mathematics

MDPI

Article

Bivariate Mixed Poisson and Normal Generalised Linear Models with Sarmanov Dependence—An Application to Model Claim Frequency and Optimal Transformed Average Severity

Ramon Alemany [1,†], Catalina Bolancé [1,*,†], Roberto Rodrigo [1,†] and Raluca Vernic [2,3,†]

[1] Department of Econometrics, Statistics and Applied Economics, Riskcenter-IREA University of Barcelona, Av. Diagonal, 690, 08034 Barcelona, Spain; ralemany@ub.edu (R.A.); rrodrima122@alumnes.ub.edu (R.R.)
[2] Faculty of Mathematics and Computer Science, Ovidius University of Constanta, 124 Mamaia, Constanta, Romania; rvernic@univ-ovidius.ro
[3] Gheorghe Mihoc-Caius Iacob Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, Calea 13 Septembrie 13, 050711 Bucharest, Romania
* Correspondence: bolance@ub.edu; Tel.: +34-9-3402-4320
† These authors contributed equally to this work.

Abstract: The aim of this paper is to introduce dependence between the claim frequency and the average severity of a policyholder or of an insurance portfolio using a bivariate Sarmanov distribution, that allows to join variables of different types and with different distributions, thus being a good candidate for modeling the dependence between the two previously mentioned random variables. To model the claim frequency, a generalized linear model based on a mixed Poisson distribution -like for example, the Negative Binomial (NB), usually works. However, finding a distribution for the claim severity is not that easy. In practice, the Lognormal distribution fits well in many cases. Since the natural logarithm of a Lognormal variable is Normal distributed, this relation is generalised using the Box-Cox transformation to model the average claim severity. Therefore, we propose a bivariate Sarmanov model having as marginals a Negative Binomial and a Normal Generalized Linear Models (GLMs), also depending on the parameters of the Box-Cox transformation. We apply this model to the analysis of the frequency-severity bivariate distribution associated to a pay-as-you-drive motor insurance portfolio with explanatory telematic variables.

Keywords: Box-Cox transformation; dependence; bivariate Sarmanov distribution; motor insurance; telematic data

MSC: 60E05; 62H05; 62H10

check for updates

Citation: Alemany, R.; Bolancé, C.; Rodrigo, R.; Vernic, R. Bivariate Mixed Poisson and Normal Generalised Linear Models with Sarmanov Dependence—An Application to Model Claim Frequency and Optimal Transformed Average Severity. *Mathematics* **2021**, *9*, 73. https://doi.org/10.3390/math9010073

Received: 16 November 2020
Accepted: 25 December 2020
Published: 31 December 2020

1. Introduction

Calculating premiums is a fundamental task for an insurance company. To this purpose, a simple procedure consists of considering the aggregate claims as the product of the random variable (r.v.) number of claims and of the r.v. average cost of these claims, then of fitting appropriate distributions to these two random variables; if, moreover, the premium is evaluated for a given policyholder, some of its characteristics are often included in the calculation as covariates in Generalized Linear Models (GLMs) used for both variables. Then the pure premium is obtained as the product of the means of the number of claims and of the average claim cost, procedure that implies assuming that the two r.v.s are uncorrelated; however, there are evidences in the literature showing how, in practice, the correlation between both variables is not zero (see References [1–4] for some illustrations using GLM). Therefore, in this paper, we shall assume that the two just mentioned r.v.s are dependent and jointly distributed using a bivariate Sarmanov distribution with GLM marginals.

Because it provides a flexible structure in joining different types of marginals, the bivariate Sarmanov distribution has recently found its place in actuarial studies like—modeling continuous claim costs [5], modeling discrete claim frequencies [1,6], modeling dependence between discrete frequency and continuous claims [7], ruin probability calculation [8,9] or modeling telematic variables [10].

The traditional approach to modeling count data by means of Poisson regression failed due to over or under dispersion of the data. Therefore, in practice, count data models based on mixed Poisson distributions are used for the number of claims (or frequency) (see Reference [11]. As an alternative, generalized Poisson regression models have also been considered in the literature (see Reference [12]). Specifically, for this variable, in this paper a GLM based on the Negative Binomial distribution will be used.

To model the claim cost r.v. (severity), Gamma and Lognormal distributions are the most common choices (see Reference [7] for comparing both cost distributions in a real motor data set). However, it may be the case that neither of these two distributions fits the data well enough. It may even be difficult to find a known distribution for the cost of claims. To overcome this problem, in this paper it is assumed that by applying a Box-Cox transformation to the claim cost r.v., the transformed data follow a Normal distribution. In the insurance context, Harrington [13] proposed the Box-Cox transformation to generalize the log-linear model in order to calculate pure premiums (see also Reference [14]).

The Box-Cox transformation [15] provides a family of power transformations that is useful in linear regression when the normality assumption is violated, its aim being to obtain a transformed r.v. that is Normal distributed. Furthermore, the Gamma and Lognormal distributions, which have been traditionally used to adjust the distribution of claims costs, are particular cases of this family; therefore, if the original r.v. follows one of these two distributions, there is an optimal transformation so that the transformed variable becomes Normal. This idea is generalized for some unknown distribution for which we can find an optimal transformed r.v. that follows a Normal distribution. Using a Box-Cox transformation of the distribution of the variable in the original scale can be considered as a generalization of the Lognormal distribution, allowing a longer and heavier right tail as λ_1 decreases, and shifting the mode as λ_2 increases.

To conclude, in this paper it is considered a bivariate Sarmanov distribution that allows us to join a Normal GLM distribution for the transformed average claim cost r.v. (see Reference [16] for Sarmanov with Normal marginal distributions) with a Negative Binomial GLM distribution for the number of claims (see Reference [6] for Sarmanov with Negative Binomial GLM marginals). This model can be used to obtain the distribution of the total cost of claims based on the collective model, for a policyholder with specific characteristics. The bivariate Sarmanov distribution allows to fit a non-linear dependence between frequency and severity, and to analyze if the riskier profiles implies larger dependency. Furthermore, the marginal severity distribution based on a Box-Cox transformation allows a longer and heavier right tail than classical models like Gamma and Lognormal.

As numerical illustration, a database containing information on a portfolio of policyholders that have contracted an auto insurance which involved the installation of a GPS/inertial device in their vehicle is used. This device provides a source of information (telematic variables) to motor insurers that complements the variables that have traditionally been used in a-*priori* pricing in auto insurance. The telematic variables provide an innovative way to calculate car insurance pricing (see References [17–22] where a similar database is considered and the telematic variables are used to predict accident rate). Furthermore, for each policyholder in the database, the number of claims and their mean cost are known; these variables are right skewed and have excess of zeros, characteristics that are common in insurance auto databases.

Traditionally, tariffication in auto insurance has been based on the total number of claims of the policyholders. Given the large number of zeros, the models commonly used for fitting the claim frequency variable are the NB and the Zero Inflated Poisson [22,23]. Alternatively, some works analyze the number of claims using multivariate models, that is, considering that there are different types of claims—with civil liability, with personal injury, and so forth [1,6,24]. More recently, tariffication based on the collective model has been proposed, that is, the premium is deduced from the distribution of the total cost variable, which in turn is deduced from the bivariate distribution of frequency and severity claims. Two alternative approaches have been proposed—on the one hand, the bivariate model is deduced from the frequency distribution and from the severity distribution conditioned to frequency defined for the average claim size distribution using the number of claims as covariate [2,4]. On the other hand, Czado et al. [25] proposed bivariate models based on copulae (see also References [26] for a more general approach). Furthermore, the bivariate Sarmanov model has also been analyzed in the collective model framework [7].

The rest of the paper is organized as follows—in Section 2.1, we describe the proposed Sarmanov model relating the r.v. number of claims and the r.v. average claim cost; its properties and particular cases for the proposed marginal distributions are presented in Section 2.2; moreover, some characteristics of the inference from margin (IFM) estimation procedure are pointed out in Section 2.3. The application on a real database using telematic variables to predict the total cost of claims of a given insured is discussed in Section 3. Finally, Section 4 concludes. This paper ends with Appendixes A and B containing the proofs and additional results.

2. The Dependence Model and Its Properties

Let N_i, $i = 1, ..., m$, be the counting r.v. representing the number of claims of individual i and let Y_i be the r.v. representing the average claim cost per insured. Then, the resulting aggregate claims of individual i can be represented as

$$S_i = N_i Y_i, \tag{1}$$

where the usual assumption under which this model is considered is that Y_i is independent of N_i. We assume that $Y_i > 0$ if and only if $N_i > 0$; otherwise, both variables take the value zero. Therefore, we define the r.v. $\tilde{Y}_i$ representing $Y_i > 0$ and, based on it, the expectation and variance of S_i (which are useful to obtain a-*priori* ratemaking in insurance) are given by:

$$\begin{aligned} \mathbb{E}S_i &= \mathbb{E}N_i \mathbb{E}\tilde{Y}_i, \\ VarS_i &= \mathbb{E}\left[\tilde{Y}_i^2\right] VarN_i + (\mathbb{E}N_i)^2 Var\tilde{Y}_i. \end{aligned}$$

However, in practice, the independence assumption is not always true, in which case the moments of S_i must be deduced from a bivariate distribution associated to the random vector (N_i, Y_i), distribution that takes into account the dependence between the two variables.

Moreover, we assume that for each i, the distributions of both random variables in vector (N_i, Y_i) depend on a set of k quantitative or binary covariates, which are represented by the vector $\mathbf{X}_i = (X_{i1}, ..., X_{ik})'$. For simplicity, the covariates are assumed to be common, but they can also be different between the two variables. In practice, we specify the relation as a GLM and define the linear predictor $\mathbf{X_i}'\beta^j$, where $\beta^j = \left(\beta_1^j, ..., \beta_k^j\right)$, $j \in \{N, Y\}$, are vectors of parameters to be estimated; note that throughout the paper, such a $j \in \{N, Y\}$ should be interpreted as an upper index and not as a power.

