International Journal of Mathematical, Engineering and Management Sciences
Vol. 9, No. 2, 224-243, 2024
https://doi.org/10.33889/IJMEMS.2024.9.2.012

Ram Arti Publishers

Enhanced Insurance Risk Assessment using Discrete Four-Variate Sarmanov Distributions and Generalized Linear Models

Piriya Prunglerdbuathong
Department of Mathematics,
Khon Kaen University, Khon Kaen, Thailand.
E-mail: pr.piriya@kkumail.com

Tippatai Pongsart
Department of Statistics,
Khon Kaen University, Khon Kaen, Thailand.
Corresponding author: tipppo@kku.ac.th

Weenakorn Ieosanurak
Department of Mathematics,
Khon Kaen University, Khon Kaen, Thailand.
E-mail: weenie@kku.ac.th

Watcharin Klongdee
Department of Mathematics,
Khon Kaen University, Khon Kaen, Thailand.
E-mail: kwatch@kku.ac.th

(Received on August 16, 2023; Revised on November 6, 2023 & January 4, 2024 & January 26, 2024;
Accepted on February 3, 2024)

Abstract
This research paper investigated multivariate risk assessment in insurance, focusing on four risks of a singular person and their interdependence. This research examined various risk indicators in non-life insurance which was under-writing for organizations with clients that purchase several non-life insurance policies. The risk indicators are probabilities of frequency claims and correlations of two risk lines. The closed forms of probability mass functions evaluated the probabilities of frequency claims. Three generalized linear models of four-variate Sarmanov distributions were proposed for marginals, incorporating various characteristics of policyholders using explanatory variables. All three models were discrete models that were a combination of Poisson and Gamma distributions. Some properties of four-variate Sarmanov distributions were explicitly shown in closed forms. The dataset spanned a decade and included the exposure of each individual to risk over an extended period. The correlations between the two risk types were evaluated in several statistical ways. The parameters of the three Sarmanov model distributions were estimated using the maximum likelihood method, while the results of the three models were compared with a simpler four-variate negative binomial generalized linear model. The research findings showed that Model 3 was the most accurate of all three models since the AIC and BIC were the lowest. In terms of the correlation, it was found that the risk of claiming auto insurances was related to claiming home insurances. Model 1 could be used for the risk assessment of an insurance company that had customers who held multiple types of insurances in order to predict the risks that may occur in the future. When the insurance company can forecast the risks that may occur in the future, the company will be able to calculate appropriate insurance premiums.

Keywords- Multivariate Sarmanov distribution, Negative binomial distribution, Generalized linear model, Non-life insurance, Claim frequency.

1. Introduction

Calculating the insurance premium requires evaluating the risk value related to an accident. Non-life

insurance firms usually focus on the risk associated with the policy contracted over a certain time period, sometimes one year. Because insurance companies have a duty to pay compensation to customers who purchase insurance claims, evaluating the risks posed by customers is extremely necessary. If customers hold many insurance policies with a single provider, then the company needs to prioritize these individuals. Therefore, analyzing the risk profiles of clients with multiple insurance policies from the same provider is both complex and intriguing. We integrate data from two distinct portfolios onto the auto and home insurance lines. It is a challenge to analyze the risk associated with customers who hold multi-line contracts. However, a multivariate analysis of policyholders' accident risks has not yet been developed. We can extend the work about this to many other research areas. This study analyzed the accident risk of policyholders across multiple insurance lines. The interdependence of accident rates indicated the behavior of customers holding several insurance policies in both motor and home insurance across these lines. This approach is relevant as insurance companies increasingly consider clients who have multiple policies with the same insurer. By performing a long-term analysis of clients' risk behavior, insurers can develop comprehensive strategies for retaining valuable customers and adjust premiums according to actual risk profiles. To perform this kind of analysis, insurers must reevaluate and modify their data systems. This study presents a suitable mechanism to achieve this aim using a collection of readily available data.

How to respond to the following query: What is the customer's accident risk if they have multiple insurance lines from their insurer? The answer might be achieved by using univariate generalized linear models (GLMs). Assuming that the client behaves independently for each line of insurance, we could estimate one model for each line. On the other hand, we used a multivariate generalized linear model (GLM) to create a joint distribution. That is specific to each client and takes into account the dependence between the accident risks covered by different insurance lines. The research team chose to adopt multivariate GLMs to jointly evaluate the dependence across multiple insurance lines.

Previous analyses considered data that included various policies of the same policyholder. These policies represented several insurance coverages. The analysis focused on the policyholder's profitability and loyalty. It also took into account the overall perspective of a client of an insurance company. Table 1 shows the related works, classified based on the number of variables and the distribution of random variables.

Table 1. The related works classified according to the number of variables and the distribution of random variables.

Number of variables	Distribution	Detail
Bivariate models in auto insurance	Other distributions	Bolancé et al. (2008) demonstrated the application of the Conditional Tail Expectation (CTE) risk metric on a bivariate real dataset containing two categories of auto insurance claim expenses. They fitted several continuous bivariate distributions, including normal, lognormal, and skew-normal, along with an alternative log-skew-normal to the data. Additionally, they introduced a bivariate nonparametric transformed kernel estimation. CTE formulas for all these distributions were provided, and numerical outcomes from the real data were analyzed and compared.
		Bermudez and Karlis (2011) introduced different multivariate Poisson regression models in order to relax the independence assumption, including zero-inflated models to account for excess of zeros and overdispersion. These models had been largely ignored to date, mainly because of their computational difficulties. Finally, these models were applied to an automobile insurance claims database with three different types of claims. They analyzed the consequences for pure and loaded premiums when the independence assumption is relaxed by using different multivariate Poisson regression models together with their zero-inflated versions.
		Boucher and Inoussa (2014) introduced a novel approach to handling bonus-malus systems in the presence of panel data. This approach was exemplified using car insurance data, including a numerical example featuring both at-fault and non-at-fault claims from a Canadian insurance company. Although the model was applied to car insurance, they asserted that if any other line of business relies on past claim experience for premium setting, a similar method to the one proposed should be employed.

Table 1 continued…

	Sarmanov distribution	Abdallah et al. (2016) suggested a bivariate dynamic model for claim counts, where the previous claims history of one type of claim is utilized to improve the prediction of another type of claim. This novel bivariate dynamic distribution for claim counts was founded on random effects originating from the Sarmanov family of multivariate distributions. Their proposed model offered enhanced flexibility in calculating predictive premiums, as closed-form expressions were readily obtainable for the predictive distribution, the moments, and the predictive moments.
		Bolancé and Vernic (2020) employed the Sarmanov distribution to establish a correlation between the number of claims and the respective severities of individual claims. The Poisson and Negative Binomial (NB) distributions were employed as marginal models for the claim count, while the Gamma and Lognormal distributions were utilized for modeling the cost of claims. They discussed their maximum likelihood estimation in the context of the Sarmanov framework. A real-world application using Spanish insurance data provided empirical validation.
		Vernic et al. (2022) presented such a model, which can link various kinds of marginals in flexible dependence structures, and was based on Sarmanov's bivariate distribution. More specifically, they utilized this distribution to combine the average severity and the frequency of claims. In addition, they proposed and demonstrated a maximum likelihood estimation approach for parameter estimation on both simulated and actual data.
	Sarmanov distribution with GLM marginals	Bolancé et al. (2020) created a bivariate model using the Sarmanov distribution alongside marginal Beta GLMs, enabling us to effectively model two key variables within contemporary motor insurance telematics databases. Their proposal was the derivation of closed-form expressions for various important quantities, including the bivariate cumulative distribution function and conditioned moments, as well as covariance and correlation. These quantities were essential in the context of risk analysis. They demonstrated that their Sarmanov-Beta-GLM model provided superior fits compared to earlier proposals that also relied on the Sarmanov distribution.
		Alemany et al. (2021) suggested a bivariate Sarmanov model that depends on the parameters of the Box-Cox transformation and had a Normal GLM and a NB as marginals. Using explanatory telematic variables, they applied this model to investigate the frequency-severity bivariate distribution linked to a pay-as-you-drive car insurance portfolio.
Bivariate models in entropy	Sarmanov distribution	Alawady et al. (2022) offered the marginal distributions of concomitants of k-record values based on Sarmanov family of bivariate distributions. Additionally, for this family, they derived the joint distribution of concomitants of k-record values. Furthermore, certain novel and practical characteristics of information measures were examined, including the cumulative entropy, cumulative residual entropy, inaccuracy measure, Shannon entropy, and cumulative residual Fisher information. Lastly, they provided several instances and numerical studies supporting the theoretical results.
		Barakat et al. (2022) presented the bivariate Sarmanov family. The Sarmanov family was one of the most adaptable and practical extended families of the conventional Farlie–Gumbel–Morgenstern family. They explored the distribution theory of order statistics concomitants within a specified family. They also investigated information measures such as Shannon entropy, inaccuracy measure, and Fisher information number, both theoretically and numerically. Two real-world bivariate datasets were analyzed and demonstrated satisfactory performance.
Bivariate models in Local Government Property Insurance	Other distributions	Jeong et al. (2023) offered a regression model for multivariate claim frequency data that accounts for overdispersion from the unobserved heterogeneity caused by systematic effects in the data, as well as dependence structures across the claim count responses, which might vary in sign and range. The bivariate Poisson-lognormal regression model with different dispersion was taken into consideration. To determine the best values for the model's parameters, the researchers used a novel Monte Carlo method, which harnessed the power of randomness to make difficult computations more manageable. They demonstrated our methodology using Wisconsin Local Government Property Insurance Fund data. Their results were an adequate performance.
Trivariate models in auto insurance	Other distributions	Shi and Valdez (2014) looked into different methods for building multivariate count models based on the negative binomial distribution in response to the peculiarities of an insurance dataset. The first used a blend of max-id copulas to directly work with discrete count data, allowing for flexible pair-wise association as well as tail and global dependence. The second used elliptical copulas to combine continuous data while keeping the original counts' dependent structure. The empirical analysis considered the claim frequency of three types of claims (third-party property damage, own damage, and third-party bodily injury) in a portfolio of motor insurance policies from a Singapore insurer. The results showed that copula-based techniques outperform the standard shock model. Finally, they put the various models to use in loss prediction applications.
Trivariate models in auto and home insurance	Sarmanov distribution with GLM marginals	Bolancé and Vernic (2019) proposed using three trivariate GLMs based on the Sarmanov distribution to capture the dependence between accident rates in different insurance lines for each policyholder. Driven by a single dataset, they considered three different kinds of accident risks: two for motor vehicles and one for homes. The three models were ultimately compared in a quantitative manner with both the elliptical copula-based models and the more straightforward trivariate negative binomial generalized linear model (NB GLM). The model 2 was the best fit.

