

# Aggregate claim models with one-way and two-way dependence among individual claims

Xueyuan Wu<sup>1,\*</sup>, Kam Chuen Yuen<sup>2</sup>, Pengcheng Zhang<sup>1</sup>

<sup>1</sup>*Department of Economics, The University of Melbourne, VIC 3010, AUS*

<sup>2</sup>*Department of Statistics and Actuarial Science, HKU, Pokfulam, HK*

**Abstract** Motivated by some real life correlations among insurance claims, we consider three aggregate claim models with dependence in this paper. Model one considers the dependence caused by a common index among indexed insurance benefits; model two takes into account the correlation arisen from common fixed costs; model three covers both types of dependence. Two random variables,  $Y_1$  denoting a random index and  $Y_2$  denoting a random cost, form the center part in the above three dependence models and detailed discussions are given on how the aggregate claims amount interacts with these sources of dependence. Theoretical results of these aggregate claim distributions are derived and algorithms for computational purposes are also provided. Some numerical results are presented for the compound Poisson case together with discussions and comparisons regarding the three dependence cases.

**Keywords** Aggregate claim model, Recursive calculation, Dependence, Indexed benefits, Fixed costs, Inflation

**AMS 2010 subject classifications** 62P05

**DOI:** 10.19139/soic.v6i3.584

## 1. Introduction

One of the major research interests in actuarial science field is to model aggregate claims amount for various insurance business. The collective risk model is one well-known and widely used approach. A convenient yet unrealistic assumption is to assume independence between number of claims and individual claim amounts as well as among individual claim amounts. To make the collective risk models fit real insurance claims data better, researchers have been developing models with various types of dependence in the past few decades.

Early research outputs on dependence among numbers of claims include [3], who considered a general method of constructing a vector of  $p$  ( $p \geq 2$ ) dependent claim numbers from a vector of independent random variables, and derived formulas to compute the aggregate claim distribution for the book of  $p$  dependent classes of business. Motivated by these papers, researchers have developed collective risk models with dependence embedded in number of claims, for example [6], [10], [13], [18], [19], [20] and [22]. Readers can also refer to Chapter XIII of [7] for an informative survey.

Dependence between claim arrivals and claim amounts are also among the focuses of researchers in this field, and various types of dependence models have been proposed and employed in risk theory, see [1], [8] and [9] for example.

Another type of dependence is grounded on individual claim amounts only. Multivariate distributions and copulas have all been employed to model such correlations in the literature. [12] and [14] studied the stop-loss premiums

---

\*Correspondence to: Xueyuan Wu (Email: xueyuanw@unimelb.edu.au). Department of Economics, 111 Barry Street, University of Melbourne, VIC 3010, AUS.

of insurance portfolios with dependent risks and [11] examined stochastic bounds on sums of dependent risks. [2] used a simple idea of mixing to establish dependence among claim sizes and among claim inter-arrival times in collective risk models and obtained a number of explicit formulas for ruin probabilities and related quantities. More recently, [16] derived closed-form expressions for the probability distribution function of aggregated risks with multivariate dependent Pareto type II distributions proposed by [4] and [5], which is widely used in insurance and risk analysis.

In this paper we shall examine the aggregate claim distributions when the individual claim amounts follow a few simple dependence structures. These dependence structures are motivated by a couple of real-life insurance claim phenomena, i.e. the indexed insurance benefits and fixed costs in claim settlements. One typical implicit index is inflation, which affects almost all kinds of insurance products. The inflation risk is one of the most common risks that affect many industries, in particular the insurance business with long tails. Given a level of inflation, the individual and aggregate claims amount can be modeled as usual by scaling. However, when addressing the risk of inflation, actuaries also need to take into account the randomness of future inflation rates. This randomness not only changes the overall scale of insurance claims, but also increases the overall variance of the aggregate claims. Therefore, when modeling the aggregate claims amount we should introduce a common multiplicative factor, which is a random variable itself, if the individual claims generated by an insurance portfolio are explicitly or implicitly referring to a common economic index. Similarly, when settling claims there could be some common fixed costs that also vary in years, which lead to a common additive factor. Both phenomena bring dependence into the aggregate claim model alone or together.

Our main aim of this paper is to illustrate the impact of the above mentioned real-life dependence on aggregate claim distributions from computational point of view, so we shall only adopt discrete settings for the random variables involved in our models. The continuous case is also interesting but it would be much harder to develop recursive algorithms as in the discrete case. In practice, the method of discretization can be used to bridge the two cases nicely.

This paper is organized as follows. Three dependence structures are defined in Section 2 with some preliminary results. Some computational aspects are discussed in Section 3 where algorithms are provided to calculate the aggregate claim distributions. In Section 4, some numerical examples are given with detailed discussions.

## 2. Three Dependence Structures

In this section we consider three types of dependence structures: one with a common multiplicative factor, one with a common additive factor and one with common multiplicative and additive factors.

### 2.1. A Multiplicative Dependence

The first dependence structure is built on a common multiplicative factor:

$$S_1 = \sum_{i=1}^N X_i Y_1, \quad (1)$$

where  $N$  is a counting random variable (r.v.) denoting the number of claims with a probability function (p.f.)  $p_n, n \in \mathbb{N}$ ;  $X_i, i = 1, 2, \dots$ , are independent and identically distributed (i.i.d.) integer-valued random variables (r.v.'s) following a common p.f.  $f(x), x \in \mathbb{N}$ ;  $Y_1$  is a discrete r.v. which has a p.f.  $g_1(y), y \in \mathbb{A}$  where  $\mathbb{A}$  is a discrete set of values in  $\mathbb{R}^+$ . When  $N = 0$ ,  $S_1 = 0$ .

Assume that  $N, \{X_i\}_{i \in \mathbb{N}^+}$  and  $Y_1$  are all independent of each other. For the purpose of convenience, we denote  $X$  a generic r.v. of  $\{X_i\}$ . Further, we define  $\hat{f}(z) = \sum_{x=0}^{\infty} z^x f(x)$  to be the probability generating function (p.g.f.) of  $X$ . Similarly,  $\hat{p}(z)$  is the p.g.f. of  $N$ .

The model (1) is motivated by an insurance portfolio that pays indexed insurance benefits. When the individual indexed benefits  $X_i Y_1, i = 1, 2, \dots$ , are linked to the same random index  $Y_1$ , then the type of dependence emerged has the form of (1). In this context,  $S_1$  denotes the aggregate indexed claims amount and one can easily see that

it can be rewritten as  $S_1 = Y_1 Z$  where  $Z := \sum_{i=1}^N X_i$  is the original (non index-linked) aggregate claims amount with i.i.d. individual claims  $X_i, i = 1, 2, \dots$ . Here  $Y_1$  and  $Z$  are independent of each other.

Let  $S_1$  and  $Z$  have p.f.'s  $h_1(s)$  and  $h_Z(x), x \in \mathbb{N}, s \in \mathbb{B}$ , respectively, where

$$\mathbb{B} = \{s = y \times x : y \in \mathbb{A}, x \in \mathbb{N}\}.$$

Then we have:

$$\begin{aligned} E[S_1] &= E[Y_1]E[Z] = E[N]E[X]E[Y_1], \\ V[S_1] &= E[Y_1^2]E[Z^2] - E[Y_1]^2E[Z]^2 \\ &= E[Y_1^2]V[Z] + V[Y_1]E[Z]^2 \\ &= E[Y_1^2]\left(E[N]V[X] + V[N]E[X]^2\right) + V[Y_1]E[N]^2E[X]^2. \end{aligned} \quad (2)$$

Regarding the level of dependence among the individual indexed claim amounts  $X_i Y_1, i = 1, 2, \dots$ , measured by their covariances and correlation coefficients, we have, for  $i, j \in \mathbb{N}^+$  and  $i \neq j$ ,

$$\begin{aligned} COV[X_i Y_1, X_j Y_1] &= E[X_i X_j Y_1^2] - E[X]^2 E[Y_1]^2 = E[X]^2 V[Y_1], \\ \rho_1 &= \frac{COV[X_i Y_1, X_j Y_1]}{V[X Y_1]} = \frac{E[X]^2 V[Y_1]}{V[X]E[Y_1^2] + E[X]^2 V[Y_1]} \\ &= \frac{V[Y_1]}{V[X] \frac{E[Y_1^2]}{E[X]^2} + V[Y_1]} = \frac{1}{1 + CV_X^2(1 + CV_{Y_1}^{-2})}, \end{aligned} \quad (3)$$

where  $CV_X (= \sqrt{V[X]}/E[X])$  and  $CV_{Y_1}$  are the coefficients of variation of  $X$  and  $Y_1$  respectively and  $\rho_1$  is the correlation coefficient of each pair of individual indexed claim amounts.

