Insurance: Mathematics and Economics 109 (2023) 29–51

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Dependence modeling of frequency-severity of insurance claims using waiting time

Guangyuan Gao [a], Jiahong Li [b,*]

[a] Center for Applied Statistics and School of Statistics, Renmin University of China, 100872 Beijing, China
[b] School of Mathematical Sciences, Peking University, 100871 Beijing, China

A R T I C L E I N F O

Article history:
Received April 2022
Received in revised form December 2022
Accepted 21 December 2022
Available online 2 January 2023

JEL classification:
300

Keywords:
Insurance claims
Frequency-severity dependence
Copula regression
Poisson process
Waiting time

A B S T R A C T

Mixed copula approach has been used to jointly model discrete variable of claim counts and continuous variable of claim amounts. We propose to use a copula to link two continuous variables of the waiting time for the second claim and the average claim size. The frequency-severity dependence can be derived using the relationship between the waiting time and the counts of a Poisson process. Assuming a Gaussian copula and a log-normal distributed average claim size, we can investigate the effect of claim counts on the conditional claim severity analytically, which would be difficult in the mixed copula approach. We propose a Monte Carlo algorithm to simulate from the predictive distribution of the aggregated claims amount. In an empirical example, we illustrate the proposed method and compare with other competing methods. It shows that our proposed method provides quite competitive results.

1. Introduction

Pricing insurance contracts is an essential task in non-life actuarial science. Generalized linear models (GLMs) are often used to model the number of claims and the individual claim amount. In this way, the corresponding expectations are modeled by the linear combination of actuarial risk factors. In the classical approach of random sum, claim counts and individual claim amount are assumed to be independent, so the pure premium is simply the product of the expected claim frequency and the expected claim severity.

Some studies indicate that the assumption of independence is not always satisfied in some cases. Garrido et al. (2016) showed that claim frequency and severity are negatively associated in the collision automobile insurance. We discuss two possible reasons for the frequency-severity dependence. For example, there may be an unobserved covariate related to both frequency and severity but in different directions. Missing this covariate in the modeling would lead to a negative frequency-severity dependence. For another example, policyholders would balance between reporting a claim with a higher renewal premium and not reporting a claim with a lower renewal premium (due to bonus-malus system). Such behavior would induce a negative frequency-severity dependence since small claims would not be reported. There is a need for flexible models to accommodate the dependence between claims frequency and severity.

According to the current literature, there are two main dependence modeling approaches: conditional approach and mixed copula approach. In the conditional approach, the number of claims is used as a covariate in the GLM for the average claim size; see Frees et al. (2011a) and Garrido et al. (2016). Lee et al. (2019) extended Garrido et al. (2016) by allowing varying dispersion parameters in the GLM for average claim size. Jeong and Valdez (2020) consider the longitudinal property of a P&C dataset in a collective risk model and extend Garrido et al. (2016)'s argument to the credibility premium.

In the mixed copula approach (Song et al., 2009), two marginal GLMs are fitted to claim counts and average claim size. A mixed copula is employed to link the discrete distribution of claim counts and the continuous distribution of average claim size; see Czado et al. (2012) and Krämer et al. (2013). Shi and Zhao (2020) extended the mixed copula approach to collective risk models and proposed a copula-linked compound distribution. Shi et al. (2015) proposed a three-part model, which splits the frequency model into a binary

* Corresponding author.
E-mail address: lijiahong@pku.edu.cn (J. Li).

https://doi.org/10.1016/j.insmatheco.2022.12.006
0167-6687/

classification model and a zero-truncated claim counts model. Genest and Nešlehová (2007) are concerned that such a mixed copula is not unique. To circumvent this issue, Frees and Wang (2006) assumed continuous latent factor as the random effect in the claim frequency model, while Shi and Valdez (2014) and Oh et al. (2020) used jittered claim counts. Another concern is that it is challenging to study the frequency-severity dependence analytically using copulas. The frequency-severity dependence induced by copula is not clear, and the effect of frequency-severity dependence on the aggregated claims amount is not apparent.

Our work differs from and contributes to the existing literature of the copula approach in the sense that we link two continuous variables by a copula, and we get analytical results for the frequency-severity dependence; see equations (3.6) and (3.11) and Fig. 7. Our proposed method is summarized as follows. First, we link a latent variable of waiting time for a second claim and average claim size by an elliptical copula. Second, we derive the frequency-severity joint model using the relationship between claim counts and waiting time of a Poisson process. The parameters are estimated by the inference functions for margins (IFM) method. Third, we employ a Monte Carlo algorithm to simulate the aggregated claims amount for new policyholders. In addition, by assuming a log-normal distributed average claim size and a Gaussian copula, we can analytically study the effect of dependence on the conditional severity given the claims count.

The rest of the paper is structured as follows. In Section 2, we review two dependence modeling approaches: the conditional approach and the copula approach. In Section 3, we propose a new dependence modeling method using latent variable of waiting time and derive analytic results under certain model assumptions. In Section 4, we conduct an empirical study of a real insurance data `ausprivauto0405` from the *CASdatasets* R package. In Section 5, we conclude the paper with some important findings. The code for this paper is available in https://github.com/Richardljh/Dependence-Modelling.

2. Revision of frequency-severity dependence modeling

We assume to have a sufficiently rich probability space $(\Omega, \mathcal{F}, P)$, carrying all objects to be studied. For each policyholder $i = 1, \ldots, n$, we denote the risk factors by $\boldsymbol{x}_i \in \mathcal{X}$, the number of claims by $N_i \in \mathbb{N}$ (natural numbers), the k-th individual claim amount by $Y_{i,k} \in \mathbb{R}_+$ (positive real numbers) for $k = 1, \ldots, N_i$, and the aggregated claims amount by $S_i \in \mathbb{R}_0$ (non-negative real numbers). We have the following random sum equation between those variables

$$S_i = \begin{cases} 0, & \text{if } N_i = 0 \\ \sum_{k=1}^{N_i} Y_{i,k} & \text{otherwise.} \end{cases} \tag{2.1}$$

For the number of claims, we normally employ a Poisson regression to estimate the claims frequency $\lambda : \mathcal{X} \to \mathbb{R}_+,\ \boldsymbol{x} \mapsto \lambda(\boldsymbol{x})$:

$$N_i \overset{\text{ind.}}{\sim} \text{Poi}(\lambda(\boldsymbol{x}_i)), \ \text{ for } i \in \mathcal{I}, \tag{2.2}$$

where λ is the mean parameter and $\mathcal{I} = \{1, 2, \ldots, n\}$ is the index set of policyholders.

For the individual claim amount, we normally employ a gamma regression to estimate the claim severity $\mu : \mathcal{X} \to \mathbb{R}_+,\ \boldsymbol{x} \mapsto \mu(\boldsymbol{x})$:

$$Y_{i,k} \overset{\text{ind.}}{\sim} \text{Gamma}(\mu(\boldsymbol{x}_i), \phi), \ \text{ for } i \in \mathcal{I}_+,\ k = 1, \ldots, N_i, \tag{2.3}$$

where $\mu(\boldsymbol{x}_i)$ is the mean parameter, ϕ is the dispersion parameter and $\mathcal{I}_+ = \{i : N_i > 0\}$ is the index set of policyholders with claims. We also denote the index set of policyholders without any claims by $\mathcal{I}_0 = \{i : N_i = 0\}$. Note that $\mathcal{I} = \mathcal{I}_+ \cup \mathcal{I}_0 = \{1, 2, \ldots, n\}$. Remark that we consider all the available risk factors $\boldsymbol{x}$ in both models for frequency and severity, however, after model fitting and variable selection, the two models may contain different risk factors.

It can be shown that the sufficient statistics in the model (2.3) is the average claim size $Y_i = S_i/N_i$, i.e., we only need the average claim size rather than individual claim amount to estimate the parameters. The model (2.3) is equivalent to the regression of average claim size $Y_i = S_i/N_i$ on the risk factors $\boldsymbol{x}_i$ with prior weights N_i:

$$Y_i|N_i \overset{\text{ind.}}{\sim} \text{Gamma}(\mu(\boldsymbol{x}_i), \phi/N_i), \ \text{ for } i \in \mathcal{I}_+. \tag{2.4}$$

This equivalence is essential since in some cases only the number of claims and the aggregated claims amount are available. This weighted regression (2.4) can be approximated by the un-weighted version:

$$Y_i|N_i \overset{\text{ind.}}{\sim} \text{Gamma}(\mu(\boldsymbol{x}_i), \phi), \ \text{ for } i \in \mathcal{I}_+. \tag{2.5}$$

This approximation is often accurate for two reasons: (1) the dispersion has a secondary effect on the estimation of the mean parameter; (2) in most cases a majority of policyholders make no more than one claim.

Under the assumption of independence between the number of claims N_i and the individual claims amount $Y_{i,k}$, we have the expected aggregated claims amount (pure premium) as the product of frequency and severity:

$$\mathbb{E}(S_i|\boldsymbol{x}_i) = \mathbb{E}(N_i|\boldsymbol{x}_i)\mathbb{E}(Y_{i,k}|\boldsymbol{x}_i) = \lambda(\boldsymbol{x}_i)\mu(\boldsymbol{x}_i) = \lambda_i\mu_i.$$

There are two existing approaches to relax the constraint of independence between N_i and $Y_{i,k}$. The first approach incorporates the positive number of claims $N_i|N_i > 0$ into the severity regression function $\mu(\cdot)$ as a covariate (Garrido et al., 2016), while the second approach employs a copula to jointly model two marginal regressions (2.2) and (2.5) (Czado et al., 2012).

2.1. First approach: conditional dependence modeling

The first approach considers the systematic effect of claim counts on average claim size. Garrido et al. (2016) incorporate the number of claims into the conditional claim severity regression model (2.4) as follows:

$$Y_i|N_i \overset{\text{ind.}}{\sim} \text{Gamma}(\mu(\boldsymbol{x}_i, N_i), \phi), \text{ for } i \in \mathcal{I}_+.$$

Under this setting, we have the expected aggregated claims amount as

$$\mathbb{E}(S_i|\boldsymbol{x}_i) = \mathbb{E}(\mathbb{E}(S_i|N_i, \boldsymbol{x}_i)|\boldsymbol{x}_i) = \mathbb{E}(\mathbb{E}(N_i Y_i|N_i, \boldsymbol{x}_i)|\boldsymbol{x}_i) = \mathbb{E}(N_i \mu(\boldsymbol{x}_i, N_i)|\boldsymbol{x}_i). \tag{2.6}$$

Garrido et al. (2016) showed that if $\mu(\cdot)$ is specified as $\mu(\boldsymbol{x}_i, N_i) = \tilde{\mu}(\boldsymbol{x}_i)e^{\theta N_i}$, the expected aggregated claims amount is $\mathbb{E}(S_i|\boldsymbol{x}_i) = \tilde{\mu}(\boldsymbol{x}_i)M'_{N_i}(\theta)$, where $M'_{N_i}(\theta)$ is the derivative of the moment generating function of N_i w.r.t. θ. This particular model restricts the dependence as log-linearity.

2.2. Second approach: dependence modeling with copulas

The second approach considers the joint distribution of claim counts and average claim size under the copula framework, where the empirical dependence is modeled implicitly in the joint distribution. Czado et al. (2012) and Shi et al. (2015) apply a mixed copula (Song et al., 2009) to jointly model the frequency and the severity. While Czado et al. (2012) use the same marginal regressions as (2.2) and (2.5), Shi et al. (2015) split the frequency model into a logistic model and a zero-truncated negative binomial model. A weakness of this approach is that we do not have any analytical results. Unlike conditional dependence modeling, it is unclear how the number of claims affects the severity of claims.

3. Jointly modeling of waiting time for claim and average claim size by copulas

In this paper, we propose to use the continuous variable of waiting time rather than the discrete variable of claim count in the copula. The relationship between waiting time and Poisson count is well established in a Poisson process, and it is straightforward to get the (induced) joint distribution of claim count and average claim size. It is worth noting that theoretically, the non-uniqueness of the copula issue still exists since only the discrete claim counts are observed rather than the waiting time. However, in our application, only parametric copulas are used, and the marginal regression models contain continuous covariates, so the identifiability is not a big issue in this paper.

3.1. Model specification

We first introduce the latent variable of waiting time, then we propose the jointly modeling under the copula framework, finally we derive the analytical results for the frequency-severity dependence under certain model specification.

3.1.1. Latent variable of waiting time

We assume that the claim arrival process follows a Poisson process with annual intensity of λ. For policyholder i, we denote the waiting time between the $k-1$ and k-th claim by $T_{i,k}$. For policyholder i, the waiting times are independent and identically distributed as an exponential distribution:

$$T_{i,k} \overset{\text{i.i.d.}}{\sim} \text{Exp}(1/\lambda(\boldsymbol{x}_i)), \text{ for } k = 1, 2, \ldots, \tag{3.1}$$

where we model the annual intensity (claims frequency) by $\lambda(\boldsymbol{x}_i)$.

We consider policies with one year coverage and denote the censored number of claims in one policy year by $N_i \in \{0, 1, 2+\}$. We have the following correspondence between the censored number of claims and the waiting times:

$$N_i = \begin{cases} 0, & \text{if and only if } T_{i,1} > 1; \\ 1, & \text{if and only if } T_{i,1} \le 1, \ T_{i,1} + T_{i,2} > 1; \\ 2+, & \text{if and only if } T_{i,1} \le 1, \ T_{i,1} + T_{i,2} \le 1, \end{cases} \tag{3.2}$$

where $N_i = 2+$ denotes the event that the number of claims is larger than 1. For most third-party motor liability insurance datasets, a tiny proportion of policyholders make more than 2 claims. So it is reasonable to combine those rare cases. However, a limitation of the proposed method is that it is not suitable for some claim frequency datasets with a high proportion of more than 2 claims.

Note that we have slightly abused the notation N_i; it denotes the uncensored number in Section 2. We have the following probability mass function for the censored number of claims

$$\Pr(N_i = m) = \begin{cases} e^{-\lambda(\boldsymbol{x}_i)}, & \text{for } m = 0; \\ \lambda(\boldsymbol{x}_i)e^{-\lambda(\boldsymbol{x}_i)}, & \text{for } m = 1; \\ 1 - e^{-\lambda(\boldsymbol{x}_i)} - \lambda(\boldsymbol{x}_i)e^{-\lambda(\boldsymbol{x}_i)}, & \text{for } m = 2+. \end{cases} \tag{3.3}$$

We denote by $N_i^* \in \{1, 2+\}$ the censored number of claims given that at least one claim occurs in a year. This random variable N_i^* is $\sigma(\{N_i > 0\})$-measurable. We denote by $T_i \in (0, \infty)$ the total waiting time until 2 claims occur given that the first claim occurs in a year. This random variable T_i is $\sigma(\{T_{i,1} \le 1\})$-measurable. In Appendix A, we derive the distribution of T_i as

$$F_{T_i}(t) = \begin{cases} \frac{1-e^{-\lambda(\boldsymbol{x}_i)t}-\lambda(\boldsymbol{x}_i)te^{-\lambda(\boldsymbol{x}_i)t}}{1-e^{-\lambda(\boldsymbol{x}_i)}}, & \text{for } t \le 1; \\ \frac{1-e^{-\lambda(\boldsymbol{x}_i)}-\lambda(\boldsymbol{x}_i)e^{-\lambda(\boldsymbol{x}_i)t}}{1-e^{-\lambda(\boldsymbol{x}_i)}}, & \text{for } t > 1. \end{cases} \tag{3.4}$$

The distribution (3.4) is a key component in our proposed method since it will appear in the joint likelihood function of N_i^* and Y_i; see (3.8). Since $\sigma(\{N_i > 0\})$ equals to $\sigma(\{T_{i,1} \le 1\})$ and both of them are sub-σ-algebras of $\mathcal{F}$, the two variables N_i^* and T_i are defined on the same measurable space $(\Omega, \sigma(\{N_i > 0\}))$. The connection between them is

$$N_i^* = \begin{cases} 1, & \text{if and only if } T_i > 1; \\ 2+, & \text{if and only if } T_i \le 1. \end{cases} \tag{3.5}$$

Remark that if over-dispersion is observed in the claim counts, one could consider using a mixture of Poisson regressions or a negative binomial regression. Another concern is that the claim counts contain too many zeros compared to the Poisson distribution. One could consider using a zero-inflated Poisson distribution, which is a mixture of a Poisson distribution and a probability mass at 0 (a special case of the mixture of Poisson distributions; see Appendix B).

