Master in Actuarial and Financial Sciences

MASTER THESIS

Title: Parameter Estimation Optimization for the Sarmanov Distribution: A Comparative Analysis of Strategies

Author: Albert Martos Ramírez

Advisor: Catalina Bolancé Losilla

Academic year: 2022-2023

UNIVERSITAT DE BARCELONA

Facultat d'Economia i Empresa

Màster de Ciències Actuarials i Financeres

Faculty of Economics and Business

Universitat de Barcelona

Master thesis

Master in Actuarial and Financial Sciences

Parameter Estimation Optimization for the Sarmanov Distribution: A Comparative Analysis of Strategies

Author: Albert Martos Ramírez

Advisor: Catalina Bolancé Losilla

The content of this document is the sole responsibility of the author, who declares that he/she has not incurred plagiarism and that all references to other authors have been expressed in the text.

Abstract

This study focuses on the optimization of parameter estimation for the Sarmanov distribution, aiming to enhance the fit of the modeling with real data. A comparative analysis of different strategies is conducted to identify the most effective approaches. The research explores the use of partial derivatives associated with those parameters that are directly linked to the dependency structure and statistical techniques to estimate these parameters with a higher precision. By evaluating and comparing the results obtained from these strategies, valuable insights are gathered regarding their performance and applicability. The findings contribute to the optimization of the parameter estimation methods for the Sarmanov distribution, leading to improve its modeling and opening new fields of investigation.

Keywords: Sarmanov Distribution, Dependency, Frequency, Severity, Optimization

Resumen

Este estudio se centra en la optimización de la estimación de parámetros para la distribución de Sarmanov, con el objetivo de mejorar los resultados de la modelización con datos reales. Se lleva a cabo un análisis comparativo de diferentes estrategias para identificar el enfoque que genera un mejor ajuste. La investigación explora el uso de derivadas parciales asociadas a aquellos parámetros que están directamente vinculados a la estructura de la dependencia y de técnicas estadísticas para estimar estos parámetros con mayor precisión. Al evaluar y comparar los resultados obtenidos a partir de estas estrategias, se obtienen conclusiones determinantes sobre las estimaciones y su aplicabilidad. Los hallazgos contribuyen a la optimización de los métodos de estimación de parámetros para la distribución de Sarmanov, mejorando así futuras modelizaciones y a la apertura de nuevos campos de investigación.

Palabras clave: Distribución de Sarmanov, Dependencia, Frecuencia, Severidad, Optimización

Contents

Acknowledgements

I would like to express my heartfelt gratitude to all those who have contributed to the completion of this project. Without their support and guidance, this achievement would not have been possible. I extend my sincere thanks to:

- My supervisor, Catalina Bolancé, for their invaluable guidance, mentorship, and support throughout my master project journey. Their expertise, feedback, and patience have been instrumental in shaping the direction and quality of the research.

- Friends and family members who have provided unwavering support, encouragement, and motivation throughout my Master's thesis journey. Their belief in my abilities, understanding, and patience have been invaluable in keeping me motivated and focused. Especially, I wanted to express my sincere gratitude to Sergi Marquès for assisting me in parts of the project and last but not least Emma Busquets, Maria Buxó and Patrícia Fernández for their unconditional support they have given to me these past months.

I am deeply grateful to all those mentioned above and anyone else who has contributed in any way to the completion of my project.

1. Introduction

In the financial and actuarial sector, the most common modeling approach for calculating the pure premium value is to assume that both the number of claims and their respective costs act independently. This simplifies the procedure by considering the assumption of independence, making the process simpler, exporting great results and making the interpretations easier, which are, especially the latter, a key aspect in the insurance industry. The previous methodology is, in fact, adopted by a large number of insurance and financial entities.

However, this assumption of independence is not exactly true, as there may exist at least minimal dependence between both frequency and severity. To illustrate this from the perspective of motor insurance, in large urban areas, high claim frequencies are often associated with relatively low costs, while in secondary roads where regular routes are usually taken, high frequencies are associated with higher claim amounts. This implies that current models could improve their accuracy by considering the dependence between variables. Working with dependence introduces additional difficulties in the process of calculating the pure premium. Nevertheless, this does not mean that an effective methodology cannot be developed by considering the existence of dependency. This is where the Sarmanov distribution comes into play; proposing to use the Sarmanov distribution with exponential kernel to model this dependence.

This distribution, despite being a high-powered distribution with multiple uses and benefits, it is not exempt from being complex and with estimation problems, particularly with regard to the parameters associated with the dependency that also depends on the individuals distributions. The main criticism of this distribution is that it imposes limits on the dependency. Three strategies are analyzed that address different approaches to estimating dependency, which depend on the parameters of the distributions, exponential kernel functions, and Laplace transforms of the marginal distributions. In any case, the study of such strategies has been perfomed considering the formulas and parameter estimation processes given in Vernic et al. (2021).

In this paper, a thorough study has been made to establish a new methodology to define the dependency parameter by imposing a interval calculated using the partial derivatives of the random variables distributions. Not only the partial derivatives are included, but also the whole process of the parameters estimation has been taking into account to define the best strategy to model both random variables. Our aim is to facilitate future applications of the Sarmanov distribution by considering this new approach.

The remainder of the paper is organized as follows. Section 2 presents the methodology that has been used to achieve the results. In Section 3 the Sarmanov distribution is defined with all details including the definition of the problem itself and the basic formulas to consider when using this statistic distribution. In Section 4 the new approach is defined with all the calculations that have been necessary to get to the results. In Section 5, the strategies to model the dependency are defined and, with real data, are analyzed to find the one that performs the best estimations based on the log-likelihood. In Section 6 we conclude with the main points of the paper and future works that should be needed in order to achieve a better performance when using the Sarmanov distribution.

2. Prior work

This project discusses the possibility to optimize and improve the performance of the Sarmanov distribution for the number of claims and their respective cost. A comprehensive review of the literature on the Sarmanov distribution to establish a solid understanding of its concepts is necessary to achieve these results.

To conduct the numerical analyses, a database containing real policy data provided by an insurance entity, including information on the number of claims and their respective costs, has been used. Adjustments have been made to address issues of outliers and repeated values in the claim costs due to the presence of deductibles. The cost values associated with policies with deductibles have been modified by generating random values from a Bivariate Gamma distribution considering the fact that costs can vary depending on the number of claims. The estimation of the parameters for each distribution has been calculated based on the remaining observations and using statistical methods.

To conduct the analysis, R Studio software has been used. In R Studio, codes have been generated to iteratively reproduce the Sarmanov process, which facilitates the proper development of the project. The parameters of the Sarmanov distribution are estimated using appropriate techniques, such as maximum likelihood estimation (MLE), also using R. The estimated parameters are interpreted in the context of the research objectives and research questions. Visualizations, such as the behavior of the different parameters of the distribution based on other parameters, are generated using R to visually summarize the characteristics of the Sarmanov distribution.

3. The Sarmanov Distribution

3.1. Fundaments

The Sarmanov distribution is frequently used in actuarial, insurance, and other areas of risk management and quantitative modeling when it is necessary to take into consideration the dependence between two or more numerical variables. It is especially helpful for modeling scenarios where the variables show some degree of dependence or correlation but are not entirely independent. So, this distribution operates with the dependency factor to obtain improved estimates. As seen in 1, it has been stated the existence of a minimal dependence between the discrete variable number of claims and the continuous individual cost. Thus, the use of the Sarmanov distribution for the calculus of the pure premium in the insurance field is justified.

If described in a mathematical way, Sarmanov could be defined as a multivariate probability distribution defined by a joint probability density function (PDF) for two or more numerical variables. As other statistical distributions, it has its own characteristics.

- Marginal functions: The Sarmanov distribution, considering actuarial modeling, is composed of a probability mass function (discrete distributions) for the number of claims or any other discrete random variable and probability density functions (continuous distributions) for the costs of individual claims or continuous random variables.

- Flexibility: One of the greatest advantages it possesses is that can be adapted to different types of data and insurance situations. Different distributions can be used to model the number of claims and the costs of claims, allowing for customization of the model to the specific characteristics of the insurance portfolio under study.