We propose a general method regarding  $h_1(s), s \in \mathbb{B}$ , that can be useful for calculating its values. For  $y \in \mathbb{A}$ , we define  $\mathbb{N}_y = \{0, y, 2y, \dots\}$ . Clearly,  $\mathbb{N}_y \subseteq \mathbb{B}$ . Let  ${}_1 f_y(x), y \in \mathbb{A}, x \in \mathbb{N}_y$ , be the p.f. of the r.v.  $yX$ . Then for  $x \in \mathbb{N}_y$ ,

$${}_1 f_y(x) = f\left(\frac{x}{y}\right). \quad (4)$$

Let  $\{{}_1 h_Z(x; y)\}_{x \in \mathbb{N}_y, y \in \mathbb{A}}$  denote the p.f. of the compound distribution consisting of  $\{p_n\}$  and  $\{{}_1 f_y(x)\}$ , and  ${}_1 h_Z(s; y) \equiv 0$  when  $s \notin \mathbb{N}_y$ . Further,  ${}_1 h_Z(s; y) = h_Z\left(\frac{s}{y}\right)$  when  $s \in \mathbb{N}_y$ . Then we have, for  $s \in \mathbb{B}$  and  $s > 0$ ,

$$\begin{aligned} h_1(s) &= \Pr(Y_1 Z = s) = E\left[\Pr\left(Z = \frac{s}{Y_1} \mid Y_1\right)\right] \\ &= \sum_{y \in \mathbb{A}} {}_1 h_Z(s; y) g_1(y) = \sum_{y \in \mathbb{A}; y \leq s} {}_1 h_Z(s; y) g_1(y) \end{aligned} \quad (5)$$

$$= \sum_{y \in \mathbb{A}; y \leq s} \mathbf{1}_{\{s \in \mathbb{N}_y\}} h_Z\left(\frac{s}{y}\right) g_1(y) \quad (6)$$

$$= \sum_{y \in \mathbb{A}; y \leq s} \mathbf{1}_{\{\lfloor \frac{s}{y} \rfloor = \frac{s}{y}\}} h_Z\left(\frac{s}{y}\right) g_1(y) \quad (7)$$

with  $h_1(0) = h_Z(0)$ , where  $\lfloor x \rfloor$  is the floor function of  $x$ . The second equality in (5) holds because the largest value of  $Y_1$  that does not void  $\{Y_1 Z = s\}$  is  $s$ . The result (7) specifies more clearly how to judge whether or not  $s \in \mathbb{N}_y$  when programming.

## 2.2. An Additive Dependence

The second dependence structure is built on a common additive factor:

$$S_2 = \sum_{i=1}^N (X_i + Y_2), \quad (8)$$

where  $N, \{X_i\}$  are the same as in model (1).  $Y_2$  is a r.v. independent of  $N$  and  $\{X_i\}$ , and has p.f.  $g_2(y), y \in \mathbb{N}$ . Note that  $S_2 = Z + NY_2$  but here  $Z$  and  $NY_2$  are dependent of each other.

Different from the model (1), (8) is motivated by an insurance portfolio with fixed costs (like paperwork, admin or transaction costs etc) on claim settlements. In a general setting, one could use a r.v. to denote this fixed cost when settling a claim that varies year by year. As settling each individual claim will incur the same fixed costs in the same time period, a dependence arises as in model (8). This is also why we assume  $Y_2$  only takes non-negative integer values.

Similarly, we can obtain the following results based on model (8):

$$E[S_2] = E[Z] + E[N]E[Y_2] = E[N](E[X] + E[Y_2]), \tag{9}$$

$$\begin{aligned} V[S_2] &= E[V[S_2|N]] + V[E[S_2|N]] \\ &= E[NV[X] + N^2V[Y_2]] + V[N(E[X] + E[Y_2])] \\ &= (E[X] + E[Y_2])^2V[N] + E[N]V[X] + E[N^2]V[Y_2]. \end{aligned} \tag{10}$$

Again, we consider the level of dependence among the individual claim sizes with fixed costs, for  $i, j \in \mathbb{N}^+$  and  $i \neq j$ ,

$$\begin{aligned} COV[X_i + Y_2, X_j + Y_2] &= V[Y_2], \\ \rho_2 = \frac{COV[X_i + Y_2, X_j + Y_2]}{V[X + Y_2]} &= \frac{V[Y_2]}{V[X] + V[Y_2]}. \end{aligned} \tag{11}$$

Let  $h_2(s)$  be the p.f. of  $S_2$ . To calculate  $h_2(s)$ , we shall derive a general result based on model (8). Similar to  ${}_1f_y(x)$ , we define  $g_2(n; y)$  as the p.f. of the r.v.  $nY_2, n \in \mathbb{N}^+$ . One can see that

$$g_2(n; y) = \begin{cases} g_2(k) & \text{if } y = kn, \quad k \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

In particular,  $g_2(0; y)$  is a degenerate p.f. at the value 0. Then we have the following result, for  $s \in \mathbb{N}^+$ :

$$h_2(s) = \sum_{n=1}^{\infty} f^{*n} * g_2(n; s)p_n, \tag{13}$$

where  $f^{*n}(x)$  is the  $n$ -fold convolution of  $f(x)$  and  $f^{*n} * g_2(n; x)$  is the convolution of  $f^{*n}(x)$  and  $g_2(n; x)$ . Also,  $h_2(0) = p_0 + g_2(0) \sum_{n=1}^{\infty} f^n(0)p_n$ .

Alternatively, let  ${}_2f_y(x)$  be the p.f. of  $X + y$ . Define set  $\mathbb{N}^{y+} = \{y, y + 1, y + 2, \dots\}, y \in \mathbb{N}$ , then we have

$${}_2f_y(x) = \begin{cases} f(x - y) & x \in \mathbb{N}^{y+}, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly,  ${}_2f_0(x) = f(x)$ . Let  $\{{}_2h_Z(x; y)\}_{x, y \in \mathbb{N}}$  denote the p.f. of the compound distribution consisting of  $\{p_n\}$  and  $\{{}_2f_y(x)\}$ , and clearly  ${}_2h_Z(x; y) = 0$  for all  $0 < x < y$ . Then we have the second result regarding  $h_2(s)$ , for  $s \in \mathbb{N}^+$ ,

$$h_2(s) = \sum_{y \in \mathbb{N}} {}_2h_Z(s; y)g_2(y) = \sum_{y=0}^s {}_2h_Z(s; y)g_2(y). \tag{14}$$

Obviously, result (14) is of higher computational importance than (13) due to its finite summation.

### 2.3. A Two-way Dependence

When both the random index and random fixed costs are in place, we propose the third dependence structure as follows:

$$S_3 = \sum_{i=1}^N (Y_1 X_i + Y_2), \tag{15}$$

where  $N, \{X_i\}_{i \in \mathbb{N}^+}$  and  $Y_j, j = 1, 2$ , have been defined in the above two models, and here they are all independent of each other. Under model (15), the individual claim sizes  $X_i, i \in \mathbb{N}^+$ , are affected not only by the common multiplicative factor  $Y_1$ , but also by the common additive factor  $Y_2$ , so it forms a two-way dependence structure. Obviously,  $S_3$  has a relationship with  $Z$  in the form of  $S_3 = Y_1 Z + NY_2$  where (1) and (8) are two special cases when we take  $Y_2 \equiv 0$  and  $Y_1 \equiv 1$  respectively.