Table 1 continued…

Bivariate and trivariate models in auto insurance	Sarmanov distribution with GLM marginals	Wang (2023) presented the Sarmanov distribution and two-stage inference. First, they used GLMs to fit the marginals and extract the relevant residuals. Second, the Sarmanov family of bivariate distributions connected these marginals via the residuals' rank. The trivariate Sarmanov model with rank-based technique outperformed the bivariate Sarmanov model.
Bivariate models in life insurance	Sarmanov distribution with Phase-Type marginal	Moutanabbir and Abdelrahman (2022) used the bivariate Sarmanov distribution with Phase-Type marginal distributions to create the model. They provided several practical closed-form distributional and interest-quantity formulations for multiple-life insurance contracts. They were able to increase the achievable correlation range with the implementation of this new kernel function.

The Sarmanov distribution with GLM marginal is proper to the interested problem and improves a fit from the previous studies. The advantage of the Sarmanov distribution is its flexible dependent structure. The Sarmanov distribution is a mixed model combining the Poisson and Gamma distributions, which is suitable for the insurance claim data. We choose the GLM since the dataset includes the information of policyholders. This article presents three four-variate GLMs based on the Sarmanov distribution to model the joint accident rate of policyholders with diverse risk coverages, considering various assumptions for capturing the dependence between the number of claims in each coverage. The novelty of this research is studying a discrete four-variate Sarmanov distribution for three models and its parameter estimation. The discrete multivariate Sarmanov with NB GLM marginals was used for the first model. In contrast, the second model followed Abdallah et al. (2016) for the bivariate situation, with some modifications, by combining a multivariate Sarmanov distribution with Gamma distributed marginals with a multivariate model of separate Poisson distributions. The third model mixed Gamma distributions with a discrete multivariate Sarmanov distribution with Poisson marginals. A method based on conditional likelihood was used to estimate the parameters of the Sarmanov-based models which yielded comparable results to the maximum likelihood estimation, particularly for large datasets. Because this work has never been done previously, the constraints and scope of the study are discussed in the next paragraph.

These three proposed Sarmanov-based models were illustrated using real-world aggregate data and artificial data that combined information from motor and home insurance lines. Three risk lines data were collected over a ten-year period from several national insurance companies in Spain. Data for the fourth risk line are created. We need to generate data due to a lack of available data. This dataset aggregation could lead to reduced numbers of zeros and increased over-dispersion but our models demonstrated flexibility in their marginals and dependency structure, making them suitable for more extensive and diverse datasets. The scope of the study are as follows: (i) all three studied models are four-variate discrete Sarmanov distributions with NB GLM marginals but different kernel functions. (ii) Obtaining statistics for a client who maintained four different types of insurance with the same company over a decade is challenging. Data on claims related to property damage (PD), bodily injury (BI), and home (H) were collected from 80,924 insurance providers in Spain who had both auto and home insurance agreements from 2006 to 2015. The third-party (TP) insurance is the fourth category. The count of claims in the TP insurance was determined using the negative binomial distribution.

This paper presented a comprehensive analysis of accident risk for insurance customers who hold policies in four lines. By utilizing multivariate Sarmanov-based models, insurers can gain valuable insights into the joint risk behavior of policyholders, and design effective bonus and penalty strategies to optimize customer retention and premium adjustments. The numerical application of real insurance data highlighted the practicality and superiority of our three proposed models, emphasizing their potential applicability in actuarial settings.

The remainder of this article is organized as follows. Section 2 describes the mixed models that yielded three multivariate Sarmanov distributions with NB GLM marginals as well as the multivariate NB GLM. Aspects of the specificity of Sarmanov distributions were investigated and a technique for modeling estimation was suggested. Estimation parameter methods are described in section 3, with the dataset and numerical application outcomes, including a predictive analytic presented in section 4. Section 5 displays some conclusions, while the Appendix contains all the supporting tables.

2. Multivariate Mixed Distribution

Subsection 2.1 reviews the multivariate Sarmanov distribution and its properties. The multivariate NB is described in subsection 2.2, and three models based on Sarmanov distribution with different kernel functions are proposed. Marginal values of these models are considered as a multivariate negative binomial distribution.

2.1 Multivariate Sarmanov Distribution

Sarmanov (1966) initially proposed Sarmanov family of bivariate distributions. The multivariate Sarmanov distribution, suggested by Lee (1996), is a powerful tool to ensure that a model has a good fit. This distribution comprises both a continuous version and a discrete version. In the continuous case, a d-variate random variable $\boldsymbol{Y}$ is assumed to be a continuous multivariate Sarmanov distribution. The joint probability density function (p.d.f.) of $\boldsymbol{Y}$ is defined for $y \in \mathbb{R}^d$ by

$$h_{Sarm}(y) = \prod_{j=1}^{d} h_j(y_j)\left(1 + \sum_{k=2}^{d}\sum_{1\le j_1<\cdots<j_k\le d} \omega_{j_1\cdots j_k}\phi_{j_1}(y_{j_1})\cdots\phi_{j_k}(y_{j_k})\right) \quad (1)$$

where, $(h_j)_{j=1}^{d}$ are the marginal p.d.f.s, $(\phi_j)_{j=1}^{d}$ are bounded non-constant kernel functions and $\omega_{j_1\cdots j_k}$ are real numbers such that

$$\begin{cases} \int_{\mathbb{R}} \phi_j(y)h_j(y)\,dy = 0, \quad \text{for } j = 1,\dots,d; \\ 1 + \sum_{k=2}^{d}\sum_{1\le j_1<\cdots<j_k\le d} \omega_{j_1\cdots j_k}\phi_{j_1}(y_{j_1})\cdots\phi_{j_k}(y_{j_k}) \ge 0, \quad \forall y \in \mathbb{R}^d \end{cases} \quad (2)$$

The correlation coefficient between two marginal variables is associated with the parameters $\omega_{j,k}$ and the kernel functions ϕ_j by:

$$corr(Y_j, Y_k) = \omega_{j,k}\frac{\mathbb{E}[Y_j\phi_j(Y_j)]\mathbb{E}[Y_k\phi_k(Y_k)]}{\sqrt{Var(Y_j)Var(Y_k)}} \quad (3)$$

Several forms of the kernel function (ϕ_j) are satisfied by the conditions expressed in (2). This study considered the exponential kernel, i.e.,

$$\phi_j(y) = e^{-y} - \mathcal{L}_{Y_j}(1) \quad (4)$$

where, $\mathcal{L}_{Y_j}$ is the Laplace transform of Y_j : $\int_0^{\infty} e^{-ty_j}h_j(y_j)dy_j$.

Equation (4) is bounded as it considers only non-negative values. Furthermore, it is decreasing, therefore
$m_j = inf_{y\ge 0}\phi_j(y) = \phi_j(\infty) = -\mathcal{L}y_j(1)$, and
$M_j = sup_{y\ge 0}\phi_j(y) = \phi_j(0) = 1 - \mathcal{L}y_j(1), j = 1,\dots,d$ (5)

This study focused on the four-variate dataset and considered the four-variate Sarmanov distribution, with joint p.d.f. described by:

$$h_{Sarm}(\boldsymbol{y}) = \prod_{j=1}^{4} h_j(y_j) \times \{1 + \sum_{1\le j<k\le 4} \omega_{j,k}\phi_j(y_j)\phi_k(y_k) + \sum_{1\le j_1}^{2}\sum_{<j_2}^{3}\sum_{<j_3\le 4}^{4} \omega_{j_1,j_2,j_3}\phi_{j_1}(y_{j_1})\phi_{j_2}(y_{j_2})\phi_{j_3}(y_{j_3}) + \omega_{1,2,3,4}\prod_{i=1}^{4}\phi_i(y_i)\} \quad (6)$$

***Proposition* 1.** The dependent parameter $\omega_{1,2,3,4}$ can be calculated by the following formula:

$$\omega_{1,2,3,4} = \frac{\mathbb{E}\left[\prod_{j=1}^{4}(Y_j - \mathbb{E}Y_j)\right]}{\prod_{j=1}^{4}\mathbb{E}[Y_j\emptyset_j(Y_j)]} \quad (7)$$

Therefore, the conditions (2) in the four-variate version make these restrictions:

$$1 + \omega_{j,k}\varepsilon_j\varepsilon_k \ge 0, 1 \le j < k \le 4 \quad (8)$$

$$1 + \sum_{1\le j<k\le 4} \omega_{j,k}\varepsilon_j\varepsilon_k + \omega_{j,k,h}\varepsilon_j\varepsilon_k\varepsilon_h \ge 0, 1 \le j < k < h \le 4 \quad (9)$$

$$1 + \sum_{1\le j<k\le 4} \omega_{j,k}\varepsilon_j\varepsilon_k + \sum_{1\le j_1}^{2}\sum_{<j_2}^{3}\sum_{<j_3\le 4}^{4} \omega_{j_1,j_2,j_3}\varepsilon_{j_1}\varepsilon_{j_2}\varepsilon_{j_3} + \omega_{1,2,3,4}\,\varepsilon_1\varepsilon_2\varepsilon_3\varepsilon_4 \ge 0 \quad (10)$$

where, $\varepsilon_j \in \{m_j, M_j\}, j = 1,2,3,4.$

In the discrete version, the joint probabilities of the multivariate Sarmanov distribution are defined at $\boldsymbol{n} \in \mathbb{N}^d$ by,

$$\Pr_{Sarm}(\boldsymbol{N} = \boldsymbol{n}) = \prod_{j=1}^{d}\Pr(N_j = n_j) \times \left(1 + \sum_{k=2}^{d}\sum_{1\le j_1<\cdots<j_k\le d} \omega_{j_1\cdots j_2}\phi_{j_1}(n_{j_1})\cdots\phi_{j_k}(n_{j_k})\right) \quad (11)$$

where, $\left(\Pr(N_j = n_j)\right)_{j=1}^{d}$ are the marginal p.f.s.

