On Copula-based Collective Risk Models

Rosy Oh[a], Jae Youn Ahn[a,*], Woojoo Lee[b,*]

[a] *Department of Statistics, Ewha Womans University, 11-1 Daehyun-Dong, Seodaemun-Gu, Seoul 120-750, Korea.*
[b] *Department of Statistics, Inha University, 235 Yonghyun-Dong, Nam-Gu, Incheon 402-751, Korea.*

Abstract

Several collective risk models have recently been proposed by relaxing the widely used but controversial assumption of independence between claim frequency and severity. Approaches include the bivariate copula model, random effect model, and two-part frequency-severity model. This study focuses on the copula approach to develop collective risk models that allow a flexible dependence structure for frequency and severity. We first revisit the bivariate copula method for frequency and average severity. After examining the inherent difficulties of the bivariate copula model, we alternatively propose modeling the dependence of frequency and individual severities using multivariate Gaussian and t-copula functions. The proposed copula models have computational advantages and provide intuitive interpretations for the dependence structure. Our analytical findings are illustrated by analyzing automobile insurance data.

Keywords: Collective risk model, Frequency-severity Dependence, Copula, Gaussian copula
JEL Classification: C300

*Corresponding Authors

Email addresses: rosy.oh5@gmail.com (Rosy Oh), jaeyahn@ewha.ac.kr (Jae Youn Ahn), lwj221@gmail.com (Woojoo Lee)

1. Introduction

The collective risk model, defined as the sum of the severities or the average of the severities, is an important tool for decision making in the insurance sector. Traditionally, two types of independence assumptions are assumed in collective risk models: one is independence between claim frequency and each individual severity and the other is independence among individual severities, as discussed by Klugman et al. (2012). In this paper, we call a collective risk model with these two independence assumptions the "independent collective risk model."

Recently, researchers have relaxed those independence assumptions using flexible statistical models such as a shared random effect model (Hernández-Bastida et al., 2009; Baumgartner et al., 2015) and a copula model (Czado et al., 2012; Krämer et al., 2013; Frees et al., 2016; Cossette et al., 2018). Specifically, the copula models of Czado et al. (2012) and Krämer et al. (2013) introduce the dependence between frequency (N) and average severity (M) via a parametric copula family including a Gaussian copula; these models are also used by Frees et al. (2016), Cossette et al. (2018), and Lee and Shi (2019). Throughout this paper, we call (N, M) summarized data. Alternatively, Frees et al. (2014) propose the so-called *two-step frequency-severity model* to provide the dependence between frequency and severity. Various applications of their two-step frequency-severity model can be found in Shi et al. (2015), Garrido et al. (2016), Park et al. (2018), and Jeong et al. (2019).

Throughout this paper, we call $(N, Y_1, \cdots, Y_N)$ micro-level data, where Y_i denotes the individual severity. Dependence models for micro-level data have also been studied in the literature. For example, under a Sparre Andersen-type dependence structure, the dependence between frequency and severity is explained using the dependence between interclaim time and severity (Albrecher and Teugels, 2006; Boudreault et al., 2006; Cossette et al., 2008, 2010; Asimit and Badescu, 2010; Landriault et al., 2014). Recently, Liu and Wang (2017) and Cossette et al. (2018) present copula models for micro-level data focusing on the structural property of the aggregate sum. However, important issues about parameter estimation and how to interpret the dependence structure have not been investigated in depth. In addition, the existing literature does not discuss the important differences between micro-level and summarized data when modeling the dependence between frequency and severity.

In this paper, we first explain the need for a copula model for micro-level data by explaining the inherent difficulties in finding a suitable copula function for summarized data. Specifically, we show that even under the independent collective risk model, existing copula families may not suitably describe the empirical properties of summarized data because of the intrinsic dependence structure between N and M. As an alternative to the copula model for summarized data, we introduce the Gaussian and t-copula models for analyzing micro-level data. To make these models concrete, we investigate useful correlation matrices to define the dependence in the copula model and find conditions for the correlation matrix to be positive definite. Using the proposed correlation matrix and corresponding Gaussian or t-copula, we show how to model the various dependence structures in the collective risk model. Our model can accommodate two types of dependence: the dependence between claim frequency and each individual severity and the dependence among individual severities. In addition, we show how to extend the proposed copula model to accommodate the regression setting. The proposed models have computational advantages in terms of parameter estimation and provide an intuitive interpretation of the dependence structure.

The remainder of this paper is organized as follows. Section 2 defines the frequently used notations. The difficulties in finding a suitable copula function for summarized data are explained in Section 3. Before proposing our copula model, we first study the correlation matrices used to explain the dependence structure between frequency and individual severities and that among individual severities. In particular, conditions are provided to guarantee when they are positive definite in Section 4. Section 5 deals with a copula model for positive frequency data and Section 6 extends it to observed data including zero frequency. Some regression settings are also discussed. The numerical study is described in Section 7 and our analytical findings are illustrated by analyzing automobile insurance data in Section 8, followed by concluding remarks.

2. Symbols

Let $\mathcal{N}$ be a set of positive integers and $\mathcal{N}_0$ be a set of non-negative integers. Let N_i represent the number of claims (*frequency*) of the i-th policyholder and Y_{ij} indicate the claim size (*individual severity*) in the j-th claim of the i-th policyholder. For a non-negative integer k, we define

$$\boldsymbol{Y}_i^{[k]} := \begin{cases} (Y_{i1}, \cdots, Y_{ik})^{\mathrm{T}}, & k > 0; \\ \text{null}, & k = 0. \end{cases}$$

We further define two quantities:

$$S_i := \begin{cases} \sum_{j=1}^{N_i} Y_{ij}, & N_i > 0; \\ \text{null}, & N_i = 0; \end{cases} \quad \text{and} \quad M_i := \begin{cases} \frac{\sum_{j=1}^{N_i} Y_{ij}}{N_i}, & N_i > 0; \\ \text{not defined}, & N_i = 0. \end{cases} \tag{1}$$

Here, S_i and M_i are called the *aggregated severity* and *average severity*, respectively. They are linked as

$$M_i = \frac{S_i}{N_i} \quad , N_i > 0.$$

We use n_i, $\boldsymbol{y}_i^{[k]}$, s_i, and m_i as the realization of N_i, $\boldsymbol{Y}_i^{[k]}$, S_i, and M_i, respectively.

For the given frequencies $\{n_1, \cdots, n_l\}$ from l policyholders, define

$$\mathcal{I}_l := \left\{ i \in \{1, \cdots, l\} \middle| n_i \neq 0 \right\}.$$

Furthermore, we call

$$\left\{ (n_i, \boldsymbol{y}_i^{\mathrm{T}}) \middle| i = 1, \cdots, l \right\} \tag{2}$$

full data and

$$\left\{ (n_i, m_i) \middle| i = 1, \cdots, l \right\} \tag{3}$$

summarized data. Summarized data (3) are understood as

$$\left\{ n_i \middle| i \notin \mathcal{I}_l \right\} \cup \left\{ (n_i, m_i) \middle| i \in \mathcal{I}_l \right\} \tag{4}$$

because m_i is not defined when $n_i = 0$. When the context is clear, we drop the subscript i to simplify the notations. For example, we denote m_i, n_i, and $\boldsymbol{y}_i^{[n_i]}$ by m, n, and $\boldsymbol{y}^{[n]}$, respectively if it is clear that these notations are defined for the i-th policyholder.

For the frequency part, we allow any non-negative integer-valued distribution including distributions in the (reproductive) *exponential dispersion family* (EDF) and zero-inflated count distributions (Yip and Yau, 2005). We use $F_1(x; \lambda, \psi_1)$ and $f_1(x; \lambda, \psi_1)$ to denote the cumulative distribution function and probability mass function, respectively. Here, λ and ψ_1 correspond to the parameter of interest and nuisance parameter(s), respectively. For the severity part, to simplify the model, we only consider continuous positive distributions with a probability density function, including distributions belonging to the continuous EDF and heavy-tailed distributions. $F_2(x;\xi,\psi_2)$ and $f_2(x;\xi,\psi_2)$ are used to denote the cumulative distribution function and probability density function, respectively. Similar to the frequency part, ξ and ψ_2 are the parameter of interest and nuisance parameter(s), respectively. In a clear context, we simply use F_1, f_1, F_2, and f_2 for $F_1(x;\lambda,\psi_1)$, $f_1(x;\lambda,\psi_1)$, $F_2(x;\xi,\psi_2)$, and $f_2(x;\xi,\psi_2)$, respectively.

3. Dependence in Collective Risk Models

One of the key assumptions frequently used in classical collective risk models is the independence of frequency and individual severities and the independence assumption among individual severities. However, recent studies (Czado et al., 2012; Krämer et al., 2013; Frees et al., 2014; Baumgartner et al., 2015; Shi et al., 2015; Garrido et al., 2016; Lee et al., 2016; Park et al., 2018; Jeong et al., 2019) have reported evidence against the independence assumption.

To capture the dependence between frequency and severity or among individual severities, Hernández-Bastida et al. (2009) and Baumgartner et al. (2015) use a shared random effect model, and Frees et al. (2014), Shi et al. (2015), Garrido et al. (2016), Lee et al. (2016), Park et al. (2018), and Jeong et al. (2019) use a frequency model to predict severities in the regression setting. On the contrary, Czado et al. (2012), Krämer et al. (2013), Frees et al. (2016), Cossette et al. (2018), and Lee and Shi (2019) adopt a parametric copula approach, including a Gaussian copula, to show the dependence between frequency and average severity. While the copula is a widely used tool for modeling dependence, the choice of a suitable copula family is often a more difficult problem than the choice of a suitable marginal distribution family. In particular, when modeling the dependence between frequency and average severity, the choice of a suitable copula family can be even harder. The following example shows that most existing copula families including Gaussian and Archimedean copulas cannot accommodate the dependence between frequency and average severity properly, even under the simplest assumption where frequency and individual severities are assumed to be independent.

Example 1. *Consider the classical collective risk model, where frequency N and the individual severity Y_i's are assumed to be independent. We further assume that N is a zero-truncated Poisson distribution with*

$$\mathbb{P}\left(N=n\right)=\frac{\lambda^n}{(e^\lambda-1)n!}$$

and

$$Y_1,\cdots,Y_N\big|N\overset{\text{i.i.d.}}{\sim}\text{Gamma}(\xi,\psi).$$

Then, we have

$$M\big|N\sim\text{Gamma}(\xi,\psi/N).$$

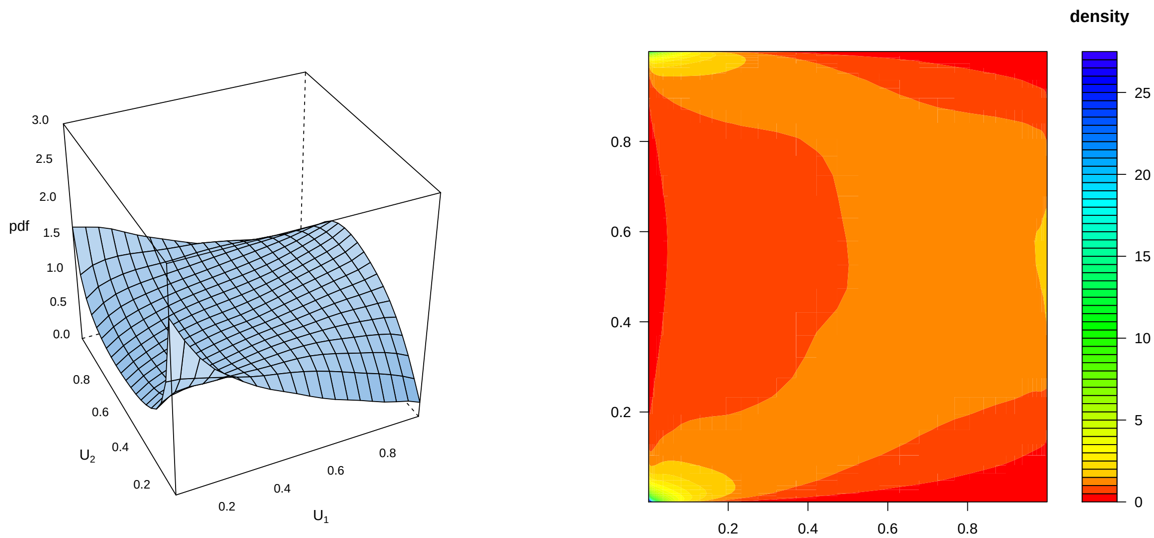
Clearly, N and M are not independent even though frequency and individual severities are independent.