Firstly, we show some basic results:

$$\begin{aligned}
 E[S_3] &= E[Y_1]E[Z] + E[N]E[Y_2] = E[N](E[X]E[Y_1] + E[Y_2]), & (16) \\
 V[S_3] &= E[V[Y_1 Z + NY_2|N]] + V[E[Y_1 Z + NY_2|N]] \\
 &= E[V[Y_1 Z|N]] + E[V[NY_2|N]] + V[N(E[Y_1]E[X] + E[Y_2])] \\
 &= E[E[Y_1^2]V[Z|N] + V[Y_1]E[Z|N]^2] + E[N^2]V[Y_2] \\
 &\quad + V[N](E[Y_1]E[X] + E[Y_2])^2 \\
 &= E[Y_1^2]E[N]V[X] + V[Y_1]E[N^2]E[X]^2 + E[N^2]V[Y_2] \\
 &\quad + V[N](E[Y_1]E[X] + E[Y_2])^2. & (17)
 \end{aligned}$$

For  $i, j \in \mathbb{N}^+$  and  $i \neq j$ , we have

$$\begin{aligned}
 COV[Y_1 X_i + Y_2, Y_1 X_j + Y_2] &= COV[X_i Y_1, X_j Y_1] + V[Y_2] \\
 &= E^2[X]V[Y_1] + V[Y_2], \\
 \rho_3 = \frac{COV[Y_1 X_i + Y_2, Y_1 X_j + Y_2]}{V[X Y_1 + Y_2]} &= \frac{E[X]^2 V[Y_1] + V[Y_2]}{E[Y_1^2]V[X] + E[X]^2 V[Y_1] + V[Y_2]} \\
 &= \frac{1}{1 + \frac{1 + CV_{Y_1}^{-2}}{\beta + CV_X^{-2}}}, & (18)
 \end{aligned}$$

where  $\beta = \frac{V[Y_2]}{V[X]V[Y_1]}$ .

Because of the two-way dependence existed in model (15), the valuation spaces of  $Y_1$  and  $Y_2$ , i.e. the positive real-valued discrete set  $\mathbb{A}$  and the non-negative integer set  $\mathbb{N}$ , are mixed together, which brings complications into the computational aspect of the new model. The formulas provided in the previous two models, (6) and (14), can not be employed directly and more assumptions are needed for the purpose of simplification.

To calculate the p.f. of  $S_3$ , denoted by  $\{h_3(s)\}_{s \in \mathbb{B}^*}$  where  $\mathbb{B}^*$  is the domain of  $S_3$  and yet to determine, we shall restrict the domain of  $Y_1$  to a set of positive rational numbers, denoted by  $\mathbb{A}_r \subseteq \mathbb{Q}^+$  and  $\mathbb{Q}^+$  denotes the whole set of positive rational numbers. In addition, we assume that  $\mathbb{A}_r$  only has a finite number of elements, denoted by  $M$ . Then we can express all its elements in terms of fractions in simplest form, i.e.  $\mathbb{A}_r = \{\frac{\delta_i}{\eta_i}, i = 1, \dots, M\}$  where  $\delta_i, \eta_i \in \mathbb{N}^+$ .

Conditional on  $Y_1 = \frac{\delta_i}{\eta_i}$  and  $Y_2 = y, y \in \mathbb{N}, i = 1, \dots, M$ , our model (15) becomes

$$S_3 = \sum_{j=1}^N (\frac{\delta_i}{\eta_i} X_j + y) = \frac{1}{\eta_i} \sum_{j=1}^N (\delta_i X_j + y \eta_i). \tag{19}$$

In the above new expression of  $S_3$ ,  $\sum_{j=1}^N (\delta_i X_j + y \eta_i)$  is an integer-valued random variable and so the domain of  $S_3, \mathbb{B}^*$ , can be defined as

$$\mathbb{B}^* = \bigcup_{i=1}^M \left\{ s = \frac{x}{\eta_i} : x \in \mathbb{N} \right\}.$$

Note that the real domain of  $S_3$  is a subset in the above defined set  $\mathbb{B}^*$  that contains a number of elements with zero probability mass. However, these unnecessary elements can be easily eliminated when conducting numerical calculations.

For  $x \in \mathbb{N}$ , we define the p.f. of  $\delta_i X_j + y \eta_i$  by

$${}_3f_{\delta_i, \eta_i, y}(x) = \begin{cases} f\left(\frac{x-y\eta_i}{\delta_i}\right) & \frac{x-y\eta_i}{\delta_i} \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

Clearly,  ${}_3f_{1,1,y}(x) = {}_2f_y(x)$  and  ${}_3f_{\delta, \eta, 0}(x) = {}_1f_\delta(x)$ . Let  $\{{}_3h_Z(x; \delta_i, \eta_i, y)\}$ ,  $x \in \mathbb{N}$ , denote the p.f. of the compound distribution consisting of  $\{p_n\}$  and  $\{{}_3f_{\delta_i, \eta_i, y}(x)\}$ , then we have, for  $s \in \mathbb{B}^*$  and  $s > 0$ ,

$$\begin{aligned} h_3(s) &= \sum_{y \in \mathbb{N}} g_2(y) \sum_{i=1}^M 1_{\{s\eta_i \in \mathbb{N}\}} {}_3h_Z(s\eta_i; \delta_i, \eta_i, y) g_1\left(\frac{\delta_i}{\eta_i}\right) \\ &= \sum_{y=0}^s g_2(y) \sum_{i=1}^M 1_{\{s\eta_i \in \mathbb{N}\}} {}_3h_Z(s\eta_i; \delta_i, \eta_i, y) g_1\left(\frac{\delta_i}{\eta_i}\right) \end{aligned} \tag{21}$$

with  $h_3(0) = p_0 + g_2(0) \sum_{n=1}^\infty f^n(0) p_n$ .

**Remark.** An alternative structure of the two-way dependence can be defined as

$$S'_3 = \sum_{i=1}^N Y_1(X_i + Y_2) = Y_1(Z + NY_2),$$

which allows the interaction between factors  $Y_1$  and  $Y_2$ . One interpretation of this slightly different dependence structure is that, in practice, the fixed costs normally are affected by inflation as well. However, this version is easier to compute than (15), as one can deal with  $Z + NY_2$  first and then incorporate  $Y_1$  in the second step. Therefore, we shall only discuss the version given in (15) within the rest of this paper.

### 3. Computational Aspects of the Models

From the key results (6), (14) and (21) obtained in Section 2 we can see that to calculate the aggregate claim distributions  $h_i(s)$ ,  $i = 1, 2, 3$ , we need to calculate the aggregate claim distributions  $h_Z(x)$ ,  ${}_2h_Z(x; y)$  and  ${}_3h_Z(x; \delta_i, \eta_i, y)$  first, which are all compound distributions under independence setting. In the actuarial literature, there have been many papers discussing how to compute the aggregate claim distributions under independence setting and recursive calculation is one typical approach. Based on the well-known Panjer's recursion, researchers have developed a number of generalized recursive methods for various types of claim number distributions, to name a few, the  $(a, b, 0)$  class, the  $(a, b, 1)$  class and the  $(A, B, 0)$  class, etc. Useful references include [15], [17], [21] and the references therein.

In the following, we shall discuss how to calculate the distributions  $h_i(s)$ ,  $i = 1, 2, 3$ , using the  $(a, b, 0)$  class as an example, where

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \quad n \in \mathbb{N}^+.$$

The only non-trivial distributions in this class are Poisson, binomial and negative binomial.