2.2 Multivariate Negative Binomial Distribution

The multivariate negative binomial distribution is the well-known multivariate Poisson GLM mixed with Gamma distribution. It is beneficial to eliminate one of the Poisson properties so that its mean is equal to its variance.

To achieve this, let the random variable (r.v.) $N_j \sim Poisson\ (\mu_j\theta)$ with μ_j a fixed positive parameter, $j = 1, \dots, d$, and θ as an unpredictability parameter which can represent the risk habit of the policyholder. Then, the parameter θ can be a positive continuous r.v. Θ. Further, assume that conditionally on $\Theta = \theta$, the random variables N_j are independent for each $j = 1, \dots, d$.

According to assumptions above, we let $\Theta \sim Gamma(\alpha, \alpha)$ and $\alpha > 0$, be a Gamma distributed r.v., then the joint probability function (p.f.) of $\mathbf{N} = (N_1, \dots, N_d)$ becomes:

$$\Pr(\mathbf{N} = \mathbf{n}) = \int_0^\infty \Pr(\mathbf{N} = \mathbf{n}|\Theta = \theta)\mathrm{h}(\theta)d\theta$$

$$= \frac{\alpha^\alpha}{\Gamma(\alpha)}\left(\prod_{j=1}^{d}\frac{\mu_j^{n_j}}{n_j!}\right)\int_0^\infty \theta^{\sum_{j=1}^{d} n_j + \alpha - 1} \times e^{-\theta\left(\sum_{j=1}^{d}\mu_j + \alpha\right)}d\theta$$

$$= \frac{\Gamma\left(\alpha + \sum_{j=1}^{d} n_j\right)}{\Gamma(\alpha)\prod_{j=1}^{d} n_j!}\left(\frac{\alpha}{\alpha + \sum_{k=1}^{d}\mu_k}\right)^\alpha \times \prod_{j=1}^{d}\left(\frac{\alpha}{\alpha + \sum_{k=1}^{d}\mu_k}\right)^{n_j}, \mathbf{n} \in \mathbb{N}^{\boldsymbol{d}} \quad (12)$$

where, h is probability density function (p.d.f.) of Θ. The equation (12) is multivariate NB distribution (see, e.g., Johnson et al., 1997):

$$NB_d\left(\alpha; \frac{\alpha}{\alpha + \sum_{j=1}^{d}\mu_j}, \left(\frac{\mu_j}{\alpha + \sum_{j=1}^{d}\mu_j}\right)_{j=1,\dots,d}\right).$$

For convenient calculation, we give the index to specify information on the exposure of each person; that is, letting by I the number of each insured under study and using the subscript i associated with the individual, we have that for $\mathbf{N}_i = (N_{i1}, \dots, N_{id}), i = 1, \dots, I$.

$$\mathbf{N}_i \sim NB_d\left(\alpha; \frac{\alpha}{\alpha+\sum_{k=1}^{d}(E_{ik}\mu_{ik})}, \left(\frac{E_{ij}\mu_{ij}}{\alpha+\sum_{k=1}^{d}(E_{ik}\mu_{ik})}\right)_{j=1,\dots,d}\right) \tag{13}$$

The correlation coefficient between two marginals for each i is:

$$corr(N_{ij}, N_{ik}) = \sqrt{\frac{E_{ij}\mu_{ij}E_{ik}\mu_{ik}}{(E_{ij}\mu_{ij}+\alpha)(E_{ik}\mu_{ik}+\alpha)}},\ 1 \le j \le k \le d \tag{14}$$

2.3 Multivariate Sarmanov Distributions with GLMs Marginals

All three models are described from Bolancé and Vernic (2019). Three Sarmanov-based four-variate models were proposed with identical NB GLM marginals but various dependence architectures. The initial model (Model 1) was a basic four-variate Sarmanov distribution with NB GLM marginals, while the other two models (Model 2 and Model 3) were created by combining three independent Poisson distributions with a Sarmanov distribution with Gamma marginals.

2.3.1 Model 1

For each i, we assume that $\mathbf{N}_i$ is the discrete multivariate Sarmanov distribution as shown in (11) with Negative Binomial GLM marginals.

Let $\boldsymbol{X}_i = \left(1, X_{i1}, \dots, X_{ip}\right)'$ be a column vector with values of the explanatory variables of person i and $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{pj})'$ be the parameter vector related to the random discrete variables N_{ij}. The link function $\boldsymbol{X}_i'\boldsymbol{\beta}_j = \ln(\mu_{ij})$ or $\mu_{ij} = \exp(\boldsymbol{X}_i'\boldsymbol{\beta}_j)$, with E_{ij} as the exposure. The jth marginal distribution of the ith person then becomes the mixed Poisson-Gamma distribution (or negative binomial):

$$\Pr(N_{ij} = n) = \frac{\Gamma(\alpha_j + n)}{n!\,\Gamma(\alpha_j)}\left(\frac{\alpha_j}{\alpha_j + E_{ij}\mu_{ij}}\right)^{\alpha_j}\left(\frac{E_{ij}\mu_{ij}}{\alpha_j + E_{ij}\mu_{ij}}\right)^{n}$$
$$= \frac{\Gamma(\alpha_j+n)}{n!\Gamma(\alpha_j)}\,\frac{{\alpha_j}^{\alpha_j}\exp\{n(\ln(E_{ij}+\boldsymbol{X}_i'\boldsymbol{\beta}_j))\}}{(\alpha_j+\exp\{\ln(E_{ij})+\boldsymbol{X}_i'\boldsymbol{\beta}_j\})^{\alpha_j+n}} \tag{15}$$

where, $\alpha_j > 0$ is the Gamma parameter. Thus $N_{ij} \sim NB(\alpha_j, \tau_{ij})$, where, $\tau_{ij} = \frac{\alpha_j}{\alpha_j+E_{ij}\mu_{ij}}$.

The kernel function is $\phi_{ij}(n_j) = e^{-n_j} - \mathcal{L}_{N_{ij}}(1),\ j = 1,2,3,4$, where, $\mathcal{L}_{N_{ij}}(1) = \left(\frac{\alpha_j}{\alpha_j+E_{ij}\mu_{ij}(1-e^{-1})}\right)^{\alpha_j}$.

Next proposition is described the boundary conditions of the dependent parameters $\omega_{j,k}$ where, $1 \le j < k \le 4$, $\omega_{j,k,h}$ where $1 \le j < k < h \le 4$, and $\omega_{1,2,3,4}$.

***Proposition* 2.** The following conditions must be fulfilled for all $i = 1, \dots, n$:

$$\max_{1\le j<k\le 4}\left\{\frac{-1}{M_{ij}M_{ik}}, \frac{-1}{m_j m_k}\right\} \le \omega_{j,k} \le \min_{1\le j<k\le 4}\left\{\frac{-1}{M_{ij}m_k}, \frac{-1}{m_j M_{ik}}\right\},$$
$$\max_{1\le j<k<h\le 4}\left\{\frac{-1}{\prod_{n=1}^{3}M_{il_n}} - \frac{\omega_{j,k}}{M_{ih}} - \frac{\omega_{j,h}}{M_{ik}} - \frac{\omega_{h,k}}{M_{ij}}, \frac{-1}{m_j m_k M_{ih}} - \frac{\omega_{j,k}}{M_{ih}} - \frac{\omega_{j,h}}{m_k} - \frac{\omega_{k,h}}{m_j}\right\} \le \omega_{j,k,h},$$

$$\omega_{j,k,h} \leq \min_{1\leq j<k<h\leq 4}\left\{\frac{-1}{\prod_{n=1}^{3} m_{l_n}} - \frac{\omega_{j,k}}{M_{ih}} - \frac{\omega_{j,h}}{M_{ik}} - \frac{\omega_{h,k}}{M_{ij}}, \frac{-1}{M_{ij}M_{ik}m_h} - \frac{\omega_{j,k}}{m_h} - \frac{\omega_{j,h}}{M_{ik}} - \frac{\omega_{k,h}}{M_{ij}}\right\},$$

$$\max_{\substack{1\leq j<k<p\leq 4\\ h=10-j-k-p}}\left\{\begin{array}{l}\frac{-1}{\prod_{l=1}^{4} m_l} - \sum_{\substack{1\leq j<k\leq 4\\ p,h\neq j,k}} \frac{\omega_{j,k}}{m_p m_h} - \sum_{\substack{1\leq j<k<p\leq 4\\ h=10-j-k-p}} \frac{\omega_{j,k,p}}{m_h}, \frac{-1}{\prod_{l=1}^{4} M_{il}} - \sum_{\substack{1\leq j<k\leq 4\\ p,h\neq j,k}} \frac{\omega_{j,k}}{M_{ip}M_{ih}}\\ - \sum_{\substack{1\leq j<k<p\leq 4\\ h=10-j-k-p}} \frac{\omega_{j,k,p}}{M_{ih}}, \frac{-1}{m_j m_k M_{ip} M_{ih}} - \frac{\omega_{j,k}}{M_{ip}M_{ih}} - \frac{\omega_{j,p}}{m_k M_{ih}} - \frac{\omega_{j,h}}{m_k M_{ip}} - \frac{\omega_{k,p}}{m_j M_{ih}} - \frac{\omega_{k,h}}{m_j M_{ip}}\\ - \frac{\omega_{p,h}}{m_j m_k} - \frac{\omega_{j,k,p}}{M_{ih}} - \frac{\omega_{j,k,h}}{M_{ip}} - \frac{\omega_{j,p,h}}{m_k} - \frac{\omega_{k,p,h}}{m_j}\end{array}\right\} \leq \omega_{1,2,3,4},$$