Now, we want to visualize the density function of a suitable copula family for (N,M) under the assumption that frequency and individual severities are independent. Let F_N and F_M denote the distribution functions for N and M, respectively. $\text{Ran}F$ *means range of F. Since the copula of (N,M) is unique only on $\text{Ran}F_N\times\text{Ran}F_M$ as shown by Sklar (1959), the corresponding copula density function is not easily visualized. Instead, we define the alternative random vector (N^*,M) as*

$$N^*:=N+Z\quad\textit{and}\quad M\big|N\sim\text{Gamma}(\xi,\psi/N),\tag{5}$$

where $Z \sim \text{Unif}[0,1]$ and Z are independent of N and $(Y_1, \cdots, Y_N)$. Clearly, (N^*, M) is a continuous random vector, and the corresponding copula is uniquely determined on $[0,1] \times [0,1]$. While the copula of (N^*, M) is different to the corresponding (sub)copula of (N, M), we can have an important insight into the shape of the (sub)copula of (N, M) by examining that of (N^*, M).*

Since the corresponding copula of (N^*, M) is implicitly defined, in this example, we estimate the corresponding copula using a kernel density estimation of the copula with simulated samples (Gijbels and Mielniczuk, 1990; Chen and Huang, 2007a). Figure 1 shows the kernel density function of the copula $\widehat{C}^*$, using $n = 2,000$ pairs of the i.i.d. random vector from (N^*, M), and the corresponding contour plot.



(a) 3D Plot: Density of the Copula

(b) Contour: Density of the Copula

Figure 1: Density estimation of the copula of (N^*, M) using a kernel density estimation

Let (U_1, U_2) be a random vector sampled from $\widehat{C}^*$ in Figure 1. As shown from the figure, for the lower U_1, the density of the copula tends to be smaller at the center of U_2. On the contrary, for the higher U_1, the density of the copula tends to be larger at the center of U_2. Clearly, the density plots in Figure 1 reflect the fact that the conditional variance of M shrinks as N rises, which captures the most eminent feature of the copula of (N, M).

In conclusion, the choice of the copula family to provide a suitable dependence structure between N and M described in Example 1 can be difficult. We consider that no existing parametric copula family can reflect the property described in Example 1 accurately. One may consider using the non-parametric bivariate copula approach as in Chen and Huang (2007b); however, this also has difficulties to provide a straightforward interpretation of the dependence, as shown in Example 1, as long as (N, M) is modeled.

In the subsequent sections, rather than directly providing the dependence of the summarized data (N, M), we provide the dependence structure of the micro-level data $(N, Y_1, \cdots, Y_N)$ using a copula method. Such an approach requires access to full data, whereas the approach in Example 1 only requires the summarized data.

4. Correlation Matrix for Frequency and Individual Severities

This section presents two useful correlation matrices to describe the dependence of $(N, Y_1, \cdots, Y_N)$.

4.1. Equicorrelation matrix

We study a correlation matrix that has a common pairwise correlation for individual severities $(Y_1, \cdots, Y_N)$. Based on this, we investigate an extended correlation matrix for $(N, Y_1, \cdots, Y_N)$.

Definition 1. *For $\rho_1, \rho_2 \in [-1, 1]$, define the following matrices.*

i. For any positive integer k, define a $k \times k$ matrix $\boldsymbol{\Sigma}_{\rho_2}^{[k,1]}$

$$\left[\boldsymbol{\Sigma}_{\rho_2}^{[k,1]}\right]_{i,j} := \begin{cases} 1, & \text{if} \quad i = j \\ \rho_2, & \text{if} \quad i \neq j \end{cases}$$

for $i, j = 1, \cdots, k$.

ii. For any non-negative integer k, define a $(k+1) \times (k+1)$ matrix $\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,1]}$

$$\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,1]} := \begin{cases} \begin{pmatrix} 1 & \rho_1 \left(\mathbf{1}_k\right)^{\mathrm{T}} \\ \rho_1 \mathbf{1}_k & \boldsymbol{\Sigma}_{\rho_2}^{[k,1]} \end{pmatrix}, & k = 1, 2, \cdots, ; \\ 1, & k = 0; \end{cases}$$

where $\mathbf{1}_k$ is a column vector of 1 with length k.

In matrix form, $\boldsymbol{\Sigma}_{\rho_2}^{[k,1]}$ and $\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,1]}$ in Definition 1 are written as

$$\boldsymbol{\Sigma}_{\rho_2}^{[k,1]} = \begin{pmatrix} 1 & \rho_2 & \rho_2 & \cdots & \rho_2 \\ \rho_2 & 1 & \rho_2 & \cdots & \rho_2 \\ \rho_2 & \rho_2 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_2 & \rho_2 & \rho_2 & \cdots & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,1]} = \begin{pmatrix} 1 & \rho_1 & \rho_1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \rho_2 & \rho_2 & \cdots & \rho_2 \\ \rho_1 & \rho_2 & 1 & \rho_2 & \cdots & \rho_2 \\ \rho_1 & \rho_2 & \rho_2 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_1 & \rho_2 & \rho_2 & \rho_2 & \cdots & 1 \end{pmatrix}.$$

The following proposition provides the determinants of $\boldsymbol{\Sigma}_{\rho_2}^{[k,1]}$ and $\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,1]}$.

Proposition 1. *For any non-negative integer k and $\rho_1, \rho_2 \in (-1, 1)$, we have*

$$\det\left(\boldsymbol{\Sigma}_{\rho_2}^{[k,1]}\right) = \left(1 + (k-1)\rho_2\right)\left(1 - \rho_2\right)^{k-1} \tag{6}$$

and

$$\det\left(\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,1]}\right) = \left[1 + (k-1)\rho_2 - k(\rho_1)^2\right]\left(1 - \rho_2\right)^{k-1}.$$

The proof is given in Appendix A. We also provide the inverse of $\boldsymbol{\Sigma}_{\rho_2}^{[k,1]}$ and $\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,1]}$.

Proposition 2. *Let k be any non-negative integer and $\rho_1, \rho_2 \in (-1, 1)$. If $\det\left(\mathbf{\Sigma}_{\rho_2}^{[k,1]}\right) \neq 0$, then the inverse matrix of $\mathbf{\Sigma}_{\rho_2}^{[k,1]}$ is*

$$\left(\mathbf{\Sigma}_{\rho_2}^{[k,1]}\right)^{-1} = \frac{1}{1-\rho_2}\left[\boldsymbol{I}_k - \frac{\rho_2}{1+(k-1)\rho_2}\boldsymbol{J}_{k\times k}\right], \tag{7}$$

where $\boldsymbol{J}_{k\times k}$ is a $k \times k$ matrix of ones. Furthermore, if $\det\left(\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,1]}\right) \neq 0$, then

$$\left[\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,1]}\right]^{-1} = \begin{pmatrix} \frac{\rho_1^2 k}{1+(k-1)\rho_2-k\rho_1^2} & \frac{\rho_1}{1+(k-1)\rho_2-k\rho_1^2}\mathbf{1}_k^{\mathrm{T}} \\ \frac{\rho_1}{1+(k-1)\rho_2-k\rho_1^2}\mathbf{1}_k & \frac{1}{1-\rho_2}\left[\boldsymbol{I}_k - \frac{\rho_2-\rho_1^2}{1+(k-1)\rho_2-k\rho_1^2}\boldsymbol{J}_{k\times k}\right] \end{pmatrix}. \tag{8}$$

The proof is given in Appendix A. The following theorem provides the condition for $\mathbf{\Sigma}_{\rho_2}^{[k,1]}$ and $\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,1]}$ to be positive definite.

Theorem 1. *Let k be any positive integer and $\rho_1, \rho_2 \in (-1, 1)$. Then, $\mathbf{\Sigma}_{\rho_2}^{[k,1]}$ is positive definite if and only if ρ_2 satisfies*

$$1 + \rho_2(k-1) > 0. \tag{9}$$

Similarly, $\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,1]}$ is positive definite if and only if ρ_1 and ρ_2 satisfy

$$k(\rho_1)^2 - 1 < (k-1)\rho_2. \tag{10}$$

The proof is given in Appendix A. For any positive integer k and $\rho_1 \in (-1, 1)$, define

$$L_{\rho_1,1}(k) = \begin{cases} \frac{k(\rho_1)^2-1}{k-1}, & k > 1; \\ -1, & \text{otherwise.} \end{cases}$$

In Figure 2, the shaded area of (ρ_1, ρ_2) guarantees that $\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,1]}$ is positive definite for different k. As shown in the figure, the shaded area shrinks as k increases. The following corollary, which extends Theorem 1, formally describes such an observation.

Corollary 1. *Let $k_1, \cdots, k_z$ be non-negative integers. Then, we have the following results.*

i. For $\rho_1, \rho_2 \in [-1, 1]$, the correlation matrices

$$\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k_1,1]}, \cdots, \mathbf{\Sigma}_{\rho_1,\rho_2}^{[k_z,1]}$$

are positive definite if and only if ρ_1 and ρ_2 satisfy

$$L_{\rho_1,1}(\max\{k_1, \cdots, k_z\}) < \rho_2 < 1. \tag{11}$$

ii. If ρ_1 and ρ_2 satisfy

$$\rho_1^2 < \rho_2 < 1,$$

then

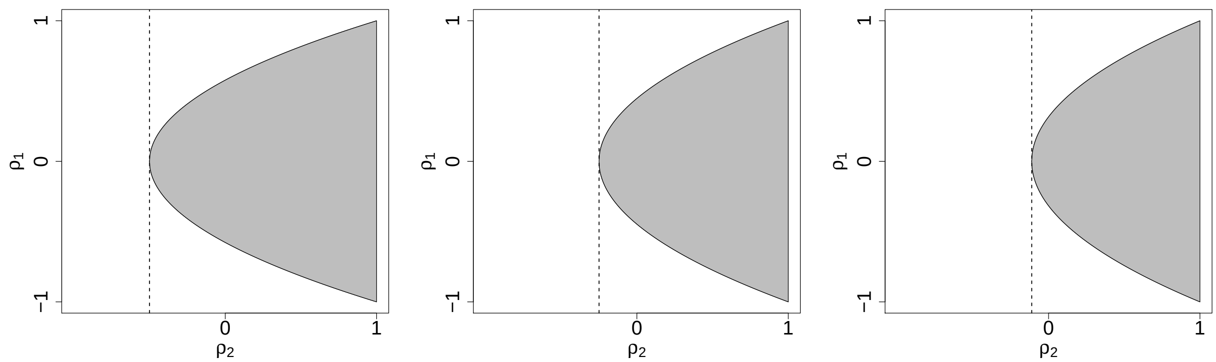
$$\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,1]}$$

is positive definite for any non-negative integer k.

For the proof, the first part comes from the fact that for $\rho_1 \in (-1,1)$, $L_{\rho_1,1}(k)$ is a non-decreasing function of k for $k \geq 1$. The second part is the result from

$$\lim_{k\to\infty} L_{\rho_1,1}(k) = \rho_1^2$$

and the first part.



(a) $k = 3$ (b) $k = 5$ (c) $k = 10$

Figure 2: Area of (ρ_1, ρ_2) satisfying (10) for different values of the k's

4.2. Autoregressive correlation matrix

We study a correlation matrix that has an autoregressive correlation structure for individual severities $(Y_1, \cdots, Y_N)$. Based on this, we investigate an extended correlation matrix for $(N, Y_1, \cdots, Y_N)$.

Definition 2. *For $\rho_1, \rho_2 \in [-1, 1]$, define the following matrices.*

i. For any positive integer k, define a $k \times k$ matrix $\mathbf{\Sigma}_{\rho_2}^{[k,2]}$ as

$$\left[\mathbf{\Sigma}_{\rho_2}^{[k,2]}\right]_{i,j} := \begin{cases} 1, & \text{if} \quad i = j \\ \rho_2^{|i-j|}, & \text{if} \quad i \neq j \end{cases}$$

for $i, j = 1, \cdots, k$.

ii. For any non-negative integer k, define a $(k+1) \times (k+1)$ matrix $\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,2]}$ as

$$\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,2]} := \begin{cases} \begin{pmatrix} 1 & \rho_1 (\mathbf{1}_k)^{\mathrm{T}} \\ \rho_1 \mathbf{1}_k & \mathbf{\Sigma}_{\rho_2}^{[k,2]} \end{pmatrix}, & k = 1, 2, \cdots, ; \\ 1, & k = 0; \end{cases}$$

where $\mathbf{1}_k$ is a column vector of 1 with length k.

In matrix form, $\mathbf{\Sigma}_{\rho_2}^{[k,2]}$ and $\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,2]}$ in Definition 2 are written as

$$\mathbf{\Sigma}_{\rho_2}^{[k,2]} = \begin{pmatrix} 1 & (\rho_2)^1 & (\rho_2)^2 & \cdots & (\rho_2)^{k-1} \\ (\rho_2)^1 & 1 & (\rho_2)^1 & \cdots & (\rho_2)^{k-2} \\ (\rho_2)^2 & (\rho_2)^1 & 1 & \cdots & (\rho_2)^{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (\rho_2)^{k-1} & (\rho_2)^{k-2} & (\rho_2)^{k-3} & \cdots & 1 \end{pmatrix}$$

and

$$\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,2]} = \begin{pmatrix} 1 & \rho_1 & \rho_1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & (\rho_2)^1 & (\rho_2)^2 & \cdots & (\rho_2)^{k-1} \\ \rho_1 & (\rho_2)^1 & 1 & (\rho_2)^1 & \cdots & (\rho_2)^{k-2} \\ \rho_1 & (\rho_2)^2 & (\rho_2)^1 & 1 & \cdots & (\rho_2)^{k-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_1 & (\rho_2)^{k-2} & (\rho_2)^{k-3} & \cdots & 1 & \end{pmatrix}.$$

The following proposition provides the determinants of $\mathbf{\Sigma}_{\rho_2}^{[k,2]}$ and $\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,2]}$.

