{E}[X])\,\mathbf{P}[M\geq T^*(c)]\frac{M_X(h)^{T^*(c)}}{e^{h\sum_{i=1}^{T^*(c)}X_i^*}}\Big], \quad (3.61)
\end{aligned}$$

where A^* is defined by

$$A^* = \sum_{i=1}^{T^*(c)} X_i^* - c. \quad (3.62)$$

Therefore, we can construct an estimator

$$\hat{\tau}_{I+CD} = (A^*+(\mathbf{E}[M-T^*(c)\mid M\geq T^*(c)])\mathbf{E}[X])\,\mathbf{P}[M\geq T^*(c)]\frac{M_X(h)^{T^*(c)}}{e^{h\sum_{i=1}^{T^*(c)}X_i^*}}. \quad (3.63)$$

The quantity $\mathbf{E}[M-T^*(c)\mid M\geq T^*(c)]$ is related to a size-biased transform (see Denuit (2020)). We have

$$\mathbf{E}[M\mid M\geq T^*(c)] = \frac{\mathbf{E}[M1\{M\geq T^*(c)\}]}{\mathbf{P}[M\geq T^*(c)]} = \mathbf{E}[M]\frac{\mathbf{P}[\tilde{M}\geq T^*(c)]}{\mathbf{P}[M\geq T^*(c)]}, \quad (3.64)$$

where $\tilde{M}$ is the size-biased version of M with distribution function

$$\mathbf{P}[\tilde{M}=k] = \frac{k\mathbf{P}[M=k]}{\mathbf{E}[M]} \quad k=0,1,\ldots. \quad (3.65)$$

Remark 3.4.1. *Equation* (3.64) *can be calculated efficiently because the distribution of* $\tilde{M}$ *and* M *are often related. For example, as shown in Ren (2021), if* M *belongs to the* $(a,b,0)$ *class with parameter* (a,b)*, then* $\tilde{M}-1$ *is in the* $(a,b,0)$ *class with parameter* $(a,a+b)$*. Particularly, if* M *follows a Poisson distribution with mean* λ*, then* $\tilde{M}-1$ *also follows a Poisson distribution with mean* λ*.*

Remark 3.4.2. *Control variates can be utilized to improve the results further. For example, let*

$$W_1 = \sum_{i=1}^{T^*(c)}(X_i^*-\mathbf{E}[X^*]), \quad (3.66)$$

and

$$W_2 = A^* - \mathbf{E}[A^*]. \tag{3.67}$$

Then the estimator

$$\hat{\tau}_{I+CD+CV} = \hat{\tau}_{I+CD} - \gamma_1 W_1 - \gamma_2 W_2. \tag{3.68}$$

can be used to improve $\hat{\tau}_{I+CD}$*. The numerical example in the next section illustrate the results.*

3.4.5 Numerical experiments

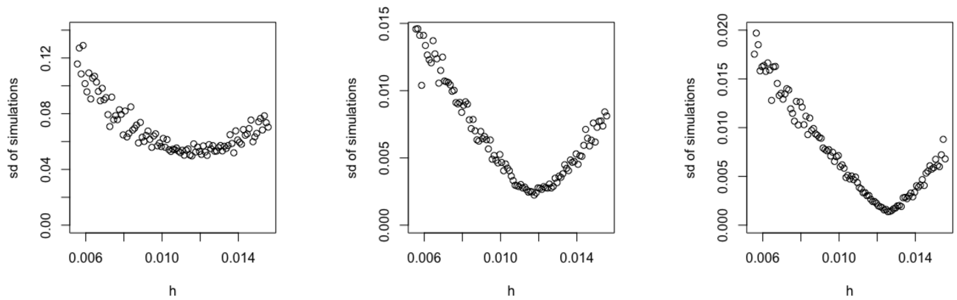
This section compares the different methods for simulating the mean excess losses by using the example described in Section 3.3.6. The results in Table 3.2 show that the combinations “I+S" and “I+CD+CV" are the most efficient for simulating mean excess losses.

c		1000	1200	1400
Technique 1	mean	14.4631	1.5426	9.4252e-02
(C)	sd	1.5972	4.8686e-01	9.5210e-02
	sd/mean	1.1044e-01	3.1561e-01	1.0102
Technique 2	mean	14.6203	1.5694	9.5366e-02
(I)	sd	4.6658e-01	4.9749e-02	4.1447e-03
	sd/mean	3.1913e-02	3.1698e-02	4.3461e-02
Technique 3	mean	14.4795	1.5728	9.5332e-02
(I+S)	sd	6.7333e-02	5.8922e-03	2.8451e-04
	sd/mean	4.6503e-03	3.7463e-03	2.9844e-03
Technique 4	mean	14.4677	1.5813	9.5695e-02
(CD)	sd	2.3288e-01	3.8123e-02	3.1565e-03
	sd/mean	1.6096e-02	2.4109e-02	3.2985e-02
Technique 5	mean	14.4938	1.5749	9.5259e-02
(CD+CV)	sd	5.1961e-02	1.3857e-02	1.4700e-03
	sd/mean	3.5851e-03	8.7989e-03	1.5431e-02
Technique 6	mean	14.4950	1.5738	9.5352e-02
(I+CD)	sd	8.2956e-02	8.2084e-03	3.7392e-04
	sd/mean	5.7230e-03	5.2158e-03	3.9215e-03
Technique 7	mean	14.4985	1.5746	9.5390e-02
(I+CD+CV)	sd	4.3460e-03	9.4551e-04	7.4557e-05
	sd/mean	2.9976e-04	6.0048e-04	7.8160e-04

Table 3.2: Comparison of the simulation methods for $\mathbf{E}[(S-c)_+]$

Remark 3.4.3. *When carrying out importance sampling related methods, we have set the the value of the tilting parameter to be the same as that for estimating the tail probability. That is, h is such that* $\mathbf{E}[S^*] = c$*. However, this may not be the optimal choice.*

For example, when $c = 1200$*, we have used* $h = 0.009561118$ *in the above. To explore the optimal value of h, we experimented and plotted in the following the standard deviation of the estimator (based on repeating the each method 100 times) against the value of h. Figure 3.1 shows that comparing with the tail probability case, one may want to set h to greater value* *when estimating mean excess losses. Determining theoretical results for the optimal value of the tilting parameter would be a suitable topic for future research.*



(a) I (b) I+S (c) I+CD

Figure 3.1: Standard deviation of the simulation results with different values of h

3.5 Simulation of tail probabilities-the two dimensional case

This section studies methods for simulating the tail probability

$$\theta = \mathbf{P}[S_1 > c, S_2 > d]$$

for the two-dimensional compound variable defined in (3.7). The estimator for θ in the raw simulation approach is just

$$\hat{\theta}_0 = \mathbb{I}\left(\sum_{i=1}^{M} X_i > c,\ \sum_{j=1}^{N} Y_j > d\right),$$

To perform the raw simulation, one first generates the value of M and N, say $M = m$ and $N = n$, then generates the values of $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ and use them to determine the value of $\hat{\theta}_0$. The average value of $\hat{\theta}_0$ over many such runs would then give an estimate for θ.

In this section, we introduce several variance reduction methods for simulating the tail probability $\mathbf{P}[S_1 > c, S_2 > d]$ and discuss their efficiency.

3.5.1 Importance sampling

Let $\mathbf{S} = (S_1, S_2)$ and $\mathbf{c} = (c, d)$. When using importance sampling to simulate $\mathbf{P}(\mathbf{S} > \mathbf{c})$, instead of sampling $\mathbb{I}(\mathbf{S} > \mathbf{c})$, one samples

$$\hat{\theta}_I = \mathbb{I}(\mathbf{S}^* > \mathbf{c})L(\mathbf{S}^*),$$

where $\mathbf{S}^*$ has p.d.f. $f_{\mathbf{S}^*}$ and $L(\mathbf{S}^*) = \frac{f_{\mathbf{S}}(\mathbf{S}^*)}{f_{\mathbf{S}^*}(\mathbf{S}^*)}$ is the likelihood ratio evaluated at $\mathbf{S}^*$.

A commonly used choice for $\mathbf{S}^*$ is the Esscher transform (exponential tilting) of $\mathbf{S}$. That is,

$$f_{\mathbf{S}}(\mathbf{x}) = \frac{M(\mathbf{h})}{e^{\mathbf{h}\cdot\mathbf{x}}} f_{\mathbf{S}^*}(\mathbf{x}),$$

where $M(\mathbf{h}) = \mathbf{E}[e^{\mathbf{h}\cdot\mathbf{S}}]$ and $\mathbf{h}\cdot\mathbf{S} = h_1S_1 + h_2S_2$ is the dot product of $\mathbf{h}$ and $\mathbf{S}$.

Thus,

$$L(\mathbf{x}) = \frac{M(\mathbf{h})}{e^{\mathbf{h}\cdot\mathbf{x}}},$$

Note that $\mathbf{S}^*$ has the m.g.f.

$$M_{\mathbf{S}^*}(\mathbf{t}) = \mathbf{E}[e^{t_1S_1^*+t_2S_2^*}] = \frac{\mathbf{E}[e^{(t_1+h_1)S_1+(t_2+h_2)S_2}]}{\mathbf{E}[e^{h_1S_1+h_2S_2}]} = \frac{M_{\mathbf{S}}(\mathbf{t}+\mathbf{h})}{M_{\mathbf{S}}(\mathbf{h})}. \tag{3.69}$$

Let $P_{(M,N)}(z_1,z_2) = \mathbf{E}[z_1^M z_2^N]$ be the p.g.f. of (M,N). Then (3.69) can be further written as

$$\begin{aligned} M_{S_1^*,S_2^*}(t_1,t_2) &= \frac{M_{S_1,S_2}(t_1+h_1,t_2+h_2)}{M_{S_1,S_2}(h_1,h_2)} \\ &= \frac{P_{M,N}(M_X(t_1+h_1),M_Y(t_2+h_2))}{P_{M,N}(M_X(h_1),M_Y(h_2))} \\ &= \frac{P_{M,N}(M_X(t_1)M_{X^*}(t_1),M_Y(t_2)M_{Y^*}(t_2))}{P_{M,N}(M_X(t_1),M_Y(t_2))}. \end{aligned} \tag{3.70}$$

Let $a(h_1) = \log(M_X(h_1))$ and $b(h_2) = \log(M_Y(h_2))$ and (M^*,N^*) be the 2d Esscher transform of (M,N) with parameter $a(h_1)$ and $b(h_2)$, respectively. Then, the p.g.f. of (M^*,N^*) is

$$P_{M^*,N^*}(z_1,z_2) = \frac{\mathbf{E}[(e^{a(h_1)}z_1)^M(e^{b(h_2)}z_2)^N]}{\mathbf{E}[e^{a(h_1)M+b(h_2)N}]} = \frac{P_{M,N}(M_X(t_1)z_1,M_Y(t_2)z_2)}{P_{M,N}(M_X(t_1),M_Y(t_2))}. \tag{3.71}$$

Therefore, (3.70) can be written as

$$M_{S_1^*,S_2^*}(t_1,t_2) = P_{M^*,N^*}\Big(M_{X^*}(t_1),M_{Y^*}(t_2)\Big). \tag{3.72}$$

which shows that (S_1^*,S_2^*) is again a bivariate compound sum variable with bivariate claim number random variable (M^*,N^*) and claim size random variables X^* and Y^*.

We state the result in the following Proposition.

Proposition 3.5.1. *The 2d Esscher transform of* $(S_1,S_2) = (\sum_{i=1}^M X_i, \sum_{j=1}^N Y_j)$ *with parameter* (h_1,h_2) *is*

$$(S_1^*,S_2^*) = \Big(\sum_{i=1}^{M^*} X_i^*, \sum_{j=1}^{N^*} Y_j^*\Big), \tag{3.73}$$

where X^*, Y^* *are the Esscher transforms of* X, Y *with parameter* h_1, h_2*, respectively, and* (M^*,N^*) *is the 2d Esscher transform of* (M,N) *with parameter* $(\log(M_X(h_1)),\log(M_Y(h_2)))$.

Example 3.5.1. *Consider the model in* (3.7)*. Assume that* $\Lambda \sim Ga(\alpha,\beta)$*. Conditional on* Λ*, claim frequencies* $M \sim Po(\lambda_1\Lambda)$ *and* $N \sim Po(\lambda_2\Lambda)$*. Claim sizes* X_i*,* $i = 1,2,\ldots$ *are i.i.d. and follow a Gamma distribution with parameters* (α_1,β_1)*. Claim sizes* Y_i*,* $i = 1,2,\ldots$ *are i.i.d. and* *follow a Gamma distribution with parameters* (α_2, β_2). X_i*'s and* Y_i*'s are mutually independent. They are independent of* (M, N). *With this setup, we have*

$$\begin{aligned} M_{S_1,S_2}(t_1,t_2) &= \mathbf{E}[e^{t_1S_1+t_2S_2}] \\ &= \mathbf{E}_\Lambda[\mathbf{E}[e^{t_1S_1+t_2S_2} \mid \Lambda]] \\ &= \mathbf{E}_\Lambda[P_{M|\Lambda}(M_X(t_1)) \cdot P_{N|\Lambda}(M_Y(t_2))] \\ &= \mathbf{E}_\Lambda[e^{\lambda_1\Lambda(M_X(t_1)-1)+\lambda_2\Lambda(M_Y(t_2)-1)}] \\ &= M_\Lambda(\lambda_1(M_X(t_1)-1)+\lambda_2(M_Y(t_2)-1)) \\ &= \left(1-\frac{\lambda_1M_X(t_1)+\lambda_2M_Y(t_2)-\lambda_1-\lambda_2}{\beta}\right)^{-\alpha}. \end{aligned} \tag{3.74}$$

Therefore,

$$\begin{aligned} M_{S_1^*,S_2^*}(t_1,t_2) &= \frac{M_{S_1,S_2}(t_1+h_1,t_2+h_2)}{M_{S_1,S_2}(h_1,h_2)} \\ &= \left(1-\frac{\lambda_1M_X(t_1+h_1)+\lambda_2M_Y(t_2+h_2)-\lambda_1M_X(h_1)-\lambda_2M_Y(h_2)}{\beta+\lambda_1+\lambda_2-\lambda_1M_X(h_1)-\lambda_2M_Y(h_2)}\right)^{-\alpha} \\ &= \left(1-\frac{\lambda_1M_X(h_1)(\frac{M_X(t_1+h_1)}{M_X(h_1)}-1)+\lambda_2M_Y(h_2)(\frac{M_Y(t_2+h_2)}{M_Y(h_2)}-1)}{\beta+\lambda_1+\lambda_2-\lambda_1M_X(h_1)-\lambda_2M_Y(h_2)}\right)^{-\alpha}. \end{aligned} \tag{3.75}$$

This implies that (S_1^*, S_2^*) *has the representation*

$$(S_1^*, S_2^*) = (\sum_{i=1}^{M^*} X_i^*, \sum_{i=1}^{N^*} Y_i^*),$$

where conditional on Λ^*, $M^* \sim Po(\lambda_1M_X(h_1)\Lambda^*)$ *and* $N^* \sim Po(\lambda_2M_Y(h_2)\Lambda^*)$, *and* $\Lambda^* \sim Ga(\alpha, \beta+\lambda_1+\lambda_2-\lambda_1M_X(h_1)-\lambda_2M_Y(h_2))$.

For the claim sizes, $X_i^* \sim Ga(\alpha_1, \beta_1-h_1)$ *and* $Y_j^* \sim Ga(\alpha_2, \beta_2-h_2)$. *This result means that* (S_1^*, S_2^*) *can be easily simulated.*

Example 3.5.2. *Consider the case when the claim frequencies* (M, N) *have a common shock, that is* $M = M_1 + M_0$ *and* $N \sim M_2 + M_0$, *where* M_0, M_1, M_2 *are independent. Let* $a = \ln M_X(t_1)$ *and* $b = \ln M_Y(t_2)$, *we have*

$$\begin{aligned} M_{S_1^*,S_2^*}(t_1,t_2) &= \frac{P_{M,N}(e^aM_{X^*}(t_1), e^bM_{Y^*}(t_2))}{P_{M,N}(e^a,e^b)} \\ &= \frac{P_{M_1}(e^aM_{X^*}(t_1))}{P_{M_1}(e^a)}\frac{P_{M_2}(e^bM_{Y^*}(t_2))}{P_{M_2}(e^b)}\frac{P_{M_0}(e^{a+b}M_{X^*}(t_1)M_{Y^*}(t_2))}{P_{M_0}(e^{a+b})}. \end{aligned} \tag{3.76}$$

Now, we consider $M^* = M_1^* + M_0^*$, $N^* = M_2^* + M_0^*$, *where* M_0^*, M_1^*, M_2^* *are Esscher transforms* *of M_0, M_1 and M_2 with parameter $a+b$, a, and b, respectively. We have*

$$\begin{aligned}
P_{M^*,N^*}(z_1,z_2) &= \mathbf{E}[z_1^{M^*} z_2^{N^*}] \\
&= \mathbf{E}[z_1^{M_1^*}]\mathbf{E}[z_2^{M_2^*}]\mathbf{E}[(z_1z_2)^{M_0^*}] \\
&= \frac{\mathbf{E}[e^{aM_1}z_1^{M_1}]}{\mathbf{E}[e^{aM_1}]}\frac{\mathbf{E}[e^{bM_2}z_2^{M_2}]}{\mathbf{E}[e^{bM_2}]}\frac{\mathbf{E}[e^{(a+b)M_0}(z_1z_2)^{M_0}]}{\mathbf{E}[e^{(a+b)M_0}]} \\
&= \frac{P_{M_1}(e^a z_1)}{P_{M_1}(e^a)}\frac{P_{M_2}(e^b z_2)}{P_{M_2}(e^b)}\frac{P_{M_0}(e^{a+b}z_1z_2)}{P_{M_0}(e^{a+b})}.
\end{aligned} \tag{3.77}$$

Therefore,

$$M_{S_1^*,S_2^*}(t_1,t_2) = P_{M^*,N^*}(M_{X^*}(t_1), M_{Y^*}(t_2)). \tag{3.78}$$

Consequently, (S_1^*, S_2^*) has the same distribution as $(\sum_{i=1}^{M^*} X_i^*, \sum_{j=1}^{N^*} Y_j^*)$. For example,

- *if $M_0 \sim Poisson(\lambda_0)$, $M_1 \sim Poisson(\lambda_1)$, $M_2 \sim Poisson(\lambda_2)$, then $M_0^* \sim Poisson(\lambda_0 M_X(h_1) M_Y(h_2))$, $M_1^* \sim Poisson(\lambda_1 M_X(h_1))$, $M_2^* \sim Poisson(\lambda_2 M_Y(h_2))$. Thus (M^*, N^*) are still bivariate Poisson distribution with common shock;*

- *if $M_0 \sim NB(r_0, p)$, $M_1 \sim NB(r_1, p)$, $M_2 \sim NB(r_2, p)$, then $M_0^* \sim NB(r_0, pM_X(h_1)M_Y(h_2))$, $M_1^* \sim NB(r_1, pM_X(h_1))$, $M_2^* \sim NB(r_2, pM_Y(h_2))$. Thus, (M^*, N^*) may no longer follow a bivariate negative binomial distribution .*

The parameter $\mathbf{h}$ needs to be selected to use the importance sampling method. Similarly to the univariate case,

$$\mathbb{I}(S_1^* > c, S_2^* > d) M_{S_1,S_2}(h_1,h_2) e^{-h_1S_1^* - h_2S_2^*} \le M_{S_1,S_2}(h_1,h_2)e^{-h_1c-h_2d}. \tag{3.79}$$

A suitable choice of $\mathbf{h}$ ought to maximize the upper bound. To this end, taking derivative of the upper bound on the right hand side with respect to h_1 and h_2 and setting the result to 0, we have

$$c = \frac{\frac{\partial}{\partial h_1} M_{S_1,S_2}(h_1,h_2)}{M_{S_1,S_2}(h_1,h_2)} = \frac{\mathbf{E}[S_1 e^{h_1S_1+h_2S_2}]}{\mathbf{E}[e^{h_1S_1+h_2S_2}]} = \mathbf{E}[S_1^*], \tag{3.80}$$

and

$$d = \mathbf{E}[S_2^*]. \tag{3.81}$$

The values of h_1, h_2 can be determined from (3.80) and (3.81).

3.5.2 Importance and stratified sampling

Similarly to those in Section 3.3.2, we define

$$g_{m,n}^*(x_1,\ldots,x_m,y_1,\ldots,y_n) = \begin{cases} \frac{M_{S_1,S_2}(h_1,h_2)}{e^{h_1\sum_{i=1}^m x_i + h_2\sum_{j=1}^n y_j}}, & \text{if } \sum_{i=1}^m x_i > c \text{ and } \sum_{j=1}^n y_j > d \\ 0, & \text{if otherwise} \end{cases} \tag{3.82}$$

and

$$p_{m,n}^* = \mathbf{P}[M^* = m, N^* = n], \quad m,n = 0,1,2,\ldots. \tag{3.83}$$

For given m_l, n_l such that $1 - \mathbf{P}[M^* \leq m_l, N^* \leq n_l]$ is small and (m', n') generated from (M^*, N^*) conditional on at least one of m', n' exceeding m_l, n_l, respectively, the estimator for $\mathbf{P}[S_1 > c, S_2 > d]$ from a run ought to be

$$\mathcal{E} = \sum_{m=1}^{m_l}\sum_{n=1}^{n_l} g^*_{m,n}(X^*_1, \ldots, X^*_m, Y^*_1, \ldots, Y^*_m)p^*_{m,n} + g^*_{m',n'}(X^*_1, \ldots, X^*_{m'}, Y^*_1, \ldots, Y^*_{m'})(1 - \sum_{m=0}^{m_l}\sum_{n=0}^{n_l} p^*_{m,n}). \tag{3.84}$$

Note that in each run we need to generate $X^*_1, \ldots, X^*_{\max(m_l,m')}$ and $Y^*_1, \ldots, Y^*_{\max(n_l,n')}$ for use.

