risks

Article

Rank-Based Multivariate Sarmanov for Modeling Dependence between Loss Reserves

Anas Abdallah * and Lan Wang

Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada; wangl139@mcmaster.ca
* Correspondence: anasabdallah@mcmaster.ca

Abstract: The interdependence between multiple lines of business has an important impact on determining loss reserves and risk capital, which are crucial for the solvency of a property and casualty (P&C) insurance company. In this work, we introduce the two-stage inference method using the Sarmanov family of multivariate distributions to the actuarial literature. In fact, we study rank-based methods using the Sarmanov distribution to adequately estimate the loss reserves and properly capture the dependence between lines of business. An inadequate choice of the dependence structure may negatively impact the estimation of the marginals and, hence, the reserve. Thus, we propose a two-stage inference strategy in this research to address this, while taking advantage of the flexibility of the Sarmanov distribution. We show that this strategy leads to a more robust estimation, and better captures the dependence between the risks. We also show that it generates smaller risk capital and a better diversification benefit. We extend the model to the multivariate case with more than two lines of business. To illustrate and validate our methods, we use three different sets of real data from both a major US property–casualty insurer and a large Canadian insurance company.

Keywords: rank-based methods; Sarmanov family of multivariate distributions; loss reserving; dependence; risk capital

check for updates

Citation: Abdallah, Anas, and Lan Wang. 2023. Rank-Based Multivariate Sarmanov for Modeling Dependence between Loss Reserves. *Risks* 11: 187. https://doi.org/10.3390/risks11110187

Academic Editors: Tak Kuen Ken Siu and Hailiang Yang

Received: 21 August 2023
Revised: 10 October 2023
Accepted: 21 October 2023
Published: 26 October 2023

1. Introduction

Insurance companies have an inverted production cycle, where they receive the premium (product price) before knowing the cost (claims). As a result, insurers must estimate these costs and set aside sufficient funds to meet their commitments to policyholders and claimants, creating what is known as reserves. Traditional reserving methods often assume independence among portfolio risk components. However, practical experience shows that risks are frequently interconnected, and this interdependence, represented by correlations between various lines of business, plays a crucial role in determining the overall portfolio reserve.

Dependence modeling plays a pivotal role within the insurance industry and the broader field of risk analysis. It is essential to comprehend the relationships among various variables or events. Dependence modeling serves as a valuable tool for quantifying and characterizing these relationships, ultimately enhancing the precision of risk assessments by accounting for dependencies that can either magnify or mitigate risks.

Furthermore, it facilitates superior portfolio diversification by offering insights into asset inter-dependencies, thereby reducing overall portfolio risk. Consequently, investors and financial institutions rely on dependence modeling to evaluate the risks associated with portfolios comprising multiple assets or financial instruments.

In domains such as financial markets, there exists a category of infrequent yet highly impacting events known as “tail events”, which significantly influence risk. Dependence modeling is instrumental in identifying these tail dependencies, a critical aspect of managing and mitigating tail risks.

Insurance companies harness dependence modeling to establish premium rates and effectively manage their risk exposure. Through an understanding of the correlations between different events or claims, insurers can accurately price policies and allocate capital to adequately cover potential losses.

Lastly, dependence modeling is of paramount importance in the realm of regulation and stress testing. Stress tests are conducted to assess the performance of a system or portfolio under adverse conditions. Dependence modeling is indispensable for crafting realistic stress scenarios that consider the intricate interplay between various risk factors.

In the context of loss reserving, understanding dependencies aids in predicting the necessary risk capital, which serves as a buffer for property and casualty (P&C) insurers against potential losses stemming from extreme and adverse events.

To calculate loss reserves, we utilize aggregated data, referred to as loss triangles. In these triangles, rows represent accident years, while columns represent development periods. The lower section of the triangle, which we aim to predict, represents future (unpaid) claims.

There are two main approaches used to capture the dependencies between different loss triangles. The first one focuses on distribution-free multivariate reserving methods. For example, Braun (2004) showed the effectiveness of the multivariate chain-ladder method using simulated data, demonstrating an increased estimation accuracy of the prediction error when accounting for the correlation between loss triangles. Merz and Wüthrich (2008) also studied the prediction error of a modified multivariate chain-ladder model proposed by Schmidt (2006) and incorporated a dependence structure into their model.