- Parametrization: The Sarmanov distribution can be parameterized using parameters that describe the behavior of the number of claims and the costs of claims. In this manner, there will be a minimum of as many parameters as the total number of parameters of the n included random variables. These parameters can be estimated from observed data, allowing to fit the model with real data.

- Homogeneity Assumption: The assumption of homogeneity, which assumes that claims distributions are consistent across different subpopulations or segments of the portfolio, may not always hold in practice. In reality, different segments of the portfolio may present different claim behaviors, and ignoring such heterogeneity may result in biased modeling results.

- Limited Applicability: Despite presenting many advantageous properties, it may not be suitable for all insurance scenarios. It is primarily used for modeling the joint distribution of claim frequency and severity, and may not be appropriate for situations where other factors, such as policyholder characteristics, external covariates, or time-varying effects, need to be considered. In such cases, alternative modeling approaches, as the generalized linear models (GLM), could be more appropriate.

3.2. Actuarial Collective Risk Model

Also known as the collective risk model or the classical collective risk model, the actuarial collective risk model, is a mathematical model used in actuarial science to study and analyze the collective risk of a portfolio of insurance policies.

In the collective risk model, the basic assumption is that the losses incurred by the policyholders or participants during the term of the policy are random variables that follow a certain probability distribution, typically assumed to be a Poisson or Negative Binomial distribution for the number of losses and a distribution from the family of exponential or a Log-Normal that characterizes the severity or amount of each loss. The collective risk is the sum of all individual losses incurred during a specified time period. The model considers both the frequency and severity of losses to calculate the aggregate risk.

Firstly, we define both random variables number of claims as N and their average claim size with X of a portfolio or of a certain policy. Let us denote S for the aggregate claims, which can be written as follows:

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

To develop future expressions, it is necessary to define the probability functions. Let p denote the probability mass function (PMF) of the discrete random variable N. Regarding to the random variable X associated with severity, its distribution will have a continuous component with probability density function (PDF) denoted by f_X and a probability mass at 0 due to large amounts of 0 when no claims are reported. As S is calculated by the product of the previous random variables, its distribution also has a probability mass at 0 and a probabiliy density function defined by f_S. Additionally, it is necessary to express the accumulative distribution function by F for any random variable. To express it, it is only necessary to add a subscript with the corresponding letter of the random variable such as F_S for the aggregate loss.

3.3. Sarmanov Structure

Following the principles of the collective risk model, the density function can be defined by incorporating the Sarmanov's dependence parameter and kernel functions as

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

where $p(0)$ represents the probability of having zero claims, $p(n)$ and $f(x)$ represent the density functions associated with their respective random variables. Additionally, we define ψ and ϕ as bounded and non-constant kernel functions, and ω as the dependence parameter that we will work with, such that $\omega \in \mathbb{R}$. It is important to note that if there are no claims the cost is zero, so $N = X = 0$ and S, defined as the product of both, equals zero $(S = 0)$. In this manner, by combining two random variables, it results in a mixed density function.

Nevertheless, it is necessary to define two conditions shown in (3) and (4) for the aforementioned probability function to be valid:

$$\begin{aligned} \sum_{n\geq 1} \psi\left(n\right) p\left(n\right) &= \int_{\mathbb{R}} \phi\left(x\right) f\left(x\right) dx = 0 \text{ and} \qquad (3) \\ 1 + \omega\psi\left(n\right)\phi\left(x\right) &\geq 0, \text{ for all } \ n \geq 1, \ x > 0. \qquad (4) \end{aligned}$$

Hereafter are presented the most important propositions associated with the work carried out in the project: the distribution of S and the computation of the correlation between frequency and severity, taking into account the exponential kernel functions and the dependence parameter. These propositions can be found in their entirety in Vernic et al. (2021).

Proposition 1 *Under the Sarmanov dependence condition (2), the expected value and variance of S can be expressed as*

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

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

$$corr\left(X,N\right) = \frac{\omega\mathbb{E}\left[N\psi\left(N\right)\right]\mathbb{E}\left[Y\phi\left(Y\right)\right] + p\left(0\right)\mathbb{E}N\mathbb{E}Y}{\sqrt{\left(1-p\left(0\right)\right)\left(VarY + p\left(0\right)\mathbb{E}^2\left[Y\right]\right)\ VarN}}. \qquad (5)$$

Proposition 3 *The correlation coefficient of the pdf (2) given in (5) when $X > 0$ is the following*

$$corr_{x>0}\left(X,N\right) = \frac{\omega\mathbb{E}\left[N\psi\left(N\right)\right]\mathbb{E}\left[X\phi\left(X\right)\right]}{\sqrt{VarX\ VarN}}. \qquad (6)$$

Further information about the formulas associated with the Sarmanov probability functions is fully described in the articles written by Lee (1996) and Kotz et al. (2000).

The most notable aspect to work on are the limits imposed on the parameter ω representing the Sarmanov dependence. To this end, we define $m_1 = \inf\limits_{n\geq 1} \psi\left(n\right), m_2 = \inf\limits_{x>0} \phi\left(x\right), M_1 =$

$\sup_{n\geq 1} \psi(n)$, $M_2 = \sup_{x>0} \phi(x)$, whose expressions are essential to restrict dependence. The mathematical expressions of m_1, m_2, M_1 and M_2 depend solely, for each one, on one of the distributions associated with the random variables of the number and cost of claims, N and X respectively. Section 4 details this thoroughly assuming multiple statistical distributions. Thus, the interval for ω is

$$\max\left\{-\frac{1}{m_1 m_2}, -\frac{1}{M_1 M_2}\right\} \leq \omega \leq \min\left\{-\frac{1}{m_1 M_2}, -\frac{1}{M_1 m_2}\right\}. \tag{7}$$

Several common types of Sarmanov kernels exist, and among them, we observe in Lee (1996) the following: first, there are kernels founded on cdfs that result in the Farlie-Gumbel-Morgenstern distribution. However, this distribution has a correlation coefficient restricted to $\left|\frac{1}{3}\right|$ which can not be accurate for the dependency between frequency and severity. Second, there are kernels based on moments of the distributions that necessitate truncating the distributions in order to be bounded. This would entail complex adaptation of the distributions, which would result in a loss of efficiency. Finally, there is the exponential kernel, which is naturally bounded and straightforward to manipulate for our specific distributions. Thus, we suggest employing exponential kernels. Concerning Sarmanov's pdf in (2), we concentrate on exponential kernels that fulfill the condition (3). To simplify the notation, let us denote Y as a random variable characterized as the positive severity consisting of all costs greater than 0: $x > 0$. Then, the exponential kernels for Y can be represented as

$$\phi(y, \gamma) = e^{-\gamma y} - \mathcal{L}Y(\gamma), \tag{8}$$

where $\mathcal{L}Y$ denotes the Laplace transform of the random variable Y, and $\phi(y)$ includes the kernel parameter γ. Furthermore, we define $\psi(n, \delta) = e^{-\delta n} - k$ as the kernel function represented by the parameter δ for the discrete random variable. The parameter k can be written as $k = \frac{\mathcal{L}N(\delta) - p(0)}{1 - p(0)}$. The exact expression for k can be found in Vernic et al. (2021). Therefore,

$$\psi(n, \delta) = e^{-\delta n} - \frac{\mathcal{L}N(\delta) - p(0)}{1 - p(0)}. \tag{9}$$

The Laplace operators have two parameters, δ and γ, that directly impact the interval defined in (7). Larger values of these parameters result in a wider interval and a greater scale effect on the dependence parameter. However, if the values are too large, the estimates of the dependency parameter may be inefficient. On the other hand, if the values are too small, the dependency parameter may be biased downwards. In Section 4, the interrelationships between them and the formulas to be used to limit the values of δ and γ according to the distributions, which can be estimated from claims data, are detailed. This will allow acceptable values of dependence to be obtained. In Section 5, these formulas will be applied using a real database with information on frequency and severity, and the results of the estimates will be compared, considering fixed parameters or those determined by the formulas of the previous section.