Denote $m'_l = \max(m_l, m')$ and $n'_l = \max(n_l, n')$. A second estimator is

$$\mathcal{E}' = \sum_{m=1}^{m_l}\sum_{n=1}^{n_l} g^*_{m,n}(X^*_{m'_l}, \ldots, X^*_{m'_l-m+1}, Y^*_{n'_l}, \ldots, Y^*_{n'_l-n+1})p^*_{m,n} + g^*_{m',n'}(X^*_{m'_l}, \ldots, X^*_{m'_l-m'+1}, Y^*_{n'_l}, \ldots, Y^*_{n'_l-n'+1})(1 - \sum_{m=0}^{m_l}\sum_{n=0}^{n_l} p^*_{m,n}). \tag{3.85}$$

Then we have

$$\begin{aligned}\hat{\theta}_{I+S} &= \frac{1}{2}(\mathcal{E} + \mathcal{E}') \\ &= \frac{1}{2}\Big(\mathcal{E} + \sum_{m=1}^{m_l}\sum_{n=1}^{n_l} g^*_{m,n}(X^*_{m'_l}, \ldots, X^*_{m'_l-m+1}, Y^*_{n'_l}, \ldots, Y^*_{n'_l-n+1})p^*_{m,n} \\ &\quad + g^*_{m',n'}(X^*_{m'_l}, \ldots, X^*_{m'_l-m'+1}, Y^*_{n'_l}, \ldots, Y^*_{n'_l-n'+1})(1 - \sum_{m=0}^{m_l}\sum_{n=0}^{n_l} p^*_{m,n})\Big).\end{aligned} \tag{3.86}$$

3.5.3 Conditioning

The conditioning method introduced in Section 3.3.3 can be extended to the bivariate case as follows.

Let $S_{1,m} = X_1 + X_2 + \cdots + X_m$ and $S_{2,n} = Y_1 + Y_2 + \cdots + Y_n$. Then $S_1 = S_{1,M}$ and $S_2 = S_{2,N}$. Define

$$T_1(c) = \min\,(m; S_{1,m} > c), \tag{3.87}$$

and

$$T_2(d) = \min\,(n; S_{2,n} > d). \tag{3.88}$$

Then we have

$$\mathbb{I}(S_1 > c, S_2 > d) \Leftrightarrow \mathbb{I}(M \geq T_1(c), N \geq T_2(d)) \tag{3.89}$$

Hence,

$$\mathbf{E}[\mathbb{I}(S_1 > c, S_2 > d) \mid T_1(c), T_2(d)] = \mathbf{P}[M \geq T_1(c), N \geq T_2(d) \mid T_1(c), T_2(d)]. \tag{3.90}$$

Therefore, because $(T_1(c), T_2(d))$ and (M, N) are independent,

$$\mathbf{E}[\mathbb{I}(S_1 > c, S_2 > d)] = \mathbf{E}\left[\mathbf{P}[M \geq T_1(c), N \geq T_2(d) \mid T_1(c), T_2(d)]\right]. \tag{3.91}$$

For a theoretical verification of the conditioning method, one can refer to Appendix A.2.

Consequently, an estimator for θ using conditioning method is

$$\hat{\theta}_{CD} = \mathbf{P}[M \geq T_1(c), N \geq T_2(d)].$$

To use the estimator to simulate $\mathbf{P}[S_1 > c, S_2 > d]$, we firstly generate $T_1(c)$, $T_2(d)$. If the generated value is $t_{1,c}$ and $t_{2,d}$ respectively, then we use $\mathbf{P}[M \geq t_{1,c}, N \geq t_{2,d}]$ as the estimate for this run.

3.5.4 Combining the conditioning method with control variates

The conditioning method can be improved by a control variable. For example, as in the discussions in Section 3.4.3, we can introduce the control variates

$$W_1 = \sum_{i=1}^{T_1(c)} (X_i - \mathbf{E}[X]) \tag{3.92}$$

and

$$W_2 = \sum_{j=1}^{T_2(d)} (Y_j - \mathbf{E}[Y]) \tag{3.93}$$

which are strongly positively related to $\mathbf{P}[M \geq T_1(c), N \geq T_2(d) \mid T_1(c), T_2(d)]$ and construct the estimator

$$\hat{\theta}_{CD+CV} = \hat{\theta}_C - \gamma_1 W_1 - \gamma_2 W_2,$$

where the parameter γ_1 and γ_2 can be estimated by running the regression:

$$\hat{\theta}_{CD} = \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2 + \epsilon.$$

3.5.5 Combining importance sampling with conditioning

Define

$$T_1^*(c) = \min\left(m : \sum_{i=1}^{m} X_i^* > c\right), \tag{3.94}$$

$$T_2^*(d) = \min\left(m : \sum_{j=1}^{m} Y_j^* > d\right), \tag{3.95}$$

and

$$S_{1,M}^* = \sum_{i=1}^{M} X_i^*, \tag{3.96}$$

$$S_{2,N}^* = \sum_{j=1}^{N} Y_j^*, \tag{3.97}$$

where X_i^* and Y_j^* are Esscher transforms of X_i and Y_j with parameters h_1 and h_2, respectively.

We have

$$\begin{aligned}
\mathbf{E}[1\{S_1 > c, S_2 > d\}] &= \mathbf{E}\Big[1\{S_1^* > c, S_2^* > d\}\frac{\mathbf{E}[e^{h_1 S_1 + h_2 S_2}]}{e^{h_1 S_1^* + h_2 S_2^*}}\Big] \\
&= \mathbf{E}\Big[1\{S_{1,M}^* > c, S_{2,N}^* > d\}\frac{\mathbf{E}[e^{h_1 S_1 + h_2 S_2}]}{e^{h_1 S_{1,M}^* + h_2 S_{2,N}^*}}\frac{M_X(h_1)^M M_Y(h_2)^N}{\mathbf{E}[M_X(h_1)^M M_Y(h_2)^N]}\Big] \\
&= \mathbf{E}\Big[1\{S_{1,M}^* > c, S_{2,N}^* > d\}\frac{M_X(h_1)^M M_Y(h_2)^N}{e^{h_1 S_{1,M}^* + h_2 S_{2,N}^*}}\Big] \\
&= \mathbf{E}\Big[1\{M \geq T_1^*(c), N \geq T_2^*(d)\}\frac{M_X(h_1)^M M_Y(h_2)^N}{e^{h_1 S_{1,M}^* + h_2 S_{2,N}^*}}\Big] \\
&= \mathbf{E}\Big[\mathbf{E}\Big[1\{M \geq T_1^*(c), N \geq T_2^*(d)\}\frac{M_X(h_1)^{M-T_1^*(c)} M_Y(h_2)^{N-T_2^*(d)}}{e^{h_1 \sum_{i=T_1^*(c)+1}^{M} X_i^* + h_2 \sum_{j=T_2^*(d)+1}^{N} Y_j^*}} \mid \triangle\Big] \\
&\quad * \frac{M_X(h_1)^{T_1^*(c)} M_Y(h_2)^{T_2^*(d)}}{e^{h_1 \sum_{i=1}^{T_1^*(c)} X_i^* + h_2 \sum_{j=1}^{T_2^*(d)} Y_j^*}}\Big] \\
&= \mathbf{E}\Big[\mathbf{E}\Big[\frac{M_X(h_1)^{M-T_1^*(c)} M_Y(h_2)^{N-T_2^*(d)}}{e^{h_1 \sum_{i=T_1^*(c)}^{M} X_i^* + h_2 \sum_{j=T_2^*(d)}^{M} Y_j^*}} \mid T_1^*(c), T_2^*(d), M \geq T^*(c), N \geq T_2^*(d)\Big] \\
&\quad * \mathbf{P}[M \geq T_1^*(c), N \geq T_2^*(d)]\frac{M_X(h_1)^{T_1^*(c)} M_Y(h_2)^{T_2^*(d)}}{e^{h_1 \sum_{i=1}^{T_1^*(c)} X_i^* + h_2 \sum_{j=1}^{T_2^*(d)} Y_j^*}}\Big] \\
&= \mathbf{E}\Big[\mathbf{P}[M \geq T_1^*(c), N \geq T_2^*(d)]\frac{M_X(h_1)^{T_1^*(c)} M_Y(h_2)^{T_2^*(d)}}{e^{h_1 \sum_{i=1}^{T_1^*(c)} X_i^* + h_2 \sum_{j=1}^{T_2^*(d)} Y_j^*}}\Big],
\end{aligned} \tag{3.98}$$

where $\triangle = \{T_1^*(c), T_2^*(d), X_1, \ldots, X_{T_1^*(c)}, Y_1, \ldots, Y_{T_2^*(d)}\}$.

In the last line, we have used the fact that

$$\mathbf{E}\Big[\frac{M_X(h_1)^{M-T_1^*(c)} M_Y(h_2)^{N-T_2^*(d)}}{e^{h_1 \sum_{i=T_1^*(c)+1}^{M} X_i^* + h_2 \sum_{j=T_2^*(d)+1}^{N} Y_j^*}} \mid T_1^*(c), T_2^*(d), M \geq T^*(c), N \geq T_2^*(d)\Big] = 1,$$

which is true because for any value of $M \geq T_1^*(c)$ and $N \geq T_2^*(d)$,

$$\mathbf{E}\Big[\frac{M_X(h_1)^{M-T_1^*(c)} M_Y(h_2)^{N-T_2^*(d)}}{e^{h_1 \sum_{i=T_1^*(c)+1}^{M} X_i^* + h_2 \sum_{j=T_2^*(d)+1}^{N} Y_j^*}} \mid T_1^*(c), T_2^*(d)\Big] = 1.$$

Overall, when combining importance sampling and conditioning method, the simulation estimator is given by

$$\hat{\theta}_{I+CD} = \mathbf{P}[M \geq T_1^*(c), N \geq T_2^*(d)]\frac{M_X(h_1)^{T_1^*(c)} M_Y(h_2)^{T_2^*(d)}}{e^{h_1 \sum_{i=1}^{T_1^*(c)} X_i^* + h_2 \sum_{j=1}^{T_2^*(d)} Y_j^*}}. \tag{3.99}$$

Therefore, compared to classical importance sampling as described in Section 3.5.1 and importance and stratified sampling described in Section 3.5.2, the combination of importance sampling with conditioning has the advantage that, we do not require the distribution or simulation of (M^*, N^*) in our procedure. Instead, we are still working on original (M, N). For some bivariate variables (M, N) whose distribution or survival function is already known, this is more convenient and we do not need to worry about the distribution or simulation of (M^*, N^*) any more.

Remark 3.5.1. *Additionally, control variates can be used to improve this method further. For example, we may use*

$$W_1 = \sum_{i=1}^{T_1^*(c)} (X_i^* - \mathbf{E}[X^*]), \tag{3.100}$$

and

$$W_2 = \sum_{j=1}^{T_2^*(d)} (Y_j^* - \mathbf{E}[Y^*]). \tag{3.101}$$

This results in

$$\hat{\theta}_{I+CD+CV} = \hat{\theta}_{I+CD} - \gamma_1 W_1 - \gamma_2 W_2. \tag{3.102}$$

The numerical example in the next section will illustrate the refinement.

3.5.6 Numerical experiments

Let $\Lambda \sim \text{Ga}(10, 0.5)$ and assume that conditional on Λ, claim frequencies M and N follow Poisson distributions with mean Λ and 0.75Λ, respectively. Claim severities X_i and Y_j follow Gamma distributions with parameters $(20, 0.5)$ and $(30, 0.6)$, respectively. For stratified sampling, $m_l = n_l = 150$. The values of θ simulated using the methods described above are reported in Table 3.3.

As opposed to the univariate case, the combinations "CD+CV", "I+CD" and "I+CD+CV" perform well. The "I+CD+CV" method does not provide enough improvement over the "I+CD" method to warrant the additional steps.

3.6 Simulation of mean excess losses-the two dimensional case

This section introduces variance reduction methods for simulating the bivariate mean excess losses

$$\tau = \mathbf{E}[(S_1 - c)_+ \times (S_2 - d)_+]. \tag{3.103}$$

3.6.1 Importance sampling & importance and stratified sampling

We can simply replace $\mathbb{I}(\mathbf{S} > \mathbf{c})$ by $(S_1^* - c)_+ \times (S_2^* - d)_+$ in Section 3.5.1 to obtain the simulation method via importance sampling.

To combine importance and stratified sampling, one simply replace the function $g_{m,n}^*$ in (3.82) by

$$g_{m,n}^*(x_1, \ldots, x_m, y_1, \ldots, y_n) = \left(\sum_{i=1}^{m} x_i - c\right)_+ \times \left(\sum_{j=1}^{n} y_j - d\right)_+ M_{S_1,S_2}(h_1, h_2) e^{-h_1 \sum_{i=1}^m x_i - h_2 \sum_{j=1}^n y_j}.$$

(c,d)		(1000,1000)	(1400,1400)	(1600,1600)
Technique 1	mean	1.1728e-01	1.2560e-02	3.3000e-03
(C)	sd	1.1097e-02	3.6218e-03	2.1296e-03
	sd/mean	9.4616e-02	2.8836e-01	6.4534e-01
Technique 2	mean	1.1686e-01	1.1996e-02	3.2343e-03
(I)	sd	4.9200e-03	8.0878e-04	2.4345e-04
	sd/mean	4.2103e-02	6.7421e-02	7.5272e-02
Technique 3	mean	1.1645e-01	1.2029e-02	3.2298e-03
(I+S)	sd	2.6964e-04	4.1684e-05	2.5491e-05
	sd/mean	2.3155e-03	3.4652 e-03	7.8923e-03
Technique 4	mean	1.1647e-01	1.2036e-02	3.2302e-03
(CD)	sd	5.4867e-04	7.8795e-05	2.4247e-05
	sd/mean	4.7108e-03	6.5468e-03	7.5063e-03
Technique 5	mean	1.1645e-01	1.2035e-02	3.2270e-03
(CD+CV)	sd	1.9324e-04	3.2022e-05	8.0387e-06
	sd/mean	1.6595e-03	2.6608e-03	2.4911e-03
Technique 6	mean	1.1651e-01	1.2035e-02	3.2271e-03
(I+CD)	sd	3.5603e-04	3.3553e-05	8.1848e-06
	sd/mean	3.0559e-03	2.7881e-03	2.5363e-03
Technique 7	mean	1.1644e-01	1.2030e-02	3.2282e-03
(I+CD+CV)	sd	2.1043e-04	2.9138e-05	7.5760e-06
	sd/mean	1.8071e-03	2.4222e-03	2.3469e-03

Table 3.3: Comparison of the simulation methods for $P[S_1 > c, S_2 > d]$

3.6.2 Conditioning

We use the notation in Section 3.5.3. In addition, define

$$A = S_{1,T_1(c)} - c, \tag{3.104}$$

and

$$B = S_{2,T_2(d)} - d. \tag{3.105}$$

Since

$$\tau = \mathbf{E}\Big[\sum_i \sum_j (S_{1,i} - c)^+ (S_{2,j} - d)^+ \mathbf{P}[M = i, N = j]\Big], \tag{3.106}$$

an estimator for τ using conditioning method is

$$\hat{\tau}_{CD} = \mathbf{E}\Big[\sum_i \sum_j (S_{1,i} - c)^+ (S_{2,j} - d)^+ \mathbf{P}[M = i, N = j] \mid T_1(c), T_2(d), A, B\Big]. \tag{3.107}$$

After simulating the values of $T_1(c), T_2(d), A$ and B, $\hat{\tau}_{CD}$ can be evaluated as follows

$$\hat{\tau}_{CD} = \mathbf{E}\Big[\sum_i \sum_j (S_{1,i} - c)^+ (S_{2,j} - d)^+ \mathbf{P}[M = i, N = j] \mid T_1(c), T_2(d), A, B\Big]$$

$$
\begin{aligned}
&= \sum_{i\geq T_1(c)} \sum_{j\geq T_2(d)} (A + (i - T_1(c))\mathbf{E}[X])(B + (j - T_2(d))\mathbf{E}[Y])\mathbf{P}[M = i, N = j] \\
&= (A - T_1(c)\mathbf{E}[X])(B - T_2(d)\mathbf{E}[Y])\mathbf{P}[M \geq T_1(c), N \geq T_2(d)] \\
&\quad + (B - T_2(d)\mathbf{E}[Y])\mathbf{E}[X] \sum_{i\geq T_1(c)} \sum_{j\geq T_2(d)} i\mathbf{P}[M = i, N = j] \\
&\quad + (A - T_1(c)\mathbf{E}[X])\mathbf{E}[Y] \sum_{i\geq T_1(c)} \sum_{j\geq T_2(d)} j\mathbf{P}[M = i, N = j] \\
&\quad + \mathbf{E}[X]\mathbf{E}[Y] \sum_{i\geq T_1(c)} \sum_{j\geq T_2(d)} ij\mathbf{P}[M = i, N = j]. \qquad (3.108)
\end{aligned}
$$

Remark 3.6.1. *Equation* (3.108) *shows that when using the conditioning method, the problem of estimating the tail moments of the aggregate loss* (S_1, S_2) *is replaced by a problem of computing the tail moments of the claim frequencies* (M, N).

3.6.3 Combining conditioning method with control variates

The estimator $\hat{\tau}_{CD}$ can be improved by introducing some control variables. For example, using

$$W_1 = \sum_{i=1}^{T_1(c)} (X_i - \mathbf{E}[X]),$$

$$W_2 = \sum_{j=1}^{T_2(d)} (Y_j - \mathbf{E}[Y]),$$

$$W_3 = A - \mathbf{E}[A],$$

and

$$W_4 = B - \mathbf{E}[B],$$

results in the estimator

$$\hat{\tau}_{CD+CV} = \hat{\tau}_{CD} - \gamma_1 W_1 - \gamma_2 W_2 - \gamma_3 W_3 - \gamma_4 W_4,$$

where the parameters $\gamma_1, \gamma_2, \gamma_3$ and γ_4 can be determined by fitting the linear regression model

$$\hat{\tau}_{CD} \sim \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2 + \gamma_3 W_3 + \gamma_4 W_4. \qquad (3.109)$$

3.6.4 Combining importance sampling with conditioning

Similarly to A and B, we define A^* and B^* by

$$A^* = S^*_{1,T_1^*(c)} - c \qquad (3.110)$$

and

$$B^* = S^*_{2,T_2^*(d)} - d. \qquad (3.111)$$

Then, we have

$$
\begin{aligned}
&\mathbf{E}[(S_1 - c)_+ \times (S_2 - d)_+\}] \\
&= \mathbf{E}\Big[(S_1 - c)_+ \times (S_2 - d)_+ \frac{\mathbf{E}[e^{h_1 S_1 + h_2 S_2}]}{e^{h_1 S_1^* + h_2 S_2^*}}\Big] \\
&= \mathbf{E}\Big[(S_1 - c)_+ \times (S_2 - d)_+ \frac{M_X(h_1)^{M^*} M_Y(h_2)^{N^*}}{e^{h_1 S_1^* + h_2 S_2^*}} \frac{M_{S_1,S_2}(h_1,h_2)}{M_X(h_1)^{M^*} M_Y(h_2)^{N^*}}\Big] \\
&= \mathbf{E}\Big[(S_{1,M}^* - c)_+ \times (S_{2,N}^* - d)_+ \frac{M_X(h_1)^{M} M_Y(h_2)^{N}}{e^{h_1 S_{1,M}^* + h_2 S_{2,N}^*}}\Big] \\
&= \mathbf{E}\Big[\mathbb{I}\{M \geq T_1^*(c), N \geq T_2^*(d)\}(A^* + \sum_{i=T_1^*(c)+1}^{M} X_i^*)(B^* + \sum_{j=T_2^*(d)+1}^{N} Y_j^*) \frac{M_X(h_1)^{M} M_Y(h_2)^{N}}{e^{h_1 S_{1,M}^* + h_2 S_{2,N}^*}}\Big] \\
&= \mathbf{E}\Big[\mathbf{E}\Big[\mathbb{I}\{M \geq T_1^*(c), N \geq T_2^*(d)\}(A^* + \sum_{i=T_1^*(c)+1}^{M} X_i^*)(B^* + \sum_{j=T_2^*(d)+1}^{N} Y_j^*) \\
&\qquad * \frac{M_X(h_1)^{M-T_1^*(c)} M_Y(h_2)^{N-T_2^*(d)}}{e^{h_1 \sum_{i=T_1^*(c)+1}^{M} X_i^* + h_2 \sum_{j=T_2^*(d)+1}^{N} Y_j^*}} \mid T_1^*(c), T_2^*(d), A^*, B^*\Big] \frac{M_X(h_1)^{T_1^*(c)} M_Y(h_2)^{T_2^*(d)}}{e^{h_1 \sum_{i=1}^{T_1^*(c)} X_i^* + h_2 \sum_{j=1}^{T_2^*(d)} Y_j^*}}\Big] \\
&= \mathbf{E}\Big[\mathbf{E}\Big[\mathbb{I}\{M \geq T_1^*(c), N \geq T_2^*(d)\}(A^* + (M - T_1^*(c))\mathbf{E}[X]) \\
&\qquad * (B^* + (N - T_2^*(d))\mathbf{E}[Y]) \mid T_1^*(c), T_2^*(d), A^*, B^*\Big] \frac{M_X(h_1)^{T_1^*(c)} M_Y(h_2)^{T_2^*(d)}}{e^{h_1 \sum_{i=1}^{T_1^*(c)} X_i^* + h_2 \sum_{j=1}^{T_2^*(d)} Y_j^*}}\Big] \\
&= \mathbf{E}\Big[\Big((A^* - T_1^*(c)\mathbf{E}[X])(B^* - T_2^*(d)\mathbf{E}[Y])\mathbf{P}[M \geq T_1^*(c), N \geq T_2^*(d)]\Big] \\
&\qquad + \mathbf{E}[X](B^* - T_2^*(d)\mathbf{E}[Y])\mathbf{E}[M\mathbb{I}\{M \geq T_1^*(c), N \geq T_2^*(d)\}] \\
&\qquad + \mathbf{E}[Y](A^* - T_1^*(c)\mathbf{E}[X])\mathbf{E}[N\mathbb{I}\{M \geq T_1^*(c), N \geq T_2^*(d)\}] \\
&\qquad + \mathbf{E}[X]\mathbf{E}[Y]\mathbf{E}[MN\mathbb{I}\{M \geq T_1^*(c), N \geq T_2^*(d)\}]\Big) \frac{M_X(h_1)^{T_1^*(c)} M_Y(h_2)^{T_2^*(d)}}{e^{h_1 \sum_{i=1}^{T_1^*(c)} X_i^* + h_2 \sum_{j=1}^{T_2^*(d)} Y_j^*}}\Big],
\end{aligned}
\tag{3.112}
$$

To implement the simulation, we need to calculate the following quantities: $\mathbf{E}[M\mathbb{I}\{M \geq m, N \geq n\}]$, $\mathbf{E}[N\mathbb{I}\{M \geq m, N \geq n\}]$ and $\mathbf{E}[MN\mathbb{I}\{M \geq m, N \geq n\}]$ for any positive integers m, n. Although we can use crude simulation to calculate them, here we provide a more efficient way to deal with these quantities.

