Sarmanov distribution for modeling dependence between the frequency and the average severity of insurance claims

Raluca Vernic[a,*], Catalina Bolancé[b], Ramon Alemany[b]

[a]*Faculty of Mathematics and Informatics, Ovidius University of Constanta, 124 Mamaia Blvd., 8700 Constanta, Romania*
[b]*Department of Econometrics, Riskcenter-IREA, University of Barcelona, Av. Diagonal, 690, 08034 Barcelona, Spain*

Abstract

Real data studies emphasized situations where the classical independence assumption between the frequency and the severity of claims does not hold in the collective model. Therefore, there is an increasing interest in defining models that capture this dependence. In this paper, we introduce such a model based on Sarmanov's bivariate distribution, which has the ability of joining different types of marginals in flexible dependence structures. More precisely, we join the claims frequency and the average severity by means of this distribution. We also suggest a maximum likelihood estimation procedure to estimate the parameters and illustrate it both on simulated and real data.
Keywords: dependence, Sarmanov distribution, frequency, severity, parameters estimation

2000 MSC: 62P05, 91B30, 91B70

---

[*]Corresponding author.
Email address: rvernic@univ-ovidius.ro (Raluca Vernic)

1. Introduction

When modeling aggregate claims with the classical collective model, the usual assumption is that claim frequency and severity are independent, an assumption which facilitates the corresponding computations. In practice, however, claim frequency and severity tend to be dependent, albeit minimally. For example, in auto insurance data, some negative or positive dependence could be found; on one hand, a high frequency can be associated with an urban driving area where the costs are low or, on the other hand, the same high frequency can be associated with daily journeys on secondary roads where accident costs are usually higher. Another example is found in health insurance data, where the dependence between frequency and severity is usually positive. Furthermore, the sample estimation of the dependence between these two variables is not easy to measure; classical correlation coefficient can provide distorted results that can be affected by a few events. For all these reasons, recently, there is an increasing interest in exploring models that account the dependence between frequency and severity. In this sense, two different approaches can be distinguished: on one hand, a model is defined for the average claim size distribution using the number of claims as covariate (see Frees and Wang, 2006; Gschlößl and Czado, 2007; Frees et al., 2011; Garrido et al., 2016; Valdez et al., 2018); as a second approach, the frequency and severity (or average severity) components are related through a copula (see Erhardt and Czado, 2012; Czado et al., 2012; Krämer et al., 2013; Hua, 2015; Lee and Shi, 2019; Oh et al., 2020; Shi et al., 2015). Alternatively, in this paper, we propose the bivariate Sarmanov distribution to model the bivariate distribution relating the frequency and the average severity of claims; our main motivation is that, similarly to copulas, this distribution allows us to separate the dependent structure from the marginal distributions and, in the same way as the copula-based models, we can easily fit the joint behavior of different marginal distributions, continuous or discrete. Furthermore, unlike copula-based models, the Sarmanov distribution does not add difficulty to the estimation of discrete marginals.

Thus, as in Czado et al. (2012), we introduce dependence between the number of claims and the corresponding average claim size, but, in contrast to these authors, who modeled this dependence by a Gaussian copula, we assume a Sarmanov dependence between the frequency and the average severity. As Czado et al. (2012) did, to estimate the parameters we propose a maximization by parts of the log-likelihood function, but given our bounded parametric space, to optimize each part we use the `optim()` function of `R` and validate our algorithm with a simulation study.

Due to its ability to join different marginals in flexible dependence structures and to its tractability, Sarmanov's multivariate distribution (see Sarmanov, 1966) recently gained a lot of attention in the actuarial literature in several aspects, like: modeling continuous claim sizes (see Bahraoui et al., 2015); modeling discrete claim frequencies (see Abdallah et al., 2016; Bolancé and Vernic, 2019); in the evaluation of ruin probabilities (see, for example, Yang and Yuen, 2016; Guo et al., 2017), etc. In some of the just mentioned papers, the Sarmanov distribution has been fitted in its bivariate and trivariate forms to real insurance data and it proved to provide a better fit than other distributions, including copula ones. In Bolancé and Vernic (2019) and Abdallah et al. (2016), the flexibility of the Sarmanov distribution allows to consider generalized linear model for the marginals and to use a Bayesian approach for credibility models based on the number of claims. Moreover, regarding the alternative copula approach (e.g., elliptical), a discussion in

Bolancé and Vernic (2019) emphasizes some disadvantages of this approach (e.g., elliptical copulas) compared with Sarmanov, especially when working with discrete variables. We focus on obtaining pure and risk premiums for a homogeneous portfolio, using the collective risk model and assuming dependence between the number and the average cost of claims; to this purpose, the proposed bivariate Sarmanov model provides closed type expressions for both the mean and variance of the aggregate claims.

In this paper, we make particular use of the special capacity of the Sarmanov distribution to join marginals of different types, more precisely, one marginal will be of discrete type, corresponding to the claim frequency, and a second marginal will be continuous, representing the average severity. This flexibility, associated with combining various marginal distributions, allows us to propose alternative models that mix a count data distribution for the frequency with a continuous distribution for the average severity.

The assumption of independence between frequency and severity allows a direct fit of the distribution of the total cost of claims S; therefore, very extreme total costs could be observed and a heavy tail distribution could be necessary for fitting this part of the distribution (see McNeil, 1997). In this paper, we assume that $S = NX$, where N is the number of claims and X is the average cost per policyholder, with $X > 0$ if $N > 0$ and $X = 0$ if $N = 0$. In our bivariate Sarmanov model, we propose the Gamma distribution for $X > 0$, distribution that is widely used in this field (see Garrido et al., 2016; Jeong and Valdez, 2020), and we analyze alternative count distribution for N (i.e., Poisson, Negative Binomial and their zero inflated forms). Although extreme values in the mean cost variable might be smoothed, they can occur, and in this case, the Gamma distribution might not work and alternative mean cost distributions should be analyzed. Our model allows for the consideration of other such distributions, but we restricted to the Gamma distribution because it has flexibility, it is adequate to model a right skewed distribution and we were able to deduce closed type expressions for the main results on the distribution of the total cost S.

The proposed model takes into account that a cost only exists if the claim frequency is 1 or more. Therefore, it is specified in two parts: the first part corresponds to the probability of 0 frequency and severity, and the second part to the bivariate probability of frequency and severity larger than 0.

A possible limitation of our compound Sarmanov-based distributions is that the dependency is related to a bounded parameter, which in some cases does not allow fitting strong correlations. However, our experience has shown that the correlation between the number and the amount of claims is not very high - a correlation lower than 0.5 is common. For example, Czado et al. (2012), using Mixed Copula models, estimated a correlation parameter equal to 0.1366; although statistically significant, even lower correlations can be found. Specifically, we illustrate this using a real data set consisting of a random sample of auto insurance policyholders.

The rest of the paper is organized as follows: in Section 2, we describe the proposed Sarmanov distribution, its properties, particular cases and estimation procedure. In Section 3, we present the results of a simulation study to evaluate the estimated parameters using a two parts log-likelihood maximization. An application to a real data set containing auto insurance number and average cost of claims is discussed in Section 4. Finally, we conclude in Section 5. The paper ends with an appendix containing some proofs.

2. Collective model with dependent number and average size of claims

We shall introduce dependence between the number of claims N and the corresponding average claim size X of a portfolio or of a certain policy. Letting S denote the aggregate claims, clearly

$$S = NX. \tag{1}$$

We let p denote the probability mass function (pmf) of N. In respect of the random variable (r.v.) X, its distribution will have both an absolutely continuous component with probability density function (pdf) denoted by f_X and a probability mass at 0. Therefore, the distribution of S also has a probability mass at 0 and a pdf that we denote by f_S. We denote the cumulative distribution function (cdf) of a r.v. by F indexed with the name of that r.v..

2.1. *Sarmanov dependence*

We assume a Sarmanov dependence between N and X as follows

$$f_{X,N}(x,n) = \begin{cases} p(0), \; n = x = 0 \\ p(n) f(x) (1 + \omega \psi(n) \phi(x)), \; n \geq 1, \; x > 0 \end{cases}, \tag{2}$$

where f is a pdf, ψ and ϕ are bounded non-constant kernel functions and $\omega \in \mathbb{R}$. Clearly, we assume that if no claims are reported, the cost to the insurance company is zero, so that if $N = 0$, directly $X = 0$ and hence the total cost $S = 0$.

We call the pdf (2) mixed because it joins the continuous pdf f and the discrete pmf p. Also, in order for (2) to define a proper pdf, we impose the conditions

$$\sum_{n \geq 1} \psi(n) p(n) = \int_{\mathbb{R}} \phi(x) f(x) dx = 0 \text{ and} \tag{3}$$

$$1 + \omega \psi(n) \phi(x) \geq 0, \text{ for all } n \geq 1, \; x > 0. \tag{4}$$

For details on Sarmanov distribution see Kotz et al. (2000), Ting Lee (1996).

To simplify the writing, we denote by Y a r.v. having pdf f and representing $X>0$. Letting $m_1=\inf\limits_{n\geq 1}\psi(n), m_2=\inf\limits_{x>0}\phi(x), M_1=\sup\limits_{n\geq 1}\psi(n), M_2=\sup\limits_{x>0}\phi(x)$, condition (4) restricts ω to the following interval

$$\max\left\{-\frac{1}{m_1 m_2},-\frac{1}{M_1 M_2}\right\}\leq\omega\leq\min\left\{-\frac{1}{m_1 M_2},-\frac{1}{M_1 m_2}\right\}. \tag{5}$$

The following proposition presents the distributions of X, of S and conditional distributions.

Proposition 1 *Under the Sarmanov dependence condition (2), it holds that*

$$\begin{aligned}
i)\ \Pr(X=0) &= p(0),\\
f_X(x) &= (1-p(0))\, f(x),\ x>0.\\
ii)\ \Pr(X=0\,|N=n) &= \begin{cases}1,\ n=0\\ 0,\ n\geq 1\end{cases},\\
f_{X|N=n}(x) &= f(x)\left(1+\omega\psi(n)\,\phi(x)\right),\ x>0,\ n\geq 1.\\
iii)\ \Pr(N=n\,|X=x) &= \begin{cases}1,\ n=x=0\\ \frac{p(n)}{1-p(0)}\left(1+\omega\psi(n)\,\phi(x)\right),\ n\geq 1,\ x>0\end{cases}.\\
iv)\ \Pr(S=0) &= p(0),\\
f_S(s) &= \sum_{n\geq 1}\frac{p(n)}{n} f\left(\frac{s}{n}\right)\left(1+\omega\psi(n)\,\phi\left(\frac{s}{n}\right)\right),\ s>0.
\end{aligned}$$

The first two moments of S are given in the following result; note that they are expressed in terms of the r.v. Y.