Proposition 3. *For $\rho_1, \rho_2 \in (-1, 1)$ and any non-negative integer k, we have*

$$\det\left(\mathbf{\Sigma}_{\rho_2}^{[k,2]}\right) = (1-\rho_2^2)^k$$

and

$$\det\left(\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,2]}\right) = 1 - \rho_1^2\left(k - 2\rho_2(k-1) + \rho_2^2(k-2)\right).$$

The proof is given in Appendix A. We also provide the following well-known result without a proof.

Proposition 4. *If $\det\left(\mathbf{\Sigma}_{\rho_2}^{[k,2]}\right) \neq 0$, then*

$$(1-\rho_2^2)\left[\left(\mathbf{\Sigma}_{\rho_2}^{[k,2]}\right)^{-1}\right]_{i,j} = \begin{cases} 1, & (i,j) = (1,1) \quad \textit{or} \quad (i,j) = (k,k); \\ 1+\rho_2^2, & (i,j) \neq (1,1), \quad (i,j) \neq (k,k) \quad \textit{and} \quad i = j; \\ -\rho_2, & j = i+1 \quad \textit{or} \quad j = i-1; \\ 0, & \textit{otherwise}. \end{cases}$$

Although the inverse matrix of $\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,2]}$ can be represented analytically using the Schur complement (Zhang, 2006), we do not pursue it because its representation is unnecessarily complicated. By contrast, the condition for $\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,2]}$ to be positive definite is succinct, as described in the following theorem. The proof is given in Appendix A.

Theorem 2. *Consider $\rho_1, \rho_2 \in (-1, 1)$. Then, for a positive integer k, $\mathbf{\Sigma}_{\rho_2}^{[k,2]}$ is positive definite for any $\rho_1, \rho_2 \in (-1, 1)$. Furthermore, for a non-negative integer k, $\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,2]}$ is positive definite for any $\rho_1, \rho_2 \in (-1, 1)$.*

5. Conditional Collective Risk Model

This section presents the dependent collective risk model for positive frequency, which is called the "conditional collective risk model" throughout this paper. In Section 6, we provide a generalized collective risk model that allows zero frequency. Consider the positive frequency and severities:

$$N^{+} \quad \text{and} \quad \{Y_1, Y_2, \cdots\}\,. \tag{12}$$

Suppose that n and k are positive integers in $\mathbb{N} \times \mathbb{N}$. The distribution function for positive frequency and severities is denoted by

$$\mathbb{P}\left(N^{+} \le n, \boldsymbol{Y}^{[k]} \le \boldsymbol{y}^{[k]}\right). \tag{13}$$

k in (13) can be determined independently of N^{+}. Cossette et al. (2018) studies a similar model, (13), with $k = N^{+}$, but we allow k to be any positive integer. For example, we allow studying the distributions of $\left(N^{+}, Y^{[1]}\right) = (N^{+}, Y_1)$ and $\left(N^{+}, Y^{[2]}\right) = (N^{+}, Y_1, Y_2)$ at the same time. From the condition that the former should be obtained from the latter by integrating it with respect to Y_2, in general, we require the distribution in (13) to satisfy that for any $k_1 < k_2$, $\mathbb{P}\left(N^{+} \le n, \boldsymbol{Y}^{[k_1]} \le \boldsymbol{y}^{[k_1]}\right)$ is derived from $\mathbb{P}\left(N^{+} \le n, \boldsymbol{Y}^{[k_2]} \le \boldsymbol{y}^{[k_2]}\right)$ by taking the integration.

Conditional Model 1. *Let $N^{+} \sim F_1^{+}$ be a non-degenerate positive integer-valued random variable, with the probability mass function f_1. Then, we define the joint distribution of $(N^{+}, \boldsymbol{Y}^{[k]})$ as satisfying*

$$(N^{+}, \boldsymbol{Y}^{[k]}) \sim H_k = C_{k+1}(F_1^{+}, F_2, \cdots, F_2), \tag{14}$$

for any positive integer k, where F_2 is non-negative continuous distribution that has f_2 as a probability density function. Here, C_{k+1} is a $(k+1)$-dimensional copula satisfying the following inheritance property: for any $k_1 < k_2$,

$$C_{k_1+1}\left(u_1, \cdots, u_{k_1+1}\right) = C_{k_2+1}\left(u_1, \cdots, u_{k_1+1}, 1, \cdots, 1\right) \quad \textit{for} \quad u_1, \cdots, u_{k_1+1} \in [0,1]. \tag{15}$$

For the two copulas C_{k_1+1} and C_{k_2+1} satisfying the inheritance property, the corresponding distributions H_{k_1+1} and H_{k_2+1} also satisfy the inheritance property. The copula models that we present in the following subsections satisfy such an inheritance property. In contrast to ours, Cossette et al. (2018)'s model does not require condition (15) to be satisfied. In terms of parameter estimation, our model and Cossette et al. (2018)'s model, which have the same marginal distribution functions and copula functions, provide the same likelihood function because we always have $k = n^{+}$ in real observations.[1] However, to interpret and construct the dependence structure, allowing k to be any positive integer as well as having the condition in (15), which is essentially the same as adding an assumption to the unobserved data, is critical, as commented in Remark 1.

The shared random effect model in Hernández-Bastida et al. (2009), Baumgartner et al. (2015), and Oh et al. (2019) and two-step frequency-severity model of Garrido et al. (2016), Park et al. (2018), and Jeong et al. (2019) require the distribution of individual severity to be in the EDF. Since these models require modeling the dependence between frequency and average severity (or aggregate severity), as mentioned by Garrido et al. (2016) and Oh et al. (2019), the EDF assumption on individual severities is required to derive average severity. On the contrary, since our model does not require modeling the dependence between frequency and average severity, it allows F_2 to be any distribution function, including a heavy-tailed distribution function.

[1] Assume positive frequency for simplicity.

Similar to Cossette et al. (2018), for a positive integer k, the joint density function in Conditional Model 1 can be written as

$$\begin{aligned}\frac{\partial^k}{\partial y_1 \cdots \partial y_k}&\left(H_k\left(n,\boldsymbol{y}^{[k]}\right)-H_k\left(n-1,\boldsymbol{y}^{[k]}\right)\right)\\&=\left(\mathbb{P}\left(N^+\le n\big|\boldsymbol{y}^{[k]}\right)-\mathbb{P}\left(N^+\le n-1\big|\boldsymbol{y}^{[k]}\right)\right)f_{\boldsymbol{Y}^{[k]}}(\boldsymbol{y}^{[k]})\end{aligned}\tag{16}$$

for $n \in \mathbb{Z}$ and $\boldsymbol{y}^{[k]} \in \mathbb{R}^k$, where $f_{\boldsymbol{Y}^{[k]}} : \mathbb{R}_+^k \mapsto \mathbb{R}_+$ is the probability density function of $\boldsymbol{Y}^{[k]}$. The computational complexity of (16) depends on the choice of the $(k+1)$-dimensional copula C_{k+1}. In particular, the conditional distribution of N^+ in (16) is involved in the integration of the copula, which increases the complexity of the estimation procedure. In the cases of the Gaussian copula and t-copula, the conditional distribution parts in (16) are relatively easy to compute as well as provide an intuitive interpretation of the dependence structure using the form of the covariance matrix. In this section, we present the Gaussian copula and t-copula versions of Conditional Model 1 in detail.

5.1. Gaussian copula model

In the following, we provide the distribution of $(N^+, \boldsymbol{Y}^{[k]})$ for a positive integer k based on the Gaussian copula family with the equicorrelation matrix, $\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,z]}$ for $z = 1$, and with the autoregressive correlation matrix, $\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,z]}$ for $z = 2$.

Conditional Model 2. *Let $N^+ \sim F_1^+$ be a non-degenerate positive integer-valued random variable, with the probability mass function f_1. Consider the positive definite matrices*

$$\boldsymbol{\Sigma}_{\rho_2}^{[k,z]} \quad \textit{and} \quad \boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,z]}$$

for $z = 1, 2$. We assume

$$(\rho_1,\rho_2) \in \left\{(\rho_1,\rho_2) \in (-1,1)^2 \big| \rho_1^2 < \rho_2 < 1\right\}\tag{17}$$

for $z = 1$, and assume

$$(\rho_1,\rho_2) \in (-1,1)^2\tag{18}$$

for $z = 2$. Then, for the given correlation matrix structure $z \in \{1, 2\}$, we define the joint distribution of $(N^+, \boldsymbol{Y}^{[k]})$ as satisfying

$$(N^+, \boldsymbol{Y}^{[k]}) \sim H_{N^+,\boldsymbol{Y}^{[k]}} = C_{\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,z]}}(F_1^+, F_2, \cdots, F_2),\tag{19}$$

for any positive integer k, where F_2 is a non-negative continuous distribution that has f_2 as a probability density function. Here, $C_{\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,z]}}$ is a Gaussian copula, with a corresponding density function denoted as $c_{\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,z]}}$ and correlation matrix $\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,z]}$.

It is straightforward to check that the copula

$$C_{\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,z]}}, \quad \text{for} \quad z = 1, 2$$

satisfies the condition in (15). The following lemma shows the explicit form of the density function of Conditional Model 2.

Lemma 1. *Considering the frequency and severities in Conditional Model 2, for each $z = 1, 2$, we have the following results.*

i. For a positive integer k, the joint density function is given by

$$\begin{aligned} & h_{N^+, \boldsymbol{Y}^{[k]}}(n, \boldsymbol{y}^{[k]}) \\ & \quad = f_{\boldsymbol{Y}^{[k]}}(\boldsymbol{y}^{[k]}) \left(\Phi \left(\frac{\Phi^{-1}(F_1^+(n)) - \mu_{[k,z]}}{\sigma_{[k,z]}} \right) - \Phi \left(\frac{\Phi^{-1}(F_1^+(n-1)) - \mu_{[k,z]}}{\sigma_{[k,z]}} \right) \right) \end{aligned} \tag{20}$$

for $n \in \mathbb{Z}$ and $\boldsymbol{y}^{[k]} \in \mathbb{R}^k$, where $f_{\boldsymbol{Y}^{[k]}}$ is a density function given as

$$f_{\boldsymbol{Y}^{[k]}}(y_1, \cdots, y_k) = c_{\boldsymbol{\Sigma}_{\rho_2}^{[k,z]}}(F_2(y_1), \cdots, F_2(y_k)) \prod_{i=1}^{k} f_2(y_i), \tag{21}$$

which is the probability density function of the cumulative distribution function $C_{\boldsymbol{\Sigma}_{\rho_2}^{[k,z]}}(F_2, \cdots, F_2)$. Here, $\mu_{[k,z]}$ and $\sigma_{[k,z]}$ are defined as

$$\mu_{[k,z]} := \left(\rho_1 \mathbf{1}_k^{\mathrm{T}}\right) \left(\boldsymbol{\Sigma}_{\rho_2}^{[k,z]}\right)^{-1} \left(\Phi^{-1}(F_2(y_1)), \cdots, \Phi^{-1}(F_2(y_k))\right)^{\mathrm{T}} \tag{22}$$

and

$$\sigma_{[k,z]} := \sqrt{1 - \left(\rho_1 \mathbf{1}_k^{\mathrm{T}}\right) \left(\boldsymbol{\Sigma}_{\rho_2}^{[k,z]}\right)^{-1} (\rho_1 \mathbf{1}_k)}. \tag{23}$$

ii. For a positive integer k, the conditional density function of is given $\boldsymbol{Y}^{[k]}$ for given $N^+ = n$ is

$$\begin{aligned} & h_{\boldsymbol{Y}^{[k]} \mid N^+}\left(\boldsymbol{Y}^{[k]} \middle| n\right) \\ & \quad = \frac{f_{\boldsymbol{Y}^{[k]}}(\boldsymbol{y}^{[k]})}{f_1^+(n)} \left(\Phi \left(\frac{\Phi^{-1}(F_1^+(n)) - \mu_{[k,z]}}{\sigma_{[k,z]}} \right) - \Phi \left(\frac{\Phi^{-1}(F_1^+(n-1)) - \mu_{[k,z]}}{\sigma_{[k,z]}} \right) \right). \end{aligned} \tag{24}$$

The proof in given in Appendix B. Conditional Model 2 has an advantage when investigating the degree of dependence among the variables via Spearman's rho. The definition of Spearman's rho is first given below.