Example 3.6.1. *Taking the case of our numerical experiments in Section 3.5.6 for example, suppose that $M_\Lambda \sim Po(\lambda_1\Lambda)$ and $N_\Lambda \sim Po(\lambda_2\Lambda)$ are conditional Poisson variables with common mixing parameter Λ. Then, $(M,N) \sim (M_\Lambda, N_\Lambda)$ when $\Lambda \sim Ga(\alpha_0,\beta_0)$. Denote $\tilde{M}_\Lambda$, $\tilde{N}_\Lambda$ the size-biased transform of M_Λ, N_Λ conditional on Λ, respectively. Specifically, we have $\tilde{M}_\Lambda \sim$*

$Po(\lambda_1\Lambda)+1$ and $\tilde{N}_\Lambda \sim Po(\lambda_2\Lambda)+1$. Hence,

$$\begin{aligned}
\mathbf{E}[M\mathbb{I}\{M \geq m, N \geq n\}] &= \mathbf{E}_\Lambda[\mathbf{E}[M\mathbb{I}\{M \geq m\} \mid \Lambda]\mathbf{P}[N \geq n \mid \Lambda]] \\
&= \mathbf{E}_\Lambda[\mathbf{E}[M_\Lambda]\mathbf{P}[\tilde{M}_\Lambda \geq m]\mathbf{P}[N_\Lambda \geq n]] \\
&= \lambda_1\mathbf{E}_\Lambda[\Lambda\mathbf{P}[\tilde{M}_\Lambda \geq m]\mathbf{P}[N_\Lambda \geq n]] \\
&= \lambda_1\mathbf{E}[\Lambda]\mathbf{E}_\Lambda\left[\mathbf{P}[\tilde{M}_\Lambda \geq m]\mathbf{P}[N_\Lambda \geq n]\frac{\Lambda}{\mathbf{E}[\Lambda]}\right] \\
&= \lambda_1\mathbf{E}[\Lambda]\mathbf{E}_{\tilde{\Lambda}}[\mathbf{P}[\tilde{M}_{\tilde{\Lambda}} \geq m]\mathbf{P}[N_{\tilde{\Lambda}} \geq n]] \\
&= \lambda_1\mathbf{E}[\Lambda]\mathbf{E}_{\tilde{\Lambda}}[\mathbf{P}[M_{\tilde{\Lambda}} \geq m-1]\mathbf{P}[N_{\tilde{\Lambda}} \geq n]],
\end{aligned} \tag{3.113}$$

where $\tilde{\Lambda} \sim Ga(\alpha_0+1,\beta_0)$. In this case, $(M_{\tilde{\Lambda}}, N_{\tilde{\Lambda}})$ is still bivariate Poisson conditional on $\tilde{\Lambda}$, and we only change the mixing variable Λ compared to the original (M,N). We are able to calculate the p.m.f. by applying the inverse FFT method to the joint characteristic function (c.f.). Similarly, we have

$$\begin{aligned}
\mathbf{E}[MN\mathbb{I}\{M \geq m, N \geq n\}] &= \mathbf{E}_\Lambda[\mathbf{E}[M_\Lambda]\mathbf{E}[N_\Lambda]\mathbf{P}[\tilde{M}_\Lambda \geq m]\mathbf{P}[\tilde{N}_\Lambda \geq n]] \\
&= \lambda_1\lambda_2\mathbf{E}_\Lambda[\Lambda^2\mathbf{P}[\tilde{M}_\Lambda \geq m]\mathbf{P}[\tilde{N}_\Lambda \geq n]] \\
&= \lambda_1\lambda_2\mathbf{E}[\Lambda^2]\mathbf{E}_\Lambda\left[\mathbf{P}[\tilde{M}_\Lambda \geq m]\mathbf{P}[\tilde{N}_\Lambda \geq n]\frac{\Lambda^2}{\mathbf{E}[\Lambda^2]}\right] \\
&= \lambda_1\lambda_2\mathbf{E}[\Lambda^2]\mathbf{E}_{\tilde{\Lambda}^{(2)}}[\mathbf{P}[\tilde{M}_{\tilde{\Lambda}^{(2)}} \geq m]\mathbf{P}[\tilde{N}_{\tilde{\Lambda}^{(2)}} \geq n]] \\
&= \lambda_1\lambda_2\mathbf{E}[\Lambda^2]\mathbf{E}_{\tilde{\Lambda}^{(2)}}[\mathbf{P}[M_{\tilde{\Lambda}^{(2)}} \geq m-1]\mathbf{P}[N_{\tilde{\Lambda}^{(2)}} \geq n-1]],
\end{aligned} \tag{3.114}$$

where $\tilde{\Lambda}^{(2)} \sim Ga(\alpha_0+2,\beta_0)$, the second moment transform of Λ. In this case, $(M_{\tilde{\Lambda}^{(2)}}, N_{\tilde{\Lambda}^{(2)}})$ is still bivariate Poisson conditional on $\tilde{\Lambda}^{(2)}$.

Example 3.6.2. *In addition, for the setting where $M \sim M_0+M_1$ and $N \sim M_0+M_2$ with M_0 being a common shock, the required quantities are as follows:*

$$\begin{aligned}
&\mathbf{E}[M\mathbb{I}\{M \geq m, N \geq n\}] \\
&= \mathbf{E}[(M_0+M_1)\mathbb{I}\{M \geq m, N \geq n\}] \\
&= \mathbf{E}[M_0\mathbb{I}\{M_1+M_0 \geq m, M_2+M_0 \geq n\}] + \mathbf{E}[M_1\mathbb{I}\{M_1+M_0 \geq m, N \geq n\}] \\
&= \mathbf{E}[M_0]\mathbf{P}[M_1+\tilde{M}_0 \geq m, M_2+\tilde{M}_0 \geq n] + \mathbf{E}[M_1]\mathbf{P}[\tilde{M}_1+M_0 \geq m, N \geq n],
\end{aligned} \tag{3.115}$$

and

$$\begin{aligned}
&\mathbf{E}[MN\mathbb{I}\{M \geq m, N \geq n\}] \\
&= \mathbf{E}[(M_0+M_1)(M_0+M_2)\mathbb{I}\{M \geq m, N \geq n\}] \\
&= \mathbf{E}[M_0^2\mathbb{I}\{M_0+M_1 \geq m, M_0+M_2 \geq n\}] + \mathbf{E}[M_1M_2\mathbb{I}\{M_1+M_0 \geq m, M_2+M_0 \geq n\}] \\
&\quad + \mathbf{E}[M_0M_1\mathbb{I}\{M_1+M_0 \geq m, M_2+M_0 \geq n\}] + \mathbf{E}[M_0M_2\mathbb{I}\{M_1+M_0 \geq m, M_2+M_0 \geq n\}] \\
&= \mathbf{E}[M_0^2]\mathbf{P}[\tilde{M}_0^{(2)}+M_1 \geq m, \tilde{M}_0^{(2)}+M_2 \geq n] + \mathbf{E}[M_1]\mathbf{E}[M_2]\mathbf{P}[\tilde{M}_1+M_0 \geq m, \tilde{M}_2+M_0 \geq n] \\
&\quad + \mathbf{E}[M_0]\mathbf{E}[M_1]\mathbf{P}[\tilde{M}_1+\tilde{M}_0 \geq m, M_2+\tilde{M}_0 \geq n] + \mathbf{E}[M_0]\mathbf{E}[M_2]\mathbf{P}[M_1+\tilde{M}_0 \geq m, \tilde{M}_2+\tilde{M}_0 \geq n].
\end{aligned} \tag{3.116}$$

The probabilities in the above equations can be calculated by applying FFT method. This is the case because the underlying bivariate distributions are of common-shock type, whose Fourier transform can be computed easily.

Remark 3.6.2. *The method in this section can be further improved by applying the control variates*

$$W_1 = \sum_{i=1}^{T_1^*(c)} (X_i^* - \mathbf{E}[X^*]), \tag{3.117}$$

$$W_2 = \sum_{j=1}^{T_2^*(d)} (Y_j^* - \mathbf{E}[Y^*]), \tag{3.118}$$

$$W_3 = A^* - \mathbf{E}[A^*], \tag{3.119}$$

$$W_4 = B^* - \mathbf{E}[B^*]. \tag{3.120}$$

The numerical example in the next section will illustrate the refinement.

3.6.5 Numerical experiments

We keep the same assumptions as those stated in Section 3.5.6, the numerical experiments for $\mathbf{P}[S_1 > c, S_2 > d]$. The simulation results are reported in Table 3.4.

(c,d)		(1000,1000)	(1400,1400)	(1600,1600)
Technique 1	mean	11314.27	904.195	215.4179
(C)	sd	2153.544	544.6096	254.0361
	sd/mean	1.9034e-01	6.0231e-01	1.1793
Technique 2	mean	11059.18	862.5813	211.7861
(I)	sd	650.7682	48.4465	11.5515
	sd/mean	5.8844e-02	5.6165e-02	5.4543e-02
Technique 3	mean	11125.75	865.9839	210.7237
(I+S)	sd	33.6866	6.2868	2.6573
	sd/mean	3.0278e-03	7.2597e-03	1.2610e-02
Technique 4	mean	11116.76	864.5113	210.5138
(CD)	sd	61.1288	5.8745	1.6730
	sd/mean	5.4988e-03	6.7951e-03	7.9473e-03
Technique 5	mean	11119.97	864.2649	210.37
(CD+CV)	sd	8.0976	1.2639	3.7518e-01
	sd/mean	7.2821e-04	1.4624e-03	1.7834e-03
Technique 6	mean	11121.99	864.7473	210.765
(I+CD)	sd	43.6417	2.2537	4.7707e-01
	sd/mean	3.9239e-03	2.6063e-03	2.2635e-03
Technique 7	mean	11120.94	864.7559	210.7915
(I+CD+CV)	sd	3.1256	3.0802e-01	7.3821e-02
	sd/mean	2.8106e-04	3.5619e-04	3.5021e-04

Table 3.4: Comparison of the simulation methods for $\mathbf{E}[(S_1 - c)_+ \times (S_2 - d)_+]$

Similarly to Section 3.5.6, we observe that the combinations “CD+CV”,“I+CD" and “I+CD+CV" perform well. The “I+CD+CV" method does provide significant improvement over the “I+CD" method.

3.7 Conclusion

In this chapter, we gave a brief review on commonly used variance reduction techniques. By combining some of them with the importance sampling method, we came up with several novel variance reduction techniques (“I+S", “I+CD",“I+CD+CV"). In our four numerical experiments which simulate $\mathbf{P}[S > c]$, $\mathbf{E}[(S - c)_+]$, $\mathbf{P}[S_1 > c, S_2 > d]$, $\mathbf{E}[(S_1 - c)_+ \times (S_2 - d)_+]$, respectively, “I+CD" and “I+CD+CV" are always among the best three methods with the lowest simulation variance. For the remaining one of the best three methods,“I+S" performs better in univariate aggregate claim model but “CD+CV" performs better in the bivariate aggregate claim model. Compared to “CD" and “CD+CV" which are based on conditioning, the techniques combining them with importance sampling do reduce time consumption when simulating rare events. Interestingly, moment transform of claim severity distributions provides us with an alternative way to simulate mean excess accurately.

Chapter 4

Pricing principle for multivariate compound loss models

4.1 Introduction and literature review

Before issuing insurance policies, one of the main challenges facing the insurer is setting the right premium per exposure.

In this chapter, we explore a model for determining premiums when the insurer faces several uncertainties, most notably those associated with the severity of losses as well as with the number of policies issued during a given time period. Several business lines are assumed, and all of them are affected by a systemic risk. Namely, we suppose that a company runs d business lines, and for each line $k \in \{1, \ldots, d\}$, let N_k denote the number of policies to be issued, and let $X_{k,i}$ denote the loss to be incurred by the i^{th} policy holder. Hence, the company's total loss is

$$S = \sum_{k=1}^{d} S_k,$$

where

$$S_k = \sum_{i=1}^{N_k} X_{k,i}$$

is the total loss to be incurred by the k^{th} business line.

The losses $X_{k,i}$ are dependent within and between the business lines, and we assume that the dependence is due to a systemic risk, which we denote by Z. It could, for example, be a combination of various exogenous risk factors such as the company's overall status, general economic situation, and so on. The systemic risk may affect the frequencies N_k and the severities $X_{k,i}$ in a number of different ways, but in this chapter we assume that, for each business line k and conditionally on Z, the frequencies and the severities are independent, and the severities are also identically distributed. Note that in a special case when the systemic risk Z takes on a constant value, say z_0, with probability one, then the aforementioned model reduces to that based on compound sums with independent frequencies and severities, which has been a classical model in many actuarial texts (e.g., Klugman et al. (2012)).

To determine the capital needed to support the potential loss S, a premium calculation principle (pcp) is chosen. Generally speaking, the pcp is a functional π from the set of all loss random variables X to the set of all non-negative extended real-numbers $\pi[X]$ (theoretically, the premium can be infinite for some losses).

Premium calculation principles belonging to the first category are mainly studied by Goovaerts et al. (1984), which relate them to expected utility theory. Heilmann (1989) reconsidered this type of premium calculation principles and view some of them as the premiums that minimize the expected loss with a loss function $L : \mathbb{R}^2 \to \mathbb{R}$. In Heilmann (1989), another derivation of the exponential principle was provided by minimizing the expected loss $\mathbf{E}[L(X, \pi[X])]$ with a specified loss function $L(x, a) = (e^{\alpha x} - e^{\alpha a})^2$ for some $\alpha > 0$.

The second category of premium calculation principles are related to Yaari (1987) dual theory for choice under risk. Yaari's dual theory proposes the idea that attitudes toward risks can be characterized by a distortion applied to distribution functions $F_X(x)$, instead of characterizing attitude towards risk by a utility function of wealth in the expected utility theory. In a specific way, Wang (1996) applied a distortion function $g(t)$ on decumulative distribution functions $\bar{F}_X(x) := 1 - F_X(x)$ and defined a class of premium principles $H_g[X] = \int_0^{+\infty} g(\bar{F}_X(x))dx$ for some non-decreasing distortion function g such that $g(0) = 0$ and $g(1) = 1$.

Apart from those stated above, there are still some premium principles that cannot be described by the expected utility theory or distorted expectation theory. For example, Denuit et al. (2006) showed that expected shortfall, which is also called conditional value at risk (CVaR) or average value at risk (AVaR), does not belong to the two previously discussed categories. Goovaerts et al. (2003) derived many existing premium principles by minimizing a Markov bound for the tail probability and proposed a unified approach to generating premium principles. Furman and Zitikis (2008) suggested and investigated another class of premium principles called weighted premium principles, which are based on weighted loss distributions and discussed later in this chapter.

Given the total loss S, let $\pi[S_k \mid S]$ denote the capital allocated to the k^{th} line of business. As defined in advance, the total loss $S = \sum_{k=1}^{d} S_k$, and denote by $Y = (S_1, S_2, \dots, S_d)'$ the set of all portfolios of the firm. For a chosen pcp π, $\pi[S]$ is called the risk capital of the firm and an allocation principle is crucial to allocating the amount of risk capital $\pi[S]$ among the d portfolios of the company. An allocation principle is a function ρ that maps Y into a unique allocation:

$$\rho : Y \to \begin{bmatrix} \rho_1[Y] \\ \rho_2[Y] \\ \vdots \\ \rho_d[Y] \end{bmatrix} = \begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_d \end{bmatrix} \tag{4.1}$$

such that $\sum_{k=1}^{d} K_i = \pi[S]$.

Early in Bühlmann (1980) and Bühlmann (1984) , the widely used Esscher premium calculation principle nowadays was verified as a risk allocation principle and the allocation rule is stated as follows:

$$\rho_E[S_k, S] = \frac{\mathbf{E}[S_k e^{\tau S}]}{\mathbf{E}[e^{\tau S}]}. \tag{4.2}$$

Later, Wang (2002) decomposed Esscher allocation principle in (4.2) as

$$\rho_E[S_k, S] = \mathbf{E}[S_k] + \frac{\mathbf{Cov}[S_k, e^{\tau S}]}{\mathbf{E}[e^{\tau S}]} \tag{4.3}$$

and suggested allocating the cost of capital based on $\frac{\mathbf{Cov}[S_k, e^{\tau S}]}{\mathbf{E}[e^{\tau S}]}$, which is illustrated by an example assuming the multivariate normal distribution among $S_1, S_2, \ldots, S_d$.

Furman and Zitikis (2009) introduced a weighted allocation principle based on the weighted premium principle described in Furman and Zitikis (2008), which is a general class of allocation principles. Let $\mathcal{X}$ denote the set of positive risk random variables with cumulative distribution function, the weighted allocation principle is defined as follows.

Definition Let $X, Y \in \mathcal{X}$,and let $w : [0, \infty) \to [0, \infty)$ be a deterministic Borel function. Then the functional $\rho_w : \mathcal{X} \times \mathcal{X} \to [0, \infty]$, defined by the equation

$$\rho_w[X, Y] = \frac{\mathbf{E}[Xw(Y)]}{\mathbf{E}[w(Y)]}, \tag{4.4}$$

is called the weighted allocation.

A number of allocation principles coincide with the weighted allocation principle. For example, we are able to obtain Esscher's allocation principle, the TCE allocation principle, the Tail covariance allocation principle the and distorted allocation by letting $w(y) = e^{\tau y}$, $w(y) = 1\{y \geq y_q\}$, $w(y) = y1\{y \geq y_q\}$, $w(y) = g'(\bar{F}_Y(Y))$, respectively.

The remainder of this chapter is organized as follows. In Section 4.2, we clarify our principle of pricing the premiums in multiple business lines. In Section 4.3, we begin our pricing work with a special case where there is only a single business line. We reduce the model to several simple classic models under different assumptions. In Section 4.4, we develop a novel methodology of premium setting and numerically illustrate how model parameters influence the premium level. In Section 4.5, we discuss two non-parametric methods to estimate the premiums empirically, which are based on whether we are able to distinguish the data under different background risk levels or not. We also compare their performance of estimation. In Section 4.6, given that our assumptions for the risk model are followed, we provide a methodology to fit the parameter value and make assessment. Section 4.7 concludes this chapter.

4.2 Pricing

There are many pcp's and capital allocation rules. In this chapter we work with the weighted pcp (Furman and Zitikis (2008)), which is defined by the equation

$$\pi_w[S] = \frac{\mathbf{E}[S\,w(S)]}{\mathbf{E}[w(S)]} \tag{4.5}$$

and encompasses a great variety of pcp's depending on the choice of the weight function $w : [0, \infty) \to [0, \infty)$, which is usually non-decreasing. In the literature, popular examples of the weight function are $w(s) = s^\tau$, $w(s) = e^{\tau s}$, $w(s) = 1 - e^{-\tau s}$ and $w(s) = \mathbf{1}\{s > \tau\}$, where $\tau > 0$ is a parameter. Under these choices, the functional π_w reduces to the size-biased, Esscher, Kamps, and excess-of-loss pcp's, respectively (e.g., Furman and Zitikis (2009) and references therein). Given S, the corresponding capital allocated to the k^{th} line of business is

$$\pi_w[S_k \mid S] = \frac{\mathbf{E}[S_k w(S)]}{\mathbf{E}[w(S)]} \tag{4.6}$$

Holding capital is costly. Denote the opportunity cost of capital for the k^{th} line of business by r_k, which is the rate of return required by the capital providers (e.g., stock holders). Hence, the total cost of capital for the k^{th} line is

$$\begin{aligned} \mathrm{CC}_k &= r_k(\pi_w[S_k \mid S] - \mathbf{E}[S_k]) \\ &= r_k \frac{\mathbf{Cov}[S_k, w(S)]}{\mathbf{E}[w(S)]} \\ &= \beta_{k,w} r_k(\pi_w[S] - \mathbf{E}[S]) \end{aligned} \tag{4.7}$$

with the 'weighted beta' (Furman and Zitikis (2017))

$$\beta_{k,w} = \frac{\mathbf{Cov}[S_k, w(S)]}{\mathbf{Cov}[S, w(S)]}.$$

(Note that when $w(s) = x$, the weighted beta reduces to the classical beta between S_k and S.) For computational purposes, the following equations are particularly useful:

$$\begin{aligned} \beta_{k,w} &= \frac{\pi_w[S_k \mid S] - \mathbf{E}[S_k]}{\pi_w[S] - \mathbf{E}[S]} \\ &= \frac{\pi_w[S_k \mid S] - \mathbf{E}[S_k]}{\sum_{k=1}^{d} (\pi_w[S_k \mid S] - \mathbf{E}[S_k])}. \end{aligned} \tag{4.8}$$

The insurer, naturally, wishes to allocate the cost of claims and the cost of holding capital to individual policy holders, and thus, in the k^{th} line of business, charges a premium P_k per policy. The premium needs to be large enough, say such that the collected amount of money $N_k \times P_k$ would cover the amount $\mathbf{E}[S_k] + \mathrm{CC}_k$ with a sufficiently large probability. Actually, this should happen simultaneously for all business lines, and so, in summary, we wish to specify those prices P_k, $1 \le k \le d$, for which the bound

$$\mathbf{P}[N_k \times P_k \ge \mathbf{E}[S_k] + \mathrm{CC}_k,\ 1 \le k \le d] \ge 1 - \theta \tag{4.9}$$

holds with a pre-specified (small) $\theta \in (0, 1)$. Obviously, there are many prices that satisfy requirement (4.9), but business considerations would suggest that some are more attractive than others. Denote the set of all d-dimensional vectors $(P_1, \ldots, P_d)$ by $\mathcal{P}_\alpha(d)$ and call it the set of admissible prices. We wish to describe this set.

Conditionally on Z, the frequencies $N_1, N_2, \ldots, N_d$ are independent. Hence, the probability in criterion (4.9) becomes

$$\begin{aligned} &\mathbf{P}[N_k \times P_k \ge \mathbf{E}[S_k] + \mathrm{CC}_k,\ 1 \le k \le d] \\ &\qquad = \mathbf{E}[\mathbf{P}[N_k \times P_k \ge \mathbf{E}[S_k] + \mathrm{CC}_k,\ 1 \le k \le d \mid Z]] \\ &\qquad = \int \prod_{k=1}^{d} \mathbf{P}[N_k \times P_k \ge \mathbf{E}[S_k] + \mathrm{CC}_k \mid Z = z]\, \mathrm{d}F_Z(z) \\ &\qquad = \int \prod_{k=1}^{d} H_{k,z}\Big(\frac{\mathbf{E}[S_k] + \mathrm{CC}_k}{P_k}\Big)\, \mathrm{d}F_Z(z), \end{aligned} \tag{4.10}$$

where, for every $y \geq 0$,

$$H_{k,z}(y) := \mathbf{P}[N_k \geq y \mid Z = z].$$

Hence, we wish to determine the set $\mathcal{P}_\theta(d)$ of all $(P_1, \ldots, P_d)$ that satisfy the bound

$$\int \prod_{k=1}^{d} H_{k,z}\Big(\frac{\mathbf{E}[S_k] + \mathrm{CC}_k}{P_k}\Big) \mathrm{d}F_Z(z) \geq 1 - \theta \tag{4.11}$$

with a pre-specified $\theta \in (0, 1)$, such as $\theta = 0.05$, or even smaller.