Proposition 2 *Under the Sarmanov dependence condition (2), the expected value* *and variance of S are given respectively, by*

$$\begin{aligned}\mathbb{E}S &= \mathbb{E}N\mathbb{E}Y + \omega\mathbb{E}\left[N\psi(N)\right]\mathbb{E}\left[Y\phi(Y)\right],\\ VarS &= \mathbb{E}\left[Y^2\right]VarN + \mathbb{E}^2\left[N\right]VarY - \omega^2\mathbb{E}^2\left[N\psi(N)\right]\mathbb{E}^2\left[Y\phi(Y)\right]\\ &\quad +\omega\left(\mathbb{E}\left[N^2\psi(N)\right]\mathbb{E}\left[Y^2\phi(Y)\right] - 2\mathbb{E}N\,\mathbb{E}\left[N\psi(N)\right]\,\mathbb{E}Y\,\mathbb{E}\left[Y\phi(Y)\right]\right).\end{aligned}$$

Proposition 3 *The correlation coefficient of the pdf (2) is given by*

$$corr(X,N) = \frac{\omega\mathbb{E}\left[N\psi(N)\right]\mathbb{E}\left[Y\phi(Y)\right] + p(0)\,\mathbb{E}N\mathbb{E}Y}{\sqrt{(1-p(0))\left(VarY + p(0)\,\mathbb{E}^2\left[Y\right]\right)\,VarN}}. \tag{6}$$

The proofs of the previous propositions are omitted because they are rather straight forward to derive and part of them can be found in Ting Lee (1996).

The correlation defined in (6) takes into account the two parts of the distribution, i.e. $N = X = 0$ and $N,X > 0$. We note that if $\omega = 0$ then $corr(X,N)$ depends on the probability of zero claims $p(0)$; only if $p(0) = 0$ then $\omega = 0$ implies $corr(X,N) = 0$.

There are some common types of Sarmanov kernels, from which we note (see Ting Lee, 1996): the kernels based on cdfs leading to the Farlie-Gumbel-Morgenstern distribution, which, however, has a correlation coefficient limited by $1/3$; the kernels based on the moments of the distributions, which, in order to be bounded, necessitate the truncation of the distributions; the exponential kernel, which is bounded by its nature and easy to handle for our particular distributions. Therefore, we propose to use exponential kernels. Regarding Sarmanov's pdf in (2), we consider in particular the exponential kernels satisfying condition (3), and we emphasize in their notation the kernel parameter. More precisely, $\phi(y,\gamma) = e^{-\gamma y} - \mathcal{L}_Y(\gamma)$, where $\mathcal{L}_Y$ denotes the Laplace transform of the r.v. Y, and γ, the kernel parameter, is inserted into the notation $\phi(y)$. Furthermore, we let $\psi(n,\delta) = e^{-\delta n} - k$, and to find k, we write

$$\begin{aligned}\sum_{n\geq 1}\psi(n,\delta)\,p(n) &= \sum_{n\geq 1}\left(e^{-\delta n}-k\right)p(n)\\ &= \sum_{n\geq 0}e^{-\delta n}p(n)-p(0)-k\left(\sum_{n\geq 0}p(n)-p(0)\right)\\ &= \mathscr{L}_N(\delta)-p(0)-k(1-p(0)).\end{aligned}$$

Imposing the condition expressed in (3), i.e. $\sum_{n\geq 1}\psi(n,\delta)\,p(n) = 0$, we obtain $k = \frac{\mathscr{L}_N(\delta)-p(0)}{1-p(0)}$. Therefore, $\psi(n,\delta) = e^{-\delta n} - \frac{\mathscr{L}_N(\delta)-p(0)}{1-p(0)}$ because in the second formula of the pdf (2) we have $n \geq 1$ (similar to a left truncation of N in 0).

The parameters δ and γ are part of the Laplace operators whose values affect the interval defined in (5): the larger the values, the wider the interval, i.e. these parameters have a scale effect on the dependence parameter. Therefore, too large values can lead to inefficient estimates of the dependency parameter, while too small values can lead to downwardly biased dependency parameters. In the simulation study we illustrate this effect.

We also note that in model (1), when N is larger, the variance of the average severity X should become smaller; from the conditional density $f_{X|N=n}(x)$ presented in Proposition 1, it can be seen that the proposed Sarmanov model is able to capture this behavior due to the kernel function $\psi(n,\delta)$, which decreases when n increases, and which interferes in e.g., the variance of X given N.

2.2. Simulation from the collective model

To simulate values from the two parts bivariate Sarmanov distribution whose pdf is defined in (2), we use the inversion method from the conditional cdf of X

given $N=n$, which easily results from (*ii*) in Proposition 1 as

$$\begin{aligned} F_{X|N=0}(0) &= 1, \\ F_{X|N=n}(x) &= \int_0^x f(y)\left(1+\omega\psi(n,\delta)\phi(y,\gamma)\right)dy \\ &= F_Y(x)+\omega\psi(n,\delta)\int_0^x f(y)\phi(y,\gamma)\,dy,\ n\geq 1,\ x>0. \end{aligned} \tag{7}$$

Hence, we simulate the value n from the distribution of N. If $n=0$ then clearly $x=0$; otherwise, we generate an uniform $U(0,1)$ value u and solve the equation $F_{X|N=n}(x)=u$ for x. This yields the generated pair (n,x).

Moreover, the Gibbs sampler can be used by drawing iteratively from both conditional cdfs (see Casella and George, 1992). Therefore, we also need the conditional cdf of N given $X=x$, i.e.,

$$\begin{aligned} F_{N|X=0}(0) &= 1, \\ F_{N|X=x}(n) &= \sum_{k=1}^{n}\Pr(N=k|X=x)=\sum_{k=1}^{n}\frac{p(k)}{1-p(0)}\left(1+\omega\psi(k,\delta)\phi(x,\gamma)\right) \\ &= \frac{1}{1-p(0)}\left[F_N(n)-p(0)+\omega\phi(x,\gamma)\sum_{k=1}^{n}\psi(k,\delta)p(k)\right], n\geq 1, x>0. \end{aligned}$$

2.3. Parameters estimation

Let $(n_i,x_i)_{i=1}^{K}$ be a random bivariate sample of the number and average amount of claims. Let θ and ν be, respectively, the parameters vectors of the marginal distribution of N and of the continuous marginal distribution of Y, while ω is the dependence parameter of Sarmanov's distribution. Based on (2), the log-likelihood function is

$$\begin{aligned}
\ln L\left((n_i,x_i)_{i=1}^{K};\theta;\nu;\omega;\delta;\gamma\right) &= \sum_{\{i:n_i=x_i=0\}} \ln p\,(0;\theta) + \sum_{\{i:n_i\geq 1,x_i>0\}} [\ln p\,(n_i;\theta) \\
&\quad + \ln f\,(x_i;\nu) + \ln\left(1+\omega\psi\,(n_i,\delta)\,\phi\,(x_i,\gamma)\right)] \\
&= \ln L\left((n_i)_{i=1}^{K};\theta\right) + \ln L\left(\{x_i\,|x_i>0, i=1,\ldots,K\};\nu\right) \\
&\quad + \sum_{\{i:n_i\geq 1,x_i>0\}} \ln\left(1+\omega\psi\,(n_i,\delta)\,\phi\,(x_i,\gamma)\right), \qquad (8)
\end{aligned}$$

where $L\left((n_i)_{i=1}^{K};\theta\right)$ is the likelihood function corresponding to the marginal r.v. N, while $L\left(\{x_i\,|x_i>0, i=1,\ldots,K\};\nu\right)$ is the one corresponding to Y.

Maximizing the log-likelihood expressed in (8) is very difficult, mainly for two reasons. The first reason is because, given the limits of the dependency parameter ω that were defined in (5), the parametric space is bounded. The second reason is due to the strong relationship that exists between the dependence parameter and the marginal ones.

We also define the log-likelihood function in (8) assuming that some parameters are known. Let $\ln L\left((n_i,x_i)_{i=1}^{K};\theta;\nu\middle|\,\omega;\delta;\gamma\right)$ be the log-likelihood function defined in (8) given that the parameters ω, δ and γ associated to the dependence structure are known; similarly, let $\ln L\left((n_i,x_i)_{i=1}^{K};\omega;\delta;\gamma\middle|\,\theta;\nu\right)$ be the log-likelihood function defined in (8) given that the marginal parameters θ and ν are known. As in Bolancé and Vernic (2019), we propose to determine the Maximum Likelihood Estimation (MLE) of the parameters in two phases. The first phase consists of maximizing by parts the log-likelihood function in order to obtain initial parameters that will be used in the second phase to obtain a full MLE (an example in a similar context using copulas is given in Czado et al., 2012). The first phase is analogous to the Inference Function for Margins (IFM) method that is commonly used to estimate copula-based models (see Joe, 2005). The aim of second phase is to check if the parameters estimated in the first phase maximize the full log-likelihood and if the asymptotic inference can be done. We note that the simulation study and application results presented in Sections 3 and 4, respectively, show that the differences between the values of the parameters obtained in both phases are very small; changes are found in third or fourth decimal and we can conclude that the differences are due to the algorithm's precision. Bolancé and Vernic (2019) successfully used the same algorithm for estimating a trivariate Sarmanov distribution with Negative Binomial marginal distributions specified as generalized linear models. Moreover, using Sarmanov distribution has advantages over copula models, given the difficulty that is added to the estimation of copula parameters when the variables are discrete. With Sarmanov distribution, we can use the `optim()` function for maximizing partial and full log-likelihood function. The same procedure can be used for estimating distributions where the marginal distributions and dependence structure are separable in the log-likelihood function in the same way as in (8). We describe the procedure below.

Phase 1

Step 0 Using MLE, find initial values for the parameters of the univariate marginal distributions, $\hat{\theta}^0$ and $\hat{\nu}^0$. For the initial parameters in the dependence structure we assume $\omega^0 = 0$ and $\delta^0 = \gamma^0 = 1$.

Step 1 (iteration j) Given the parameters for the marginal distributions in $j-1$, find $\hat{\delta}^j$, $\hat{\gamma}^j$ and $\hat{\omega}^j$ within the interval defined in (5) for this dependence parameter, by maximizing the log-likelihood

$$\ln L\left((n_i, x_i)_{i=1}^{K}; \omega; \delta; \gamma \middle| \hat{\theta}^{j-1}; \hat{\nu}^{j-1} \right).$$

Step 2 Given $\hat{\delta}^j$, $\hat{\gamma}^j$ and $\hat{\omega}^j$, obtain new values for the parameters of the marginal distributions by maximizing the log-likelihood function

$$\ln L\left((n_i, x_i)_{i=1}^{K}; \theta; \nu \middle| \hat{\omega}^j; \hat{\delta}^j; \hat{\gamma}^j \right).$$

Given that the kernel functions also depend on the parameters of the marginal distributions, the maximization is carried out within an interval that guarantees $(1+\omega\psi(n,\delta)\phi(y,\gamma)) > 0$. In practice, we define the interval for the parameters of the marginal distributions as $\left(\hat{\theta}^{j-1}\hat{\nu}^{j-1}\right) \pm \varepsilon$, where ε is defined as $\left(\hat{\theta}^{j-1}\hat{\nu}^{j-1}\right)/a$, with $a > 0$.