2.1. The Model Relating the Counting and Average Claim Cost r.v.s

This dependence model is defined in two parts: the first part is the probability mass function (pmf) of $N_i = 0$ and the second part is the conditional bivariate density function that joins the pmf p_N of the discrete r.v. N_i with the probability density function (pdf) $f_{\tilde{Y}}$ of the continuous r.v. $\tilde{Y}_i$ corresponding to $Y_i > 0$. The model is:

$$f_{Y_i,N_i}(y, n|\mathbf{X_i}) = \left\{ \begin{array}{l} p_N\left(0|\mathbf{X}_i'\beta^N\right),\ n = y = 0 \\ p_N\left(n|\mathbf{X}_i'\beta^N\right) f_{\tilde{Y}}\left(y|\mathbf{X}_i'\beta^Y\right)\left(1 + \omega\psi\left(n|\mathbf{X}_i'\beta^N\right)\phi\left(y|\mathbf{X}_i'\beta^Y\right)\right), n \geq 1, y > 0 \end{array} \right. , \tag{2}$$

where ω is the Sarmanov dependence parameter and $\psi(\cdot)$ and $\phi(\cdot)$ are bounded kernel functions.

In order for (2) to define a proper pdf for all i, we impose the conditions:

$$\begin{aligned}\sum_{n\geq 1}\psi\left(n|\mathbf{X}_i'\beta^N\right)p_N\left(n|\mathbf{X}_i'\beta^N\right) &= \int_{\mathbb{R}^+}\phi\left(y|\mathbf{X}_i'\beta^Y\right)f_{\tilde{Y}}\left(y|\mathbf{X}_i'\beta^Y\right)dy = 0, \text{ and} \quad (3)\\ 1+\omega\psi\left(n|\mathbf{X}_i'\beta^N\right)\phi\left(y|\mathbf{X}_i'\beta^Y\right) &\geq 0, \text{ for all } n\geq 1, y>0. \quad (4)\end{aligned}$$

Then, the kernel functions limits for each i are defined as:

$$\begin{aligned} m_N(\mathbf{X}_i'\beta^N) &= \inf_{n\geq 1}\psi\left(n|\mathbf{X}_i'\beta^N\right),\ M_N(\mathbf{X}_i'\beta^N) = \sup_{n\geq 1}\psi\left(n|\mathbf{X}_i'\beta^N\right),\\ m_Y(\mathbf{X}_i'\beta^Y) &= \inf_{y>0}\phi(y|(\mathbf{X}_i'\beta^Y)),\ M_Y(\mathbf{X}_i'\beta^Y) = \sup_{y>0}\phi\left(y|\mathbf{X}_i'\beta^Y\right).\end{aligned}$$

Taking into account the conditions defined in (3) and (4), limits are also defined for the dependence parameter ω. However, since this parameter does not depend on the linear predictor, new maximum and minimum values are defined as: $m_j^\star = \max_{\forall\mathbf{X}_i} m_j(\mathbf{X}_i'\beta^j)$ and $M_j^\star = \min_{\forall\mathbf{X}_i} M_j(\mathbf{X_i}'\beta^j)$, $j\in\{N,Y\}$, so that the limits of the dependence parameter are:

$$\max\left\{-\frac{1}{m_N^\star m_Y^\star}, -\frac{1}{M_N^\star M_Y^\star}\right\} \leq \omega \leq \min\left\{-\frac{1}{m_N^\star M_Y^\star}, -\frac{1}{M_N^\star m_Y^\star}\right\}. \quad (5)$$

In the following proposition, the new formulae for the expectation and variance of S_i are presented. To simplify the notation, the functions used in the bivariate density expressed in (2) are rewritten as follows: $p_i(n) = p_N(n|\mathbf{X}_i'\beta^N)$, $f_i(y) = f_{\tilde{Y}}(y|\mathbf{X}_i'\beta^Y)$, $\psi_i(n) = \psi(n|\mathbf{X}_i'\beta^N)$ and $\phi_i(y) = \phi(y|\mathbf{X}_i'\beta^Y)$.

Proposition 1. *Under the dependence introduced by the bivariate Sarmanov distribution (2), the expected value and variance of the aggregate claims S_i defined in (1) for individual i are given respectively, by*

$$\begin{aligned}\mathbb{E}S_i &= \mathbb{E}N_i\mathbb{E}\tilde{Y}_i + \omega\mathbb{E}[N_i\psi_i(N_i)]\mathbb{E}\left[\tilde{Y}_i\phi_i(\tilde{Y}_i)\right],\\ VarS_i &= \mathbb{E}\left[\tilde{Y}_i^2\right]VarN_i + (\mathbb{E}N_i)^2 Var\tilde{Y}_i - \omega^2\mathbb{E}^2[N_i\psi_i(N_i)]\mathbb{E}^2\left[\tilde{Y}_i\phi_i(\tilde{Y}_i)\right]\\ &\quad +\omega\left(\mathbb{E}\left[N_i^2\psi_i(N_i)\right]\mathbb{E}\left[\tilde{Y}_i^2\phi_i(\tilde{Y}_i)\right] - 2\mathbb{E}N_i\ \mathbb{E}[N_i\psi_i(N_i)]\ \mathbb{E}\tilde{Y}_i\ \mathbb{E}\left[\tilde{Y}_i\phi_i(\tilde{Y}_i)\right]\right).\end{aligned}$$

From Proposition 3 in Bolancé and Vernic [7], the following correlation for each i can be easily deduced:

$$corr(Y_i, N_i) = \frac{\omega\mathbb{E}[N_i\psi_i(N_i)]\mathbb{E}\left[\tilde{Y}_i\phi_i(\tilde{Y}_i)\right] + p_i(0)\mathbb{E}N_i\mathbb{E}\tilde{Y}_i}{\sqrt{(1-p_i(0))\left(Var\tilde{Y}_i + p_i(0)\mathbb{E}^2\left[\tilde{Y}_i\right]\right)VarN_i}}. \quad (6)$$

To calculate the moments of S_i and the correlation expressed in (6), we need to define the marginal GLMs and the resulting particular bivariate Sarmanov model with given kernel functions.

We propose to use exponential kernels. For the cost per insured Y_i, the kernel function is expressed as $\phi_i(y) = e^{-\gamma y} - \mathcal{L}_{\tilde{Y}_i}(\gamma)$, where $\mathcal{L}_{\tilde{Y}_i}$ denotes the Laplace transform of $\tilde{Y}_i$.

For the number of claims with $n \geq 1$ in (2), we let $\psi_i(n) = e^{-\delta n} - k_i$, and, to find k_i, we impose condition (3) as follows

$$\begin{aligned}
\sum_{n\geq 1} \psi_i(n) p_i(n) &= \sum_{n\geq 1} \left(e^{-\delta n} - k_i\right) p_i(n) \\
&= \sum_{n\geq 0} e^{-\delta n} p_i(n) - p_i(0) - k_i \left(\sum_{n\geq 0} p_i(n) - p_i(0) \right) \\
&= \mathcal{L}_{N_i}(\delta) - p_i(0) - k_i(1 - p(0)) = 0.
\end{aligned}$$

Therefore, $k_i = \frac{\mathcal{L}_{N_i}(\delta) - p_i(0)}{1 - p_i(0)}$ and $\psi_i(n) = e^{-\delta n} - \frac{\mathcal{L}_{N_i}(\delta) - p_i(0)}{1 - p_i(0)}$, where $\mathcal{L}_{N_i}$ denotes the Laplace transform of N_i.

2.2. *Marginal Distributions*

2.2.1. Counting Distribution

We assume that the distribution of the counting process N_i is the Negative Binomial (NB), where we take $\mu_i^N = \mathbb{E}N_i = e^{\mathbf{X}_i' \beta^N}$, so that in the GLM specification $\ln \mu_i^N = \mathbf{X}_i' \beta^N$, and the pmf is:

$$p_i(n) = \frac{\Gamma(\alpha + n)}{n! \Gamma(\alpha)} \left(\frac{\alpha}{\alpha + \mu_i^N} \right)^{\alpha} \left(\frac{\mu_i^N}{\alpha + \mu_i^N} \right)^{n}, \tag{7}$$

where $\alpha > 0$. The previous model is heteroscedastic and its variance is $VarN_i = \frac{\mu_i^N(\alpha + \mu_i^N)}{\alpha}$ for each i. Furthermore, the Laplace transform is $\mathcal{L}_{N_i}(\delta) = \left(\frac{\alpha}{\alpha + \mu_i^N (1 - e^{-\delta})} \right)^{\alpha}$.

To obtain the correlation between the frequency and the severity per policyholder, and the moments of S_i from Proposition 1, we need to calculate $\mathbb{E}\left[N_i^2\right]$, $\mathbb{E}[N_i \psi_i(N_i)]$ and $\mathbb{E}\left[N_i^2 \psi_i(N_i)\right]$. The first one is direct, given that $VarN_i = \mathbb{E}\left[N_i^2\right] - (\mathbb{E}N_i)^2$, while the other two expectations can be directly deduced from the Proposition 5 of Bolancé and Vernic [7]. The results are:

$$\begin{aligned}
\mathbb{E}\left[N_i^2\right] &= \frac{\mu_i^N(\alpha + \mu_i^N + \alpha\mu_i^N)}{\alpha}, \qquad (8) \\
\mathbb{E}[N_i \psi_i(N_i)] &= \frac{\mu_i^N \alpha^{\alpha+1}}{(\alpha + \mu_i^N)\left[\alpha + \mu_i^N(1 - e^{-\delta})\right]^{\alpha}} \times \left\{ \frac{\alpha + \mu_i^N}{\alpha e^{\delta} - \mu_i^N(1 - e^{\delta})} \right. \qquad (9) \\
&- \left. \frac{(\alpha + \mu_i^N)^{\alpha+1} - (\alpha + \mu_i^N)(\alpha + \mu_i^N(1 - e^{-\delta}))^{\alpha}}{\alpha\left[(\alpha + \mu_i^N)^{\alpha} - \alpha^{\alpha}\right]} \right\}, \\
\mathbb{E}\left[N_i^2 \psi_i(N_i)\right] &= \frac{\mu_i^N \alpha^{\alpha+1}}{(\alpha + \mu_i^N)\left[\alpha + \mu_i^N(1 - e^{-\delta})\right]^{\alpha}} \times \left\{ \frac{\alpha\mu_i^N(\alpha + \mu_i^N) + (\alpha + \mu_i^N)^2 e^{\delta}}{\left[\alpha e^{\delta} - \mu_i^N(1 - e^{\delta})\right]^2} \right. \qquad (10) \\
&- \left. \left(\mu_i^N + \alpha\left(1 + \mu_i^N\right)\right) \frac{(\alpha + \mu_i^N)^{\alpha+1} - (\alpha + \mu_i^N)(\alpha + \mu_i^N(1 - e^{-\delta}))^{\alpha}}{\alpha^2\left[(\alpha + \mu_i^N)^{\alpha} - \alpha^{\alpha}\right]} \right\}.
\end{aligned}$$