For both the mixture of Poisson distributions and the negative binomial distribution, the distribution of waiting time T_i in (3.4) needs to be modified accordingly. We derive the distribution of T_i for the mixture of Poisson distributions in equation (B.1) and for the zero-inflated Poisson distribution in equation (B.2) in Appendix B. For the negative binomial distribution, we show that the waiting time $T_{i,k}$ follows a Pareto type II distribution in equation (B.3) in Appendix B; however, there is no closed form for the distribution of T_i.

3.1.2. Jointly modeling specification

We establish the (indirect) relationship between N_i^* and Y_i for $i \in \mathcal{I}_+$ via a copula on two continuous variables T_i and Y_i:

$$F_{T_i,Y_i}(t, y|\boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho) = C\left(F_{T_i}(t|\boldsymbol{x}_i; \boldsymbol{\alpha}), F_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi); \rho\right), \tag{3.6}$$

where $F_{T_i,Y_i} : \mathbb{R}_+^2 \to (0, 1)$ is the joint distribution of T_i and Y_i, $F_{T_i}(\cdot; \boldsymbol{\alpha})$ is the marginal distribution of T_i with the regression coefficients $\boldsymbol{\alpha}$, e.g., the frequency regression function can be chosen as $\lambda(\boldsymbol{x}_i) = \exp(\boldsymbol{\alpha}^\top \boldsymbol{x}_i)$, and $F_{Y_i}(\cdot; \boldsymbol{\beta}, \phi)$ is the marginal distribution of Y_i with dispersion ϕ and the regression coefficients $\boldsymbol{\beta}$, e.g., the severity regression function can be chosen as $\mu(\boldsymbol{x}_i) = \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_i)$. The dependence between waiting time and severity is opposite to the frequency-severity dependence in the direction. Thus, a negative (positive) dependence parameter ρ implies a positive (negative) frequency-severity dependence.

Following (3.5), the induced joint distribution of N_i^* and Y_i is then given by

$$\begin{aligned} &\Pr(N_i^* = m, Y_i \le y|\boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho) \\ &= \begin{cases} F_{T_i,Y_i}(1, y|\boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho), & \text{for } m = 2+; \\ F_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi) - F_{T_i,Y_i}(1, y|\boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho), & \text{for } m = 1. \end{cases} \end{aligned} \tag{3.7}$$

The mixed density $f_{N_i^*,Y_i} = \partial \Pr(N_i^* = m, Y_i \le y)/\partial y$ can be derived as

$$\begin{aligned} &f_{N_i^*,Y_i}(m, y|\boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho) \\ &= \begin{cases} C_2\left(F_{T_i}(1|\boldsymbol{x}_i; \boldsymbol{\alpha}), F_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi); \rho\right) f_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi), & \text{for } m = 2+; \\ \left[1 - C_2\left(F_{T_i}(1|\boldsymbol{x}_i; \boldsymbol{\alpha}), F_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi); \rho\right)\right] f_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi), & \text{for } m = 1, \end{cases} \end{aligned} \tag{3.8}$$

where $f_{Y_i}(\cdot; \boldsymbol{\beta}, \phi)$ is the marginal probability density function of average claim size for policyholder $i \in \mathcal{I}_+$, $C_2(\cdot, \cdot)$ is the partial derivative of copula $C(\cdot, \cdot)$ with respect to the second variable.

We focus on two elliptical copulas, Gaussian copula and t copula. For detailed properties of these two copula families, we refer to Joe (2014). The Gaussian copula $C^G(u_1, u_2)$ is given by

$$C^G(u_1, u_2) = \Phi_2\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho\right),$$

where $\Phi_2(z_1, z_2; \rho)$ is the standard bivariate normal distribution of z_1 and z_2 with correlation ρ, and $\Phi^{-1} : (0, 1) \to \mathbb{R}$ is the inverse function of the standard normal distribution. The t copula with degree of freedom ν is given by

$$C^t(u_1, u_2; \rho, \nu) = T_{2,\nu}(T_\nu^{-1}(u_1), T_\nu^{-1}(u_2); \rho),$$

where $T_{2,\nu}(w_1, w_2; \rho)$ is the standard bivariate t distribution of w_1 and w_2 with degree of freedom ν and correlation ρ, and $T_\nu^{-1} : (0, 1) \to \mathbb{R}$ is the inverse function of the univariate t distribution with degree of freedom ν.

For Gaussian copula, we can derive that (see Appendix C)

$$f_{N_i^*,Y_i}(m, y|\boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho) = \begin{cases} \Phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right) f_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi), & \text{for } m = 2+; \\ \Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right) f_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi), & \text{for } m = 1, \end{cases}$$

$z_{i,1}$ is the normal score of $u_{i,1} := \Pr(T_i \le 1|\boldsymbol{x}_i; \boldsymbol{\alpha})$,

$$z_{i,1} = \Phi^{-1}(u_{i,1}) = \Phi^{-1}(F_{T_i}(1|\boldsymbol{x}_i; \boldsymbol{\alpha})) = \Phi^{-1}(\Pr(N_i^* = 2 + |\boldsymbol{x}_i; \boldsymbol{\alpha})),$$

and $z_{i,2}$ is the normal score of $u_{i,2} := \Pr(Y_i \le y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi)$,

$$z_{i,2} = \Phi^{-1}(u_{i,2}) = \Phi^{-1}(F_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi)).$$

When $\rho = 0$, we have

$$\Phi\left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}}\right) = \Phi\left(z_{i,1}\right) = \Pr(N_i^* = 2+|\boldsymbol{x}_i; \boldsymbol{\alpha})$$

and

$$\Phi\left(-\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}}\right) = 1 - \Phi\left(z_{i,1}\right) = \Pr(N_i^* = 1|\boldsymbol{x}_i; \boldsymbol{\alpha}).$$

Therefore, when the correlation ρ between T_i and Y_i is zero, we return to the independent case with the joint probability density function as

$$f_{N_i^*,Y_i}(m, y|\boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi) = \Pr(N_i^* = m|\boldsymbol{x}_i; \boldsymbol{\alpha}) \times f_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi).$$

For t copula, we can derive that (see Appendix C)

$$f_{N_i^*,Y_i}(m, y|\boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho) = \begin{cases} T_{\nu+1}\left(\frac{w_{i,1} - \rho w_{i,2}}{\sqrt{(1-\rho^2)\left(\nu + w_{i,2}^2\right)/(\nu+1)}}\right) f_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi), & \text{for } m = 2+; \\ T_{\nu+1}\left(-\frac{w_{i,1} - \rho w_{i,2}}{\sqrt{(1-\rho^2)\left(\nu + w_{i,2}^2\right)/(\nu+1)}}\right) f_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi), & \text{for } m = 1, \end{cases}$$

where

$$w_{i,1} = T_\nu^{-1}(F_{T_i}(1|\boldsymbol{x}_i; \boldsymbol{\alpha})) = T_\nu^{-1}(\Pr(N_i^* = 2+|\boldsymbol{x}_i; \boldsymbol{\alpha}))$$

and

$$w_{i,2} = T_\nu^{-1}(F_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi)).$$

Remark that an uncorrelated bivariate t distribution with $\rho = 0$ does not imply independence (McNeil et al., 2015). From this aspect, we prefer the Gaussian copula, which nests the independence case. On the other hand, t copula has a non-zero tail dependence which may be undesirable for some datasets. We compare those two copulas in Section 4 of the empirical study.

3.1.3. Analytical results under the log-normal distributed average claim size and the Gaussian copula

Assume that the logarithm of average claim size follows a normal distribution:

$$\log Y_i \overset{\text{ind.}}{\sim} \mathcal{N}(\mu(\boldsymbol{x}_i), \sigma^2), \text{ for } i \in \mathcal{I}_+,$$

where $\mu(\boldsymbol{x}_i) = \boldsymbol{\beta}^\top \boldsymbol{x}_i$. For Gaussian copula, following (3.8), the conditional distribution of N_i^* given Y_i is

$$f_{N_i^*|Y_i}(m|y, \boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \rho) = \begin{cases} \Phi\left(\frac{z_{i,1} - \rho\left(\frac{\log y_i - \mu(\boldsymbol{x}_i)}{\sigma}\right)}{\sqrt{1-\rho^2}}\right), & \text{for } m = 2+; \\ \Phi\left(-\frac{z_{i,1} - \rho\left(\frac{\log y_i - \mu(\boldsymbol{x}_i)}{\sigma}\right)}{\sqrt{1-\rho^2}}\right), & \text{for } m = 1. \end{cases} \tag{3.9}$$

The conditional distribution of Y_i given N_i^* is

$$f_{Y_i|N_i^*}(y|m, \boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \rho) = \begin{cases} \Phi\left(\frac{z_{i,1} - \rho\left(\frac{\log y_i - \mu(\boldsymbol{x}_i)}{\sigma}\right)}{\sqrt{1-\rho^2}}\right) \frac{f_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \sigma^2)}{u_{i,1}}, & \text{for } m = 2+; \\ \Phi\left(-\frac{z_{i,1} - \rho\left(\frac{\log y_i - \mu(\boldsymbol{x}_i)}{\sigma}\right)}{\sqrt{1-\rho^2}}\right) \frac{f_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \sigma^2)}{1-u_{i,1}}, & \text{for } m = 1, \end{cases} \tag{3.10}$$

where $u_{i,1} = F_{T_i}(1|\boldsymbol{x}_i; \boldsymbol{\alpha}) = \Pr(N_i^* = 2+|\boldsymbol{x}_i; \boldsymbol{\alpha})$ and $z_{i,1} = \Phi^{-1}(u_{i,1})$. Appendix D shows that the conditional expectation of $\log Y_i$ given N_i^* is

$$\mathbb{E}(\log Y_i|N_i^* = m, \boldsymbol{x}_i) = \mu(\boldsymbol{x}_i) + \begin{cases} -\frac{\rho\sigma}{u_{i,1}}\phi\left(\Phi^{-1}(u_{i,1})\right), & \text{for } m = 2+; \\ \frac{\rho\sigma}{1-u_{i,1}}\phi\left(\Phi^{-1}(u_{i,1})\right), & \text{for } m = 1. \end{cases} \tag{3.11}$$

It is worth noting that when $\rho > 0$ ($\rho < 0$), the conditional expectation $\mathbb{E}(\log Y_i|N_i^* = 1)$ is larger (smaller) than $\mathbb{E}(\log Y_i|N_i^* = 2+)$. The conditional expectation depends on the claim frequency $\lambda(\boldsymbol{x}_i)$ via the quantity $u_{i,1}$. Also, note that the discrepancy between the two conditional expectations in (3.11) is increasing with σ and $|\rho|$, as expected. In Fig. 7 of Section 4, we show the curves of the conditional expectation (3.11) against the claims frequency. Our proposed method facilitates an analytical investigation of frequency-severity dependence.

3.2. Parameter estimation

We consider three estimation methods: global maximum likelihood estimation (MLE), two-stage estimation, and inference functions for margins (IFM) method. Those three methods are compared in Section 3.2.4 of a simulated example. Generally, we prefer the IFM method since it has the least computing burden and provides quite similar estimates compared with the other two methods; see Joe (2014). The estimation procedures for Gaussian copula and t copula are similar. In this section, we mainly focus on the Gaussian copula. Remark that for the degrees of freedom of the t copula, we can either treat it as a pre-defined hyperparameter or find its MLE directly (Demarta and McNeil, 2005); see Section 4.2.2.

3.2.1. Global maximum likelihood estimation

The joint log-likelihood for a data set containing n observations $\{(m_1, y_1, \boldsymbol{x}_1), \ldots, (m_n, y_n, \boldsymbol{x}_n)\}$ is given by

$$\begin{aligned} l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho) &= \sum_{i\in\mathcal{I}_0} \log f_{N_i}(0|\boldsymbol{x}_i; \boldsymbol{\alpha}) + \sum_{i\in\mathcal{I}_+} \log\Big((1 - f_{N_i}(0|\boldsymbol{x}_i; \boldsymbol{\alpha})) f_{N_i^*, Y_i}(m_i, y_i|\boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho)\Big) \\ &= \sum_{i\in\mathcal{I}_0} \log \Pr(N_i = 0|\boldsymbol{x}_i; \boldsymbol{\alpha}) + \sum_{i\in\mathcal{I}_+} \log \Pr(N_i > 0|\boldsymbol{x}_i; \boldsymbol{\alpha}) + \sum_{i\in\mathcal{I}_+} \log f_{Y_i}(y_i|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi) + \\ &\quad \sum_{i\in\mathcal{I}_+} \left\{\mathbb{1}_1(m_i) \log C_2\left(F_{T_i}(1|\boldsymbol{x}_i; \boldsymbol{\alpha}), F_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi); \rho\right)\right. \\ &\quad \left. + \mathbb{1}_{2+}(m_i) \log\left[1 - C_2\left(F_{T_i}(1|\boldsymbol{x}_i; \boldsymbol{\alpha}), F_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi); \rho\right)\right]\right\}, \end{aligned} \tag{3.12}$$

where C_2 is the partial derivative of either a Gaussian copula $C^G(\cdot, \cdot; \rho)$ or a t copula $C^t(\cdot, \cdot; \rho, \nu)$ with respect to the second variable.

For the Gaussian copula and $\rho = 0$, the joint log-likelihood is given by

$$\begin{aligned} l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi) &= \sum_{i\in\mathcal{I}_0} \log \Pr(N_i = 0|\boldsymbol{x}_i; \boldsymbol{\alpha}) + \sum_{i\in\mathcal{I}_+} \mathbb{1}_1(m_i) \log \Pr(N_i = m_i|\boldsymbol{x}_i; \boldsymbol{\alpha}) + \\ &\quad \sum_{i\in\mathcal{I}_+} \mathbb{1}_{2+}(m_i) \log \Pr(N_i = m_i|\boldsymbol{x}_i; \boldsymbol{\alpha}) + \sum_{i\in\mathcal{I}_+} \log f_{Y_i}(y_i|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi) \\ &= \sum_{i\in\mathcal{I}} \log f_{N_i}(m_i|\boldsymbol{x}_i; \boldsymbol{\alpha}) + \sum_{i\in\mathcal{I}_+} \log f_{Y_i}(y_i|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi), \end{aligned} \tag{3.13}$$

which corresponds to the independent case. The probability mass function f_{N_i} is specified in (3.3), i.e., we have censored the number of claims $N_i \in \{0, 1, 2+\}$. We apply the Fisher's scoring method to estimate the parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho$; see the scoring functions (E.8) in Appendix E.

3.2.2. Two-stage estimation

The joint log-likelihood (3.12) suggests a two-stage estimation strategy (Shi and Zhao, 2020). The two-stage estimation strategy is detailed as follows. First, we calculate the MLE $\hat{\boldsymbol{\alpha}}$ in the claims frequency (marginal) model (2.2). Second, holding MLE $\hat{\boldsymbol{\alpha}}$, we maximize the following log-likelihood to find the MLEs of $\boldsymbol{\beta}, \rho$:

$$\begin{aligned} l^{2s}(\boldsymbol{\beta}, \phi, \rho) &= \sum_{i\in\mathcal{I}_+} \log f_{Y_i}(y_i|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi) \\ &\quad + \sum_{i\in\mathcal{I}_+} \left\{\mathbb{1}_1(m_i) \log C_2\left(F_{T_i}(1|\boldsymbol{x}_i; \hat{\boldsymbol{\alpha}}), F_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi); \rho\right)\right. \\ &\quad \left. + \mathbb{1}_{2+}(m_i) \log\left[1 - C_2\left(F_{T_i}(1|\boldsymbol{x}_i; \hat{\boldsymbol{\alpha}}), F_{Y_i}(y|\boldsymbol{x}_i; \boldsymbol{\beta}, \phi); \rho\right)\right]\right\}. \end{aligned} \tag{3.14}$$

By adding a fixed value (not containing parameters to be estimated)

$$-\sum_{i\in\mathcal{I}_+} \left[\mathbb{1}_1(m_i) \log \Pr(N_i^* = 1|\boldsymbol{x}_i; \hat{\boldsymbol{\alpha}}) + \mathbb{1}_{2+}(m_i) \log \Pr(N_i^* = 2+|\boldsymbol{x}_i; \hat{\boldsymbol{\alpha}})\right]$$

to the right hand side of (3.14), we find that maximizing $l^{2s}(\boldsymbol{\beta}, \phi, \rho)$ is equivalent to maximizing the log-likelihood of conditional severity model as follows

$$\sum_{i\in\mathcal{I}_+} \log f_{Y_i|N_i^*}(y_i|m_i, \boldsymbol{x}_i; \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \phi, \rho).$$

Remark that the key point of two-stage estimation procedure is to determine the conditional distribution of Y_i given N_i^*.