4. Bivariate Sarmanov distribution

From an actuarial standpoint, the Sarmanov distribution can be used to model aggregate costs while taking into account the dependence between the number of claims and their respective costs. Thus, it is necessary to define the statistical distributions that could be applied to it. Initially, discrete distributions should be used for frequency and continuous distributions for severity, but not all distributions would be useful. For the behavior of frequency, the following distributions could be used:

1. Poisson Distribution: Given the random variable N, we say that $N \sim Poisson(\lambda)$. It is characterized, especially for having a variance equal to the expected value of that distribution. That is, $E[N] = \lambda$, $V[N] = \lambda$, and therefore, $E[N] = V[N]$. This characteristic limits the use of the Poisson distribution in the insurance sector, as it does not adequately represent the frequency of claims since it does not capture its asymmetry. For example, if a λ close to 0 is adopted, the vast majority of expected values will be 0 or 1. Its Laplace transform can be written as follows:

$$\mathcal{L}_N(\delta) = e^{\lambda\left(e^{-\delta}-1\right)}. \tag{10}$$

   So, considering (10), the kernel funcion associated to a Poisson distribution is

$$\psi(n,\delta) = e^{-\delta n} - \frac{e^{\lambda\left(e^{-\delta}-1\right)} - e^{-\lambda}}{1-e^{-\lambda}}. \tag{11}$$

2. Zero-Inflated Poisson Distribution: We say that $N \sim ZIP(\lambda, \pi)$. It is a modification of the Poisson distribution when there is a high number of 0s in the data. In the insurance sector, the vast majority of insureds in the Non-Motor field do not have claims, making the ZIP a distribution that better fits reality. However, it has a higher degree of complexity as it presents a parameter π that represents the probability of excess zero claims. The Laplace function according to a ZIP distribution and the resulting kernel function are

$$\mathcal{L}_{\tilde{N}}(\delta) = \pi + (1-\pi)\, e^{\lambda\left(e^{-\delta}-1\right)}, \tag{12}$$

$$\psi(n,\delta) = e^{-\delta n} - \frac{(\pi + (1-\pi)e^{\lambda(e^{-\delta}-1)}) - (\pi + (1-\pi)e^{-\lambda})}{1-(\pi + (1-\pi)e^{-\lambda})}. \tag{13}$$

3. Negative Binomial Distribution: It can be written as $N \sim NB(p, r)$. Unlike the Poisson distribution, the Negative Binomial has a more extensive application due to its good properties. We define its expected value as $E[N] = \frac{r(1-p)}{p}$ and variance as $V[N] = \frac{r(1-p)}{p^2}$,

and therefore, $E[N] < V[N]$. This distribution allows for heavier tails and does not limit the number of claims as much. The $\mathcal{L}_{\tilde{N}}$ for the Negative Binomial can be written as

$$\mathcal{L}_N(\delta) = \left(\frac{p}{1 - qe^{-\delta}}\right)^r. \tag{14}$$

By substituting the previous expression into the kernel function, we would obtain

$$\psi(n, \delta) = e^{-\delta n} - \frac{(\frac{p}{1-(1-p)e^{-\delta}})^r - p^r}{1 - p^r}. \tag{15}$$

4. Zero-Inflated Negative Binomial Distribution: Defined as $N \sim ZINB(p, r, \pi)$. Like the ZIP, the ZINB adapts better than the NB when the percentage of zeros is excessively high.

$$\mathcal{L}_N(\delta) = \pi + (1 - \pi)\left(\frac{p}{1 - qe^{-\delta}}\right)^r. \tag{16}$$

In terms of severity modeling, the following continuous distributions could be used:

1. Gamma Distribution: Denoted as $Y \sim Gamma(\alpha, \beta)$, it is one of the most commonly used distributions due to its good fit to claim costs. Its main properties include $E[Y] = \frac{\alpha}{\beta}$ and $V[Y] = \frac{\alpha}{\beta^2}$ and the Laplace transform is

$$\mathcal{L}_Y(\gamma) = \left(\frac{\beta}{\beta + \gamma}\right)^\alpha. \tag{17}$$

   Regarding the kernel function $\phi(y, \gamma)$, its expression would be defined as follows:

$$\phi(y, \gamma) = e^{-\gamma y} - \left(\frac{\beta}{\beta + \gamma}\right)^\alpha. \tag{18}$$

2. Log-Normal Distribution: An alternative to the Gamma distribution is the Log-Normal distribution. If the random variable Y follows a Log-Normal distribution, $Y \sim Log-Normal(\mu, \sigma)$, its logarithm would be distributed as a Normal random variable. With respect to Y, we would obtain $E[Y] = e^{\mu + \frac{\sigma^2}{2}}$ and $V[Y] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$.
   Regarding to its Laplace transform, as the domain of the normal distribution includes the entire set of real numbers and the kernel function f must be bounded, it is necessary to use a left truncated normal distribution that is truncated at the left point a. There is no exact value for a, but a valid option would be to consider $a = -3\sigma + \mu$. The detailed

explanation of the procedure of definition of $\mathcal{L}_Y(\gamma)$ and the formula given to a can be found in Bolancé and Vernic (2020). The definitive expression of the Laplace transform that will be used when considering a Log-Normal distribution for severity is given by

$$\mathcal{L}_Y(\gamma) = e^{-\gamma\mu + \frac{\gamma^2\sigma^2}{2}} \frac{1 - \Phi\left(\frac{a-\mu+\gamma\sigma^2}{\sigma}\right)}{K}. \tag{19}$$

By substituting the expression of $\mathcal{L}_Y(\gamma)$ with $a = -3\sigma + \mu$ and $K = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$ in the kernel function, the corresponding formula would be obtained for the Log-Normal distribution.

$$\phi(y,\gamma) = e^{-\gamma y} - e^{-\gamma\mu + \frac{\gamma^2\sigma^2}{2}} \frac{1 - \Phi(-3 + \gamma\sigma^2)}{1 - \Phi(-3)}. \tag{20}$$

3. Other distributions: Among others, the Exponential, the Inverse Gaussian or the Pareto distributions can also be used to model the cost of claims.

Due to the complexity of some of the distributions, this document will focus specifically on the Poisson, Zero Inflated Poisson and Negative Binomial distributions regarding frequency and Gamma and Log-Normal for the severity.

These distributions will directly affect the estimation and behavior of the parameters δ, γ, and ω so it is important to choose wisely which distribution should be used. In the following subsections, some constructions of bivariate Sarmanov applicable to the insurance sector are presented. The main objective of the research is to provide intervals for δ and γ that directly influence ω, which will vary depending on the starting distributions. Additionally, the relationship between the parameters of the kernels for Poisson-Gamma and Negative Binomial-Gamma cases will be detailed.

4.1. The bivariate Sarmanov Poisson - Gamma distribution

Let $N \sim Poisson(\lambda)$ and $\tilde{X} \sim Gamma(\alpha,\beta)$ be the marginal distributions in Sarmanov model. In this instance, the distribution that we have analyzed is denoted as $(N,X) \sim Sarmanov - Poisson - Gamma(\lambda,\alpha,\beta,\omega,\delta,\gamma)$ which includes six parameters. We must define first m_1 and M_1 associated with the discrete random variable which is Poisson distributed and m_2 and M_2 with the continuous Gamma distribution. For the discrete distributions, note that we define m_1 and M_1 for those cases that $N > 0$ meaning that $P[N = 0] = 0$ and the expressions are then simplified. In this way, the following expressions are obtained and can also be used if that particular individual distribution is considered:

$$\begin{aligned} m_1 &= \lim_{n\to\infty} \psi(n,\delta) = -e^{\lambda\left(e^{-\delta}-1\right)}, \\ M_1 &= \lim_{n\to 1} \psi(n,\delta) = e^{-\delta} - e^{\lambda\left(e^{-\delta}-1\right)}, \\ m_2 &= \lim_{x\to\infty} \phi(x,\delta) = 1 - \left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}, \\ M_2 &= \lim_{x\to 0} \phi(x,\delta) = -\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}. \end{aligned} \tag{21}$$

The strategy to improve the estimates of ω and therefore the results of the Sarmanov is to limit the values of the kernel functions δ and γ, which have a direct impact on the dependence as seen in (7), to a certain interval. To do so, it is proposed to calculate the first partial derivatives with respect to δ and γ of the expressions that are found in (7) for each expression of the boundaries of ω. The idea is to find those parameters of the exponential kernels in order to provide more flexibility to the Sarmanov model without having to set arbitrary maximum values to δ and γ and, in this way, we should get a more accurate estimation of the dependency. Thus, for each of the four expressions defined in the range of ω, the roots of the two partial derivatives will need to be found. These roots will be the values of δ and γ which will be used as boundaries for the parameter ranges of δ and γ. However, before that, these derivatives must be defined.