2.2.2. Severity Distribution

In what concerns the mean cost per policyholder represented by the r.v. Y_i, we assume that its distribution is unknown, but the variable $\tilde{Y}_i$ can be normalised using a Box-Cox transformation. The one parameter Box-Cox transformation is given by:

$$T_{\lambda}(y) = \begin{cases} \frac{y^{\lambda} - 1}{\lambda} \text{ if } \lambda \neq 0, \ y > 0, \\ \ln(y) \text{ if } \lambda = 0, \ y > 0, \end{cases} \tag{11}$$

while the two parameters Box-Cox transformation is:

$$T_{\lambda_1,\lambda_2}(y) = \begin{cases} \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1} \text{ if } \lambda_1 \neq 0,\ y > -\lambda_2, \\ \ln(y+\lambda_2) \text{ if } \lambda_1 = 0,\ y > -\lambda_2. \end{cases} \tag{12}$$

We assume that the two parameters Box-Cox transformation $T_{\lambda_1,\lambda_2}(\cdot)$ is applied to the r.v. $\tilde{Y}_i$ and obtain the truncated normal r.v. $Z_i = T_{\lambda_1,\lambda_2}(\tilde{Y}_i)$. Since the domain of $\tilde{Y}_i$ is left bounded (also necessary in order to have a bounded kernel function), when $\lambda_1 \geq 0$, the r.v. Z_i is left truncated normal (LTN), $Z_i \sim LTN\left(\mu_i^Z, \sigma^2; a\right)$, while when $\lambda_1 < 0$, the resulting r.v. is doubly truncated normal (DTN), $Z_i \sim DTN\left(\mu_i^Z, \sigma^2; a, b\right)$, with $a < b \leq -\lambda_1^{-1}$ as we shall see below.

Then, with $\varphi(\cdot)$ and $\Phi(\cdot)$ denoting the pdf and, respectively, the cumulative distribution function of the standard normal distribution, the density of $\tilde{Y}_i$ is given by

$$f_{\tilde{Y}_i}(y) = \begin{cases} T'_{\lambda_1,\lambda_2}(y)\varphi\left(\frac{T_{\lambda_1,\lambda_2}(y)-\mu_i^Z}{\sigma}\right)\frac{1}{\sigma\bar{\Phi}\left(z_{a,i}\right)}, \lambda_1 \geq 0 \\ T'_{\lambda_1,\lambda_2}(y)\varphi\left(\frac{T_{\lambda_1,\lambda_2}(y)-\mu_i^Z}{\sigma}\right)\frac{1}{\sigma\left(\Phi\left(z_{b,i}\right)-\Phi\left(z_{a,i}\right)\right)}, \lambda_1 < 0 \end{cases},$$

where $\bar{\Phi}(z) = 1 - \Phi(z)$ and $z_{a,i} = \frac{a-\mu_i^Z}{\sigma}, z_{b,i} = \frac{b-\mu_i^Z}{\sigma}$. More precisely:

- If $\lambda_1 > 0$, the condition $y > -\lambda_2$ implies

  $$z = T_{\lambda_1,\lambda_2}(y) = \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1} > -\frac{1}{\lambda_1} = T_{\lambda_1,\lambda_2}(-\lambda_2),$$

  hence the left truncation point a must satisfy the condition $a \geq -\frac{1}{\lambda_1}$, for which we obtain that the left truncation condition gives

  $$z = T_{\lambda_1,\lambda_2}(y) > a \Rightarrow y > (1 + a\lambda_1)^{\frac{1}{\lambda_1}} - \lambda_2.$$

  Therefore, the pdf of $\tilde{Y}_i$ is

  $$f_{\tilde{Y}_i}(y) = \frac{(y+\lambda_2)^{\lambda_1-1}}{\sigma\sqrt{2\pi}\bar{\Phi}\left(z_{a,i}\right)} e^{-\frac{1}{2\sigma^2\lambda_1^2}\left((y+\lambda_2)^{\lambda_1}-1-\lambda_1\mu_i^Z\right)^2},\ y > (1+a\lambda_1)^{\frac{1}{\lambda_1}} - \lambda_2,\ a \geq -\frac{1}{\lambda_1}. \tag{13}$$

- If $\lambda_1 = 0$, then $y > -\lambda_2$ implies

  $$z = T_{\lambda_1,\lambda_2}(y) = \ln(y+\lambda_2) > T_{\lambda_1,\lambda_2}(-\lambda_2) = -\infty,$$

  so that the left truncation point a can be any real value. Then the left truncation condition becomes

  $$z = T_{\lambda_1,\lambda_2}(y) > a \Rightarrow y > e^a - \lambda_2.$$

  The distribution of $\tilde{Y}_i$ becomes the left truncated lognormal $LTLN\left(\mu_i^Z, \sigma^2; a\right)$ having pdf

  $$f_{\tilde{Y}_i}(y) = \frac{1}{(y+\lambda_2)\sigma\sqrt{2\pi}\bar{\Phi}\left(z_{a,i}\right)} e^{-\frac{1}{2\sigma^2}\left(\ln(y+\lambda_2)-\mu_i^Z\right)^2},\ y > e^a - \lambda_2.$$

- If $\lambda_1 < 0$, the condition $y > -\lambda_2$ implies

  $$z = T_{\lambda_1,\lambda_2}(y) = \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1} < -\frac{1}{\lambda_1} = T_{\lambda_1,\lambda_2}(-\lambda_2),$$

and the left truncation point a must be such that $a < -\frac{1}{\lambda_1}$. Again, the left truncation condition gives

$$z = T_{\lambda_1,\lambda_2}(y) > a \Rightarrow y > (1 + a\lambda_1)^{\frac{1}{\lambda_1}} - \lambda_2.$$

However, note that in this case, when $y \to \infty$ then $z \to -\frac{1}{\lambda_1}$, hence the domain of Z_i is also right bounded, yielding the doubly truncated normal distribution $DTN(\mu_i^Z, \sigma^2; a, b)$ with $b = -\lambda_1^{-1}$ (note that we can take $b < -\lambda_1^{-1}$ as long as $a < b$). We therefore write the pdf of $\tilde{Y}_i$ as

$$f_{\tilde{Y}_i}(y) = \frac{(y+\lambda_2)^{\lambda_1 - 1}}{\sigma\sqrt{2\pi}(\Phi(z_{b,i}) - \Phi(z_{a,i}))} e^{-\frac{1}{2\sigma^2\lambda_1^2}\left((y+\lambda_2)^{\lambda_1} - 1 - \lambda_1\mu_i^Z\right)^2}, y > (1 + a\lambda_1)^{\frac{1}{\lambda_1}} - \lambda_2, a < -\frac{1}{\lambda_1}.$$

In order to write the exponential kernel function corresponding to $Z_i, \phi_{Z_i}(z) = e^{-\gamma z} - \mathcal{L}_{Z_i}(\gamma)$, we note that the Laplace transform of the DTN distribution $DTN(\mu_i^Z, \sigma^2; a, b)$ is

$$\mathcal{L}_{Z_i}(\gamma) = e^{-\gamma\mu_i^Z + \frac{\gamma^2\sigma^2}{2}} \frac{\Phi\left(\frac{b - \mu_i^Z + \gamma\sigma^2}{\sigma}\right) - \Phi\left(\frac{a - \mu_i^Z + \gamma\sigma^2}{\sigma}\right)}{\Phi(z_{b,i}) - \Phi(z_{a,i})}.$$

Taking $b = \infty$, we obtain the formula for the LTN distribution.
Therefore, the kernel corresponding to $\tilde{Y}_i$ becomes

$$\tilde{\phi}_i(y) = \begin{cases} e^{-\gamma\frac{(y+\lambda_2)^{\lambda_1} - 1}{\lambda_1}} - e^{-\gamma\mu_i^Z + \frac{\gamma^2\sigma^2}{2}} \frac{\bar{\Phi}\left(\frac{a - \mu_i^Z + \gamma\sigma^2}{\sigma}\right)}{\bar{\Phi}(z_{a,i})}, \lambda_1 > 0 \\ \frac{1}{(y+\lambda_2)^{\gamma}} - e^{-\gamma\mu_i^Z + \frac{\gamma^2\sigma^2}{2}} \frac{\bar{\Phi}\left(\frac{a - \mu_i^Z + \gamma\sigma^2}{\sigma}\right)}{\bar{\Phi}(z_{a,i})}, \lambda_1 = 0 \\ e^{-\gamma\frac{(y+\lambda_2)^{\lambda_1} - 1}{\lambda_1}} - e^{-\gamma\mu_i^Z + \frac{\gamma^2\sigma^2}{2}} \frac{\Phi\left(\frac{b - \mu_i^Z + \gamma\sigma^2}{\sigma}\right) - \Phi\left(\frac{a - \mu_i^Z + \gamma\sigma^2}{\sigma}\right)}{\Phi(z_{b,i}) - \Phi(z_{a,i})}, \lambda_1 < 0 \end{cases}.$$

As discussed above for the r.v. N_i, the following quantities are needed: $\mathbb{E}\tilde{Y}_i$, $\mathbb{E}[\tilde{Y}_i^2]$, $\mathbb{E}[\tilde{Y}_i\tilde{\phi}_i(\tilde{Y}_i)]$, $\mathbb{E}[\tilde{Y}_i^2\tilde{\phi}_i(\tilde{Y}_i)]$, which will be separately calculated for $\lambda_1 \neq 0$ and for $\lambda_1 = 0$.

Case $\lambda_1 \neq 0$.