$$\omega_{1,2,3,4} \leq \min_{\substack{1\leq j<k<p\leq 4\\ h=10-j-k-p}}\left\{\begin{array}{l}\frac{-1}{m_j M_{ik} M_{ip} M_{ih}} - \frac{\omega_{j,k}}{M_{ip}M_{ih}} - \frac{\omega_{j,p}}{M_{ik}M_{ih}} - \frac{\omega_{j,h}}{M_{ik}M_{ip}} - \frac{\omega_{k,p}}{m_j M_{ih}} - \frac{\omega_{k,h}}{m_j M_{ip}} - \frac{\omega_{p,h}}{m_j M_{ik}}\\ - \frac{\omega_{j,k,p}}{M_{ih}} - \frac{\omega_{j,k,h}}{M_{ip}} - \frac{\omega_{j,p,h}}{M_{ik}} - \frac{\omega_{k,p,h}}{m_j}, \frac{-1}{M_{ij}m_k m_p m_h} - \frac{\omega_{j,k}}{m_p m_h} - \frac{\omega_{j,p}}{m_k m_h} - \frac{\omega_{j,h}}{m_k m_p} - \frac{\omega_{k,p}}{M_{ij}m_h}\\ - \frac{\omega_{k,h}}{M_{ij}m_p} - \frac{\omega_{p,h}}{M_{ij}m_k} - \frac{\omega_{j,k,p}}{m_h} - \frac{\omega_{j,k,h}}{m_p} - \frac{\omega_{j,p,h}}{m_k} - \frac{\omega_{k,p,h}}{M_{ij}}\end{array}\right\},$$

where, $m_j = -\left(\frac{\alpha_j}{\alpha_j + E_{ij}\mu_{ij}(1-e^{-1})}\right)^{\alpha_j}$, $M_{ij} = 1 - \left(\frac{\alpha_j}{\alpha_j + E_{ij}\mu_{ij}(1-e^{-1})}\right)^{\alpha_j}$, $j = 1,2,3,4, i = 1, \dots, n,$ and, $\omega_{j,k} = \omega_{k,j}$.

2.3.2 Model 2

First, let N_i be a d-variate Poisson distribution with independent marginal, which is mixed with a d-variate Sarmanov distribution with Gamma marginal. This model is a form of that proposed by Abdallah et al. (2016) in the bivariate version with a distinct parameterization. Then, *Gamma* (α_j, α_j) marginals are used for the Sarmanov distribution, and the p.f. of the mixed distribution exposure is given by solving the following equation:

$$\mathrm{P}(N_i = n) = \int_0^\infty \dots \int_0^\infty \left(\prod_{j=1}^{d} e^{-E_{ij}\mu_{ij}\theta_j} \frac{(E_{ij}\mu_{ij}\theta_j)^{n_j}}{n_j!}\right) \times h_{Sarm}(\theta_1, \dots, \theta_d) d\theta_1, \dots, d\theta_d \quad (16)$$

where, h_{Sarm} is obtained in (1), and h_j is the p.d.f. of the mixing marginal r.v. $\Theta_j \sim Gamma\,(\alpha_j, \alpha_j)$, $j = 1, \dots, d$, and the kernel function $\phi_j(\theta_j) = e^{-\theta_j} - \mathcal{L}_{\Theta_j}(1)$, where, $\mathcal{L}_{\Theta_j}$ is the Laplace transform of Θ_j. The formula p.f. $\Pr(N_i = n)$ is also of the Sarmanov form, while the kernel functions are more complicated, as shown in the next proposition.

***Proposition* 3.** Assume that **N** is a multivariate random count variable. If the multivariate Poisson distribution is blended via the means of the $\theta_j s$ with the independent marginal $N_j \sim Poisson\,(\mu_j \theta_j)$, $j = 1, \dots, d$, together with a multivariate Sarmanov distribution with *Gamma* (α_j, α_j) marginals and exponential kernels, then the mixed distribution of N has the p.f.:

$$\Pr(\mathbf{N}=\mathbf{n}) = \prod_{j=1}^{d} \Pr(N_j = n_j) \times \left[1 + \sum_{k=2}^{d} \sum_{1\le j_1<\cdots<j_k\le d} \omega_{j_1\ldots j_k} \times \prod_{\ell=1}^{k} \left(\left(\frac{\alpha_{j\ell}+\mu_{j\ell}}{\alpha_{j\ell}+\mu_{j\ell}+1}\right)^{\alpha_{j\ell}+n_{j\ell}} - \left(\frac{\alpha_{j\ell}}{\alpha_{j\ell}+1}\right)^{\alpha_{j\ell}}\right)\right] \quad (17)$$

with marginals $N_j \sim NB\left(\alpha_j, \tau_j\right)$ and $\tau_j = \frac{\alpha_j}{\alpha_j+\mu_j}, j = 1, \ldots, d.$

In the four-variate case, the p.f. with the exposure for Model 2 becomes:

$$\Pr(\mathbf{N}_i=\mathbf{n}) = \prod_{i=1}^{4} \Pr(N_{ij} = n_j) \left[1 + \sum_{1\le j_1<j_2\le 4} \omega_{j_1,j_2} \prod_{k=1}^{2} \left(\left(\frac{\alpha_{j_k}+E_{ij_k}\mu_{ij_k}}{\alpha_{j_k}+E_{ij_k}\mu_{ij_k}+1}\right)^{\alpha_{j_k}+n_{j_k}} - \left(\frac{\alpha_{j_k}}{\alpha_{j_k}+1}\right)^{\alpha_{j_k}}\right) + \sum_{1\le j_1}^{2} \sum_{<j_2}^{3} \sum_{<j_3\le 4}^{4} \omega_{j_1,j_2,j_3} \prod_{\ell=1}^{3} \left(\left(\frac{\alpha_\ell+E_{i\ell}\mu_{i\ell}}{\alpha_\ell+E_{i\ell}\mu_{i\ell}+1}\right)^{\alpha_\ell+n_\ell} - \left(\frac{\alpha_\ell}{\alpha_\ell+1}\right)^{\alpha_\ell}\right) + \omega_{1,2,3,4} \prod_{j=1}^{4} \left(\left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}+1}\right)^{\alpha_j+n_j} - \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}\right)\right] \quad (18)$$

where, as before, the marginal $N_{ij} \sim NB\left(\alpha_j, \tau_{ij}\right)$ with $\tau_{ij} = \frac{\alpha_j}{\alpha_j+E_{ij}\mu_{ij}}, j = 1,2,3,4.$

The limitations of dependent parameters are the same as those given in Proposition 2, with the maxima: $M_{ij} = \left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}+1}\right)^{\alpha_j} - \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}$, while in the minimum case $m_j = -\left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}$.

2.3.3 Model 3

In the third model, $\mathbf{N}_j$ is assumed to follow a discrete multivariate Sarmanov distribution with Poisson marginals mixed with independent Gamma distributions as:

$$\Pr\left(\mathbf{N}_i = \mathbf{n}\right) = \int_0^\infty \ldots \int_0^\infty \Pr_{Sarm}(\mathbf{N}_i=\mathbf{n}) \left(\prod_{j=1}^{d} h_j(\theta_j)\right) d\theta_1, \ldots, d\theta_d \quad (19)$$

where, $\Pr_{Sarm}(\mathbf{N}_i=\mathbf{n})$ is the discrete multivariate Sarmanov distribution (11) with Poisson marginals defined by

$$\Pr_{Sarm}(\mathbf{N}_i=\mathbf{n}) = \left(\prod_{j=1}^{d} e^{-E_{ij}\mu_{ij}\theta_j} \frac{\left(E_{ij}\mu_{ij}\theta_j\right)^{n_j}}{n_j!}\right) \times \left(1 + \sum_{k=2}^{d} \sum_{1\le j_1<\cdots<j_k\le d} \omega_{j_1\ldots j_k} \phi_{j_1}(n_{j_1}) \ldots \phi_{j_k}(n_{j_k})\right) \quad (20)$$

with the kernel function $\phi_{ij}(n_{ij}) = e^{-n_{ij}} - \mathcal{L}_{N_{ij}}(1)$, where, $\mathcal{L}_{N_{ij}}$ is the Laplace transformation of a Poisson distribution with mean $E_{ij}\mu_{ij}\theta_j$. The mixing distributions h_j have $Gamma\left(\alpha_j, \alpha_j\right)$ distributions. The explicit formula of this model is shown in Proposition 4.