The other main approach for modeling dependence between business lines employs parametric methods, based on various distributional families. One commonly used method for parametric loss reserving is the copula model. For example, a Gaussian copula is used by Brehm (2002) to model the joint distribution of unpaid losses. De Jong (2012) used a Gaussian copula correlation matrix to model the dependence between lines of business. Shi et al. (2012) used multivariate Gaussian copula to capture correlation due to accounting years using loss triangles, while Merz et al. (2013) allowed the correlation matrix to vary over time and produced a more accurate depiction of dependence. Abdallah et al. (2015) used hierarchical Archimedean copulas to model dependence within and between lines of business. More recently, Shi (2017) conducted an analysis of multiple inter-company loss triangles using the Bayesian hierarchical model. Avanzi et al. (2016) introduced a multivariate Tweedie approach to capture cell-wise dependence in loss reserving, while Araiza Iturria et al. (2021) presented a stochastic model aimed at capturing dependencies between loss triangles. In their work, they opted for a Tweedie-distributed double-generalized linear model to represent the marginal distribution. Lally and Hartman (2018) used hierarchical Bayesian Gaussian process regression to estimate loss reserves across a spectrum of product lines. Additionally, Badounas and Pitselis (2020) explored the use of the quantile regression technique in the context of loss reserves.

Bootstrapping is also another popular parametric approach used for loss reserving, which involves resampling historical data to simulate and generate new (synthetic) datasets, also called pseudo-responses. Kirschner et al. (2002) proposed a synchronized bootstrap, which aimed to estimate the prediction error of a multivariate dependence model. Taylor and McGuire (2007) modified their approach to account for the additional complexity introduced by the generalized linear model framework. Shi and Frees (2011) used Frank and Gaussian copula to model the dependence between lines of business and introduced a parametric bootstrapping method to estimate the prediction error.

The contribution of this work to the actuarial literature in general, and to loss reserving in particular, is twofold. Firstly, this work introduces rank-based methods to the Sarmanov Family of distributions. This family is considered a richer and more flexible class of distributions for modeling dependence between risks, thanks to its flexible structure that nicely joins the marginals.

Second, we suggest direct pairwise dependence modeling for both bivariate and trivariate loss reserving analyses, using a rank-based Sarmanov for multivariate distributions applied to more than two lines of business.

Sarmanov's multivariate distribution, as described in Sarmanov's seminal work by Sarmanov (1966), has garnered significant attention in various corners of the actuarial literature. This distribution is noted for its tractability, its ability to accommodate a large array of flexible dependence structures, and for linking different marginal distributions. This adaptability has recently led to heightened interest across multiple realms of actuarial research.

Within the domain of rate-making, it has proven invaluable for in-depth severity analysis, as exemplified by the works of Hernández-Bastida and Fernández-Sánchez (2012) and Bahraoui et al. (2015). Additionally, it has been employed for frequency analysis, as seen in the studies by Abdallah et al. (2016a) and Bolancé and Vernic (2019). Notably, it has been instrumental in exploring the dependence between frequency and severity, as evidenced in the research by Vernic et al. (2022).

In the context of reserving, Sarmanov's distribution has been effectively utilized, as highlighted by Abdallah et al. (2016b). This application enabled the capture of dependencies between two lines of business through the incorporation of random effects. Sarmanov's distribution has also emerged as a valuable tool in analyzing ruin theory probabilities, as demonstrated by its application in studies conducted by Yang and Yuen (2016), Guo et al. (2017), and more recently, Chen et al. (2023).

Some of the referenced studies have demonstrated that, in comparison to alternative distributions like copulas, the Sarmanov family of distributions offers a superior fit to actual insurance data. For example, Bolancé and Vernic (2019) emphasize some disadvantages of the copula approach (e.g., elliptical) compared with the Sarmanov distributions. Moreover, Bahraoui et al. (2015) showed that the bivariate Sarmanov is more flexible than copulas in modeling dependence. Additionally, the correlation coefficients of Sarmanov's family of distributions have wider ranges; see Bahraoui et al. (2015) and Lee (1996) for more details. Bolancé et al. (2020) proposed a Sarmanov method with beta marginals and put it in use for motor insurance pricing.

The adaptability and extensive utility of Sarmanov's multivariate distribution have positioned it as a cornerstone in contemporary actuarial research. Its ability to navigate complex dependence structures has fostered a deeper understanding of dependencies and risk assessment in various segments of the insurance landscape.