We have

$$\mathbb{E}\tilde{Y}_i = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}\bar{\Phi}(z_{a,i})} \int_{(1+a\lambda_1)^{\frac{1}{\lambda_1}} - \lambda_2}^{\infty} y(y+\lambda_2)^{\lambda_1 - 1} e^{-\frac{1}{2\sigma^2\lambda_1^2}\left((y+\lambda_2)^{\lambda_1} - 1 - \lambda_1\mu_i^Z\right)^2} dy, \lambda_1 > 0 \\ \frac{1}{\sigma\sqrt{2\pi}(\Phi(z_{b,i}) - \Phi(z_{a,i}))} \int_{(1+a\lambda_1)^{\frac{1}{\lambda_1}} - \lambda_2}^{\infty} y(y+\lambda_2)^{\lambda_1 - 1} e^{-\frac{1}{2\sigma^2\lambda_1^2}\left((y+\lambda_2)^{\lambda_1} - 1 - \lambda_1\mu_i^Z\right)^2} dy, \lambda_1 < 0 \end{cases}.$$

We change variable $\frac{(y+\lambda_2)^{\lambda_1} - 1}{\lambda_1} - \mu_i^Z = t \Rightarrow (y + \lambda_2)^{\lambda_1 - 1} dy = dt$, hence

$$\mathbb{E}\tilde{Y}_i = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}\bar{\Phi}(z_{a,i})} \int_{a - \mu_i^Z}^{\infty} \left[(\lambda_1(t + \mu_i^Z) + 1)^{\frac{1}{\lambda_1}} - \lambda_2\right] e^{-\frac{t^2}{2\sigma^2}} dt, \lambda_1 > 0 \\ \frac{1}{\sigma\sqrt{2\pi}(\Phi(z_{b,i}) - \Phi(z_{a,i}))} \int_{a - \mu_i^Z}^{\infty} \left[(\lambda_1(t + \mu_i^Z) + 1)^{\frac{1}{\lambda_1}} - \lambda_2\right] e^{-\frac{t^2}{2\sigma^2}} dt, \lambda_1 < 0 \end{cases}.$$

In general, there is no closed type formula for this integral and for similar integrals associated with $\mathbb{E}[\tilde{Y}_i^2]$, $\mathbb{E}[\tilde{Y}_i\tilde{\phi}_i(\tilde{Y}_i)]$, $\mathbb{E}[\tilde{Y}_i^2\tilde{\phi}_i(\tilde{Y}_i)]$ and they must be evaluated numerically. However, in the following particular case some recursive formulas can be used.

Particular case: $\lambda_1 = \frac{1}{m}$, where m is a positive integer. The following notation for $k \in \mathbb{N}$, $k > 0$ is introduced:

$$
\begin{aligned}
A^i_{m,k} &= \mathbb{E}\left[\left(\frac{Z_i}{m}+1\right)^k\right], \\
B^i_{m,k} &= \mathbb{E}\left[\left(\frac{Z_i}{m}+1\right)^k e^{-\gamma Z_i}\right].
\end{aligned}
$$

Here the upper index i emphasises the connection with individual i. The proof of the following lemma is very easy and we skip it since it can also be obtained as a particular case of a result in Burkardt [27].

Lemma 1. *Let $L_j(\alpha)$ be the j-th moment of the standard left truncated normal distribution with left truncation point α, that is,*

$$
L_j(\alpha) = \frac{1}{\sqrt{2\pi}\bar{\Phi}(\alpha)} \int_\alpha^\infty z^j e^{-\frac{z^2}{2}} dz, j \in \mathbb{N}.
$$

Then $L_j(\alpha)$ can be recursively evaluated as

$$
\begin{aligned}
L_0(\alpha) &= 1, L_1(\alpha) = \frac{\varphi(\alpha)}{\bar{\Phi}(\alpha)}, \\
L_j(\alpha) &= \alpha^{j-1}\frac{\varphi(\alpha)}{\bar{\Phi}(\alpha)} + (j-1)L_{j-2}(\alpha), j \geq 2.
\end{aligned}
$$

Lemma 2. *With the above notation, for $Z_i \sim LTN\left(\mu_i^Z, \sigma^2; a\right)$ it holds that*

$$
\begin{aligned}
\mathbb{E}[Z_i^r] &= \sum_{j=0}^{r} \binom{r}{j} \sigma^j \left(\mu_i^Z\right)^{r-j} L_j\left(\frac{a-\mu_i^Z}{\sigma}\right), r \in \mathbb{N}, \\
A^i_{m,k} &= \sum_{r=0}^{k} \binom{k}{r} \frac{\mathbb{E}[Z_i^r]}{m^r}, \\
B^i_{m,k} &= e^{-\gamma\mu_i^Z + \frac{\gamma^2\sigma^2}{2}} \frac{\bar{\Phi}\left(\frac{a-\mu_i^Z}{\sigma}+\sigma\gamma\right)}{\bar{\Phi}\left(\frac{a-\mu_i^Z}{\sigma}\right)} \sum_{j=0}^{k} \binom{k}{j} \left(\frac{\mu_i^Z-\sigma^2\gamma}{m}+1\right)^{k-j} \left(\frac{\sigma}{m}\right)^j L_j\left(\frac{a-\mu_i^Z}{\sigma}+\sigma\gamma\right).
\end{aligned}
$$

The following proposition provides the needed formulas.

Proposition 2. *Let $\tilde{Y}_i$ be the r.v. with pdf (13) and $\lambda_1 = \frac{1}{m}$. Then*

$$
\begin{aligned}
\mathbb{E}\tilde{Y}_i &= A^i_{m,m} - \lambda_2, \\
\mathbb{E}\left[\tilde{Y}_i^2\right] &= A^i_{m,2m} - 2\lambda_2 A^i_{m,m} + \lambda_2^2, \\
\mathbb{E}\left[\tilde{Y}_i\tilde{\phi}_i(\tilde{Y}_i)\right] &= B^i_{m,m} - A^i_{m,m}\mathcal{L}_{Z_i}(\gamma), \\
\mathbb{E}\left[\tilde{Y}_i^2\tilde{\phi}_i(\tilde{Y}_i)\right] &= B^i_{m,2m} - A^i_{m,2m}\mathcal{L}_{Z_i}(\gamma) - 2\lambda_2\left(B^i_{m,m} - \mathcal{L}_{Z_i}(\gamma)A^i_{m,m}\right).
\end{aligned}
$$

Case $\lambda_1 = 0$.

Let now $\lambda_1 = 0$. As noted before, in this case, $\tilde{Y}_i$ follows a left truncated Lognormal distribution $LTLN\left(\mu_i^Z, \sigma^2; a\right)$ and the following result holds without further assumption on the parameter λ_1.

Proposition 3. *For $\tilde{Y}_i \sim LTLN\left(\mu_i^Z, \sigma^2; a\right)$ and assuming that $\gamma = 1$, it holds that*

$$\begin{aligned}
\mathbb{E}\tilde{Y}_i &= e^{\mu_i^Z + \frac{\sigma^2}{2}} \frac{\bar{\Phi}(z_{a,i} - \sigma)}{\bar{\Phi}(z_{a,i})} - \lambda_2, \\
\mathbb{E}\left[\tilde{Y}_i^2\right] &= e^{2\mu_i^Z + 2\sigma^2} \frac{\bar{\Phi}(z_{a,i} - 2\sigma)}{\bar{\Phi}(z_{a,i})} - 2\lambda_2 e^{\mu_i^Z + \frac{\sigma^2}{2}} \frac{\bar{\Phi}(z_{a,i} - \sigma)}{\bar{\Phi}_i(z_{a,i})} + \lambda_2^2, \\
\mathbb{E}\left[\tilde{Y}_i \tilde{\phi}_i(\tilde{Y}_i)\right] &= 1 - e^{\sigma^2} \frac{\bar{\Phi}(z_{a,i} + \sigma)\bar{\Phi}(z_{a,i} - \sigma)}{\bar{\Phi}^2(z_{a,i})}, \\
\mathbb{E}\left[\tilde{Y}_i^2 \tilde{\phi}_i(\tilde{Y}_i)\right] &= e^{\mu_i^Z + \frac{\sigma^2}{2}} \frac{\bar{\Phi}(z_{a,i} - \sigma)}{\bar{\Phi}(z_{a,i})} - e^{\mu_i^Z + \frac{5\sigma^2}{2}} \frac{\bar{\Phi}(z_{a,i} + \sigma)\bar{\Phi}(z_{a,i} - 2\sigma)}{\bar{\Phi}^2(z_{a,i})} \\
&\quad + 2\lambda_2 \left(e^{\sigma^2} \frac{\bar{\Phi}(z_{a,i} + \sigma)\bar{\Phi}(z_{a,i} - \sigma)}{\bar{\Phi}^2(z_{a,i})} - 1 \right),
\end{aligned}$$

where, as before, $z_{a,i} = \frac{a - \mu_i^Z}{\sigma}$.

2.3. Parameter Estimation

Since the distribution associated with the r.v. Y_i is unknown, the parameter estimation is based on the two parameters Box-Cox transformed variable $T_{\lambda_1,\lambda_2}\left(\tilde{Y}_i\right) = Z_i$, whose distribution is LTN or DTN. Therefore, we can estimate the bivariate Sarmanov with NB and TN marginal distributions and then the results on the original scale can be deduced as we have shown in Section 2.2. The bivariate Sarmanov distribution with NB and LTN or DTN marginals has pdf:

$$f_{Z_i,N_i}(z, n|\mathbf{X_i}) = \begin{cases} \left(\frac{\alpha}{\alpha + \mu_i^N}\right)^{\alpha}, n = z = 0 \\ \frac{\Gamma(\alpha+n)}{n!\Gamma(\alpha)} \left(\frac{\alpha}{\alpha+\mu_i^N}\right)^{\alpha} \left(\frac{\mu_i^N}{\alpha+\mu_i^N}\right)^{n} \frac{\varphi\left(\frac{z-\mu_i^Z}{\sigma}\right)}{\sigma\bar{\Phi}(z_{a,i})} (1 + \omega\psi_i(n)\phi_{Z_i}(z)), n \geq 1, z > 0, \lambda_1 \geq 0 \\ \frac{\Gamma(\alpha+n)}{n!\Gamma(\alpha)} \left(\frac{\alpha}{\alpha+\mu_i^N}\right)^{\alpha} \left(\frac{\mu_i^N}{\alpha+\mu_i^N}\right)^{n} \frac{\varphi\left(\frac{z-\mu_i^Z}{\sigma}\right)}{\sigma(\Phi(z_{b,i})-\Phi(z_{a,i}))} (1 + \omega\psi_i(n)\phi_{Z_i}(z)), n \geq 1, z > 0, \lambda_1 < 0 \end{cases}, \tag{14}$$

where we recall that $\mu_i^N = e^{\mathbf{X}_i'\beta^N}$ and $\mu_i^Z = \mathbf{X}_i'\beta^Z$. In conclusion, we have to estimate the transformation parameters λ_1 and λ_2, the vectors of parameters β^N and β^Z associated with the covariate vector, the parameter α of NB marginal distribution, the parameter σ of the LTN or DTN marginal distribution and the dependence parameter ω.