***Proposition* 4.** Let **N** be a multivariate discrete random variable. The discrete multivariate Sarmanov distribution comprises Poisson marginals mixed with Gamma distribution and exponential kernels. The means of Poisson marginals are assumed as $\mu_j\theta_j, j = 1, \ldots, d$. Then the distribution of **N** has the p.f.:

$$\Pr(\mathbf{N}=\mathbf{n}) = \prod_{j=1}^{d} \Pr\left(N_j = n_j\right) \times \left[1 + \sum_{k=2}^{d} \sum_{1\le j_1<\cdots<j_k\le d} \omega_{j_1\cdots j_k} \prod_{l=1}^{k} \left(e^{-n_{j_l}} - \left(\frac{\alpha_{j_l}+\mu_{j_l}}{\alpha_{j_l}+\mu_{j_l}(2-e^{-1})}\right)^{\alpha_{j_l}+n_{j_l}}\right)\right] \quad (21)$$

where, the marginals $N_{ij} \sim NB\left(\alpha_j, \tau_{ij}\right)$ with $\tau_{ij} = \frac{\alpha_j}{\alpha_j+\mu_{ij}}, j = 1, \ldots, d.$

In the four-variate case, the p.f. with the exposure for Model 3 is:

$$\Pr(\mathbf{N}_i=\mathbf{n}) = \prod_{i=1}^{4}\Pr\left(N_{ij}=n_j\right) \times \left[1+\sum_{1\le j_1<j_2\le 4}\omega_{j_1,j_2}\prod_{l=1}^{2}\left(\left(e^{-n_{j_l}}-\left(\frac{\alpha_{j_l}+E_{ij}\mu_{j_l}}{\alpha_{j_l}+E_{ij}\mu_{j_l}(2-e^{-1})}\right)^{\alpha_{j_l}+n_{j_l}}\right)\right)\right]+$$

$$\sum_{1\le j_1}^{2}\sum_{<j_2}^{3}\sum_{<j_3\le 4}^{4}\omega_{j_1,j_2,j_3}\prod_{l=1}^{3}\left(\left(e^{-n_{j_l}}-\left(\frac{\alpha_{j_l}+E_{ij}\mu_{j_l}}{\alpha_{j_l}+E_{ij}\mu_{j_l}(2-e^{-1})}\right)^{\alpha_{j_l}+n_{j_l}}\right)\right)+\omega_{1,2,3,4}\prod_{j=1}^{4}\left(\left(e^{-n_j}-\left(\frac{\alpha_j+E_{ij}\mu_j}{\alpha_j+E_{ij}\mu_j(2-e^{-1})}\right)^{\alpha_j+n_j}\right)\right) \quad (22)$$

where, as above, the marginals $N_{ij}\sim NB(\alpha_j,\tau_{ij})$ with $\tau_{ij}=\frac{\alpha_j}{\alpha_j+E_{ij}\mu_{ij}}, j=1,2,3,4.$

The ranges of dependent parameters are the same as those given in Proposition 2 with the maxima: $M_{ij}=1-\left(\frac{\alpha_j+E_{ij}\mu_j}{\alpha_j+E_{ij}\mu_j(2-e^{-1})}\right)^{\alpha_j}$, while for the minimum case $m_j=e^{-n_j}-\left(\frac{\alpha_j+E_{ij}\mu_j}{\alpha_j+E_{ij}\mu_j(2-e^{-1})}\right)^{\alpha_j+n_j}$.

All three ranges also rely on each customer i, with m_{ij} acquiring some value in $\mathbb{N}$ and not by letting $n_{ij}\to\infty$ as before.

3. Parameter Estimation for the Model 1, 2 and 3

All three models based on the Sarmanov distribution are complicated, making the estimation of all parameters of the three models difficult. Therefore, estimating all parameters of the model based on the Sarmanov distribution requires splitting it into two parts as the marginal distribution and the dependence structure. Suppose the NB GLM used for marginals is the accurate model. In such a situation, the variance-covariance structure of the dependent variables should not impact the point estimators of the β parameters, which are responsible for estimating the expected claims frequency. The point estimators of the β parameters produced from independent Poisson GLM, independent NB GLM, and multivariate Sarmanov-based models with NB GLM marginals should, thus, be substantially equal with a sufficiently large sample. The parameter estimation method was from Bolancé and Vernic (2019). They considered the trivariate Sarmanov models.

Assuming these assertions, the following strategy was proposed:
(i) The parameters $\beta_j, j=1,2,3,4$, of the marginal distribution were estimated and represented by $\widehat{\boldsymbol{\beta}}=(\hat{\beta}_1,\hat{\beta}_2,\hat{\beta}_3,\hat{\beta}_4)$.
(ii) The parameters of the dependent structure were also estimated as: $\alpha_1,\alpha_2,\alpha_3,\alpha_4,\omega_{1,2},\omega_{1,3},\omega_{1,4},\omega_{2,3},\omega_{2,4},\omega_{3,4},\omega_{1,2,3},\omega_{1,2,4},\omega_{1,3,4},\omega_{2,3,4}$ and $\omega_{1,2,3,4}$.

The maximum likelihood estimation (MLE) was applied to estimating the parameters $\beta_j, j=1,2,3,4$, for each univariate marginal distribution. If the NB GLM model is accurate, the estimations will be without bias. The initial estimated vector $\widehat{\boldsymbol{\alpha}}^0=(\hat{\alpha}_1^0,\hat{\alpha}_2^0,\hat{\alpha}_3^0,\hat{\alpha}_4^0)$, which serve as the starting points for the subsequent iterative algorithm and is also derived from the MLE of the univariate NB GLM.

The following two conditional likelihoods: $L(\widehat{\boldsymbol{\omega}}|\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}})$ and $L(\widehat{\boldsymbol{\alpha}}|\widehat{\boldsymbol{\omega}},\widehat{\boldsymbol{\beta}})$, where, $\widehat{\boldsymbol{\alpha}}=(\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3,\hat{\alpha}_4)$ and $\widehat{\boldsymbol{\omega}}=(\hat{\omega}_{1,2},\hat{\omega}_{1,3},\hat{\omega}_{1,4},\hat{\omega}_{2,3},\hat{\omega}_{2,4},\hat{\omega}_{3,4},\hat{\omega}_{1,2,3},\hat{\omega}_{1,2,4},\hat{\omega}_{1,3,4},\hat{\omega}_{2,3,4},\hat{\omega}_{1,2,3,4})$ are two vectors that provide us an estimates for the variance and covariance matrices, with calculated parameters established to estimate the dependent parameters.

By maximizing the conditional likelihood with the supplied parameters $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}^0$, the initial values for the dependency parameters $\widehat{\boldsymbol{\omega}}^0 = (\widehat{\omega}_{1,2}^0, \widehat{\omega}_{1,3}^0, \widehat{\omega}_{1,4}^0, \widehat{\omega}_{2,3}^0, \widehat{\omega}_{2,4}^0, \widehat{\omega}_{3,4}^0, \widehat{\omega}_{1,2,3}^0, \widehat{\omega}_{1,2,4}^0, \widehat{\omega}_{1,3,4}^0, \widehat{\omega}_{2,3,4}^0, \widehat{\omega}_{1,2,3,4}^0)$ are produced. The parameter space must also be established for the existing constraints on the ω values by identifying the signs of the parameters within $\widehat{\boldsymbol{\omega}}^0$ and their corresponding ranges. It is crucial that the procedure works for any iteration l. Sample estimators based on (8) - (10) were used to determine the signs and establish the range of variation by utilizing Proposition 2. The remaining steps of the method were split into two steps and beginning with $l = 0$:

***Step* 1:** Find l by maximizing the conditional likelihood $L(\widehat{\boldsymbol{\omega}}^l | \widehat{\boldsymbol{\alpha}}^l, \widehat{\boldsymbol{\beta}})$ within the parameter space derived from the estimated signs and intervals (refer to the procedure outlined above).

***Step* 2:** Get $\widehat{\boldsymbol{\alpha}}^{l+1}$ from maximizing the conditional likelihood $L(\widehat{\boldsymbol{\alpha}}^{l+1} | \widehat{\boldsymbol{\omega}}^l, \widehat{\boldsymbol{\beta}})$. In Step 2 if this inequality $L(\widehat{\boldsymbol{\alpha}}^{l+1} | \widehat{\boldsymbol{\omega}}^l, \widehat{\boldsymbol{\beta}}) \leq L(\widehat{\boldsymbol{\omega}}^l | \widehat{\boldsymbol{\alpha}}^l, \widehat{\boldsymbol{\beta}})$ is true, the solution is analyzed from the previous iteration; otherwise, continue to Step 1 for the following iteration.

The estimated parameters gained through this suggested two-step procedure as start values were used to carry out the complete log-likelihood maximization for the models based on the Sarmanov distribution using a finite-difference approximation of the "optim()" function in R. As previously mentioned, during the optimization process the estimated dependency parameters from the $\widehat{\boldsymbol{\omega}}$ vector must fall within the boundaries defined by equations (8) - (10). These boundaries shift with each iteration of the optimization algorithm, along with the values of the estimated coefficients $\widehat{\boldsymbol{\beta}}$ and the estimated scale parameters $\widehat{\boldsymbol{\alpha}}$.

4. Numerical Application

4.1 Data Set

This research studied the claim frequencies of four insurance policyholder types. Four dependent variables were assessed as the number of auto insurance claims involving only property damage (PD), bodily injury (BI), and third-party (TP), and home insurance (H). All the claims concerned mandatory legal liability.

Claim data for the three risk categories PD, BI, and H were sourced from numerous national companies in Spain, consisting of 80,924 insurers who held both auto and home contracts between 2006 and 2015. This duration of 10 years represented the maximum available for each portfolio of the corporation. Results gave a 10-year snapshot of auto and house insurance portfolios, even though only some of the customers had policies for the whole 10-year period at the time of data extraction. Customers that moved between insurance companies during the period examined were tagged with the same identity in the database, allowing claims to be counted on all their vehicle and house policies. The number of third-party claims was generated using the negative binomial distribution (Bülbül and Baykal, 2016). Claim frequencies of TP were not greater than the claim frequencies of PD for each contract. When PD claims were zero, TP claims were also zero. Data characteristics are presented in Table 2.

Data collected over a 10-year period contained a wide range of alterations owing to shifts in product or coverage offerings, consumer habits, and political-social-economic environments. Therefore, to ensure the consistency of a specific sort of aggregate accident rate across all policies in an insurance line purchased by a consumer, the legal culpability for both vehicle and home lines was examined for consistency across the 10-year period. Customer exposure was computed as the number of days the contract was valid for each insurance of the same kind during the observed time. For customers with many policies in the same insurance line, the contract lengths were summed without considering whether

or not the policies were concurrently valid (i.e. holding more than one policy at the same time).

Table 2. Frequency of claims.