In this paper, we specifically employ the Sarmanov distribution as we transition from the one-stage inference technique, where we simultaneously estimate both the marginals and the dependence parameters, to a two-stage inference modeling approach. Indeed, altering the dependence structure can result in distinct parameter estimations for the marginals, potentially leading to a different total reserve estimation. As a consequence, this method has the undesired effect of violating the linearity property of the mean. Therefore, we suggest employing a two-stage inference approach, commonly known as the rank-based method, utilizing the Sarmanov family of multivariate distributions. In the initial step, we fit generalized linear models (GLMs) to the individual marginals, establishing fixed parameters for the marginals and reserve estimations. Subsequently, we establish connections between the dependencies of these GLMs using the rank-based method, employing bivariate and trivariate Sarmanov distributions. It is worth noting that a similar approach has previously been explored using copula models. For example, Genest and Nešlehová (2014) discussed the rank-based methods for copula estimation, while Côté et al. (2016) introduced the rank-based methods for loss reserving, using nested Archimedean copulas, and a copula-based risk aggregation model.

The statistical properties of rank-based methods, including the consistency and asymptotic normality of estimators, were previously established by Genest et al. (1995). They conducted a comprehensive examination of a semi-parametric approach for estimating dependence parameters within a family of multivariate distributions. In this study, we showcase the practical applications of these methods and extend their utility to the multivariate

Sarmanov distribution family. In essence, our research demonstrates that the proposed method more effectively captures the dependencies among lines of business (LOBs) and yields lower risk capital estimates compared to traditional one-stage inference models.

Section 2 provides an overview of loss triangle modeling, introducing notations and presenting a concise overview of the Sarmanov distribution. Section 3 introduces the rank-based method to the Sarmanov family of multivariate distributions. For illustration and validation, Section 4 applies the model to seven LOBs from different datasets, sourced from a major US and a large Canadian property–casualty insurer. In Section 5, we analyze the implications for risk capital and demonstrate the advantages of our methods in terms of diversification benefits. Section 6 concludes the paper.

2. Preliminary

2.1. Modeling and Reserves

In this research, we use the generalized linear model (GLM) to model the marginals for each LOB (see De Jong and Heller 2008 and McCullagh and Nelder 1989 for a full review). GLMs provide the flexibility to choose the most adequate distribution for each LOB and perform a regression analysis that captures linear and non-linear relationships between the response and predictor variables.

In our case, the response variable is the incremental loss, while the accident year (rows) and development period (columns) are the predictor variables or covariates.

In fact, in a loss triangle, the row represents the year when an accident occurred, while the column represents each year that has passed (lag) since the accident happened. We use i and j to indicate the accident year and development period, respectively. Let $\ell \in \{1, \ldots, L\}$ represent the LOB, then we denote $X_{ij}^{(\ell)}$ as the incremental payment for the i^{th} accident year and the j^{th} development period. Also, let $p_i^{(\ell)}$ be the premium for the ℓ^{th} LOB and i^{th} accident year. As earned premiums vary by accident year, and to take into account the volume of each LOB, we work with standardized payments, $Y_{ij}^{(\ell)}$, such as

$$Y_{ij}^{(\ell)} = X_{ij}^{(l)} / p_i^{(\ell)}.$$

$Y_{ij}^{(\ell)}$ is called the incremental loss ratio, and it is a key performance metric used to measure the profitability of a P&C insurance company.

In order to fit the GLM, we perform the same procedure used by Abdallah et al. (2016b). Let $s_i^{(\ell)}$ be the effect of the accident year and $t_j^{(\ell)}$ be the effect of the development period, $i, j \in \{1, 2, \ldots, n\}$, then the systematic component for the ℓ^{th} LOB can be written as:

$$\eta_{ij}^{(\ell)} = u^{(\ell)} + s_i^{(\ell)} + t_j^{(\ell)},$$

where $u^{(\ell)}$ is the intercept; for parameter identification, we set $s_i^{(\ell)}$ and $t_j^{(\ell)}$ to 0 for $i,\ j = 1$.

Throughout the remainder of this paper and in our empirical illustration, we use both log-normal and gamma distributions for the different marginals and LOBs. More details about the fit and model selection are provided in Section 4.

When the log-normal distribution is assumed for the marginals, and to ease calculations, the incremental loss ratios is denoted by $Z_{ij}^{(\ell)}$ instead of $Y_{ij}^{(\ell)}$, i.e., $Z_{ij}^{(\ell)} \sim \mathcal{LN}(a_{ij}^{(\ell)}, b^{(\ell)})$, and then we use the change of variables $Y_{ij}^{(\ell)} = \log(Z_{ij}^{(\ell)}) \sim \mathcal{N}(a_{ij}^{(\ell)}, b^{(\ell)})$. We consider

$$a_{ij}^{(\ell)} = \eta_{ij}^{(\ell)},$$

with the location (log-scale) parameter $a_{ij}^{(\ell)}$ and shape parameter (standard deviation) $b^{(\ell)}$. As for the Gamma distribution, we have $Y_{ij}^{(\ell)} \sim \mathcal{G}(\alpha^{(\ell)}, \tau_{ij}^{(\ell)})$ and use the exponential link to ensure positive means