Let (n_i, y_i), $i = 1, \ldots, m$, be a sample of observed values of frequency and severity per policyholder and let $z_i = T_{\lambda_1,\lambda_2}(y_i)$, $i = 1, \ldots, m$, be the transformed severity. The logarithm of the likelihood function $l(\Theta)$ to be maximised with respect to the vector of parameters which is defined as $\Theta = \left(\beta^N, \beta^Z, \alpha, \sigma, \lambda_1, \lambda_2, \omega\right)$ to be estimated, depends on the value of λ_1 as follows: letting m_0 be the number of insured with $n_i = 0$ and m_1 the number of insured with $n_i \geq 1$, then

- If $\lambda_1 \neq 0$ and $\lambda_2 > -\min(y_1, \ldots, y_m)$

$$\begin{aligned}
l(\Theta) &= m\alpha \ln \alpha - \alpha \sum_{i=1}^{m} \ln\left(\alpha + e^{\mathbf{X}_i'\beta^N}\right) + \sum_{i=1}^{m_1} \ln \Gamma(\alpha + n_i) - \sum_{i=1}^{m_1} \ln(n_i!) \\
&- m_1 \ln \Gamma(\alpha) + \sum_{i=1}^{m_1} n_i \mathbf{X}_i'\beta^N - \sum_{i=1}^{m_1} n_i \ln\left(\alpha + e^{\mathbf{X}_i'\beta^N}\right) \\
&+ \sum_{i=1}^{m_1} \ln\left[\varphi\left(\frac{\frac{(y_i+\lambda_2)^{\lambda_1}-1}{\lambda_1} - \mathbf{X}_i'\beta^Z}{\sigma}\right)\right] - \sum_{i=1}^{m_1} \ln\left[\Phi\left(\frac{b - \mathbf{X}_i'\beta^Z}{\sigma}\right) - \Phi\left(\frac{a - \mathbf{X}_i'\beta^Z}{\sigma}\right)\right] \\
&- m_1 \ln \sigma + \sum_{i=1}^{m_1} \ln\left[1 + \omega\psi_i(n_i)\phi_{Z_i}\left(\frac{(y_i + \lambda_2)^{\lambda_1} - 1}{\lambda_1}\right)\right],
\end{aligned} \tag{15}$$

where $b = \infty$ if $\lambda_1 \geq 0$ and $b \leq -\lambda_1^{-1}$ if $\lambda_1 < 0$.

- If $\lambda_1 = 0$ and $\lambda_2 > -\min(y_1, ..., y_m)$

$$\begin{aligned} l(\Theta) &= m\alpha \ln \alpha - \alpha \sum_{i=1}^{m} \ln\left(\alpha + e^{\mathbf{X}_i'\beta^N}\right) + \sum_{i=1}^{m_1} \ln \Gamma(\alpha + n_i) - \sum_{i=1}^{m_1} \ln(n_i!) \\ &- m_1 \ln \Gamma(\alpha) + \sum_{i=1}^{m_1} n_i \mathbf{X}_i'\beta^N - \sum_{i=1}^{m_1} n_i \ln\left(\alpha + e^{\mathbf{X}_i'\beta^N}\right) \\ &+ \sum_{i=1}^{m_1} \ln\left[\varphi\left(\frac{\ln(y_i + \lambda_2) - \mathbf{X}_i'\beta^Z}{\sigma}\right)\right] - \sum_{i=1}^{m_1} \ln\left[\bar{\Phi}\left(\frac{a - \mathbf{X}_i'\beta^Z}{\sigma}\right)\right] \\ &- m_1 \ln \sigma + \sum_{i=1}^{m_1} \ln[1 + \omega\psi_i(n_i)\phi_{Z_i}(\ln(y_i + \lambda_2))]. \end{aligned} \tag{16}$$

To maximise the log-likelihood functions defined in (15) and (16), a procedure similar to the one described and tested in Bolancé and Vernic [7] can be used, which is divided in two phases: the first phase consists in using the inference from margins (IFM) technique to estimate the marginal parameters, and the second phase in using these results as starting values to obtain estimations for all parameters based on the maximization of the full log-likelihood.

The IFM is a two step estimation method which starts from an initial estimation of the marginal parameters, estimation obtained by the maximum likelihood (ML) method for the parameters of the two GLMs—GLM of the frequency variable with NB distribution, and GLM of the severity variable whose distribution is defined by the transformation parameters and by the parameters of the truncated normal distribution with given truncation values a and b. We will take the left truncation value such that $\Phi\left(\frac{a - \mathbf{X}_i'\beta^Z}{\sigma}\right)$ is almost zero; note that since we consider heterogeneity between policyholders, in practice, a good choice will be $a = \min(a_1, ..., a_m)$, where $a_i = -3\sigma + \mathbf{X}_i'\beta^Z$ (another simple choice for a would be the minimum of the transformed data) (see References [28] for ML estimation of the univariate model based on Box-Cox transformation). However, note that for (15) we must check that $a \geq -\frac{1}{\lambda_1}$ when $\lambda_1 \geq 0$ and that $a < -\frac{1}{\lambda_1}$ if $\lambda_1 < 0$. In practice, for $\lambda_1 < 0$, we used the maximum value for the right truncation point $b = -\lambda_1^{-1}$, which does not affect the likelihood value because $\Phi((b - mu)/sigma) \approx 1$.

Therefore, for the Sarmanov distribution defined in (14), each iteration of the IFM consists of the following two steps:

Step 1 (iteration r) Given the parameters of the marginal distributions obtained at iteration $r - 1$, find the estimation $\hat{\omega}^r$ of the dependence parameter within the interval defined in (5), estimation that maximises the log-likelihood in (15) or (16).

Step 2 Given $\hat{\omega}^r$ obtained in Step 1, find new values for the parameters of the marginals that maximise the log-likelihood function in (15) or (16).

In practice, in the numerical analysis that will be presented in Section 3, two assumptions on the Box-Cox transformation parameters are used. On the one hand, as we have described before, we can include the transformation parameters in the optimization procedure. On the other hand, we a-*priori* select the values of the transformation parameters, which can be calculated by maximizing the log-likelihood of the normal distribution associated with the transformed variable. The results obtained in both cases are practically the same.

3. Numerical Analysis

In this section, a database corresponding to a car insurance portfolio is analyzed, in which some of the variables have been measured via a telematic system. It is interesting to check if, in the collective risk model context, the telematic variables can replace or complement classical a-*priori* ratemaking variables, as the age, the driving license age, the power of the car, and so forth. The dependent variables are the frequency and mean severity of claims per policyholder. Information on 25,014 policyholders with a car insurance is available, who have contracted a policy that incorporates a GPS device in the vehicle. The data were provided by a Spanish insurer. The information corresponds to all the drivers in the insurer's database who had a telematics insurance product in 2010. In general, this type of insurance is mainly chosen by young drivers who value the fact that their car can be located in case of an accident. So, the group of drivers is generally composed of new drivers. Our dataset contains all the available drivers; most of them did not report any claim during 2010, but a few reported at least one accident. Only accidents at fault are considered, and the cost of the accident was also collected. The cost of the accident is generally known once damages, medical expenses and bodily injury compensations are paid. The sum of all these costs was included in the dataset in a variable measuring the cost of claim.

Traditionally, the variables used to model the a-*priori* premium in auto insurance have been classified in policyholder and auto characteristics; furthermore, if there is some experience with the insureds, that is, they are not new in the company, some tariff variables also could be considered (see Reference [11], for some examples, [chapters 12 and 13]). In practice, these variables are those available for the insurance company. In our case, we have a portfolio with new policyholders in the company and for whom we know their age, their driving license age, and if night parking is used; we also know the gender, but this variable is not included in the analysis because official regulations prevent differentiating premiums based on the gender of the insured. About the auto of the policyholder, it is know that all are cars of private use and the available variables are the age of the car and its power. More recently, thanks to GPS devices, telematic variables are available for the insurance company; in general, the variables that are used in the premium calculation are the total of kilometers (km) driven and three percentages: in urban area, at night and over the speed limit (see Reference [22] for an example with similar portfolio).

Table 1 shows the main descriptive statistics of the dependent variables and of the covariates used in the bivariate Sarmanov model defined in (14). It can be noticed that the dependent variables have right skewness. Considering this shape and comparing with alternative count data models like the Poisson or Zero Inflated Poisson, the NB GLM defined in (7) proved to be the best option for our data. Focusing on the mean severity, the distributions that have been classically used are the Gamma or the Lognormal (see Reference [7] for an example). However, in our case, the shape of the severity distribution is unknown and has a heavier tail than the Gamma and the Lognormal; in this case, the Box-Cox transformation allows us to work with the normal distribution for the transformed variable. Regarding the explanatory variables, it can be checked that the portfolio is made up of young policyholders with a symmetric age distribution. The variables with the largest skewness, which in our case is positive, are the % of km driven over the speed limit (X_8) and the % of km driven at night (X_9); these variables are those directly related with the risk exposure and, therefore, with the frequency and the severity of claims.

At the bottom of Table 1, the linear Pearson correlation coefficient and its confidence interval are shown. The value of this coefficient is low, but significant. Given the right skewness shape of both dependent variables, the linear correlation is not the best dependence measure. Alternatively, Reference [25] proposed fitting copula models with GLM marginals, and, using Gaussian copula, they obtain a dependence parameter equal to 0.13, that is, the dependence between frequency and severity is not very strong, but it affects the insurance premium considerably. The Sarmanov model with GLM marginals proposed in this work is an alternative way to copula for modelling dependence; furthermore, it allows heterogeneity dependence between policyholders with different covariate values.

Table 1. Definition of variables and descriptive statistics: mean, median, standard deviation (STD), minimum (Min), maximum (Max), skewness (Skew) and kurtosis (Kur). The last row shows the linear correlation between the dependent variables and a confidence interval (CI) at a 95% level.