Steps 1 and 2 are repeated until convergence. Furthermore, the interval for the dependence parameter ω in Step 1 has to be calculated at each iteration j using the parameters of the marginal distributions and of the kernel functions estimated on the previous iteration $j-1$. In Step 0, the initial values of the parameters δ and γ are fixed at 1; this affects the initial interval of the dependence parameter, which could be too narrow. Therefore, if the dependence parameter is located at an extreme of the interval, the initial values of the parameters δ and γ must be increased.

Phase 2 Starting with the initial parameters estimated in Phase 1, perform full MLE.

Given our bounded parametric space, optimizations in the two phases were carried out using the `optim()` function of `R` with the method `L-BFGS-B` (Byrd et al., 1995).

2.4. *Particular cases*

2.4.1. *Counting distributions*

For the r.v. number of claims, we consider four different distributions: Poisson, Negative Binomial, and their zero inflated forms, Zero Inflated Poisson (ZIP) and Zero Inflated Negative Binomial (ZINB).

If N is Poisson distributed, $N \sim Po(\lambda)$, $\lambda > 0$, we recall that

$$\mathbb{E}N = VarN = \lambda,\ \mathbb{E}\left[N^2\right] = \lambda + \lambda^2,\ \mathscr{L}_N(\delta) = e^{\lambda\left(e^{-\delta}-1\right)}.$$

Assuming that N is Negative Binomial distributed, $N \sim NB(r,p)$, $r > 0$, $p \in (0,1)$, then, with $q = 1-p$,

$$\begin{aligned} \Pr(N=n) &= \frac{\Gamma(r+n)}{n!\Gamma(r)} p^r q^n,\ n \in \mathbb{N}, \\ \mathbb{E}N &= \frac{rq}{p},\ \mathbb{E}\left[N^2\right] = \frac{rq(1+qr)}{p^2},\ VarN = \frac{rq}{p^2},\ \mathscr{L}_N(\delta) = \left(\frac{p}{1-qe^{-\delta}}\right)^r. \end{aligned}$$

If N follows a certain discrete distribution with support $\mathbb{N}$ and $\tilde{N}$ follows the same distribution in the zero inflated form with parameter $\pi \in (0,1)$ (the probability of extra zeros), then the following relations hold

$$\begin{aligned} \Pr\left(\tilde{N}=n\right) &= \begin{cases} \pi + (1-\pi)\Pr(N=0),\ n=0 \\ (1-\pi)\Pr(N=n),\ n \geq 1 \end{cases}, \\ \mathbb{E}\tilde{N} &= (1-\pi)\mathbb{E}N,\ \mathbb{E}\left[\tilde{N}^2\right] = (1-\pi)\mathbb{E}\left[N^2\right],\ Var\tilde{N} = (1-\pi)\left(VarN + \pi\mathbb{E}^2 N\right), \\ \mathscr{L}_{\tilde{N}}(\delta) &= \pi + (1-\pi)\mathscr{L}_N(\delta). \end{aligned}$$

Note that by taking $\pi = 0$ in the above formulas, we obtain the corresponding formulas for the original (not inflated) distribution. Therefore, in the following, we consider that $\pi \in [0,1)$ and present the results for the general inflated forms; in this sense, for simplicity, we drop the tilde from $\tilde{N}$.

Proposition 4 *Let* $\psi(n,\delta) = e^{-\delta n} - \frac{\mathscr{L}_N(\delta)-p(0)}{1-p(0)}$ *be the exponential kernel and* $\pi \in [0,1)$.

i) If $N \sim ZIP(\lambda, \pi)$, then

$$\begin{aligned}
\mathbb{E}[N\psi(N,\delta)] &= (1-\pi)\lambda e^{-\lambda}\left(e^{\lambda e^{-\delta}-\delta} - \frac{e^{\lambda e^{-\delta}}-1}{1-e^{-\lambda}}\right), \\
\mathbb{E}\left[N^2\psi(N,\delta)\right] &= (1-\pi)\lambda e^{-\lambda}\left[e^{\lambda e^{-\delta}-\delta}\left(\lambda e^{-\delta}+1\right) - (\lambda+1)\frac{e^{\lambda e^{-\delta}}-1}{1-e^{-\lambda}}\right].
\end{aligned}$$

ii) If $N \sim ZINB(r, p, \pi)$, then

$$\begin{aligned}
\mathbb{E}[N\psi(N,\delta)] &= (1-\pi)\frac{rqp^r}{\left(1-qe^{-\delta}\right)^r}\left(\frac{1}{e^{\delta}-q} - \frac{1-\left(1-qe^{-\delta}\right)^r}{p(1-p^r)}\right), \\
\mathbb{E}\left[N^2\psi(N,\delta)\right] &= (1-\pi)\frac{rqp^r}{\left(1-qe^{-\delta}\right)^r}\left[\frac{rq+e^{\delta}}{\left(e^{\delta}-q\right)^2} - (1+qr)\frac{1-\left(1-qe^{-\delta}\right)^r}{p^2(1-p^r)}\right].
\end{aligned}$$

2.4.2. Gamma severity distribution

Let Y be Gamma distributed, $Y \sim Ga(\alpha, \beta)$, $\alpha,\ \beta > 0$, where β is the rate parameter. We recall that

$$\mathbb{E}Y = \frac{\alpha}{\beta},\ \mathbb{E}\left[Y^2\right] = \frac{\alpha(\alpha+1)}{\beta^2},\ VarY = \frac{\alpha}{\beta^2},\ \mathscr{L}_Y(\gamma) = \left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}.$$

The following result is needed to evaluate the expected value and variance of S.

Proposition 5 *Let $Y \sim Ga(\alpha, \beta)$, $\alpha,\ \beta > 0$, and let $\phi(x,\gamma) = e^{-\gamma x} - \mathscr{L}_Y(\gamma)$ be the exponential kernel. Then*

$$\begin{aligned}
\mathbb{E}[Y\phi(Y,\gamma)] &= -\frac{\alpha\gamma\beta^{\alpha-1}}{(\beta+\gamma)^{\alpha+1}}, \\
\mathbb{E}\left[Y^2\phi(Y,\gamma)\right] &= -\frac{\alpha(\alpha+1)\gamma\beta^{\alpha-2}(2\beta+\gamma)}{(\beta+\gamma)^{\alpha+2}}.
\end{aligned}$$

We note that the Gamma distribution is a particular case for the mean cost Y and alternative distributions with bounded Laplace transformation can also be used; this could be the subject of future research.

2.4.3. Particular compound distributions

By combining the above discussed counting distributions with the Gamma severity distribution, we obtain four particular compound distributions: compound Poisson-Gamma, compound Zero Inflated Poisson-Gamma, compound Negative Binomial-Gamma and compound Zero Inflated Negative Binomial-Gamma. The next proposition presents pdfs for the general inflated forms; the proof is immediate, hence we omit it.

Proposition 6 *Let $Y \sim Ga(\alpha, \beta)$ and let $\psi(n,\delta) = e^{-\delta n} - \frac{\mathscr{L}_N(\delta) - p(0)}{1-p(0)}$, $\phi(x,\gamma) = e^{-\gamma x} - \left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}$ be the exponential kernels. Then, with $\pi \in [0,1)$:*
i) If $N \sim ZIP(\lambda, \pi)$, then the compound zero inflated Poisson-Gamma pdf is

$$f_{X,N}(x,n) = \begin{cases} \pi + (1-\pi)e^{-\lambda},\ n = x = 0 \\ (1-\pi)\frac{\beta^{\alpha}e^{-\lambda}}{\Gamma(\alpha)}\frac{\lambda^n}{n!}x^{\alpha-1}e^{-\beta x}\left[1+\omega\left(e^{-\delta n} - \pi - (1-\pi)e^{\lambda\left(e^{-\delta}-1\right)}\right)\phi(x,\gamma)\right], \\ \qquad n \geq 1,\ x > 0. \end{cases}$$

ii) If $N \sim ZINB(r,p,\pi)$, then the compound zero inflated Negative Binomial-Gamma pdf is

$$f_{X,N}(x,n) = \begin{cases} \pi + (1-\pi)p^r, n = x = 0 \\ (1-\pi)\frac{\beta^{\alpha}p^r\Gamma(r+n)}{\Gamma(\alpha)\Gamma(r)n!}q^n x^{\alpha-1}e^{-\beta x}\left[1+\omega\left(e^{-\delta n} - \pi - \frac{(1-\pi)p^r}{\left(1-qe^{-\delta}\right)^r}\right)\phi(x,\gamma)\right], \\ \qquad n \geq 1,\ x > 0. \end{cases}$$

To simulate values from such compound distributions by inversion, we use formula (7) of the conditional cdf under the assumptions that $Y \sim Ga(\alpha, \beta)$ and $\phi(x,\gamma) = e^{-\gamma x} - \left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}$. We have

$$\begin{aligned} \int_0^x f(y)\phi(y,\gamma)\,dy &= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\int_0^x y^{\alpha-1}e^{-\beta y}\left(e^{-\gamma y} - \left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}\right)dy \\ &= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\left[\int_0^x \left(y^{\alpha-1}e^{-(\beta+\gamma)y}\right)dy - \left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}\int_0^x y^{\alpha-1}e^{-\beta y}dy\right], \end{aligned}$$

hence, letting $F_{Ga(\alpha,\beta)}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^x y^{\alpha-1} e^{-\beta y} dy$ denote the $Ga(\alpha,\beta)$ cdf, this yields for $n \geq 1$, $x > 0$,

$$F_{X|N=n}(x) = \left[1 - \omega\psi(n,\delta)\left(\frac{\beta}{\beta+\gamma}\right)^\alpha\right] F_{Ga(\alpha,\beta)}(x) + \omega\psi(n,\delta)\left(\frac{\beta}{\beta+\gamma}\right)^\alpha F_{Ga(\alpha,\beta+\gamma)}(x).$$

Therefore, as discussed before, to simulate a pair (n,x), we first simulate the value n from the distribution of N, and if $n \geq 1$, we generate an uniform $U(0,1)$ value u and solve the equation $F_{X|N=n}(x) = u$ for x.