$$\tau_{ij}^{(\ell)} = exp(\eta_{ij}^{(\ell)})/\alpha^{(\ell)},$$

where the non-zero $\alpha^{(\ell)}$ is the shape parameter and $\tau_{ij}^{(\ell)}$ is the scale parameter. For parameter estimation, We use the maximum likelihood estimation (MLE), which is often favored over other classical estimation methods for its efficiency and asymptotic properties, as well as for the consistency and invariance of the estimators. Furthermore, MLE naturally gives rise to likelihood ratio tests, which serve as potent tools for conducting hypothesis tests and making informed decisions in model selection. These tests assist us in evaluating whether one model significantly outperforms another in our empirical illustration.

With the estimated parameters, the total reserve can be obtained as follows:

$$\sum_{\ell=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i^{(\ell)} E(y_{ij}^{(\ell)}), \tag{1}$$

where $E(y_{ij}^{(\ell)})$ is the projected unpaid loss ratio. More specifically, for log-normal distribution, we have

$$E(y_{ij}^{(\ell)}) = exp\left[a_{ij}^{(\ell)} + \frac{(b^{(\ell)})^2}{2}\right],$$

while for the gamma distribution, we have

$$E(y_{ij}^{(\ell)}) = \tau_{ij}^{(\ell)} \alpha^{(\ell)}.$$

2.2. *Sarmanov Distribution*

Sarmanov (1966) introduced Sarmanov's bivariate distribution to the literature, and Cohen (1984) suggested a more general form of bivariate Sarmanov in physics. A multivariate version was proposed by Lee (1996), who found applications in the medical area. As Sarmanov's distribution has a flexible structure, it attracted the attention of a wide range of applied studies. Johnson and Kott (1975) introduced the multivariate Farlie–Gumbel–Morgenstern (FGM) distribution, while Tank and Gebizlioglu (2004) proposed the Sarmanov class with FGM distribution properties for dependent risks. A bivariate Sarmanov model was used by Schweidel et al. (2008) to capture the relationship between a customer's waiting time and the actual service duration. Miravete (2009) used the Sarmanov model to compare the tariff plans between two related cellular telephone companies, and Danaher and Smith (2011) discussed some applications of Sarmanov to marketing. Bairamov et al. (2011) introduced a class of bivariate distributions, which generalizes the Sarmanov class.

In the insurance field, Sarmanov distributions have been used for pricing, reserving, and evaluating ruin probabilities. Detailed contributions to actuarial science were presented earlier in Section 1.

In this paper, we use the Sarmanov distribution to capture the pairwise dependence between two or more LOBs, in a loss-reserving context, which is presented in this section. We introduce the rank-based methods to the Sarmanov distribution, which is described in the next section.

2.2.1. Bivariate Sarmanov Distribution

In the case of two LOBs ($L = 2$), with $Y_{ij}^{(\ell)}$ denoting the incremental loss ratios from each LOB ($\ell \in \{1, 2\}$), let $f^{(\ell)}$ be the univariate probability density function, and $\psi^{(\ell)}(y_{ij}^{(\ell)})$ be non-constant functions, such that

$$\int_{-\infty}^{\infty} \psi^{(\ell)}(t) f^{(\ell)}(t) dt = 0.$$

Then the probability density function of the bivariate Sarmanov distribution is defined by

$$f^S(y_{ij}^{(1)}, y_{ij}^{(2)}) = f^{(1)}(y_{ij}^{(1)}) f^{(2)}(y_{ij}^{(2)}) \left(1 + \omega_{1,2} \psi^{(1)}(y_{ij}^{(1)}) \psi^{(2)}(y_{ij}^{(2)})\right), \tag{2}$$

with the mixing function:

$$\psi^{(\ell)}(y_{ij}^{(\ell)}) = exp(-y_{ij}^{(\ell)}) - L^{(\ell)}(1), \tag{3}$$

as proposed in Corollary 2 by Lee (1996), where $L^{(\ell)}$ is the Laplace transform of $f^{(\ell)}$, evaluated at 1. In (2), $\omega_{1,2}$ is the dependence parameter between LOBs 1 and 2.