4.3 Pricing a single business line

When $d = 1$, our model simply reduces to a single collective risk model

$$S = S_1 = \sum_{i=1}^{N_1} X_{1,i}, \tag{4.12}$$

where $X_{1,i}, i = 1, 2, \ldots, N_1$ are identically distributed and have the same distributions. The set $\mathcal{P}_\theta(1)$ consists of all the premiums per policy P_1 that satisfy the bound

$$P_1 \geq \frac{1}{n_{1,\theta}}(\mathbf{E}[S_1] + \mathrm{CC}_1) \tag{4.13}$$

with

$$n_{1,\theta} = \sup\{n \ : \ \mathbf{P}[N_1 \geq n] \geq 1 - \theta\}.$$

Consider a few special cases that have been particularly discussed in the literature. We start with the simplest case when the frequencies and the background are fixed.

4.3.1 Fixed N_1 and Z

Let $N_1 = n_1^*$ and $Z = z^*$ for some n_1^* and z^*. In this case, irrespective of the value of $\theta \in (0, 1)$, we have $n_{1,\theta} = n_1^*$, and we also have $\mathbf{E}[S_1] = n_1^* \mathbf{E}[X_{1,1}]$ and $\pi_w[S_1] = n_1^* \pi_w[X_{1,1} \mid S_1]$. Hence, the premium P_1 satisfies criterion (4.13) if and only if

$$P_1 \geq \mathbf{E}[X_{1,1}] + r_1(\pi_w[X_{1,1} \mid S_1] - \mathbf{E}[X_{1,1}]). \tag{4.14}$$

Note that bound (4.14) can be rewritten as

$$\begin{aligned} P_1 &\geq \mathbf{E}[X_{1,1}] + r_1 \frac{\mathbf{Cov}[X_{1,1}, w(S_1)]}{\mathbf{E}[w(S_1)]} \\ &= \mathbf{E}[X_{1,1}] + r_1 \beta_w (\pi_w[S_1] - \mathbf{E}[S_1]) \end{aligned} \tag{4.15}$$

with the weighted beta

$$\begin{aligned} \beta_w &= \frac{\pi_w[X_{1,1} \mid S_1] - \mathbf{E}[X_{1,1}]}{\pi_w[S_1] - \mathbf{E}[S_1]} \\ &= \frac{1}{n_1^*}. \end{aligned} \tag{4.16}$$

Example 4.3.1. *We apply exponential tilting $w(s) = e^{\tau s}$. Let the severity $X_{1,1}$ follow a Gamma $Ga(\alpha_1,\beta_1)$ distribution with mean $\mu_1 = \alpha_1/\beta_1$ and variance $\sigma_1^2 = \alpha_1/\beta_1^2$. Hence, $\mathbf{E}[S_1] = n^*\alpha_1/\beta_1$. To calculate $\pi_w[S_1]$, we start with the equation*

$$\mathbf{E}[S_1 e^{\tau S_1}] = \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{E}[e^{\tau S_1}] = \frac{\mathrm{d}}{\mathrm{d}\tau}(M_{X_{1,1}}(\tau))^{n_1^*} = \frac{\alpha_1 n_1^*}{\beta_1}\left(1 - \frac{\tau}{\beta_1}\right)^{-\alpha_1 n_1^* - 1} \tag{4.17}$$

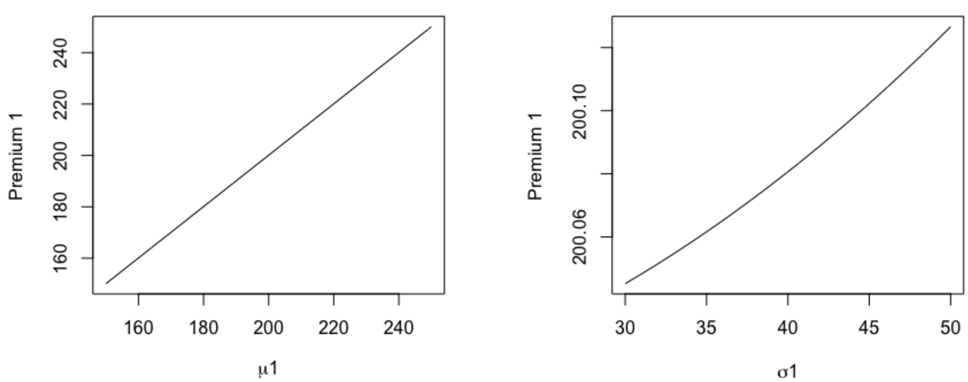
for all $\tau < \beta$. Hence

$$\pi_w[S_1] = \frac{\alpha_1 n_1^*}{\beta_1}\left(1 - \frac{\tau}{\beta_1}\right)^{-1}. \tag{4.18}$$

In view of the above, bound (4.14) becomes equivalent to

$$\begin{aligned} P_1 &\geq \frac{\alpha_1}{\beta_1}\left[1 + r_1\left\{\left(1 - \frac{\tau}{\beta_1}\right)^{-1} - 1\right\}\right] \\ &= \frac{\alpha_1}{\beta_1}\left(1 + r_1\frac{\tau}{\beta_1 - \tau}\right) \\ &= \mu_1\left(1 + r_1\frac{\tau\sigma_1^2}{\mu_1 - \tau\sigma_1^2}\right) \end{aligned} \tag{4.19}$$

for $\tau < \beta_1$. Figure 4.1 depicts how P_1 changes with respect to the mean and standard deviations of the loss severities.



(a) mean of severity (b) sd of severity

Figure 4.1: Fixed N_1 and Z in a single business line

4.3.2 Random N_1 and fixed Z

In this scenario, we let N_1 be any non-negative integer-valued random variable, but we still assume that the background does not change, that is, $Z = z^*$ for some z^*. Hence, the premium

P_1 satisfies criterion (4.13) if and only if

$$P_1 \geq \frac{1}{n_{1,\theta}}\Big(\mathbf{E}[S_1] + r_1(\pi_w[S_1] - \mathbf{E}[S_1])\Big)$$
$$= \frac{\mathbf{E}[N_1]}{n_{1,\theta}}\mathbf{E}[X_{1,1}] + \frac{r_1}{n_{1,\theta}}(\pi_w[S_1] - \mathbf{E}[S_1]) \tag{4.20}$$

Obviously, when $N_1 = n_1^*$, we have $n_{1,\theta} = n_1^*$, and thus inequality (4.20) reduces to (4.15).

Example 4.3.2. *We keep the assumptions in Example 4.3.1. However, we now assume that the frequency N_1 is random and follows a Binomial distribution $B(n_1, p_1)$ with positive integer n_1 and parameter $p_1 > 0$, which implies that $\mathbf{E}[N_1] = n_1 p_1$. Hence, $\mathbf{E}[S_1] = \mathbf{E}[N_1]\mathbf{E}[X_{1,1}] = n_1 p_1 \alpha_1/\beta_1$. To calculate $\pi_w[S_1]$, we start with the equation*

$$\mathbf{E}[S_1 e^{\tau S_1}] = \frac{\mathrm{d}}{\mathrm{d}\tau} P_{N_1}(M_{X_{1,1}}(\tau)). \tag{4.21}$$

Under the above specified distributions of $X_{1,1}$ and N_1, the moment generating function is $M_{X_{1,1}}(\tau) = (1 - \tau/\beta_1)^{-\alpha_1}$ for all $\tau < \beta_1$, and the probability generating function is $P_{N_1}(x) = [1 - p_1 + p_1 x]^{n_1}$. Consequently, equation (4.21) implies

$$\mathbf{E}[S_1 e^{\tau S_1}] = \frac{n_1 p_1 \alpha_1}{\beta_1}\Big(1 - \frac{\tau}{\beta_1}\Big)^{-\alpha_1 - 1}\Big[1 - p_1 + p_1\Big(1 - \frac{\tau}{\beta_1}\Big)^{-\alpha_1}\Big]^{n_1 - 1} \tag{4.22}$$

for all $\tau < \beta_1$, and hence

$$\pi_w[S_1] = \frac{n_1 p_1 \alpha_1}{\beta_1}\Big(1 - \frac{\tau}{\beta_1}\Big)^{-\alpha_1 - 1}\Big[1 - p_1 + p_1\Big(1 - \frac{\tau}{\beta_1}\Big)^{-\alpha_1}\Big]^{-1}. \tag{4.23}$$

In view of the above, bound (4.20) becomes equivalent to

$$P_1 \geq \frac{n_1 p_1 \alpha_1}{n_{1,\theta}\beta_1}\Big(1 + r_1\Big\{\Big(1 - \frac{\tau}{\beta_1}\Big)^{-\alpha_1 - 1}\Big[1 - p_1 + p_1\Big(1 - \frac{\tau}{\beta_1}\Big)^{-\alpha_1}\Big]^{-1} - 1\Big\}\Big)$$
$$= \frac{n_1 p_1 \mu_1}{n_{1,\theta}}\Big(1 + r_1\Big\{\Big(1 - \frac{\tau\sigma_1^2}{\mu_1}\Big)^{-\mu_1^2/\sigma_1^2 - 1}\Big[1 - p_1 + p_1\Big(1 - \frac{\tau\sigma_1^2}{\mu_1}\Big)^{-\mu_1^2/\sigma_1^2}\Big]^{-1} - 1\Big\}\Big) \tag{4.24}$$

whenever $\tau < \beta$. Figure 4.2 depicts how P_1 changes with respect to parameters.

4.3.3 Random N_1 and Z

In this scenario, we assume that both N_1 and Z are random. Consequently, instead of using criterion (4.13), we now have to go back to the original criterion (4.11), which under the specification $d = 1$ becomes

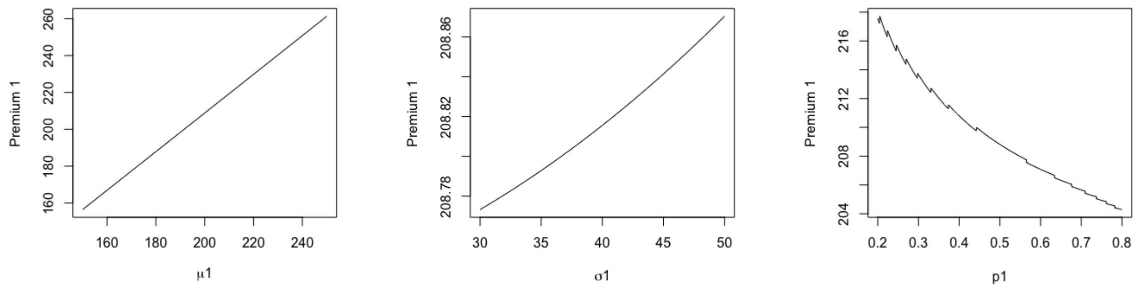
$$\int H_{1,z}\Big(\frac{\mathbf{E}[S_1] + \mathrm{CC}_1}{P_1}\Big)\mathrm{d}F_Z(z) \geq 1 - \theta. \tag{4.25}$$

The premium P_1 satisfies criterion (4.25) if and only if

$$\int H_{1,z}\Big(\frac{1}{P_1}\Big(\mathbf{E}[S_1] + r_1(\pi_w[S_1] - \mathbf{E}[S_1])\Big)\Big)\mathrm{d}F_Z(z) \geq 1 - \theta. \tag{4.26}$$

Note that bound (4.26) can be rewritten as

$$\int H_{1,z}\Big(\frac{1}{P_1}\Big(\mathbf{E}[S_1] + r_1\frac{\mathbf{Cov}[S_1, w(S_1)]}{\mathbf{E}[w(S_1)]}\Big)\Big)\mathrm{d}F_Z(z) \geq 1 - \theta. \tag{4.27}$$



(a) mean of severity (b) sd of severity (c) prob parameter in Bino distr

Figure 4.2: Random N_1 and fixed Z in a single business line

Example 4.3.3. *As in the previous example, we work with the exponential tilting $w(s) = e^{\tau s}$. We assume that the conditional severities $[X_{k,i} \mid Z]$ follow a Gamma distribution $Ga(\alpha_{1,Z}, \beta_{1,Z})$ with parameters $\alpha_{1,Z} > 0$ and $\beta_{1,Z} > 0$. Furthermore, we assume that the conditional frequencies follow a Binomial distribution $B(n_1, p_{1,Z})$ with positive integer n_1 parameter $p_{1,Z} = p_1 Z > 0$. Note that when $Z = z^*$ almost surely for a constant z^*, the current scenario reduces to that considered in Section 4.3.2. In the current section, we consider the more general scenario by assuming that there are constants z^* and z^{**} such that $Z = z^* + (z^{**} - z^*)B(1, p)$, where $B(1, p)$ represents the Bernoulli distribution with parameters $p > 0$. Consequently,*

$$H_{1,z}(y) \triangleq \mathbf{P}[N_1 \geq y \mid Z = z] = \sum_{m=\lceil y \rceil}^{n_1} \binom{n_1}{m} p_{1,z}^m (1 - p_{1,z})^{n_1 - m},$$

where $\lceil y \rceil$ is the ceiling of y. With the above assumptions, we have

$$\begin{aligned} \mathbf{E}[S_1] &= \mathbf{E}[\mathbf{E}[N_1 \mid Z]\mathbf{E}[X_{1,1} \mid Z]] \\ &= \frac{n_1 p_{1,z^*} \alpha_{1,z^*}}{\beta_{1,z^*}}(1 - p) + \frac{n_1 p_{1,z^{**}} \alpha_{1,z^{**}}}{\beta_{1,z^{**}}} p. \end{aligned} \tag{4.28}$$

The calculation of

$$\pi_w[S_1] = \frac{\mathbf{E}[S_1 e^{\tau S_1}]}{\mathbf{E}[e^{\tau S_1}]}$$

is more complex. First, since $\mathbf{E}[e^{\tau S_1} \mid Z] = P_{N_1|Z}(M_{X_{1,1}|Z}(\tau))$, we have

$$\mathbf{E}[e^{\tau S_1}] = \int P_{N_1|Z=z}(M_{X_{1,1}|Z=z}(\tau)) \mathrm{d}F_Z(z), \tag{4.29}$$

where $P_{N_1|Z}(x) = \mathbf{E}[x^{N_1} \mid Z]$ and $M_{X_{1,1}|Z}(\tau) = \mathbf{E}[e^{\tau X_{1,1}} \mid Z]$ are the (conditional) probability and moment generating functions, respectively. Under the above distributional assumptions, we have $M_{X_{1,1}|Z}(\tau) = (1 - \tau/\beta_{1,Z})^{-\alpha_{1,Z}}$ and $P_{N_1|Z}(x) = (1 - p_{1,Z} + p_{1,Z} x)^{n_1}$, and thus from equation (4.29) we have

$$\begin{aligned} \mathbf{E}[e^{\tau S_1}] = &\left[1 - p_{1,z^*} + p_{1,z^*}\left(1 - \frac{\tau}{\beta_{1,z^*}}\right)^{-\alpha_{1,z^*}}\right]^{n_1}(1 - p) \\ &+ \left[1 - p_{1,z^{**}} + p_{1,z^{**}}\left(1 - \frac{\tau}{\beta_{1,z^{**}}}\right)^{-\alpha_{1,z^{**}}}\right]^{n_1} p. \end{aligned} \tag{4.30}$$

Furthermore, we have

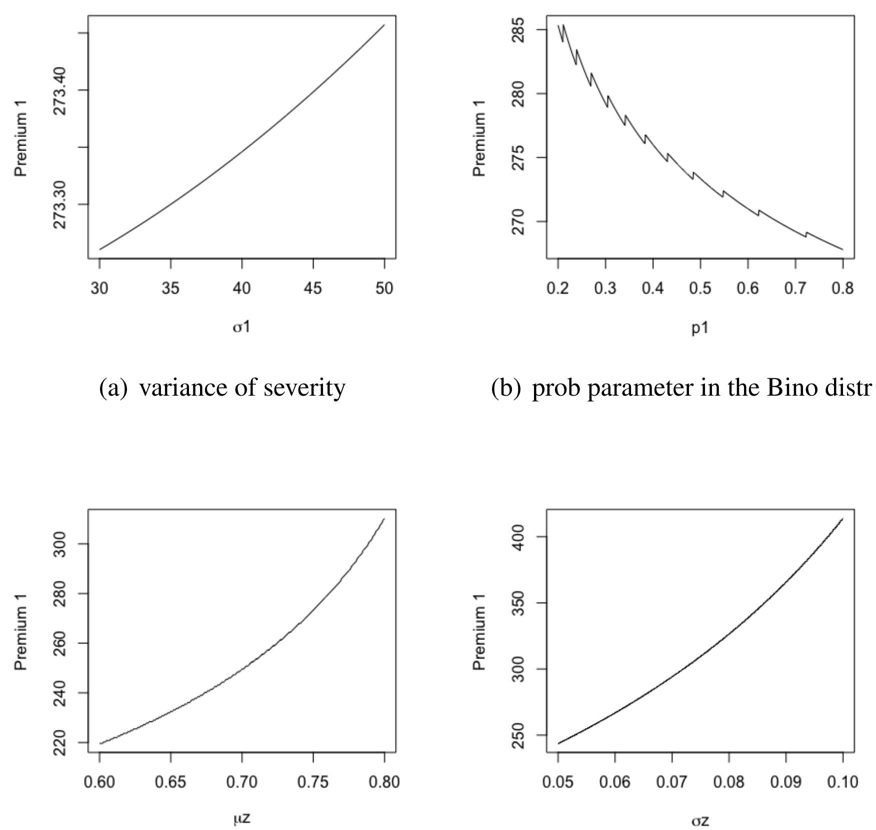
$$\mathbf{E}[S_1 e^{\tau S_1} \mid Z] = \frac{\mathrm{d}}{\mathrm{d}\tau} P_{N_1|Z}(M_{X_{1,1}|Z}(\tau)), \tag{4.31}$$

and so, under the above distributional assumptions, we have

$$\mathbf{E}[S_1 e^{\tau S_1}] = \mathbf{E}[S_1 e^{\tau S_1} \mid Z = z^*](1-p) + \mathbf{E}[S_1 e^{\tau S_1} \mid Z = z^{**}]p \tag{4.32}$$

with

$$\mathbf{E}[S_1 e^{\tau S_1} \mid Z = z] = \frac{n_1 p_{1,z} \alpha_{1,z}}{\beta_{1,z}} \Big(1 - \frac{\tau}{\beta_{1,z}}\Big)^{-\alpha_{1,z}-1} \Big[1 - p_{1,z} + p_{1,z}\Big(1 - \frac{\tau}{\beta_{1,z}}\Big)^{-\alpha_{1,z}}\Big]^{n_1-1}. \tag{4.33}$$



(a) variance of severity (b) prob parameter in the Bino distr

(c) mean of background risk (d) sd of background risk

Figure 4.3: General consideration in a single business line

4.4 Pricing multiple business lines

4.4.1 Parametric setting

In this section we illustrate how the above developed pricing technique can work in practical calculations. We make the following distributional assumptions:

- Assume that conditional on the background risk Z, the loss severities $[X_{k,i} \mid Z]$ follow gamma distributions with parameters $\alpha_{k,Z} > 0$ and $\beta_{k,Z} > 0$. In addition, assume that the conditional mean and variance of the average severity are

$$\mathbf{E}[X_{k,1} \mid Z] = c_k Z^{\kappa_1} \quad \text{and} \quad \mathbf{Var}[X_{k,1} \mid Z] = \sigma_k^2 Z^{\kappa_2}$$

  for some positive constants $c_k > 0$ and $\sigma_k^2 > 0$. Under this specification, the gamma parameters are

$$\alpha_{k,Z} = \frac{c_k^2}{\sigma_k^2} Z^{2\kappa_1 - \kappa_2} \quad \text{and} \quad \beta_{k,Z} = \frac{c_k}{\sigma_k^2} Z^{\kappa_1 - \kappa_2}$$

- Assume that the conditional frequencies $[N_k \mid Z]$ follow Binomial distributions with parameters n_k and $p_{k,Z} > 0$. In particular, when $p_{k,Z} = p_k Z$ with $p_k > 0$, the mean and variance of the Binomial distribution are $\mathbf{E}[N_k \mid Z] = n_k p_k Z$ and $\mathbf{Var}[N_k] = n_k p_k Z(1 - p_k Z)$ respectively.

- As to the background risk, we consider two cases:
  1) $Z = z^*$ almost surely for a constant z^*
  2) For some constants z^* and z^{**} such that $Z = z^* + (z^{**} - z^*)\mathrm{B}(1, p)$ where $\mathrm{B}(1, p)$ represents the Bernoulli distribution with probability $p > 0$.