In order to apply Gibbs sampler, we also need $F_{N|X=x}$, which, for the exponential kernel and $n \geq 1$, $x > 0$, is given by

$$\begin{aligned} F_{N|X=x}(n) &= \frac{1}{1-p(0)}\left[F_N(n) - p(0) + \omega\phi(x,\gamma)\sum_{k=1}^{n}\left(e^{-\delta k} - \frac{\mathscr{L}_N(\delta) - p(0)}{1-p(0)}\right)p(k)\right] \\ &= \frac{1}{1-p(0)}\left[(F_N(n) - p(0))\left(1 - \omega\phi(x,\gamma)\frac{\mathscr{L}_N(\delta) - p(0)}{1-p(0)}\right)\right. \\ &+ \left.\omega\phi(x,\gamma)\sum_{k=1}^{n} e^{-\delta k} p(k)\right]. \end{aligned}$$

This will be particularized for a certain distribution of N (with special attention to the zero inflated forms).

3. Simulation Study

To evaluate our proposed estimation procedure, we summarize the results of a simulation study. We compare the Root Mean Square Relative Error (RMSRE) and the Mean Absolute Percentage Error (MAPE) of the estimated parameters associated to the different bivariate Sarmanov distributions that we have analyzed in the previous sections for modeling the dependence between claims frequency and claims average severity. Given that the absolute values of these errors do not carry much meaning, we estimated empirical bootstrap confidence intervals (EBCIs) at 95% confidence level, using $1,000$ resamples with replacement.

Using the Gibbs method (Casella and George, 1992) we generated $1,000$ bivariate samples of sizes $K = 500$ and $K = 5,000$ from the following compound Sarmanov models: Poisson-Gamma (CPG), Negative Binomial-Gamma (CNBG), Zero Inflated Poisson-Gamma (CZIPG) and Zero Inflated Negative Binomial-Gamma (CZINBG). We have selected different parameters for the analyzed distributions such that the expected number of claims is around 0.1 or 0.2. In all the simulated models, we assumed the same parameters for the Gamma marginal distribution: shape $\alpha = 0.3$ and rate $\beta = 0.0006$. Concerning the claim frequency distribution, the kernel parameters δ and γ and the dependence parameter ω, we used those shown in Table 1; we considered four distinct cases for each compound model that we denoted as Mi.1, $i = 1,...,4$, for $\delta = \gamma = 1$ and Mi.2, $i = 1,...,4$, for $\delta = \gamma = 2$. Comparing both groups of models, Mi.1 and Mi.2, we observe the effect of the kernel parameters on the bounds defined in expression (5): the larger the parameters values, the wider is the interval of the dependence parameter ω. In practice, this implies that if the kernel parameters δ and γ are undervalued, the estimated dependence parameter could be biased; on the contrary, the overvaluation of δ and γ will imply a larger dispersion of the estimated dependence parameter.

We have obtained the EBCIs at 95% confidence level of the RMSRE and MAPE for the estimated parameters of the CPG, CNBG, CZIPG and CZINBG distributions, respectively; given the tables we obtained are very large, they are displayed in the Appendix (Tables 7, 8, 9 and 10). The estimated parameters for each sample are obtained using the procedure described in Subsection 2.3; we have noticed that the estimated parameters obtained with this procedure depend very closely on the parameters used for the margins and for the kernel functions in Step 0 of Phase 1. To obtain simulation results for the CPG and CNBG distributions, for all replicates in Step 0, we have used the MLE of the parameters associated with the univariate marginal distribution and the true values for the parameters of the kernel functions. For the the CZIPG and CZINBG distributions, the univariate estimation failed in a small number of replicates (5 for CZIPG and 18 for CZINBG); in these cases, we decided to use in Step 0 the true values of parameters of the marginal distributions.

In general, the obtained EBCIs are narrow. From the results displayed for the CPG and CNBG distributions in Tables 7 and 8 of the Appendix, it can be seen that for the parameters associated to the marginal distributions and kernel functions, in almost all cases, the RMSRE and MAPE have upper confidence interval limits below or near 0.5 for $K = 500$ and below or near 0.15 for $K = 5,000$. The relative errors of the dependence parameter are larger than the ones obtained for the parameters associated to the marginal distributions and kernel functions. This parameter has to be within the limits defined in expression (5). These limits are very sensitive to the parameters associated to the marginal distributions and kernel functions, so that these larger errors are expected. Furthermore, larger values for the kernel parameters δ and γ tend to increase the errors given the larger dispersion.

In what concerns the compound zero inflated distributions, CZIPG and CZINBG, from the results shown in Tables 9 and 10 of the Appendix, we note that in some cases, the relative errors of the parameters of the marginal distributions and kernel functions decrease very slightly when the sample size increases; this is due to the larger error associated with the parameters estimated at Step 0. On the contrary, the results for the dependence parameter lead to similar comments as for the CPG and CNBG distributions.

We also mention that the runtime is fast: to obtain $1,000$ replicates with $K = 5,000$, we need around 10 minutes (i7-7700 CPU, 3.60GHz).

Table 1: Parameters of the bivariate compound Sarmanov models. The Gamma parameters are the same in all the cases: $\alpha = 0.3$ and $\beta = 0.0006$. Dependence bounds between parentheses.

				Mi.1: $\delta=\gamma=1$	Mi.2: $\delta=\gamma=2$
CPG	λ			ω (-26.85,3.25)	ω (-91.99,8.85)
M1.j	0.2			-7	
M2.j	0.2			3	
	λ			ω (-25.99,3.15)	ω (-87.99,8.46)
M3.j	0.1			-7	
M4.j	0.1			3	
CNBG	r	p		ω (-15.45,3.80)	ω (-32.55,10.78)
M1.j	0.3	0.6		-12	
M2.j	0.3	0.6		3	
	r	p		ω (-17.39,3.69)	ω (-36.46,10.41)
M3.j	0.15	0.6		-12	
M4.j	0.15	0.6		3	
CZIPG	λ	π		ω (-24.61,3.48)	ω (-49.30,9.69)
M1.j	0.4	0.5		-12	
M2.j	0.4	0.5		3	
	λ	π		ω (-26.85,3.25)	ω (-91.99,8.85)
M3.j	0.2	0.5		-12	
M4.j	0.2	0.5		3	
CZINBG	r	p	π	ω (-9.79,4.43)	ω (-21.51,12.99)
M1.j	0.3	0.43	0.5	-8	
M2.j	0.3	0.43	0.5	3	
	r	p	π	ω (-17.39,3.69)	ω (-36.46,10.41)
M3.j	0.15	0.6	0.5	-8	
M4.j	0.15	0.6	0.5	3	
i=1,2,3,4 and j=1,2					

4. Numerical example

We now analyze a data set of auto insurance policyholders of an international company. This data set contains a sample of $K = 99,972$ Spanish insureds. This data are specifically designed for this numerical example and represent around 25% of the total policies considered as study object. We have selected annual policies in force in 2013 that have been renewed for at least one time, i.e. the policyholders have been with the company for more than one year. All the selected insureds drive a car for private use. For each individual we have information on the number and on the average cost of claims; these variables are calculated taking into account only the civil liability coverage and at fault material damage claims. We assume that they have a homogeneous risk profile. Our aim is to fit the bivariate Sarmanov distribution and to check the effect of dependence between frequency and severity on the risk premium.

In Table 2, we display the results of the initial analysis that consisted in obtaining the basic descriptives and estimated initial parameters for the marginal distributions assuming independence. At the top of this table, we present the analysis of the number of claims. From the values of the Chi-square statistic, we can see that the best fits are obtained with the NB and ZINB distributions, being somewhat better for the NB. Below the double line in Table 2, we show the basic descriptive statistics for the average cost of claims, together with the estimated parameters of the Gamma distribution for this variable. We also compared the log-likelihood value of the Gamma distribution with some alternative distributions with different tail shapes (and same number of parameters): Weibull, Log-Normal and Log-Logistic; the results are shown in Table 3. We can see that for these data, the best fit is provided by the Gamma distribution.

Table 2: Results of basic descriptive analysis and initial parameters for marginal distributions.

		Po	NB	ZIPo	ZINB
	Initial Parameters	$\lambda = 0.0887$	$r = 0.3171$	$\lambda = 0.3647$	$r = 11.1344$
			$p = 0.7814$	$\pi = 0.7567$	$p = 0.9705$
					$\pi = 0.7374$
Frequency	TRUE				
0	92538.00	91482.28	92524.63	92538.00	92537.99
1	6166.00	8118.58	6285.65	6160.47	6172.32
2	1122.00	360.24	950.48	1123.51	1103.16
3	125.00	10.66	170.11	136.60	142.28
4	18.00	0.24	32.81	12.46	14.81
5	3.00	0.00	1.73	0.06	0.11
Chi-Square	99972.00	6761.20	52.81	152.02	77.09

	Gamma			
Initial Parameters	$\alpha = 0.1881$			
	$\beta = 0.0003$			
	Mean	Median	STDEV	Skewness
Severity	685.63	441.00	1580.81	15.73

The Pearson correlation coefficient between the frequency and severity is 0.4152.

Table 3: Comparing distributions for average severity per policyholder.

	Gamma	Weibull	Log-Normal	Log-Logistic
log-likelihood	51234.75	51213.25	33323.50	49430.50

Table 4 contains the results of the estimated parameters for the bivariate Sarmanov for CNBG and CZINBG; as expected, given the results in Table 2, the results for CNPG and CZIPG were worse, so we did not display them. The starting values of the kernel parameters used to obtain the results in Table 4 were $\delta = \gamma = 1$. Furthermore, the parameters were also estimated using different initial values for the kernel parameters, i.e. $\delta = \gamma = 2$, and the results were practically the same.

We also compared the results obtained using the Sarmanov distribution with the results obtained for the bivariate Gaussian copula (see Czado et al., 2012, who proposed a copula based model with Gamma and Poisson marginal distributions) and with the proposal of Garrido et al. (2016) based on the conditional distribution of the mean severity given the frequency of claims. In both cases, the authors assume $X > 0$ for $N > 0$. We have assumed the same marginal distributions as in Table 4: Gamma for the mean severity and NB and ZINB for the frequency. However, the proposals of Czado et al. (2012) and Garrido et al. (2016) are based on the particular case where the number of claims follows a Poisson distribution and the mean cost per policyholder is Gamma distributed; both papers propose MLE algorithms. Since in our case the distributions that better fit the number of claims are the NB and the ZINB, to estimate the parameters we used an algorithm similar to the one proposed in Subsection 2.3. The models were defined in two parts: for $X = N = 0$ and for $X, N > 0$. In the Appendix, we describe in more details the alternative models and the estimation algorithms. The AIC and BIC values for each estimated model included in Table 5 show that the Sarmanov based models provide the best fit for our data set.

Focusing on the estimated bivariate Sarmanov distributions that are shown in

Table 4, based on the AIC and BIC values, we note that the best fit is obtained with the CZINBG, although the difference from the CNBG model is minimal. In both cases, we obtain a positive and statistically significant positive dependence between the frequency and average severity of claims. Furthermore, the dependence parameter is within the interval defined in (5), which indicates that the estimated Sarmanov models work. The effect of this dependence on risk premium is analyzed below.