3.2.3. Inference functions for margins (IFM) method

The inference functions for margins (IFM) method uses the "plug-in" normal scores when calculating the MLE of the copula association parameter ρ. The IFM method is detailed as follows. First, we calculate the MLEs $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ in the two marginal models (2.2) and (2.5), respectively, i.e., we maximize the independent log-likelihood (3.13). Second, we find the MLE of the copula association parameter ρ:

$$\hat{\rho} = \arg\max_{\rho} \sum_{i \in \mathcal{I}_+} \Big\{ \mathbb{1}_1(m_i) \log C_2 \left(F_{T_i}(1|\boldsymbol{x}_i; \hat{\boldsymbol{\alpha}}), F_{Y_i}(y|\boldsymbol{x}_i; \hat{\boldsymbol{\beta}}, \hat{\phi}); \rho \right) + \mathbb{1}_{2+}(m_i) \log \Big[1 - C_2 \left(F_{T_i}(1|\boldsymbol{x}_i; \hat{\boldsymbol{\alpha}}), F_{Y_i}(y|\boldsymbol{x}_i; \hat{\boldsymbol{\beta}}, \hat{\phi}); \rho \right) \Big] \Big\}. \tag{3.15}$$

The IFM method has the least computing burden among the three estimation methods.

3.2.4. A simulation study

In this section we examine the finite-sample properties of three estimation methods discussed above. We assume that the average claim size follows either a gamma distribution or a log-normal distribution. We generate the covariates x_{i1} from a uniform distribution U(18, 65) and x_{i2} from a Bernoulli distribution B(1, 0.5) for $i = 1, \dots, n$. We set the regression function for claim frequency as

$$\lambda(\boldsymbol{x}_i) = \exp\{-0.5 - 0.04x_{i1} + 0.3x_{i2}\}, \quad i = 1, \dots, n.$$

For the gamma distributed claim size, we set the dispersion as $\phi = 2$ and the mean as

$$\mu(\boldsymbol{x}_i) = \exp\{-1 + 0.1x_{i1} - 0.2x_{i2}\}, \quad i = 1, \dots, n.$$

For the log-normal distributed claim size, we set the dispersion as $\sigma^2 = 2^2$ and the logged mean as

$$\mu(\boldsymbol{x}_i) = -1 + 0.1x_{i1} - 0.2x_{i2}, \quad i = 1, \dots, n.$$

We consider three different Gaussian copulas with the association value $\rho \in \{0.2, 0.5, 0.9\}$. Thus, we have six different scenarios in total. We simulate a data set of sample size $n = 5000$ for 50 times for each scenario.

The estimation results are summarized in Tables 7 and 8. We have similar observations for both severity distributions. We find no substantial difference among the estimated parameters using the three estimation methods when $\rho = 0.2, 0.5$. For the case of $\rho = 0.9$, the estimated parameters in the frequency model using the three estimation methods are quite close, while the estimated parameters in the severity model using the IFM method have a larger bias than those from the other two estimation methods. According to the current literature, for most data sets the association between frequency and severity is not very high (e.g. 0.9). Hence, we conclude that the IFM method performs as well as both the global MLE and the two-stage estimation when the proposed model is applied to a real data set. Furthermore, a t copula with the degree of freedom 10 is fitted, and the estimated parameters are quite close to those from the Gaussian copula. So we do not provide detailed results for the t copula.

3.3. Predictive distribution of the aggregated claims amount

The predictive distribution of aggregated claim amount is crucial for risk management. Two Monte Carlo algorithms are proposed to simulate the aggregated claims amount. The first one is for the general case while the second one is for the model discussed in Section 3.1.3.

3.3.1. General case

We employ the following Monte Carlo algorithm to simulate the correlated number of claims and average claim size for a new policyholder with risk factors $\boldsymbol{x}_{n+1}$.

Algorithm 1 The first Monte Carlo simulation.

Generate the waiting time for the first claim $T_{n+1,1} \sim \text{Exp}(1/\hat{\lambda}(\boldsymbol{x}_{n+1}))$.
if $T_{n+1,1} > 1$ **then**
  return The number of claims as $N_{n+1} = 0$ and the aggregated claims amount as $S_{n+1} = 0$.
else
  Generate a pair of correlated normal scores $(z_{n+1,1}, z_{n+1,2})$ or $(w_{n+1,1}, w_{n+1,2})$ from a bivariate normal distribution or a bivariate t distribution with correlation $\hat{\rho}$.
  Recover T_{n+1} by inverting its distribution function $F_{T_{n+1}}(t|\boldsymbol{x}_{n+1}; \hat{\boldsymbol{\alpha}})$.
  Recover Y_{n+1} by inverting its distribution function $F_{Y_{n+1}}(y|\boldsymbol{x}_{n+1}; \hat{\boldsymbol{\beta}}, \hat{\phi})$.
  Set $T \leftarrow T_{n+1}$, $N \leftarrow 2$.
  if $T > 1$ **then**
    return The number of claims as $N_{n+1} = N - 1$ and the aggregated claims amount as $S_{n+1} = N_{n+1}Y_{n+1}$.
  else
    repeat
      Generate the next waiting time $T_{n+1,*}$ from $\text{Exp}(1/\hat{\lambda}(\boldsymbol{x}_{n+1}))$.
      Set $T \leftarrow T + T_{n+1,*}$, $N \leftarrow N + 1$.
    until $T > 1$
    return The number of claims as $N_{n+1} = N - 1$ and the aggregated claims amount as $S_{n+1} = N_{n+1}Y_{n+1}$.
  end if
end if

Repeating Algorithm 1 for many times leads to an empirical predictive distribution of the aggregated claims amount for policyholder $n+1$ with risk factor $\boldsymbol{x}_{n+1}$. This simulated predictive distribution only incorporates the process variance induced by the assumed stochastic model. The estimation variance can be incorporated by an additional step of sampling regression coefficients from a multivariate normal distribution with mean of the MLEs and covariance of the inverse Fisher's information matrix. The estimation variance accounts for the estimation error due to finite sample size.

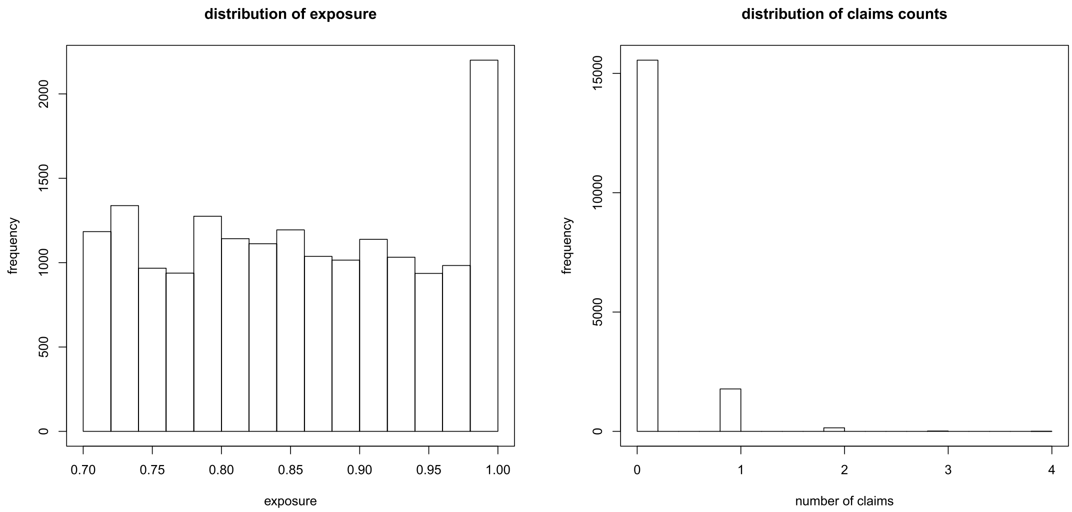

Fig. 1. Left: The distribution of exposures. Right: The distribution of claim counts.

3.3.2. *Log-normal distributed average claim size and Gaussian copula*

In Section 3.1.3, the conditional distributions can be derived under the assumption of the log-normal distributed average claim size and the Gaussian copula. So we can employ the following Monte Carlo simulation which is much easier and faster than Algorithm 1 for the general case.

Algorithm 2 The second Monte Carlo simulation.

1: Generate the number of claims N_{n+1} from a Poisson distribution with mean of $\hat{\lambda}(\boldsymbol{x}_{n+1})$.
2: **if** $N_{n+1} = 0$ **then**
3: **return** The number of claims as N_{n+1} and the aggregated claims amount as $S_{n+1} = 0$.
4: **else**
5: Generate the average claim size Y_{n+1} from the conditional distribution (3.10) given the number of claims.
6: **return** The number of claims as N_{n+1} and the aggregated claims amount as $S_{n+1} = N_{n+1}Y_{n+1}$.
7: **end if**

4. An empirical study

We consider the ausprivauto0405 data set in the *CASdatasets* R package. It is based on one-year compulsory third party car insurance policies taken out in 2004 and 2005. The number of claims and the total claims amounts are available for those policies. For each policy, five risk factors are provided, including vehicle value, vehicle age, vehicle body, gender and driver age.

4.1. *Data description*

A preliminary analysis shows that the annual claims frequency and average claim size are negatively related to the exposure for the partially exposed policies (i.e., less than 0.7 years exposure). Including those partially exposed policies would distort the frequency-severity dependence. Thus, we only consider the nearly fully exposed policies in the following, i.e., the $n = 1,7491$ policies with more than 0.7 years exposure. For those nearly fully exposed policies, the annual claims frequency and average claim size are not strongly correlated with the exposure, which is commonly observed in a typical insurance claim data set.

The total exposure of the portfolio is $\sum_{i=1}^{n} e_i = 14,964.57$ years. The total claim count is $\sum_{i=1}^{n} N_i = 2,112$. The empirical claims frequency is $\bar{\lambda} = \sum_{i=1}^{n} N_i / \sum_{i=1}^{n} e_i = 14.11\%$. The empirical claim severity is $\bar{\mu} = \sum_{i \in \mathcal{I}_+} Y_i / |\mathcal{I}_+| = 1,486.00$. We show the distributions of exposure, claim counts and average claim size in Figs. 1 and 2 (left). A large proportion of policies are fully exposed with 1 year. Most policies do not make a claim and very rare policies make more than 2 claims. The distribution of average claim size has a heavy right tail.

We investigate five available actuarial risk factors: vehicle value VehValue, vehicle age VehAge (4 levels), vehicle body VehBody (11 levels), gender Gender (2 levels) and driver age DrivAge (6 levels). Except for vehicle value VehValue, all the other risk factors are categorical variables. The distributions of the five risk factors are shown in Figs. 2 (right), 3 and 4, respectively. The range of vehicle values is from AUD 0.19 thousands to AUD 34.56 thousands. The old vehicles are nearly twice as many as the youngest ones. Most vehicles are sedan, hatchback, or station wagon. Female drivers are more than male drivers. Older working drivers and working drivers are more than other age groups. We select the level containing the most policyholders as the reference level, i.e., older working female drivers with old sedan cars. Applying the dummy coding, the risk factors $(\texttt{VehValue}_i, \texttt{VehAge}_i, \texttt{VehBody}_i, \texttt{Gender}_i, \texttt{DrivAge}_i)$ can be encoded in a 20-dimensional vector $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,20}) \in \mathbb{R}_+ \times \{0,1\}^{19}$, where $x_{i,1} = \texttt{VehValue}_i$.

As a preliminary analysis of frequency-severity dependence, we draw the boxplot of the logged average claim size for different numbers of claims in Fig. 5. It shows that the average claim size tends to increase with the number of claims.

We denote the index set of the whole portfolio by $\mathcal{I} = \{1, 2, \ldots, n = 17491\}$, and the index set of 1,938 policies with at least one claim by $\mathcal{I}_+ = \{i \in \mathcal{I} : N_i > 0\}$. We split the whole portfolio $\mathcal{I}$ into a learning set $\mathcal{I}_L$ of 13,995 policies (80% of the whole portfolio) and its complement set as a test set $\mathcal{I}_T$, i.e., $\mathcal{I} - \mathcal{I}_L = \mathcal{I}_T$. The empirical claim frequencies on the learning set $\mathcal{I}_L$ and the test set $\mathcal{I}_T$

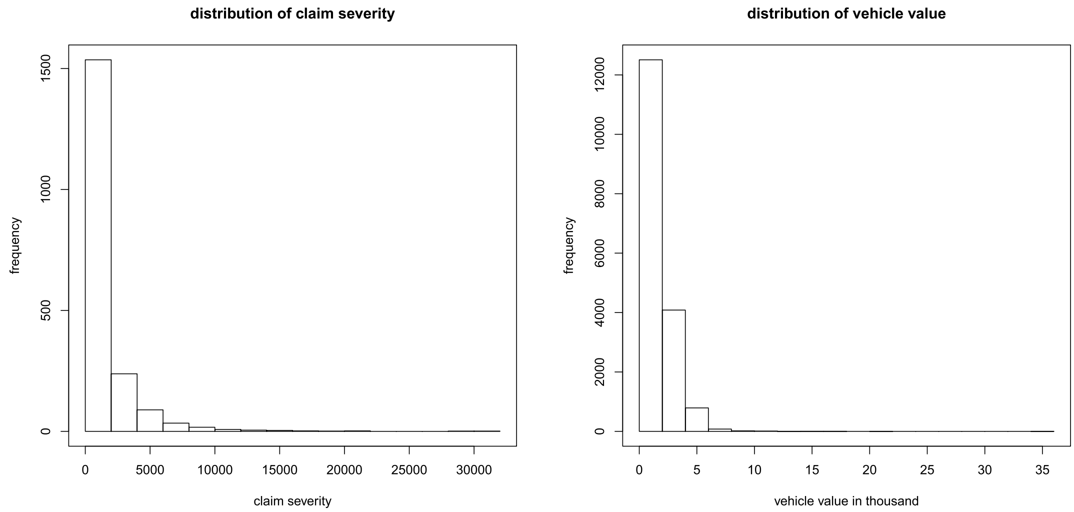

Fig. 2. Left: The distribution of average claim size. Right: The distribution of vehicle values.

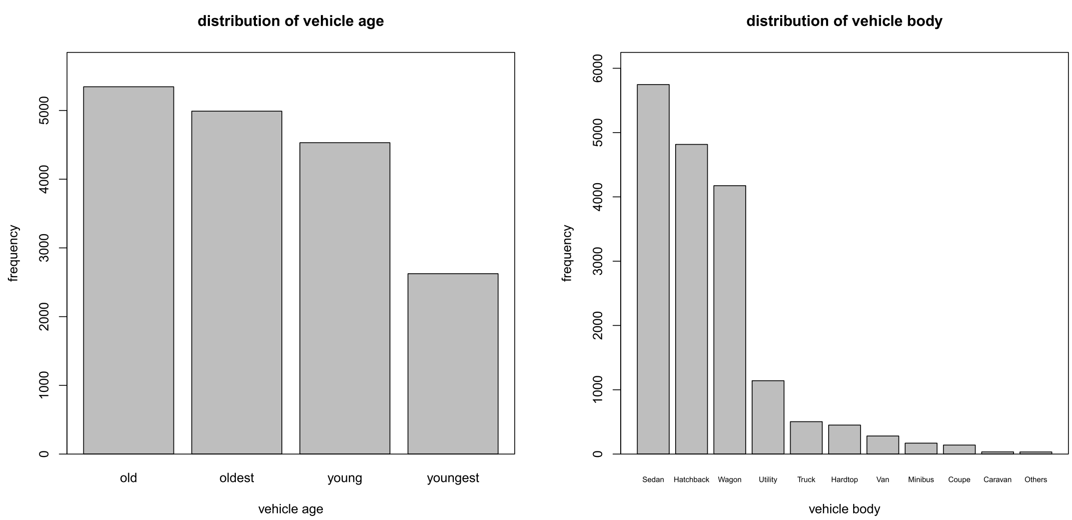

Fig. 3. Left: The distribution of vehicle age. Right: The distribution of vehicle body.