When the marginal distribution follows a Gamma distribution, i.e., $Y_{ij}^{(\ell)} \sim \mathcal{G}(\alpha^{(\ell)}, \tau_{ij}^{(\ell)})$, then the mixing function is expressed as follows:

$$\psi^{(\ell)}(y_{ij}^{(\ell)}) = exp(-y_{ij}^{(\ell)}) - (1 + \tau_{ij}^{(\ell)})^{-\alpha^{(\ell)}}, \ \ \ell = 1, 2.$$

When the marginal distribution follows a log-normal distribution, as mentioned in Section 2.1, we have $Y_{ij}^{(\ell)} = \log(Z_{ij}^{(\ell)}) \sim \mathcal{N}(a_{ij}^{(\ell)}, b^{(\ell)})$. Consequently, the mixing function can be obtained as follows:

$$\psi^{(\ell)}(y_{ij}^{(\ell)}) = exp(-y_{ij}^{(\ell)}) - exp\Big(-a_{ij}^{(\ell)} + \frac{(b^{(\ell)})^2}{2}\Big), \ \ \ell = 1, 2.$$

The variable $\omega_{1,2}$ in (2) should be a real number that requires the constraint

$$1 + \omega_{1,2} \psi^{(1)}(y_{ij}^{(1)}) \psi^{(2)}(y_{ij}^{(2)}) \geq 0,$$

for all $y_{ij}^{(1)}, y_{ij}^{(2)}$.

For convenience, from now on, we denote $a_{ij}^{(\ell)}$ as a_ℓ, $b^{(\ell)}$ as b_ℓ, $\alpha^{(\ell)}$ as α_ℓ, and $\tau_{ij}^{(\ell)}$ as τ_ℓ.

As shown by Abdallah et al. (2016b), the bounds of the dependence parameter $\omega_{1,2}$ of the Sarmanov bivariate distribution, in the case of normal and gamma marginals for LOBs 1 and 2, respectively, are obtained as follows

$$-\frac{1}{b_1 exp(-a_1 + b_1^2/2)\sqrt{\alpha_2}\tau_2(1+\tau_2)^{-\alpha_2-1}} \leq \omega_{1,2} \leq \frac{1}{b_1 exp(-a_1 + b_1^2/2)\sqrt{\alpha_2}\tau_2(1+\tau_2)^{-\alpha_2-1}}.$$

Similarly, if the two LOBs both follow gamma distribution, then $\omega_{1,2}$ is bounded as follows

$$-\frac{1}{\sqrt{\alpha_1}\tau_1(1+\tau_1)^{-\alpha_1-1}\sqrt{\alpha_2}\tau_2(1+\tau_2)^{-\alpha_2-1}} \leq \omega_{1,2} \leq \frac{1}{\sqrt{\alpha_1}\tau_1(1+\tau_1)^{-\alpha_1-1}\sqrt{\alpha_2}\tau_2(1+\tau_2)^{-\alpha_2-1}}.$$

The proof of these results directly follows from Theorem 2 by Lee (1996).

2.2.2. Trivariate Sarmanov Distribution

The Sarmanov distribution can easily be generalized to the trivariate case thanks to its flexible structure. In this section, we introduce the Sarmanov distribution to capture dependence between more than two LOBs to the loss reserving literature. As such, we now work with three LOBs, with $Y_{ij}^{(\ell)}$, $\ell \in \{1, 2, 3\}$. The probability density function is then given as follows

$$f^S(y_{ij}^{(1)}, y_{ij}^{(2)}, y_{ij}^{(3)}) = f^{(1)}(y_{ij}^{(1)}) f^{(2)}(y_{ij}^{(2)}) f^{(3)}(y_{ij}^{(3)}) \times \big(1 + \omega_{1,2}\psi^{(1)}(y_{ij}^{(1)})\psi^{(2)}(y_{ij}^{(2)}) + \omega_{1,3}\psi^{(1)}(y_{ij}^{(1)})\psi^{(3)}(y_{ij}^{(3)}) + \omega_{2,3}\psi^{(2)}(y_{ij}^{(2)})\psi^{(3)}(y_{ij}^{(3)}) + \omega_{1,2,3}\psi^{(1)}(y_{ij}^{(1)})\psi^{(2)}(y_{ij}^{(2)})\psi^{(3)}(y_{ij}^{(3)})\big), \tag{4}$$

which is proposed by Theorem 4 by Lee (1996).