Table 4: Estimation results of bivariate Sarmanov distributions for CNBG and CZINBG models

	CNBG	CZINBG
r	0.2994	11.1291
p	0.7703	0.9695
π	0.0000	0.7453
α	0.2783	0.2756
β	0.0004	0.0004
δ	1.0519	1.1180
γ	0.6806	0.6970
ω	2.0863*	2.4814*
Min(ω)	-24.9979	-27.1313
Max(ω)	3.67676	4.0042114
$corr(X,N)$	0.4159	0.4208
AIC	157,508.0	157,442.6
BIC	157,574.6	157,518.7

*Statistically significant positive dependence at 99% confidence level.

Table 5: Comparing bivariate models.

	CNBG		CZINBG	
	AIC	BIC	AIC	BIC
Sarmanov	157,508.0	157,574.6	157,442.6	157,518.7
Gaussian Copula	157,654.2	157,688.7	157,571.1	157,612.6
Garrido et al.	157,844.7	157,892.2	157,791.0	157,838.5

4.1. Effect on pure and risk premiums

In insurance, the pure premium is calculated as the expected cost of the reported claims, i.e. $\mathbb{E}S = \mathbb{E}[NX]$ in our case, while the risk premium commonly consists of adding the effect of the dispersion of this variable, i.e. $VarS = Var[NX]$. For example, if we use the standard deviation criterion, we obtain the risk premium formula $\rho_R = \mathbb{E}S + \eta\sqrt{VarS}$, where $\eta > 0$ is a loading constant. Therefore, for calculating this premium, we need to know the distribution of S and especially its first two moments. For our numerical example, we present in Table 6 the total pure and risk premiums evaluated for the $K = 99,972$ policyholders in two cases: if $N > 0$ and $X > 0$ were independent (i.e., $\omega = 0$), and by assuming that $N > 0$ and $X > 0$ are Sarmanov distributed with $\omega > 0$ and with $\mathbb{E}S$ and $VarS$ calculated as in Proposition 2. We used the models whose parameters are shown in Table 4 and assumed $\eta = 0$ (pure premium) and $\eta = 1$. If we compare the evaluated premiums without and with dependence, we can observe the effect of the dependence: the dependence between frequency and severity leads to an increase in premiums that could improve the company solvency, reducing hence the ruin probability.

Table 6: Premiums obtained with CNBG and CZINBG models using $\omega = 0$ and $\omega > 0$, for $K = 99,972$ policyholders.

	$\eta = 0$		$\eta = 1$	
	CNBG	CZINBG	CNBG	CZINBG
ρ_R with $\omega = 0$	6,209,898	6,142,407	58,304,175	57,789,996
ρ_R with $\omega > 0$	6,266,396	6,200,767	58,978,497	58,038,198
Difference	56,498	58,360	674,322	248,202

5. Conclusions

In this paper, we have shown how Sarmanov distribution allows us to mix continuous and discrete marginal distributions and to model their dependence. Specifically, we have obtained four bivariate particular cases where we assumed the Gamma distribution for the continuous marginal, and Poisson, Zero Inflated Poisson, Negative Binomial and, respectively, Zero Inflated Negative Binomial distribution for the discrete marginal. Furthermore, a two part maximum likelihood estimation method was proposed and evaluated using a simulation study. We concluded that our proposed method is consistent in terms of the considered error metrics of the estimated parameters for the four proposed particular cases.

As a direct application, we used our model to introduce dependence between the frequency and severity of claims in the collective model. We numerically illustrated this on an auto insurance data set, for which we obtained low, but significant positive dependence between frequency and severity. We concluded that with our model, this dependence between frequency and severity can lead to changes in premiums that could improve the company's performance.

In a further work, we intend to also consider other distributions for the claim frequency and severity, such as mixture distributions, which are challenging in what concerns parameters estimation. Also, introducing regression components is another aspect that we take into account, as well as a Bayesian approach.

Appendix

Proofs

The following lemmas will be needed to prove Proposition 4; although the first lemma is given for the continuous r.v. Y, it holds for any r.v., including a discrete r.v. N, assuming that the involved expected values exist. The proof of this lemma is immediate, hence we omit it.

Lemma 1 *Let Y be some r.v. and let $\psi(x,\delta) = e^{-\delta x} - \mathscr{L}_Y(\delta)$ be the corresponding exponential kernel. Then*

$$\mathbb{E}\left[Y\psi(Y,\delta)\right] = \mathbb{E}\left[Ye^{-\delta Y}\right] - \mathscr{L}_Y(\delta)\,\mathbb{E}\left[Y\right], \tag{9}$$

$$\mathbb{E}\left[Y^2\psi(Y,\delta)\right] = \mathbb{E}\left[Y^2e^{-\delta Y}\right] - \mathscr{L}_Y(\delta)\,\mathbb{E}\left[Y^2\right]. \tag{10}$$

Lemma 2 *If the r.v. N follows a certain discrete distribution with support $\mathbb{N}$ and $\tilde{N}$ follows the same distribution in the zero inflated form with parameter $\pi \in (0,1)$, then*

$$\mathbb{E}\left[\tilde{N}\psi\left(\tilde{N},\delta\right)\right] = (1-\pi)\,\mathbb{E}\left[N\psi(N,\delta)\right],$$

$$\mathbb{E}\left[\tilde{N}^2\psi\left(\tilde{N},\delta\right)\right] = (1-\pi)\,\mathbb{E}\left[N^2\psi(N,\delta)\right],$$

where $\psi(N,\delta) = e^{-\delta N} - \frac{\mathscr{L}_N(\delta)-p(0)}{1-p(0)}$ and $\psi\left(\tilde{N},\delta\right) = e^{-\delta\tilde{N}} - \frac{\mathscr{L}_{\tilde{N}}(\delta)-\tilde{p}(0)}{1-\tilde{p}(0)}, \tilde{p}(0) =$ $\Pr\left(\tilde{N}=0\right)$.

Proof of Lemma 2. The first formula easily results by applying formula (9),

$$\begin{aligned}
\mathbb{E}\left[\tilde{N}\psi\left(\tilde{N},\delta\right)\right] &= \mathbb{E}\left[\tilde{N}e^{-\delta\tilde{N}}\right]-\frac{\mathscr{L}_{\tilde{N}}(\delta)-\tilde{p}(0)}{1-\tilde{p}(0)}\mathbb{E}\tilde{N}=(1-\pi)\sum_{n\geq 1}ne^{-\delta n}p(n)\\
&\quad -\frac{\pi+(1-\pi)\mathscr{L}_N(\delta)-\pi-(1-\pi)p(0)}{1-\pi-(1-\pi)p(0)}(1-\pi)\mathbb{E}N\\
&= (1-\pi)\left(\mathbb{E}\left[Ne^{-\delta N}\right]-\frac{\mathscr{L}_N(\delta)-p(0)}{1-p(0)}\mathbb{E}N\right)\\
&= (1-\pi)\mathbb{E}\left[N\psi(N,\delta)\right].
\end{aligned}$$

The proof of the second formula is similar, based on formula (10). $\square$

Proof of Proposition 4. *i)* We start by proving the case $\pi=0$. When $N\sim Po(\lambda)$, from the proof of Lemma 4.1 in Tamraz and Vernic (2018) we know that $\mathbb{E}\left[Ne^{-\delta N}\right]=\lambda e^{\lambda\left(e^{-\delta}-1\right)-\delta}$, hence, applying also formula (9),

$$\mathbb{E}\left[N\psi(N,\delta)\right]=\lambda e^{\lambda\left(e^{-\delta}-1\right)-\delta}-\lambda\frac{e^{\lambda\left(e^{-\delta}-1\right)}-e^{-\lambda}}{1-e^{-\lambda}}=\lambda e^{-\lambda}\left(e^{\lambda e^{-\delta}-\delta}-\frac{e^{\lambda e^{-\delta}}-1}{1-e^{-\lambda}}\right).$$

For the second formula, we use

$$\begin{aligned}
\mathbb{E}\left[N^2e^{-\delta N}\right] &= e^{-\lambda}\sum_{n=0}^{\infty}\frac{n^2\lambda^n}{n!}e^{-\delta n}=e^{-\lambda}\sum_{n=1}^{\infty}\frac{(n-1+1)\left(\lambda e^{-\delta}\right)^n}{(n-1)!}\\
&= e^{-\lambda}\left[\left(\lambda e^{-\delta}\right)^2\sum_{n=2}^{\infty}\frac{\left(\lambda e^{-\delta}\right)^{n-2}}{(n-2)!}+\lambda e^{-\delta}\sum_{n=1}^{\infty}\frac{\left(\lambda e^{-\delta}\right)^{n-1}}{(n-1)!}\right]\\
&= e^{-\lambda}\left(\left(\lambda e^{-\delta}\right)^2e^{\lambda e^{-\delta}}+\lambda e^{-\delta}e^{\lambda e^{-\delta}}\right)=\lambda e^{\lambda e^{-\delta}-\lambda-\delta}\left(\lambda e^{-\delta}+1\right),
\end{aligned}$$

that we insert into (10) and obtain

$$\begin{aligned}
\mathbb{E}\left[N^2\psi(N,\delta)\right] &= \lambda e^{\lambda e^{-\delta}-\lambda-\delta}\left(\lambda e^{-\delta}+1\right)-\lambda(\lambda+1)\frac{e^{\lambda\left(e^{-\delta}-1\right)}-e^{-\lambda}}{1-e^{-\lambda}}\\
&= \lambda e^{-\lambda}\left[e^{\lambda e^{-\delta}-\delta}\left(\lambda e^{-\delta}+1\right)-(\lambda+1)\frac{e^{\lambda e^{-\delta}}-1}{1-e^{-\lambda}}\right].
\end{aligned}$$

The formulas for $N \sim ZIP(\lambda, \pi)$ easily result by applying Lemma 2.