4.4.2 The case of $w(x) = x$

We have

$$\mathbf{E}[S] = \int \sum_{k=1}^{d} \mathbf{E}[S_k \mid Z] \mathrm{d}F_Z(z), \tag{4.34}$$

with

$$\mathbf{E}[S_k \mid Z] = \mathbf{E}[N_k \mid Z]\mathbf{E}[X_{k,1} \mid Z] \quad \text{for} \quad k = 1, \ldots, d \tag{4.35}$$

Equations (4.34)–(4.35) give the following formulas:

- If $Z = z^*$ almost surely, then

$$\mathbf{E}[S] = \sum_{k=1}^{d} \frac{n_k p_{k,z^*} \alpha_{k,z^*}}{\beta_{k,z^*}}; \tag{4.36}$$

- If $Z = z^* + (z^{**} - z^*)\mathrm{B}(1, p)$, then

$$\mathbf{E}[S] = (1 - p)\mathbf{E}[S \mid Z = z^*] + p\mathbf{E}[S \mid Z = z^{**}]. \tag{4.37}$$

In addition, we have

$$\begin{aligned} \mathbf{E}[S_k S \mid Z] &= \mathbf{E}[S_k^2 \mid Z] + \sum_{l \neq k} \mathbf{E}[S_k \mid Z]\mathbf{E}[S_l \mid Z] \\ &= \mathbf{Var}[S_k \mid Z] + \sum_{l=1}^{d} \mathbf{E}[S_k \mid Z]\mathbf{E}[S_l \mid Z], \end{aligned} \tag{4.38}$$

with

$$\mathbf{Var}[S_k \mid Z] = \mathbf{E}[N_k \mid Z]\mathbf{Var}[X_{k,1} \mid Z] + \mathbf{Var}[N_k \mid Z]\mathbf{E}[X_{k,1} \mid Z]^2. \tag{4.39}$$

Under the above specifications of the distributions of $[X_{k,1} \mid Z]$ and $[N_k \mid Z]$, from equations (4.38) and (4.39), we have

$$\mathbf{E}[S_k S \mid Z] = \frac{n_k p_{k,Z}\alpha_{k,Z} + n_k p_{k,Z}(1-p_{k,Z})\alpha_{k,Z}^2}{\beta_{k,Z}^2} + \sum_{l=1}^{d} \frac{n_l p_{l,Z}\alpha_{l,Z}}{\beta_{l,Z}}. \tag{4.40}$$

4.4.3 The case of $w(x) = e^{\tau x}$

When $w(x) = e^{\tau x}$, we have

$$\mathbf{E}[e^{\tau S}] = \int \prod_{k=1}^{d} P_{N_k|Z=z}(M_{X_{k,1}|Z=z}(\tau))\mathrm{d}F_Z(z), \tag{4.41}$$

because $\mathbf{E}[e^{\tau S} \mid Z]$ is the product of $\mathbf{E}[e^{\tau S_k} \mid Z]$, $k = 1, \ldots, d$, and

$$\mathbf{E}[e^{\tau S_k} \mid Z] = P_{N_k|Z}(M_{X_{k,1}|Z}(\tau)), \tag{4.42}$$

where $P_{N_k|Z}(x) = \mathbf{E}[x^{N_k} \mid Z]$ and $M_{X_{k,1}|Z}(\tau) = \mathbf{E}[e^{\tau X_{k,1}} \mid Z]$ are the (conditional) probability and moment generating functions, respectively. Under the Gamma assumption for the conditional severities, we have

$$M_{X_{k,1}|Z}(\tau) = \left(1 - \frac{\tau}{\beta_{k,Z}}\right)^{-\alpha_{k,Z}} \quad \text{for} \quad \tau < \beta_{k,Z}. \tag{4.43}$$

Under the Binomial assumption for the conditional frequencies, we have

$$P_{N_k|Z}(x) = (1 - p_{k,Z} + p_{k,Z}x)^{n_k}. \tag{4.44}$$

Equations (4.41)–(4.44) give the following formulas:

- If $Z = z^*$ almost surely, then

$$\mathbf{E}[e^{\tau S}] = \prod_{k=1}^{d} \left[1 - p_{k,z^*} + p_{k,z^*}\left(1 - \frac{\tau}{\beta_{k,z^*}}\right)^{-\alpha_{k,z^*}}\right]^{n_k} \tag{4.45}$$

  for all $\tau > 0$ such that

$$\tau < \min_k\{\beta_{k,z^*}\}. \tag{4.46}$$

- If $Z = z^* + (z^{**} - z^*)\mathrm{B}(1, p)$, then

$$\mathbf{E}[e^{\tau S}] = (1-p)\mathbf{E}[e^{\tau S} \mid Z = z^*] + p\mathbf{E}[e^{\tau S} \mid Z = z^{**}] \tag{4.47}$$

  for all $\tau > 0$ such that condition (4.46) is satisfied.

In addition, we have

$$\begin{aligned}\mathbf{E}[S_k e^{\tau S} \mid Z] &= \mathbf{E}[S_k e^{\tau S_k} \mid Z] \prod_{l \neq k} \mathbf{E}[e^{\tau S_l} \mid Z] \\ &= \mathbf{E}[S_k e^{\tau S_k} \mid Z] \prod_{l \neq k} P_{N_l|Z}(M_{X_{l,1}|Z}(\tau)), \end{aligned} \tag{4.48}$$

with the product on the right-hand side calculated in the previous section. For the expectation, we have

$$\mathbf{E}[S_k e^{\tau S_k} \mid Z] = \frac{\mathrm{d}}{\mathrm{d}\tau} P_{N_k|Z}(M_{X_{k,1}|Z}(\tau)). \tag{4.49}$$

Under the above specifications of the distributions of $X_{k,1}$ and N_k given Z, from equations (4.48) and (4.49), we have

$$\begin{aligned}\mathbf{E}[S_k e^{\tau S} \mid Z] = \frac{n_k p_{k,Z} \alpha_{k,Z}}{\beta_{k,Z}} \left(1 - \frac{\tau}{\beta_{k,Z}}\right)^{-\alpha_{k,Z}-1} \left[1 - p_{k,Z} + p_{k,Z}\left(1 - \frac{\tau}{\beta_{k,Z}}\right)^{-\alpha_{k,Z}}\right]^{n_k-1} \\ \times \prod_{l \neq k} \left[1 - p_{l,Z} + p_{l,Z}\left(1 - \frac{\tau}{\beta_{l,Z}}\right)^{-\alpha_{l,Z}}\right]^{n_l} \end{aligned} \tag{4.50}$$

for all $\tau > 0$ that satisfy condition (4.46).

4.4.4 Numerical parametric analysis

Numerical illustrations are given in this example. We assume that the background variable Z follows the distribution $z^* + (z^{**} - z^*)\mathrm{B}(1, p)$. Here, z^* is the current value for the background risk, which might relate to current interest rate and is already known. Then Z is a two-value random variable which equals z^* with probability $1 - p$ and z^{**} with probability p. Hence, we can readily obtain that $\mathbf{E}[Z] = (1 - p)z^* + pz^{**}$ and $\mathbf{Var}[Z] = (z^{**} - z^*)^2 p(1 - p)$, denoted by μ_Z and σ_Z^2, respectively. We can also obtain the distribution of Z from the mean and variance, that is,

$$p = 1/\left(\frac{\mathbf{Var}[Z]}{(\mathbf{E}[Z] - z^*)^2} + 1\right),$$

and

$$z^{**} = \frac{\mathbf{E}[Z] - (1 - p)z^*}{p}.$$

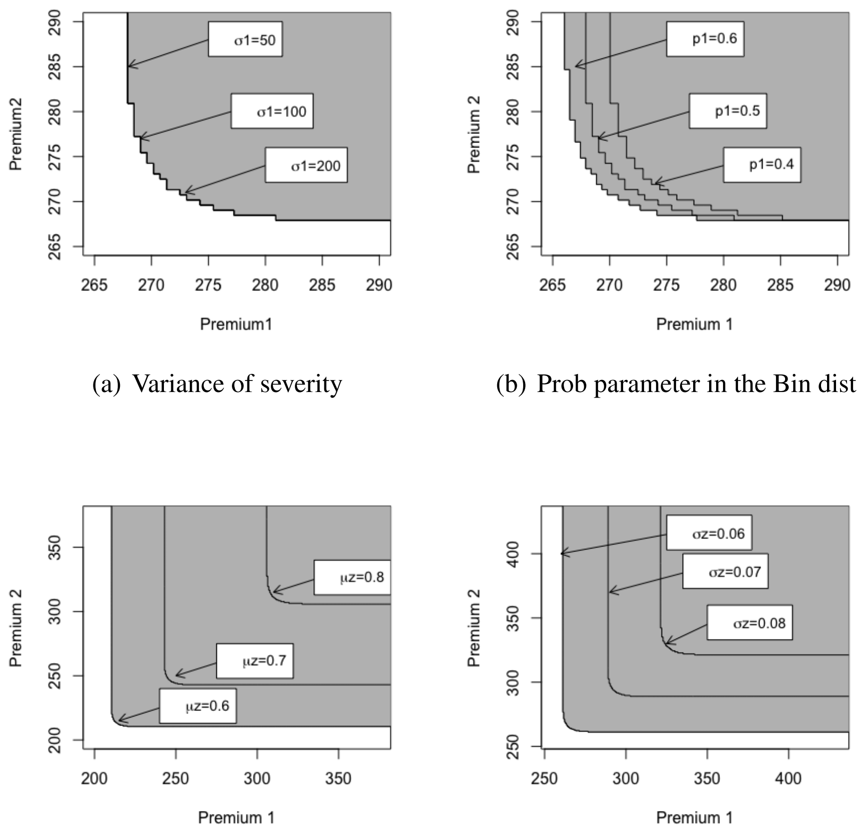
Conditional on Z, $X_{k,i}$ follows a Gamma distribution with mean and variance

$$\mathbf{E}[X_{k,1} \mid Z] = c_k Z \quad \text{and} \quad \mathbf{Var}[X_{k,1} \mid Z] = \sigma_k^2 Z^2$$

for some positive constants $c_k > 0$ and $\sigma_k^2 > 0$.

Then, we explore how these parameters will influence the admissible prices when $d = 2$. Here, we ignore the parameter c_k because an increase of c_k only indicates larger severity mean, which will definitely results in higher premium.

When $w(x) = x$, the variance of the severity has little influence on the premiums. The frequency distribution which is more skewed to the left has higher premium than others. Higher variance of background risk will result in higher premiums. Figure 4.4 depicts the results.



(a) Variance of severity (b) Prob parameter in the Bin dist

(c) Mean of background risk (d) sd of background risk

Figure 4.4: Parametric analysis of the case $w(x) = x$

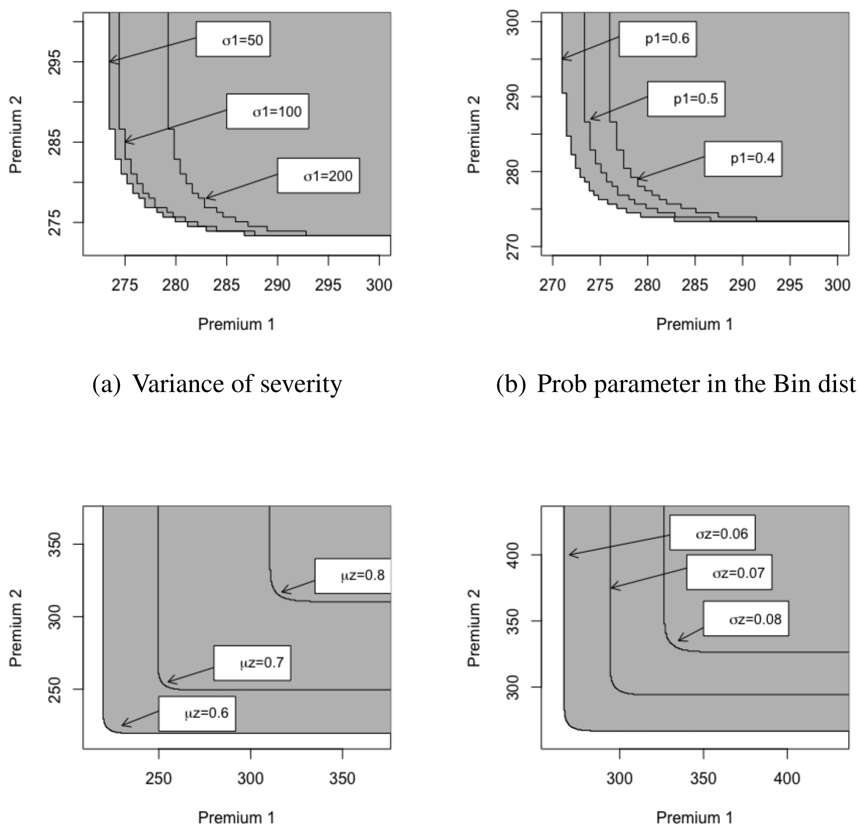
When $w(x) = e^{\tau x}$, the variance of the severity has a more obvious influence on premiums compared to the former case, and higher variance of severity will cause higher premium. However for other parameters, they keep the same rules as in the former case. Figure 4.5 depicts the results.

4.5 Non-parametric pricing methods

Firstly, we consider the situation when the parameters of the model are unknown. Suppose that we observe each of the d lines of businesses for n years. For year j, the k^{th} line has policy number $n_{k,j}$ and the claim severities are $x_{k,1}, \ldots, x_{k,j}$, so that $s_{k,j} = \sum_{i=1}^{n_{k,j}} x_{k,i}$, which represents the total claim amount in k^{th} business line, $s_{t,j} = \sum_{k=1}^{d} s_{k,j}$, representing the total claim amount in all business lines for year j. We need to find $P_1, P_2, \ldots, P_d$ such that

$$\mathbf{P}[N_k \times P_k \geq \mathbf{E}[S_k] + \mathrm{CC}_k,\ 1 \leq k \leq d] \geq 1 - \theta. \tag{4.51}$$

For convenience, we now consider the case of two business lines. The mean losses $\mathbf{E}[S_k], k =$



(a) Variance of severity

(b) Prob parameter in the Bin dist

(c) Mean of background risk

(d) sd of background risk

Figure 4.5: Parametric analysis of the case $w(x) = e^{\tau x}$

$1, 2$, can be estimated empirically by

$$\overline{S}_k = \frac{1}{n} \sum_{j=1}^{n} s_{k,j}, \quad k = 1, 2. \tag{4.52}$$

In addition, $\mathbf{E}[w(S)]$ can be estimated by

$$\overline{w(S)} = \frac{1}{n} \sum_{j=1}^{n} w(s_{t,j}), \tag{4.53}$$

then $\mathrm{CC}_k = r_k \frac{\mathbf{Cov}[S_k, w(S)]}{\mathbf{E}[w(S)]}$, $k = 1, 2$ can be estimated by

$$\overline{\mathrm{CC}}_k = \frac{r_k}{n-1} \sum_{j=1}^{n} \frac{(s_{k,j} - \overline{S}_k)(w(s_{t,j}) - \overline{w(S)})}{\overline{w(S)}}, \quad k = 1, 2. \tag{4.54}$$

We need to find P_1, P_2 such that

$$\frac{1}{n} \sum_{j=1}^{n} \mathbb{I}\left\{ n_{1,j} \geq \frac{\overline{S}_1 + \overline{\mathrm{CC}}_1}{P_1}, n_{2,j} \geq \frac{\overline{S}_2 + \overline{\mathrm{CC}}_2}{P_2} \right\} \geq 1 - \theta, \tag{4.55}$$

which can be rewritten as

$$P_1 \geq \frac{1}{n_{1,*}}(\overline{S}_1 + \overline{\mathrm{CC}}_1), \tag{4.56}$$

$$P_2 \geq \frac{1}{n_{2,*}}(\overline{S}_2 + \overline{\mathrm{CC}}_2), \tag{4.57}$$

where $(n_{1,*}, n_{2,*})$ satisfy

$$\frac{\sum_{j=1}^{n} \mathbb{I}\{n_{1,j} \geq n_1, n_{2,j} \geq n_2\}}{n} \geq 1 - \theta \tag{4.58}$$

and can be obtained empirically.

In summary, the empirical estimations for P_1 and P_2 are

$$\hat{P}_1 = \frac{1}{n_{1,*}}\Big(\frac{1}{n}\sum_{j=1}^{n} s_{1,j} + r_1 * \frac{n}{n-1}\frac{\sum_{i=1}^{n}\big(s_{1,j} - \sum_{j=1}^{n} s_{1,j}/n\big)\big(w(s_{t,j}) - \sum_{j=1}^{n} w(s_{t,j})/n\big)}{\sum_{j=1}^{n} w(s_{t,j})}\Big), \tag{4.59}$$

$$\hat{P}_2 = \frac{1}{n_{2,*}}\Big(\frac{1}{n}\sum_{j=1}^{n} s_{2,j} + r_1 * \frac{n}{n-1}\frac{\sum_{j=1}^{n}\big(s_{2,j} - \sum_{j=1}^{n} s_{2,j}/n\big)\big(w(s_{t,j}) - \sum_{j=1}^{n} w(s_{t,j})/n\big)}{\sum_{j=1}^{n} w(s_{t,j})}\Big). \tag{4.60}$$

When considering d business line in total, we have for $k = 1, \dots, d$,

$$\hat{P}_k = \frac{1}{n_{k,*}}\Big(\frac{1}{n}\sum_{j=1}^{n} s_{k,j} + r_1 * \frac{n}{n-1}\frac{\sum_{j=1}^{n}\big(s_{k,j} - \sum_{j=1}^{n} s_{k,j}/n\big)\big(w(s_{t,j}) - \sum_{j=1}^{n} w(s_{t,j})/n\big)}{\sum_{j=1}^{n} w(s_{t,j})}\Big). \tag{4.61}$$

The key is to determine $(n_{1,*}, \dots, n_{d,*})$ which are obtained from the boundary points of the region in which $(n_1, \dots, n_d)$'s satisfy

$$\frac{\sum_{j=1}^{n} \mathbb{I}\{n_{1,j} \geq n_1, \dots, n_{d,j} \geq n_d\}}{n} \geq 1 - \theta. \tag{4.62}$$

Secondly, we consider the case when the risk level Z can be observed. That is, we are able to classify the data into different risk levels $z_1, z_2, \dots$. Then, the inequality (4.51) becomes

$$\int \prod_{k=1}^{d} \mathbf{P}[N_k \times P_k \geq \mathbf{E}[S_k] + \mathrm{CC}_k \mid Z = z]\mathrm{d}F_Z(z) \geq 1 - \theta. \tag{4.63}$$

The empirical estimates for $\mathbf{E}[S_k]$, $\mathbf{E}[w(S)]$ and $\mathbf{Cov}[S_k, w(S)]$ are kept the same as in the former section, referring to (4.52), (4.53) and (4.54) . Suppose that there are $n_{z_1}, n_{z_2}, \dots, n_{z_r}$ data points under risk level $Z = z_1, Z = z_2, \dots, Z = z_r$ such that $n_{z_1} + n_{z_2} + \cdots + n_{z_r} = n$, the former inequality then becomes

$$\sum_{l=1}^{r}\Big(\prod_{k=1}^{d} \mathbf{P}[N_k \times P_k \geq \mathbf{E}[S_k] + \mathrm{CC}_k \mid Z = z_l]\Big)\frac{n_{z_l}}{n} \geq 1 - \theta. \tag{4.64}$$

Taking the case of $d = 2$ as an example, we need to find P_1, P_2 such that

$$\sum_{l=1}^{r}\Big(\prod_{k=1}^{2} \mathbf{P}\Big[N_k \geq \frac{\mathbf{E}[S_k] + \mathrm{CC}_k}{P_k} \mid Z = z_l\Big]\Big)\frac{n_{z_l}}{n} \geq 1 - \theta. \tag{4.65}$$

This means empirically

$$\sum_{l=1}^{r}\left(\frac{\sum_{j=1}^{n_{z_l}}\mathbb{I}\{n_{1,z_l,j}\geq\frac{\mathbf{E}[S_1]+\mathrm{CC}_1}{P_1}\}}{n_{z_l}}\times\frac{\sum_{j=1}^{n_{z_l}}\mathbb{I}\{n_{2,z_l,j}\geq\frac{\mathbf{E}[S_2]+\mathrm{CC}_2}{P_2}\}}{n_{z_l}}\right)\frac{n_{z_l}}{n}\geq 1-\theta. \tag{4.66}$$

Simplifying the above inequality and replacing terms with empirical estimates we obtain

$$\frac{1}{n}\sum_{l=1}^{r}\frac{1}{n_{z_l}}\left(\sum_{j=1}^{n_{z_l}}1\left\{n_{1,j,z_l}\geq\frac{\overline{S}_1+\overline{\mathrm{CC}}_1}{P_1}\right\}\times\sum_{j=1}^{n_{z_l}}1\left\{n_{2,j,z_l}\geq\frac{\overline{S}_2+\overline{\mathrm{CC}}_2}{P_2}\right\}\right)\geq 1-\theta. \tag{4.67}$$

where n_{1,j,z_l} and n_{2,j,z_l}, $j=1,\ldots,n_{z_l}$ are the claim frequencies of 1^{st} and 2^{nd} business line from the j^{th} data whose risk level is $Z=z_l$.

That is to say, different from the previous section, here we obtain $(n_{1,*},n_{2,*})$ empirically from the boundary points of the region in which (n_1,n_2) satisfy

$$\frac{1}{n}\sum_{l=1}^{r}\frac{1}{n_{z_l}}\left(\sum_{j=1}^{n_{z_l}}1\{n_{1,j,z_l}\geq n_1\}\times\sum_{j=1}^{n_{z_l}}1\{n_{2,j,z_l}\geq n_2\}\right)\geq 1-\theta. \tag{4.68}$$

Theoretically, a straightforward method can be used to compare the performance of two non-parametric quantile boundaries. Since all the data points (n_1,n_2) lie in the lower left side of the quantile boundary satisfy

$$\mathrm{P}(N_1\geq n_1,N_2\geq n_2)\geq 1-\theta, \tag{4.69}$$

we can use the difference between areas constructed by theoretical quantile boundary and non-parametric quantile boundary for evaluation purposes given that we know the exact distributions for severities and frequencies. We illustrate the idea by Figure 4.6. By default, if not

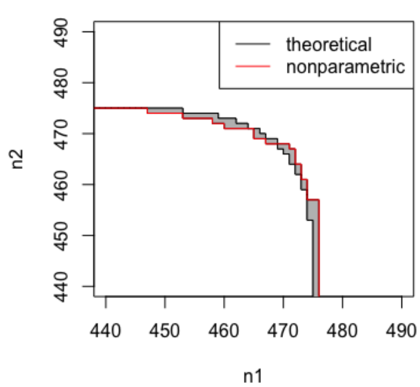
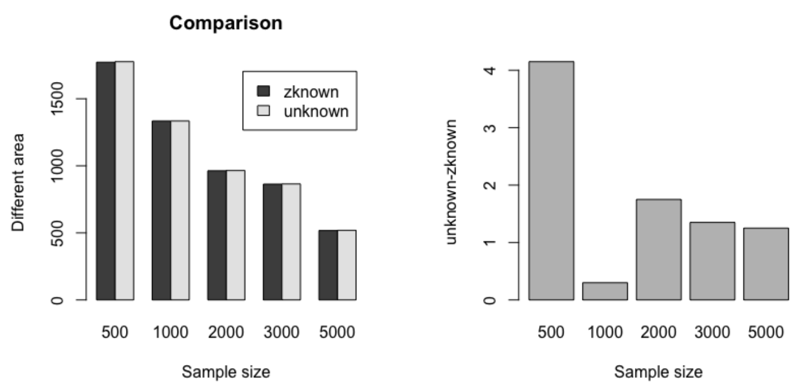


Figure 4.6: Difference of areas between the theoretical and non-parametric boundaries

specified, the following sample datasets being used are generated with parameter values $z^*=1$,

$z^{**} = 0.5$, $p = 0.5$ and $n_k = 2000, p_k = 0.5$ for $k = 1, 2$. We use the size of the shaded region determined by two quantile boundaries to represent the fitting performance compared to theoretical quantile boundary.