Number of claims	0	1	2	3	4	5	≥ 6
Auto property damage	68,679	7,850	2,581	941	422	184	267
Auto bodily injury	78,427	2,249	215	24	6	0	3
Auto third party	75,875	4,558	435	52	3	1	0
Home	69,109	8,707	2,099	647	208	88	66

Results in Table 2 show the frequency of claims for each category of risk. For BI and TP, policyholders reported a maximum of six and five claims, respectively with the highest values for PD and H at 40 and 23, respectively. The total number of active days for all policies contracted during the examined period was adjusted by the exposure for each client in each insurance line during the full 10-year period. The maximum potential exposure in each category was equal to the number of signed policies multiplied by the number of days from January 1, 2006 to December 31, 2015. In practical terms, relative exposure was derived by dividing the overall exposure by the total number of days during the study period, with exposures for the auto and home insurance lines differing.

Table 3. Dependence analysis (p-values).

Chi-Squared Statistic (p-value)				
	PD	BI	TP	H
PD		64090(0.000)	378784(0.000)	233.67(0.239)
BI	64090(0.000)		35049(0.000)	38.523(0.375)
TP	378784(0.000)	35049(0.000)		121.94(0.001)
H	233.67(0.239)	38.523(0.375)	121.94(0.001)	
Pearson (p-value)				
	PD	BI	TP	H
PD		0.452(0.000)	0.865(0.000)	0.016(0.000)
BI	0.452(0.000)		0.531(0.000)	0.012(0.001)
TP	0.865(0.000)	0.531(0.000)		0.026(0.000)
H	0.016(0.000)	0.012(0.001)	0.026(0.000)	
Kendall (p-value)				
	PD	BI	TP	H
PD		0.395(0.000)	0.653(0.000)	0.011(0.001)
BI	0.395(0.000)		0.512(0.000)	0.013(0.000)
TP	0.653(0.000)	0.512(0.000)		0.029(0.000)
H	0.011(0.001)	0.013(0.000)	0.029(0.000)	
Spearman (p-value)				
	PD	BI	TP	H
PD		0.403(0.000)	0.665(0.000)	0.011(0.001)
BI	0.403(0.000)		0.513(0.000)	0.013(0.000)
TP	0.665(0.000)	0.513(0.000)		0.029(0.000)
H	0.011(0.001)	0.013(0.000)	0.029(0.000)	

The relationship between the number of claims for each risk type, calculated using four distinct statistics is shown in Table 3, with the p-value showing the significance for each statistic also provided. The statistics employed included Chi-Squared (top) to test for dependency between two categorical variables, the Kendall and Spearman coefficients (third and fourth, respectively) to test for non-linear correlation, and the Pearson coefficient (second) to test for linear dependence. As shown in Table 3, all the statistics suggested dependency among the different types of accident rates, with the sole exception being the Chi-Squared statistic for two pairs: (PD, H) and (BI, H). A positive and significant correlation suggested that the likelihood of filing claims in one insurance category, given that claims had already been recorded in another category, differed from the likelihood that no claims had been filed in the other category. This must be considered when calculating the client's risk. The predictive analysis, described later, examined

how the joint and conditional probabilities associated with various customer profiles varied when taking into account the various dependence structures represented by the three alternative Sarmanov-based models.

Results in Table 4 show the explanatory factors (covariates) used, along with their means and variances. These variables were based on the latest available information for each client and were used consistently across the three dependent variables. The covariates were chosen based on customer characteristics and consisted of

(i) Gender (X_1): This variable is not used for calculating insurance premiums in the Spanish market but included in the risk analysis, (Gender of the policyholder: X_1=1 if woman, X_1= 0 if man).
(ii) Area of residence: There are two variables defined as
- The size of cities (X_2), big cities (Barcelona and Madrid), and others,
  (Area of residence: X_2 =1 if big city, X_2= 0 if other),
- The specific weather (X_3), the north of Spain and other,
  (Area of residence: X_3=1 if north, X_3= 0 if other).

(iii) Age (X_4): The policyholder's age is considered as a whole number.
(iv) Other policies (X_5): The policyholder had contracted policies in other lines, e.g., accident insurance, life insurance, pension plans, etc., (Client has other policies in the same company: X_5=1 if yes, X_5= 0 if no).

The changes in covariates over the 10 years were mostly irrelevant. For instance, gender did not change, and most policyholders had other policies throughout the period. The only variables with significant changes were those related to the area of residence. However, after analyzing migration figures between Spanish provinces during the study period, the probability of an individual changing regions (and thus areas) was around 0.007 and considered insignificant.

Table 4. Explanatory variables of the three proposed models (values represent the most recent information available for each policyholder).

Variable	Description	Mean	Variance
X_1	Gender of the policyholder: =1 if woman, = 0 if man	0.237	0.181
X_2	Area of residence: =1 if big city, = 0 if other	0.197	0.158
X_3	Area of residence: =1 if north, = 0 if other	0.289	0.205
X_4	Age of policyholder	53.242	172.123
X_5	Client has other polices in the same company: =1 if yes, = 0 if no	0.219	0.430

4.2 Parameter Estimation Results

Table 5 displays the estimated parameter outcomes of the four-variate NB GLM, incorporating the relationship between the numbers of claims in various kinds of insurance policies. The model had independent marginals, and the estimated parameters obtained by fitting three independent univariate NB GLMs are also shown in Table 5. For both models, the estimated parameter values in the vectors $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$ corresponding to the covariates were comparable. Both the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) of the estimated four-variate model with dependent marginals were lower, indicating that the inclusion of dependency between marginals improved the fit.

The estimated parameters of the three models based on the four-variate Sarmanov distributions with NB GLM marginals are shown in Table 6. Table 10 in the Appendix also displays the boundaries for the

parameters $\omega_{1,2}, \omega_{1,3}, \omega_{1,4}, \omega_{2,3}, \omega_{2,4}, \omega_{3,4}, \omega_{1,2,3}, \omega_{1,2,4}, \omega_{1,3,4}, \omega_{2,3,4}$ and $\omega_{1,2,3,4}$, as well as the linear correlation coefficients. In Table 10, the experimental linear correlations fell within these bounds.

Table 5. Estimation results of the four-variate negative binomial generalized linear model assuming dependence (top) and independence (bottom).

	Dependent marginal distributions							
	Estimated parameters				Standard errors			
	PD	BI	TP	H	PD	BI	TP	H
Intercept	-7.2230	-9.6112	-9.4368	-7.5229	0.0439	0.0925	0.0638	0.0448
X_1	1.3267	-1.6637	0.4475	-0.3395	0.0210	0.0775	0.0310	0.0238
X_2	0.9506	0.1851	0.7340	1.4707	0.0245	0.0697	0.0340	0.0231
X_3	0.7997	1.7734	0.0594	-1.9351	0.0214	0.0454	0.0337	0.0316
X_4	-0.0684	-0.0554	-0.0344	-0.0423	0.0008	0.0018	0.0012	0.0008
X_5	0.2413	-0.3287	-1.2759	0.3803	0.0121	0.0327	0.0442	0.0143
α	0.6754							
Log-Likelihood = -121,197.8								
AIC = 242,407.6								
BIC = 242,366.2								
	Independent marginal distributions							
	Estimated parameters				Standard errors			
	PD	BI	TP	H	PD	BI	TP	H
Intercept	-6.4529	-7.2500	-10.5004	-7.0337	0.0464	0.0941	0.0715	0.0460
X_1	-0.2508	-0.9589	0.5307	-0.8099	0.0237	0.0582	0.0339	0.0253
X_2	0.4941	-4.1915	1.2174	-1.9984	0.0247	0.2297	0.0380	0.0336
X_3	-1.3444	-1.5011	0.7942	-0.8453	0.0243	0.0569	0.0352	0.0231
X_4	-0.0557	-0.0762	-0.0302	-0.0386	0.0008	0.0019	0.0013	0.0008
X_5	-0.6976	-0.3015	0.3884	0.0931	0.0158	0.0324	0.0173	0.0141
α	0.5060	0.5066	0.5000	0.4675				
Log-Likelihood = -160,869.2								
AIC = 321,750.4								
BIC = 321,709.0								

In Table 6, Model 3 had the best fit, according to AIC and BIC. According to Bolancé and Vernic (2019), the best-fit model was Model 2. Therefore, the trivariate Sarmanov models improved the trivariate NB GLM model. Furthermore, Models 2 and 3 improved Model 1. It was similar to four-variate models. All three models, derived from the Sarmanov distribution, produced comparable outcomes in terms of the significance of the parameters $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$. These findings suggest that the kind of coverage affects the variables' impact. For instance, the impact of gender is negative and significant when it comes to property damage claims, bodily injury, and home, meaning that women tend to file fewer claims of this nature. Residing in big cities has a positive impact on the quantity of property damage and third-party claims, but it negatively influences the number of bodily injury and home claims. On the other hand, residing in the northern part of the country adversely affects both types of claims. As age increases, the number of claims in the auto and home line decreases. At last, purchasing additional products from the same company impacts both the auto line and the home line, resulting in a positive effect on third-party and home claims and a negative effect on property damage and bodily injury claims.

The expected dependence between the examined coverages was an additional distinction between the four estimated multivariate models. Each Sarmanov model was linked to a different dependency structure, with the relevance of dependence parameter values $\omega_{j,k}, 1 \leq j < k \leq 4, \omega_{j,k,h}, 1 \leq j < k < h \leq 4$ and $\omega_{1,2,3,4}$ varying among the models. For each individual i, Model 1 suggested direct dependency between the NB random variables $N_{ij} \sim NB(\alpha_j, \tau_{ij})$, $j = 1,2,3,4$ with $\tau_{ij} = \frac{\alpha_j}{\alpha_j + \mu_{ij}}$, while Model 2 considered the dependence of the unobserved Gamma random variable $\Theta_j \sim Gamma(\alpha_j, \alpha_j)$. Finally, in Model 3, the

relationship between the Poisson random variables was presupposed as $\widetilde{N}_{ij} \sim Poisson(\mu_{ij})$, where, $N_{ij} = \widetilde{N}_{ij}\Theta_{ij}$.

Table 6. Results of the three model estimations using the NB GLM for marginals of the Sarmanov distributions.