#### 3.1. The Calculation of $h_1(s)$

According to the Panjer's Recursion,  $h_Z(x)$  satisfies

$$h_Z(x) = \frac{1}{1 - af(0)} \sum_{k=1}^x \left(a + \frac{bk}{x}\right) f(k) h_Z(x - k), \quad x \in \mathbb{N}^+, \tag{22}$$

with  $h_Z(0) = \hat{p}(f(0))$ . Bear in mind that when dealing with an infinite series, we often omit the tail piece according to certain accuracy requirements. This argument applies within the rest of this paper.

To calculate  $h_1(s), s \in \mathbb{B}$ , from (6) we can see that, for any given  $y \in \mathbb{A}, y \leq s$ , we need to check whether  $s$  is a multiple of  $y$ , i.e.  $s \in \mathbb{N}_y$ . If the answer is yes, then the probability  $h_Z(\frac{s}{y})$  associated with the  $y$  value is involved in  $h_1(s)$  and can be determined using the Panjer's recursion. Quite often, the number of relevant  $y$  values for each  $s \in \mathbb{B}$  is very small, so the additional computations from  $h_Z$  to  $h_1$  is very reasonable. For example, if  $\mathbb{A} = \{1.05, 1.1, 1.15\}$ , for  $s = 10.5$  and  $s = 23.1$  we have

- $h_1(10.5) = h_Z(10)g_1(1.05)$ ; and
- $h_1(23.1) = h_Z(22)g_1(1.05) + h_Z(21)g_1(1.1)$ .

More importantly, the computation of  $h_1(s)$  does not require values of  $h_Z(k)$  for  $k > \frac{s}{y^*}$  where  $y^*$  is the minimum realization of  $Y_1$ . Also, when  $Y_1 \equiv 1, h_1(s) = h_Z(s)$  for  $s \in \mathbb{N}$ . A suggested algorithm to calculate  $h_1(s)$  is given below:

- Step 1. Using  $\mathbb{N}$  and the given domain  $\mathbb{A}$  to generate domain  $\mathbb{B}$ . There should be no duplicate values in  $\mathbb{B}$ .
- Step 2. Calculate and store  $h_Z(x), x \in \mathbb{N}$  using appropriate methods. In this case, the Panjer's recursion is an obvious choice.
- Step 3. For each  $s \in \mathbb{B}$  and  $s > 0$ , calculate  $h_1(s)$  using formula (7).

Note that to obtain  $H_1(s), s \in \mathbb{N}$ , we need to calculate  $H_1(s) = \sum_{x \in \mathbb{B}; x \leq s} h_1(x)$ .

**3.2. The Calculation of  $h_2(s)$**

The calculation of  $h_2(s)$  is based on  ${}_2h_Z(x; y)$  that involves a modified individual claim amount distribution  ${}_2f_y(x)$ . It is actually a type of truncation by letting  ${}_2f_y(x) = f(x - y)$  for  $x \in \mathbb{N}^{y+}$ . When  $\{p_n\}$  belongs to the  $(a, b, 0)$  class and  $y \in \mathbb{N}^+, {}_2h_Z(x; y)$  satisfies

$${}_2h_Z(x; y) = \sum_{k=y}^x \left( a + \frac{bk}{x} \right) f(k - y) {}_2h_Z(x - k; y), \quad x \geq y, \tag{23}$$

with  ${}_2h_Z(0; y) = p_0$  and  ${}_2h_Z(x; y) = 0$  for  $x = 1, \dots, y - 1$ . Further, when  $Y_2 \equiv 0, h_2(x) = h_Z(x)$  for  $x \in \mathbb{N}$ . Assume that  $Y_2$  takes values in  $\mathbb{D} \subseteq \mathbb{N}$ . When  $\mathbb{D}$  has only limited number of elements, a reasonable algorithm is given below:

- Step 1. For each  $y \in \mathbb{D}$ , generate  ${}_2f_y(x), x \in \mathbb{N}$ .
- Step 2. Calculate and store  ${}_2h_Z(x; y), x \in \mathbb{N}$  using the Panjer's recursion. Note that we need to store  ${}_2h_Z(x; y), x \in \mathbb{N}$ , for every  $y \in \mathbb{D}$ .
- Step 3. For each  $s \in \mathbb{B}$ , calculate

$$h_2(s) = \sum_{y \in \mathbb{D}; y \leq s} {}_2h_Z(s; y)g_2(y).$$

**3.3. The Calculation of  $h_3(s)$**

The calculation of  $h_3(x)$  is much more complicated, which is a combination of the above two cases. Similar to Model 2, we assume that  $Y_2$  takes values in  $\mathbb{D} \subseteq \mathbb{N}$  where  $\mathbb{D}$  has only limited number of elements. According to (20) and (21) we propose an algorithm as follows:

- Step 1. For the given domain of  $Y_1, \mathbb{A}_r$ , determine  $\delta_i$  and  $\eta_i$  for all  $i = 1, \dots, M$ .
- Step 2. Given  $Y_2 = y \in \mathbb{D}$ , for each  $i = 1, \dots, M$ , generate  ${}_3f_{\delta_i, \eta_i, y}(x), x \in \mathbb{N}$ .

Step 3. Calculate and store  ${}_3h_Z(x; \delta_i, \eta_i, y), i = 1, \dots, M, x \in \mathbb{N}^+$ , using the formula

$${}_3h_Z(x; \delta_i, \eta_i, y) = \frac{\sum_{k=1}^x (a + \frac{bk}{x}) \times {}_3f_{\delta_i, \eta_i, y}(k) \times {}_3h_Z(x - k; \delta_i, \eta_i, y)}{1 - a \times {}_3f_{\delta_i, \eta_i, y}(0)}$$

with  ${}_3h_Z(0; \delta_i, \eta_i, y) = \hat{p}({}_3f_{\delta_i, \eta_i, y}(0))$ . Note that we need to store  ${}_3h_Z(x; \delta_i, \eta_i, y), x \in \mathbb{N}$ , for all  $y \in \mathbb{D}$  and  $i = 1, \dots, M$ .

Step 4. For  $s \in \mathbb{B}^*$ , calculate

$$h_3(s) = \sum_{y=0}^s g_2(y) \sum_{i=1}^M 1_{\{s\eta_i \in \mathbb{N}\}} {}_3h_Z(s\eta_i; \delta_i, \eta_i, y) g_1\left(\frac{\delta_i}{\eta_i}\right).$$

### 4. Numerical Studies

In this section, we shall provide some numerical examples to demonstrate the impact of the three dependence structures proposed in Section 2 on the aggregate claim distributions. To make the following numerical discussions more reasonable, we shall propose a condition of fairness, i.e. the aggregate claim models under discussion all have equal means.

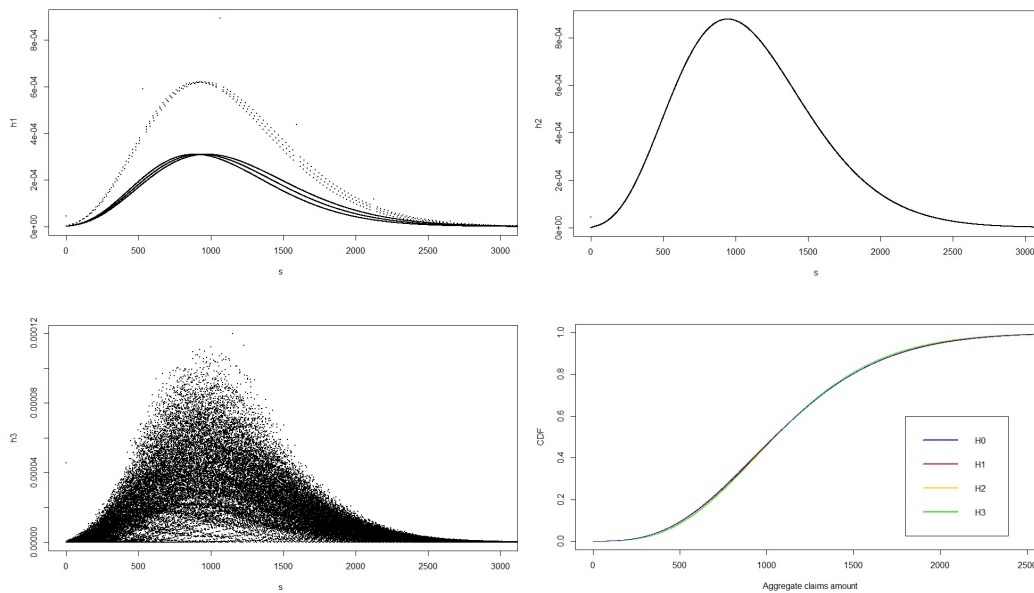


Figure 1. The graph of  $h_i, i = 1, 2, 3$ , and  $H_i, i = 0, 1, 2, 3$ , in Case 1