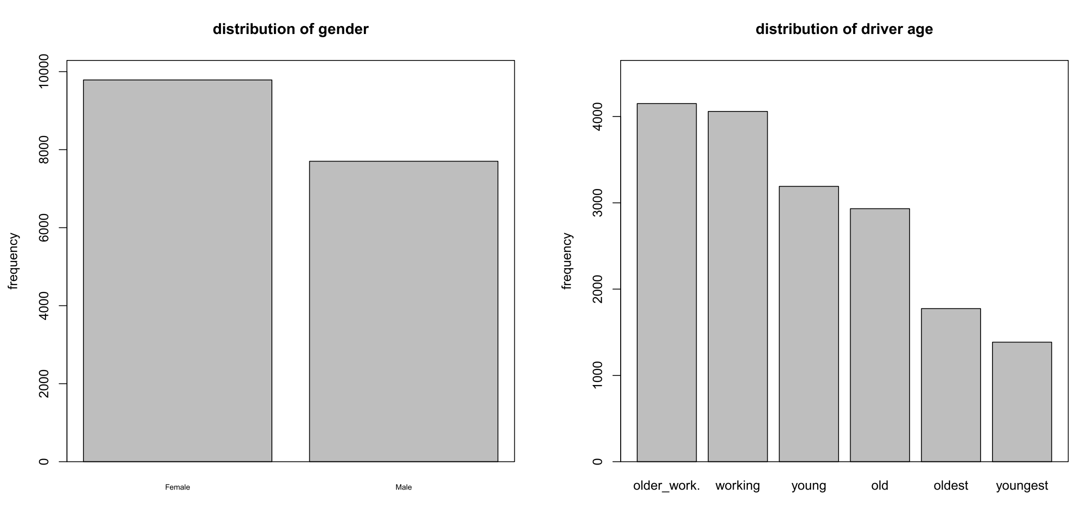

Fig. 4. Left: The distribution of gender. Right: The distribution driver age.

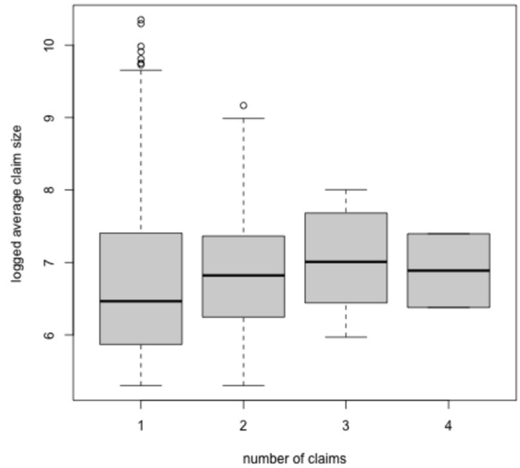

Fig. 5. The boxplot of the logged average claim size.

Table 1
The estimated coefficients in the marginal claims frequency model (4.1).

	estimated	std. error	z value	p-value
Intercept	-2.0348	0.0528	-38.54	$<$ 2e-16
LnVehValue	0.2302	0.0405	5.69	1.3e-08
DrivAge old	-0.1748	0.0808	-2.16	0.031
DrivAge oldest	-0.1310	0.0959	-1.37	0.172
DrivAge working	-0.0359	0.0707	-0.51	0.612
DrivAge young	0.0603	0.0737	0.82	0.413
DrivAge youngest	0.1642	0.0930	1.77	0.078

are 14.13% and 14.04%, respectively. We denote the learning policies with claims by $\mathcal{I}_{L+} = \mathcal{I}_L \cap \mathcal{I}_+$ and the test policies with claims by $\mathcal{I}_{T+} = \mathcal{I}_T \cap \mathcal{I}_+$.

We will establish two GLMs for the claim frequency and severity, respectively. While the learning set $\mathcal{I}_L$ is used for the claim frequency modeling, the set $\mathcal{I}_{L+} \subset \mathcal{I}_L$ is used for both the average claim size modeling and the dependence modeling. The test set $\mathcal{I}_T$ is used to evaluate the out-of-sample prediction performance of the claim frequency model and the aggregated claims amount model, while the set $\mathcal{I}_{T+} \subset \mathcal{I}_T$ is used to evaluate the out-of-sample prediction performance of the claim severity model.

4.2. Jointly modeling of frequency-severity

We first establish the marginal GLMs for the claims frequency and severity. A Poisson GLM, a negative binomial GLM and a mixture of Poisson regressions are used and compared for the claims frequency modeling, while a gamma GLM and a log-normal GLM are used and compared for the claims severity modeling. We then investigate the frequency-severity dependence using the proposed method.

4.2.1. Marginal models

The Poisson GLM is specified in (2.2) with the claims frequency modeled by:

$$\log \lambda(\boldsymbol{x}_i) = \log e_i + \alpha_0 + \alpha_1 \log x_{i,1} + \sum_{k=2}^{20} \alpha_k x_{i,k}, \tag{4.1}$$

where e_i is the exposure, $x_{i,1}$ is the vehicle value $\texttt{VehValue}_i$ and $(x_{i,k})_{k=2:20}$ is the dummy coding of $\texttt{VehAge}_i$, $\texttt{VehBody}_i$, $\texttt{Gender}_i$, $\texttt{DrivAge}_i$. Note that we have investigated the possible non-linear effect of vehicle value in a generalized additive model, and it turns out that the logged claims frequency is linearly related to the logged vehicle value. We fit the model to the learning data set $\mathcal{I}_L$ and apply the backward elimination algorithm by the AIC to select the relevant covariates. The final model contains the vehicle value and the driver age. Using the Pearson's residuals, the dispersion is estimated as 1.0629, indicating no obvious over-dispersion.

The estimated coefficients are listed in Table 1. The claim frequency is increasing with the vehicle value. Youngest drivers have the highest claim frequency among all the age groups, while old drivers have the lowest claim frequency. Vehicle age, vehicle body and gender are not important in predicting the claim frequency (at least according to the AIC). The out-of-sample Poisson deviance loss on $\mathcal{I}_T$ is 0.5305, compared with 0.5347 for a null homogeneous model without any covariates.

We refine the model by searching for important interaction terms. We first incorporate all the bivariate interaction terms having more than 1,000 exposure years into the claim frequency model, then we perform the backward elimination algorithm by the AIC. The out-of-sample Poisson deviance loss on $\mathcal{I}_T$ is 0.5315, even worse than the model with the main terms.

For comparison, two competing models (a negative binomial regression and a mixture of two Poisson regressions) are fitted to the number of claims. The mean regression function in the negative binomial model is the same as in the Poisson model (4.1), while the dispersion parameter is assumed to be constant. The mean regression functions in the two component regressions of the mixture Poisson model are the same as in the Poisson model (4.1). We use one prediction as the baseline and the other as alternative predictions. We examine whether the insurer could profit by switching from the base to the alternatives. A large relative Gini index (Frees et al., 2011b)

Table 2
The matrix of the Gini indices of marginal claims frequency models.

base model	relative Gini index		
	Poisson	negative binomial	mixture of Poisson
Poisson	0	-0.72	-1.93
negative binomial	0.79	0	-1.58
mixture of Poisson	2.30	1.94	0

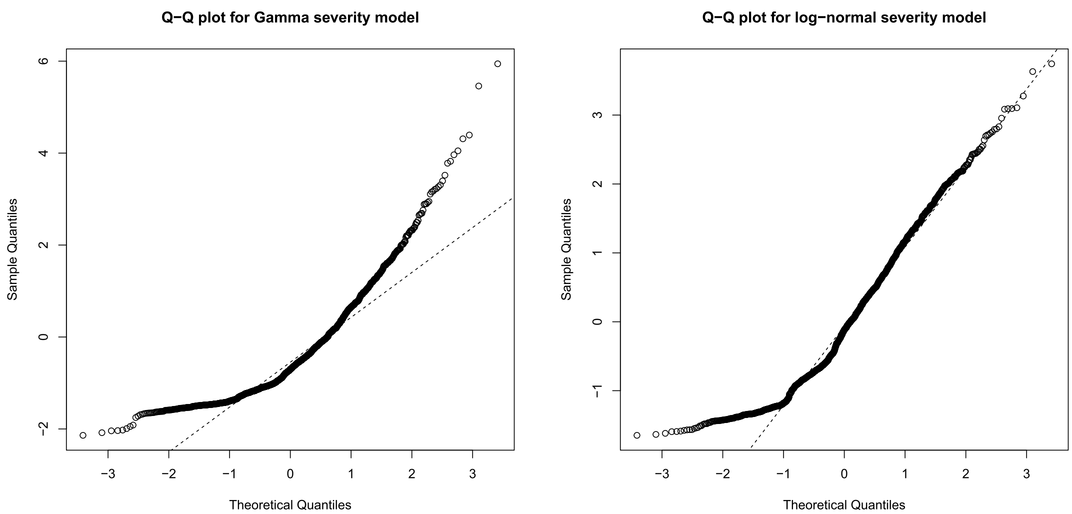

Fig. 6. Left: The Q-Q plot of deviance residuals in the gamma severity model. Right: The Q-Q plot of deviance residuals in the log-normal severity model.

Table 3
The estimated coefficients in the average claim size model (4.2). The Fano factor σ^2 is estimated as 1.1109. Note that the covariate of claim counts (in the last row) is not included in the marginal severity model in the proposed copula method.

	estimated	std. error	z value	p-value
Intercept	6.3249	0.1093	57.84	<2e-16
LnVehValue	0.1759	0.0592	2.97	0.003
VehAge oldest	0.1616	0.0792	2.04	0.042
VehAge young	-0.1231	0.0719	-1.71	0.087
VehAge youngest	-0.1388	0.0861	-1.61	0.107
Gender Male	0.1076	0.0549	1.96	0.050
claim counts	0.2021	0.0844	2.39	0.017

rejects the base and suggests the alternative. The matrix of the (out-of-sample) relative Gini indices is shown in Table 2. It turns out that the Poisson model performs at least as well as the two alternatives.

Next, we establish the GLMs for the average claim size. Preliminary modeling using a gamma distribution and a log-normal distribution indicates that the gamma distribution cannot capture the right heavy tail of the average claim size; see the Q-Q plots of the deviance residuals in Fig. 6. So we implement a log-normal GLM for the average claim size. Remark that the log-normal regression model is still insufficient for the claims severity data. One may implement a mixture of distributions (e.g. a mixture of gamma and Pareto distributions) to the claims amount data.

The log-normal claims severity model is specified as follows:

$$\log Y_i \overset{\text{ind.}}{\sim} \mathcal{N}(\mu(\boldsymbol{x}_i), \sigma^2), \quad \text{for } i \in \mathcal{I}_+, \tag{4.2}$$

where $\mu(\boldsymbol{x}_i) = \beta_0 + \beta_1 \log x_{i,1} + \sum_{k=2}^{20} \beta_k x_{i,k} + \beta_{21} e_i$. Note that we have investigated the possible non-linear effect of the exposure and the vehicle value in a generalized additive model, and it turns out that the logged average claim size is linearly related to the exposure and the logged vehicle value. We fit the model to the learning set $\mathcal{I}_{L+}$ and apply a backward elimination algorithm by the AIC to select the relevant covariates. The final model contains vehicle value, vehicle age and gender. The estimated coefficients are listed in Table 3. The average claim size is positively related to vehicle value and vehicle age. Male drivers tend to have a higher severity than female drivers. Vehicle body and driver age are not important for predicting the average claim size (at least according to the AIC).

In the conditional modeling of frequency-severity (Garrido et al., 2016), the number of claims is incorporated into the average claim size model as a covariate; see equation (2.6). We implement this approach and the estimated coefficient is added to the last line of Table 3. Thus, the average claim size is positively related to the number of claims. The out-of-sample mean squared error on $\mathcal{I}_{T+}$ is 1.1412 for the severity model without claim counts and 1.1390 for the one with claim counts. Thus, by incorporating claim counts as a covariate we improve the out-of-sample predictive performance of the average claim size model. As in the claim frequency modeling, we

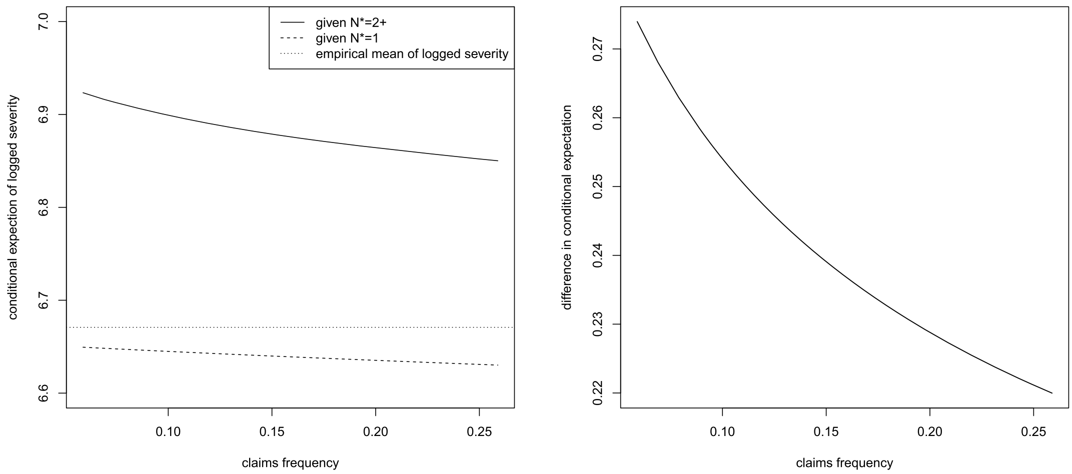

Fig. 7. Left: The conditional expectation of logged severity $\mathbb{E}(\log Y_i|N_i^*=m,\boldsymbol{x}_i)$ against the estimated claims frequency $\hat{\lambda}(\boldsymbol{x}_i)$. Right: The difference between the conditional expectations of logged severity given $N_i^*=1$ or $N_i^*=2+$ against the estimated claims frequency.

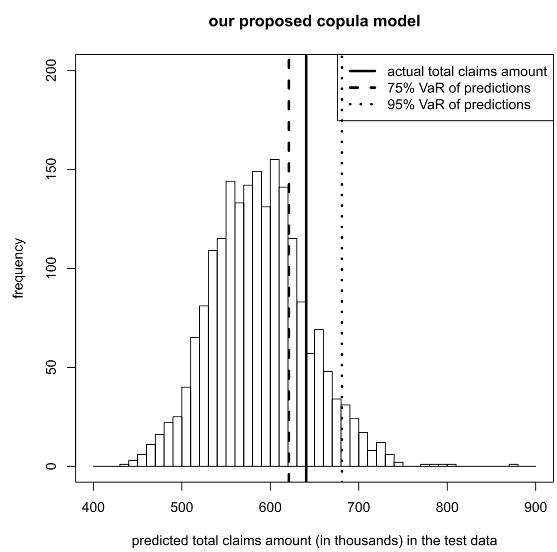

Fig. 8. The predictive distribution of the total claims amounts on the test data set $\mathcal{I}_T$.

refine the average claim size model by incorporating bivariate interaction terms. However, there is no improvement on the model fitting or out-of-sample prediction performance, and the above statements still hold.

4.2.2. The copula model

By fixing the estimated coefficients $\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}},\hat{\sigma}$, we maximize the function (3.15) w.r.t. ρ to find the IFM estimate $\hat{\rho}=-0.1106$ for a Gaussian copula. The estimated 95% confidence interval of ρ is $(-0.1957,-0.0254)$. Therefore, the waiting time and the average claim size have a significantly negative relationship (i.e. positive frequency-severity dependence), which aligns with the finding from the conditional modeling.

We investigate the conditional expectation $\mathbb{E}(\log Y_i|N_i^*=m,\boldsymbol{x}_i)$ given the claim counts $m\in\{1,2+\}$. According to (3.11), this conditional expectation depends on the estimated claim frequency $\hat{\lambda}(\boldsymbol{x}_i)$ from the marginal frequency model. Suppose that for a policyholder i, the (unconditional) expected logged claim severity $\hat{\mu}(\boldsymbol{x}_i)$ is equal to the empirical mean of logged severity 6.67. We draw the conditional expectation $\mathbb{E}(\log Y_i|N_i^*=m,\boldsymbol{x}_i)$ against the estimated claim frequency $\hat{\lambda}(\boldsymbol{x}_i)$ in Fig. 7 (left).