	Model 1							
	Estimated parameters				Standard errors			
	PD	BI	TP	H	PD	BI	TP	H
Intercept	-6.4529	-7.2500	-10.5004	-7.0337	0.0464	0.0941	0.0715	0.0460
X_1	-0.2508	-0.9589	0.5307	-0.8099	0.0237	0.0582	0.0339	0.0253
X_2	0.4941	-4.1915	1.2174	-1.9984	0.0247	0.2297	0.0380	0.0336
X_3	-1.3444	-1.5011	0.7942	-0.8453	0.0243	0.0569	0.0352	0.0231
X_4	-0.0557	-0.0762	-0.0302	-0.0386	0.0008	0.0019	0.0013	0.0008
X_5	-0.6976	-0.3015	0.3884	0.0931	0.0158	0.0324	0.0173	0.0141
$\omega_{1,2} = -0.6425, \omega_{1,3} = 2.6397, \omega_{1,4} = -0.1943, \omega_{2,3} = 0.6801, \omega_{2,4} = 0.5968, \omega_{3,4} = 0.0857,$ $\omega_{1,2,3} = 1.3866, \omega_{1,2,4} = -1.0895, \omega_{1,3,4} = -1.0538, \omega_{2,3,4} = -0.8753, \omega_{1,2,3,4} = 2.2396$								
Log-Likelihood = -160,869.2, AIC = 321,750.4, BIC = 321,709.0								
	Model 2							
	Estimated parameters				Standard errors			
	PD	BI	TP	H	PD	BI	TP	H
Intercept	-6.4530	-7.2500	-10.5005	-7.0336	0.0461	0.0942	0.0716	0.0459
X_1	-0.2507	-0.9589	0.5307	-0.8100	0.0236	0.0581	0.0338	0.0254
X_2	0.4941	-4.1916	1.2173	-1.9983	0.0248	0.2298	0.0380	0.0337
X_3	-1.3445	-1.5012	0.7942	-0.8454	0.0245	0.0568	0.0353	0.0232
X_4	-0.0558	-0.0762	-0.0302	-0.0386	0.0008	0.0018	0.0011	0.0007
X_5	-0.6976	-0.3015	0.3885	0.0931	0.0156	0.0325	0.0175	0.0142
$\omega_{1,2} = -0.8003, \omega_{1,3} = 5.0473, \omega_{1,4} = -0.3813, \omega_{2,3} = 0.7640, \omega_{2,4} = 1.0752, \omega_{3,4} = 0.4318,$ $\omega_{1,2,3} = 2.0763, \omega_{1,2,4} = 1.7522, \omega_{1,3,4} = 0.6192, \omega_{2,3,4} = -2.5218, \omega_{1,2,3,4} = -1.1423$								
Log-Likelihood = -161,809.1, AIC = 323,630.2, BIC = 323,588.8								
	Model 3							
	Estimated parameters				Standard errors			
	PD	BI	TP	H	PD	BI	TP	H
Intercept	-6.4530	-7.2500	-10.5005	-7.0336	0.0462	0.0941	0.0715	0.0460
X_1	-0.2507	-0.9589	0.5307	-0.8100	0.0238	0.0582	0.0338	0.0254
X_2	0.4941	-4.1916	1.2173	-1.9983	0.0247	0.2298	0.0380	0.0336
X_3	-1.3445	-1.5012	0.7942	-0.8454	0.0244	0.0568	0.0352	0.0231
X_4	-0.0558	-0.0762	-0.0302	-0.0386	0.0008	0.0019	0.0013	0.0008
X_5	-0.6976	-0.3015	0.3885	0.0931	0.0157	0.0324	0.0173	0.0141
$\omega_{1,2} = 0.5989, \omega_{1,3} = 7.4883, \omega_{1,4} = -0.7965, \omega_{2,3} = 2.2689, \omega_{2,4} = -2.7598, \omega_{3,4} = 0.1943,$ $\omega_{1,2,3} = -2.6922, \omega_{1,2,4} = -0.8146, \omega_{1,3,4} = 1.1925, \omega_{2,3,4} = 0.0764, \omega_{1,2,3,4} = -2.3407$								
Log-Likelihood = -160,104, AIC = 320,220, BIC = 320,178.6								

To examine the dependence patterns across the models, the mean of each individual correlation value was determined using formula (14) for the four-variate NB model and formula (3) for the Sarmanov models (Table 7). All three models showed the lowest dependent relationships between TP and H.

Table 7. Correlations derived from the three estimated models with four variables.

	Model 1			
	PD	BI	TP	H
PD	1.0000	-0.0111	0.0904	-0.0096
BI	-0.0111	1.0000	0.0071	0.0090
TP	0.0904	0.0071	1.0000	0.0025
H	-0.0096	0.0090	0.0025	1.0000
	Model 2			
	PD	BI	TP	H
PD	1.0000	-0.0568	0.0787	0.1619
BI	-0.0568	1.0000	0.2495	-0.0276
TP	0.0787	0.2495	1.0000	0.0453
H	0.1619	-0.0276	0.0453	1.0000

Table 7 continued…

	Model 3			
	PD	BI	TP	H
PD	1.0000	0.0081	0.2459	-0.0357
BI	0.0081	1.0000	0.0181	-0.0300
TP	0.2459	0.0181	1.0000	0.0052
H	-0.0357	-0.0300	0.0052	1.0000

4.3 Predictive Analytic

A predictive analytic was performed by splitting the dataset into two parts as the training sample and the test sample to test the accuracy of the models. This procedure was implemented to determine the best model according to the provided dataset and performed by dividing 80% of the data into training data and the remaining 20% of the data into test data. Then, the probability obtained from the test data was analyzed against the accident rates of the training data. In the training sample, policyholders with (0,0,0,0) had a probability of around 0.73 (Table 9 in the Appendix). If a model predicted a probability equal to or greater than 0.73 for policyholders with (0,0,0,0), these policyholders were considered as being well classified. This analysis was consistently applied to each group of policyholders to determine the percentage of accurate classifications within each group according to their respective accident rates. In every instance, a frequency of well-classified predictions closer to 1 indicated a more accurate prediction. Results in Table 8 show the frequency of the successes obtained for each calculated Sarmanov four-variate model, taking into account the circumstances where the predicted probability was higher than the corresponding relative frequency. The success percentage of reported claims (PD, BI, H, PD) = (>0,0,0,0) of Model 2 was zero (Table 8), indicating that no customer had predicted probability equal to or greater than 0.00175 (Table 9). The weighted mean of all frequencies for each model was given as the total row, while the annual exposure (365 days) given by the number contracted in each line was also considered when arriving at these results. When considering the total row, the best results were obtained for Model 1.

Table 8. Success percentage attained by the three Sarmanov four-variate models.

Reported claims (PD, BI, H, TP)	Sample frequency	Model 1	Model 2	Model 3
(0,0,0,0)	58,611	0.99974	1	0.99957
(>0,0,0,0)	142	0.14286	0	0
(0,>0,0,0)	5,796	0	0	0
(0,0,>0,0)	0	0	0	0
(0,0,0,>0)	9,906	0	0	0
(>0,>0,0,0)	449	0	0	0
(>0,0,>0,0)	0	0	0	0
(>0,0,0,>0)	20	0	0	0
(0,>0,>0,0)	2,635	0	0	0
(0,>0,0,>0)	951	0	0	0
(>0,>0,>0,0)	0	0	0	0
(>0,0,>0,>0)	1,476	0	0	0
(0,>0,>0,>0)	0	0	0	0
(>0,>0,>0,>0)	0	0	0	0
Total	80,924	0.72434	0.72427	0.72396

The result of this section can be applied to indicate the relation of the four risk types. The risk of claiming home insurance is related to claiming car insurance. The amount of risk for insurance can be calculated from the proposed model. Furthermore, the risk prediction can also be applied.

5. Conclusions

This research was studied about a risk assessment from a customer who held many policies of the same company. The way to assess the risk was to evaluate probability of claim frequencies of occurred

accidents. The accident insurances that we were interested were a motor insurance (property damage, bodily injury, and third-party) and a home insurance. Therefore, the discrete four-variate Sarmanov distributions with NB GLM were utilized to indicate the probability of four types of claim frequencies for accidents. The research findings can be summarized as follows: Proposition 1 formulated the closed form of the discrete four-variate Sarmanov distribution for the dependent parameters ($\omega_{1,2,3,4}$); Proposition 2 determined the restrictions on dependent parameters; Propositions 3 and 4 presented the explicit forms of Model 2 and Model 3, respectively. According to the Chi-Squared statistics, it was found that the number of claims in property damage (PD) was dependent on the number of claims in the home (H), and the number of claims in the bodily injury (BI) was related to the number of claims in the home (H). The maximum likelihood results indicated that the two mixed models incorporating the Sarmanov distribution (Models 3) significantly improved the fit. All three models showed the lowest dependent relationships between TP and H by calculating the mean of correlation. Additionally, Model 1 demonstrated the most effective predictive performance among the three models.

The advantage of closed forms of the discrete four-variate Sarmanov distributions was comfortable to calculate the probability of the frequency claims in the accident coverages. In comparison to simpler models such as the multivariate discrete Sarmanov distribution with NB GLM marginals, the numerical application revealed that the multivariate Sarmanov distribution-based mixing models substantially enhanced the model fit and the flexibility. The advantage of this model was that it could evaluate all four risks without having to evaluate each risk separately. In addition, these models were used to assess the number of claims produced by four different risk types. The experimental results displayed the risk of claiming car insurances depended on the claims of the home insurances. The behavior of the customers who have claimed the home insurances affects the claims of the car insurances.

Given the limitations of finding out the real data of customers who had four types of insurances with the same insurance company, the data needed to be generated in one risk line. Using a mixed dataset between the real reported claims from Spanish insurance and the generated data, the three suggested models were compared with the four-variate NB GLM.