In the following, we shall consider four aggregate claim models,  $S_i, i = 0, 1, 2, 3$ , where  $S_1, S_2, S_3$  have been defined in Section 2 and  $S_0 = Z$ , which is the aggregate claims amount with i.i.d. individual claims. We first make the following assumptions:

- $E[N] = 10$ ;
- the i.i.d. random variables  $\{X_i\}$  in all models follow the geometric distribution with p.f.

$$f(x) = (1 - q)q^{x-1}, \quad x \in \mathbb{N}^+.$$



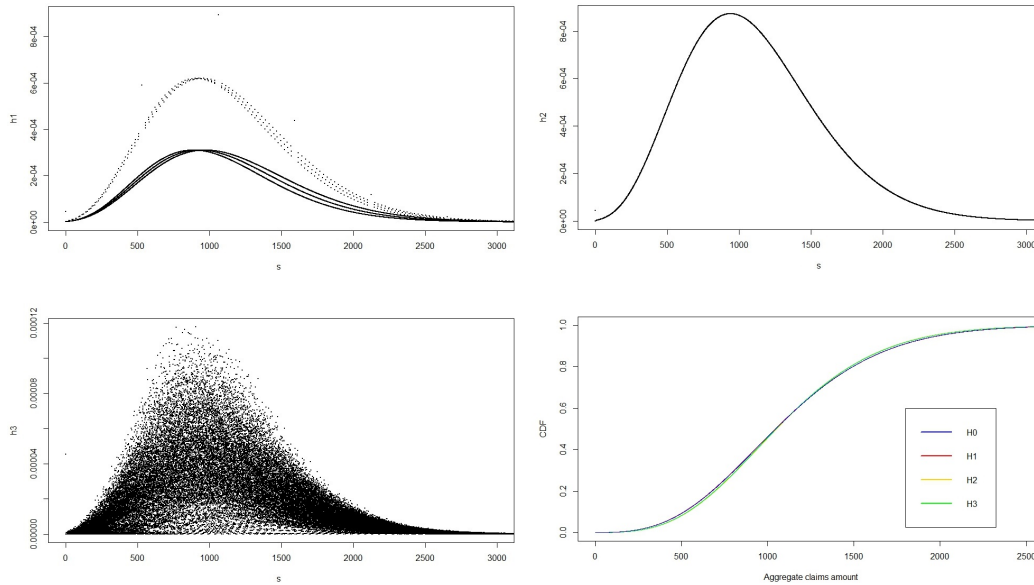


Figure 2. The graph of  $h_i, i = 1, 2, 3$ , and  $H_i, i = 0, 1, 2, 3$ , in Case 2

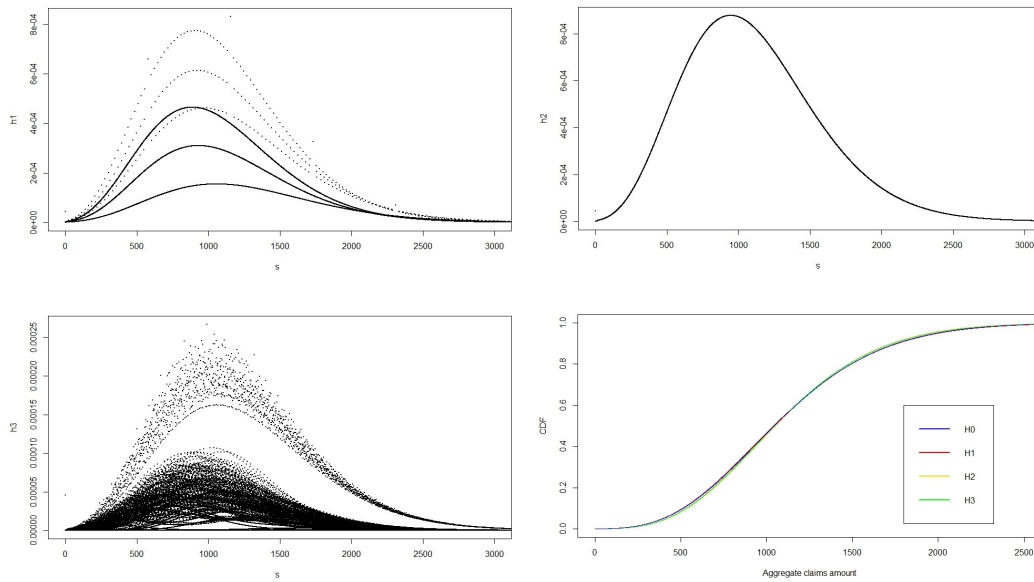


Figure 3. The graph of  $h_i, i = 1, 2, 3$ , and  $H_i, i = 0, 1, 2, 3$ , in Case 3

Let  $q_i$  be the geometric parameter in model  $S_i, i = 0, 1, 2, 3$ , and let  $q_0 = \frac{109}{110}, q_1 = q_2 = 0.99, q_3 = 0.989$ ;

- $E[Y_1] = 1.1$  and  $E[Y_2] = 10$ .

Under the above assumptions one can easily verify that  $E[S_i] = 1100, i = 0, 1, 2, 3$ .

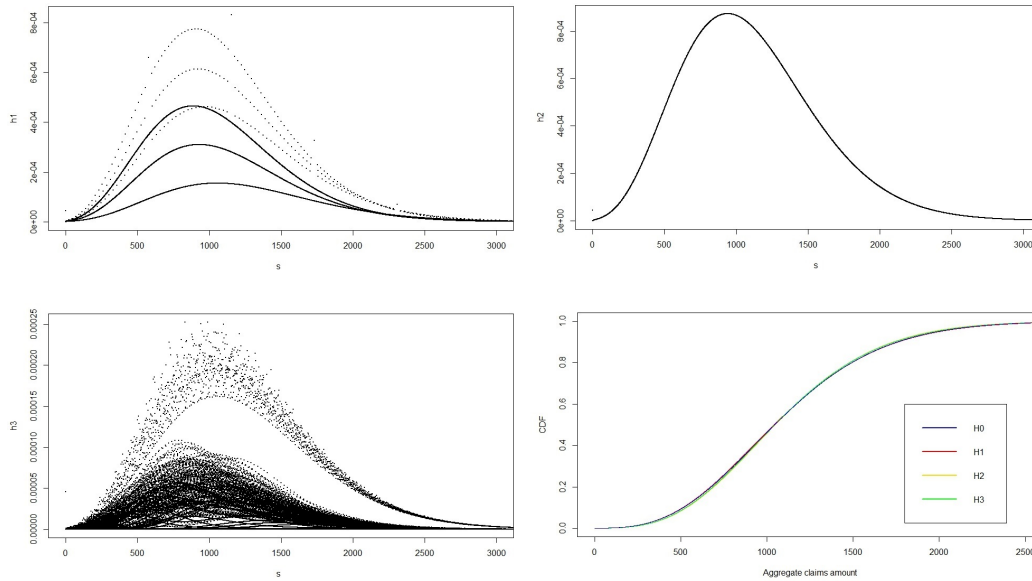


Figure 4. The graph of  $h_i, i = 1, 2, 3$ , and  $H_i, i = 0, 1, 2, 3$ , in Case 4