Then, we aim to verify that the second non-parametric estimation constructed when the background risk is observable, performs better than the first non-parametric estimation. We simulate 20 times for different sample sizes, and calculate the mean of different areas. The results are reported in Figure 4.7, The time used for non-parametric methods are shown in



(a) Area size of two methods (b) Difference of areas

Figure 4.7: Comparison of areas between non-parametric methods

Table 4.1. In the following table and figures, “z-known" represents the case that we can observe the risk level Z, while “z-unknown" represents the case that we cannot.

sample size	500	1000	2000	3000	5000
z known	0.186	0.179	0.670	0.295	1.019
z unknown	2.788	4.909	9.808	15.224	25.164

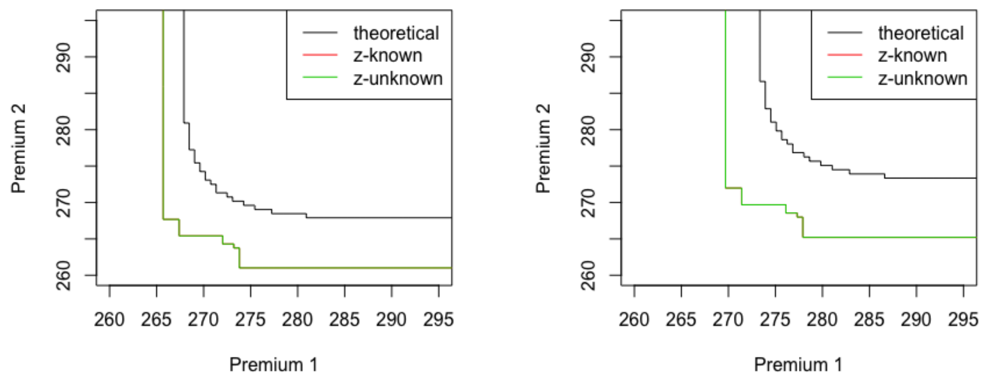
Table 4.1: Comparison of computing time (seconds)

We can see clearly that the second empirical estimation of quantile boundary hugely reduces the computation time, but only performs slightly better in the estimating results. Both non-parametric methods perform better as sample size increases.

The overall performance of the estimated premiums is shown in the remaining part of this subsection. Here we assume that the background variable Z follows the distribution $z^* + (z^{**} - z^*)\mathrm{B}(1, p)$, where $z^* = 1$, $z^{**} = 0.5$, $p = 0.5$. Conditional on Z, $X_{k,i} \sim \mathrm{Ga}(\alpha_{k,Z}, \beta_{k,Z})$. We set average severity and variance as

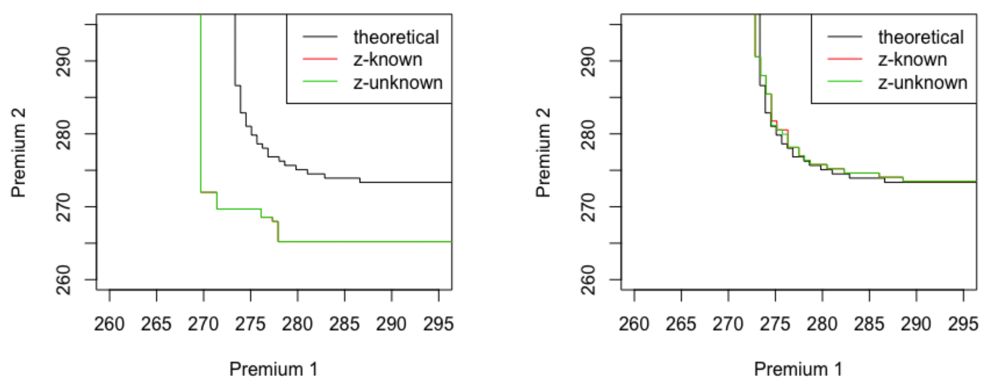
$$\mathbf{E}[X_{k,1} \mid Z] = \frac{\alpha_{k,Z}}{\beta_{k,Z}} = c_k Z \quad \text{and} \quad \mathbf{Var}[X_{k,1} \mid Z] = \frac{\alpha_{k,Z}}{\beta_{k,Z}^2} = \sigma_k^2 Z^2$$

for some positive constants $c_k = 200$ and $\sigma_k = 40$. Also, we assume that the conditional distributions of the loss frequencies $[N_k \mid Z]$ follow the Binomial distribution $\mathrm{B}(n_k, p_{k,Z})$, $k = 1, 2$ with $n_k = 2000$ and $p_k = 0.5$. The results are shown in Figure 4.8, Figure 4.9 and Figure 4.10. Note that the green and red lines are almost overlapping.



(a) $w(x) = x$ (b) $w(x) = e^{\tau x}$

Figure 4.8: Comparison of the theoretical model with empirical methods when $n = 100$



(a) $w(x) = x$ (b) $w(x) = e^{\tau x}$

Figure 4.9: Comparison of the theoretical model with empirical methods when $n = 1000$

4.6 Parametric model fitting method

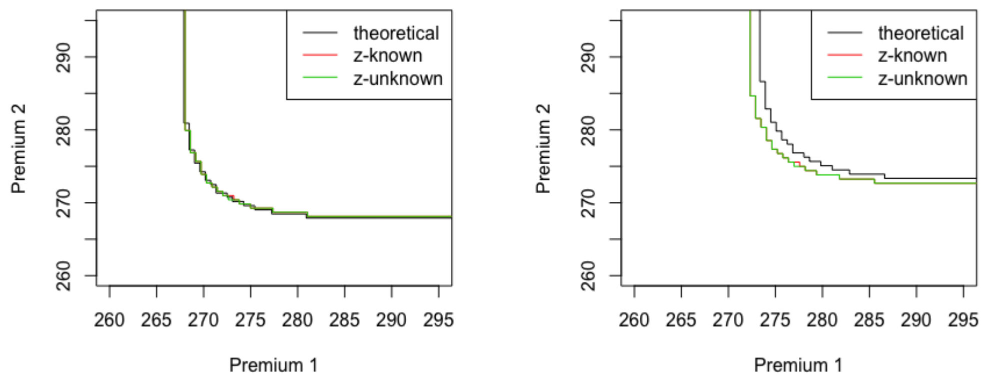
4.6.1 Relationship between the model parameters

According to the law of total variance, we have

$$\mathbf{Cov}[X, Y] = \mathbf{E}[\mathbf{Cov}[X, Y \mid Z]] + \mathbf{Cov}[\mathbf{E}[X \mid Z], \mathbf{E}[Y \mid Z]]. \tag{4.70}$$

Because of our former assumptions, we can theoretically obtain that, for any $i \neq j$, and $i, j \in \{1, \ldots, d\}$,

$$\begin{aligned}\mathbf{Cov}[N_i, N_j] &= \mathbf{E}[\mathbf{Cov}[N_i, N_j \mid Z]] + \mathbf{Cov}[\mathbf{E}[N_i \mid Z], \mathbf{E}[N_j \mid Z]] \\ &= \mathbf{Cov}[\mathbf{E}[N_i \mid Z], \mathbf{E}[N_j \mid Z]]. \end{aligned} \tag{4.71}$$



(a) $w(x) = x$ (b) $w(x) = e^{\tau x}$

Figure 4.10: Comparison of the theoretical model with empirical methods when $n = 10000$

Similarly, we can also obtain

$$\mathbf{Cov}[X_{i,1}, X_{j,1}] = \mathbf{Cov}[\mathbf{E}[X_{i,1} \mid Z], \mathbf{E}[X_{j,1} \mid Z]] \tag{4.72}$$

and

$$\mathbf{Cov}[N_i, X_{j,1}] = \mathbf{Cov}[\mathbf{E}[N_i \mid Z], \mathbf{E}[X_{j,1} \mid Z]]. \tag{4.73}$$

We can use the empirical method to estimate those quantities and assess our parametric fitting. We illustrate this with an example of two business lines.

Keeping our former assumptions, $[X_{k,i} \mid Z]$, $k = 1, 2$, follow a Gamma distribution with average severity and variance

$$\mathbf{E}[X_{k,1} \mid Z] = c_k Z \quad \text{and} \quad \mathbf{Var}[X_{k,1} \mid Z] = \sigma_k^2 Z^2. \tag{4.74}$$

As well, the frequencies $[N_k \mid Z]$, $k = 1, 2$, follow a Binomial distribution $B(n_k, p_k Z)$, and the conditional average frequency and variance are given by

$$\mathbf{E}[N_k \mid Z] = n_k p_k Z \quad \text{and} \quad \mathbf{Var}[N_k \mid Z] = n_k p_k Z(1 - p_k Z). \tag{4.75}$$

Suppose that we only know the exact value of $n_k, k = 1, 2$. Then we will use empirical estimation to obtain as much information as we can from the data. Theoretically, we have

$$\mathbf{E}[X_{k,1}] = c_k \mathbf{E}[Z] \quad k = 1, 2, \tag{4.76}$$

$$\mathbf{E}[N_k] = n_k p_k \mathbf{E}[Z] \quad k = 1, 2, \tag{4.77}$$

$$\mathbf{Cov}[N_k, X_{k,1}] = n_k p_k c_k \mathbf{Var}[Z] \quad k = 1, 2, \tag{4.78}$$

$$\mathbf{Cov}[N_1, N_2] = n_1 p_1 n_2 p_2 \mathbf{Var}[Z], \tag{4.79}$$

$$\mathbf{Cov}[X_{1,1}, X_{2,1}] = c_1 c_2 \mathbf{Var}[Z], \tag{4.80}$$

$$\mathbf{Cov}[X_{k,1}, X_{k,2}] = c_k^2 \mathbf{Var}[Z] \quad k = 1, 2, \tag{4.81}$$

$$\mathbf{Var}[X_{k,1}] = c_k^2\mathbf{Var}[Z] + \sigma_k^2\mathbf{E}[Z^2] \quad k = 1, 2, \tag{4.82}$$

$$\mathbf{E}[S_k] = n_k p_k c_k \mathbf{E}[Z^2] \quad k = 1, 2. \tag{4.83}$$

From the above statements, we have the following relationship between parameters:

$$\frac{c_1}{c_2} = \frac{\mathbf{E}[X_{1,1}]}{\mathbf{E}[X_{2,1}]}, \tag{4.84}$$

where $\mathbf{E}[X_{k,1}]$, $k = 1, 2$ can be roughly estimated with the data. Similar approximate relationship among c_1, c_2, p_1, p_2 can also be obtained by evaluation of the ratios. For example,

$$\frac{\mathbf{Cov}[N_k, X_{k,1}]}{\mathbf{E}[N_k]\mathbf{E}[X_{k,1}]} = \frac{\mathbf{Var}[Z]}{\mathbf{E}[Z]^2}, \tag{4.85}$$

and

$$\frac{\mathbf{E}[S_k]}{\mathbf{E}[N_k]\mathbf{E}[X_{k,1}]} = \frac{\mathbf{E}[Z^2]}{\mathbf{E}[Z]^2}, \tag{4.86}$$

from which we can estimate ratios between mean, variance and second moment of background risk Z empirically.

Since all the parameters mentioned above relate closely with each other, we are able to roughly determine them through empirical estimation given any one of the parameters. Now, suppose we set the mean of background risk to a specific value $\mathbf{E}[Z] = \mu_Z$, then for $k = 1, 2$ other quantities can be determined as

$$c_k = \frac{\mathbf{E}[X_{k,1}]}{\mu_Z}, \tag{4.87}$$

$$p_k = \frac{\mathbf{E}[N_k]}{n_k \mu_Z}, \tag{4.88}$$

$$\mathbf{Var}[Z] = \frac{\mathbf{Cov}[N_k, X_{k,1}]}{\mathbf{E}[N_k]\mathbf{E}[X_{k,1}]}\mu_Z^2, \tag{4.89}$$

$$\sigma_k = \sqrt{\frac{\mathbf{Var}[X_{k,1}] - \mathbf{Cov}[X_{k,1}, X_{k,2}]}{\mathbf{E}[Z^2]}} \tag{4.90}$$

with

$$\mathbf{E}[Z^2] = \frac{\mathbf{E}[S_k]}{\mathbf{E}[N_k]\mathbf{E}[X_{k,1}]}\mu_Z^2, \tag{4.91}$$

where all the quantities on the right hand side can be estimated empirically.

We may change the value of μ_Z and choose the most suitable group of parameters according to the data, the details for which will be discussed in the subsequent section.

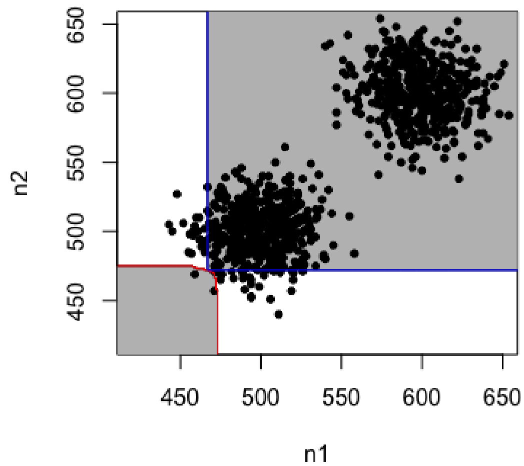


Figure 4.11: Illustration of our assessment method

4.6.2 Parametric fitting method and assessment

In this section, we introduce an assessment method for a group of estimated parameters for the proposed model. We illustrate our main idea using Figure 4.11.

Supposing that we have a group of estimated parameters for a predetermined μ_Z, the region bounded by the red boundary are those (n_1, n_2) that satisfy

$$P(N_1 \geq n_1, N_2 \geq n_2) \geq 1 - \theta, \tag{4.92}$$

for such an estimated parametric model. Given that we have n data points $(n_{1,j}, n_{2,j})$, $j = 1, \ldots, n$ for frequencies (N_1, N_2) recorded from the true model, all the points (n_1, n_2) satisfying

$$\frac{\sum_{j=1}^{n} 1\{n_{1,j} \geq n_1, n_{2,j} \geq n_2\}}{n} \geq 1 - \theta \tag{4.93}$$

should exactly be located in this region if the estimated parameters match the true model well. The equality should approximately hold for those (n_1, n_2) on the boundary of the red region. In other words, for any point (n_1, n_2) on the red quantile boundary, denoted by $(n_{1,l,*}, n_{2,l,*})$, $l = 1, \ldots, h$, the number of data points in the blue region should account for almost $1 - \theta$ of the total samples. Ideally, we aim to select the parametric model with quantile boundary $\{\wp_l, l = 1, \ldots, h\}$, such that

$$\wp_l = \frac{\sum_{j=1}^{n} 1\{n_{1,j} \geq n_{1,l,*}, n_{2,j} \geq n_{2,l,*}\}}{n} \sim 1 - \theta. \tag{4.94}$$

Hence, we are able to define mean squared error and mean absolute error for a parametric model with any parameter values as

$$MSE_q = \sum_{l=1}^{h} (\wp_l - 1 + \theta)^2, \tag{4.95}$$

and

$$MAE_q = \sum_{l=1}^{h} |\Theta p_l - 1 + \theta|. \tag{4.96}$$

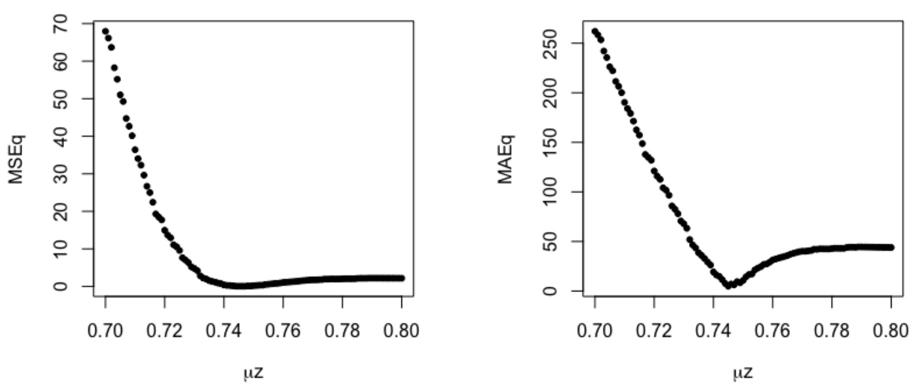
Example 4.6.1. *Keeping the same assumptions, we continue with the example given in the former subsection. We assume that the background risk Z follows the distribution* $Z \sim z^* + (z^{**} - z^*)B(1, p)$*, with* $z^* = 1$ *already known.*

In addition, suppose that we have for $k = 1, 2$, $c_k = 200$, $\sigma_k = 40$, $n_k = 2000$, $p_k = 0.5$, $z^{**} = 0.5$ *and* $p = 0.5$, *yielding* $\mathbf{E}[Z] = 0.75$ *and* $\mathbf{Var}[Z] = 0.0625$. *We simulate a sample of size* $n = 1000$ *and see how our parametric fitting and assessment perform in Table 4.2.*

Predetermined μ_Z	c_1	c_2	σ_1	σ_2	p_1	p_2	σ_Z^2	MSE_q	MAE_q
Theoretical 0.75	200	200	40	40	0.5	0.5	0.0625	6.238	0.0442
0.7	214.3	216.9	44.495	43.33	0.534	0.533	0.0549	67.949	261.994
0.75	200.02	202.4	41.529	40.442	0.498	0.498	0.063	0.195	10.705
0.8	187.52	189.75	38.93	37.91	0.467	0.467	0.0717	2.175	43.923

Table 4.2: Performance of MSE_q and MAE_q for some specific μ_Z

With more predetermined μ_Z increasing with a smaller step, we obtain Figure 4.12.



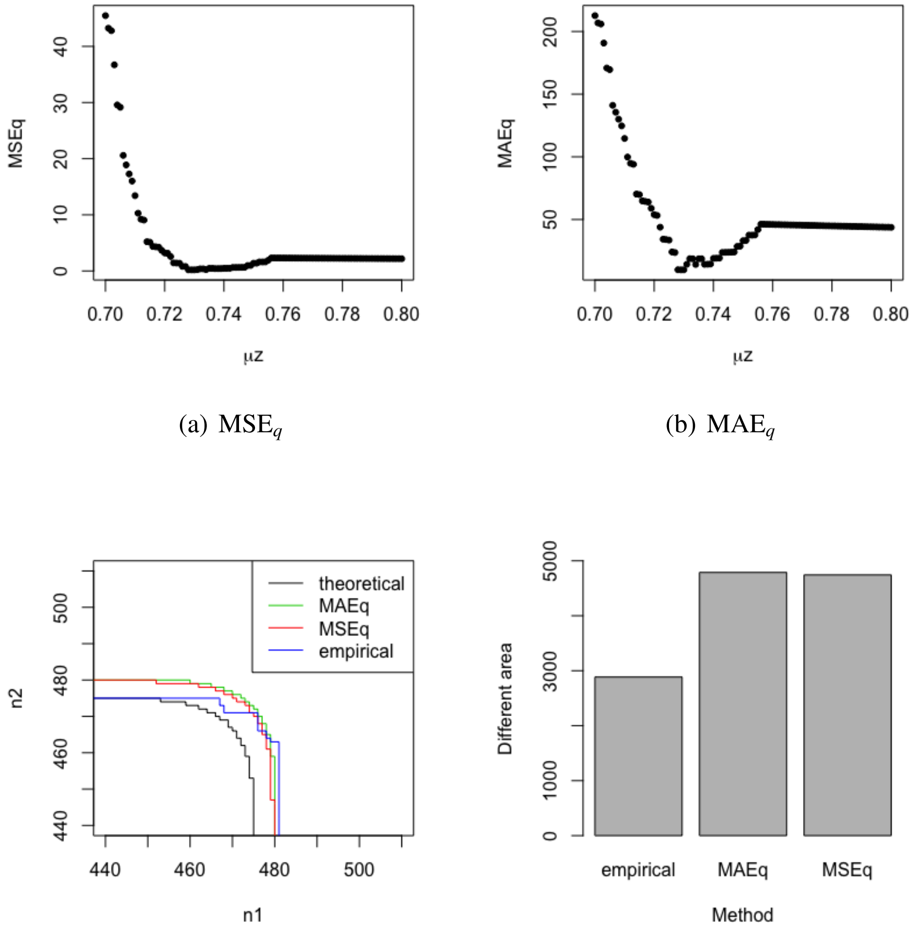
(a) MSE_q (b) MAE_q

Figure 4.12: MSE_q and MAE_q when μ_Z varies

The predetermined μ_Z which minimizes MAE_q is 0.745, and that minimizes MSE_q is 0.747. Both of them are around the real $\mu_Z = 0.75$. Although μ_Z obtained by minimizing MSE_q is closer to the real μ_Z than that obtained by minimizing MAE_q, MAE_q has a more obvious trend when changing the predetermined μ_Z, which may be more convincing.

We also explore the performance of MAE_q and MSE_q for different sample sizes.

When $n = 100$, the μ_Z that minimizes MSE_q is 0.73, and the μ_Z that minimizes MAE_q is 0.729 in Figure 4.13. The empirical method performs better than the parametric fitting method.



(a) MSE_q

(b) MAE_q

(c) Comparison of quantile boundaries

(d) Difference of area

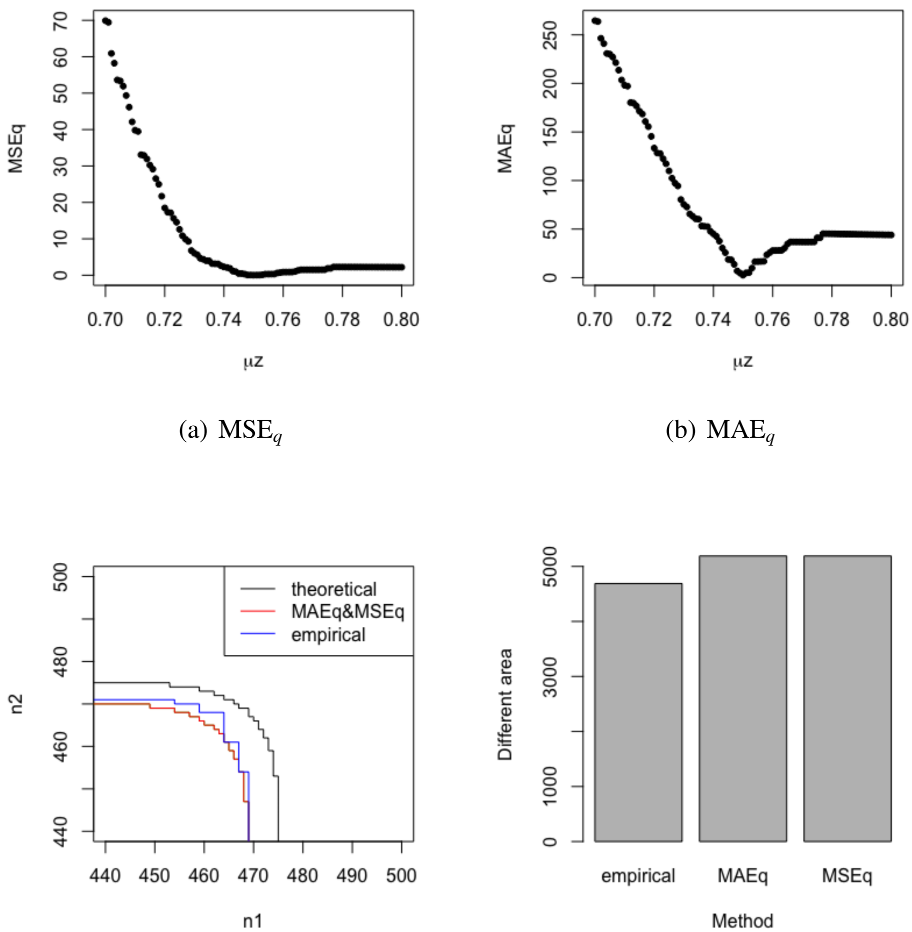
Figure 4.13: Parametric fitting of the case $n = 100$

When $n = 200$, the μ_Z that minimizes MSE_q is 0.75, and the μ_Z that minimizes MAE_q is 0.75 in Figure 4.14. Although, we obtain perfect estimation for μ_Z, the empirical method still works better than the parametric fitting method. This might be caused by poor estimations for $\mathbf{E}[N_k]$ and $\mathbf{Cov}[N_1, N_2]$.