Then, the structure of the partial derivatives will always be the same. Only δ appears as the parameter of the discrete distribution kernel, and γ as the parameter of the continuous distribution, which allows for a simpler calculation of the first partial derivative. If it is done in terms of m_1 represented by δ, and m_2 represented by γ, we would obtain:

$$\begin{aligned} \frac{\partial \frac{-1}{m_1 m_2}}{\partial \delta} &= \frac{\frac{\partial m_1}{\partial \delta}}{m_1^2 m_2}, \\ \frac{\partial \frac{-1}{m_1 m_2}}{\partial \gamma} &= \frac{\frac{\partial m_2}{\partial \gamma}}{m_1 m_2^2}. \end{aligned} \tag{22}$$

One can define the partial derivatives in this way because the parameter δ appears only in those expressions with subscript $_{(1)}$, and γ in the subscripts $_{(2)}$. This fact simplifies the partial derivatives which only one of them needs to be differentiated. In this manner, it is necessary to calculate the partial derivatives with respect to δ in the expressions m_1 and M_1 and with respect to γ for m_2 and M_2.

Thus, for the case in which Poisson and Gamma distributions are considered, the final expressions of the partial derivatives are

$$
\begin{aligned}
\frac{\partial}{\partial\delta}\left(-\frac{1}{m_1m_2}\right) &= \frac{-\lambda e^{-\lambda(e^{-\delta}-1)-\delta}(\beta+\gamma)^{\alpha}}{\beta^{\alpha}},\\
\frac{\partial}{\partial\delta}\left(-\frac{1}{M_1M_2}\right) &= \frac{-e^{-\delta}+\lambda e^{\lambda(e^{-\delta}-1)-\delta}}{(1-(\frac{\beta}{\beta+\gamma})^{\alpha})(e^{-\delta}-e^{\lambda(e^{-\delta}-1)})^2},\\
\frac{\partial}{\partial\delta}\left(-\frac{1}{m_1M_2}\right) &= \frac{\lambda e^{-\delta(e^{-\delta}-1)-\delta}}{\left(1-\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}\right)},\\
\frac{\partial}{\partial\delta}\left(-\frac{1}{M_1m_2}\right) &= -\frac{-e^{-\delta}+\lambda e^{\lambda\left(e^{-\delta}-1\right)-\delta}\left(\beta+\gamma\right)^{\alpha}}{\beta^{\alpha}\left(e^{-\delta}-e^{\lambda\left(e^{-\delta}-1\right)}\right)^2},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{m_1m_2}\right) &= -\frac{\alpha(\beta+\gamma)^{(\alpha-1)}}{e^{\lambda\left(e^{-\delta}-1\right)}\beta^{\alpha}},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{M_1M_2}\right) &= \frac{\alpha\beta^{\alpha}(\beta+\gamma)^{(\alpha-1)}}{e^{-\delta}-e^{\lambda\left(e^{-\delta}-1\right)}\left((\beta+\gamma)^{\alpha}-\beta^{\alpha}\right)^2},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{m_1M_2}\right) &= \frac{\alpha\beta^{\alpha}(\beta+\gamma)^{(\alpha-1)}}{e^{\lambda\left(e^{-\delta}-1\right)}\left((\beta+\gamma)^{\alpha}-\beta^{\alpha}\right)^2},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{M_1m_2}\right) &= -\frac{\alpha(\beta+\gamma)^{(\alpha-1)}}{\left(e^{-\delta}-e^{\lambda\left(e^{-\delta}-1\right)}\right)\beta^{\alpha}}.
\end{aligned}
\tag{23}
$$

To exemplify, and considering one of the components of the lower limit of the parameter ω, $-\frac{1}{m_1m_2}$, the expressions for the partial derivatives with respect to δ and γ are calculated. Subsequently, using the estimated parameters from the distributions associated with frequency and severity, the following system of equations needs to be solved:

$$
\begin{aligned}
\frac{\partial}{\partial\delta}\left(-\frac{1}{m_1m_2}\right) &= 0,\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{m_1m_2}\right) &= 0
\end{aligned}
$$

where only δ and γ are unknown. This system is solved for each of the components that limit ω, resulting in different estimations for the parameters of the exponential kernels. In this way, and considering the same approach in (7), the maximum values are taken for the left interval of the dependency (negative dependency), while the minimum values are taken for the upper limit (positive dependency). To carry out this procedure, the R package **rootSolve** and the function **multiroot** are employed to find the roots of the system.

For the other bivariate Sarmanov distribution cases, only the partial derivatives are shown as the procedure to estimate the values of δ and γ is the same although the expressions vary from each other.

4.1.1. Delta and Gamma Laplace transform parameters relationship

Studying the behavior of the parameters of the kernels δ and γ can assist in establishing new dependency estimation strategies and may also be helpful to determine fixed values for these components. Therefore, an analysis is performed on the impact that each parameter has on the other, taking into account the effects they produce on the ω interval. In other words, it is analyzed how each parameter affects the amplitude of the dependency interval.

To do so, we will fix one of the two parameters to 1 while the other will take multiple values up to a maximum limit of 20. We will define that, if $\delta = 1$, then $\gamma \in (0, 20)$ or $\gamma = 1$ with $\delta \in (0, 10)$. It is also essential to define the parameters of the initial distributions, λ by the Poisson distribution, and α and β by the Gamma distribution. Multiple realistic scenarios have been collected to detail this behavior. On page 18, twelve combinations for these parameters are shown in a table with their representative number that appears in the results graph on the same page.

In the two upper graphs, we observe how the parameter γ affects the dependency interval while holding δ fixed. There are no relatively high differences in the effects that it has on the lower and upper bounds of the interval of ω, from the first observed γ value to the last. For the lower bound, there are two scenarios in which it behaves differently from the rest, where γ becomes more relevant. If we look at the upper limit, small values of γ contrast with large ones, where the former enlarge the interval and the latter make it smaller. However, in general, this parameter has a significant impact on defining the interval.

The differences arise in the values of the parameter δ. In this case, higher values of δ for a fixed γ result in a wider interval in comparison with lower values of the same parameter. Therefore, the δ component, assuming that frequency follows a Poisson distribution and severity follows a Gamma distribution, takes on greater importance than γ in defining the interval associated with dependence.

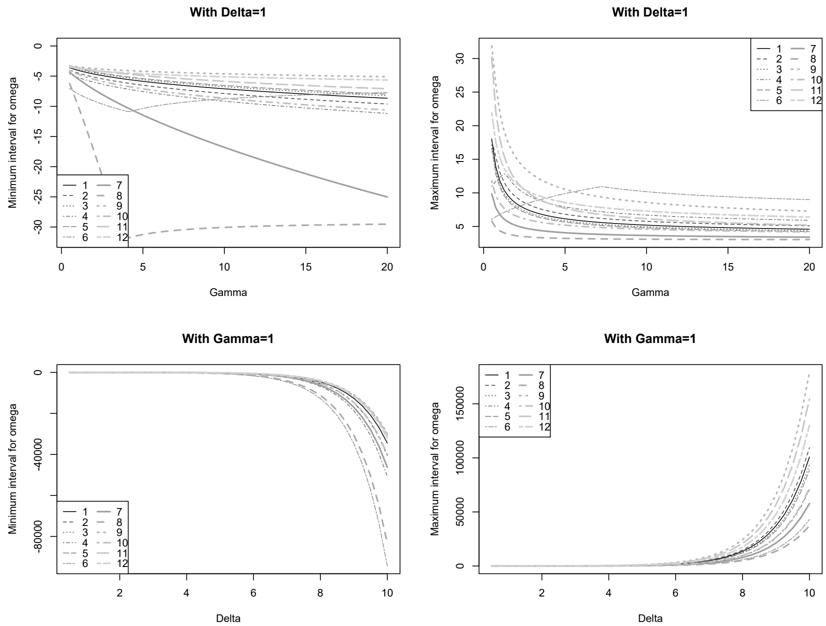

Figure 1: Relationship between δ and γ to the interval values of ω.