Definition 3. *Define Spearman's rho of $(V_1, V_2) \sim C(G_1, G_2)$ for some bivariate copula C and the marginal distributions G_1 and G_2 as*

$$\rho(V_1, V_2) = 12 \int \int C(G_1(v_1), G_2(v_2)) - G_1(v_1) G_2(v_2) \mathrm{d}v_1 \mathrm{d}v_2.$$

For the details on Spearman's rho, see Nelsen (2006). Spearman's rho in Conditional Model 2 can be obtained from the following results.

Corollary 2. *Consider the frequency and severities defined in Conditional Model 2. Then, for each $z = 1, 2$, we have the following results.*

i. Spearman's rho between N^+ and Y_j can be calculated as

$$\rho(N^+, Y_j) = \sum_{n=0}^{\infty} \int_0^{\infty} 12 \left[C_{\rho_1}(F_1^+(n), F_2(y)) - F_1^+(n) F_2(y) \right] f_1(n) f_2(y) \mathrm{d}y$$

, where C_{ρ_1} is the Gaussian copula with the correlation coefficient ρ_1.

ii. Spearman's rho between Y_{j_1} and Y_{j_2} can be calculated as

$$\rho(Y_{j_1}, Y_{j_2}) = \frac{6}{\pi} \arcsin\left(\frac{\rho_2}{2}\right), \quad \textit{for} \quad z = 1$$

and

$$\rho(Y_{j_1}, Y_{j_2}) = \frac{6}{\pi} \arcsin\left(\frac{\rho_2^{|k_1 - k_2|}}{2}\right), \quad \textit{for} \quad z = 2$$

for the positive integers j_1 and j_2 satisfying $j_1 \neq j_2$.

The proof of the first part is immediate from the definition, and the proof of the second part can be found in Kruskal (1958).

Remark 1. *In Conditional Model 1, allowing k to be any integer value makes the definition of the dependence measures such as*

$$\rho\left(N^+, Y_j\right) \quad \textit{and} \quad \rho\left(Y_{j_1}, Y_{j_2}\right), \quad j_1 \neq j_2 \tag{25}$$

in Corollary 2 well defined and interpreted straightforwardly. On the contrary, fixing $k = N^+$ as in Cossette et al. (2018) may complicate the interpretation.[2] *Indeed, under the model with fixed $k = N^+$ only, the definition of the (marginal) dependence measures in* (25) *may be loose because Y_j is defined only when $N^+ \geq j$. Therefore, the model with fixed $k = N^+$ only is suitable to discuss the following conditional versions of the dependence measures*

$$\rho\left(N^+, Y_j \middle| N^+ \geq j\right) \quad \textit{and} \quad \rho\left(Y_{j_1}, Y_{j_2} \middle| N^+ \geq \max\{j_1, j_2\}\right), \quad j_1 \neq j_2. \tag{26}$$

However, interpreting the dependence structure with such dependence measures (26) *can be difficult. As a consequence, the interpretation of the (marginal) correlation matrix*

$$\mathbf{\Sigma}_{\rho_1, \rho_2}^{[k,z]}, \quad z = 1, 2$$

introduced in Section 4 may be difficult under Conditional Model 1 with fixed $k = N^+$ only.

5.2. T-copula model

The Gaussian copula is not an inevitable choice for modeling frequency and severities. However, for the convenience of the statistical modeling and estimation, we focus on the copula family specified by the covariance (correlation) matrix only. The elliptical copula family has such a property. Among them, we focus on the t-copula that has gained broad popularity in recent studies.

[2] The main purpose of Cossette et al. (2018) is the construction of the collective risk model based on a hierarchical Archimedean copula, which does not require the risk measure such as (25).

We first define the k-dimensional multivariate t-distribution $\boldsymbol{Z} \sim \mathrm{MVT}(\mathbf{0}_k, \boldsymbol{\Sigma}, \mathrm{df})$, with scale matrix $\boldsymbol{\Sigma}$, and degrees of freedom df, which has the following probability density function:

$$f(\boldsymbol{z}) = \frac{\Gamma\left(\frac{k+\mathrm{df}}{2}\right)}{\Gamma\left(\frac{\mathrm{df}}{2}\right)(\pi \mathrm{df})^{k/2}\left|\boldsymbol{\Sigma}\right|^{1/2}}\left(1+\frac{1}{\mathrm{df}}\boldsymbol{k}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{k}\right)$$

for $\boldsymbol{z} = (z_1, \cdots, z_k) \in \mathbb{R}^k$. We denote the corresponding t-copula as $C_{\mathrm{df},\boldsymbol{\Sigma}^{[k,1]}_{\rho_1,\rho_2}}$. We also represent Φ_{df} and ϕ_{df} by the cumulative density function and probability density function of the univariate Student's t-distribution, respectively. The t-copula version of Conditional Model 1 is provided below.

Conditional Model 3. *Let $N^+ \sim F_1^+$ be a positive integer-valued random variable, with the probability mass function f_1^+, and assume* (19) *and* (18) *for $z = 1$ and $z = 2$, respectively. Then, for each $z = 1, 2$, define the joint distribution of $(N, \boldsymbol{Y}^{[k]})$ as in*

$$(N^+, \boldsymbol{Y}^{[k]}) \sim C_{\mathrm{df},\boldsymbol{\Sigma}^{[k,z]}_{\rho_1,\rho_2}}(F_1^+, F_2, \cdots, F_2),$$

for any positive integer k, where F_2 is a non-negative continuous distribution that has f_2 as a probability density function. Here, $C_{\mathrm{df},\boldsymbol{\Sigma}^{[k,z]}_{\rho_1,\rho_2}}$ is a t-copula, which has the corresponding density function denoted as $c_{\mathrm{df},\boldsymbol{\Sigma}^{[k,z]}_{\rho_1,\rho_2}}$, with scale matrix $\boldsymbol{\Sigma}^{[k,z]}_{\rho_1,\rho_2}$ and degree of freedom df.

The following result is the t-copula version of Lemma 1.

Lemma 2. *Considering the frequency and severities in Conditional Model 3 with $z = 1$ and $z = 2$, we have the following results.*

i. For a positive integer k, the joint density function is given by

$$\begin{aligned} &h_{N^+,\boldsymbol{Y}^{[k]}}(n, \boldsymbol{y}^{[k]}) \\ &\quad = f^*_{\boldsymbol{Y}^{[k]}}(\boldsymbol{y}^{[k]})\left(\Phi_{\mathrm{df}+k}\left(\frac{\Phi_{\mathrm{df}}^{-1}(F_1^+(n)) - \mu^*_{[k,z]}}{\sigma^*_{[k,z]}}\right) - \Phi_{\mathrm{df}+k}\left(\frac{\Phi_{\mathrm{df}}^{-1}(F_1^+(n-1)) - \mu^*_{[k,z]}}{\sigma^*_{[k,z]}}\right)\right) \end{aligned} \tag{27}$$

for $n \in \mathbb{N}^+$ and $\boldsymbol{y}^{[k]} \in \mathbb{R}^k$, where $f^*_{\boldsymbol{Y}^{[k]}}(\boldsymbol{y}^{[k]})$ is a density function given as

$$f^*_{\boldsymbol{Y}^{[k]}}(y_1, \cdots, y_k) = c_{\mathrm{df},\boldsymbol{\Sigma}^{[k,z]}_{\rho_2}}(F_2(y_1), \cdots, F_2(y_k))\prod_{i=1}^{k} f_2(y_i),$$

which is the probability density function of the cumulative distribution function $C_{\mathrm{df},\boldsymbol{\Sigma}^{[k,z]}_{\rho_2}}(F_2, \cdots, F_2)$. Here, $\mu^*_{[k,z]}$ and $\sigma^*_{[k,z]}$ are defined as

$$\mu^*_{[k,z]} := \left(\rho_1 \mathbf{1}_k^{\mathrm{T}}\right)\left(\boldsymbol{\Sigma}^{[k,z]}_{\rho_2}\right)^{-1}\left(\Phi_{\mathrm{df}}^{-1}(F_2(y_1)), \cdots, \Phi_{\mathrm{df}}^{-1}(F_2(y_k))\right)^{\mathrm{T}} \tag{28}$$

and

$$\sigma^*_{[k,z]} := \sigma_{[k,z]} \times \sqrt{\frac{\mathrm{df} + \left(\Phi^{-1}_{\mathrm{df}}(F_2(y_1)), \cdots, \Phi^{-1}_{\mathrm{df}}(F_2(y_k))\right)\left(\boldsymbol{\Sigma}^{[k,z]}_{\rho_2}\right)^{-1}\left(\Phi^{-1}_{\mathrm{df}}(F_2(y_1)), \cdots, \Phi^{-1}_{\mathrm{df}}(F_2(y_k))\right)^{\mathrm{T}}}{\mathrm{df}+k}}. \tag{29}$$

ii. *For a positive integer k, the conditional density function of is given $\boldsymbol{Y}^{[k]}$ for given $N^+ = n$ is*

$$h_{\boldsymbol{Y}^{[k]}\big|N^+}\left(\boldsymbol{Y}^{[k]}\middle|n\right) = \frac{f^*_{\boldsymbol{Y}^{[k]}}(\boldsymbol{y}^{[k]})}{f_1^+(n)}\left(\Phi_{\mathrm{df}+k}\left(\frac{\Phi^{-1}_{\mathrm{df}}(F_1^+(n)) - \mu^*_{[k,z]}}{\sigma^*_{[k,z]}}\right) - \Phi_{\mathrm{df}+k}\left(\frac{\Phi^{-1}_{\mathrm{df}}(F_1^+(n-1)) - \mu^*_{[k,z]}}{\sigma^*_{[k,z]}}\right)\right). \tag{30}$$

For brevity, we omit the proof because it is similar to that of Lemma 1.

6. Collective Risk Model with the Observed Data

Conditional Models 2 and 3 cannot be directly applied to real data including zero frequency because they assume that the frequency is positive. In the following, we explain how to modify Conditional Models 2 and 3 to accommodate zero frequency.

Model 1 (Dependent collective risk model)**.** *For each correlation matrix $z = 1, 2$, consider the frequency and severities defined in Conditional Model 2 or 3. We assume that R is a Bernoulli random variable with success probability p, and it is mutually independent of $(N^+, Y_1, \cdots, Y_k)$ for any $k > 0$. Define*

$$N := \begin{cases} N^+, & R = 1; \\ 0, & R = 0; \end{cases}$$

and its cumulative distribution function and density function as F_1 and f_1, respectively. Then, define

$$(N, \boldsymbol{Y}^{[N]}) := \begin{cases} (N, Y_1, \cdots, Y_N), & N \geq 1; \\ 0, & N = 0; \end{cases}$$

as a dependent collective risk model.

Model 1 is called the "dependent collective risk model" throughout the paper. The following theorems, which are the corollaries of Lemmas 1 and 2, show the joint density function of $(N, \boldsymbol{Y}^{[N]})$ in Model 1.

Theorem 3. *Consider Model 1 along with Conditional Model 2. Then, for each $z = 1, 2$, the joint density function of the discrete margin N and continuous margins $\boldsymbol{Y}_N$ is given by*

$$h(n, \boldsymbol{y}^{[n]}) = \begin{cases} p\, f_{\boldsymbol{Y}^{[n]}}(\boldsymbol{y}^{[n]})\left(\Phi\left(\frac{\Phi^{-1}(F_1^+(n)) - \mu_{[n,z]}}{\sigma_{[n,z]}}\right) - \Phi\left(\frac{\Phi^{-1}(F_1^+(n-1)) - \mu_{[n,z]}}{\sigma_{[n,z]}}\right)\right), & n \in \mathbb{N}, \quad \boldsymbol{y}^{[n]} \in \mathbb{R}^n \\ 1 - p, & n = 0, \end{cases} \tag{31}$$

where $\mu_{[n,z]}$ and $\sigma_{[n,z]}$ are defined in (22) *and* (23).

The proof is provided here. (31) is trivial for $n = 0$. Now, consider the case for $n > 0$. For an observation $(n, y_1, \cdots, y_n)$, the corresponding likelihood function is

$$\begin{aligned} h(n, \boldsymbol{y}^{[n]}) &= p\, f_1^{+}(n) h_{\boldsymbol{Y}^{[n]}\big|N^+}(y_1, \cdots, y_n \big| n) \\ &= p\, h_{N^+, \boldsymbol{Y}^{[n]}}(n, \boldsymbol{y}^{[n]}), \end{aligned}$$

where the second equality is from part ii of Lemma 1.

Theorem 4. *Consider Model 1 along with Conditional Model 3. Then, for each $z = 1, 2$, the joint density function of the discrete margin N and continuous margins $\boldsymbol{Y}_N$ is given by*

$$\begin{aligned} & h(n, \boldsymbol{y}^{[n]}) \\ &= \begin{cases} p\, f_{\boldsymbol{Y}^{[n]}}(\boldsymbol{y}^{[n]}) \left(\Phi_{\mathrm{df}+n} \left(\frac{\Phi_{\mathrm{df}}^{-1}(F_1^{+}(n)) - \mu^{*}_{[n,z]}}{\sigma^{*}_{[n,z]}} \right) - \Phi_{\mathrm{df}+n} \left(\frac{\Phi_{\mathrm{df}}^{-1}(F_1^{+}(n-1)) - \mu^{*}_{[n,z]}}{\sigma^{*}_{[n,z]}} \right) \right), & n \in \mathbb{N}, \quad \boldsymbol{y}^{[n]} \in \mathbb{R}^n \\ 1 - p, & n = 0, \end{cases} \end{aligned} \tag{32}$$

where $\mu^{*}_{[n,z]}$ and $\sigma^{*}_{[n,z]}$ are defined in (28) *and* (29).