Given $N_i^*=2+$, the conditional expected logged severity is larger than the unconditional expected logged severity $\hat{\mu}(\boldsymbol{x}_i)$, and the difference is decreasing with the claim frequency $\hat{\lambda}(\boldsymbol{x}_i)$. Given $N_i^*=1$, the conditional expected logged severity is slightly smaller than the unconditional expected logged severity $\hat{\mu}(\boldsymbol{x}_i)$, and the difference is slightly increasing with the estimated claim frequency $\hat{\lambda}(\boldsymbol{x}_i)$. In Fig. 7 (right), the difference between the conditional expectation of logged severity given $N_i^*=2+$ or $N_i^*=1$ is decreasing from 0.27 to 0.22 with the estimated claims frequency.

We apply the Monte Carlo algorithm in Section 3.3 to simulate a sample of the aggregated claims amount for each policyholder on the test data set $\mathcal{I}_T$ for 2,000 times. Hence, we get an empirical predictive distribution of the aggregated claims amount for each policyholder $i\in\mathcal{I}_T$ on the test data set. The predictive distribution of the total aggregated claims amount on the test data set is shown in Fig. 8. The actual total claims amount is between the predicted 75% and 95% VaRs.

Finally, we compare with a t copula, which might be more desirable than the Gaussian copula if the data indicated a non-zero tail dependence. The degree of freedom of t copula is either predetermined or estimated by maximizing the joint likelihood. We try both ways

Table 4
Dependence measures for the estimated t and Gaussian copulas. Remark that a negative ρ indicates a negative waiting time-severity dependence. Note that this dependence parameter has a different interpretation from that in Model 2 discussed later.

copula family	df	$\hat{\rho}$	Kendall's tau	coefficient of tail dependence
t	10	-0.0811	-0.0517	0.0042
t	15	-0.0914	-0.0583	0.0005
t	20	-0.0964	-0.0615	0.0001
Gaussian	-	-0.1106	-0.0705	0

Table 5
The matrix of the Gini indices of frequency-severity models.

base model	relative Gini index									
	0	1	2	3	0*	1*	2*	0**	1**	2**
0	0	8.67	**11.60**	**12.77**	2.56	3.89	5.57	**11.81**	**12.92**	**13.38**
1	0.84	0	9.58	**11.17**	0.84	3.70	6.28	**11.41**	**11.26**	**11.41**
2	-2.77	-0.09	0	6.57	-2.77	-1.18	1.21	4.42	6.28	6.00
3	-3.73	-2.10	3.38	0	-3.73	-2.65	-1.34	2.70	2.89	3.07
0*	2.56	8.67	**11.60**	**12.77**	0	3.89	5.57	**11.81**	**12.92**	**13.38**
1*	5.45	6.23	**10.10**	**11.54**	5.45	0	5.30	8.90	**12.24**	**10.54**
2*	3.94	3.65	8.50	**11.06**	3.94	4.22	0	**10.93**	**11.45**	9.91
0**	-3.04	-2.05	5.15	6.53	-3.04	0.48	-1.89	0	5.60	4.36
1**	-3.46	-1.03	3.49	7.10	-3.46	-2.75	-1.23	3.85	0.00	4.77
2**	-4.39	-1.81	3.54	6.76	-4.39	-1.91	-0.58	4.60	5.31	0

here. In the first way, we predetermine the degrees of freedom at $\{10, 15, 20\}$, i.e., we fit three t copulas with three different degrees of freedom. The corresponding dependence measures are shown in Table 4, indicating that all the copula families lead to a similar fit. The tail dependence is rather weak for this data set. In the second way, we calculate the MLEs of both the association parameter ρ and the degree of freedom ν. The range of the degree of freedom is restricted to $[1, 1000]$ to reduce computing time. The MLEs are calculated as $\hat{\rho} = -0.1106$ and $\hat{\nu} = 1000$. With such a large degree of freedom, we get a similar fit to the Gaussian copula. We conclude that there is no obvious tail dependence for this data, and the Gaussian copula is a reasonable choice.

4.3. Comparison with the competing models

We consider nine competing models: the independent frequency-severity models, denoted by Models $0, 0^*, 0^{**}$; the conditional dependence models by Garrido et al. (2016), denoted by Models $1, 1^*, 1^{**}$; the mixed copula models by Czado et al. (2012), denoted by Models $2, 2^*, 2^{**}$. Our proposed model is called as Model 3. Model 0 consists of the same two marginal models as (4.1) and (4.2). Model 1 consists of the same marginal frequency model as (4.1) and the conditional severity model as shown in Table 3. Model 2 consists of the same two marginal models as (4.1) and (4.2) in a mixed copula. Models $0^*, 1^*, 2^*$, are similar to Models $0, 1, 2$ but with a negative binomial regression model for claims frequency, while Models $0^{**}, 1^{**}, 2^{**}$ with a mixture of Poisson regressions for claims frequency. We estimate the Gaussian copula association parameter in Model 2 as $\hat{\rho} = 0.1058$. Hence, all Models $1, 2, 3$ indicate that frequency and severity are positively related. We compare those models in terms of two out-of-sample metrics on the test data set $\mathcal{I}_T$: relative Gini indices (Frees et al., 2011b) and percentiles of actual claims amounts in the predictive distribution for grouped policyholders.

The matrix of the relative Gini indices is shown in Table 5, with the standard errors shown in Table 6. We bold the Gini indices larger than 10 (around double the standard errors). All the bolded numbers are in the columns for Models $2, 3, 0^{**}, 1^{**}$ and 2^{**}. Thus, we have two facts: the copula models 2 and 3 are better than the independent models 0 and 0^* and the conditional models 1 and 1^*; the mixture of Poisson regressions in Models $0^{**}, 1^{**}$ and 2^{**} are more suitable than the Poisson regression or the negative binomial regression in Models $0, 1, 0^*, 1^*$ and 2^*. We investigate the Gini indices for Models 2 and 3 in Table 5. It shows that we could identify more profit opportunities by switching from Model 2 to 3 than from Model 3 to 2, but such a difference between Models 2 and 3 is not statistically significant. Then we investigate the Gini indices for Models $3, 0^{**}, 1^{**}$ and 2^{**}. It shows that Model 3 is the best (although not statistically significant). We conclude that for this particular data, our Model 3 provides quite competitive results compared with the other competing methods.

Since most policyholders do not make any claims, it is not suitable to compare the actual claims amount with the predicted claims amount at the individual policyholder level. We cluster the policyholders in $\mathcal{I}_T$ into 100 groups and compare the actual claims amount with the predictions at the group level. Ideally, the percentiles of the grouped actual claims amount in the predictive distribution should be uniformly distributed in $(0, 1)$. We perform the Kolmogorov-Smirnov tests for those 100 percentiles against the uniform distribution $(0, 1)$. The p-values for Models $0-3$ are $0.31, 0.34, 0.42, 0.40$, respectively, under the null hypothesis that the percentiles follow a uniform distribution $(0, 1)$. We conclude that Models 2 and 3 are better than Models 0 and 1 in terms of this out-of-sample measure.

5. Conclusions

In this paper, we propose to investigate the frequency-severity dependence by using the waiting time for claim. The proposed model utilizes the relationship between counts and the waiting time of a Poisson process. The copula links two continuous variables of the waiting time and average claim size, which induces the frequency-severity dependence. Given the claim counts, the conditional expectation of the logged severity (3.11) is derived, which is related to the claims frequency, e.g. see Fig. 7.

Table 6
The standard errors of the Gini indices of frequency-severity models.

base model	relative Gini index									
	0	1	2	3	0*	1*	2*	0**	1**	2**
0	0	4.51	5.11	5.21	5.26	4.78	4.96	5.31	4.73	5.17
1	4.49	0	4.78	5.37	4.49	4.86	4.59	4.86	4.34	4.44
2	5.17	4.77	0	4.42	5.17	5.65	5.06	4.63	4.34	4.63
3	5.39	5.49	4.49	0	5.39	5.98	5.3	5.07	4.42	4.49
0*	5.26	4.51	5.11	5.21	0	4.78	4.96	5.31	4.73	5.17
1*	4.81	4.89	5.6	5.82	4.81	0	4.84	5.26	5.41	5.89
2*	5.05	4.57	4.98	5.16	5.05	4.86	0	5.50	4.60	4.98
0**	5.32	4.84	4.62	5.02	5.32	5.38	5.56	0.00	4.91	5.01
1**	4.85	4.38	4.35	4.47	4.85	5.47	4.69	4.92	0.00	3.99
2**	5.27	4.44	4.61	4.53	5.27	5.95	5.09	4.99	4.04	0.00

In a simulation study, we learn the finite sample properties of parameter estimation methods, including global MLE, two-stage estimation and the IFM method. We observe that the estimated coefficients from those estimation methods are quite close under a moderate dependence. We prefer the IFM method since it has the least computing burden.

In an empirical study, we compare the Gaussian copula with the t copula, and both copula models lead to a similar fit for this particular data set. We point out a weakness of the t copula that it does not nest the independent case. We contrast the proposed model with nine competing models: the independent models, the conditional models by Garrido et al. (2016) and the mixed copula models by Czado et al. (2012). The relative Gini indices and the grouped percentiles on the test data indicate that the proposed model has a competitive generalization ability for this particular data compared with other methods.

Declaration of competing interest

There is no competing interest.

Data availability

The data that has been used is confidential.

Acknowledgement

Guangyuan Gao gratefully acknowledges financial support from the National Natural Science Foundation of China (71901207). We particularly thank the Editor and two anonymous referees for providing valuable and insightful comments on earlier drafts.

Appendix A. Proof of (3.4)

Proof.

$$F_{T_i}(t) = \Pr(T_i \le t) = \Pr(T_{i,1} + T_{i,2} \le t | T_{i,1} \le 1) = \frac{\Pr(T_{i,1} + T_{i,2} \le t, T_{i,1} \le 1)}{\Pr(T_{i,1} \le 1)}.$$

If $t \le 1$, we get

$$\begin{aligned} F_{T_i}(t) &= \frac{\Pr(T_{i,1} + T_{i,2} \le t, T_{i,1} \le 1)}{\Pr(T_{i,1} \le 1)} \\ &= \frac{\Pr(T_{i,1} + T_{i,2} \le t)}{\Pr(T_{i,1} \le 1)} \\ &= \frac{1 - e^{-\lambda(\boldsymbol{x}_i)t} - \lambda(\boldsymbol{x}_i)te^{-\lambda(\boldsymbol{x}_i)t}}{1 - e^{-\lambda(\boldsymbol{x}_i)}}. \end{aligned}$$

If $t > 1$, we get

$$\begin{aligned} F_{T_i}(t) &= \frac{\Pr(T_{i,1} + T_{i,2} \le t, T_{i,1} \le 1)}{\Pr(T_{i,1} \le 1)} \\ &= \frac{\Pr(T_{i,1} \le 1) - \Pr(T_{i,1} + T_{i,2} > t, T_{i,1} \le 1)}{\Pr(T_{i,1} \le 1)} \\ &= \frac{\Pr(T_{i,1} \le 1) - \int_0^1 \Pr(T_{i,2} > t - s)\lambda(\boldsymbol{x}_i)e^{-\lambda(\boldsymbol{x}_i)s}ds}{\Pr(T_{i,1} \le 1)} \\ &= \frac{\Pr(T_{i,1} \le 1) - \int_0^1 e^{-\lambda(\boldsymbol{x}_i)(t-s)}\lambda(\boldsymbol{x}_i)e^{-\lambda(\boldsymbol{x}_i)s}ds}{\Pr(T_{i,1} \le 1)} \\ &= \frac{1 - e^{-\lambda(\boldsymbol{x}_i)} - \lambda(\boldsymbol{x}_i)e^{-\lambda(\boldsymbol{x}_i)t}}{1 - e^{-\lambda(\boldsymbol{x}_i)}}. \quad \square \end{aligned}$$

Appendix B. Waiting time associated with mixture Poisson distribution and negative binomial distribution

B.1. Mixture Poisson distribution

We consider the mixture of two exponential distributed waiting times, which leads to a mixture of Poisson distributions. Choose a Bernoulli random variable $H_i \in \{0, 1\}$ with $\Pr(H_i = 0) = \pi_0 > 0$. Given $\{H_i = h\}$, the waiting time between the $k - 1$ and k-th claim of policy i is independently and identically distributed as an exponential distribution:

$$T_{i,k}|H_i = h \overset{\text{i.i.d.}}{\sim} \text{Exp}(1/\lambda_h(\boldsymbol{x}_i)), \quad \text{for } k = 1, 2, \ldots \text{ and } h = 0, 1.$$

The correspondence between the censored number of claims N_i and the waiting times $T_{i,k}$ is the same as (3.2). By Partition Equation, we gained that

$$\begin{aligned}
\Pr(N_i = 0) &= \Pr(N_i = 0|H_i = 0)\Pr(H_i = 0) + \Pr(N_i = 0|H_i = 1)\Pr(H_i = 1)\\
&= \pi_0 \Pr(T_{i,1} > 1|H_i = 0) + (1 - \pi_0)\Pr(T_{i,1} > 1|H_i = 1)\\
&= \pi_0 e^{-\lambda_0(\boldsymbol{x}_i)} + (1 - \pi_0)e^{-\lambda_1(\boldsymbol{x}_i)},\\
\Pr(N_i = 1) &= \Pr(N_i = 1|H_i = 0)\Pr(H_i = 0) + \Pr(N_i = 1|H_i = 1)\Pr(H_i = 1)\\
&= \pi_0 \lambda_0(\boldsymbol{x}_i)e^{-\lambda_0(\boldsymbol{x}_i)} + (1 - \pi_0)\lambda_1(\boldsymbol{x}_i)e^{-\lambda_1(\boldsymbol{x}_i)},\\
\Pr(N_i = 2+) &= \Pr(N_i = 2 + |H_i = 0)\Pr(H_i = 0) + \Pr(N_i = 2 + |H_i = 1)\Pr(H_i = 1)\\
&= \pi_0\left(1 - e^{-\lambda_0(\boldsymbol{x}_i)} - \lambda_0(\boldsymbol{x}_i)e^{-\lambda_0(\boldsymbol{x}_i)}\right) + (1 - \pi_0)\left(1 - e^{-\lambda_1(\boldsymbol{x}_i)} - \lambda_1(\boldsymbol{x}_i)e^{-\lambda_1(\boldsymbol{x}_i)}\right).
\end{aligned}$$

Recall that

$$F_{T_i}(t) = \Pr(T_i \le t) = \Pr(T_{i,1} + T_{i,2} \le t|T_{i,1} \le 1) = \frac{\Pr(T_{i,1} + T_{i,2} \le t, T_{i,1} \le 1)}{\Pr(T_{i,1} \le 1)}.$$

The conditional distribution of T_i given $\{H_i = h\}$ is the same as (3.4). For $t > 0$, by Partition Equation, the unconditional CDF of T_i is given by

$$\begin{aligned}
F_{T_i}(t) &= \Pr(T_i \le t) = \Pr(T_i \le t|H_i = 0)\Pr(H_i = 0) + \Pr(T_i \le t|H_i = 1)\Pr(H_i = 1)\\
&= \begin{cases} \pi_0 \frac{1-e^{-\lambda_0(\boldsymbol{x}_i)t} - \lambda_0(\boldsymbol{x}_i)te^{-\lambda_0(\boldsymbol{x}_i)t}}{1-e^{-\lambda_0(\boldsymbol{x}_i)}} + (1-\pi_0)\frac{1-e^{-\lambda_1(\boldsymbol{x}_i)t} - \lambda_1(\boldsymbol{x}_i)te^{-\lambda_1(\boldsymbol{x}_i)t}}{1-e^{-\lambda_1(\boldsymbol{x}_i)}}, & \text{for } t \le 1;\\ \pi_0 \frac{1-e^{-\lambda_0(\boldsymbol{x}_i)} - \lambda_0(\boldsymbol{x}_i)e^{-\lambda_0(\boldsymbol{x}_i)t}}{1-e^{-\lambda_0(\boldsymbol{x}_i)}} + (1-\pi_0)\frac{1-e^{-\lambda_1(\boldsymbol{x}_i)} - \lambda_1(\boldsymbol{x}_i)e^{-\lambda_1(\boldsymbol{x}_i)t}}{1-e^{-\lambda_1(\boldsymbol{x}_i)}}, & \text{for } t > 1. \end{cases}
\end{aligned} \tag{B.1}$$