Number	λ	α	β
1	0.30	0.30	0.6
2	0.60	0.30	0.6
3	0.15	0.30	0.6
4	1.00	0.30	0.6
5	0.05	0.30	0.6
6	2.00	0.30	0.6
7	0.30	0.60	0.6
8	0.30	1.20	0.6
9	0.30	0.15	0.6
10	0.30	0.30	0.3
11	0.30	0.30	1.2
12	0.30	0.15	0.3

Table 1: Twelve scenarios with the parameters of the Poisson-Gamma Sarmanov Distribution considered for the analysis.

4.2. The bivariate Sarmanov Zero Inflated Poisson - Gamma distribution

Consider the Sarmanov model with marginal distributions $N \sim ZIP(\lambda,\pi)$ and $\tilde{X} \sim Gamma(\alpha,\beta)$. Let (N,X) denote the joint distribution under analysis, and we denote it as $(N,X) \sim Sarmanov - ZIP - Gamma(\lambda,\pi,\alpha,\beta,\omega,\delta,\gamma)$. The associated exponential kernels m_1 and M_1 for the Zero Inflated Poisson are as follows:

$$\begin{aligned} m_1 &= \lim_{n\to\infty} \psi(n,\delta) = -(\pi + (1-\pi)e^{\lambda(e^{-\delta}-1)}), \\ M_1 &= \lim_{n\to 1} \psi(n,\delta) = e^{-\delta} - (\pi + (1-\pi)e^{\lambda(e^{-\delta}-1)}) \end{aligned} \tag{24}$$

The expressions of the exponential kernels for the Gamma distribution are in the previous point.

This document presents the partial derivatives for this set of distributions for the Sarmanov following the principles in (22). However, the use of the Zero-Inflated Poisson is not as common as the Poisson or the Negative Binomial, and therefore, this distribution combination will not be detailed extensively. For this reason, a future study could be to analyze the Zero-Inflated Poisson with the Gamma or any continuous distribution that fits the costs of claims properly.

Then, their first partial derivatives can be defined as

$$\frac{\partial}{\partial\delta}\left(-\frac{1}{m_1m_2}\right) = \frac{(\pi-1)\,\lambda e^{\lambda\left(e^{-\delta}-1\right)-\delta}}{\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}\left((\pi-1)e^{\lambda\left(e^{-\delta}-1\right)}-\pi\right)^2},$$

$$\frac{\partial}{\partial\delta}\left(-\frac{1}{M_1M_2}\right) = -\frac{(1-\pi)\,\lambda e^{\lambda\left(e^{-\delta}-1\right)-\delta}+e^{\delta}}{\left(1-\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}\right)\left(-(1-\pi)\,e^{\lambda\left(e^{-\delta}-1\right)}+e^{\delta}-\pi\right)^2},$$

$$\frac{\partial}{\partial\delta}\left(-\frac{1}{m_1M_2}\right) = -\frac{(\pi-1)\,\lambda e^{\lambda\left(e^{-\delta}-1\right)-\delta}}{\left(\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}-1\right)\left((\pi-1)e^{\lambda\left(e^{-\delta}-1\right)}-\pi\right)^2}$$

$$\frac{\partial}{\partial\delta}\left(-\frac{1}{M_1m_2}\right) = -\frac{(1-\pi)\,\lambda e^{\lambda\left(e^{-\delta}-1\right)-\delta}+e^{\delta}}{\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}\left((\pi-1)e^{\lambda\left(e^{-\delta}-1\right)}+e^{\delta}-\pi\right)^2},$$

$$\frac{\partial}{\partial\gamma}\left(-\frac{1}{m_1m_2}\right) = \frac{\alpha}{\left((\pi-1)e^{\lambda\left(e^{-\delta}-1\right)}-\pi\right)(\beta+\gamma)\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}},$$

$$\frac{\partial}{\partial\gamma}\left(-\frac{1}{M_1M_2}\right)=-\frac{\alpha\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}}{\left((\pi-1)e^{\lambda\left(e^{-\delta}-1\right)}+e^{\delta}-\pi\right)(\beta+\gamma)\left(\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}-1\right)^{2}},$$

$$\frac{\partial}{\partial\gamma}\left(-\frac{1}{m_1M_2}\right)=\frac{\alpha\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}}{\left((1-\pi)e^{\lambda\left(e^{-\delta}-1\right)}+\pi\right)(\beta+\gamma)\left(1-\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}\right)^{2}},$$

$$\frac{\partial}{\partial\gamma}\left(-\frac{1}{M_1m_2}\right)=\frac{\alpha}{\left((\pi-1)e^{\lambda\left(e^{-\delta}-1\right)}+e^{\delta}-\pi\right)(\beta+\gamma)\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}}. \tag{25}$$

4.3. The bivariate Sarmanov Negative Binomial - Gamma distribution

Let $N\sim NB(r,p)$ and $\tilde{X}\sim Gamma(\alpha,\beta)$ be the marginal distributions in Sarmanov model. In this case we denote our analyzed distribution as $(N,X)\sim Sarmanov-NegativeBinomial-Gamma(r,p,\alpha,\beta,\omega,\delta,\gamma)$. The expressions of the exponential kernels m_1 and M_1 for the discrete distribution Negative Binomial associated with the number of claim are:

$$\begin{aligned}
m_1 &= \lim_{n\to\infty}\psi(n,\delta)=-\left(\frac{p}{1-qe^{-\delta}}\right)^{r},\\
M_1 &= \lim_{n\to 1}\psi(n,\delta)=e^{-\delta}-\left(\frac{p}{1-qe^{-\delta}}\right)^{r}.
\end{aligned} \tag{26}$$

The partial derivatives to use when considering the Negative Binomial and Gamma Sarmanov distribution are the following:

$$\frac{\partial}{\partial\delta}\left(-\frac{1}{m_1m_2}\right)=\frac{(p-1)\,r}{\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}\left(\frac{p}{1-(1-p)e^{-\delta}}\right)^{r}(e^{\delta}+p-1)},$$

$$\frac{\partial}{\partial\delta}\left(-\frac{1}{M_1M_2}\right)=\frac{e^{\delta}\left(e^{\delta}\left((p-1)r\left(\frac{p}{e^{\delta}+p-1}\right)^{r}e^{r\delta}+1\right)+p-1\right)}{\left(\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}-1\right)(e^{\delta}+p-1)\left(\left(\frac{p}{e^{\delta}+p-1}\right)^{r}e^{r\delta+\delta}-1\right)^{2}},$$

$$
\begin{aligned}
\frac{\partial}{\partial\delta}\left(-\frac{1}{m_1 M_2}\right) &= -\frac{(p-1)\,r}{\left(\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}-1\right)\left(\frac{p}{1-(1-p)e^{-\delta}}\right)^{r}(e^{\delta}+p-1)},\\
\frac{\partial}{\partial\delta}\left(-\frac{1}{M_1 m_2}\right) &= \frac{e^{\delta}\left(e^{\delta}\left((p-1)r\left(\frac{p}{e^{\delta}+p-1}\right)^{r}e^{r\delta}+1\right)+p-1\right)}{\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}(e^{\delta}+p-1)\left(\left(\frac{p}{e^{\delta}+p-1}\right)^{r}e^{r\delta+\delta}-1\right)^{2}},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{m_1 m_2}\right) &= -\frac{\alpha}{\left(\frac{p}{1-e^{-\delta}(1-p)}\right)^{r}(\beta+\gamma)\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{M_1 M_2}\right) &= \frac{\alpha\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}}{(e^{-\delta}-\left(\frac{p}{1-e^{-\delta}(1-p)}\right)^{r})(\beta+\gamma)(1-\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha})^{2}},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{m_1 M_2}\right) &= \frac{\alpha\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}}{\left(\frac{p}{1-e^{-\delta}(1-p)}\right)^{r}(\beta+\gamma)\left(\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}-1\right)^{2}},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{M_1 m_2}\right) &= \frac{\alpha}{(e^{-\delta}-\left(\frac{p}{1-e^{-\delta}(1-p)}\right)^{r})(\beta+\gamma)\left(\frac{\beta}{\beta+\gamma}\right)^{\alpha}}.
\end{aligned}
\tag{27}
$$

In terms of frequency and severity of the claims, this pair of distributions closely aligns with reality. Therefore, considering these two distributions together is a good option, and it is the one that will be further explored in subsequent sections with greater depth in a practical application with real data.