Additionally, as proposed by Drouet Mari and Kotz (2001) and mentioned by Ratovomirija et al. (2017), it is often assumed that $\omega_{1,\ldots,L} = 0$ for $\ell \geq 3$. Therefore, (4) is simplified to

$$f^S(y_{ij}^{(1)}, y_{ij}^{(2)}, y_{ij}^{(3)}) = f^{(1)}(y_{ij}^{(1)}) f^{(2)}(y_{ij}^{(2)}) f^{(3)}(y_{ij}^{(3)}) \times \big(1 + \omega_{1,2}\psi^{(1)}(y_{ij}^{(1)})\psi^{(2)}(y_{ij}^{(2)}) + \omega_{1,3}\psi^{(1)}(y_{ij}^{(1)})\psi^{(3)}(y_{ij}^{(3)}) + \omega_{2,3}\psi^{(2)}(y_{ij}^{(2)})\psi^{(3)}(y_{ij}^{(3)})\big). \tag{5}$$

The mixing function $\psi^{(\ell)}(y_{ij}^{(\ell)})$ is the same as (3). The dependence parameters $\omega_{1,2}, \omega_{1,3}$, and $\omega_{2,3}$ in (5) should be real numbers that require the following condition:

$$1 + \omega_{1,2}\psi^{(1)}(y_{ij}^{(1)})\psi^{(2)}(y_{ij}^{(2)}) + \omega_{1,3}\psi^{(1)}(y_{ij}^{(1)})\psi^{(3)}(y_{ij}^{(3)}) + \omega_{2,3}\psi^{(2)}(y_{ij}^{(2)})\psi^{(3)}(y_{ij}^{(3)}) \geq 0,$$

as shown by Ratovomirija et al. (2017). Also, Bolancé and Vernic (2019) showed that each bivariate case condition still needs to be applied. As such, we add the following restrictions for trivariate distribution:

$$1 + \omega_{c,d}\psi^{(c)}(y_{ij}^{(c)})\psi^{(d)}(y_{ij}^{(d)}) \geq 0, \quad 1 \leq c < d \leq 3.$$

2.3. One-Stage Inference for the Dependence Structure

For the bivariate case, the one-stage inference method estimates the two marginals and the dependence parameter $\omega_{1,2}$ simultaneously using maximum likelihood estimation. The log-likelihood of the bivariate Sarmanov distribution is given as follows:

$$\mathcal{L} = \sum_{i=1}^{n}\sum_{j=1}^{n+1-i} \log\,(f^{(1)}(y_{ij}^{(1)}) f^{(2)}(y_{ij}^{(2)})) + \sum_{i=1}^{n}\sum_{j=1}^{n+1-i} \log h(y_{ij}^{(1)}, y_{ij}^{(2)}; \omega_{1,2}),$$

where

$$h(y_{ij}^{(1)}, y_{ij}^{(2)}; \omega_{1,2}) = 1 + \omega_{1,2}\psi^{(1)}(y_{ij}^{(1)})\psi^{(2)}(y_{ij}^{(2)})$$

is the probability density function of the Sarmanov distribution dependence component.

Similarly, the one-stage inference method for trivariate Sarmanov distribution can be performed using the following log-likelihood function

$$\mathcal{L} = \sum_{i=1}^{n}\sum_{j=1}^{n+1-i} \log f^{(1)}(y_{ij}^{(1)}) f^{(2)}(y_{ij}^{(2)}) f^{(3)}(y_{ij}^{(3)}) + \sum_{i=1}^{n}\sum_{j=1}^{n+1-i} \log h(y_{ij}^{(1)}, y_{ij}^{(2)}, y_{ij}^{(3)}; \vec{\omega}), \tag{6}$$

where

$$h(y_{ij}^{(1)}, y_{ij}^{(2)}, y_{ij}^{(3)}; \vec{\omega}) = 1 + \omega_{1,2}\psi^{(1)}(y_{ij}^{(1)})\psi^{(2)}(y_{ij}^{(2)}) + \omega_{1,3}\psi^{(1)}(y_{ij}^{(1)})\psi^{(3)}(y_{ij}^{(3)}) + \omega_{2,3}\psi^{(2)}(y_{ij}^{(2)})\psi^{(3)}(y_{ij}^{(3)}),$$

is the probability density function of the Sarmanov distribution dependence component and $\vec{\omega} = (\omega_{1,2}, \omega_{1,3}, \omega_{2,3})$.

Again, from the trivariate log-likelihood function above, we estimate the dependence parameters $\vec{\omega}$ and the marginal parameters, simultaneously.

3. Rank-Based Sarmanov

3.1. Rank-Based Method and Risk Analysis

Rank-based methods offer numerous advantages over one-step inference techniques in statistical analysis. They exhibit exceptional resilience to outliers and extreme data values, minimizing the impact of data anomalies, which is a valuable trait when handling atypical data. Additionally, rank-based methods operate with fewer strict distributional assumptions or without them altogether. This versatility enables their effective application even when dealing with complex or undisclosed data distributions.