For brevity, we omit the proof because it is the same as in Theorem 3.

6.1. Derivation of some useful quantities

Using Lemma 5 in the Appendix B, the following proposition shows how to derive some useful quantities from Model 1.

Proposition 5. *Consider Model 1 along with either Conditional Model 2 or Conditional Model 3. Then, we have*

$$\mathbb{E}\left[S\right] = \mathbb{E}\left[\sum_{j=1}^{N} Y_j\right] = p \sum_{n=1}^{\infty} n \int_0^{\infty} y\, h_{N^+, \boldsymbol{Y}^{[1]}}(n, y) \mathrm{d}y$$

and

$$\mathrm{cov}\left[N, S\right] = p \sum_{n=1}^{\infty} n^2 \int_0^{\infty} y h_{N^+, \boldsymbol{Y}^{[1]}}(n, y) \mathrm{d}y - \left(p \sum_{n=1}^{\infty} n f_{N^+}(n) \right) \left(p \sum_{n=1}^{\infty} n \int_0^{\infty} y\, h_{N^+, \boldsymbol{Y}^{[1]}}(n, y) \mathrm{d}y \right),$$

where $h_{N^+, \boldsymbol{Y}^{[1]}}$ is defined in (20) *or* (27)*, depending on the assumptions of Conditional Model 2 or Conditional Model 3 as well as the type of covariance matrix. Furthermore, we have*

$$\mathrm{cov}\left[N, M \big| N > 0\right] = \sum_{n=1}^{\infty} n \int_0^{\infty} y h_{N^+, \boldsymbol{Y}^{[1]}}(n, y) \mathrm{d}y - \left(\sum_{n=1}^{\infty} n f_{N^+}(n) \right) \left(\sum_{n=1}^{\infty} n \int_0^{\infty} y h_{N^+, \boldsymbol{Y}^{[1]}}(n, y) \mathrm{d}y \right).$$

The detailed derivation steps of Proposition 5 are given in Appendix C.

6.2. Extension to regression models

We provide regression models for the dependent collective risk model below.

Model 2. *For each individual $i = 1, \cdots, I$, consider the dependent collective risk model for*

$$(N_i, \boldsymbol{Y}_i^{[N_i]}) := \begin{cases} (N_i, Y_{i1}, \cdots, Y_{iN_i}), & N_i \geq 1; \\ 0, & N_i = 0; \end{cases}$$

in Model 1. Let $(\boldsymbol{x}_i, \boldsymbol{x}_i^*, \boldsymbol{w}_i)$ be the given characteristics of the i-th policyholder. Consider the following model.

i. For the frequency part, use the hurdle regression model with N_{it}. Specifically, use $\boldsymbol{x}_i$ and $\boldsymbol{x}_i^*$ as the explanatory variables for the zero and positive frequency parts, respectively. Denote $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ as the corresponding sets of regression coefficients. Specifically,

$$N_i^+ \sim F_1^+ \left(\cdot; \lambda_i, \psi_1\right), \quad \textit{with} \quad \eta_1(\lambda_i) = \boldsymbol{x}_i \boldsymbol{\beta}$$

and

$$R_i \sim \mathrm{Ber}\left(p_i\right), \quad \textit{with} \quad \eta_1^*(p_i) = \boldsymbol{x}_i^* \boldsymbol{\beta}^*$$

for some link functions η_1 and η_1^*.

ii. For the severity part, use the regression model with Y_{it} and $\boldsymbol{w}_i$ as the dependent variable and the set of explanatory variables, respectively. Denote $\boldsymbol{\gamma}$ as the corresponding sets of regression coefficients. Specifically, for $N_i > 0$,

$$Y_{it} \sim F_2\left(\cdot\,; \xi_i, \psi_2\right), \quad \textit{with} \quad \eta_2(\xi_i) = \boldsymbol{w}_i \boldsymbol{\gamma}$$

for $t \in \mathcal{N}$ with some link function η_2.

Model 2 is a flexible regression model for frequency and severities. Its marginal distribution of frequency is given as

$$F_1(n; \lambda_i, \psi_1) = 1 - p_i + p_i F_1^+(n; \lambda_i, \psi_1)$$

for $n = 0, 1, \ldots$. A special case of Model 2 is particularly interesting because it is convenient for statistical estimation. Consider any distribution function F_1 that can explain the frequency part including zero. Define

$$p_i = 1 - F_1\left(0; \lambda_i, \psi_1\right)$$

and

$$F_1^+(n; \lambda_i, \psi_1) = \frac{F_1\left(n; \lambda_i, \psi_1\right) - F_1\left(0; \lambda_i, \psi_1\right)}{1 - F_1\left(0; \lambda_i, \psi_1\right)}$$

for $n \in \mathbb{N}_+$. Then, we have

$$N_i \sim F_1(\cdot; \lambda_i, \psi_1).$$

We use this model for the simulation and data analysis in the following sections. Specifically, we use the Poisson distribution with mean $\lambda_i = \exp(\boldsymbol{x}_i\boldsymbol{\beta})$ for N_i, and Gamma distribution $\xi_i = \exp(\boldsymbol{w}_i\boldsymbol{\gamma})$ for Y_{it}.

7. Numerical Study

We conduct a simulation study to investigate the finite sample properties of the parameter estimates and effect of the dependence between frequency and severity on them for the proposed model. The portfolio of policyholders of size $I = 5000$ are generated from the proposed model under 12 scenarios motivated by the real data analysis in Section 8. Table 1 provides the details of the parameter settings. In each simulation, two predictors $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ are used and generated from Bernoulli(0.5) independently.

Table 1: Parameter settings for the 12 scenarios

	Parameter								
Scenario	β_0	β_1	β_2	γ_0	γ_1	γ_2	ν	ρ_1	ρ_2
1	-2.5	0.5	1.0	8	-0.1	0.3	0.7	-0.05	0.10
2	-2.5	0.5	1.0	8	-0.1	0.3	0.7	-0.05	0.05
3	-2.5	0.5	1.0	8	-0.1	0.3	0.7	0.05	0.10
4	-2.5	0.5	1.0	8	-0.1	0.3	0.7	0.05	0.05
5	-2.5	0.5	1.0	8	-0.1	0.3	0.7	0.10	0.10
6	-2.5	0.5	1.0	8	-0.1	0.3	0.7	0.10	0.05
7	-2.5	0.5	1.5	8	-0.1	0.3	0.7	-0.05	0.10
8	-2.5	0.5	1.5	8	-0.1	0.3	0.7	-0.05	0.05
9	-2.5	0.5	1.5	8	-0.1	0.3	0.7	0.05	0.10
10	-2.5	0.5	1.5	8	-0.1	0.3	0.7	0.05	0.05
11	-2.5	0.5	1.5	8	-0.1	0.3	0.7	0.10	0.10
12	-2.5	0.5	1.5	8	-0.1	0.3	0.7	0.10	0.05

For each scenario, Table 2 and Table 3 summarize the simulation results from 500 independent Monte Carlo samples, including the relative bias and mean squared error (MSE) of the parameter estimates. Table 2 indicates that in all the scenarios, the estimates are close to the true parameters of the proposed model and shows that the relative bias and MSE are small. A relative bias larger than 10% is only observed for ρ_2 in scenario 5, which has relatively high correlations for ρ_1 and ρ_2.

Table 2: Relative bias in $\%$ for all the parameters from the 12 scenarios

	Relative Bias ($\%$)								
Scenario	β_0	β_1	β_2	γ_0	γ_1	γ_2	ν	ρ_1	ρ_2
1	0.06	0.34	0.01	0.01	-1.57	-0.10	0.19	5.79	8.61
2	0.10	0.35	0.11	0.11	-0.55	-0.59	0.16	0.36	0.42
3	-0.12	-0.57	-0.17	-0.17	-0.60	-0.73	0.23	1.74	6.77
4	0.01	-0.36	0.11	0.11	-4.00	0.46	0.43	0.94	-2.63
5	0.00	0.32	-0.05	-0.05	0.89	-0.77	0.11	1.58	14.90
6	0.02	-0.13	-0.07	-0.07	-1.62	-1.50	0.31	-1.45	-0.83
7	0.01	-0.07	-0.03	-0.03	2.98	1.15	0.12	0.66	-0.08
8	0.03	0.10	-0.08	-0.08	-2.36	0.05	0.11	-1.92	-2.80
9	0.03	0.43	0.04	0.04	0.68	1.50	0.26	1.14	4.61
10	0.16	0.07	0.22	0.22	-1.39	-0.72	0.04	0.98	-0.73
11	-0.10	-0.30	-0.15	-0.15	-0.95	0.40	0.08	1.89	3.16
12	-0.04	-0.03	-0.18	-0.18	-0.80	0.88	0.28	0.68	-2.55

Table 3: Mean absolute error for all the parameters from the 12 scenarios

Scenario	MSE								
	β_0	β_1	β_2	γ_0	γ_1	γ_2	ν	ρ_1	ρ_2
1	0.0028	0.0020	0.0023	0.0039	0.0033	0.0037	0.0004	0.0008	0.0023
2	0.0028	0.0023	0.0026	0.0041	0.0034	0.0035	0.0003	0.0008	0.0038
3	0.0025	0.0023	0.0024	0.0037	0.0032	0.0035	0.0004	0.0007	0.0027
4	0.0025	0.0022	0.0023	0.0035	0.0030	0.0037	0.0004	0.0008	0.0035
5	0.0022	0.0020	0.0024	0.0038	0.0030	0.0039	0.0003	0.0008	0.0024
6	0.0028	0.0022	0.0025	0.0035	0.0029	0.0037	0.0004	0.0009	0.0034
7	0.0022	0.0016	0.0022	0.0033	0.0022	0.0031	0.0003	0.0005	0.0014
8	0.0025	0.0014	0.0022	0.0031	0.0018	0.0032	0.0002	0.0005	0.0015
9	0.0022	0.0015	0.0022	0.0029	0.0020	0.0032	0.0003	0.0005	0.0013
10	0.0023	0.0013	0.0021	0.0034	0.0023	0.0029	0.0003	0.0006	0.0016
11	0.0023	0.0015	0.0023	0.0038	0.0021	0.0033	0.0002	0.0004	0.0013
12	0.0025	0.0014	0.0022	0.0036	0.0021	0.0030	0.0003	0.0004	0.0016

8. Real Data Analysis

To see the usefulness of the proposed model for examining the dependence structure between (a) frequency and severity and (b) severities, we analyze a real automobile insurance dataset.

8.1. Data

We use the automobile insurance data provided by the Massachusetts Executive Office of Energy and Environmental Affairs, which were used by Ferreira Jr and Minikel (2012). The data contain the history of automobile insurance claims in 2006 in the state of Massachusetts. With 15 variables, the data consist of information on $3,991,012$ insured persons. The dataset also shows the $681,423$ claims of liability and personal injury protection coverage claim information with 10 variables. Each policyholder has information on the number of claims, individual claim amounts with the date of accidents, and covariates. Among the observations, we randomly sample $1,500,000$ policyholders whose accidents occurred before 2008 and who have automobile insurance providing third party liability coverage for property damage and bodily injury. The first $1,000,000$ observations are used as the training data to develop the model and the rest is reserved as the hold-out sample for validation purposes. We use the following two covariates, CLASS and TERRITORY, for the risk classification. CLASS denotes five groups divided by policyholder characteristics (A: adults, B: business, I: $<$ 3 years of experience, M: 3 $\sim$ 6 years of experience, S: senior citizens). TERRITORY denotes six territory groups divided by the driving characteristics (1: least risky to 6: most risky territory). Table 4 shows the mean frequency and mean severity per claim of the data categorized by CLASS and TERRITORY. Full details of the covariates used can be found in the online supplement and in Ferreira Jr and Minikel (2012).