When $n = 500$, the μ_Z that minimizes MSE_q is 0.733, and the μ_Z that minimizes MAE_q is 0.733 in Figure 4.15. The parametric fitting method starts working better than the empirical method.

When $n = 1000$, the μ_Z that minimizes MSE_q is 0.747, and the μ_Z that minimizes MAE_q is 0.745 in Figure 4.16. The parametric model obtained by minimizing MSE_q works very well to estimate the theoretical quantile boundary. However, the parametric model obtained by minimizing MAE_q does not work well compared to the empirical quantile boundary.

When $n = 1500$, the μ_Z that minimizes MSE_q is 0.748, and the μ_Z that minimizes MAE_q is 0.748 in Figure 4.17. Obviously, the parametric fitting method works better than empirical



(a) MSE_q (b) MAE_q

(c) Comparisons of quantile boundaries (d) Difference of area

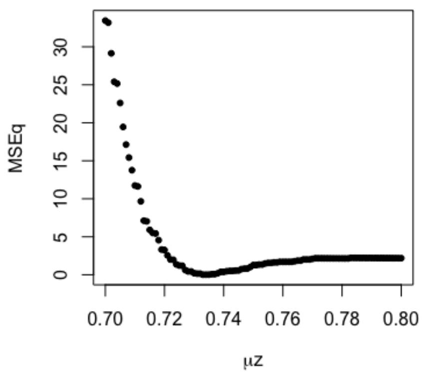
Figure 4.14: Parametric fitting of the case $n = 200$

method.

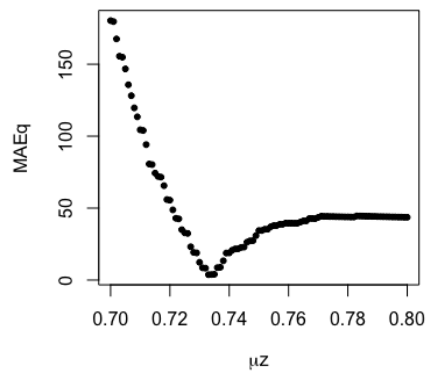
From the above figures, we see the necessity of parametric estimating method. Although the empirical method works better when sample size is small, the parametric estimation supported by assessment has a better accuracy for large sample sizes.

4.7 Conclusion

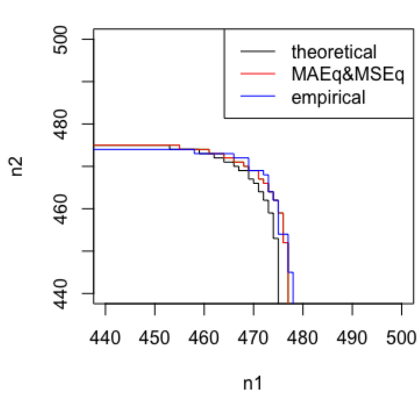
In this chapter, we studied premium principles for multiple business lines when the claim frequencies and claim severities are correlated via a background risk. We developed a novel methodology of premium setting and numerically illustrated how model parameters influence the premiums level. As one might expect, higher mean of severity, mean of background risk and standard deviation of background risk will result in higher premiums in a certain business line. In addition, a larger number of policies (frequencies) (larger parameter p_1 in the Binomial



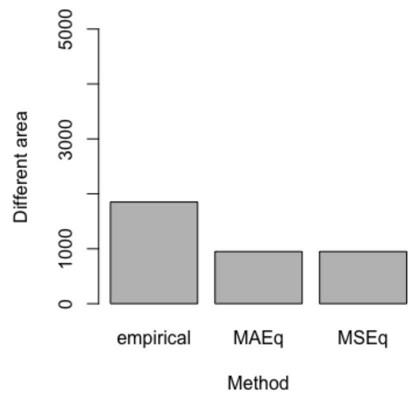
(a) MSE_q



(b) MAE_q



(c) Comparisons of quantile boundaries

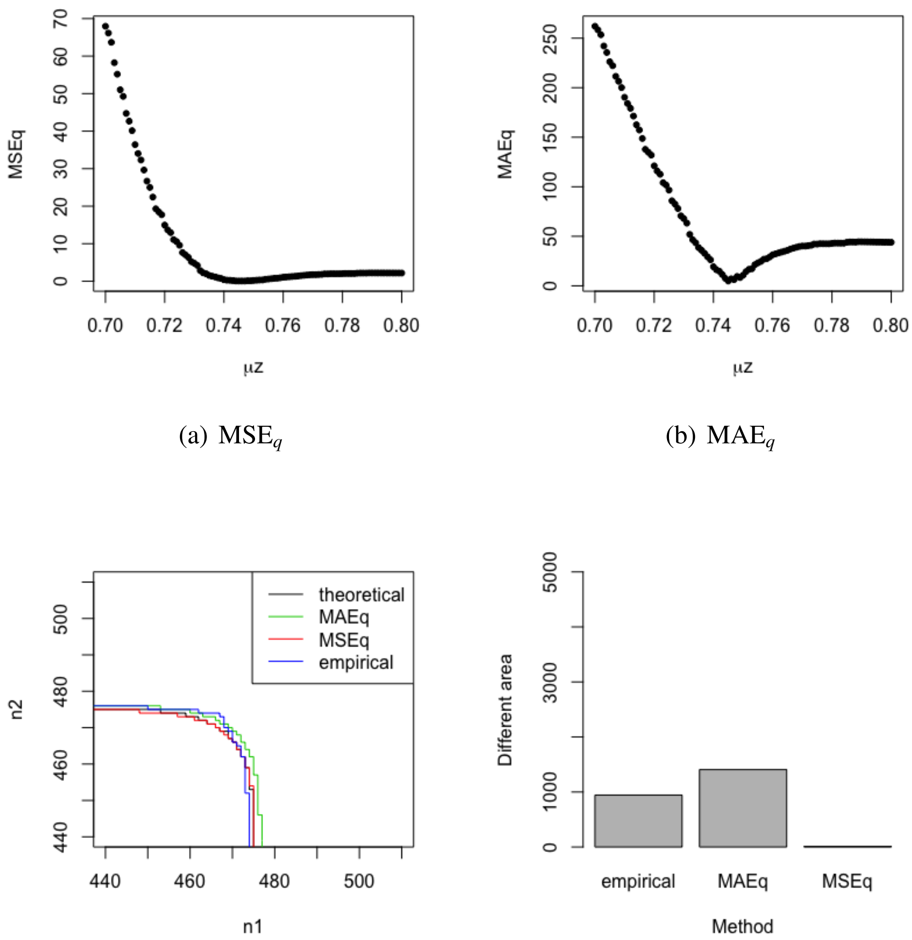


(d) Difference of area

Figure 4.15: Parametric fitting of the case $n = 500$

distribution) can lead to a relatively lower premium per policy. Depending on the choice of weight function, the variance of the severity may have different influence on premiums.

If the background risk levels can be observed along with the loss data, a less time-consuming non-parametric method can be used to empirically estimate premiums. However, the improvement on accuracy is not significant. Given that our assumptions for the risk model are correct, the proposed parameter fitting method works well when the sample size is large. Otherwise, the empirical methods are recommended.



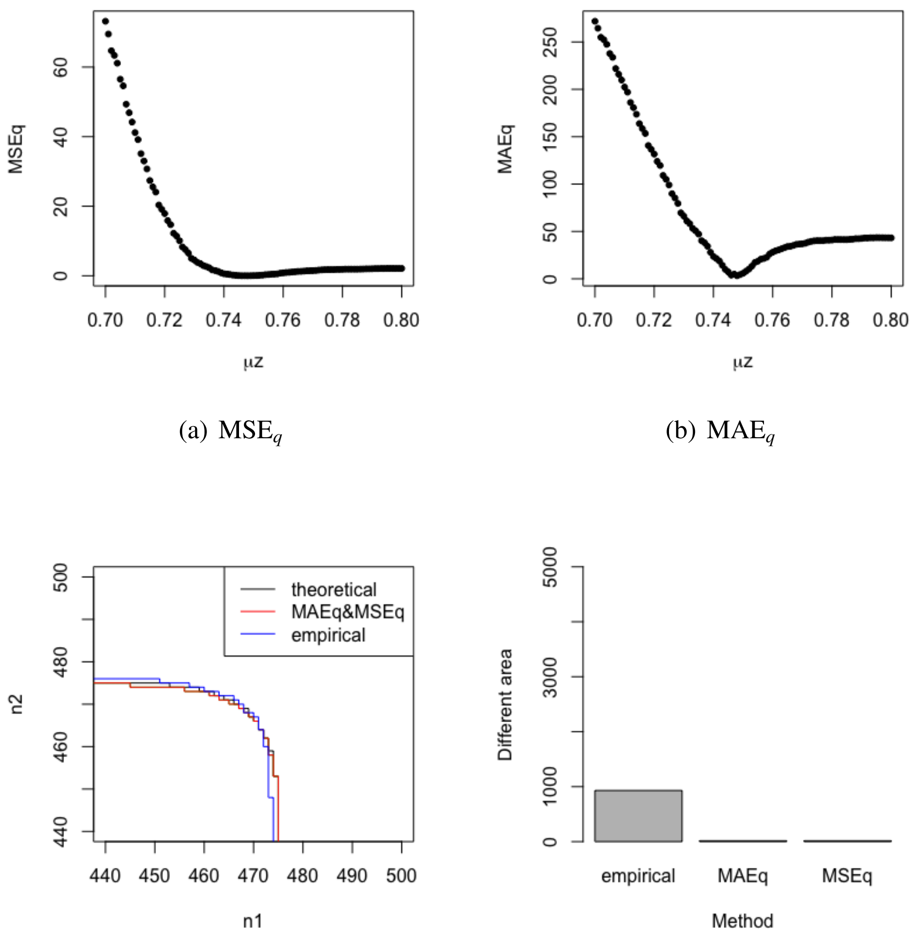
(a) MSE_q

(b) MAE_q

(c) Comparison of quantile boundaries

(d) Difference of area

Figure 4.16: Parametric fitting of the case $n = 1000$



(a) MSE_q

(b) MAE_q

(c) Comparison of quantile boundaries

(d) Difference of area

Figure 4.17: Parametric fitting of the case $n = 1500$

Chapter 5

Summary and future research topics

In this thesis we investigated the univariate and bivariate collective risk models and made contributions in the following aspects.

In Chapter 2, we provide theoretical upper bounds for the tail probability of the bivariate compound distributions. The results generalize those of Willmot and Lin (1994) and Willmot et al. (2001) for univariate compound distributions. Exponential or generalized bounds are derived depending on whether the claim size distributions have finite moment generating functions or not. In order for the bound to apply, some quite restrictive conditions for the claim number distributions have to be satisfied. However, we were able to make the constraints as relaxed as possible, so that the derived upper bounds are applicable to some commonly used bivariate discrete distributions, such as a specific category of bivariate compound distributions. The results are useful in measuring and managing the risk of insurance companies.

In Chapter 3, we developed several novel variance reduction techniques for simulating tail probability and mean excess loss of the classic univariate collective risk model. We also extended them to bivariate aggregate claims model whose claim severities are identically independent distributed and independent of claim frequencies. The numerical experiments show that our new simulation methods are highly effective. It is worth mentioning that, such variance reduction techniques can naturally be applied to simulating the tail probability of bivariate compound distributions discussed in the first part of our thesis.

In Chapter 4, we investigate a novel multiple collective risk model, in which the claim frequencies and claim severities are correlated via a background risk. Considering a model with two business lines, we numerically illustrate how model parameters influence the premiums level. We developed two empirical methods of parameter estimation based on whether the background risk levels can be observed along with the loss data, Furthermore, we developed a parametric fitting method, which outperforms empirical methods when the sample size is large enough, and of course, if the model specification is correct.

For future research, several topics can be considered:

- In Chapter 2, we only numerically verified that several commonly used bivariate counting distributions satisfy the required constraints for the bound to hold. Some work could be done to explore such properties in various bivariate distributions analytically.

- In Chapter 3, we make use of a size-biased transform and a second moment transform to simulate mean excess loss alternatively. We also verify that these can be applied to the bivariate Poisson distribution with mixing parameter or common shock. Moment transform on bivariate distributions could be considered in detail.

- In Chapter 4, we mainly discuss premium pricing via numerical methodologies because of the complexity of our assumptions. We may consider finding some special cases of our model to determine a potential analytical allocation rule.

Bibliography

Asmussen, S. and P. W. Glynn (2007). *Stochastic Simulation: Algorithms and Analysis*, Volume 57. Springer Science & Business Media.

Barlow, R. E. and F. Proschan (1975). Statistical theory of reliability and life testing: probability models. Technical report, Florida State Univ Tallahassee.

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

Bryan, K. M. (1999). Elementary inversion of the Laplace transform. *Mathematical Sciences Technical Reports (MSTR)*, Paper 114.

Bühlmann, H. (1980). An economic premium principle. *ASTIN Bulletin: The Journal of the IAA 11*(1), 52–60.

Bühlmann, H. (1984). The general economic premium principle. *ASTIN Bulletin: The Journal of the IAA 14*(1), 13–21.

Cai, J. and Y. Wu (1997). Some improvements on the Lundberg bound for the ruin probability. *Statistics & Probability Letters 33*(4), 395–403.

Cossette, H., E. Marceau, and F. Marri (2008). On the compound poisson risk model with dependence based on a generalized Farlie–Gumbel–Morgenstern copula. *Insurance: Mathematics and Economics 43*(3), 444–455.

Cossette, H., E. Marceau, I. Mtalai, and D. Veilleux (2018). Dependent risk models with Archimedean copulas: A computational strategy based on common mixtures and applications. *Insurance: Mathematics and Economics 78*, 53–71.

Czado, C., R. Kastenmeier, E. C. Brechmann, and A. Min (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal 2012*(4), 278–305.

Davidson, R. and J. G. MacKinnon (1992). Regression-based methods for using control variates in Monte Carlo experiments. *Journal of Econometrics 54*(1-3), 203–222.

Denuit, M. (2020). Size-biased risk measures of compound sums. *North American Actuarial Journal 24*(4), 512–532.

Denuit, M., J. Dhaene, M. Goovaerts, and R. Kaas (2006). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley & Sons.

Dhaene, J., G. Willmot, and B. Sundt (1999). Recursions for distribution functions and stop-loss transforms. *Scandinavian Actuarial Journal 1999*(1), 52–65.

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

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Volume 1. John Wiley & Sons,.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Volume 2. John Wiley & Sons,.

Furman, E. and R. Zitikis (2008). Weighted risk capital allocations. *Insurance: Mathematics and Economics 43*(2), 263–269.

Furman, E. and R. Zitikis (2009). Weighted pricing functionals with applications to insurance: an overview. *North American Actuarial Journal 13*(4), 483–496.

Furman, E. and R. Zitikis (2017). Beyond the pearson correlation: Heavy-tailed risks, weighted Gini correlations, and a Gini-type weighted insurance pricing model. *ASTIN Bulletin: The Journal of the IAA 47*(3), 919–942.

Gerber, H. U. (1979). An introduction to mathematical risk theory. *Huebner Foundation Monograph. 8*.

Gerber, H. U. (1994). Martingales and tail probabilities. *ASTIN Bulletin: The Journal of the IAA 24*(1), 145–146.

Glasserman, P., P. Heidelberger, and P. Shahabuddin (2000). Efficient Monte Carlo methods for value-at-risk. *Master. Risk 2*.

Goovaerts, M., F. E. De Vylder, and J. Haezendonck (1984). *Insurance Premiums*. North-Holland.

Goovaerts, M. J., R. Kaas, J. Dhaene, and Q. Tang (2003). A unified approach to generate risk measures. *ASTIN Bulletin: The Journal of the IAA 33*(2), 173–191.

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

Heilmann, W.-R. (1989). Decision theoretic foundations of credibility theory. *Insurance: Mathematics and Economics 8*(1), 77–95.

Hesselager, O. (1996). Recursions for certain bivariate counting distributions and their compound distributions. *ASTIN Bulletin: The Journal of the IAA 26*(1), 35–52.

Jin, T. and J. Ren (2010). Recursions and fast Fourier transforms for certain bivariate compound distributions. *Journal of Operational Risk 4*, 19.

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

Kousky, C. and R. M. Cooke (2009). The unholy trinity: fat tails, tail dependence, and micro-correlations. *Resources for the Future Discussion Paper*, 09–36.

Krämer, N., E. C. Brechmann, D. Silvestrini, and C. Czado (2013). Total loss estimation using copula-based regression models. *Insurance: Mathematics and Economics 53*(3), 829–839.

Lavenberg, S. S. and P. D. Welch (1981). A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Management Science 27*(3), 322–335.

Lin, X. (1996). Tail of compound distributions and excess time. *Journal of Applied Probability 33*(1), 184–195.

Martel-Escobar, M., A. Hernández-Bastida, and F. J. Vázquez-Polo (2012). On the independence between risk profiles in the compound collective risk actuarial model. *Mathematics and Computers in Simulation 82*(8), 1419–1431.

Meyers, G. G. (2007). The common shock model for correlated insurance losses. *Variance 1*(1), 40–52.

Panjer, H. H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin: The Journal of the IAA 12*(1), 22–26.

Peköz, E. and S. M. Ross (2004). Compound random variables. *Probability in the Engineering and Informational Sciences 18*(4), 473.

Ren, J. (2021). Tail moments of compound distributions. *Available at SSRN 3880127*.

Robertson, J. (1992). The computation of aggregate loss distributions. *Proceedings of the Casualty Actuarial Society 79*(150), 57–133.

Ross, S. M. (2013). *Simulation, 5th Edision*. Academic Press.

Ross, S. M., J. J. Kelly, R. J. Sullivan, W. J. Perry, D. Mercer, R. M. Davis, T. D. Washburn, E. V. Sager, J. B. Boyce, and V. L. Bristow (1996). *Stochastic Processes*, Volume 2. Wiley New York.

Sarabia, J. M., E. Gómez-Déniz, F. Prieto, and V. Jordá (2016). Risk aggregation in multivariate dependent pareto distributions. *Insurance: Mathematics and Economics 71*, 154–163.

Shi, P., X. Feng, and A. Ivantsova (2015). Dependent frequency–severity modeling of insurance claims. *Insurance: Mathematics and Economics 64*, 417–428.

Sundt, B. (1999). On multivariate Panjer recursions. *ASTIN Bulletin: The Journal of the IAA 29*(1), 29–45.

Sundt, B. and R. Vernic (2009). *Recursions for Convolutions and Compound Distributions with Insurance Applications*. Springer Science & Business Media.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin: The Journal of the IAA 26*(1), 71–92.

Wang, S. (1998). Aggregation of correlated risk portfolios: models and algorithms. *Proceedings of the Casualty Actuarial society 85*(163), 848–939.

Wang, S. (2002). A set of new methods and tools for enterprise risk capital management and portfolio optimization. *CAS Forum Summer*, 43–78.

Willmot, G. E. (1994). Refinements and distributional generalizations of Lundberg's inequality. *Insurance: Mathematics and Economics 15*(1), 49–63.

Willmot, G. E. and X. Lin (1994). Lundberg bounds on the tails of compound distributions. *Journal of Applied Probability 31*(3), 743–756.

Willmot, G. E. and X. Lin (1997). Simplified bounds on the tails of compound distributions. *Journal of Applied Probability 34*(1), 127–133.

Willmot, G. E., X. S. Lin, and X. S. Lin (2001). *Lundberg Approximations for Compound Distributions with Insurance Applications*, Volume 156. Springer Science & Business Media.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society 55*(1), 95–115.

Appendix A

Supplement materials

A.1 Proof of the Remark 3.3.2 in Section 3.3.5

From

$$
\begin{aligned}
&\mathbf{Var}\Big[\mathbf{P}[M \geq T^*(c)]\frac{M_X(h)^{T^*(c)}}{e^{h\sum_{i=1}^{T^*(c)} X_i^*}}\Big] \\
&= \mathbf{Var}\Big[\mathbf{E}\Big[\mathbb{I}\{M \geq T^*(c)\}\frac{M_X(h)^M}{e^{h\sum_{i=1}^{M} X_i^*}} \mid T^*(c), X_1^*, \ldots, X_{T^*(c)}^*\Big]\Big] \\
&= \mathbf{Var}\Big[\mathbf{E}\Big[\mathbb{I}\{M \geq T^*(c)\}\frac{\mathbf{E}[M_X(h)^M]}{e^{h\sum_{i=1}^{M} X_i^*}}\frac{M_X(h)^M}{\mathbf{E}[M_X(h)^M]} \mid T^*(c), X_1^*, \ldots, X_{T^*(c)}^*\Big]\Big] \\
&= \mathbf{Var}\Big[\mathbf{E}\Big[\mathbb{I}\{M^* \geq T^*(c)\}\frac{\mathbf{E}[M_X(h)^M]}{e^{h\sum_{i=1}^{M^*} X_i^*}} \mid T^*(c), X_1^*, \ldots, X_{T^*(c)}^*\Big]\Big] \\
&= \mathbf{Var}\Big[\mathbf{E}\Big[\mathbb{I}\{S^* > c\}\frac{\mathbf{E}[e^{hS}]}{e^{hS^*}} \mid T^*(c), X_1^*, \ldots, X_{T^*(c)}^*\Big]\Big] \\
&= \mathbf{Var}\Big[\mathbb{I}\{S^* > c\}\frac{\mathbf{E}[e^{hS}]}{e^{hS^*}}\Big] - \mathbf{E}\Big[\mathbf{Var}\Big[\mathbb{I}\{S^* > c\}\frac{\mathbf{E}[e^{hS}]}{e^{hS^*}} \mid T^*(c), X_1^*, \ldots, X_{T^*(c)}^*\Big]\Big] \\
&\leq \mathbf{Var}\Big[\mathbb{I}\{S^* > c\}\frac{\mathbf{E}[e^{hS}]}{e^{hS^*}}\Big], \qquad \text{(A.1)}
\end{aligned}
$$

the h minimizing the variance of $\mathbf{P}[M \geq T^*(c)]\frac{M_X(h)^{T^*(c)}}{e^{h\sum_{i=1}^{T^*(c)} X_i^*}}$ is roughly the same as that minimizing the variance of $\mathbb{I}\{S^* > c\}\frac{\mathbf{E}[e^{hS}]}{e^{hS^*}}$.