	Description	Mean	Median	STD	Min	Max	Skew	Kur
N	Number of claims per policyholder	0.106	0.000	0.370	0.000	5.000	4.005	21.598
Y	Cost of claims per policyholder *	1444.175	696.720	4141.511	17.750	130870.360	23.365	678.291
X_1	Age of the driver	27.565	27.491	3.094	19.849	36.903	−0.059	2.077
X_2	Age of driver License	7.174	6.616	3.053	1.810	15.910	1.402	9.959
X_3	Age of the vehicle	8.749	7.775	4.174	1.938	20.468	0.770	2.996
X_4	Power of the vehicle	97.225	97.000	27.773	12.000	500.000	−1.309	2.713
X_5	Night parking (1 = yes, 0 = no)	0.774	1.000	0.418	0.000	1.000	−1.467	9.955
X_6	Total of km. driven in logarithm	8.681	8.774	0.698	0.466	10.820	1.049	4.126
X_7	% of km. driven at urban area	25.875	22.922	14.357	0.000	100.000	2.382	10.764
X_8	% of km. driven over the speed limit	6.332	3.999	6.827	0.000	70.433	1.830	9.512
X_9	% of km. driven at night	6.908	5.148	6.351	0.000	100.000	0.497	2.393
ρ	Correlation between N and Y (CI)				0.094 (0.052, 0.135)			

* Given $N > 0$ equal to 2177 policyholders.

Focusing on the marginal distributions for the frequency and severity variables, the NB and Gamma GLM models (see Reference [2] for example), respectively, are the most used to model the variables number and cost of claims in auto insurance. Alternatively, in some cases, the Zero Inflated Poisson (ZIP) and Lognormal distributions could improve the fit of NB and Gamma, respectively (see References [7,29], for example). In Table 2 the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are compared for alternative univariate GLM models considered for marginals, including the Poisson, the ZIP and the NB for frequency, and the Gamma, Lognormal and Box-Cox transformation based model for severity. These results show that ZIP is better than NB if AIC is used, but, if BIC is used, the best model is the NB. For the severity variable, the best model is the one based on the Box-Cox transformation.

Table 2. AIC and BIC for Generalized Linear Models (GLMs) for frequency (N) and severity (Y) variables.

	N			Y		
	Poisson	**ZIP**	**NB**	**Gamma**	**Lognormal**	**Box-Cox**
AIC	17,372	16,861	16,906	36,009	35,312	25,835
BIC	17,454	17,024	16,996	36,071	35,374	25,898

Table 3 presents the results of the estimated bivariate Sarmanov model used to fit the joint distribution of the frequency and of the Box-Cox transformation of the mean severity of claims for a given insured with the characteristics represented in a given vector of covariates. The transformation parameters are calculated a-*priori* by maximizing the normal log-likelihood function of the transformed variable; alternatively, the transformation parameters are included in step 2 of the estimation procedure described in Section 2.3, and it is observed that the results do not improve. A negative value for λ_1 is obtained, indicating that the distribution of the mean cost per policyholder has a longer and heavier tail than the Lognormal distribution. The estimated model with all the explanatory variables that are shown in Table 3 is compared with alternative proposals that can be found in the literature (see References [2,7,25]), to estimate collective risk model distributions (results are shown in Appendix B); for the data used in this paper, these models do not improve our proposed model.

Three Sarmanov GLM models are estimated—Model I incorporates all the explanatory variables, Model II only includes classical a-*priori* characteristics of the policyholder and car, and, finally, Model III only includes telematic variables. Furthermore, Models I and II are estimated without the variable X_1 (Age of the driver) given its high correlation with X_2 (Age of driver license), and noticed that the signs and p-values of the parameters of the remaining variables change only slightly. The dependence parameter ω is positive and indicates a significant dependence between frequency and severity that need to be considered in this collective model in order to calculate insurance premiums.

Comparing the three models, we can see that the best one is Model I, this model incorporating non telematic and telematic variables. The effect of the driving experience on the frequency and severity is negative; on the contrary, the effects of the car characteristics are positive in the sense that older car and higher power increase the accident rates. The coefficients associated with the telematic variables are positive, their p-values indicating significant effects on the frequency. For the severity, the significant effects are associated to X_8 (Percentage of kilometers driven over the speed limit) and X_9 (Percentage of kilometers driven at night).

Focusing on Models II and III, we note that the best fit is obtained with the telematic variables; although, as we have seen in Model I, these variables are complemented by non telematic variables improving the goodness of fit in the full model. In short, the effects of the covariates hardly change if we compare models I, II and III, which implies that multicollinearity hardly affects the results, and that telematic and non-telematic variables complement each other and take into account different characteristics of the insured. The telematic variables are taking into account the exposure to risk that the non-telematic variables do not detect by themselves.

Table 3. Parameter estimates (p-values) and goodness of fit statistics for the bivariate Sarmanov models with GLM marginals for number (N) and mean cost (Y) per policyholder.

	$\lambda_1 = -0.0745$, $\lambda_2 = 13.2501$					
	Model I		**Model II**		**Model III**	
	N	Y	N	Y	N	Y
Int.	−8.073 (<0.001)	4.990 (<0.001)	−1.794 (<0.001)	5.171 (<0.001)	−8.277 (<0.001)	5.016 (<0.001)
X_1	0.006 (0.259)	−0.004 (0.260)	−0.007 (0.227)	−0.005 (0.194)		
X_2	−0.083 (<0.001)	−0.001 (0.480)	−0.081 (<0.001)	0.002 (0.488)		
X_3	0.012 (0.013)	0.007 (0.021)	0.008 (0.082)	0.008 (0.015)		
X_4	0.002 (0.037)	0.001 (0.071)	0.003 (0.001)	0.001 (0.019)		
X_5	−0.025 (0.321)	−0.009 (0.396)	−0.022 (0.339	−0.005 (0.446)		
X_6	0.603 (<0.001)	0.012 (0.323)			0.603 (<0.001)	0.010 (0.344)
X_7	0.022 (<0.001)	0.001 (0.200)			0.024 (<0.001)	0.001 (0.153)
X_8	0.007 (0.024)	0.004 (0.041)			0.006 (0.046)	0.005 (0.018)
X_9	0.007 (0.026)	0.004 (0.051)			0.009 (0.007)	0.004 (0.031)
α	0.407 (<0.001)		0.339 (<0.001)		0.381 (<0.001)	
σ		0.650 (<0.001)		0.652 (<0.001)		0.652 (<0.001)
ω	110.236 (0.036)		91.508 (0.065)		129.375 (0.019)	
$\hat{l}(\Theta)$	−10,555.071		−10,706.315		−10,605.962	
AIC	21,156.143		21,442.630		21,237.923	
BIC	242,786.895		246,265.501		243,957.371	

In Table 4, we show the mean of S_i, that is, the pure premium, for four different insured profiles, using the full sample (with extremes) and with the positive dependence parameter ω estimated in Model I; this mean $E(S_i)$, is compared with the one obtained without dependence (i.e., dependence parameter equal to zero); we also calculate the differences between these two pure premiums, with and without frequency and severity dependence. The first and second profiles correspond to policyholders with high risk—they are 25 years old, with 6 years old driver license, with a car with 150 horsepower, 8100 yearly kilometers driven, 50% on urban area, 40% over the speed limit and 20% at night; the difference between these two profiles is that the first one does not use night parking, while the second does. The profiles 3 and 4 are policyholders with lower risk, having the same personal and car characteristics, same total km driven as the profiles 1 and 2, but with only 5% over the speed limit and 10% at night. We observe how the Sarmanov dependence parameter affects the pure premium. The difference between premiums calculated with and without dependence increases with risk, from 5.69 Euros for the lower risk profile (Profile 4) to 12.51 Euros for the higher risk profile (Profile 1).

Table 4. Pure premium in collective model using parameters of Model I.

	Profile 1	**Profile 2**	**Profile 3**	**Profile 4**
$E(S_i), \omega = 110.2364$	687.7501	659.5172	359.0895	344.4958
$E(S_i), \omega = 0$	675.2375	647.6507	353.0899	338.8105
Difference	12.51264	11.86648	5.999624	5.685362

4. Conclusions

In this paper, dependence between the claim frequency and the average severity of a policyholder was introduced, using a bivariate Sarmanov distribution with Negative Binomial GLM and Normal GLM marginals. The Normal GLM distribution was obtained by applying a Box-Cox transformation with two parameters on the original distribution, motivated by the fact that, in the collective risk model context, the claim cost distribution may not coincide with the traditionally used distributions (Exponential, Gamma, Lognormal, etc.). For some specific values of the transformation parameters, some useful closed type expressions for calculating the moments of the aggregate claims were obtained.

The proposed model was fitted on a sample of policyholders from an auto insurance portfolio. The peculiarity of these data is that they are associated with an insurance which involved the installation of a GPS/inertial device in the vehicle, device that allows the collection of telematic information related to the total kilometers and to the way which they are traveled. We have shown how these variables complement and do not substitute the variables that have traditionally been used in car insurance pricing. Furthermore, we have analyzed the importance of incorporating the dependence between frequency and severity in the collective model for calculating the premium, and how our Sarmanov model allows this dependence to be incorporated in a simple way.

In general, the bivariate Sarmanov distribution shows how the effect of dependency between frequency and severity is different depending on the risk profiles. For our dataset, the results show that riskier profiles have a greater effect of dependency. Furthermore, the Box-Cox transformation improves the distribution fit of the cost of claims variable, given that it allows a longer and heavier right tail than classical models like Gamma and Lognormal.

Author Contributions: All authors contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors gratefully acknowledge the help of the four referees in significantly improving the paper. They also thank Fundación BBVA, Equipos de Investigación Científica en Big Data 2018 for support.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: The used database is available from the authors.