Note that when $\lambda_0(\boldsymbol{x}_i) \to 0$, the distribution of N_i converges to (right-censored) zero-inflated Poisson distribution. The unconditional CDF of T_i in a zero-inflated Poisson distribution is given by

$$\begin{aligned}
F_{T_i}(t) &= \Pr(T_i \le t) = \Pr(T_i \le t|H_i = 1)\Pr(H_i = 1)\\
&= \begin{cases} (1-\pi_0)\frac{1-e^{-\lambda_1(\boldsymbol{x}_i)t} - \lambda_1(\boldsymbol{x}_i)te^{-\lambda_1(\boldsymbol{x}_i)t}}{1-e^{-\lambda_1(\boldsymbol{x}_i)}}, & \text{for } t \le 1;\\ (1-\pi_0)\frac{1-e^{-\lambda_1(\boldsymbol{x}_i)} - \lambda_1(\boldsymbol{x}_i)e^{-\lambda_1(\boldsymbol{x}_i)t}}{1-e^{-\lambda_1(\boldsymbol{x}_i)}}, & \text{for } t > 1. \end{cases}
\end{aligned} \tag{B.2}$$

B.2. Negative binomial distribution

The negative binomial distribution is known as a compound distribution of Poisson distribution with the mixing gamma distribution of the Poisson mean. Assume that the Poisson mean λ_i follows a gamma distribution with shape s and rate $r_i = r(\boldsymbol{x}_i)$, and given λ_i the number of claims N_i follows a Poisson distribution with mean λ_i. The unconditional distribution of N_i is a negative binomial distribution given by

$$\begin{aligned}
f(N_i = m|s, r_i) &= \int_0^\infty \frac{\lambda_i^m}{m!}e^{-\lambda_i} \cdot \lambda_i^{s-1}\frac{e^{-\lambda_i r_i}}{r_i^{-s}\Gamma(s)}d\lambda_i\\
&= \frac{r_i^s}{m!\Gamma(s)}\int_0^\infty \lambda_i^{s+m-1}e^{-\lambda_i(1+r_i)}d\lambda_i\\
&= \frac{r_i^s}{m!\Gamma(s)}(1+r_i)^{-s-m}\Gamma(s+m)\int_0^\infty \frac{\lambda_i^{s+m-1}e^{-\lambda_i(1+r_i)}}{(1+r_i)^{-s-m}\Gamma(s+m)}d\lambda_i\\
&= \frac{r_i^s}{m!\Gamma(s)}(1+r_i)^{-s-m}\Gamma(s+m)\\
&= \frac{\Gamma(s+m)}{m!\Gamma(s)}\left(\frac{1}{1+r_i}\right)^m\left(\frac{r_i}{1+r_i}\right)^s
\end{aligned}$$

$$=\frac{\Gamma(s+m)}{m!\Gamma(s)}p_i^m(1-p_i)^s,$$

where $p_i=1/(1+r_i)$. Note that we do not require s as an integer.

Given λ_i, the waiting time $T_{i,k}$ follows an exponential distribution with mean $1/\lambda_i$. The unconditional distribution of waiting time is derived as

$$\begin{aligned}
f(T_{ik}=t|s,r_i) &= \int_0^\infty \lambda_i e^{-\lambda_i t}\cdot\lambda_i^{s-1}\frac{e^{-\lambda_i r_i}}{r_i^{-s}\Gamma(s)}d\lambda_i \\
&= \frac{r_i^s}{\Gamma(s)}\int_0^\infty \lambda_i^s e^{-\lambda_i(t+r_i)}d\lambda_i \\
&= \frac{r_i^s}{\Gamma(s)}(t+r_i)^{-s-1}\Gamma(s+1)\int_0^\infty \frac{\lambda_i^s e^{-\lambda_i(t+r_i)}}{(t+r_i)^{-s-1}\Gamma(s+1)}d\lambda_i \\
&= \left(\frac{t+r_i}{r_i}\right)^{-(s+1)}\frac{\Gamma(s+1)}{r_i\Gamma(s)} \\
&= \frac{s}{r_i}\left(1+\frac{t}{r_i}\right)^{-(s+1)}.
\end{aligned} \tag{B.3}$$

Thus, the waiting time T_{ik} follows a Pareto Type II distribution with shape s and scale r_i. Recall that

$$F_{T_i}(t)=\Pr(T_i\le t)=\Pr(T_{i,1}+T_{i,2}\le t|T_{i,1}\le 1)=\frac{\Pr(T_{i,1}+T_{i,2}\le t,T_{i,1}\le 1)}{\Pr(T_{i,1}\le 1)}.$$

Unfortunately we do not have an explicit expression for the distribution of the sum of i.i.d Pareto variables $T_{i,1}+T_{i,2}$ (Ramsay, 2006). Therefore, we do not have a closed form for the CDF $F_{T_i}(t)$.

Appendix C. Proof of (3.8) for Gaussian copula and *t* copula (revisited)

Proof.

$$f_{N_i^*,Y_i}(m,y|\boldsymbol{x}_i;\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho)=\begin{cases} C_2\left(F_{T_i}(1|\boldsymbol{x}_i;\boldsymbol{\alpha}),F_{Y_i}(y|\boldsymbol{x}_i;\boldsymbol{\beta},\phi);\rho\right)f_{Y_i}(y|\boldsymbol{x}_i;\boldsymbol{\beta},\phi), & \text{for } m=2+;\\ \left[1-C_2\left(F_{T_i}(1|\boldsymbol{x}_i;\boldsymbol{\alpha}),F_{Y_i}(y|\boldsymbol{x}_i;\boldsymbol{\beta},\phi);\rho\right)\right]f_{Y_i}(y|\boldsymbol{x}_i;\boldsymbol{\beta},\phi), & \text{for } m=1,\end{cases}$$

For the partial derivative of bivariate Gaussian copula, we refer to Joe Joe (2014) Section 4.3:

$$\frac{\partial}{\partial u_2}C^G(u_1,u_2;\rho)=\frac{\partial}{\partial u_2}\Phi_2(\Phi^{-1}(u_1),\Phi^{-1}(u_2);\rho)=\Phi\left(\frac{\Phi^{-1}(u_1)-\rho\Phi^{-1}(u_2)}{\sqrt{1-\rho^2}}\right).$$

For Gaussian copula, since $z_{i,1}=\Phi^{-1}(u_{i,1})$ and $z_{i,2}=\Phi^{-1}(u_{i,2})$, we can gain that

$$f_{N_i^*,Y_i}(m,y|\boldsymbol{x}_i;\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho)=\begin{cases}\Phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)f_{Y_i}(y|\boldsymbol{x}_i;\boldsymbol{\beta},\phi), & \text{for } m=2+;\\ \Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)f_{Y_i}(y|\boldsymbol{x}_i;\boldsymbol{\beta},\phi), & \text{for } m=1.\end{cases}$$

For the partial derivative of bivariate t copula, we refer to Joe Joe (2014) Section 4.13:

$$\frac{\partial}{\partial u_2}C^t(u_1,u_2;\rho,\nu)=\frac{\partial}{\partial u_2}T_{2,\nu}(T_\nu^{-1}(u_1),T_\nu^{-1}(u_2);\rho)=T_{\nu+1}\left(\frac{T_\nu^{-1}(u_1)-\rho T_\nu^{-1}(u_2)}{\sqrt{(1-\rho^2)\left(\nu+\left[T_\nu^{-1}(u_2)\right]^2\right)/(\nu+1)}}\right).$$

For t copula, since $w_{i,1}=T_\nu^{-1}(u_{i,1})$ and $w_{i,2}=T_\nu^{-1}(u_{i,2})$, we can gain that

$$f_{N_i^*,Y_i}(m,y|\boldsymbol{x}_i;\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho)=\begin{cases}T_{\nu+1}\left(\frac{w_{i,1}-\rho w_{i,2}}{\sqrt{(1-\rho^2)\left(\nu+w_{i,2}^2\right)/(\nu+1)}}\right)f_{Y_i}(y|\boldsymbol{x}_i;\boldsymbol{\beta},\phi), & \text{for } m=2+;\\ T_{\nu+1}\left(-\frac{w_{i,1}-\rho w_{i,2}}{\sqrt{(1-\rho^2)\left(\nu+w_{i,2}^2\right)/(\nu+1)}}\right)f_{Y_i}(y|\boldsymbol{x}_i;\boldsymbol{\beta},\phi), & \text{for } m=1.\quad\square\end{cases}$$

Appendix D. Proof of (3.11)

Proof. Let $Z_{i,2} = \frac{\log Y_i - \mu(\boldsymbol{x}_i)}{\sigma}$. By (3.10), we can compute the conditional density of $Z_{i,2}$.

When $m = 2+$, we get

$$\begin{aligned}\Pr(Z_{i,2} \le z_{i,2} | N_i^* = m, \boldsymbol{x}_i) &= \Pr(Y_i \le e^{\mu(\boldsymbol{x}_i)+z_{i,2}\sigma} | N_i^* = m, \boldsymbol{x}_i) \\ &= \int_0^{e^{\mu(\boldsymbol{x}_i)+z_{i,2}\sigma}} \Phi\left(\frac{z_{i,1} - \rho\left(\frac{\log y - \mu(\boldsymbol{x}_i)}{\sigma}\right)}{\sqrt{1-\rho^2}}\right) \frac{1}{u_{i,1}} \frac{1}{\sqrt{2\pi}\, y\sigma} e^{-\frac{[\log y - \mu(\boldsymbol{x}_i)]^2}{2\sigma^2}} dy,\end{aligned}$$

where $z_{i,1} = \Phi^{-1}(u_{i,1})$. Thus, the conditional density of $Z_{i,2}$ is given by

$$f_{Z_{i,2}}(z_{i,2} | N_i^* = m, \boldsymbol{x}_i) = \frac{\partial}{\partial z_{i,2}} \Pr(Z_{i,2} \le z_{i,2} | N_i^* = m, \boldsymbol{x}_i) = \Phi\left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}}\right) \frac{1}{u_{i,1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_{i,2}^2}{2}}.$$

The conditional expectation of $Z_{i,2}$ is given by

$$\begin{aligned}\mathbb{E}(Z_{i,2} | N_i^* = m, \boldsymbol{x}_i) &= \int_{-\infty}^{\infty} \Phi\left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}}\right) \frac{1}{u_{i,1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_{i,2}^2}{2}} z_{i,2} dz_{i,2} \\ &= \frac{1}{u_{i,1}} \int_{-\infty}^{\infty} \Phi\left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}}\right) d\left(-\frac{1}{\sqrt{2\pi}} e^{-\frac{z_{i,2}^2}{2}}\right) \\ &= \frac{1}{u_{i,1}} \left\{ -\frac{1}{\sqrt{2\pi}} e^{-\frac{z_{i,2}^2}{2}} \Phi\left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\Bigg|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_{i,2}^2}{2}} d\Phi\left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}}\right) \right\} \\ &= \frac{1}{u_{i,1}} \left\{ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_{i,2}^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z_{i,1} - \rho z_{i,2})^2}{2(1-\rho^2)}} \left(\frac{-\rho}{\sqrt{1-\rho^2}}\right) dz_{i,2} \right\} \\ &= \frac{-\rho}{u_{i,1}} \left\{ \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(z_{i,1}^2 - 2\rho z_{i,2} z_{i,1} + z_{i,2}^2)}{2(1-\rho^2)}} dz_{i,2} \right\}.\end{aligned}$$

Since $\frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(z_{i,1}^2 - 2\rho z_{i,2} z_{i,1} + z_{i,2}^2)}{2(1-\rho^2)}}$ is the joint density of two standard Gaussian random variables, we get

$$\mathbb{E}(Z_{i,2} | N_i^* = m, \boldsymbol{x}_i) = \frac{-\rho}{u_{i,1}} \phi(z_{i,1}) = \frac{-\rho}{u_{i,1}} \phi(\Phi^{-1}(u_{i,1})).$$

Because $\log Y_i = \mu(\boldsymbol{x}_i) + \sigma Z_{i,2}$, the conditional expectation of $\log Y_i$ is given by

$$\mathbb{E}(\log Y_i | N_i^* = m, \boldsymbol{x}_i) = \mu(\boldsymbol{x}_i) + \sigma \mathbb{E}(Z_{i,2} | N_i^* = m, \boldsymbol{x}_i) = \mu(\boldsymbol{x}_i) - \frac{\rho\sigma}{u_{i,1}} \phi(\Phi^{-1}(u_{i,1})).$$

When $m = 1$, we get

$$\begin{aligned}\Pr(Z_{i,2} \le z_{i,2} | N_i^* = m, \boldsymbol{x}_i) &= \Pr(Y_i \le e^{\mu(\boldsymbol{x}_i)+z_{i,2}\sigma} | N_i^* = m, \boldsymbol{x}_i) \\ &= \int_0^{e^{\mu(\boldsymbol{x}_i)+z_{i,2}\sigma}} \Phi\left(-\frac{z_{i,1} - \rho\left(\frac{\log y - \mu(\boldsymbol{x}_i)}{\sigma}\right)}{\sqrt{1-\rho^2}}\right) \frac{1}{1-u_{i,1}} \frac{1}{\sqrt{2\pi}\, y\sigma} e^{-\frac{(\log y - \mu(\boldsymbol{x}_i))^2}{2\sigma^2}} dy,\end{aligned}$$

where $z_{i,1} = \Phi^{-1}(u_{i,1})$. Thus, the conditional density of $Z_{i,2}$ is given by

$$f_{Z_{i,2}}(z_{i,2} | N_i^* = m, \boldsymbol{x}_i) = \frac{\partial}{\partial z_{i,2}} \Pr(Z_{i,2} \le z_{i,2} | N_i^* = m, \boldsymbol{x}_i) = \Phi\left(-\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}}\right) \frac{1}{1-u_{i,1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_{i,2}^2}{2}}.$$

The conditional expectation of $Z_{i,2}$ is given by

$$
\begin{aligned}
\mathbb{E}(Z_{i,2}|N_i^*=m,\boldsymbol{x}_i) &= \int_{-\infty}^{\infty} \Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\frac{1}{1-u_{i,1}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z_{i,2}^2}{2}} z_{i,2}dz_{i,2} \\
&= \frac{1}{1-u_{i,1}}\int_{-\infty}^{\infty} \Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right) d\left(-\frac{1}{\sqrt{2\pi}}e^{-\frac{z_{i,2}^2}{2}}\right) \\
&= \frac{1}{1-u_{i,1}}\left\{-\frac{1}{\sqrt{2\pi}}e^{-\frac{z_{i,2}^2}{2}}\Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\Bigg|_{-\infty}^{\infty} + \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{z_{i,2}^2}{2}} d\Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right\} \\
&= \frac{1}{1-u_{i,1}}\left\{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{z_{i,2}^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{(z_{i,1}-\rho z_{i,2})^2}{2(1-\rho^2)}}\left(\frac{\rho}{\sqrt{1-\rho^2}}\right)dz_{i,2}\right\} \\
&= \frac{\rho}{1-u_{i,1}}\left\{\int_{-\infty}^{\infty}\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{(z_{i,1}^2-2\rho z_{i,2}z_{i,1}+z_{i,2}^2)}{2(1-\rho^2)}}dz_{i,2}\right\}.
\end{aligned}
$$

Similar to the case of $m=2+$, the conditional expectation of $\log Y_i$ for $m=1$ can be gained as

$$
\mathbb{E}(\log Y_i|N_i^*=m,\boldsymbol{x}_i)=\mu(\boldsymbol{x}_i)+\sigma\mathbb{E}(Z_{i,2}|N_i^*=m,\boldsymbol{x}_i)=\mu(\boldsymbol{x}_i)+\frac{\rho\sigma}{1-u_{i,1}}\phi(\Phi^{-1}(u_{i,1})). \quad \square
$$