8.2. Estimation results

We apply Model 2 to the Massachusetts automobile data, where N_i and Y_{it} follow a Poisson distribution and a gamma distribution, respectively. Table 5 summarizes the estimation results for the model. Based on the results, the class and territory groups are important factors for both the frequency and the severity parts. The regression coefficients for the territory group show increasing patterns from the least risky to

Table 4: Table of mean frequency and mean severity per claim (percentage of observations in brackets)

		CLASS				
		A	B	I	M	S
TERRITORY	1	0.046 / 3241 (13.69%)	0.046 / 1513 (0.31%)	0.078 / 4186 (1.17%)	0.058 / 4547 (0.87%)	0.048 / 3426 (2.89%)
	2	0.046 / 3833 (13.82%)	0.046 / 5516 (0.30%)	0.070 / 3091 (1.13%)	0.066 / 3501 (0.94%)	0.049 / 3006 (3.16%)
	3	0.048 / 3933 (8.19%)	0.033 / 2180 (0.15%)	0.085 / 4724 (0.58%)	0.055 / 3977 (0.55%)	0.048 / 4196 (1.67%)
	4	0.051 / 3702 (14.88%)	0.060 / 4431 (0.27%)	0.097 / 3809 (1.00%)	0.061 / 3601 (0.96%)	0.060 / 4153 (3.01%)
	5	0.055 / 4042 (14.11%)	0.083 / 2949 (0.21%)	0.093 / 4052 (0.84%)	0.084 / 4538 (0.94%)	0.061 / 3812 (2.72%)
	6	0.062 / 4249 (8.87%)	0.111 / 2979 (0.11%)	0.107 / 4264 (0.60%)	0.092 / 4968 (0.71%)	0.070 / 4412 (1.35%)

the most risky area in both the frequency and the severity parts. The class group with less experience of driving (class=I) shows more accidents and the claim amount in that group tends to be higher than that in the other groups. In the dependent collective risk model, the dependence between frequency and severity is measured by the parameter ρ_1. Its estimate is -0.018 with a 95% confidence interval of $(-0.031, -0.005)$, suggesting a significant negative correlation between the number of accidents and claim size. Furthermore, this dependence seems weaker than the dependence between severities ρ_2.

To examine how well the proposed model and Tweedie's compound Poisson model fit the training dataset, we consider two quantities: expected aggregate severity, $\mathbb{E}[S]$, and the value at risk of aggregate severity at the confidence level $\alpha = 0.995$, $VaR_{0.995}(S)$, by risk group defined by the CLASS and TERRITORY variables. Figure 3 and Table 6 report the results. For comparison purposes, we also report the empirical values of $\mathbb{E}[S]$ and $VaR_{0.995}(S)$ (i.e., model-free estimates) for each risk group. It is expected that good models produce $\mathbb{E}[S]$ and $VaR_{0.995}(S)$ close to the corresponding empirical values. Figure 3 shows that both models provide similar estimates of $\mathbb{E}[S]$ and they are close to the empirical values. Although the MSEs of the two estimates of $\mathbb{E}[S]$, at the bottom of Table 6, do not show large differences, the estimates of $VaR_{0.995}(S)$ of both models show substantial differences in MSEs. Specifically, regarding $VaR_{0.995}(S)$, the risk group with class I shows high values within each territory group and it tends to be increasing as the territory becomes riskier. This pattern is also found in Tweedie's model; however, it overestimates them for most of the risk groups, which makes its MSE larger than that of our proposed model.

8.3. Validation results

For validation purposes, we compare the two models in terms of the total loss prediction for the 500,000 policyholders in the hold-out sample. Figure 4 presents the predictive distributions of the various models, which are based on 5,000 Monte Carlo simulations under the estimation result from each model. In the figure, the dotted vertical line indicates the actual amount of losses and dashed vertical line indicates the mean estimated total loss. The predictive distribution from our proposed collective risk model has less variation, and its mean is closer to the actual total loss of the hold-out sample. This result can be explained by the fact that there is a significant negative correlation between frequency and individual severity, which is reflected appropriately in our model that allows such dependence.

Table 5: Estimation results

Parameter	Est	Std. error	95% CI	
Frequency part				
Intercept	-3.295	0.012	-3.319	-3.270
territory=2	0.084	0.016	0.052	0.116
territory=3	0.121	0.018	0.085	0.157
territory=4	0.213	0.016	0.182	0.243
territory=5	0.326	0.015	0.296	0.356
territory=6	0.455	0.016	0.423	0.487
class=B	0.306	0.036	0.235	0.377
class=I	1.029	0.016	0.999	1.060
class=M	0.497	0.018	0.462	0.533
class=S	-0.017	0.013	-0.043	0.008
Severity part				
Intercept	8.067	0.014	8.039	8.095
territory=2	0.081	0.019	0.043	0.118
territory=3	0.062	0.022	0.020	0.104
territory=4	0.144	0.018	0.108	0.179
territory=5	0.154	0.018	0.120	0.189
territory=6	0.296	0.019	0.259	0.333
class=B	0.087	0.042	0.005	0.170
class=I	0.115	0.018	0.080	0.151
class=M	0.133	0.021	0.092	0.175
class=S	-0.105	0.015	-0.135	-0.075
ν	0.738	0.004	0.730	0.746
Copula part				
ρ_1	-0.018	0.006	-0.031	-0.005
ρ_2	0.027	0.001	0.026	0.029

9. Conclusion

We propose copula-based dependent collective risk models that allow the dependence between frequency and individual severities and that among individual severities to be separate. We also provide the conditions for the two correlation matrices used to describe the dependence to be positive definite. In particular, we emphasize that using the Gaussian copula or t-copula has computational advantages because they allow an analytic form for the conditional distribution of frequency given the individual severities.

Various extensions of our proposed models are possible as future research topics. First, it would be interesting to find appropriate general copula classes that could be used for dependent collective risk models. Although we pursue some special copulas based on computational convenience, if they cannot satisfactorily explain the given data, other complex copula functions may be necessary. Consequently, the copula choice problem (i.e., model selection problem) becomes an important issue here. Second, it would be interesting to model repeated measurements of frequency and individual severities over time. Our copula model should thus be extended to take into account the fact that the measurements from different time points are correlated. A promising approach would be to use a vine copula to combine several dependent collective risk models.

Table 6: Comparison of $\mathbb{E}[S]$ and $VaR_{0.995}(S)$ by risk group for the training dataset from the Massachusetts automobile data: Empirical value, Proposed model, and Tweedie model

Risk Group	$TERRITORY$	$CLASS$	Data		Proposed		Tweedie	
			$E[S]$	$VaR_{\alpha}(S)$	$E[S]$	$VaR_{\alpha}(S)$	$E[S]$	$VaR_{\alpha}(S)$
1	1	A	121	5693	119	6845	117	7903
2	1	B	96	3137	171	8542	162	10133
3	1	I	437	18275	356	11718	356	20618
4	1	M	299	15756	227	10357	227	13917
5	1	S	109	4425	107	6263	104	6900
6	2	A	142	6706	139	7760	141	9262
7	2	B	146	5904	219	10508	184	11842
8	2	I	473	18603	432	13255	432	23934
9	2	M	226	11415	260	11286	279	16734
10	2	S	153	7136	126	7071	122	8361
11	3	A	145	6692	142	7738	141	9357
12	3	B	179	8332	202	9800	217	12778
13	3	I	561	25000	436	13807	466	27135
14	3	M	315	16081	275	11290	273	16172
15	3	S	129	5270	127	6928	126	8456
16	4	A	163	8178	171	8969	168	10928
17	4	B	185	9404	244	10663	259	15605
18	4	I	519	18828	529	15908	534	28154
19	4	M	331	16643	313	12778	324	19202
20	4	S	150	6339	148	7797	146	9498
21	5	A	186	9429	191	9439	193	12277
22	5	B	366	22817	305	12494	278	16734
23	5	I	537	20277	615	16528	606	32221
24	5	M	327	15567	356	13709	359	20675
25	5	S	159	7846	168	8381	171	11167
26	6	A	263	14963	252	11698	250	15310
27	6	B	443	25328	378	14252	333	19741
28	6	I	655	24436	775	19352	780	38271
29	6	M	379	19188	466	16152	461	25125
30	6	S	221	12124	218	10163	231	13991
MSE				-	469	4192541	476	9759489

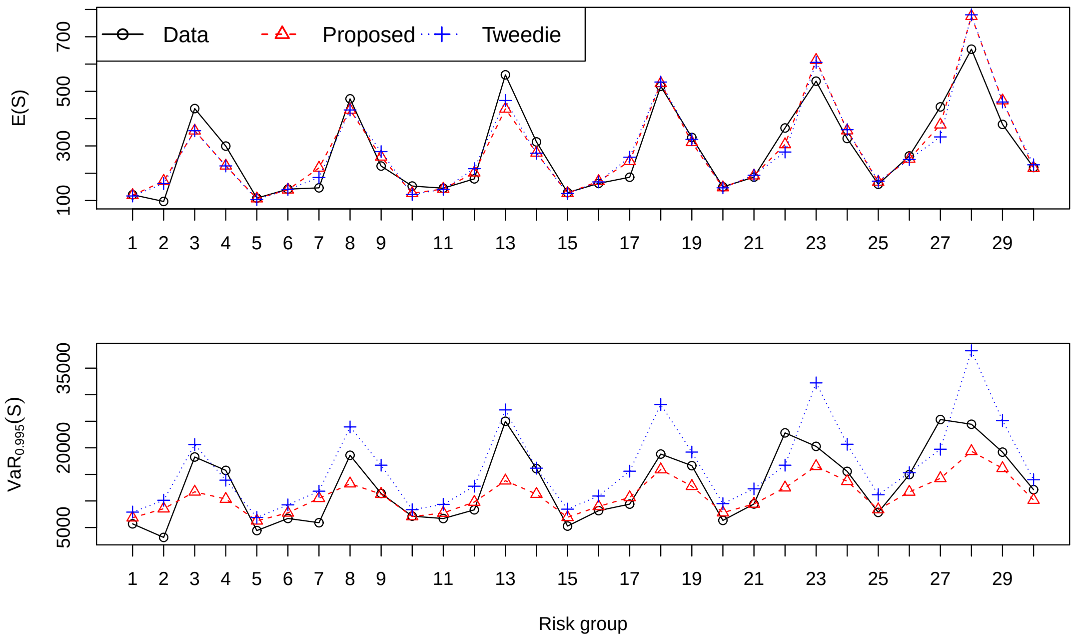


Figure 3: Comparison of $\mathbb{E}[S]$ and $VaR_{0.995}(S)$ by risk group for the training dataset from the Massachusetts automobile data: Empirical value, Proposed model, and Tweedie model

Acknowledgements

Woojoo Lee was supported by a Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A1B03936100). Jae Youn Ahn was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (NRF-2017R1D1A1B03032318).

References

Albrecher, H. and Teugels, J. L. (2006). Exponential behavior in the presence of dependence in risk theory. *Journal of Applied Probability*, 43(1):257–273.

Asimit, A. V. and Badescu, A. L. (2010). Extremes on the discounted aggregate claims in a time dependent risk model. *Scandinavian Actuarial Journal*, 2010(2):93–104.

Baumgartner, C., Gruber, L. F., and Czado, C. (2015). Bayesian total loss estimation using shared random effects. *Insurance: Mathematics and Economics*, 62:194–201.

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

Chen, S. X. and Huang, T.-M. (2007a). Nonparametric estimation of copula functions for dependence modelling. *Canadian Journal of Statistics*, 35(2):265–282.

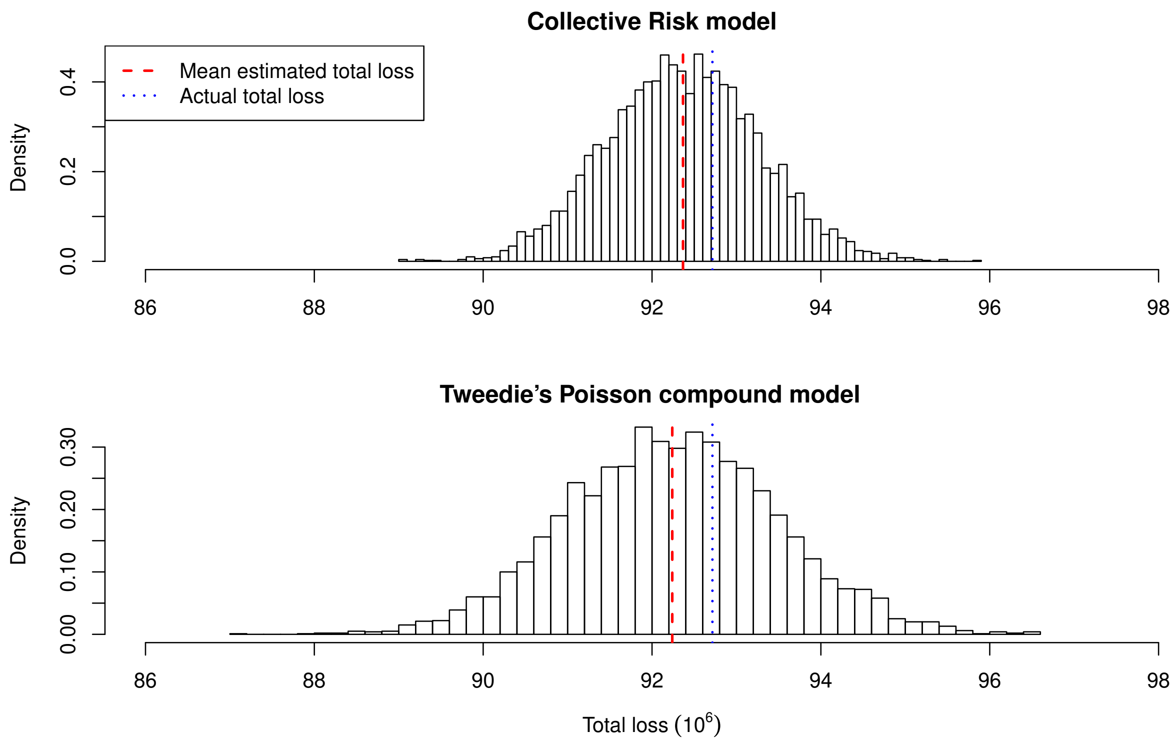


Figure 4: Predictive distribution of the total aggregated losses from the hold-out sample

Chen, S. X. and Huang, T.-M. (2007b). Nonparametric estimation of copula functions for dependence modelling. *Canadian Journal of Statistics*, 35(2):265–282.