Furthermore, rank-based methods reduce the need for rigid model assumptions, granting greater flexibility to model intricate data relationships accurately. In multivariate analysis, rank-based methods shine by adeptly capturing dependencies between variables, especially in scenarios where these dependencies are nonlinear or non-monotonic.

In the context of risk analysis, where financial data can be influenced by outliers and extreme events, rank-based methods maintain their robustness. Their reduced sensitivity to extreme values makes them a robust choice for capturing the overall risk profile of portfolios and investments. Moreover, when constructing risk models, rank-based methods ease the burden of strict model assumptions, allowing for adaptability to various risk scenarios. In risk simulations and stress testing, rank-based methods prove invaluable for generating scenarios and evaluating the repercussions of extreme events, essential for effective risk management and capital allocation. The accessible and explicable nature of rank-based methods also facilitates comprehension by risk analysts and decision-makers, empowering them to make well-informed and timely decisions based on the results.

Also, risk analysis often deals with financial data that can be influenced by outliers and extreme events. Rank-based methods are less sensitive to extreme values, making them more robust at capturing the overall risk profile of a portfolio or investment. Risk models often involve assumptions about asset returns and correlations. Rank-based methods reduce the need for strict model assumptions, providing flexibility to adapt to different risk scenarios. Assessing tail risk, such as extreme losses, is a critical aspect of risk analysis. Rank-based methods are particularly effective at estimating tail risk measures, like value-at-risk (VaR) and tail value-at-risk (TVaR). In risk simulations and stress testing, rank-based methods are valuable for generating scenarios and assessing the impact of extreme events, which are essential for risk management and capital allocation. Rank-based methods often result in transparent and interpretive risk assessments, making it easier for risk analysts and decision-makers to understand the results and take appropriate actions. The risk capital implications are examined in Section 5.

In this section, we will leverage another crucial advantage of the rank-based method in risk analysis, specifically in the context of estimating reserves: its robustness in total reserve estimation. In fact, when using the one-stage inference method described in the previous section, the total reserve estimate in the presence of dependence does not equate to the sum of the marginal reserves estimated assuming independence. This is an aftereffect of the simultaneous estimation of the marginal and dependence parameters. An inadequate choice of the marginals may have an undesirable effect on the estimation of the dependence structure, and vice versa. Therefore, as mentioned in Section 2.1, once we estimate the parameters from the independence model, the total reserve can be calculated as follows

$$\sum_{\ell=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i^{(\ell)} E[y_{ij}^{(\ell)}].$$

However, in the one-stage inference, the marginal parameters in the presence of dependence may change, and deviate from those obtained with the independence model. This violates the linear property of the mean, as

$$E[\sum_{\ell=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}^{(\ell)}]$$

produced using the dependence model does not equal the total reserve

$$\sum_{\ell=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n}E[y_{ij}^{(\ell)}]$$

obtained with independence.

This paper addresses this inferential issue using the Sarmanov family of multivariate distributions. Thus, we propose using an alternative two-stage inference strategy, in which generalized linear models (GLMs) are first fitted to the marginals; in that way, we can fix the estimates of the reserves. In the second step, standardized residuals from those models are linked through a Sarmanov distribution to estimate the dependence structure using rank-based methods. This rank-based general approach has already been used in the copula modeling literature, we refer the reader to Genest and Favre (2007) or Genest and Nešlehová (2014) for a full review. However, to our knowledge, these techniques have never been applied to the Sarmanov family of multivariate distributions.

Therefore, we present a more robust estimation approach employing rank-based methods. We compare its outcomes with those of the one-stage inference strategy and evaluate its influence on both dependence estimation and risk capital analyses.

3.2. Rank-Based Method for Multivariate Sarmanov Distribution

As described above, using rank-based methods requires a two-stage inference method. First, we estimate the parameters of the marginals by maximizing the following log-likelihood of the marginals for the bivariate case

$$\mathcal{L}_{marginals} = \sum_{\ell=1}^{2}\sum_{i=1}^{n}\sum_{j=1}^{n+1-i}\log f^{(\ell)}(y_{ij}^{(\ell)}). \tag{7}$$

Next, we use the rank of residuals R_{ij} to estimate the dependence parameter, separately. The residuals of both log-normal and gamma distributions are expressed as follows, respectively,

$$r_{ij}^{(\ell)} = \frac{\log(y_{ij}^{(\ell)}) - a_{ij}^{(\ell)}}{b^{(\ell)}},$$

and

$$r_{ij}^{(\ell)} = \frac{y_{ij}^{(\ell)}}{\tau_{ij}^{(\ell)}}.$$

Starting from the residuals, we obtain the following rank of residuals

$$R_{ij}^{(\ell)} = \frac{1}{55+1}\sum_{i*=1}^{10}\sum_{j*=1}^{11-i*}\mathbf{1}(r_{i*j*}^{(\ell)} \leq r_{ij}^{(\ell)}), \tag{8}$$

with $\mathbf{1}(A)$ denoting the indicator function.