4.3.1. Delta and Gamma Laplace transform parameters relationship

The same analysis carried out for the case of the bivariate Poisson-Gamma Sarmanov on page 17 can be done for the Negative Binomial-Gamma. As we are working with a different discrete distribution, the parameters associated with this distribution are different and its parameter values will be different according to the scenario. The number of scenarios is not changed as twelve of them are considered.

Despite changing the discrete distribution, the results obtained are very similar; the parameter γ does not have such a remarkable effect as δ on the ω intervals. The larger the value of γ, the smaller the lower value of the interval, while the upper value is reduced. Regarding δ, it is observed that the greater it grows, the greater the amplitude of the interval. The graphs and parameters for each scenario can be viewed on page 22.

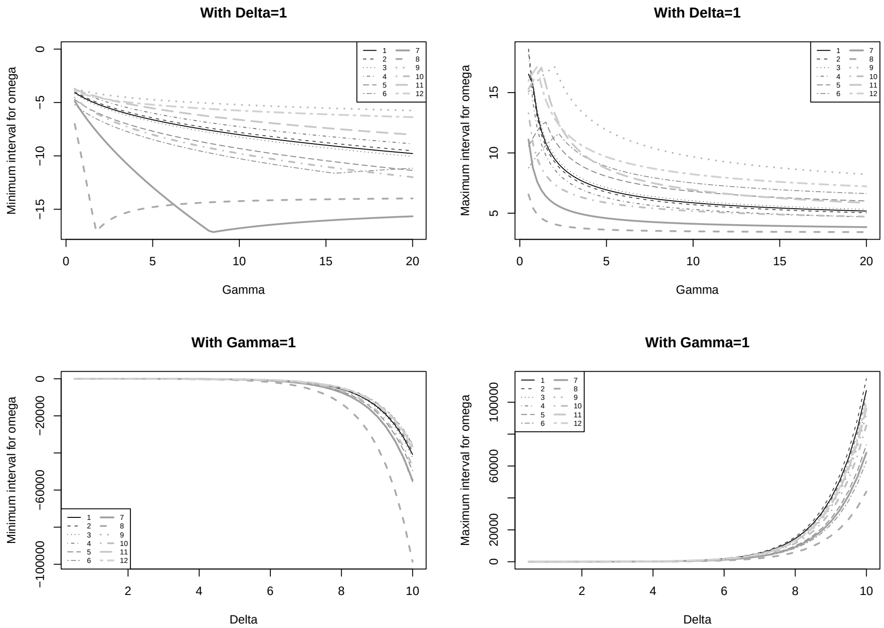

Figure 2: Relationship between δ and γ to the interval values of ω.

Number	r	p	α	β
1	0.30	0.60	0.30	0.6
2	0.15	0.60	0.30	0.6
3	0.45	0.60	0.30	0.6
4	0.30	0.75	0.30	0.6
5	0.20	0.40	0.30	0.6
6	0.15	0.30	0.30	0.6
7	0.30	0.60	0.60	0.6
8	0.30	0.60	1.20	0.6
9	0.30	0.60	0.15	0.6
10	0.30	0.60	0.30	0.3
11	0.30	0.60	0.30	1.2
12	0.30	0.60	0.15	0.3

Table 2: Twelve scenarios with the parameters of the Negative Binomial-Gamma Sarmanov Distribution considered for the analysis.

4.4. The bivariate Sarmanov Negative Binomial - Log-Normal distribution

Let $N \sim NB(r,p)$ and $\tilde{X} \sim Log-Normal(\alpha,\beta)$ be the marginal distributions in Sarmanov model. In this case we denote our analysed distribution as $(N,X) \sim Sarmanov-NegativeBinomial-Log-Normal(r,p,\mu,\sigma,\omega,\delta,\gamma)$. The exponential kernels m_2 and M_2 for the continuous distribution to use in this situation for the Log-Normal distribution, are:

$$\begin{aligned} m_2 &= \lim_{x\to\infty} \phi(x,\delta) = -e^{-\gamma\mu+\frac{\gamma^2\sigma^2}{2}} \frac{1-\Phi(-3+\gamma\sigma^2)}{1-\Phi(-3)}, \\ M_2 &= \lim_{x\to 0} \phi(x,\delta) = 1-e^{-\gamma\mu+\frac{\gamma^2\sigma^2}{2}} \frac{1-\Phi(-3+\gamma\sigma^2)}{1-\Phi(-3)}. \end{aligned} \tag{28}$$

The expressions for m_2 and M_2 are more complex than those used for the Gamma distribution as they include integrals, making it difficult to compute the expressions for the partial derivatives. Nevertheless, the same methodology used for the other distribution combinations has been followed. Therefore, the partial derivatives to be used are:

$$\frac{\partial}{\partial\delta}\left(-\frac{1}{m_1 m_2}\right) = \frac{(p-1)\,r}{\left(\frac{p}{1-(1-p)e^{-\delta}}\right)^r (e^\delta+p-1)\left(-e^{\gamma\mu+\frac{\gamma^2\sigma^2}{2}}\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}} e^{-t^2}dt}{0.99865\sqrt{\pi}}\right]\right)},$$

$$\frac{\partial}{\partial\delta}\left(-\frac{1}{M_1 M_2}\right) = \frac{\left(e^\delta\left(\left(e^\delta\left((p-1)r\left(\frac{p}{e^\delta+p-1}\right)^r e^{r\delta}+1\right)\right)+p-1\right)\right)}{\left(\left(\frac{p}{e^\delta+p-1}\right)^r e^{r\delta+\delta}-1\right)^2\left(1-e^{\gamma\mu+\frac{\gamma^2\sigma^2}{2}}\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}} e^{-t^2}dt}{0.99865\sqrt{\pi}}\right]\right)},$$

$$\frac{\partial}{\partial\delta}\left(-\frac{1}{m_1 M_2}\right) = \frac{(p-1)\,r}{\left(\frac{p}{1-(1-p)e^{-\delta}}\right)^r (e^\delta+p-1)\left(1-e^{\gamma\mu+\frac{\gamma^2\sigma^2}{2}}\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}} e^{-t^2}dt}{0.99865\sqrt{\pi}}\right]\right)}$$

$$\frac{\partial}{\partial\delta}\left(-\frac{1}{M_1 m_2}\right) = \frac{\left(e^\delta\left(e^\delta\left((p-1)r\left(\frac{p}{e^\delta+p-1}\right)^r e^{r\delta}+1\right)+p-1\right)\right)}{\left(\left(\frac{p}{e^\delta+p-1}\right)^r e^{r\delta+\delta}-1\right)^2\left(e^{\gamma\mu+\frac{\gamma^2\sigma^2}{2}}\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}} e^{-t^2}dt}{0.99865\sqrt{\pi}}\right]\right)},$$