Cossette, H., Marceau, E., and Marri, F. (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., Marceau, E., and Marri, F. (2010). Analysis of ruin measures for the classical compound poisson risk model with dependence. *Scandinavian Actuarial Journal*, 2010(3):221–245.

Cossette, H., Marceau, E., and Mtalai, I. (2018). Collective risk models with dependence. *Available at SSRN: https://ssrn.com/abstract=3104912 or http://dx.doi.org/10.2139/ssrn.3104912*.

Czado, C., Kastenmeier, R., Brechmann, E. C., and Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scand. Actuar. J.*, (4):278–305.

Ferreira Jr, J. and Minikel, E. (2012). Measuring per mile risk for pay-as-you-drive automobile insurance. *Transportation Research Record: Journal of the Transportation Research Board*, (2297):97–103.

Frees, E. W., Derrig, R. A., and Meyers, G. (2014). *Predictive Modeling Applications in Actuarial Science*, volume 1. Cambridge University Press.

Frees, E. W., Lee, G., and Yang, L. (2016). Multivariate frequency-severity regression models in insurance. *Risks*, 4(1):4.

Garrido, J., Genest, C., and Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, 70:205 – 215.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Communications in Statistics-Theory and Methods*, 19(2):445–464.

Hernández-Bastida, A., Fernández-Sánchez, M. P., and Gómez-Déniz, E. (2009). The net Bayes premium with dependence between the risk profiles. *Insurance Math. Econom.*, 45(2):247–254.

Jeong, H., Valdez, E. A., Ahn, J. Y., and Park, S. (2019). Generalized linear mixed models for dependent compound risk models.

Johnson, R. A. and Wichern, D. W. (2007). *Applied multivariate statistical analysis*. Pearson Prentice Hall, Upper Saddle River, NJ, sixth edition.

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

Krämer, N., Brechmann, E. C., Silvestrini, D., and Czado, C. (2013). Total loss estimation using copula-based regression models. *Insurance: Mathematics and Economics*, 53(3):829–839.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814–861.

Landriault, D., Lee, W. Y., Willmot, G. E., and Woo, J.-K. (2014). A note on deficit analysis in dependency models involving coxian claim amounts. *Scandinavian Actuarial Journal*, 2014(5):405–423.

Lee, G. Y. and Shi, P. (2019). A dependent frequency–severity approach to modeling longitudinal insurance claims. *Insurance: Mathematics and Economics*, 87:115–129.

Lee, W., Park, S. C., , and Ahn, J. Y. (2016). Investigating dependence between frequency and severity via simple generalized linear models. *Working Paper*.

Liu, H. and Wang, R. (2017). Collective risk models with dependence uncertainty. *ASTIN Bulletin: The Journal of the IAA*, 47(2):361–389.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.

Oh, R., Shi, P., and Ahn, J. Y. (2019). Implementation of frequency-severity association in bms ratemaking. *Working Paper*.

Park, S. C., Kim, J. H., and Ahn, J. Y. (2018). Does hunger for bonuses drive the dependence between claim frequency and severity? *Insurance: Mathematics and Economics*, 83:32–46.

Shi, P., Feng, X., and Ivantsova, A. (2015). Dependent frequency–severity modeling of insurance claims. *Insurance: Mathematics and Economics*, 64:417–428.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Institute of Statistics of the University of Paris*, 8:229–231.

Yip, K. C. and Yau, K. K. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, 36(2):153–163.

Zhang, F. (2006). *The Schur complement and its applications*, volume 4. Springer Science & Business Media.

Zhang, Y. (2013). Likelihood-based and bayesian methods for tweedie compound poisson linear mixed models. *Statistics and Computing*, 23(6):743–757.

Appendix A. Proofs on Covariance Matrix

Proof of Proposition 1. Proof of (6) is a well known result from the elementary matrix algebra. For the proof of the second equation, we may use (6) and Schur complement (Zhang, 2006).

□

Proof of Proposition 2. Proof of (7) is from the classical matrix algebra. For $k = 0$, (8) is trivial. For $k > 0$, with the following partitioned matrix representation

$$\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,1]} = \begin{pmatrix} 1 & \rho_1 \left(\mathbf{1}_k\right)^{\mathrm{T}} \\ \rho_1 \mathbf{1}_k & \mathbf{\Sigma}_{\rho_2}^{[k,1]} \end{pmatrix}$$

and Schur complement (Zhang, 2006), we have

$$\left(\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,1]}\right)^{-1} = \begin{pmatrix} 1 + \left(\rho_1 \boldsymbol{I}_k\right) \boldsymbol{E}^{-1} \left(\rho_1 \boldsymbol{I}_k\right)^{\mathrm{T}} & -\left(\rho_1 \boldsymbol{I}_k\right)^{\mathrm{T}} \boldsymbol{E}^{-1} \\ -\boldsymbol{E}^{-1} \left(\rho_1 \boldsymbol{I}_k\right) & \boldsymbol{E}^{-1} \end{pmatrix} \tag{A.1}$$

where

$$\boldsymbol{E} := \mathbf{\Sigma}_{\rho_2}^{[k,1]} - \left(\rho_1 \boldsymbol{I}_k\right) \left(\rho_1 \boldsymbol{I}_k\right)^{\mathrm{T}}.$$

Furthermore, using (7), we have

$$\boldsymbol{E}^{-1} = \frac{1}{1-\rho_2} \left[\boldsymbol{I}_k - \frac{\rho_2 - \rho_1^2}{1 + (k-1)\rho_2 - k\rho_1^2} \boldsymbol{J}_{k\times k}\right] \tag{A.2}$$

where the non-singularity of $\boldsymbol{E}^{-1}$ is guaranteed by $\det\left(\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,1]}\right) \neq 0$. Hence, simple algebraic manipulations using (A.1) and (A.2) conclude the proof.

□

Proof of Theorem 1. The first part is the classical result in matrix algebra. Now, we move to the second part. Since the proof is trivial for $k = 0$ or 1, we only consider a positive integer $k > 1$. By Schur complement (Zhang, 2013), we have that $\mathbf{\Sigma}_{\rho_1,\rho_2}^{[k,1]}$ is positive definite if and only if

$$1 - \left(\rho_1 \mathbf{1}_k\right)^{\mathrm{T}} \left(\mathbf{\Sigma}_{\rho_2}^{[k,1]}\right)^{-1} \left(\rho_1 \mathbf{1}_k\right) \tag{A.3}$$

and

$$\mathbf{\Sigma}_{\rho_2}^{[k,1]} \tag{A.4}$$

are positive definite.

From the following calculation

$$-\left(1-\rho_2\right)\left(1+(k-1)\rho_2\right)\left(1-\left(\rho_1\mathbf{1}_k\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\rho_2}^{[k,1]}\right)^{-1}\left(\rho_1\mathbf{1}_k\right)\right)=\left(1-\rho_2\right)\left(k\rho_1^2-1-(k-1)\rho_2\right),$$

we have (A.3) is positive definite if and only if (10) holds. Furthermore, since (A.4) is positive definite if and only if (9) holds, we have that $\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,1]}$ is positive definite if and only if (10) and (9) holds. Now, we conclude the proof by observing the intersection of (10) and (9) is (10).

□

Proof of Proposition 3. The first equation is well known in the classical matrix algebra, and the second equation is from the first equation and Schur complement (Zhang, 2006). □

Proof of Theorem 2. The first part is the classical result in matrix algebra. Now, we move to the second part. Since the proof is trivial for $k=0$ or 1, we only consider a positive integer $k>1$.

By Schur complement (Zhang, 2013), we have that $\boldsymbol{\Sigma}_{\rho_1,\rho_2}^{[k,2]}$ is positive definite if and only if

$$1-\left(\rho_1\mathbf{1}_k\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\rho_2}^{[k,2]}\right)^{-1}\left(\rho_1\mathbf{1}_k\right) \tag{A.5}$$

and

$$\boldsymbol{\Sigma}_{\rho_2}^{[k,2]} \tag{A.6}$$

are positive definite, where (A.6) is positive definite from the first part.

From the following calculation

$$-\left(1-\rho_2^2\right)\left(1-\left(\rho_1\mathbf{1}_k\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\rho_2}^{[k,2]}\right)^{-1}\left(\rho_1\mathbf{1}_k\right)\right)=\rho_2^2\left((k-2)\rho_1^2+1\right)-2(k-1)\rho_1^2\rho_2+(k\rho_1^2-1),$$

we have that (A.5) is positive definite if and only if

$$\rho_2^2\left((k-2)\rho_1^2+1\right)-2(k-1)\rho_1^2\rho_2+(k\rho_1^2-1)<0. \tag{A.7}$$

Since (A.7) is evident for $k=2$, it is enough to prove (A.7) for a positive integer $k>2$. From the following observation

$$-1<\frac{(k-2)\rho_1^2+\rho_1^2}{(k-2)\rho_1^2+1}<1$$

and the fact that the left side of inequality in (A.7) is a quadratic equation for $k>2$, we have (A.7) if and only if

$$g(\rho_1,\rho_2,k)<0, \quad \text{for} \quad \rho_2=-1, \quad 1, \quad \frac{(k-2)\rho_1^2+\rho_1^2}{(k-2)\rho_1^2+1} \tag{A.8}$$

where

$$g(\rho_1,\rho_2,k):=\rho_2^2\left((k-2)\rho_1^2+1\right)-2(k-1)\rho_1^2\rho_2+(k\rho_1^2-1).$$

Based on this, (A.8) always holds.

□

Appendix B. Auxilary Results on the Dependent Collective Risk Model

For the proof of the Lemma 1, the following classical results on the conditional distribution in multivariate normal distribution are applied to the special covariance structure $\boldsymbol{\Sigma}^{[k,1]}_{\rho_1,\rho_2}$.

Lemma 3. *Let $k_0 \in \mathbb{N}_+$, and consider $\rho_1, \rho_2 \in (-1, 1)$ satisfying* (19) *and* (18) *for $z = 1$ and $z = 2$, respectively. If we consider $(Y_0, Y_1, \cdots, Y_k)^{\mathrm{T}} \sim \mathrm{MVN}\left(\mathbf{0}_{k+1}, \boldsymbol{\Sigma}^{[k,z]}_{\rho_1,\rho_2}\right)$ with positive integer $k \le k_0$ and for $z \in \{1, 2\}$, the conditional distribution of Y_0 for given $(Y_1, \cdots, Y_k) = (y_1, \cdots, y_k)$ is*

$$Y_0 \big| (Y_1, \cdots, Y_k) = (y_1, \cdots, y_k) \sim \mathrm{N}(\mu_1, \sigma_1^2)$$

where

$$\mu_1 = \left(\rho_1 \mathbf{1}_n^{\mathrm{T}}\right) \left(\boldsymbol{\Sigma}^{[k,z]}_{\rho_2}\right)^{-1} (y_1, \cdots, y_k)^{\mathrm{T}} \quad \text{and} \quad \sigma_1^2 = 1 - \left(\rho_1 \mathbf{1}_k^{\mathrm{T}}\right) \left(\boldsymbol{\Sigma}^{[k,z]}_{\rho_2}\right)^{-1} \left(\rho_1 \mathbf{1}_k\right).$$

For the proof, see Johnson and Wichern (2007).

Lemma 4. *Consider Conditional Model 2 for for $z \in \{1, 2\}$. The conditional distribution of N is given as*

$$\mathbb{P}\left(N \le n \big| \boldsymbol{y}^{[k]}\right) = \Phi\left(\frac{\Phi^{-1}(F_1(n)) - \mu_{[k,1]}}{\sigma_{[k,1]}}\right)$$

for a positive integer k and $\boldsymbol{y}^{[k]} \in \mathbb{R}^k_+$, where $\mu_{[k,z]}$ and $\sigma_{[k,z]}$ are defined in (22) *and* (23).