ii) We first prove the case $\pi = 0$. For $N \sim NB(r, p)$, from the proof of Lemma 4.1 from Tamraz and Vernic (2018) we have that $\mathbb{E}\left[Ne^{-\delta N}\right] = \frac{rqp^r e^{-\delta}}{\left(1-qe^{-\delta}\right)^{r+1}}$. Then, based on formula (9),

$$\begin{aligned}
\mathbb{E}[N\psi(N,\delta)] &= \frac{rqp^r e^{-\delta}}{\left(1-qe^{-\delta}\right)^{r+1}} - \frac{rq}{p}\frac{\left(\frac{p}{1-qe^{-\delta}}\right)^r - p^r}{1-p^r} \\
&= \frac{rqp^r}{\left(1-qe^{-\delta}\right)^r}\left(\frac{e^{-\delta}}{1-qe^{-\delta}} - \frac{1-\left(1-qe^{-\delta}\right)^r}{p(1-p^r)}\right),
\end{aligned}$$

yielding the first formula. To obtain the second stated formula, we first evaluate

$$\begin{aligned}
\mathbb{E}\left[N^2 e^{-\delta N}\right] &= \sum_{n=0}^{\infty}\frac{\Gamma(r+n)}{n!\Gamma(r)}n^2 p^r\left(qe^{-\delta}\right)^n = \sum_{n=1}^{\infty}\frac{\Gamma(r+n)(n-1+1)}{(n-1)!\Gamma(r)}p^r\left(qe^{-\delta}\right)^n \\
&= \frac{p^r}{\left(1-qe^{-\delta}\right)^r}\left[\sum_{n=2}^{\infty}\frac{\Gamma(r+n)}{(n-2)!\Gamma(r)}\left(1-qe^{-\delta}\right)^r\left(qe^{-\delta}\right)^n\right. \\
&\quad \left. + \sum_{n=1}^{\infty}\frac{\Gamma(r+n)}{(n-1)!\Gamma(r)}\left(1-qe^{-\delta}\right)^r\left(qe^{-\delta}\right)^n\right] \\
&= \frac{p^r}{\left(1-qe^{-\delta}\right)^r}\left[\frac{r(r+1)\left(qe^{-\delta}\right)^2}{\left(1-qe^{-\delta}\right)^2} + \frac{rqe^{-\delta}}{1-qe^{-\delta}}\right] \\
&= \frac{rqp^r e^{-\delta}\left(rqe^{-\delta}+1\right)}{\left(1-qe^{-\delta}\right)^{r+2}}.
\end{aligned}$$

Therefore, based on (10), we have

$$\begin{aligned}
\mathbb{E}\left[N^2\psi(N,\delta)\right] &= \frac{rqp^r e^{-\delta}\left(rqe^{-\delta}+1\right)}{\left(1-qe^{-\delta}\right)^{r+2}} - \frac{rq(1+qr)}{p^2}\frac{\left(\frac{p}{1-qe^{-\delta}}\right)^r - p^r}{1-p^r} \\
&= \frac{rqp^r}{\left(1-qe^{-\delta}\right)^r}\left[\frac{e^{-\delta}\left(rqe^{-\delta}+1\right)}{\left(1-qe^{-\delta}\right)^2} - \frac{1+qr}{p^2}\frac{1-\left(1-qe^{-\delta}\right)^r}{1-p^r}\right],
\end{aligned}$$

which easily yields the stated formula.

The general case $N \sim ZINB(r, p, \pi)$ follows from Lemma 2, which completes the proof. □

Proof of Proposition 5. We start with

$$\mathbb{E}\left[Ye^{-\gamma Y}\right] = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_{0}^{\infty} y^{\alpha+1-1} e^{-(\beta+\gamma)y} dy = \frac{\alpha\beta^{\alpha}}{(\beta+\gamma)^{\alpha+1}},$$

that we insert into (9) and obtain

$$\mathbb{E}\left[Y\phi(Y,\gamma)\right] = \frac{\alpha\beta^{\alpha}}{(\beta+\gamma)^{\alpha+1}} - \frac{\alpha}{\beta}\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha},$$

hence the first stated formula.

Also,

$$\mathbb{E}\left[Y^{2}e^{-\gamma Y}\right] = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_{0}^{\infty} y^{\alpha+2-1} e^{-(\beta+\gamma)y} dy = \frac{\alpha(\alpha+1)\beta^{\alpha}}{(\beta+\gamma)^{\alpha+2}},$$

hence, according to (10),

$$\mathbb{E}\left[Y^{2}\phi(Y,\gamma)\right] = \frac{\alpha(\alpha+1)\beta^{\alpha}}{(\beta+\gamma)^{\alpha+2}} - \frac{\alpha(\alpha+1)}{\beta^{2}}\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha},$$

from where we easily obtain the second stated formula. □

Alternative Models

Based on the idea of Garrido et al. (2016), in the same context of this work and using our notation, the dependence between frequency and severity can be modeled by adding to the severity model the number of claims as an explanatory variable; i.e., the r.v. N follows a counting distribution between those defined in Section 2.4.1, while the r.v. Y, defined only for $N > 0$, is specified as a Generalized Linear Model (GLM), where the following parameterization of the Gamma distribution is considered

$$E(Y|N) = \mu = \frac{\alpha}{\beta} \Rightarrow \beta = \frac{\alpha}{\mu}.$$

Therefore, the Gamma pdf is

$$f_{Y|N}(y|n) = \frac{1}{\Gamma(\alpha)} \left(\frac{\alpha}{\mu}\right) \left(\frac{\alpha}{\mu} y\right)^{\alpha-1} e^{-\frac{\alpha}{\mu} y},$$

where $\mu = e^{\lambda N}$ and λ is the parameter that induces a degree of dependence between the number of claims and the average severity. The parameters are estimated by maximizing the joint log-likelihood function

$$\begin{aligned} \ln L\left((n_i, y_i)_{i=1}^{K}; \theta; \nu; \alpha; \lambda\right) &= \sum_{\{i: n_i = y_i = 0\}} \ln p(0;\theta) + \sum_{\{i: n_i \geq 1, y_i > 0\}} [\ln p(n_i;\theta) \\ &\quad + \ln f_{Y_i|N_i}(y_i|n_i)], \end{aligned} \tag{11}$$

where θ and ν are, respectively, the parameters vectors of the marginal distribution of N and of the average cost per policyholder Y. Garrido et al. (2016) proposed an estimation procedure for the Poisson-Gamma particular case. For alternative counting distributions such as the Negative Binomial and the Zero Inflated models, we have maximized the joint log-likelihood function by using the `optim()` function of `R` with the method `L-BFGS-B`. The initial parameters were obtained from the independent case. The optimization procedure is iterated until an optimum is reached. To check the optimal result, we considered different bounds for the method `L-BFGS-B`.

Based on Czado et al. (2012), in the same context of this work and using our notation, we considered the following Copula model

$$F_{X,N}(x,n|\nu;\theta;\rho) = C(u_1,u_2|\rho) = \Phi_2\left[\Phi^{-1}(u_1), \Phi^{-1}(u_2)\middle|\rho\right], \tag{12}$$

where ρ is the dependency parameter, $u_1 = F_X(x|\nu)$ is the cdf of the average severity r.v. and $u_2 = Pr(N \leq n|\theta)$ is the cdf of the counting variable. The conditional likelihood based on the conditional random variable $N|N > 0$ is maximized in two parts (see Czado et al., 2012, for expressions): the first part is associated with the marginal distribution and the second part with the dependence structure. These authors proposed an estimation procedure for the Poisson-Gamma particular case. For our alternative counting distributions (Negative Binomial and the Zero Inflated models), we used a procedure similar to the one described in Section 2.3, given that the likelihood has a similar decomposition.

Simulation Results Tables

Acknowledgments

The authors gratefully acknowledge the anonymous referees for their valuable remarks and comments, which significantly contributed to the paper improvement.

Catalina Bolancé gratefully acknowledge support from the BBVA Foundation and the Spanish Ministry of Science and Innovation grant PID2019-105986GB-C21.

References

Abdallah, A., Boucher, J., Cossette, H., 2016. Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. Insurance: Mathematics and Economics 68, 120–133.

Bahraoui, Z., Bolancé, C., Pelican, E., Vernic, R., 2015. On the bivariate distribution and copula. an application on insurance data using truncated marginal distributions. Statistics and Operations Research Transactions, SORT 39, 209–230.

Table 7: Results of lower (LCI) and upper (UCI) EBCIs (1000 bootstrap resamples) of RMSRE and MAPE for compound Poisson-Gamma distributions (CPG).

			Poisson		Gamma				Dependece					
			λ		α		β		δ		γ		ω	
			LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI
CPG	K								$\gamma = \delta = 1$					
M1.1	500	RMSRE	0.094	0.104	0.125	0.139	0.237	0.268	0.124	0.138	0.195	0.214	1.019	1.121
		MAPE	0.074	0.082	0.096	0.107	0.182	0.203	0.083	0.094	0.150	0.166	0.841	0.922
	5000	RMSRE	0.029	0.032	0.041	0.044	0.076	0.084	0.059	0.066	0.074	0.082	0.765	0.803
		MAPE	0.024	0.026	0.032	0.035	0.059	0.065	0.039	0.044	0.054	0.060	0.684	0.727
M2.1	500	RMSRE	0.094	0.103	0.124	0.138	0.236	0.266	0.147	0.159	0.170	0.186	1.344	1.635
		MAPE	0.073	0.081	0.095	0.106	0.182	0.204	0.107	0.119	0.129	0.143	0.700	0.856
	5000	RMSRE	0.030	0.032	0.040	0.044	0.077	0.084	0.059	0.066	0.069	0.077	0.874	0.948
		MAPE	0.024	0.026	0.032	0.035	0.059	0.066	0.040	0.045	0.050	0.056	0.682	0.749
M3.1	500	RMSRE	0.137	0.150	0.182	0.206	0.371	0.426	0.144	0.157	0.203	0.221	1.627	1.915
		MAPE	0.108	0.119	0.136	0.152	0.277	0.309	0.102	0.115	0.153	0.170	1.401	1.523
	5000	RMSRE	0.044	0.048	0.054	0.059	0.105	0.116	0.057	0.068	0.096	0.108	0.789	0.843
		MAPE	0.035	0.039	0.043	0.047	0.081	0.089	0.032	0.039	0.069	0.079	0.665	0.720
M4.1	500	RMSRE	0.137	0.150	0.182	0.205	0.370	0.430	0.152	0.164	0.192	0.211	2.654	3.393
		MAPE	0.108	0.119	0.135	0.152	0.279	0.311	0.115	0.128	0.143	0.160	1.037	1.382
	5000	RMSRE	0.044	0.047	0.054	0.059	0.105	0.115	0.065	0.076	0.091	0.103	1.055	1.188
		MAPE	0.035	0.039	0.043	0.047	0.080	0.089	0.038	0.045	0.064	0.072	0.679	0.785
									$\gamma = \delta = 2$					
M1.2	500	RMSRE	0.299	0.357	0.353	0.402	0.409	0.455	0.288	0.349	0.290	0.350	1.885	2.108
		MAPE	0.152	0.188	0.233	0.268	0.296	0.333	0.113	0.150	0.128	0.163	1.552	1.686
	5000	RMSRE	0.033	0.036	0.056	0.061	0.082	0.090	0.010	0.014	0.031	0.036	0.919	0.983
		MAPE	0.026	0.029	0.044	0.048	0.064	0.070	0.005	0.006	0.018	0.022	0.769	0.831
M2.2	500	RMSRE	0.094	0.103	0.132	0.146	0.239	0.271	0.139	0.151	0.151	0.164	3.050	3.667
		MAPE	0.073	0.081	0.100	0.112	0.185	0.206	0.100	0.113	0.114	0.126	2.129	2.429
	5000	RMSRE	0.030	0.032	0.040	0.043	0.076	0.083	0.020	0.027	0.052	0.060	1.395	1.548
		MAPE	0.024	0.026	0.032	0.035	0.059	0.065	0.011	0.014	0.033	0.038	1.099	1.213
M3.2	500	RMSRE	0.138	0.151	0.311	0.350	0.481	0.552	0.070	0.074	0.056	0.060	3.834	4.594
		MAPE	0.109	0.119	0.234	0.261	0.355	0.398	0.055	0.060	0.038	0.043	2.854	3.203
	5000	RMSRE	0.045	0.050	0.078	0.086	0.116	0.128	0.020	0.026	0.037	0.042	1.241	1.344
		MAPE	0.037	0.040	0.059	0.065	0.090	0.099	0.008	0.011	0.023	0.027	1.014	1.105
M4.2	500	RMSRE	0.138	0.151	0.182	0.205	0.370	0.427	0.194	0.210	0.135	0.154	6.365	8.093
		MAPE	0.109	0.120	0.135	0.152	0.278	0.310	0.149	0.165	0.086	0.099	3.137	3.934
	5000	RMSRE	0.044	0.047	0.054	0.059	0.105	0.115	0.051	0.062	0.061	0.073	2.056	2.319
		MAPE	0.035	0.038	0.042	0.047	0.080	0.089	0.022	0.029	0.037	0.044	1.535	1.715