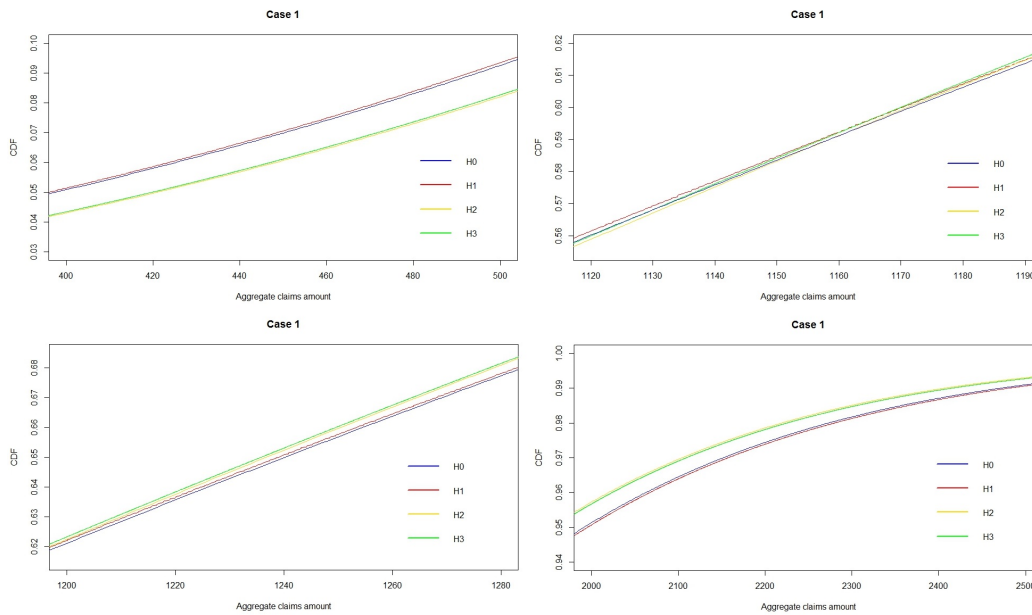


Figure 5. The detailed c.d.f. of  $S_0, S_1, S_2$  and  $S_3$  in Case 1

As mentioned in Section 2.1,  $Y_1$  is likely to be some real-life index when considering the indexed insurance benefits, so we suggest a distribution with a finite support for  $Y_1$ . Bear in mind that here we are considering the aggregate claims amount for a given time period, so the distribution of the index may change over time. Also, Section 2.2 gave one motivation for  $Y_2$ , i.e. the random fixed cost on claim settlements for an insurance portfolio. Again, a distribution with a finite support should be a reasonable assumption. To better demonstrate the impact of different level of dependence on the aggregate claim distributions, we consider the following four cases:

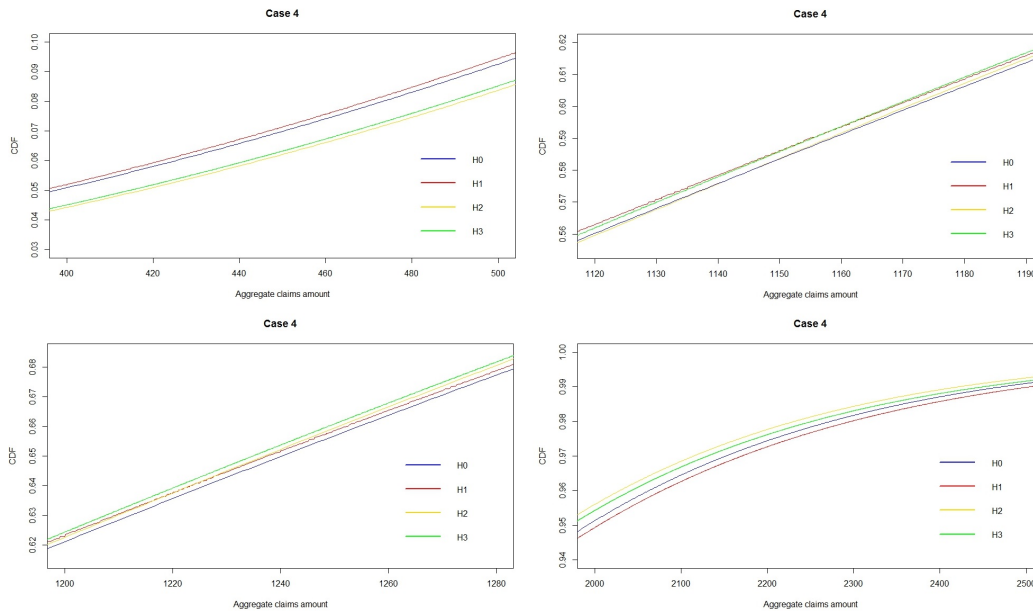


Figure 6. The detailed c.d.f. of  $S_0, S_1, S_2$  and  $S_3$  in Case 4

- Case 1.  $\mathbb{A} = \{1.05, 1.1, 1.15\}, \{g_1(1.05) = \frac{1}{3}, g_1(1.1) = \frac{1}{3}, g_1(1.15) = \frac{1}{3}\};$   
 $\mathbb{D} = \{5, 10, 15\}, \{g_2(5) = \frac{1}{3}, g_2(10) = \frac{1}{3}, g_2(15) = \frac{1}{3}\};$
- Case 2.  $\mathbb{A} = \{1.05, 1.1, 1.15\}, \{g_1(1.05) = \frac{1}{3}, g_1(1.1) = \frac{1}{3}, g_1(1.15) = \frac{1}{3}\};$   
 $\mathbb{D} = \{5, 10, 25\}, \{g_2(5) = \frac{1}{2}, g_2(10) = \frac{1}{3}, g_2(25) = \frac{1}{6}\};$
- Case 3.  $\mathbb{A} = \{1.05, 1.1, 1.25\}, \{g_1(1.05) = \frac{1}{2}, g_1(1.1) = \frac{1}{3}, g_1(1.25) = \frac{1}{6}\};$   
 $\mathbb{D} = \{5, 10, 15\}, \{g_2(5) = \frac{1}{3}, g_2(10) = \frac{1}{3}, g_2(15) = \frac{1}{3}\};$
- Case 4.  $\mathbb{A} = \{1.05, 1.1, 1.25\}, \{g_1(1.05) = \frac{1}{2}, g_1(1.1) = \frac{1}{3}, g_1(1.25) = \frac{1}{6}\};$   
 $\mathbb{D} = \{5, 10, 25\}, \{g_2(5) = \frac{1}{2}, g_2(10) = \frac{1}{3}, g_2(25) = \frac{1}{6}\}.$

**Example 1.** In the first example, we assume that  $N$  follows a Poisson(10) distribution. Table 1 summarises some empirical results regarding the aggregate claim amounts  $S_i, i = 0, \dots, 3$ , under the above four cases. In Table 1, there is one italic correlation coefficient in each case which emphasizes the dependence model that is significantly different from the other two in respect of the degree of dependence. Also, tripling the variance of  $Y_1$  or  $Y_2$  roughly doubled  $\rho_3$  while tripling the variance of both  $Y_1$  and  $Y_2$  tripled  $\rho_3$  overall. Further, under the above assumptions, the degree of dependence in the two-way dependence model  $S_3$  is always the largest in all four cases which is consistent with the fact that it has two sources of correlation embedded in the model.