Appendix A

Proof of Proposition 1. The expected value results easily from

$$\begin{aligned} \mathbb{E}S_i &= \mathbb{E}[N_iY_i] = \sum_{n\geq 1}\int_0^\infty nyp_i(n)f_i(y)(1+\omega\psi_i(n)\phi_i(y))dy \\ &= \mathbb{E}N_i\mathbb{E}\tilde{Y}_i + \omega\sum_{n\geq 1} np_i(n)\psi_i(n)\int_0^\infty yf_i(y)\phi_i(y)dy. \end{aligned}$$

For the variance, we start with

$$\begin{aligned} \mathbb{E}\left[S_i^2\right] &= \mathbb{E}\left[N_i^2Y_i^2\right] = \sum_{n\geq 1}\int_0^\infty n^2y^2p_i(n)f_i(y)(1+\omega\psi_i(n)\phi_i(y))dy \\ &= \mathbb{E}\left[N_i^2\right]\mathbb{E}\left[\tilde{Y}_i^2\right] + \omega\sum_{n\geq 1} n^2p_i(n)\psi_i(n)\int_0^\infty y^2f_i(y)\phi_i(y)dy \\ &= \mathbb{E}\left[N_i^2\right]\mathbb{E}\left[\tilde{Y}_i^2\right] + \omega\mathbb{E}\left[N_i^2\psi_i(N_i)\right]\mathbb{E}\left[\tilde{Y}_i^2\phi_i(\tilde{Y}_i)\right]. \end{aligned}$$

Therefore, the variance follows from

$$\begin{aligned} VarS_i &= \mathbb{E}\left[S_i^2\right] - \mathbb{E}^2[S_i] = \mathbb{E}\left[N_i^2\right]\mathbb{E}\left[\tilde{Y}_i^2\right] + \omega\mathbb{E}\left[N_i^2\psi_i(N_i)\right]\mathbb{E}\left[\tilde{Y}_i^2\phi_i(\tilde{Y}_i)\right] \\ &\quad -\left(\mathbb{E}^2[N_i]\,\mathbb{E}^2\left[\tilde{Y}_i\right] + 2\omega\mathbb{E}N_i\mathbb{E}\tilde{Y}_i\mathbb{E}[N_i\psi_i(N_i)]\mathbb{E}\left[\tilde{Y}_i\phi_i(\tilde{Y}_i)\right] + \omega^2\mathbb{E}^2[N_i\psi_i(N_i)]\mathbb{E}^2\left[\tilde{Y}_i\phi_i(\tilde{Y}_i)\right]\right) \\ &= \left(\mathbb{E}\left[N_i^2\right] - \mathbb{E}^2[N_i]\right)\mathbb{E}\left[\tilde{Y}_i^2\right] + \mathbb{E}^2[N_i]\left(\mathbb{E}\left[\tilde{Y}_i^2\right] - \mathbb{E}^2\left[\tilde{Y}_i\right]\right) - \omega^2\mathbb{E}^2[N_i\psi_i(N_i)]\mathbb{E}^2\left[\tilde{Y}_i\phi_i(\tilde{Y}_i)\right] \\ &\quad +\omega\left(\mathbb{E}\left[N_i^2\psi_i(N_i)\right]\mathbb{E}\left[\tilde{Y}_i^2\phi_i(\tilde{Y}_i)\right] - 2\mathbb{E}N_i\mathbb{E}\tilde{Y}_i\mathbb{E}[N_i\psi_i(N_i)]\mathbb{E}\left[\tilde{Y}_i\phi_i(\tilde{Y}_i)\right]\right). \end{aligned}$$

This completes the proof. □

Proof of Lemma 2. Since $Z_i \sim LTN\left(\mu_i^Z, \sigma^2; a\right)$, we can write $Z_i = \mu_i^Z + \sigma Z_i'$, where $Z_i' \sim LTN\left(0, 1; \frac{a-\mu_i^Z}{\sigma}\right)$ and the first formula is immediate with $\alpha = z_{a,i} = \frac{a-\mu_i^Z}{\sigma}$. The formula of $A_{m,k}^i$ is also immediate. To prove the formula of $B_{m,k}^i$ we write

$$\begin{aligned} B_{m,k}^i &= \frac{1}{\sigma\sqrt{2\pi}\bar{\Phi}(z_{a,i})}\int_a^\infty\left(\frac{z}{m}+1\right)^k e^{-\gamma z-\frac{\left(z-\mu_i^Z\right)^2}{2\sigma^2}}dz \\ &= \frac{1}{\sigma\sqrt{2\pi}\bar{\Phi}(z_{a,i})}\int_a^\infty\left(\frac{z}{m}+1\right)^k e^{-\frac{1}{2\sigma^2}\left(z-\mu_i^Z+\sigma^2\gamma\right)^2+\frac{1}{2}\left(\sigma^2\gamma^2-2\mu_i^Z\gamma\right)}dz. \end{aligned}$$

Changing variable $x = \frac{z-\mu_i^Z+\sigma^2\gamma}{\sigma}$ results in

$$\begin{aligned} B_{m,k}^i &= \frac{e^{-\gamma\mu_i^Z+\frac{\gamma^2\sigma^2}{2}}}{\sqrt{2\pi}\bar{\Phi}\left(z_{a,i}\right)}\int_{\frac{a-\mu_i^Z+\sigma^2\gamma}{\sigma}}^\infty\left(\frac{\sigma}{m}x+\frac{\mu_i^Z-\sigma^2\gamma}{m}+1\right)^k e^{-\frac{x^2}{2}}dx \\ &= e^{-\gamma\mu_i^Z+\frac{\gamma^2\sigma^2}{2}}\sum_{j=0}^k\left(\begin{array}{c} k \\ j \end{array}\right)\left(\frac{\mu_i^Z-\sigma^2\gamma}{m}+1\right)^{k-j}\left(\frac{\sigma}{m}\right)^j\frac{1}{\sqrt{2\pi}\bar{\Phi}\left(\frac{a-\mu_i^Z}{\sigma}\right)}\int_{\frac{a-\mu_i^Z+\sigma^2\gamma}{\sigma}}^\infty x^je^{-\frac{x^2}{2}}dx, \end{aligned}$$

and, by considering the definition of L_j, we obtain the stated result. □

Proof of Proposition 2. Since

$$\tilde{Y}_i = (1+\lambda_1 Z_i)^{\frac{1}{\lambda_1}} - \lambda_2 = \left(\frac{Z_i}{m}+1\right)^m - \lambda_2,$$

we easily obtain that $\mathbb{E}\tilde{Y}_i = A^i_{m,m} - \lambda_2$. Also,

$$\tilde{Y}_i^2 = \left(\frac{Z_i}{m}+1\right)^{2m} - 2\lambda_2\left(\frac{Z_i}{m}+1\right)^m + \lambda_2^2,$$

which yields the formula of $\mathbb{E}\left[\tilde{Y}_i^2\right]$.

For the formula of $\mathbb{E}\left[\tilde{Y}_i\tilde{\phi}_i\left(\tilde{Y}_i\right)\right]$, we write

$$\begin{aligned}
\mathbb{E}\left[\tilde{Y}_i\tilde{\phi}_i\left(\tilde{Y}_i\right)\right] &= \mathbb{E}\left[\left(\left(\frac{Z_i}{m}+1\right)^m - \lambda_2\right)\left(e^{-\gamma Z_i} - \mathcal{L}_{Z_i}(\gamma)\right)\right] \\
&= \mathbb{E}\left[\left(\frac{Z_i}{m}+1\right)^m e^{-\gamma Z_i}\right] - \mathcal{L}_{Z_i}(\gamma)\mathbb{E}\left[\left(\frac{Z_i}{m}+1\right)^m\right] - \lambda_2\left(\mathbb{E}\left(e^{-\gamma Z_i}\right) - \mathcal{L}_{Z_i}(\gamma)\right) \\
&= B^i_{m,m} - \mathcal{L}_{Z_i}(\gamma)A^i_{m,m}.
\end{aligned}$$

Also, for $\mathbb{E}\left[\tilde{Y}_i^2\tilde{\phi}_i\left(\tilde{Y}_i\right)\right]$ we have

$$\begin{aligned}
\mathbb{E}\left[\tilde{Y}_i^2\tilde{\phi}_i\left(\tilde{Y}_i\right)\right] &= \mathbb{E}\left[\left(\left(\frac{Z_i}{m}+1\right)^m - \lambda_2\right)^2\left(e^{-\gamma Z_i} - \mathcal{L}_{Z_i}(\gamma)\right)\right] \\
&= \mathbb{E}\left[\left(\left(\frac{Z_i}{m}+1\right)^{2m} - 2\lambda_2\left(\frac{Z_i}{m}+1\right)^m + \lambda_2^2\right)\left(e^{-\gamma Z_i} - \mathcal{L}_{Z_i}(\gamma)\right)\right] \\
&= \mathbb{E}\left[\left(\frac{Z_i}{m}+1\right)^{2m} e^{-\gamma Z_i}\right] - \mathcal{L}_{Z_i}(\gamma)\mathbb{E}\left[\left(\frac{Z_i}{m}+1\right)^{2m}\right] + \lambda_2^2\left(\mathbb{E}\left(e^{-\gamma Z_i}\right) - \mathcal{L}_{Z_i}(\gamma)\right) \\
&\quad -2\lambda_2\mathbb{E}\left[\left(\frac{Z_i}{m}+1\right)^m e^{-\gamma Z_i}\right] + 2\lambda_2\mathcal{L}_{Z_i}(\gamma)\mathbb{E}\left[\left(\frac{Z_i}{m}+1\right)^m\right],
\end{aligned}$$

which easily yields the result. □

Proof of Proposition 3. Since $Z_i = \ln\left(\tilde{Y}_i + \lambda_2\right) \Leftrightarrow \tilde{Y}_i = e^{Z_i} - \lambda_2$, we have that

$$\begin{aligned}
\mathbb{E}\tilde{Y}_i &= \mathbb{E}\left[e^{Z_i} - \lambda_2\right] = \mathcal{L}_{Z_i}(-1) - \lambda_2 = e^{-(-1)\mu_i^Z + \frac{(-1)^2\sigma^2}{2}} \frac{\bar{\Phi}\left(\frac{a-\mu_i^Z-\sigma^2}{\sigma}\right)}{\bar{\Phi}\left(\frac{a-\mu_i^Z}{\sigma}\right)} - \lambda_2, \\
\mathbb{E}\left[\tilde{Y}_i^2\right] &= \mathbb{E}\left[\left(e^{Z_i} - \lambda_2\right)^2\right] = \mathbb{E}\left[e^{2Z_i} - 2\lambda_2 e^{Z_i} + \lambda_2^2\right] = \mathcal{L}_{Z_i}(-2) - 2\lambda_2\mathcal{L}_{Z_i}(-1) + \lambda_2^2 \\
&= e^{2\mu_i^Z + \frac{4\sigma^2}{2}} \frac{\bar{\Phi}\left(\frac{a-\mu_i^Z-2\sigma^2}{\sigma}\right)}{\bar{\Phi}\left(\frac{a-\mu_i^Z}{\sigma}\right)} - 2\lambda_2 e^{\mu_i^Z + \frac{\sigma^2}{2}} \frac{\bar{\Phi}(z_{a,i} - \sigma)}{\bar{\Phi}(z_{a,i})} + \lambda_2^2,
\end{aligned}$$