A.2 Theoretical verification of conditioning method for the tail probability in the two dimensional case

To verify the feasibility of the conditioning method for $\mathbf{P}[S_1 > c, S_2 > d]$, we extend the proposition in Lin (1996) .

Assuming that $a_{m,n} = \mathbf{P}[M > m, N > n]$, we have

$$\mathbf{P}[T_1(c) = m] = \overline{F}^{*m}(c) - \overline{F}^{*(m-1)}(c)$$

and

$$\mathbf{P}[T_2(d) = n] = \overline{G}^{*n}(d) - \overline{G}^{*(n-1)}(d),$$

for $m, n = 1, 2, \ldots$ where $\overline{F}^{*m}(x) = 1 - F^{*m}(x)$, $\overline{G}^{*n}(y) = 1 - G^{*n}(y)$, with $F^{*m}(x)$, $G^{*n}(y)$ being the distribution functions (d.f.) of $S_{1,m}$, $S_{2,n}$ respectively. Therefore we can obtain the following proposition.

Proposition A.2.1.

$$\mathbf{P}[S_1 > c, S_2 > d] = \mathbf{E}[a_{T_1(c)-1,T_2(d)-1}] \tag{A.2}$$

Proof. Directly from Lemma 2.2.1, we have

$$\begin{aligned}
\mathbf{P}[S_1 > c, S_2 > d] &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{m,n}\{\overline{F}^{*(m+1)}(c)\overline{G}^{*(n+1)}(d) + \overline{F}^{*m}(c)\overline{G}^{*n}(d) \\
&\qquad - \overline{F}^{*m}(c)\overline{G}^{*(n+1)}(d) - \overline{F}^{*(m+1)}(c)\overline{G}^{*n}(d)\} \\
&= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{m-1,n-1}\{\overline{F}^{*m}(c)\overline{G}^{*n}(d) + \overline{F}^{*(m-1)}(c)\overline{G}^{*(n-1)}(d) \\
&\qquad - \overline{F}^{*(m-1)}(c)\overline{G}^{*n}(d) - \overline{F}^{*m}(c)\overline{G}^{*(n-1)}(d)\} \\
&= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{m-1,n-1}[\overline{F}^{*m}(c) - \overline{F}^{*(m-1)}(d)][\overline{G}^{*n}(d) - \overline{G}^{*(n-1)}(d)] \\
&= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{m-1,n-1}\mathbf{P}[T_1(c) = m]\mathbf{P}[T_2(d) = n] \\
&= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{m-1,n-1}\mathbf{P}[T_1(c) = m, T_2(d) = n] \\
&= \mathbf{E}[a_{T_1(c)-1,T_2(d)-1}],
\end{aligned} \tag{A.3}$$

since $T_1(c)$ and $T_2(d)$ are independent.

Appendix B

Algorithms of simulation methods

In each of following algorithms, the estimator of the underlying quantity of interest is the average of generate n samples.

B.1 Simulation of the tail probability of the univariate compound sum

Algorithm 1 Importance sampling

1: Apply Esscher transform to S with parameter h and calculate the distributions of M^*, X^*.
2: Let $\mathbf{E}[M^*]\mathbf{E}[X^*] = c$ and determine the parameter value of h.
3: Generate M^*.
4: Generate $X_1^*, \ldots, X_{M^*}^*$ independently from X^* and calculate the sum, denoted by $S^* = \sum_{i=1}^{M^*} X_i^*$ and calculate $\hat{\theta}_I = \mathbb{I}(S^* > c)$.
5: Repeat steps 3-4 for n times and take the average of $\hat{\theta}_I$.

Algorithm 2 Importance and stratified sampling

1: Apply Esscher transform to S with parameter h and calculate the distributions of M^*, X^*.
2: Let $\mathbf{E}[M^*]\mathbf{E}[X^*] = c$ and determine the parameter value of h.
3: Fix a value of m_l such that $\mathbf{P}[M^* > m_l]$ is small.
4: Generate M^* conditional on it exceeding m_l.
5: Generate $X_1^*, \ldots, X_{M^*}^*$ independently from X^* and calculate the estimator $\hat{\theta}_{I+S}$ in this run.
6: Repeat steps 4-5 for n times and take the average of $\hat{\theta}_{I+S}$.

Algorithm 3 Conditioning

1: Generate $X_1, X_2, \ldots$ independently from X until their sum exceeds c.
2: Record the stopping time $T(c)$ and calculate $\hat{\theta}_{CD} = \mathbf{P}[M \geq T(c)]$.
3: Repeat steps 1-2 for n times and take the average of $\hat{\theta}_{CD}$.

Algorithm 4 Combining conditioning method with control variates

1: Generate $X_1, X_2, \dots$ independently from X until their sum exceeds c.
2: Record the stopping time $T(c)$ and calculate $\hat{\theta}_{CD} = \mathbf{P}[M \geq T(c)]$.
3: Calculate the control variate W in this run.
4: Repeat steps 1-3 for n times and calculate $\gamma = cov(\hat{\theta}_{CD}, W)/Var(W)$ using the simulated values of $\hat{\theta}_{CD}$ and W.
5: Take the average of $\hat{\theta}_{CD+CV} = \hat{\theta}_{CD} - \gamma W$.

Algorithm 5 Combining importance sampling with conditioning

1: Apply Esscher transform to S with parameter h and calculate the distributions of M^*, X^*.
2: Let $\mathbf{E}[M^*]\mathbf{E}[X^*] = c$ and determine the parameter value of h.
3: Generate $X_1^*, X_2^*, \dots$ independently from X^* until their sum exceeds c.
4: Record the stopping time $T^*(c)$ and calculate $\hat{\theta}_{I+CD}$ in this run.
5: Repeat steps 3-4 for n times and take the average of $\hat{\theta}_{I+CD}$.

Algorithm 6 Combining importance sampling, conditioning and control variates

1: Apply Esscher transform to S with parameter h and calculate the distributions of M^*, X^*.
2: Let $\mathbf{E}[M^*]\mathbf{E}[X^*] = c$ and determine the parameter value of h.
3: Generate $X_1^*, X_2^*, \dots$ independently from X^* until their sum exceeds c.
4: Record the stopping time $T^*(c)$ and calculate $\hat{\theta}_{I+CD}$ in this run.
5: Calculate the control variate W in this run.
6: Repeat steps 3-5 for n times and calculate $\gamma = cov(\hat{\theta}_{I+CD}, W)/Var(W)$ using the simulated values of $\hat{\theta}_{I+CD}$ and W.
7: Take the average of $\hat{\theta}_{I+CD+CV} = \hat{\theta}_{I+CD} - \gamma W$.

B.2 Simulation of the mean excess loss of the univariate compound distribution

Algorithm 7 Importance sampling

1: Apply Esscher transform to S with parameter h and calculate the distributions of M^*, X^*.
2: Let $\mathbf{E}[M^*]\mathbf{E}[X^*] = c$ and determine the parameter value of h.
3: Generate M^*.
4: Generate $X_1^*, \dots, X_{M^*}^*$ independently from X^* and calculate the sum, denoted by $S^* = \sum_{i=1}^{M^*} X_i^*$.
5: Repeat steps 3-4 for n times and take the average of $\hat{\tau}_I$.

Algorithm 8 Importance and stratified sampling

1. Apply Esscher transform to S with parameter h and calculate the distributions of M^*, X^*.
2. Let $\mathbf{E}[M^*]\mathbf{E}[X^*] = c$ and determine the parameter value of h.
3. Fix a value of m_l such that $\mathbf{P}[M^* > m_l]$ is small.
4. Generate M^* conditional on it exceeding m_l.
5. Generate $X_1^*, \dots, X_{M^*}^*$ independently from X^* and calculate the estimator $\hat{\tau}_{I+S}$ in this run.
6. Repeat steps 4-5 for n times and take the average of $\hat{\tau}_{I+S}$.

Algorithm 9 Conditioning

1. Generate $X_1, X_2, \dots$ independently from X until their sum exceeds c.
2. Record the stopping time $T(c)$ and the quantity A.
3. Use $T(c)$ and A to calculate $\hat{\tau}_{CD}$ in this run.
4. Repeat steps 1-3 for n times and take the average of $\hat{\tau}_{CD}$.

Algorithm 10 Combining conditioning method with control variates

1. Generate $X_1, X_2, \dots$ independently from X until their sum exceeds c.
2. Record the stopping time $T(c)$ and the quantity A.
3. Use $T(c)$ and A to calculate $\hat{\tau}_{CD}$ and W_1 in this run.
4. Repeat steps 1-3 for n times and calculate W_2 by deducting sample mean $\mathbf{E}[A]$ from simulated A's.
5. Obtain γ_1, γ_2 by running the linear regression $\hat{\tau}_{CD} = \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2 + \epsilon$.
6. Take the average of $\hat{\tau}_{CD+CV} = \hat{\tau}_{CD} - \gamma_1 W_1 - \gamma_2 W_2$.

Algorithm 11 Combining importance sampling with conditioning

1. Apply Esscher transform to S with parameter h and calculate the distributions of M^*, X^*.
2. Let $\mathbf{E}[M^*]\mathbf{E}[X^*] = c$ and determine the parameter value of h.
3. Generate $X_1^*, X_2^*, \dots$ independently from X^* until their sum exceeds c.
4. Record the stopping time $T^*(c)$ and calculate the quantity A^*.
5. Use $T^*(c)$ and A^* to calculate $\hat{\tau}_{I+CD}$ in this run.
6. Repeat steps 3-5 for n times and take the average of $\hat{\tau}_{I+CD}$.

Algorithm 12 Combining importance sampling, conditioning and control variates

1. Apply Esscher transform to S with parameter h and calculate the distributions of M^*, X^*.
2. Let $\mathbf{E}[M^*]\mathbf{E}[X^*] = c$ and determine the parameter value of h.
3. Generate $X_1^*, X_2^*, \dots$ independently from X^* until their sum exceeds c.
4. Record the stopping time $T^*(c)$ and the quantity A^*.
5. Use $T^*(c)$ and A^* to calculate $\hat{\tau}_{I+CD}$ and W_1 in this run.
6. Repeat steps 3-5 for n times and calculate W_2 by deducting sample mean $\mathbf{E}[A^*]$ from simulated A^*'s.
7. Obtain γ_1, γ_2 by running the linear regression $\hat{\tau}_{I+CD} = \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2 + \epsilon$.
8. Take the average of $\hat{\tau}_{I+CD+CV} = \hat{\tau}_{I+CD} - \gamma_1 W_1 - \gamma_2 W_2$.

B.3 Simulation of the tail probability of the bivariate compound sum

Algorithm 13 Importance sampling

1: Apply Esscher transform to (S_1, S_2) with parameter (h_1, h_2) and calculate the distributions of X^*, Y^* and (M^*, N^*).
2: Let $\mathbf{E}[S_1^*] = c$, $\mathbf{E}[S_2^*] = d$ and determine the parameter value of h_1, h_2.
3: Generate (M^*, N^*).
4: Generate $X_1^*, \ldots, X_{M^*}^*$ independently from X^* and $Y_1^*, \ldots, Y_{N^*}^*$ independently from Y^*. Calculate the sum, denoted by $S_1^* = \sum_{i=1}^{M^*} X_i^*$ and $S_2^* = \sum_{j=1}^{N^*} Y_j^*$, respectively.
5: Repeat steps 3-4 for n times and take the average of $\hat{\theta}_I$.

Algorithm 14 Importance and stratified sampling

1: Apply Esscher transform to (S_1, S_2) with parameter (h_1, h_2) and calculate the distributions of X^*, Y^* and (M^*, N^*).
2: Let $\mathbf{E}[S_1^*] = c$, $\mathbf{E}[S_2^*] = d$ and determine the parameter value of h_1, h_2.
3: Fix values of m_l, n_l such that $1 - \mathbf{P}[M^* \leq m_l, N^* \leq n_l]$ is small.
4: Generate (M^*, N^*) conditional on $M^* > m_l$, $N^* > n_l$.
5: Generate $X_1^*, \ldots, X_{\max(m_l, M^*)}^*$ independently from X^* and $Y_1^*, \ldots, Y_{\max(n_l, N^*)}^*$ independently from Y^*. Calculate the estimator $\hat{\theta}_{I+S}$ in this run.
6: Repeat steps 4-5 for n times and take the average of $\hat{\theta}_{I+S}$.

Algorithm 15 Conditioning

1: Generate $X_1, X_2, \ldots$ independently from X until their sum exceeds c and $Y_1, Y_2, \ldots$ independently from Y until their sum exceeds d.
2: Record the stopping times $T_1(c)$ and $T_2(d)$. Calculate $\hat{\theta}_{CD} = \mathbf{P}[M \geq T_1(c), N \geq T_2(d)]$.
3: Repeat steps 1-2 for n times and take the average of $\hat{\theta}_{CD}$.

Algorithm 16 Combining conditioning method with control variates

1: Generate $X_1, X_2, \ldots$ independently from X until their sum exceeds c and $Y_1, Y_2, \ldots$ independently from Y until their sum exceeds d.
2: Record the stopping times $T_1(c)$ and $T_2(d)$. Calculate $\hat{\theta}_{CD} = \mathbf{P}[M \geq T_1(c), N \geq T_2(d)]$.
3: Calculate the control variates W_1 and W_2 in this run.
4: Repeat steps 1-3 for n times and obtain γ_1, γ_2 by running the linear regression $\hat{\theta}_{CD} = \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2 + \epsilon$.
5: Take the average of $\hat{\theta}_{CD+CV} = \hat{\tau}_{CD} - \gamma_1 W_1 - \gamma_2 W_2$.

Algorithm 17 Combining importance sampling with conditioning

1: Apply Esscher transform to (S_1, S_2) with parameter (h_1, h_2) and calculate the distributions of X^*, Y^* and (M^*, N^*).
2: Let $\mathbf{E}[S_1^*] = c$, $\mathbf{E}[S_2^*] = d$ and determine the parameter value of h_1, h_2.
3: Generate $X_1^*, X_2^*, \ldots$ independently from X^* until their sum exceeds c and $Y_1^*, Y_2^*, \ldots$ independently from Y^* until their sum exceeds d.
4: Record the stopping times $T_1^*(c)$ and $T_2^*(d)$. Calculate the estimator $\hat{\theta}_{I+CD}$ in this run.
5: Repeat steps 3-4 for n times and take the average of $\hat{\theta}_{I+CD}$.

Algorithm 18 Combining importance sampling, conditioning and control variates

1: Apply Esscher transform to (S_1, S_2) with parameter (h_1, h_2) and calculate the distributions of X^*, Y^* and (M^*, N^*).
2: Let $\mathbf{E}[S_1^*] = c$, $\mathbf{E}[S_2^*] = d$ and determine the parameter value of h_1, h_2.
3: Generate $X_1^*, X_2^*, \ldots$ independently from X^* until their sum exceeds c and $Y_1^*, Y_2^*, \ldots$ independently from Y^* until their sum exceeds d.
4: Record the stopping times $T_1^*(c)$ and $T_2^*(d)$. Calculate the estimator $\hat{\theta}_{I+CD}$ in this run.
5: Calculate the control variates W_1 and W_2 in this run.
6: Repeat steps 3-5 for n times and obtain γ_1, γ_2 by running the linear regression $\hat{\theta}_{I+CD} = \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2 + \epsilon$.
7: Take the average of $\hat{\theta}_{I+CD+CV} = \hat{\tau}_{I+CD} - \gamma_1 W_1 - \gamma_2 W_2$.

B.4 Simulation of the mean excess loss of the bivariate compound distribution

Algorithm 19 Importance sampling

1: Apply Esscher transform to (S_1, S_2) with parameter (h_1, h_2) and calculate the distributions of X^*, Y^* and (M^*, N^*).
2: Let $\mathbf{E}[S_1^*] = c$, $\mathbf{E}[S_2^*] = d$ and determine the parameter value of h_1, h_2.
3: Generate (M^*, N^*).
4: Generate $X_1^*, \ldots, X_{M^*}^*$ independently from X^* and $Y_1^*, \ldots, Y_{N^*}^*$ independently from Y^*. Calculate the sum, denoted by $S_1^* = \sum_{i=1}^{M^*} X_i^*$ and $S_2^* = \sum_{j=1}^{N^*} Y_j^*$, respectively.
5: Repeat steps 3-4 for n times and take the average of $\hat{\tau}_I$.

Algorithm 20 Importance and stratified sampling

1: Apply Esscher transform to (S_1, S_2) with parameter (h_1, h_2) and calculate the distributions of X^*, Y^* and (M^*, N^*).
2: Let $\mathbf{E}[S_1^*] = c$, $\mathbf{E}[S_2^*] = d$ and determine the parameter value of h_1, h_2.
3: Fix values of m_l, n_l such that $1 - \mathbf{P}[M^* \leq m_l, N^* \leq n_l]$ is small.
4: Generate (M^*, N^*) conditional on $M^* > m_l$, $N^* > n_l$.
5: Generate $X_1^*, \ldots, X^*_{\max(m_l, M^*)}$ independently from X^* and $Y_1^*, \ldots, Y^*_{\max(n_l, N^*)}$ independently from Y^*. Calculate the estimator $\hat{\tau}_{I+S}$ in this run.
6: Repeat steps 4-5 for n times and take the average of $\hat{\tau}_{I+S}$.

Algorithm 21 Conditioning

1: Generate $X_1, X_2, \ldots$ independently from X until their sum exceeds c and $Y_1, Y_2, \ldots$ independently from Y until their sum exceeds d.
2: Record the stopping times $T_1(c)$, $T_2(d)$ and the quantities A, B.
3: Calculate the conditional estimator $\hat{\tau}_{CD}$ in this run.
4: Repeat steps 1-3 for n times and take the average of $\hat{\tau}_{CD}$.

Algorithm 22 Combining conditioning method with control variates

1: Generate $X_1, X_2, \ldots$ independently from X until their sum exceeds c and $Y_1, Y_2, \ldots$ independently from Y until their sum exceeds d.
2: Record the stopping times $T_1(c)$, $T_2(d)$ and the quantities A, B.
3: Calculate the conditional estimator $\hat{\tau}_{CD}$ and the control variates W_1, W_2 in this run.
4: Repeat steps 1-3 for n times and calculate W_3, W_4 by deducting sample means $\mathbf{E}[A]$, $\mathbf{E}[B]$ from simulated A's, B's, respectively.
5: Obtain $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ by fitting the regression model $\hat{\tau}_{CD} \sim \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2 + \gamma_3 W_3 + \gamma_4 W_4$.
6: Take the average of $\hat{\tau}_{CD+CV} = \hat{\tau}_{CD} - \gamma_1 W_1 - \gamma_2 W_2 - \gamma_3 W_3 - \gamma_4 W_4$.

Algorithm 23 Combining importance sampling with conditioning

1: Apply Esscher transform to (S_1, S_2) with parameter (h_1, h_2) and calculate the distributions of X^*, Y^* and (M^*, N^*).
2: Let $\mathbf{E}[S_1^*] = c$, $\mathbf{E}[S_2^*] = d$ and determine the parameter value of h_1, h_2.
3: Generate $X_1^*, X_2^*, \ldots$ independently from X^* until their sum exceeds c and $Y_1^*, Y_2^*, \ldots$ independently from Y^* until their sum exceeds d.
4: Record the stopping times $T_1^*(c)$, $T_2^*(d)$ and the quantities A^*, B^*.
5: Calculate the conditional estimator $\hat{\tau}_{I+CD}$ in this run.
6: Repeat steps 3-5 for n times and take the average of $\hat{\tau}_{I+CD}$.

Algorithm 24 Combining importance sampling, conditioning and control variates

1: Apply Esscher transform to (S_1, S_2) with parameter (h_1, h_2) and calculate the distributions of X^*, Y^* and (M^*, N^*).
2: Let $\mathbf{E}[S_1^*] = c$, $\mathbf{E}[S_2^*] = d$ and determine the parameter value of h_1, h_2.
3: Generate $X_1^*, X_2^*, \ldots$ independently from X^* until their sum exceeds c and $Y_1^*, Y_2^*, \ldots$ independently from Y^* until their sum exceeds d.
4: Record the stopping times $T_1^*(c)$, $T_2^*(d)$ and the quantities A^*, B^*.
5: Calculate the conditional estimator $\hat{\tau}_{I+CD}$ and the control variates W_1, W_2 in this run.
6: Repeat steps 3-5 for n times and calculate W_3, W_4 by deducting sample means $\mathbf{E}[A^*]$, $\mathbf{E}[B^*]$ from simulated A^*'s, B^*'s, respectively.
7: Obtain $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ by fitting the regression model $\hat{\tau}_{I+CD} \sim \gamma_0+\gamma_1 W_1+\gamma_2 W_2+\gamma_3 W_3+\gamma_4 W_4$.
8: Take the average of $\hat{\tau}_{I+CD+CV} = \hat{\tau}_{I+CD} - \gamma_1 W_1 - \gamma_2 W_2 - \gamma_3 W_3 - \gamma_4 W_4$.

Curriculum Vitae

Name:	Ang Li
Post-Secondary Education and Degrees:	Shandong University Jinan, CN 2012 - 2016 B.S.
	University of Western Ontario London, ON 2016 - 2017 M.S.
	University of Western Ontario London, ON 2017 - 2021 Ph.D.
Honours and Awards:	Hua Loo-Keng Scholarship 2015-2016
	DSAS Graduate Teaching Assistant Award 2018-2019
Related Work Experience:	Teaching Assistant The University of Western Ontario 2016 - 2021

Publications and submissions:

Li, A. and Ren, J. Simulation methods for compound distributions (Under review by the journal: *Variance*)