To better explain the computation procedure of  $h_i(s), i = 1, 2, 3$ , we further specify the domains involved in the calculations as follows:

- Case 1&2.  $\mathbb{B} = \{0, 1.05, 1.10, 1.15, 2.10, 2.20, 2.30, \dots\}; \mathbb{A}_r = \{\frac{21}{20}, \frac{11}{10}, \frac{23}{20}\};$   
 $\mathbb{B}^* = \{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, \dots\};$
- Case 3&4.  $\mathbb{B} = \{0, 1.05, 1.10, 1.25, 2.10, 2.20, 2.50, \dots\}; \mathbb{A}_r = \{\frac{21}{20}, \frac{11}{10}, \frac{5}{4}\};$   
 $\mathbb{B}^* = \{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, \dots\}.$

Having all parameters assumed and making use of the algorithms given in Section 3, we calculate the p.f. of  $S_i, i = 0, \dots, 3$ , i.e.  $h_Z$  and  $h_i, i = 1, 2, 3$ , in R. It can be seen that the Panjer’s recursion with  $a = 0$  and  $b = 10$  is

Table 1. Some preliminary results

	Case 1	Case 2	Case 3	Case 4
$V[Y_1]$	0.00167	0.00167	0.00500	0.00500
$CV_{Y_1}$	0.03711	0.03711	0.06428	0.06428
$V[Y_2]$	16.66667	50.00000	16.66667	50.00000
$CV_{Y_2}$	0.40825	0.70711	0.40825	0.70711
$V[S_1]$	242,788	242,788	246,785	246,785
$\rho_1$	0.00139	0.00139	0.00414	0.00414
$V[S_2]$	221,833	225,500	221,833	225,500
$\rho_2$	0.00168	0.00503	0.00168	0.00503
$V[S_3]$	223,385	227,051	226,688	230,354
$\rho_3$	0.00306	0.00640	0.00581	0.00911

the core in all calculations. The details of how to conduct the computational tasks have been given in Section 3, so we need not repeat here. Based on the calculated results, we draw the following Figures 1-6 that visually present the p.f. and cumulative distribution function (c.d.f.) of  $S_i$ ,  $i = 0, \dots, 3$ , under the above presumed four cases. According to the descriptive statistics summarised in Table 1, one can see that the multiplicative factor  $Y_1$  in Case 3 and 4 have higher variations and the additive factor  $Y_2$  in Case 2 and 4 have higher variations. Overall, the variances of  $S_i$ ,  $i = 1, 2, 3$ , in Case 1 are the lowest and the variances of  $S_i$  in Case 4 are the highest. One can see from the figures that:

- First of all, the non-smooth graphs of  $h_1$  and  $h_3$  look strange comparing with those of  $h_2$ . It is the consequence of the multiplicative factor  $Y_1$ , as different combinations of  $Z$  and  $Y_1$  values could obtain the same value for  $S_1$ . An example has been given in Section 3.1. Therefore, the probability masses of  $Z$  have been reallocated under the framework of  $S_1$  or  $S_3$  and resulted in the non-smooth dot plots of  $h_1$  and  $h_3$ . On the contrary,  $Y_2$  has only a shifting effect on the aggregate claim distributions, so it did not affect the smoothness in the graphs of  $h_2$ .
- The impact of the multiplicative factor  $Y_1$  on the shape of p.f.'s is vertical. For Case 3 and 4 which have more variable  $Y_1$  values, the p.f.'s of  $S_1$  and  $S_3$  change a lot vertically comparing with the other two cases. Also, together with the additive factor  $Y_2$ , the p.f. of  $S_3$  fluctuate a lot such that the probability masses of  $h_3$  almost fill in certain regions. The number of values that  $S_3$  can take given  $S_3 \leq s$  is much bigger than the maximum value  $s$ .
- No matter how fluctuating the p.f. curves are, the c.d.f.'s are all smooth functions. Under the presumed numerical cases and the scale of the figures, the differences among the four c.d.f.'s are not significant in Figure 1-4.
- To better show the relationships among the four c.d.f.'s we show some detailed graphs for Case 1 and Case 4, which are the cases with the biggest difference, in Figure 5 and 6 respectively. One can see that for smaller aggregate claims amount, under our model assumptions, the independent model  $S_0$  tends to underestimate their associated probabilities if the multiplicative dependence exists; it tends to overestimate the associated probabilities if either the additive or the two-way dependence exists. Figure 5 and 6 also show that the impact of the additive factor  $Y_2$  is more obvious when the multiplicative factor  $Y_1$  has a small variance, as the curves for  $S_0$  and  $S_1$  (without additive factor) are more separated from  $S_2$  and  $S_3$  (with additive factor) in Case 1 than in Case 4.
- Figure 6 shows the same trends as the ones shown by Figure 5. Due to the increased variance of  $Y_1$ , its impact on the c.d.f.'s is more easily seen.

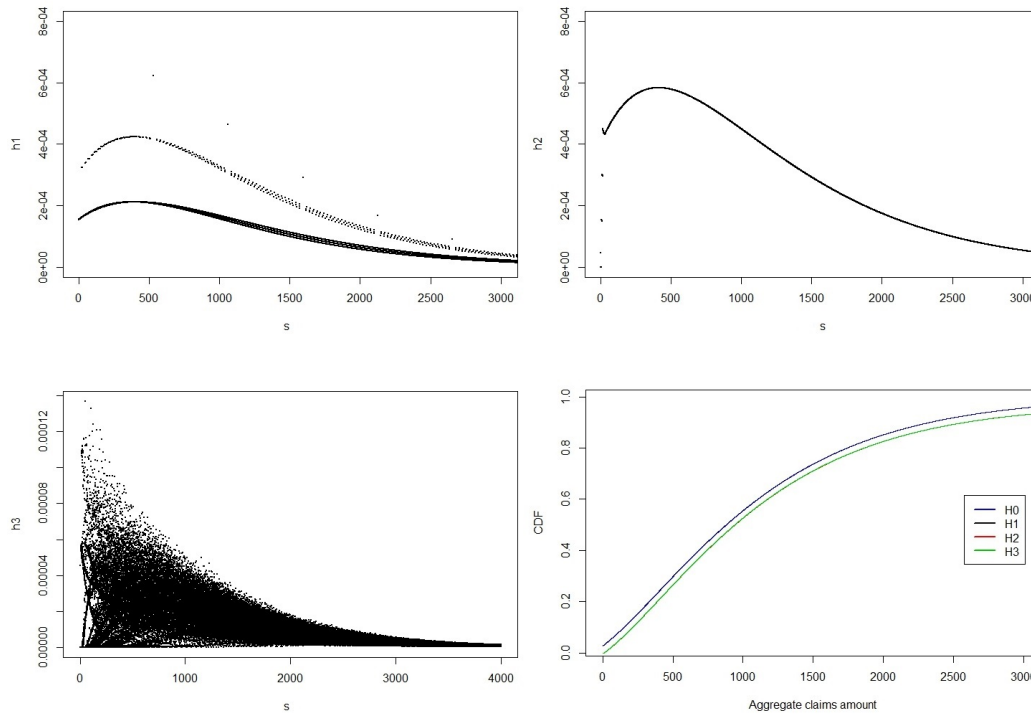


Figure 7. The graph of  $h_i, i = 1, 2, 3$ , and  $H_i, i = 0, 1, 2, 3$ , in Case 1

**Example 2.** To further our findings in the previous compound Poisson case, we explore the compound negative binomial case in the following. Assume that  $N$  follows a negative binomial distribution with size 2 and probability  $\frac{1}{6}$ . It can be verified that  $E[N] = 10$  and  $V[N] = 60$ . Table 2 summarises the empirical results regarding the aggregate claim amounts  $S_i, i = 0, \dots, 3$ , under the four cases specified in Example 1. One can see that the aggregate claims amounts  $S_i, i = 1, 2, 3$ , have much higher variances under the compound negative binomial assumption than under the compound Poisson assumption, but the correlation among claim sizes is not affected by the number of claims assumption due to the independence between number of claims and individual claim sizes.