Proof. For convenience, define

$$\mathcal{I}_N := \{n \in \mathcal{N}_0 \big| \mathbb{P}(N = n) > 0\}.$$

For the calculation of

$$\mathbb{P}\left(N \le n \big| y_1, \cdots, y_k\right)$$

let $N^* \sim F_1^*$ be continuous cumulative distribution satisfying

$$F_1^*(-1) = 0 \quad \text{and} \quad F_1^*(n) = F_1(n)$$

for positive integer $n \in \mathcal{I}_N$ where $F_1^*(\cdot)$ being a strictly increasing function on $(-1, N_{\sup})$, where $N_{\sup}$ is defined as the essential supremum of N. Existence of such F_1^* is guaranteed by linear interpolation on the fixed points

$$\{(-1, 0)\} \cup \{(n, F_1(n)) \big| n \in \mathcal{I}_N\}.$$

Now, consider the joint distribution function of N^* and $\boldsymbol{Y}^{[k]}$

$$\mathbb{P}\left((N^*, Y_1, \cdots, Y_k) \le (n, y_1, \cdots, y_k)\right) := C_{\boldsymbol{\Sigma}^{[k,1]}_{\rho_1,\rho_2}}\left(F_1^*(n), F_2(y_1), \cdots, F_2(y_k)\right)$$

for any integer n and $(y_1, \cdots, y_k) \in \mathbb{R}^k$. Then, from Lemma 3, we have

$$\left(\Phi^{-1}(F_1^*(N^*)), \Phi^{-1}(F_2(Y_1)), \cdots, \Phi^{-1}(F_2(Y_k))\right) \sim \mathrm{MVN}\left(\mathbf{0}_{z+1}, \boldsymbol{\Sigma}^{(z)}_{\rho_1,\rho_2}\right)$$

and

$$\Phi^{-1}(F_1^*(N^*)) \big| Y_1 = y_1, \cdots, Y_k = y_k \sim \mathrm{N}(\mu_{[k,1]}, (\sigma_{[k,1]})^2)$$

where $\mu_{[k,1]}$ and $\sigma_{[k,1]}$ are defined in (22) and (23).

Hence, we have

$$
\begin{aligned}
\mathbb{P}\left(N \le n \middle| \boldsymbol{y}^{[k]}\right) &= \frac{\frac{\partial^z}{\partial y_1 \cdots, \partial y_k} C_{\boldsymbol{\Sigma}^{[k_2]}_{\rho_1,\rho_2}}\left(F_1(n), F_2(y_1), \cdots, F_2(y_k)\right)}{f_{\boldsymbol{Y}}(\boldsymbol{y}^{[k]})} \\
&= \frac{\frac{\partial^z}{\partial y_1 \cdots, \partial y_k} C_{\boldsymbol{\Sigma}^{[k_2]}_{\rho_1,\rho_2}}\left(F_1^*(n), F_2(y_1), \cdots, F_2(y_k)\right)}{f_{\boldsymbol{Y}}(\boldsymbol{y}^{[k]})} \\
&= \mathbb{P}\left(N^* \le n \middle| \boldsymbol{y}^{[k]}\right) \\
&= \mathbb{P}\left(\frac{\Phi^{-1}(F_1^*(N^*)) - \mu_{[k,1]}}{\sigma_{[k,1]}} \le \frac{\Phi^{-1}(F_1^*(n)) - \mu_{[k,1]}}{\sigma_{[k,1]}} \middle| \boldsymbol{y}^{[k]}\right) \\
&= \Phi\left(\frac{\Phi^{-1}(F_1^*(n)) - \mu_{[k,1]}}{\sigma_{[k,1]}}\right) \\
&= \Phi\left(\frac{\Phi^{-1}(F_1(n)) - \mu_{[k,1]}}{\sigma_{[k,1]}}\right).
\end{aligned}
\tag{B.1}
$$

for an integer $n \in \mathcal{I}_N \cup \{-1\}$. Finally, the following observation with (B.1)

$$
\mathbb{P}\left(N \le n \middle| \boldsymbol{y}^{[k]}\right) = \mathbb{P}\left(N \le n-1 \middle| \boldsymbol{y}^{[k]}\right) \quad \text{for positive integer } n \notin \mathcal{I}_N \cup \{-1\}
$$

concludes the proof. □

Proof of Lemma 1. We start from the proof of part i. First, for $z = 0$, we have

$$
h_{N,\boldsymbol{Y}^{[k]}}(n, \boldsymbol{y}^{[k]}) = f_1(0)
$$

by definition in Conditional Model 2.

Now, let z be any positive integer. Then, the joint density function of the discrete margin N and the continuous margins $Y_1, \cdots, Y_k$ is given by

$$
h_{N,\boldsymbol{Y}^{[k]}}(n, \boldsymbol{y}^{[k]}) := \mathrm{P}_k\left(n, \boldsymbol{y}^{[k]}\right) - \mathrm{P}_k\left(n-1, \boldsymbol{y}^{[k]}\right)
$$

for integer $n \in \mathbb{Z}$ and $\boldsymbol{y}^{[k]} \in \mathbb{R}^k$, where

$$
\mathrm{P}_k\left(n, y_1, \cdots, y_k\right) := \frac{\partial^z}{\partial y_1 \cdots, \partial y_k} \mathbb{P}\left(N \le n, Y_1 \le y_1, \cdots, Y_k \le y_k\right).
$$

Here, $\mathrm{P}_k\left(n, y_1, \cdots, y_k\right)$ can be written as

$$
\mathrm{P}_k\left(n, \boldsymbol{y}^{[k]}\right) = \mathbb{P}\left(N \le n \middle| y_1, \cdots, y_k\right) f_{\boldsymbol{Y}^{[k]}}(y_1, \cdots, y_k) \tag{B.2}
$$

where $f_{\boldsymbol{Y}^{[k]}}$ is the density function of $\boldsymbol{Y}^{[k]}$. Using Lemma 4, we have

$$
\begin{aligned}
h_{N,\boldsymbol{Y}^{[k]}}(n,\boldsymbol{y}^{[k]}) &= \mathbb{P}\left(N\le n\middle|\boldsymbol{y}^{[k]}\right) f_{\boldsymbol{Y}}(\boldsymbol{y}^{[k]}) - \mathbb{P}\left(N\le n-1\middle|\boldsymbol{y}^{[k]}\right) f_{\boldsymbol{Y}}(\boldsymbol{y}^{[k]}) \\
&= f_{\boldsymbol{Y}^{[k]}}(\boldsymbol{y}^{[k]})\left(\Phi\left(\frac{\Phi^{-1}(F_1(n))-\mu_{[k,1]}}{\sigma_{[k,1]}}\right)-\Phi\left(\frac{\Phi^{-1}(F_1(n-1))-\mu_{[k,1]}}{\sigma_{[k,1]}}\right)\right)
\end{aligned}
$$

for an integer $n\in\mathbb{Z}$ and $\boldsymbol{y}^{[k]}\in\mathbb{R}^k$.

Note that since a continuous random vector $\boldsymbol{Y}^{[k]}$ follows

$$
\boldsymbol{Y}^{[k]} \sim C_{\boldsymbol{\Sigma}_{\rho_2}^{(z)}}(F_2,\cdots,F_2)
$$

for any positive integer z, then the density function of $\boldsymbol{y}^{[k]}$ is given as in (21). The proof of Part i is complete. For brevity, we omit the proof of Part ii because it is trivial from Part i. $\square$

Lemma 5. *Consider the frequency and severities*

$$
(N,\boldsymbol{Y}_N) := \begin{cases}(N,Y_1,\cdots,Y_N), & N\ge 1;\\ 0, & N=0;\end{cases}
$$

in a dependent collective risk model in Model 1 with the assumptions employed in Conditional Model 2 or Conditional Model 3. Then, for a positive integer n and k satisfying $k\le n$, we have

$$
\mathbb{E}\left[Y_k\middle|N=n\right]\mathbb{P}\left(N=n\middle|N>0\right)=\int_0^\infty y\,h_{N^+,\boldsymbol{Y}^{[1]}}(n,y)\mathrm{d}y \tag{B.3}
$$

and

$$
\mathbb{E}\left[S\middle|N=n\right]\mathbb{P}\left(N=n\middle|N>0\right)=n\int_0^\infty y h_{N^+,\boldsymbol{Y}^{[1]}}(n,y)\mathrm{d}y \tag{B.4}
$$

where $h_{N^+,\boldsymbol{Y}^{[1]}}$ is defined in (20) or (27), depending on Conditional Model 2 or Conditional Model 3 assumptions as well as the type of the covariance matrix.

Proof. For a positive integer n and k satisfying $k\le n$, we have

$$
\begin{aligned}
\mathbb{E}\left[Y_k\middle|N=n\right]\mathbb{P}\left(N=n\middle|N>0\right) &= \mathbb{E}\left[Y_k\middle|R=1,N^+=n\right]\mathbb{P}\left(R=1,N^+=n\middle|R=1\right)\\
&= \mathbb{E}\left[Y_k\middle|R=1,N^+=n\right]\mathbb{P}\left(N^+=n\middle|R=1\right)\\
&= \mathbb{E}\left[Y_k\middle|N^+=n\right]\mathbb{P}\left(N^+=n\right)\\
&= \int_0^\infty y\,h_{N^+,\boldsymbol{Y}^{[1]}}(n,y)\mathrm{d}y
\end{aligned}
$$

where the third equality is from the independence assumption between R and $(N^+,Y_1,\cdots,Y_k)$ for any positive integer k satisfying $k<n_{\text{sup}}+1$, and the last equality is from Lemma 1. Finally, the equation (B.3)

derives

$$\mathbb{E}\left[S\middle|N=n\right]\mathbb{P}\left(N=n\middle|N>0\right)=\sum_{j=1}^{n}\mathbb{E}\left[Y_j\middle|N=n\right]\mathbb{P}\left(N=n\middle|N>0\right)$$
$$=n\int_0^\infty y\,h_{N^+,\boldsymbol{Y}^{[1]}}(n,y)\mathrm{d}y.$$

□

Appendix C. Proofs on Proposition 5

Proof. The proof of the first part is from the following equation

$$\begin{aligned}
\mathbb{E}[S] &= \mathbb{E}\left[\mathbb{E}\left[S\middle|N\right]\right]\\
&= \sum_{n=0}^{\infty}\mathbb{E}\left[S\middle|N=n\right]\mathbb{P}(N=n)\\
&= \sum_{n=1}^{\infty}\mathbb{E}\left[S\middle|N=n\right]\mathbb{P}(N=n)\\
&= \sum_{n=1}^{\infty}\mathbb{E}\left[S\middle|N=n\right]\mathbb{P}\left(N^+=n,R=1\right)\\
&= p\sum_{n=1}^{\infty}\mathbb{E}\left[S\middle|N=n\right]\mathbb{P}\left(N=n\middle|N>0\right)\\
&= p\sum_{n=1}^{\infty}n\int_0^\infty y\,h_{N^+,\boldsymbol{Y}^{[1]}}(n,y)\mathrm{d}y.
\end{aligned}$$

where the last equality is from (B.4). The proof of the second part follows from the following equation

$$\begin{aligned}
\mathbb{E}[NS] &= \mathbb{E}\left[\mathbb{E}\left[NS\middle|N\right]\right]\\
&= \mathbb{E}\left[N\mathbb{E}\left[S\middle|N\right]\right]\\
&= \sum_{n=1}^{\infty}n\mathbb{E}\left[S\middle|N=n\right]\mathbb{P}(N=n)\\
&= p\sum_{n=1}^{\infty}n\mathbb{E}\left[S\middle|N=n\right]\mathbb{P}\left(N=n\middle|N>0\right)\\
&= p\sum_{n=1}^{\infty}n^2\int_0^\infty yh_{N^+,\boldsymbol{Y}^{[1]}}(n,y)\mathrm{d}y
\end{aligned}$$

where the last equality is from (B.4).

Finally, the proof of the second part is from the following observations:

$$
\begin{aligned}
\mathbb{E}\left[NM\middle|N>0\right] &= \mathbb{E}\left[S\middle|N>0\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left[S\middle|N>0, N=n\right] \mathbb{P}\left(N=n\middle|N>0\right) \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left[S\middle|N=n\right] \mathbb{P}\left(N=n\middle|N>0\right) \\
&= \sum_{n=1}^{\infty} n \int_0^{\infty} y h_{N+,\boldsymbol{Y}^{[1]}}(n,y)\mathrm{d}y
\end{aligned}
$$

where the last equality is from (B.4) and

$$
\begin{aligned}
\mathbb{E}\left[M\middle|N>0\right] &= \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N} Y_j\middle|N>0\right] \\
&= \sum_{n=1}^{\infty} \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}\left[Y_j\middle|N>0, N=n\right] \mathbb{P}\left(N=n\middle|N>0\right) \\
&= \sum_{n=1}^{\infty} \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}\left[Y_j\middle|N=n\right] \mathbb{P}\left(N=n\middle|N>0\right) \\
&= \sum_{n=1}^{\infty} \frac{1}{n} \sum_{j=1}^{n} \int_0^{\infty} y\, h_{N+,\boldsymbol{Y}^{[1]}}(n,y)\mathrm{d}y \\
&= \sum_{n=1}^{\infty} \int_0^{\infty} y\, h_{N+,\boldsymbol{Y}^{[1]}}(n,y)\mathrm{d}y
\end{aligned}
$$

where the last equality is from (B.3).

$\square$