$$
\begin{aligned}
\frac{\partial}{\partial\gamma}\left(-\frac{1}{m_1 m_2}\right) &= \frac{\frac{e^{-\gamma\mu+\frac{\gamma^2\sigma^2}{2}}}{0.99865}\left(\frac{\sigma e^{-\frac{1}{2}(\gamma\sigma-3)}}{\sqrt{2\pi}}-(\sigma^2\gamma-\mu)\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}}e^{-t^2}dt}{\sqrt{\pi}}\right]\right)}{\left(\frac{p}{1-e^{-\delta}(1-p)}\right)^r\left(-e^{\gamma\mu+\frac{\gamma^2\sigma^2}{2}}\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}}e^{-t^2}dt}{0.99865\sqrt{\pi}}\right]\right)^2},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{M_1 M_2}\right) &= \frac{\frac{e^{-\gamma\mu+\frac{\gamma^2\sigma^2}{2}}}{0.99865}\left(\frac{\sigma e^{-\frac{1}{2}(\gamma\sigma-3)}}{\sqrt{2\pi}}-(\sigma^2\gamma-\mu)\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}}e^{-t^2}dt}{\sqrt{\pi}}\right]\right)}{(e^{-\delta}-\left(\frac{p}{1-e^{-\delta}(1-p)}\right)^r)\left(1-e^{\gamma\mu+\frac{\gamma^2\sigma^2}{2}}\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}}e^{-t^2}dt}{0.99865\sqrt{\pi}}\right]\right)^2},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{m_1 M_2}\right) &= \frac{\frac{e^{-\gamma\mu+\frac{\gamma^2\sigma^2}{2}}}{0.99865}\left(\frac{\sigma e^{-\frac{1}{2}(\gamma\sigma-3)}}{\sqrt{2\pi}}-(\sigma^2\gamma-\mu)\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}}e^{-t^2}dt}{\sqrt{\pi}}\right]\right)}{\left(\frac{p}{1-e^{-\delta}(1-p)}\right)^r\left(1-e^{\gamma\mu+\frac{\gamma^2\sigma^2}{2}}\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}}e^{-t^2}dt}{0.99865\sqrt{\pi}}\right]\right)^2},\\
\frac{\partial}{\partial\gamma}\left(-\frac{1}{M_1 m_2}\right) &= \frac{\frac{e^{-\gamma\mu+\frac{\gamma^2\sigma^2}{2}}}{0.99865}\left(\frac{\sigma e^{-\frac{1}{2}(\gamma\sigma-3)}}{\sqrt{2\pi}}-(\sigma^2\gamma-\mu)\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}}e^{-t^2}dt}{\sqrt{\pi}}\right]\right)}{(e^{-\delta}-\left(\frac{p}{1-e^{-\delta}(1-p)}\right)^r)\left(-e^{\gamma\mu+\frac{\gamma^2\sigma^2}{2}}\left[\frac{\frac{1}{2}-\int_0^{\frac{\gamma\sigma-3}{\sqrt{2}}}e^{-t^2}dt}{0.99865\sqrt{\pi}}\right]\right)^2}.
\end{aligned}
\tag{29}
$$

Notice that instead of using the standard Normal distribution function, the Gauss error function (erf) has been considered, extracted from the integral of the distribution function. In this way, a slightly more reduced expression is obtained. Additionally, in (29), the value 0.99865 appears, which is the value obtained in $1-\Phi(3)$.

The Log-Normal distribution is frequently used to model the cost of claims. Additionally, the Gamma distribution is widely acknowledged for its good fit to severity insurance data. Therefore, this study focuses more on the latter distribution, although the procedure is also replicable for the Log-Normal distribution.

5. Application

In this section, the theoretical framework proposed to work on the dependency of the Sarmanov distribution is put into practice. As this is a complex method, it is important to find tools that facilitate its use in order to increase its applicability. However, to date, there is no software that can directly model the frequency and severity of an insurance portfolio considering this approach, and therefore, the technical structure of the modeling process has been developed by ourselves.

5.1. Optimization strategy

To model the frequency and severity of claims using the Sarmanov distribution, it is necessary to establish a parameter estimation strategy based on the initial distributions of the random variables. Such strategy follows the steps defined in Vernic et al. (2021), which involve obtaining the parameters by maximizing the log-likelihood, which could be defined as

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

where θ and ν are the parameters of the two random variables, frequency and severity, respectively. This strategy assumes that most of the parameters are known. Given this, an iterative strategy can be made following these steps:

- Parameter estimation from data. It is necessary to input the values of the parameters of the discrete and continuous distribution. To do so, data is needed to estimate the parameters for each distribution. The **fitdistrplus** package can be used, which allows the user to estimate them using the maximum likelihood approach.

- It is also necessary to give values to δ, γ, and ω. Initially, ω is set to 0, and there will not be a fixed value for the kernel parameters, but rather an iterative process will be followed such that, for a given value of one parameter, the other will take any value. In fact, the maximum values that δ and γ could take would be the values found at the roots of the partial derivatives. The importance in selecting the values for those parameters is due to the different results obtained for each possible combination. In our case, a grid will be defined such that $\delta \in [0.1, 21]$ and $\gamma \in [0.1, 21]$.

- So, for each combination on δ and γ an iterative process is defined that must be continued until the parameters converge such that:

- With all the parameters defined in the previous sections, we proceed to estimate the dependence $\hat{\omega}$, which is obtained by maximizing the log-likelihood given all the parameters.

- Based on this estimation of the dependence, the rest of the parameters, including those of the initial distributions and the kernel parameters, are jointly re-estimated. Once re-estimated, the process is repeated.

However, the parameters from the last point could also be estimated separately, that is, estimating on one hand the parameters related to the number and cost of claims, and on the other hand, $\hat{\delta}$ and $\hat{\gamma}$.

5.2. Application with the bivariate Sarmanov Negative Binomial - Gamma distribution

To put into practice the developed strategy and apply it in a real situation, a database provided by an insurance entity containing 65533 policies with respective information on the number of claims and average cost has been used. However, the values have been slightly altered to ensure the anonymity of the insureds. The data distribution is described as follows:

Number of claims	Number of observations	Mean cost per claim
0	57935	0
1	5838	932.75
2	1535	1654.78
3	189	1048.53
4	29	862.08
5	7	319.92

Table 3: Data table with 3 columns

An interesting aspect to quantify is the correlation between the number of accidents and their respective severities in those cases where accidents have occurred. When working with a bivariate distribution that assumes dependence between variables, it is important to detect whether there is any type of relationship between variables, even if they are not linearly connected. However, a small positive correlation is observed, implying that the higher the number of accidents, the higher the costs which can be resumed as a positive correlation. Overall, with the Sarmanov, this relationship will be captured through the parameter associated with the dependence. Note that, given the positive correlations, the parameter associated with the dependence ω will also take positive values.

Linear	Kendall	Spearman
0.12867	0.14578	0.18129

Table 4: Linear correlation between number and cost of claims for those policies that $X > 0$.

It is also important to determine which individual distributions are most appropriate for the data being used. In section 4, multiple distributions were defined that could potentially fit the insureds data adequately, but the best-fitting distribution can usually be identified. This can be done by comparing the different Akaike Information Criteria (AIC) obtained when fitting distributions to the data. Through such comparison, which we should look for the minimum value, it is observed that the Negative Binomial distribution is the best fit for the number of accidents, while the Gamma distribution is a better fit for the costs.

	Poisson	Negative Binomial	Gamma	Log-Normal
AIC	59237.591	56999.476	15537.317	16903.445

Table 5: Akaike Information Criteria for the distributions fitted to the data.

That is why, once the individual distributions to be used are determined, the values of their respective parameters must be estimated according to the strategy defined in the previous section by MLE methodology. These parameters will be called initial parameters. We refer to them as initial values because in subsequent stages, they will be re-estimated according to the optimization algorithm. However, the final estimation does not differ significantly from the initial values.

	$\hat{r}$	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$
Values	0.3731	0.7175	0.6844	0.6334

Table 6: Estimated parameters for the discrete and continuous random variables.

With the new approach that include the use of expressions obtained from partial derivatives for the Bivariate Negative Binomial - Gamma Sarmanov distribution, the roots of equating the partial derivatives are found, considering the previous parameters for their calculus. These roots will define the maximum threshold that the parameters of the kernels δ and γ can take. In this way, we limit their values for the entire process. Maximum values are obtained as $\delta_{max} = 21.48$ and $\gamma_{max} = 142.05$. This implies that the iterative process needs to cover all possible values of $\delta, \gamma \in (0, (\delta_{max}, \gamma_{max})]$. However, analyzing all possible combinations of δ and γ entails a high computational cost. Therefore, the maximum value of the interval for both parameters has been set to 21, so that $\delta, \gamma \in (0, 21]$.

This represents a step forward in the investigation of the limits of the Sarmanov parameters. In previous studies, as can be found in Vernic et al. (2021), relatively small values were used for both upper limits. Another option was to impose maximum values without any theoretical foundation. This, in a way, restricts giving more freedom to the algorithm and therefore makes it difficult to obtain a better-fitting model if the AIC is taken into account.