For quantitative risk, the relationship between the auto and the home claims reported by customers was significant and positive, supporting the proposed models. The risk assessment by the insurers must consider the likelihood of both motor and home reported claims by each policyholder because customers who reported claims for the auto insurances also had the right to make claims for their home insurances. Thus, if an insurer intends to pursue a customer-focused strategy that includes an integrated premium, the total prospective risk premium will be higher than the premium earned assuming independence. The company must carefully verify whether the total sum of annual risk premiums falls below or exceeds the overall risk premium for the entire period, while the interrelation between certain aspects over a short-term (one year) might go unnoticed.

Insurance providers can leverage research effectively. The proposed model can be utilized to assess the probability of customers maintaining insurance coverage for up to four risks with the same company. Based on this assessment, the company can make informed decisions to raise or lower insurance premiums for individual customers. Modifying insurance premiums to align with the needs of the insured is an effective strategy to retain customers and prevent them from purchasing insurance policies from competing companies. Insurers should consider the risks of purchasing multiple insurance types with the same company. Moreover, if the customer has a good background, they can get various forms of insurance from the same company. If customers have a history of claims, they should purchase insurance from several companies.

The Sarmanov distribution can be used to solve problems in many fields of science, such as hydrology (Sarmanov, 1974) and entropy (Alawady et al., 2022).

Future works will consider a different kernel function of the Sarmanov distribution and apply it to the continuous Sarmanov distributions for claim severity problems. Another kernel function is $\phi(y) = y - \mathbb{E}[Y]$. The interesting continuous distributions are the Pareto and Weibull distribution because they are used in insurance claims.

Conflict of Interest
The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgments
This research was supported by the Development and Promotion of Science and Technology Talents Project (DPST) from the Institute for the Promotion of Teaching Science and Technology (IPST).

Appendix

List of abbreviations

BI	bodily injury
GLM	generalized linear model
GLMs	generalized linear models
H	home
NB	Negative binomial
NB GLM	Negative binomial generalized linear model
p.d.f.	probability density function
p.f.	probability function
p.f.s.	probability functions
p.m.f.	probability mass function
PD	property damage
r.v.	random variable
r.v.s.	random variables
TP	third-party
vs.	vice versa

Table 9. Customer counts for each combined accident rate in training, test, and overall samples.

(BI,PD,TP,H)	Training sample		Test sample		Full sample	
	Number	Frequency	Number	Frequency	Number	Frequency
(0,0,0,0)	46,889	0.72428	11,722	0.72425	58,611	0.72427
(>0,0,0,0)	114	0.00176	28	0.00173	142	0.00175
(0,>0,0,0)	4,637	0.07163	1,159	0.07161	5,796	0.07162
(0,0,>0,0)	0	0	0	0	0	0
(0,0,0,>0)	7,925	0.12241	1,981	0.12240	9,906	0.12241
(>0,>0,0,0)	359	0.00555	90	0.00556	449	0.00555
(>0,0,>0,0)	0	0	0	0	0	0
(>0,0,0,>0)	16	0.00025	4	0.00025	20	0.00025
(0,>0,>0,0)	2,108	0.03256	527	0.03256	2,635	0.03256
(0>,0,>0,0)	761	0.01175	190	0.01174	951	0.01175
(0,>0,0,>0)	0	0	0	0	0	0
(0,0,>0,>0)	1,181	0.01824	295	0.01823	1,476	0.01824
(>0,>0,>0,0)	0	0	0	0	0	0
(>0,>0,0,>0)	0	0	0	0	0	0
(0,>0,>0,>0)	422	0.00652	106	0.00655	528	0.00652
(>0,>0,>0,>0)	328	0.00507	82	0.00507	410	0.00507
Total	64,739	1	16,185	1	80,924	1

Table 10. Limits of dependent parameters and linear correlations in Sarmanov distribution-based models.

	Model 1		Model 2		Model 3	
ω_{12}	-0.9997	4.3378	-0.9997	4.3378	-0.9231	4.6911
ρ_{12}	-0.1596	0.6926	-0.0774	0.3359	-0.1162	0.5903
ω_{13}	-1.0026	4.1177	-1.0026	4.1177	-0.7828	4.6036
ρ_{13}	-0.0863	0.3545	-0.0528	0.2167	-0.0321	0.1886
ω_{14}	-1.0172	4.2632	-1.0172	4.2632	-0.8105	4.7663
ρ_{14}	-0.0277	0.1161	-0.0247	0.1035	-0.0564	0.3320
ω_{23}	-1.0028	4.1188	-1.0028	4.1188	-0.7831	4.6056
ρ_{23}	-0.0254	0.1042	-0.01116	0.04585	-0.0632	0.3718
ω_{24}	-1.0175	4.3047	-1.0175	4.3047	-0.8129	5.2926
ρ_{24}	-0.0814	0.3445	-0.0522	0.2210	-0.1116	0.7264
ω_{34}	-1.0146	4.0479	-1.0146	4.0479	-0.9946	4.7291
ρ_{34}	-0.0438	0.1747	-0.0354	0.1412	-0.0444	0.2113
ω_{123}	-1.0944	2.4174	-1.0944	2.4174	3.0196	4.7502
ω_{124}	-0.5704	1.6293	-0.5704	1.6293	-1.2067	0.6787
ω_{134}	-3.7222	1.4844	-3.7222	1.4844	-3.4517	4.7477
ω_{234}	-12.1990	14.1847	-12.1990	14.1847	-27.0343	8.2993
ω_{1234}	-24.7270	18.7063	-14.0164	25.8349	-10.9654	37.7003

References

Abdallah, A., Boucher, J.P., & Cossette, H. (2016). Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insurance: Mathematics and Economics, 68*, 120-133. https://doi.org/10.1016/j.insmatheco.2016.01.003.

Alawady, M.A., Barakat, H.M., Mansour, G.M., & Husseiny, I.A. (2022). Information measures and concomitants of k-record values based on Sarmanov family of bivariate distributions. *Bulletin of the Malaysian Mathematical Sciences Society*, *46*(1), Article 9. https://doi.org/10.1007/s40840-022-01396-9.

Alemany, R., Bolancé, C., Rodrigo, R., & Vernic, R. (2021). Bivariate mixed poisson and normal generalised linear models with Sarmanov dependence-an application to model claim frequency and optimal transformed average severity. *Mathematics*, *9*(1), 73. https://doi.org/10.3390/math9010073.

Barakat, H.M., Alawady, M.A., Husseiny, I.A., & Mansour, G.M. (2022). Sarmanov family of bivariate distributions: statistical properties - concomitants of order statistics - information measures. *Bulletin of the Malaysian Mathematical Sciences Society*, *45*(Suppl 1), 49-83. https://doi.org/10.1007/s40840-022-01241-z.

Bermudez, L., & Karlis, D. (2011). Bayesian multivariate Poisson models for insurance ratemaking. *Insurance: Mathematics and Economics, 48*(2), 226-236. https://doi.org/10.1016/j.insmatheco.2010.11.001.

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. https://doi.org/10.1016/j.insmatheco.2019.01.001.

Bolancé, C., & Vernic, R. (2020). Frequency and severity dependence in the collective risk model: An approach based on Sarmanov distribution. *Mathematics, 8*(9), 1400. https://doi.org/10.3390/math8091400.

Bolancé, C., Guillen, M., & Pitarque, A. (2020). A Sarmanov distribution with beta marginals: An application to motor insurance pricing. *Mathematics, 8*(11), 2020. https://doi.org/10.3390/math8112020.

Bolancé, C., Guillén, M., Pelican, E., & Vernic, R. (2008). Skewed bivariate models and nonparametric estimation for CTE risk measure. *Insurance: Mathematics and Economics*, 43(3), 386-393. https://doi.org/10.1016/j.insmatheco.2008.07.005.

Boucher, J.P., & Inoussa, R. (2014). A posteriori ratemaking with panel data. *Astin Bulleti*, *44*(3), 587-612. https://doi.org/10.1017/asb.2014.11.

Bülbül, S.E., & Baykal, K.B. (2016). Optimal bonus-malus system design in motor third-party liability insurance in Turkey: Negative binomial model. *International Journal of Economics and Finance, 8*(8), 205-211. https://doi.org/10.5539/ijef.v8n8p205.

Jeong, H., Tzougas, G., & Fung, T.C. (2023). Multivariate claim count regression model with varying dispersion and dependence parameters. *Journal of the Royal Statistical Society. Series A: Statistics in Society, 186*(1), 61-83. https://doi.org/10.1093/jrsssa/qnac010.

Johnson, N.L., Kotz, S., & Balakrishnan, N. (1997). *Discrete multivariate distributions*. Wiley, New York. ISBN: 0471128449.

Lee, M.L.T., (1996). Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics-Theory and Methods, 25*(6), 1207-1222. https://doi.org/10.1080/03610929608831759.

Moutanabbir, K., & Abdelrahman, H. (2022). Bivariate Sarmanov phase-type distributions for joint lifetimes modeling. *Methodology and Computing in Applied Probability, 24*(2) 1093-1118. https://doi.org/10.1007/s11009-021-09875-5.

Sarmanov, I.O. (1974). New forms of correlation relationships between positive quantities applied in hydrology. *Mathematical Models in Hydrology, 100*, 104-109.

Sarmanov, O.V. (1966). Generalized normal correlation and two-dimensional Frechet classes. *Doklady Akademii Nauk SSSR, 168*(1), 32-35.

Shi, P., & Valdez, E.A. (2014). Multivariate negative binomial models for insurance claim counts. *Insurance: Mathematics and Economics, 55*, 18-29. https://doi.org/10.1016/j.insmatheco.2013.11.011.

Vernic, R., Bolancé, C., & Alemany, R. (2022). Sarmanov distribution for modeling dependence between the frequency and the average severity of insurance claims. *Insurance: Mathematics and Economics, 102*, 111-125. https://doi.org/10.1016/j.insmatheco.2021.12.001.

Wang, L. (2023). *Rank-based multivariate Sarmanov for modeling dependence between loss reserves* (Master's thesis, pp 1-56). McMaster University, Canada.