Table 8: Results of lower (LCI) and upper (UCI) EBCIs (1000 bootstrap resamples) of RMSRE and MAPE for compound Negative Binomial-Gamma distributions (CNBG).

			NB				Gamma				Dependece					
			r		p		α		β		δ		γ		ω	
			LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI
CNBG	K										$\gamma=\delta=1$					
M1.1	500	RMSRE	0.438	0.550	0.136	0.148	0.314	0.350	0.451	0.523	0.022	0.023	0.023	0.023	0.776	0.839
		MAPE	0.304	0.347	0.108	0.119	0.242	0.268	0.331	0.373	0.021	0.022	0.022	0.023	0.631	0.689
	5000	RMSRE	0.096	0.105	0.040	0.044	0.092	0.101	0.115	0.126	0.016	0.017	0.017	0.018	0.692	0.725
		MAPE	0.075	0.083	0.032	0.036	0.071	0.079	0.090	0.099	0.013	0.014	0.014	0.015	0.617	0.658
M2.1	500	RMSRE	0.414	0.487	0.133	0.145	0.311	0.347	0.444	0.507	0.021	0.022	0.021	0.022	1.012	1.165
		MAPE	0.298	0.338	0.105	0.116	0.240	0.265	0.322	0.362	0.019	0.021	0.019	0.020	0.649	0.750
	5000	RMSRE	0.097	0.107	0.042	0.045	0.093	0.102	0.115	0.127	0.017	0.018	0.017	0.018	0.805	0.857
		MAPE	0.075	0.083	0.033	0.036	0.072	0.080	0.089	0.099	0.013	0.014	0.014	0.015	0.688	0.739
M3.1	500	RMSRE	0.731	1.110	0.174	0.193	0.420	0.471	0.700	0.808	0.041	0.042	0.041	0.042	0.586	0.651
		MAPE	0.444	0.542	0.137	0.152	0.321	0.357	0.482	0.549	0.036	0.038	0.036	0.039	0.445	0.497
	5000	RMSRE	0.136	0.153	0.057	0.062	0.118	0.131	0.158	0.175	0.026	0.029	0.029	0.031	0.559	0.599
		MAPE	0.105	0.117	0.045	0.050	0.092	0.102	0.123	0.136	0.020	0.022	0.022	0.025	0.446	0.489
M4.1	500	RMSRE	0.744	1.025	0.181	0.197	0.471	0.521	0.711	0.823	0.028	0.029	0.028	0.029	1.686	2.004
		MAPE	0.457	0.546	0.143	0.157	0.358	0.396	0.490	0.560	0.025	0.027	0.025	0.027	0.945	1.136
	5000	RMSRE	0.122	0.135	0.054	0.060	0.118	0.131	0.153	0.171	0.023	0.024	0.023	0.025	0.730	0.799
		MAPE	0.097	0.107	0.043	0.048	0.090	0.100	0.120	0.132	0.018	0.020	0.019	0.021	0.557	0.617
											$\gamma=\delta=2$					
M1.2	500	RMSRE	0.437	0.546	0.136	0.148	0.307	0.343	0.440	0.513	0.021	0.022	0.025	0.026	0.908	0.982
		MAPE	0.302	0.346	0.108	0.119	0.238	0.263	0.324	0.366	0.016	0.018	0.021	0.023	0.726	0.795
	5000	RMSRE	0.096	0.105	0.041	0.044	0.090	0.099	0.114	0.125	0.011	0.013	0.015	0.017	0.705	0.747
		MAPE	0.075	0.083	0.033	0.036	0.071	0.078	0.089	0.098	0.007	0.009	0.011	0.012	0.610	0.655
M2.2	500	RMSRE	0.480	0.628	0.174	0.193	0.427	0.479	0.699	0.804	0.026	0.027	0.027	0.028	3.762	4.392
		MAPE	0.426	0.465	0.138	0.152	0.326	0.363	0.489	0.555	0.022	0.023	0.023	0.025	2.810	3.154
	5000	RMSRE	0.096	0.106	0.041	0.045	0.090	0.099	0.114	0.125	0.011	0.013	0.014	0.015	1.055	1.155
		MAPE	0.076	0.083	0.033	0.036	0.071	0.077	0.090	0.099	0.007	0.008	0.009	0.010	0.832	0.916
M3.2	500	RMSRE	0.733	1.132	0.175	0.193	0.429	0.481	0.707	0.812	0.025	0.026	0.027	0.028	1.314	1.460
		MAPE	0.445	0.544	0.138	0.153	0.329	0.367	0.490	0.557	0.021	0.022	0.023	0.025	1.061	1.162
	5000	RMSRE	0.136	0.153	0.058	0.063	0.120	0.132	0.159	0.177	0.016	0.018	0.024	0.026	0.644	0.694
		MAPE	0.106	0.118	0.046	0.051	0.093	0.103	0.125	0.138	0.010	0.012	0.016	0.018	0.522	0.569
M4.2	500	RMSRE	0.729	1.130	0.174	0.193	0.427	0.479	0.699	0.804	0.026	0.027	0.027	0.028	3.762	4.392
		MAPE	0.444	0.543	0.138	0.152	0.326	0.363	0.489	0.555	0.022	0.023	0.023	0.025	2.810	3.154
	5000	RMSRE	0.137	0.154	0.058	0.064	0.125	0.138	0.164	0.182	0.014	0.016	0.017	0.019	1.351	1.469
		MAPE	0.107	0.118	0.046	0.051	0.097	0.108	0.128	0.142	0.009	0.011	0.012	0.013	1.082	1.188

Table 9: Results of lower (LCI) and upper (UCI) EBCIs (1000 bootstrap resamples) of RMSRE and MAPE for compound Zero-Inflated-Poisson-Gamma distributions (CZIPG).

			ZI-Poisson				Gamma				Dependece					
			λ		π		α		β		δ		γ		ω	
			LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI
CZIPG	K										$\gamma = \delta = 1$					
M1.1	500	RMSRE	0.313	0.342	0.311	0.346	0.201	0.224	0.311	0.352	0.134	0.137	0.142	0.144	0.845	0.910
		MAPE	0.247	0.271	0.230	0.258	0.155	0.171	0.230	0.257	0.126	0.132	0.137	0.142	0.698	0.757
	5000	RMSRE	0.179	0.190	0.157	0.166	0.047	0.052	0.082	0.090	0.131	0.134	0.140	0.142	0.660	0.690
		MAPE	0.156	0.167	0.144	0.152	0.036	0.040	0.065	0.072	0.124	0.129	0.135	0.139	0.593	0.628
M2.1	500	RMSRE	0.534	0.596	0.515	0.557	0.242	0.276	0.439	0.510	0.199	0.205	0.218	0.224	1.804	2.176
		MAPE	0.410	0.455	0.396	0.438	0.175	0.198	0.318	0.360	0.186	0.195	0.210	0.217	0.823	1.034
	5000	RMSRE	0.287	0.298	0.232	0.239	0.047	0.051	0.086	0.096	0.207	0.215	0.219	0.226	0.782	0.842
		MAPE	0.271	0.283	0.227	0.233	0.037	0.041	0.066	0.073	0.192	0.202	0.205	0.215	0.631	0.689
M3.1	500	RMSRE	0.508	0.563	0.516	0.559	0.303	0.344	0.493	0.581	0.120	0.125	0.140	0.142	0.927	1.044
		MAPE	0.401	0.441	0.388	0.430	0.221	0.248	0.350	0.401	0.109	0.115	0.135	0.139	0.752	0.828
	5000	RMSRE	0.190	0.206	0.186	0.205	0.082	0.092	0.122	0.134	0.086	0.089	0.095	0.096	0.631	0.669
		MAPE	0.152	0.167	0.143	0.157	0.062	0.069	0.096	0.106	0.081	0.085	0.092	0.095	0.533	0.573
M4.1	500	RMSRE	0.357	0.386	0.332	0.360	0.154	0.172	0.274	0.313	0.204	0.210	0.219	0.225	1.129	1.350
		MAPE	0.287	0.312	0.280	0.302	0.117	0.130	0.205	0.230	0.193	0.201	0.207	0.216	0.626	0.752
	5000	RMSRE	0.320	0.342	0.264	0.280	0.058	0.064	0.112	0.122	0.188	0.196	0.216	0.223	0.907	1.030
		MAPE	0.273	0.295	0.239	0.253	0.046	0.051	0.088	0.097	0.173	0.182	0.203	0.212	0.628	0.716
											$\gamma = \delta = 2$					
M1.2	500	RMSRE	0.275	0.301	0.285	0.316	0.195	0.218	0.308	0.348	0.088	0.093	0.127	0.131	0.956	1.051
		MAPE	0.216	0.237	0.217	0.239	0.147	0.164	0.228	0.255	0.073	0.079	0.118	0.124	0.761	0.833
	5000	RMSRE	0.175	0.185	0.156	0.163	0.045	0.049	0.081	0.089	0.077	0.082	0.119	0.123	0.741	0.779
		MAPE	0.154	0.164	0.144	0.151	0.035	0.039	0.064	0.071	0.062	0.068	0.107	0.113	0.658	0.699
M2.2	500	RMSRE	0.355	0.383	0.331	0.359	0.153	0.171	0.274	0.312	0.166	0.174	0.195	0.202	2.598	3.044
		MAPE	0.286	0.312	0.280	0.302	0.116	0.129	0.204	0.228	0.142	0.153	0.176	0.186	1.823	2.067
	5000	RMSRE	0.286	0.298	0.232	0.239	0.046	0.050	0.084	0.093	0.110	0.119	0.170	0.179	1.205	1.324
		MAPE	0.271	0.282	0.227	0.233	0.036	0.039	0.066	0.073	0.086	0.095	0.146	0.157	0.935	1.033
M3.2	500	RMSRE	0.468	0.519	0.499	0.540	0.298	0.338	0.488	0.577	0.106	0.111	0.123	0.127	1.798	2.096
		MAPE	0.370	0.408	0.374	0.415	0.217	0.245	0.349	0.400	0.091	0.097	0.114	0.119	1.422	1.573
	5000	RMSRE	0.221	0.240	0.200	0.216	0.066	0.075	0.111	0.123	0.062	0.068	0.121	0.125	0.776	0.829
		MAPE	0.179	0.196	0.163	0.178	0.051	0.056	0.088	0.097	0.048	0.053	0.110	0.116	0.634	0.688
M4.2	500	RMSRE	0.534	0.594	0.511	0.552	0.241	0.275	0.436	0.508	0.190	0.197	0.180	0.188	4.572	5.824
		MAPE	0.414	0.457	0.392	0.432	0.175	0.198	0.315	0.358	0.173	0.182	0.163	0.173	2.532	3.064
	5000	RMSRE	0.319	0.340	0.266	0.281	0.058	0.065	0.110	0.121	0.095	0.105	0.170	0.179	1.641	1.847
		MAPE	0.271	0.292	0.240	0.254	0.045	0.050	0.087	0.096	0.069	0.077	0.148	0.158	1.248	1.395