Table 2. Some preliminary results

	Case 1	Case 2	Case 3	Case 4
$V[Y_1]$	0.00167	0.00167	0.00500	0.00500
$CV_{Y_1}$	0.03711	0.03711	0.06428	0.06428
$V[Y_2]$	16.66667	50.00000	16.66667	50.00000
$CV_{Y_2}$	0.40825	0.70711	0.40825	0.70711
$V[S_1]$	848,622	848,622	854,285	854,285
$\rho_1$	0.00139	0.00139	0.00414	0.00414
$V[S_2]$	827,667	833,000	827,667	833,000
$\rho_2$	0.00168	0.00503	0.00168	0.00503
$V[S_3]$	829,907	835,240	834,587	839,920
$\rho_3$	0.00306	0.00640	0.00581	0.00911

For the purpose of illustration, we present some graphs for Case 1 and 4 only. Based on the given algorithms in Section 3 we calculate the p.f. of  $S_i, i = 0, \dots, 3$ , i.e.  $h_Z$  and  $h_i, i = 1, 2, 3$ , in R. The Panjer's recursion with

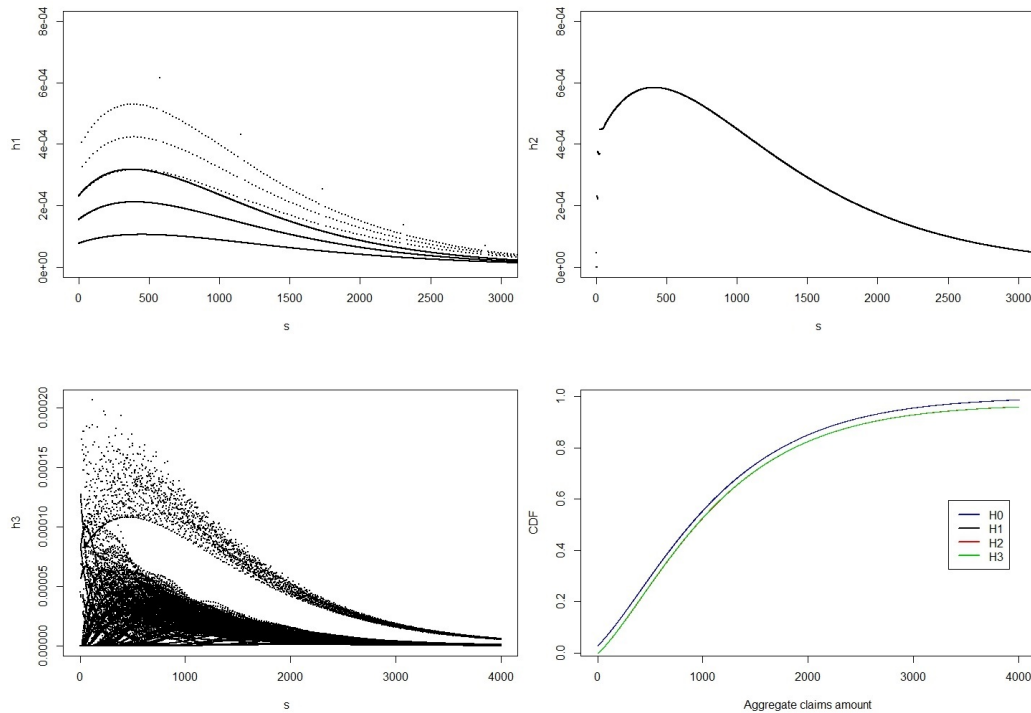


Figure 8. The graph of  $h_i, i = 1, 2, 3$ , and  $H_i, i = 0, 1, 2, 3$ , in Case 4

$a = b = \frac{51}{6}$  is the core in all calculations. Figure 7 and 8 summarize the numerical results obtained under the above assumptions. And it can be seen from Figure 7 and 8 that under the compound negative binomial assumption, the impact of the multiplicative factor  $Y_1$  and the additive factor  $Y_2$  on the distribution of aggregate claims amount  $S_i, i = 1, 2, 3$ , are similar to what we observed in Example 1. Also, the impact of  $Y_2$  on the c.d.f.'s  $H_i, i = 0, 1, 2, 3$ , is clearer than the compound Poisson case.

### Acknowledgement

The authors are grateful to the anonymous referees whose constructive comments have led to substantial improvements of the article. The research of Kam Chuen Yuen was supported by a GRF grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU17329216), and the CAE 2013 research grant from the Society of Actuaries – any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the SOA.

### REFERENCES

1. H. Albrecher, and J. Teugels, *Exponential behavior in the presence of dependence in risk theory*, Journal of Applied Probability, vol. 43, pp. 257–273, 2006.
2. H. Albrecher, C. Constantinescu, and S. Loisel, *Explicit ruin formulas for models with dependence among risks*, Insurance: Mathematics and Economics, vol. 48, pp. 265-270, 2011.
3. R.S. Ambagaspiya, *On the distribution of a sum of correlated aggregate claims*, Insurance: Mathematics and Economics, vol. 23, pp. 15-19, 1998.

4. B.C. Arnold, *Pareto Distributions*, International Cooperative Publishing House, Fairland, MD, 1983.
5. B.C. Arnold, *Pareto Distributions (second ed.)*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Boca Raton, FL, 2015.
6. A.V. Asimit, and A.L. Badescu, *Extremes on the discounted aggregate claims in a time dependent risk model*, Scandinavian Actuarial Journal, vol. 2010, pp. 93-104, 2010.
7. S. Asmussen, and H. Albrecher, *Ruin Probabilities (second ed.)*, World Scientific, New Jersey, 2010.
8. M. Boudreault, H. Cossette, D. Landriault, and É. Marceau, *On a risk model with dependence between interclaim arrivals and claim sizes*, Scandinavian Actuarial Journal, vol. 2006, pp. 265-285, 2006.
9. Y. Chen, and K.C. Yuen, *Precise large deviations of aggregate claims in a size-dependent renewal risk model*, Insurance: Mathematics and Economics, vol. 51, pp. 457-461, 2012.
10. H. Cossette, and É. Marceau, *The discrete-time model with correlated classes of business*, Insurance: Mathematics and Economics, vol. 26, pp. 133-149, 2000.
11. M. Denuit, C. Genest, and É. Marceau, *Stochastic bounds on sums of dependent risks*, Insurance: Mathematics and Economics, vol. 25, pp. 85-104, 1999.
12. J. Dhaene, and M.J. Goovaerts, *Dependency of risks and stop-loss order*, ASTIN Bulletin, vol. 26, pp. 201-212, 1996.
13. J. Li, Q. Tang, and R. Wu, *Subexponential tails of discounted aggregate claims in a time-dependent renewal risk model*, Advances in Applied Probability, vol. 42, pp. 1126-1146, 2010.
14. A. Müller, *Stop-loss order for portfolios of dependent risks*, Insurance: Mathematics and Economics, vol. 21, pp. 219-223, 1997.
15. H.H. Panjer, *Recursive Evaluation of a Family of Compound Distributions*, ASTIN Bulletin, vol. 12, pp. 22-26, 1981.
16. J.M. Sarabia, E. Gómez-Déniz, F. Prieto, and V. Jordá, *Risk aggregation in multivariate dependent Pareto distributions*, Insurance: Mathematics and Economics, vol. 71, pp. 154-163, 2016.
17. K.K. Siaw, X. Wu, D. Pitt, and Y. Wang, *Matrix-form recursive evaluation of the aggregate claims distribution revisited*, Annals of Actuarial Science, vol. 5, no. 2, pp. 163-179, 2011.
18. K.C. Yuen, J. Guo, and X. Wu, *On a correlated aggregate claims model with Poisson and Erlang risk processes*, Insurance: Mathematics and Economics, vol. 31, pp. 205-214, 2002.
19. K.C. Yuen, J. Guo, and X. Wu, *On the first time of ruin in the bivariate compound Poisson model*, Insurance: Mathematics and Economics, vol. 38, pp. 298-308, 2006.
20. G. Wang, and K.C. Yuen, *On a correlated aggregate claims model with thinning-dependence structure*, Insurance: Mathematics and Economics, vol. 36, pp. 456-468, 2005.
21. X. Wu, and S. Li, *Matrix-form recursions for a family of compound distributions*, ASTIN Bulletin, vol. 40, no. 1, pp. 351-368, 2010.
22. X. Wu, and K.C. Yuen, *A discrete-time risk model with interaction between classes of business*, Insurance: Mathematics and Economics, vol. 33, pp. 117-133, 2003.