yielding the first two formulas. The other two formulas result from

$$\begin{aligned}
\mathbb{E}\big[\tilde{Y}_i\tilde{\phi}_i(\tilde{Y}_i)\big] &= \mathbb{E}\Big[\Big(e^{Z_i}-\lambda_2\Big)\Big(e^{-Z_i}-\mathcal{L}_{Z_i}(1)\Big)\Big] = 1-\mathcal{L}_{Z_i}(1)\mathbb{E}\Big[e^{Z_i}\Big]-\lambda_2\mathbb{E}\Big[e^{-Z_i}-\mathcal{L}_{Z_i}(1)\Big] \\
&= 1-\mathcal{L}_{Z_i}(1)\mathcal{L}_{Z_i}(-1) = 1-e^{-\mu_i^Z+\frac{\sigma^2}{2}}\frac{\bar{\Phi}\Big(\frac{a-\mu_i^Z+\sigma^2}{\sigma}\Big)}{\bar{\Phi}\Big(\frac{a-\mu_i^Z}{\sigma}\Big)}e^{\mu_i^Z+\frac{\sigma^2}{2}}\frac{\bar{\Phi}\Big(\frac{a-\mu_i^Z-\sigma^2}{\sigma}\Big)}{\bar{\Phi}\Big(\frac{a-\mu_i^Z}{\sigma}\Big)}, \\
\mathbb{E}\Big[\tilde{Y}_i^2\tilde{\phi}_i(\tilde{Y}_i)\Big] &= \mathbb{E}\Big[\Big(e^{Z_i}-\lambda_2\Big)^2\Big(e^{-Z_i}-\mathcal{L}_{Z_i}(1)\Big)\Big] = \mathbb{E}\Big[\Big(e^{2Z_i}-2\lambda_2e^{Z_i}+\lambda_2^2\Big)\Big(e^{-Z_i}-\mathcal{L}_{Z_i}(1)\Big)\Big] \\
&= \mathbb{E}\Big[e^{Z_i}\Big]-\mathcal{L}_{Z_i}(1)\mathbb{E}\Big[e^{2Z_i}\Big]-2\lambda_2+2\lambda_2\mathcal{L}_{Z_i}(1)\mathbb{E}\Big[e^{Z_i}\Big]+\lambda_2^2\mathbb{E}\Big[e^{-Z_i}-\mathcal{L}_{Z_i}(1)\Big] \\
&= \mathcal{L}_{Z_i}(-1)-\mathcal{L}_{Z_i}(1)\mathcal{L}_{Z_i}(-2)+2\lambda_2\mathcal{L}_{Z_i}(1)\mathcal{L}_{Z_i}(-1)-2\lambda_2 \\
&= e^{\mu_i^Z+\frac{\sigma^2}{2}}\frac{\bar{\Phi}\Big(\frac{a-\mu_i^Z-\sigma^2}{\sigma}\Big)}{\bar{\Phi}(z_{a,i})}-e^{-\mu_i^Z+\frac{\sigma^2}{2}}\frac{\bar{\Phi}\Big(\frac{a-\mu_i^Z+\sigma^2}{\sigma}\Big)}{\bar{\Phi}(z_{a,i})}e^{2\mu_i^Z+\frac{4\sigma^2}{2}}\frac{\bar{\Phi}\Big(\frac{a-\mu_i^Z-2\sigma^2}{\sigma}\Big)}{\bar{\Phi}(z_{a,i})} \\
&\quad +2\lambda_2e^{\sigma^2}\frac{\bar{\Phi}(z_{a,i}+\sigma)\bar{\Phi}(z_{a,i}-\sigma)}{\bar{\Phi}^2(z_{a,i})}-2\lambda_2,
\end{aligned}$$

which completes the proof. □

Appendix B

Table A1. Log-likelihood, AIC and BIC for alternative multivariate models with a NB GLM marginal for frequency variable (N).

GLM Marginal for Severity (Y)	Lognormal Distribution			Box-Cox Based Distribution		
	Sarmanov Distribution	Gaussian Copula *	Conditional Distribution **	Sarmanov Distribution	Gaussian Copula *	Conditional Distribution **
log−lik	−11,636	−11,723	−11,525	−10,555	−10,574	−10,567
AIC	23,318	23,493	23,096	21,156	21,194	21,181
BIC	267,652	269,656	265,093	242,787	243,224	243,070

* Czado et al. [25] and ** Garrido et al. [2].

References

1. Abdallah, A.; Boucher, J.P.; Cossette, H. Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insur. Math. Econ.* **2016**, *68*, 120–133. [CrossRef]
2. Garrido, J.; Genest, C.; Schulz, J. Generalized linear models for dependent frequency and severity of insurance claims. *Insur. Math. Econ.* **2016**, *70*, 205–215. [CrossRef]
3. Valdez, E.A.; Jeong, H.; Ahn, J.Y.; Park, S. Generalized linear mixed models for dependent compound risk models. *Variance* **2018**.
4. Jeong, H.; Valdez, E.A. Predictive compound risk models with dependence. *Insur. Math. Econ.* **2020**, *94*, 182–195. [CrossRef]
5. Bahraoui, Z.; Bolancé, C.; Pelican, E.; Vernic, R. On the bivariate distribution and copula. An application on insurance data using truncated marginal distributions. *Stat. Oper. Res. Trans. SORT* **2015**, *39*, 209–230.
6. Bolancé, C.; Vernic, R. Multivariate count data generalized linear models: Three approaches based on the Sarmanov distribution. *Insur. Math. Econ.* **2019**, *85*, 89–103. [CrossRef]
7. Bolancé, C.; Vernic, R. Frequency and Severity Dependence in the Collective Risk Model: An Approach Based on Sarmanov Distribution. *Mathematics* **2020**, *8*, 1400. [CrossRef]
8. Guo, F.; Wang, D.; Yang, H. Asymptotic results for ruin probability in a two-dimensional risk model with stochastic investment returns. *J. Comput. Appl. Math.* **2017**, *325*, 198–221. [CrossRef]
9. Yang, Y.; Yuen, K.C. Finite-time and infinite-time ruin probabilities in a two-dimensional delayed renewal risk model with Sarmanov dependent claims. *J. Math. Anal. Appl.* **2016**, *442*, 600–625. [CrossRef]
10. Bolancé, C.; Guillen, M.; Pitarque, A. A Sarmanov Distribution with Beta Marginals: An Application to Motor Insurance Pricing. *Mathematics* **2020**, *8*, 2020. [CrossRef]
11. Frees, E.W. *Regression Modelling with Actuarial and Financial Applications*; Cambridge University Press: New York, NY, USA, 2009.
12. Ismail, N.; Jemain, A.A. Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models. In *Casualty Actuarial Society Forum*; 2007; pp. 103–158. Available online: https://www.semanticscholar.org/paper/Handling-Overdispersion-with-Negative-Binomial-and-Ismail-Jemain/2791e7be78958751709b7765d92958c0b295597c (accessed on 2 November 2020).
13. Harrington, S.E. Estimation and testing for functional form in pure premium regression models. *Astin Bull.* **1986**, *16*, 31–43. [CrossRef]

14. Jee, B. A Comparative Analysis of Alternative Pure Premium Models in the Automobile Risk Classification System. *J. Risk Insur.* **1989**, *56*, 434–459. [CrossRef]
15. Box, G.E.P.; Cox, D.R. An Analysis of Transformations. *J. R. Stat. Soc. Ser. B* **1964**, *26*, 211–252. [CrossRef]
16. Pelican, E.; Vernic, R. Parameters estimation for the bivariate Sarmanov distribution with normal-type marginals. *ROMAI J* **2013**, *9*, 155–165.
17. Sun, S.; Bi, J.; Guillen, M.; Pérez-Marín, A.M. Assessing driving risk using internet of vehicles data: An analysis based on generalized linear models. *Sensors* **2020**, *20*, 2712. [CrossRef]
18. Ayuso, M.; Guillen, M.; Nielsen, J.P. Improving automobile insurance ratemaking using telematics: Incorporating mileage and driver behaviour data. *Transportation* **2019**, *46*, 735–752. [CrossRef]
19. Pérez-Marin, A.M.; Guillen, M.; Alcañiz, M.; Bermúdez, L. Quantile regression with telematics information to assess the risk of driving above the posted speed limit. *Risks* **2019**, *7*, 80. [CrossRef]
20. Pesantez-Narvaez, J.; Guillen, M.; Alcañiz, M. Predicting motor insurance claims using telematics data-XGBoost versus logistic Regression. *Risks* **2019**, *7*, 70. [CrossRef]
21. Pérez-Marín, A.M.; Guillen, M. Semi-autonomous vehicles: Usage-based data evidences of what could be expected from eliminating speed limit violations. *Accid. Anal. Prev.* **2019**, *123*, 99–106. [CrossRef]
22. Guillen, M.; Nielsen, J.P.; Ayuso, M.; Pérez-Marín, A.M. The use of telematics devices to improve automobile insurance rates. *Risk Anal.* **2019**, *39*, 662–672. [CrossRef]
23. Bolancé, C.; Guillen, M.; Pinquet, J. On the link between credibility and frequency premium. *Insur. Math. Econ.* **2008**, *43*, 209–213. [CrossRef]
24. Bermúdez, L.; Karlis, D. Bayesian multivariate Poisson models for insurance ratemaking. *Insur. Math. Econ.* **2011**, *48*, 226–236. [CrossRef]
25. Czado, C.; Kastenmeier, R.; Brechmann, E.C.; Min, A. A mixed copula model for insurance claims and claim sizes. *Scand. Actuar. J.* **2012**, *4*, 278–305. [CrossRef]
26. Shi, P.; Feng, X.; Ivantsova, A. Dependent frequency–severity modeling of insurance claims. *Insur. Math. Econ.* **2015**, *64*, 417–428. [CrossRef]
27. Burkardt, J. The Truncated Normal Distribution, 2014. Available online: https://people.sc.fsu.edu/~jburkardt/presentations/truncated_normal.pdf (accessed on 23 October 2020).
28. Zhang, T.; Yang, B. Box-Cox Transformation in Big Data. *Technometrics* **2017**, *59*, 189–201. [CrossRef]
29. Boucher, J.P.; Denuit, M.; Guillen, M. Number of Accidents or Number of Claims? An Approach with Zero-Inflated Poisson Models for Panel Data. *J. Risk Insur.* **2009**, *76*, 821–846. [CrossRef]