Table 10: Results of lower (LCI) and upper (UCI) EBCIs (1000 bootstrap resamples) of compound Zero-Inflated-Negative Binomial-Gamma distributions (CZINBG).

			ZI-NB						Gamma				Dependece					
			r		p		π		α		β		δ		γ		ω	
			LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI
CZINBG	K												$\gamma = \delta = 1$					
M1.1	500	RMSRE	0.098	0.104	0.101	0.107	0.106	0.111	0.261	0.294	0.417	0.484	0.121	0.124	0.127	0.130	0.669	0.746
		MAPE	0.083	0.090	0.088	0.094	0.094	0.101	0.190	0.215	0.302	0.341	0.113	0.118	0.120	0.125	0.508	0.563
	5000	RMSRE	0.038	0.041	0.041	0.043	0.044	0.047	0.079	0.088	0.108	0.119	0.052	0.054	0.055	0.057	0.589	0.622
		MAPE	0.031	0.033	0.034	0.037	0.038	0.040	0.060	0.067	0.085	0.093	0.048	0.050	0.051	0.054	0.492	0.528
M2.1	500	RMSRE	0.097	0.102	0.100	0.106	0.105	0.110	0.258	0.294	0.421	0.489	0.122	0.126	0.122	0.126	0.936	1.095
		MAPE	0.083	0.089	0.086	0.093	0.094	0.100	0.186	0.210	0.303	0.343	0.115	0.120	0.114	0.119	0.591	0.690
	5000	RMSRE	0.045	0.050	0.042	0.046	0.048	0.052	0.068	0.075	0.101	0.111	0.061	0.065	0.062	0.067	0.731	0.783
		MAPE	0.032	0.036	0.033	0.036	0.037	0.041	0.052	0.058	0.079	0.087	0.051	0.056	0.053	0.057	0.609	0.659
M3.1	500	RMSRE	0.112	0.117	0.103	0.108	0.120	0.124	0.600	0.706	1.131	1.611	0.125	0.129	0.131	0.134	0.978	1.032
		MAPE	0.099	0.105	0.090	0.096	0.109	0.115	0.407	0.467	0.707	0.839	0.118	0.123	0.124	0.129	0.850	0.911
	5000	RMSRE	0.070	0.076	0.059	0.064	0.075	0.081	0.135	0.149	0.201	0.223	0.083	0.088	0.088	0.093	0.513	0.560
		MAPE	0.055	0.061	0.047	0.052	0.061	0.066	0.104	0.115	0.154	0.170	0.072	0.077	0.077	0.082	0.402	0.443
M4.1	500	RMSRE	0.113	0.118	0.102	0.107	0.120	0.124	0.576	0.673	1.104	1.486	0.124	0.127	0.126	0.129	2.105	2.453
		MAPE	0.100	0.106	0.088	0.094	0.110	0.116	0.396	0.453	0.703	0.825	0.117	0.122	0.118	0.124	0.991	1.233
	5000	RMSRE	0.073	0.079	0.059	0.064	0.074	0.080	0.136	0.151	0.202	0.226	0.084	0.089	0.084	0.089	0.699	0.790
		MAPE	0.057	0.063	0.047	0.052	0.059	0.065	0.103	0.114	0.154	0.171	0.072	0.077	0.073	0.078	0.459	0.528
													$\gamma = \delta = 2$					
M1.2	500	RMSRE	0.098	0.103	0.100	0.105	0.106	0.110	0.255	0.290	0.417	0.484	0.092	0.097	0.110	0.115	1.016	1.094
		MAPE	0.084	0.090	0.086	0.093	0.093	0.099	0.186	0.210	0.301	0.340	0.078	0.084	0.097	0.104	0.830	0.903
	5000	RMSRE	0.046	0.052	0.043	0.047	0.047	0.051	0.069	0.076	0.102	0.113	0.038	0.042	0.052	0.056	0.715	0.765
		MAPE	0.033	0.038	0.034	0.037	0.036	0.040	0.053	0.059	0.080	0.089	0.028	0.031	0.040	0.044	0.596	0.648
M2.2	500	RMSRE	0.097	0.102	0.100	0.105	0.105	0.110	0.256	0.291	0.419	0.487	0.098	0.103	0.109	0.114	2.627	2.887
		MAPE	0.083	0.089	0.086	0.092	0.093	0.099	0.186	0.210	0.303	0.342	0.085	0.091	0.095	0.102	2.101	2.315
	5000	RMSRE	0.046	0.052	0.043	0.046	0.048	0.052	0.068	0.075	0.101	0.111	0.036	0.040	0.047	0.051	1.089	1.197
		MAPE	0.033	0.038	0.033	0.037	0.037	0.041	0.052	0.058	0.080	0.088	0.025	0.028	0.034	0.038	0.864	0.952
M3.2	500	RMSRE	0.113	0.119	0.103	0.108	0.120	0.124	0.586	0.686	1.113	1.575	0.115	0.119	0.119	0.123	1.969	2.069
		MAPE	0.099	0.106	0.090	0.096	0.110	0.116	0.398	0.458	0.705	0.830	0.105	0.111	0.108	0.114	1.771	1.879
	5000	RMSRE	0.072	0.078	0.056	0.061	0.072	0.077	0.135	0.149	0.200	0.222	0.050	0.055	0.071	0.076	0.896	0.972
		MAPE	0.057	0.063	0.044	0.049	0.057	0.063	0.103	0.115	0.155	0.171	0.037	0.041	0.056	0.062	0.715	0.782
M4.2	500	RMSRE	0.114	0.119	0.102	0.107	0.120	0.124	0.580	0.681	1.110	1.539	0.120	0.124	0.116	0.120	4.733	5.310
		MAPE	0.100	0.107	0.088	0.094	0.110	0.115	0.397	0.454	0.705	0.829	0.111	0.117	0.104	0.110	3.318	3.768
	5000	RMSRE	0.072	0.078	0.058	0.062	0.074	0.079	0.135	0.150	0.200	0.223	0.052	0.057	0.065	0.070	1.674	1.856
		MAPE	0.057	0.063	0.046	0.050	0.059	0.064	0.102	0.114	0.154	0.171	0.037	0.041	0.048	0.054	1.331	1.464

Bolancé, C., Vernic, R., 2019. Multivariate count data generalized linear models: Three approaches based on the sarmanov distribution. Insurance: Mathematics and Economics 85, 89–103.

Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 16, 1190–1208.

Casella, G., George, E.I., 1992. Explaining the gibbs sampler. The American Statistician 46, 167–174.

Czado, C., Kastenmeier, R., Brechmann, E.C., Min, A., 2012. A mixed copula model for insurance claims and claim sizes. Scandinavian Actuarial Journal 4, 278–305.

Erhardt, V., Czado, C., 2012. Modeling dependent yearly claim totals including zero claims in private health insurance. Scandinavian Actuarial Journal 2, 106–129.

Frees, E., Gao, J., Rosenberg, M., 2011. Predicting the frequency and amount of health care expenditures. North American Actuarial Journal 15, 377–392.

Frees, E.W., Wang, P., 2006. Copula credibility for aggregate loss models. Insurance: Mathematics and Economics 38, 360–373.

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

Gschlößl, S., Czado, C., 2007. Spatial modelling of claim frequency and claim size in non-life insurance. Scandinavian Actuarial Journal 3, 360–373.

Guo, F., Wang, D., Yang, H., 2017. Asymptotic results for ruin probability in a two-dimensional risk model with stochastic investmentreturns. Journal of Computational and Applied Mathematics 325, 198–221.

Hua, L., 2015. Tail negative dependence and its applications for aggregate loss modeling. Insurance: Mathematics and Economics 61, 135–145.

Jeong, H., Valdez, E.A., 2020. Predictive compound risk models with dependence. Insurance: Mathematics and Economics 94, 182–195.

Joe, H., 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analysis 94, 401–419.

Kotz, S., Balakrishnan, N., Johnson, N.L., 2000. Continuous multivariate distributions, Volume 1: Models and applications. second ed., Wiley series in Probability and Statistics, New York.

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

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

McNeil, A.J., 1997. Estimating the tails of loss severity distributions using extreme value theory. ASTIN Bulletin 27, 117–137.

Oh, R., Shi, P., Ahn, J.Y., 2020. Bonus-malus premiums under the dependent frequency-severity modeling. Scandinavian Actuarial Journal 2020, 172–195.

Sarmanov, O., 1966. Generalized normal correlation and two-dimensional frechet classes. Doclady (Soviet Mathematics) 168, 596–599.

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

Ting Lee, M.L., 1996. Properties and applications of the sarmanov family of bivariate distributions. Communications in Statistics-Theory and Methods 25, 1207–1222.

Valdez, E.A., Jeong, H., Ahn, J.Y., Park, S., 2018. Generalized linear mixed models for dependent compound risk models. Variance .

Yang, Y., Yuen, K.C., 2016. Finite-time and infinite-time ruin probabilities in a two-dimensional delayed renewal risk model withsarmanov dependent claims. Journal of Mathematical Analysis and Applications 442, 600–625.