Appendix E. Fisher's scoring functions for Gaussian copula model

Consider a data set containing n observations $\{(m_1,y_1,\boldsymbol{x}_1),\dots,(m_n,y_n,\boldsymbol{x}_n)\}$. We denote the risk factors (including an intercept) used in the frequency model and the severity model by $\boldsymbol{c}_i\in\mathbb{R}^p$ and $\boldsymbol{d}_i\in\mathbb{R}^q$, respectively. We denote the parameters of the proposed model by $\boldsymbol{\theta}=(\boldsymbol{\alpha}^\top,\boldsymbol{\beta}^\top,\phi,\rho)^\top\in\mathbb{R}^{p+q+2}$. By (3.12), we have

$$
l(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho)=l_1(\boldsymbol{\alpha})+l_2(\boldsymbol{\beta},\phi)+l_3(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho),
$$

where

$$
\begin{aligned}
l_1(\boldsymbol{\alpha}) &= \sum_{i\in\mathcal{I}_0}\log\Pr(N_i=0|\boldsymbol{c}_i;\boldsymbol{\alpha})+\sum_{i\in\mathcal{I}_+}\log\Pr(N_i>0|\boldsymbol{c}_i;\boldsymbol{\alpha}),\\
l_2(\boldsymbol{\beta},\phi) &= \sum_{i\in\mathcal{I}_+}\log f_{Y_i}(y_i|\boldsymbol{c}_i;\boldsymbol{\beta},\phi),\\
l_3(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho) &= \sum_{i\in\mathcal{I}_+}\left[\mathbb{1}_1(m_i)\log\Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)+\mathbb{1}_{2+}(m_i)\log\Phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right].
\end{aligned}
$$

First, l_1 is given as

$$
l_1(\boldsymbol{\alpha})=\sum_{i\in\mathcal{I}_0}-\exp\langle\boldsymbol{\alpha},\boldsymbol{c}_i\rangle v_i+\sum_{i\in\mathcal{I}_+}\log\left(1-e^{-\exp\langle\boldsymbol{\alpha},\boldsymbol{c}_i\rangle v_i}\right).
$$

We can calculate the partial derivatives of $l_1(\boldsymbol{\alpha})$ for $1\le j\le p$

$$
\frac{\partial}{\partial\alpha_j}l_1(\boldsymbol{\alpha})=\sum_{i\in\mathcal{I}_0}-\exp\langle\boldsymbol{\alpha},\boldsymbol{c}_i\rangle v_ic_{ij}+\sum_{i\in\mathcal{I}_+}\frac{e^{-\exp\langle\boldsymbol{\alpha},\boldsymbol{c}_i\rangle v_i}}{1-e^{-\exp\langle\boldsymbol{\alpha},\boldsymbol{c}_i\rangle v_i}}\exp\langle\boldsymbol{\alpha},\boldsymbol{c}_i\rangle v_ic_{ij},
$$

where $\boldsymbol{c}_i=(c_{i0},c_{i1},\dots,c_{ip})^\top$ with $c_{i0}=1$. We can write the partial derivative into a vector version as

$$
\frac{\partial}{\partial\boldsymbol{\alpha}}l_1(\boldsymbol{\alpha})=\sum_{i\in\mathcal{I}_0}-\exp\langle\boldsymbol{\alpha},\boldsymbol{c}_i\rangle v_i\boldsymbol{c}_i+\sum_{i\in\mathcal{I}_+}\frac{e^{-\exp\langle\boldsymbol{\alpha},\boldsymbol{c}_i\rangle v_i}}{1-e^{-\exp\langle\boldsymbol{\alpha},\boldsymbol{c}_i\rangle v_i}}\exp\langle\boldsymbol{\alpha},\boldsymbol{c}_i\rangle v_i\boldsymbol{c}_i \tag{E.1}
$$

Second, l_2 is given as

$$
l_2(\boldsymbol{\beta},\phi)=\sum_{i\in\mathcal{I}_+}-\frac{m_iy_i}{\phi\exp\langle\boldsymbol{\beta},\boldsymbol{d}_i\rangle}-\frac{m_i}{\phi}\langle\boldsymbol{\beta},\boldsymbol{d}_i\rangle+\left(\frac{m_i}{\phi}-1\right)\log y_i+\frac{m_i}{\phi}\log\left(\frac{m_i}{\phi}\right)-\log\Gamma\left(\frac{m_i}{\phi}\right).
$$

We compute the partial derivatives of $l_2(\boldsymbol{\beta},\phi)$ for $1\le j\le q$

$$
\begin{aligned}
\frac{\partial}{\partial\beta_j}l_2(\boldsymbol{\beta},\phi) &= \sum_{i\in\mathcal{I}_+}\frac{m_iy_i}{\phi\exp\langle\boldsymbol{\beta},\boldsymbol{d}_i\rangle}d_{ij}-\frac{m_i}{\phi}d_{ij}\\
&=\frac{1}{\phi}\sum_{i\in\mathcal{I}_+}m_i\left(\frac{y_i}{\exp\langle\boldsymbol{\beta},\boldsymbol{d}_i\rangle}-1\right)d_{ij},
\end{aligned}
$$

where $\boldsymbol{d}_i = (d_{i0}, d_{i1}, \ldots, d_{iq})^\top$ with $d_{i0} = 1$. We can write the partial derivative into a vector version as

$$\frac{\partial}{\partial \boldsymbol{\beta}} l_2(\boldsymbol{\beta}, \phi) = \frac{1}{\phi} \sum_{i \in \mathcal{I}_+} m_i \left(\frac{y_i}{\exp\langle \boldsymbol{\beta}, \boldsymbol{d}_i \rangle} - 1 \right) \boldsymbol{d}_i. \tag{E.2}$$

We also compute the partial derivatives of $l_2(\boldsymbol{\beta}, \phi)$ w.r.t. ϕ:

$$\begin{aligned} \frac{\partial}{\partial \phi} l_2(\boldsymbol{\beta}, \phi) &= \sum_{i \in \mathcal{I}_+} \frac{m_i y_i}{\phi^2 \exp\langle \boldsymbol{\beta}, \boldsymbol{d}_i \rangle} + \frac{m_i}{\phi^2} \langle \boldsymbol{\beta}, \boldsymbol{d}_i \rangle - \frac{m_i}{\phi^2} \log y_i - \frac{m_i}{\phi^2} \log \left(\frac{m_i}{\phi} \right) - \frac{m_i}{\phi^2} + \frac{\Gamma'(m_i/\phi)}{\Gamma(m_i/\phi)} \frac{m_i}{\phi^2} \\ &= \sum_{i \in \mathcal{I}_+} \left[\frac{y_i}{\exp\langle \boldsymbol{\beta}, \boldsymbol{d}_i \rangle} + \langle \boldsymbol{\beta}, \boldsymbol{d}_i \rangle - \log y_i - 1 \right] \frac{m_i}{\phi^2} - \frac{m_i}{\phi^2} \log \left(\frac{m_i}{\phi} \right) + \frac{\Gamma'(m_i/\phi)}{\Gamma(m_i/\phi)} \frac{m_i}{\phi^2}, \end{aligned} \tag{E.3}$$

where $\Gamma'(\cdot)$ is the derivative of gamma function:

$$\Gamma'(s) = \int_0^\infty x^{s-1} e^{-x} \log s dx.$$

Third, we compute the partial derivatives of $l_3(\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho)$.

$$\begin{aligned} \frac{\partial}{\partial \alpha_j} l_3(\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho) &= \sum_{i \in \mathcal{I}_+} \Bigg\{ \mathbb{1}_1(m_i) \left[\Phi \left(-\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \right]^{-1} \phi \left(-\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \left(-\frac{1}{\sqrt{1-\rho^2}} \right) \frac{\partial z_{i,1}}{\partial \alpha_j} \\ &\quad + \mathbb{1}_{2+}(m_i) \left[\Phi \left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \right]^{-1} \phi \left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \left(\frac{1}{\sqrt{1-\rho^2}} \right) \frac{\partial z_{i,1}}{\partial \alpha_j} \Bigg\} \\ &= \sum_{i \in \mathcal{I}_+} \Bigg\{ -\mathbb{1}_1(m_i) \left[\Phi \left(-\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \right]^{-1} \phi \left(-\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \\ &\quad + \mathbb{1}_{2+}(m_i) \left[\Phi \left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \right]^{-1} \phi \left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \Bigg\} \frac{1}{\sqrt{1-\rho^2}} \frac{\partial z_{i,1}}{\partial \alpha_j}, \end{aligned}$$

where

$$\frac{\partial z_{i,1}}{\partial \alpha_j} = \frac{d z_{i,1}}{d u_{i,1}} \frac{\partial u_{i,1}}{\partial \alpha_j} = \left(\frac{d u_{i,1}}{d z_{i,1}} \right)^{-1} \frac{\partial u_{i,1}}{\partial \alpha_j} = \frac{1}{\phi(\Phi^{-1}(u_{i,1}))} \frac{\partial u_{i,1}}{\partial \alpha_j}.$$

Now, we calculate the partial derivative of $u_{i,1}$ for $1 \le j \le p$

$$\begin{aligned} \frac{\partial u_{i,1}}{\partial \alpha_j} &= \frac{\partial}{\partial \alpha_j} \Pr(N_i^* = 2 + |\boldsymbol{c}_i; \boldsymbol{\alpha}) \\ &= \frac{\partial}{\partial \alpha_j} \frac{1 - e^{-\lambda_i v_i} - \lambda_i v_i e^{-\lambda_i v_i}}{1 - e^{-\lambda_i v_i}} \\ &= \frac{\partial}{\partial \lambda_i} \left(\frac{1 - e^{-\lambda_i v_i} - \lambda_i v_i e^{-\lambda_i v_i}}{1 - e^{-\lambda_i v_i}} \right) \frac{\partial \lambda_i}{\partial \alpha_j} \\ &= \frac{\lambda_i v_i^2 e^{-\lambda_i v_i} - v_i e^{-\lambda_i v_i} + v_i e^{-2\lambda_i v_i}}{(1 - e^{-\lambda_i v_i})^2} \lambda_i c_{ij}, \end{aligned}$$

where $\lambda_i = \lambda(\boldsymbol{c}_i) = \exp\langle \boldsymbol{\alpha}, \boldsymbol{c}_i \rangle$. Therefore,

$$\begin{aligned} \frac{\partial l_3}{\partial \boldsymbol{\alpha}} &= \sum_{i \in \mathcal{I}_+} \Bigg\{ -\mathbb{1}_1(m_i) \left[\Phi \left(-\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \right]^{-1} \phi \left(-\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \\ &\quad + \mathbb{1}_{2+}(m_i) \left[\Phi \left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \right]^{-1} \phi \left(\frac{z_{i,1} - \rho z_{i,2}}{\sqrt{1-\rho^2}} \right) \Bigg\} \frac{1}{\sqrt{1-\rho^2}} \frac{1}{\phi(\Phi^{-1}(u_{i,1}))} \\ &\quad \times \frac{\exp\langle \boldsymbol{\alpha}, \boldsymbol{c}_i \rangle v_i^2 e^{-\exp\langle \boldsymbol{\alpha}, \boldsymbol{c}_i \rangle v_i} - v_i e^{-\exp\langle \boldsymbol{\alpha}, \boldsymbol{c}_i \rangle v_i} + v_i e^{-2\exp\langle \boldsymbol{\alpha}, \boldsymbol{c}_i \rangle v_i}}{(1 - e^{-\exp\langle \boldsymbol{\alpha}, \boldsymbol{c}_i \rangle v_i})^2} \exp\langle \boldsymbol{\alpha}, \boldsymbol{c}_i \rangle \boldsymbol{c}_i. \end{aligned} \tag{E.4}$$

Similarly,

$$\begin{aligned}\frac{\partial}{\partial \beta_j} l_3(\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \rho) =& \sum_{i\in\mathcal{I}_+}\left\{ \mathbb{1}_1(m_i)\left[\Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right]^{-1}\phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\left(\frac{\rho}{\sqrt{1-\rho^2}}\right)\frac{\partial z_{i,2}}{\partial \beta_j}\right.\\ &\left. +\mathbb{1}_{2+}(m_i)\left[\Phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right]^{-1}\phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\left(-\frac{\rho}{\sqrt{1-\rho^2}}\right)\frac{\partial z_{i,2}}{\partial \beta_j}\right\}\\ =& \sum_{i\in\mathcal{I}_+}\left\{ \mathbb{1}_1(m_i)\left[\Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right]^{-1}\phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right.\\ &\left. -\mathbb{1}_{2+}(m_i)\left[\Phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right]^{-1}\phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right\}\frac{\rho}{\sqrt{1-\rho^2}}\frac{\partial z_{i,2}}{\partial \beta_j},\end{aligned}$$

where

$$\frac{\partial z_{i,2}}{\partial \beta_j} = \frac{1}{\phi(\Phi^{-1}(u_{i,2}))}\frac{\partial u_{i,2}}{\partial \beta_j}.$$

Then, we calculate the partial derivative of $u_{i,2}$ for $1 \le j \le q$

$$\frac{\partial u_{i,2}}{\partial \beta_j} = \frac{\partial}{\partial \mu_i} F_{Y_i}(y_i|\boldsymbol{d}_i; \boldsymbol{\beta}, \phi)\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial}{\partial \mu_i} F_{Y_i}(y_i; \mu_i, \phi)\exp\langle \boldsymbol{\beta}, \boldsymbol{d}_i\rangle d_{ij}.$$

where $\mu_i = \mu(\boldsymbol{d}_i) = \exp\langle \boldsymbol{\beta}, \boldsymbol{d}_i\rangle$.

Now, we compute the partial derivatives of $u_{i,2}$ with respect to μ_i.

$$\begin{aligned}\frac{\partial}{\partial \mu_i}F_{Y_i}(y_i;\mu_i,\phi) &= \frac{\partial}{\partial \mu_i}\int_0^{y_i} f_{Y_i}(y;\mu_i,\phi)dy = \int_0^{y_i}\frac{\partial}{\partial \mu_i} f_{Y_i}(y;\mu_i,\phi)dy\\ &= \int_0^{y_i}\frac{\partial}{\partial \mu_i}\left[\frac{1}{\Gamma(m_i/\phi)}\left(\frac{m_i}{\phi\mu_i}\right)^{m_i/\phi} y^{m_i/\phi-1}e^{-ym_i/\phi\mu_i}\right]dy\\ &= \int_0^{y_i}\frac{m_i}{\phi\mu_i^2} f_{Y_i}(y;\mu_i,\phi)(y-\mu_i)dy\\ &= \frac{m_i}{\phi\mu_i^2}\left[\int_0^{y_i} y f_{Y_i}(y;\mu_i,\phi)dy - \int_0^{y_i}\mu_i f_{Y_i}(y;\mu_i,\phi)dy\right]\\ &= \frac{m_i}{\phi\mu_i^2}\left[\int_0^{y_i} y f_{Y_i}(y;\mu_i,\phi)dy - \mu_i F_{Y_i}(y;\mu_i,\phi)\right]\\ &= \frac{m_i}{\phi\mu_i^2}\left[\int_0^{y_i} \frac{1}{\Gamma(m_i/\phi)}\left(\frac{m_i}{\phi\mu_i}\right)^{m_i/\phi} y^{m_i/\phi+1-1}e^{-ym_i/\phi\mu_i}dy - \mu_i F_{Y_i}(y;\mu_i,\phi)\right].\end{aligned}$$

Let $a_i = m_i/\phi$ and $b_i = m_i/\phi\mu_i$, then

$$\begin{aligned} y f_{Y_i}(y;\mu_i,\phi) &= \frac{1}{\Gamma(m_i/\phi)}\left(\frac{m_i}{\phi\mu_i}\right)^{m_i/\phi} y^{m_i/\phi+1-1}e^{-ym_i/\phi\mu_i}\\ &= \frac{1}{\Gamma(a_i)} b_i^{a_i} y^{a_i+1-1}e^{-b_i y}\\ &= \frac{a_i}{b_i}\frac{1}{\Gamma(a_i+1)} b_i^{a_i+1} y^{a_i+1-1}e^{-b_i y}\\ &= \mu_i f_{Y_i}(y; a_i+1, b_i),\end{aligned}$$

where $f_{Y_i}(y; a_i+1, b_i)$ is the pdf of gamma distribution with $a_i + 1$ and b_i. Then, we have

$$\frac{\partial}{\partial \mu_i}F_{Y_i}(y_i;\mu_i,\phi) = \frac{m_i}{\phi\mu_i}[F_{Y_i}(y_i; a_i+1, b_i) - F_{Y_i}(y_i;\mu_i,\phi)].$$