Instantly, the problem that arises from this is which method for handling the dependency achieves a better fit to the data. For this reason, following the path of previous studies and the strategy explained in Section 5.1, three different estimation methods will be used. The first method significantly restricts the movement of δ and γ up to 2 and estimates the parameters in two phases, starting with the dependency and concluding with the remaining parameters. The second method follows the same principles as the first one but enlarges the intervals of the parameters of the exponential kernels by using the roots of the partial derivatives. Lastly, the last strategy sets the maximum values of the kernel parameters as in the second method, with the difference that the estimations are done in three phases: first, the dependency; then, the parameters of the discrete and continuous distributions, r, p, α, and β; and finally, δ and γ are estimated.

It is worth noting that the parameter estimation for any process is an iterative procedure that ends when convergence is reached. However, to reduce the computational cost the results presented here are obtained by applying a total of 10 iterations for each combination of δ and γ. There is no need to consider a larger number of iterations as convergence is achieved quickly in the majority of cases.

To facilitate the interpretation of the results, each model will be assigned to a number for easier understanding of future tables. This will allow easier identifications and analysis of the obtained results.

Model	**Method**
01	Standard method bounding δ and γ in the interval (0,2].
02	Standard method bounding δ and γ with the partial derivatives roots.
03	Three steps method bounding δ and γ with the partial derivatives roots.

Table 7: List of the models considered.

The final parameter estimates that maximize the log-likelihood are found in the last k iteration of the algorithm, in this case, in the tenth iteration. Table 8 provides detailed values for these estimates for each model. It can be observed that the parameter estimates for the individual distributions are very similar, with only slight differences in the decimal places. The significant differences lie in the parameters of the exponential kernels and the dependency. A common issue encountered when applying this algorithm is that excessive limitation of δ and

γ leads to underestimated estimates of ω, as well as the interval of ω itself. By expanding the upper limit of δ and γ, more flexibility is allowed for the dependence to adjust correctly to the data. Consequently, the minimum and maximum limits of ω also grow exponentially, as detailed in Table 9 using the expression (7). The difference between model 02 and model 03, which differ in the estimation approach for the parameters of the initial distributions and exponential kernels, is small. The only notable difference is found in the values of the variable associated with the dependence. In model 02, it takes a smaller value in comparison to model 03.

Model	**Initial** δ	**Initial** γ	r_{opt}	p_{opt}	α_{opt}	β_{opt}	δ_{opt}	γ_{opt}	ω_{opt}
01	1.9	1.3	0.374	0.718	0.677	0.634	2	2	12.91867
02	9.7	5.5	0.374	0.718	0.675	0.631	9.67	4.34	24630.84
03	9.7	4.3	0.373	0.718	0.684	0.633	9.69	4.35	24984.48

Table 8: Optimal estimated parameters obtained after ten iterations.

Model	ω_{min}	ω_{max}
01	-23.58	14.55
02	-80083.87	26450.23
03	-83336.03	26835.37

Table 9: Minimum and maximum values for ω at the last iterative process.

The differences in the final parameters ultimately affect the fit of the strategy, and therefore it is necessary to analyze the log-likelihoods or the AIC. Among the three strategies, the one that results in poorer results is the one that limits the parameters δ and γ to a maximum value of 2. Compared to models 02 and 03, model 01 differs the most as its AIC is the largest. On the other hand, the fit of the strategies 02 and 03 is relatively similar, although the former is slightly better. This suggests that the estimation of parameters for the initial distributions and the kernels should be optimized jointly rather than separately. To confirm this, further studies with more real data comparing both strategies should be conducted to determine which strategy should be considered for future applications. Furthermore, as seen in Table 10, obtaining a larger optimal ω does not necessarily lead to better results.

Another aspect of interest is the correlation captured by the Sarmanov model between the number of claims and their respective severity in cases where there have been claims. If independence is assumed, the correlation should be 0. However, the correlation captured by the Sarmanov model, as defined in (6), is approximately 0.022, $\rho_{X>0} \approx 0.022$, although it may vary depending on the strategy. Previously, the linear correlation between the two variables had been calculated in 4, and positive values around 0.15 were observed. Capturing this exact value of correlation in the Sarmanov distribution is not straightforward primarily because the linear component of the data is non-existent. Nevertheless, it is demonstrated that this dependency factor is being included in the estimation.

Model	**AIC**	**Sarmanov correlation with** $X > 0$
01	72187.12	0.02289
02	72111.74	0.02205
03	72113.30	0.02190

Table 10: Akaike Information Criteria and the obtained Sarmanov correlation between number and cost of claims for those policies that $X > 0$ for each estimated model.

Based on the results of the estimations for each strategy, one thing is clear: δ and γ cannot be excessively constrained. Larger intervals are needed to obtaina better fit of the data, and incorporating the roots of the partial derivatives as the maximum limit solve this issue. Nevertheless, the generated interval is quite wide, so the next steps should focus on finding the range of the initial δ and γ parameters that provide the optimal estimation of the distribution.

Regarding the strategies with two and three processes, it has been observed that the difference are small. However, considering that carbon footprint has gained significant relevance in recent years, strategy 02 is computationally more inefficient than 03. With a count of 10 iterations, model 02 takes around 2.62 seconds for each combination of initial δ and γ, while model 03 takes approximately 2.25 seconds. This implies a difference of approximately 26.72%. If the Sarmanov estimation is to be applied with thousands of combinations, strategy 02 is proved to be less efficient compared to strategy 03. This, coupled with the fact that a wider interval has been used for the parameters of the exponential kernels, emphasizes the need to narrow down their initial values and avoid having to explore every combination.

6. Conclusions

The Sarmanov distribution, despite not being widely known, can be applied in various areas of the insurance sector, such as loss and claims analysis or portfolio evaluation, among others, due to its property to incorporate a dependency factor among the involved random variables. Nevertheless, its practicality remains relatively low, primarily due to the need for further investigation of all aspects related to dependency, which still have room for improvement.

Based on the given definition of the Sarmanov distribution, the parameters δ and γ of the exponential kernels have a direct impact on the dependency factor ω. It has been observed that the parameter δ has a more significant effect on both negative and positive limits of the dependency ω, while the effects caused by γ are less notable. However, both parameters are important in the estimation of the model as they are necessary for maximizing the log-likelihood and, therefore, influence the model's performance.

The main problem lies in the limitation of the parameters δ and γ within a predefined interval. This interval is not explicitly defined, but in many cases, minimum and maximum values are imposed, which may not result in the best possible fit. Therefore, based on the analyses conducted in the study, and considering the Akaike information criterion as the metric to evaluate the model fit, expanding the intervals of δ and γ through the roots of the partial derivatives is a valid option that provides greater flexibility to the algorithm to find the optimal estimation of the kernel parameters and, consequently, of ω.

Additionally, the method frequently used to estimate or optimize the parameters has also been questioned. The approach utilized in previous studies involves an iterative process that includes two estimation phases: first, the estimation of dependency and then the estimation of the remaining parameters. The possibility of introducing a new estimation phase, where the optimal values of δ and γ are calculated separately from the parameters of the individual distributions, has also been explored. However, the fit obtained from this new strategy has been slightly worse compared to the one that use two estimation processes. Nevertheless, employing three estimation phases has one advantage: it reduces the carbon footprint by reducing its computational time.

This study is just the tip of the iceberg, opening up new research projects. In this way, the same process could be replicated considering new combinations of individual distributions, such as the Negative Binomial - Log-Normal case. The expressions for the partial derivatives are shown in this document, making their application straightforward. Other possibilities could include, by gathering new data, to verify whether the strategy with three processes provides better or similar results compared to using only two processes or developing an R package that incorporates the estimation of parameters considering the Sarmanov distribution as well. Nonetheless, there is still much work to be done in the field of research on the Sarmanov distribution to enhance its applicability.

References

Bolancé, C., Vernic, R., 2020. Frequency and severity dependence in the collective. Mathematics 2020 .

Kotz, S., Balakrishnan, N., Johnson, N., 2000. Continuous multivariate distributions, volume 1: Models and applications. .

Lee, T., 1996. Properties and applications of the sarmanov family of bivariate distributions. Theory and Methods 25, 1207-1222 .

Vernic, R., Bolancé, C., Alemany, R., 2021. Sarmanov distribution for modeling dependence between the frequency and the average severity of insurance claims. Insurance: Mathematics and Economics .