Thus, we have

$$\begin{aligned}\frac{\partial l_3}{\partial \boldsymbol{\beta}} = \sum_{i\in\mathcal{I}_+} \Bigg\{ & \mathbb{1}_1(m_i)\left[\Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right]^{-1}\phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right) \\ & -\mathbb{1}_{2+}(m_i)\left[\Phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right]^{-1}\phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\Bigg\}\frac{\rho}{\sqrt{1-\rho^2}}\frac{1}{\phi(\Phi^{-1}(u_{i,2}))} \\ & \times \frac{m_i}{\phi\mu_i}[F_{Y_i}(y_i;a_i+1,b_i)-F_{Y_i}(y_i;\mu_i,\phi)]\mu_i \boldsymbol{d}_i. \end{aligned} \tag{E.5}$$

To compute the partial derivative of l_3 with respect to ϕ, we only need to compute:

$$\frac{\partial}{\partial\phi}u_{i,2} = \frac{\partial}{\partial\phi}F_{Y_i}(y_i;\mu_i,\phi).$$

Similarly,

$$\begin{aligned}\frac{\partial}{\partial\phi}F_{Y_i}(y_i;\mu_i,\phi) &= \frac{\partial}{\partial\phi}\int_0^{y_i} f_{Y_i}(y;\mu_i,\phi)dy = \int_0^{y_i}\frac{\partial}{\partial\phi}f_{Y_i}(y;\mu_i,\phi)dy \\ &= \int_0^{y_i}\frac{\partial}{\partial\phi}\exp\left\{\log f_{Y_i}(y;\mu_i,\phi)\right\}dy \\ &= \int_0^{y_i} f_{Y_i}(y;\mu_i,\phi)\frac{\partial}{\partial\phi}\log f_{Y_i}(y;\mu_i,\phi)dy, \end{aligned}$$

where $\frac{\partial}{\partial\phi}\log f_{Y_i}(y;\mu_i,\phi)$ is given as

$$\frac{\partial}{\partial\phi}\log f_{Y_i}(y;\mu_i,\phi) = \left[\frac{y}{\exp\langle\boldsymbol{\beta},\boldsymbol{d}_i\rangle}+\langle\boldsymbol{\beta},\boldsymbol{d}_i\rangle-\log y-1\right]\frac{m_i}{\phi^2}-\frac{m_i}{\phi^2}\log\left(\frac{m_i}{\phi}\right)+\frac{\Gamma'(m_i/\phi)}{\Gamma(m_i/\phi)}\frac{m_i}{\phi^2}.$$

Thus, we have

$$\begin{aligned}\frac{\partial l_3}{\partial \phi} = \sum_{i\in\mathcal{I}_+} \Bigg\{ & \mathbb{1}_{m_i}(1)\left[\Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right]^{-1}\phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right) \\ & -\mathbb{1}_{m_i}(2+)\left[\Phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right]^{-1}\phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\Bigg\}\frac{\rho}{\sqrt{1-\rho^2}}\frac{1}{\phi(\Phi^{-1}(u_{i,2}))} \\ & \times \int_0^{y_i} f_{Y_i}(y;\mu_i,\phi)\frac{\partial}{\partial\phi}\log f_{Y_i}(y;\mu_i,\phi)dy. \end{aligned} \tag{E.6}$$

We now compute the partial derivatives of l_3 with respect to ρ.

$$\begin{aligned}\frac{\partial l_3}{\partial \rho} &= \frac{\partial}{\partial\rho}\sum_{i\in\mathcal{I}_+}\left[\mathbb{1}_1(m_i)\log\Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)+\mathbb{1}_{2+}(m_i)\log\Phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right] \\ &= \sum_{i\in\mathcal{I}_+}\Bigg\{\mathbb{1}_1(m_i)\left[\Phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right]^{-1}\phi\left(-\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\left[\frac{z_{i,2}-\rho z_{i,1}}{(1-\rho^2)^{3/2}}\right] \\ &\quad +\mathbb{1}_{2+}(m_i)\left[\Phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\right]^{-1}\phi\left(\frac{z_{i,1}-\rho z_{i,2}}{\sqrt{1-\rho^2}}\right)\left[\frac{\rho z_{i,1}-z_{i,2}}{(1-\rho^2)^{3/2}}\right]\Bigg\}. \end{aligned} \tag{E.7}$$

By applying equations (E.1), (E.2), (E.3), (E.4), (E.5), (E.6), (E.7), we get the following partial derivatives:

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\alpha}}l(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho) &= \frac{\partial}{\partial\boldsymbol{\alpha}}l_1(\boldsymbol{\alpha})+\frac{\partial}{\partial\boldsymbol{\alpha}}l_3(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho), \\ \frac{\partial}{\partial\boldsymbol{\beta}}l(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho) &= \frac{\partial}{\partial\boldsymbol{\beta}}l_2(\boldsymbol{\beta},\phi)+\frac{\partial}{\partial\boldsymbol{\beta}}l_3(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho), \\ \frac{\partial}{\partial\phi}l(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho) &= \frac{\partial}{\partial\phi}l_2(\boldsymbol{\beta},\phi)+\frac{\partial}{\partial\phi}l_3(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho), \\ \frac{\partial}{\partial\rho}l(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho) &= \frac{\partial}{\partial\rho}l_3(\boldsymbol{\rho},\boldsymbol{\beta},\phi,\rho), \end{aligned} \tag{E.8}$$

Therefore, the scoring function $s(\boldsymbol{\theta})$ of proposed model is given by

$$s(\boldsymbol{\theta})=\frac{\partial}{\partial\boldsymbol{\theta}}l(\boldsymbol{\theta})=\left(\frac{\partial}{\partial\boldsymbol{\alpha}}l(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho)^{\top},\frac{\partial}{\partial\boldsymbol{\beta}}l(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho)^{\top},\frac{\partial}{\partial\phi}l(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho),\frac{\partial}{\partial\rho}l(\boldsymbol{\alpha},\boldsymbol{\beta},\phi,\rho)\right)^{\top}.$$

For the log-normal distributed severity, we can follow the above procedure to get the corresponding scoring functions.

Appendix F. Tables

Table 7
Estimated coefficients under the gamma distributed severity.

	Global MLE			Two-stage estimation			IFM estimation		
	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
Case 1									
$\alpha_0=-0.5$	-0.4739	0.0261	0.1108	-0.4754	0.0246	0.1135	-0.4754	0.0246	0.1135
$\alpha_1=-0.04$	-0.0407	-0.0007	0.0028	-0.0407	-0.0007	0.0028	-0.0407	-0.0007	0.0028
$\alpha_2=0.3$	0.3026	0.0026	0.0677	0.3019	0.0019	0.0664	0.3019	0.0019	0.0664
$\beta_0=-1$	-1.0100	-0.0100	0.1621	-1.0099	-0.0099	0.1621	-1.0456	-0.0456	0.1686
$\beta_1=0.1$	0.1001	0.0001	0.0040	0.1001	0.0001	0.0040	0.1006	0.0006	0.0040
$\beta_2=-0.2$	-0.2055	-0.0055	0.1074	-0.2055	-0.0055	0.1075	-0.2095	-0.0095	0.1092
$\phi=2$	1.9965	-0.0035	0.0855	1.9965	-0.0035	0.0855	1.9837	-0.0163	0.0855
$\rho=0.2$	0.2032	0.0032	0.0701	0.2031	0.0031	0.0701	0.2051	0.0051	0.0721
Case 2									
$\alpha_0=-0.5$	-0.5309	-0.0309	0.1371	-0.5350	-0.0350	0.1370	-0.5350	-0.0350	0.1370
$\alpha_1=-0.04$	-0.0393	0.0007	0.0032	-0.0392	0.0008	0.0032	-0.0392	0.0008	0.0032
$\alpha_2=0.3$	0.3066	0.0066	0.0787	0.3059	0.0059	0.0773	0.3059	0.0059	0.0773
$\beta_0=-1$	-1.0199	-0.0199	0.1436	-1.0195	-0.0195	0.1435	-1.1067	-0.1067	0.1821
$\beta_1=0.1$	0.1002	0.0002	0.0036	0.1002	0.0002	0.0036	0.1015	0.0015	0.0041
$\beta_2=-0.2$	-0.1898	0.0102	0.1029	-0.1898	0.0102	0.1029	-0.1971	0.0029	0.1088
$\phi=2$	1.9935	-0.0065	0.0871	1.9932	-0.0068	0.0871	1.9600	-0.0400	0.0954
$\rho=0.5$	0.5052	0.0052	0.0616	0.5050	0.0050	0.0615	0.5064	0.0064	0.0621
Case 3									
$\alpha_0=-0.5$	-0.4899	0.0101	0.1147	-0.4895	0.0105	0.1206	-0.4895	0.0105	0.1206
$\alpha_1=-0.04$	-0.0401	-0.0001	0.0030	-0.0401	-0.0001	0.0032	-0.0401	-0.0001	0.0032
$\alpha_2=0.3$	0.2830	-0.0170	0.0626	0.2836	-0.0164	0.0656	0.2836	-0.0164	0.0656
$\beta_0=-1$	-1.0305	-0.0305	0.1377	-1.0307	-0.0307	0.1376	-1.1850	-0.1850	0.2514
$\beta_1=0.1$	0.1004	0.0004	0.0040	0.1004	0.0004	0.0040	0.1028	0.0028	0.0057
$\beta_2=-0.2$	-0.2074	-0.0074	0.1052	-0.2075	-0.0075	0.1054	-0.2238	-0.0238	0.1182
$\phi=2$	1.9910	-0.0090	0.0707	1.9908	-0.0092	0.0714	1.9430	-0.0570	0.1006
$\rho=0.9$	0.8995	-0.0005	0.0154	0.8992	-0.0008	0.0155	0.8699	-0.0301	0.0402

Table 8
Estimated coefficients under the log-normal distributed severity.

	Global MLE			Two-stage estimation			IFM estimation		
	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
Case 4									
$\alpha_0=-0.5$	-0.4736	0.0264	0.1112	-0.4754	0.0246	0.1135	-0.4754	0.0246	0.1135
$\alpha_1=-0.04$	-0.0407	-0.0007	0.0028	-0.0407	-0.0007	0.0028	-0.0407	-0.0007	0.0028
$\alpha_2=0.3$	0.3025	0.0025	0.0675	0.3019	0.0019	0.0664	0.3019	0.0019	0.0664
$\beta_0=-1$	-1.0337	-0.0337	0.2279	-1.0335	-0.0335	0.2280	-1.0353	-0.0353	0.2271
$\beta_1=0.1$	0.1008	0.0008	0.0056	0.1008	0.0008	0.0057	0.1009	0.0009	0.0055
$\beta_2=-0.2$	-0.2067	-0.0067	0.1519	-0.2066	-0.0066	0.1519	-0.2072	-0.0072	0.1555
$\sigma=2$	1.9927	-0.0073	0.0538	1.9927	-0.0073	0.0538	1.9967	-0.0033	0.0535
$\rho=0.2$	0.2047	0.0047	0.0684	0.2047	0.0047	0.0684	0.2043	0.0043	0.0683
Case 5									
$\alpha_0=-0.5$	-0.5319	-0.0319	0.1373	-0.5350	-0.0350	0.1370	-0.5350	-0.0350	0.1370
$\alpha_1=-0.04$	-0.0392	0.0008	0.0032	-0.0392	0.0008	0.0032	-0.0392	0.0008	0.0032
$\alpha_2=0.3$	0.3064	0.0064	0.0785	0.3059	0.0059	0.0773	0.3059	0.0059	0.0773
$\beta_0=-1$	-0.9962	0.0038	0.2215	-0.9955	0.0045	0.2219	-0.9996	0.0004	0.2219
$\beta_1=0.1$	0.0995	-0.0005	0.0055	0.0994	-0.0006	0.0055	0.0995	-0.0005	0.0055
$\beta_2=-0.2$	-0.1859	0.0141	0.1428	-0.1858	0.0142	0.1426	-0.1801	0.0199	0.1527
$\sigma=2$	1.9919	-0.0081	0.0603	1.9918	-0.0082	0.0603	1.9946	-0.0054	0.0609
$\rho=0.5$	0.5069	0.0069	0.0613	0.5067	0.0067	0.0613	0.5055	0.0055	0.0604
Case 6									
$\alpha_0=-0.5$	-0.4878	0.0122	0.1150	-0.4895	0.0105	0.1206	-0.4895	0.0105	0.1206
$\alpha_1=-0.04$	-0.0401	-0.0001	0.0031	-0.0401	-0.0001	0.0032	-0.0401	-0.0001	0.0032
$\alpha_2=0.3$	0.2843	-0.0157	0.0633	0.2836	-0.0164	0.0656	0.2836	-0.0164	0.0656
$\beta_0=-1$	-1.0483	-0.0483	0.1836	-1.0474	-0.0474	0.1847	-1.0631	-0.0631	0.2395
$\beta_1=0.1$	0.1011	0.0011	0.0047	0.1010	0.0010	0.0048	0.1013	0.0013	0.0064
$\beta_2=-0.2$	-0.2137	-0.0137	0.1276	-0.2132	-0.0132	0.1268	-0.2136	-0.0136	0.1637
$\sigma=2$	1.9871	-0.0129	0.0580	1.9868	-0.0132	0.0582	1.9929	-0.0071	0.0628
$\rho=0.9$	0.9008	0.0008	0.0149	0.9007	0.0007	0.0149	0.8972	-0.0028	0.0149

References

Czado, C., Kastenmeier, R., Brechmann, E.C., et al., 2012. A mixed copula model for insurance claims and claim sizes. Scandinavian Actuarial Journal 2012 (4), 278–305.
Demarta, S., McNeil, A.J., 2005. The *t* copula and related copulas. International Statistical Review 73 (1), 111–129.
Frees, E.W., Gao, J., Rosenberg, M.A., 2011a. Predicting the frequency and amount of health care expenditures. North American Actuarial Journal 15 (3), 377–392.
Frees, E.W., Meyers, G., Cummings, D.A., 2011b. Summarizing insurance scores using a Gini index. Journal of the American Statistical Association 495, 1085–1098.
Frees, E.W., Wang, P., 2006. Copula credibility for aggregate loss models. Insurance. Mathematics & Economics 38 (2), 360–373.
Garrido, J., Genest, C., Schulz, J., 2016. Generalized linear models for dependent frequency and severity of insurance claims. Insurance. Mathematics & Economics 70, 205–215.
Genest, C., Nešlehová, J., 2007. A primer on copulas for count data. ASTIN Bulletin 37 (2), 475–515.
Jeong, H., Valdez, E.A., 2020. Predictive compound risk models with dependence. Insurance. Mathematics & Economics 94, 182–195.
Joe, H., 2014. Dependence Modelling with Copulas. Chapman & Hall/CRC.
Krämer, N., Brechmann, E.C., Silvestrini, D., Czado, C., 2013. Total loss estimation using copula-based regression models. Insurance. Mathematics & Economics 53, 829–839.
Lee, W., Park, S.C., Ahn, J.Y., 2019. Investigating dependence between frequency and severity via simple generalized linear models. Journal of the Korean Statistical Society 48 (1), 13–28.
McNeil, A.J., Frey, R., Embrechts, P., 2015. Quantitative Risk Management: Concept, Techniques and Tools, revised edition. Princeton University Press.
Oh, R., Shi, P., Ahn, J.Y., 2020. Bonus-malus premiums under the dependent frequency-severity modeling. Scandinavian Actuarial Journal 2020 (3), 172–195.
Ramsay, C.M., 2006. The distribution of sums of certain i.i.d. Pareto variates. Communications in Statistics. Theory and Methods 35 (3), 395–405.
Shi, P., Feng, X.P., Ivantsova, A., 2015. Dependent frequency-severity modelling of insurance claims. Insurance. Mathematics & Economics 64, 417–428.
Shi, P., Valdez, E.A., 2014. Longitudinal modeling of insurance claim counts using jitters. Scandinavian Actuarial Journal 2014 (2), 159–179.
Shi, P., Zhao, Z.F., 2020. Regression for copula-linked compound distributions with application in modelling aggregate insurance claims. Annals of Applied Statistics 14 (1), 357–380.
Song, P.X.K., Li, M., Yuan, Y., 2009. Joint regression analysis of correlated data using Gaussian copulas. Biometrics 65, 60–68.