Consequently, the rank-based estimate $\hat{\omega}_{1,2}$ of the Sarmanov dependence parameter $\omega_{1,2}$ can be obtained from the loss-triangle data by maximizing the pseudo-log-likelihood:

$$\mathcal{L}(\omega_{1,2}) = \sum_{i=1}^{n}\sum_{j=1}^{n+1-i}\log h(R_{ij}^{(1)}, R_{ij}^{(2)}; \omega_{1,2}), \tag{9}$$

with

$$h(R_{ij}^{(1)}, R_{ij}^{(2)}; \omega_{1,2}) = (1 + \omega_{1,2}\psi^{(1)}(R_{ij}^{(1)})\psi^{(2)}(R_{ij}^{(2)})),$$

and

$$\psi^{(\ell)}(R_{ij}^{(\ell)}) = exp(-R_{ij}^{(\ell)}) - L^{(\ell)}(1). \tag{10}$$

Therefore, the bound of the parameter $\omega_{1,2}$ becomes

$$1 + \omega_{1,2}\psi^{(1)}(R_{ij}^{(1)})\psi^{(2)}(R_{ij}^{(2)}) \geq 0.$$

Similarly, for the trivariate rank-based Sarmanov distribution, we first use the maximum likelihood estimation for the parameters of the marginal

$$\mathcal{L}_{marginals} = \sum_{\ell=1}^{3}\sum_{i=1}^{n}\sum_{j=1}^{n+1-i} \log f^{(\ell)}(y_{ij}^{(\ell)}). \tag{11}$$

Then, we calculate the rank of residuals as described earlier, and then optimize the following pseudo-likelihood of the trivariate Sarmanov distribution to obtain the estimation of $\vec{\omega} = (\omega_{1,2}, \omega_{1,3}, \omega_{2,3})$

$$\mathcal{L}(\vec{\omega}) = \sum_{i=1}^{n}\sum_{j=1}^{n+1-i} \log h(R_{ij}^{(1)}, R_{ij}^{(2)}, R_{ij}^{(3)}; \vec{\omega}), \tag{12}$$

with

$$\begin{aligned} h(R_{ij}^{(1)}, R_{ij}^{(2)}, R_{ij}^{(3)}; \vec{\omega}) = {} & 1 + \omega_{1,2}\psi^{(1)}(R_{ij}^{(1)})\psi^{(2)}(R_{ij}^{(2)}) \\ & + \omega_{1,3}\psi^{(1)}(R_{ij}^{(1)})\psi^{(3)}(R_{ij}^{(3)}) + \omega_{2,3}\psi^{(2)}(R_{ij}^{(2)})\psi^{(3)}(R_{ij}^{(3)}). \end{aligned}$$

The mixing function for the trivariate case is the same as in (10).

Additionally, the bounds of the ω, for each $\omega_{c,d}$, $1 \leq c < d \leq 3$, need to satisfy the following constraints

$$1 + \omega_{1,2}\psi^{(1)}(R_{ij}^{(1)})\psi^{(2)}(R_{ij}^{(2)}) + \omega_{1,3}\psi^{(1)}(R_{ij}^{(1)})\psi^{(3)}(R_{ij}^{(3)}) + \omega_{2,3}\psi^{(2)}(R_{ij}^{(2)})\psi^{(3)}(R_{ij}^{(3)}) >= 0,$$

and

$$1 + \omega_{c,d}\psi^{(c)}(R_{ij}^{(c)})\psi^{(d)}(R_{ij}^{(d)}) \geq 0, \quad 1 \leq c < d \leq 3.$$

4. Empirical Analysis for Models Estimation

4.1. Data

To calibrate and validate our methods, we implement the models proposed in the previous sections with two sets of real data. For illustration, the model is first applied to two LOBs from a major US property–casualty insurer, and then to five LOBs from a large Canadian insurer.

4.1.1. US Schedule P Data

The first dataset comes from Schedule P of the National Association of Insurance Commissioners (NAIC) database, and was already used in the actuarial literature; see, e.g., Shi and Frees (2011) and Abdallah et al. (2016b). It consists of two loss triangles, from both personal and commercial automobile LOBs, respectively.

The NAIC is an organization created and governed by the head of insurance regulators from the whole US. It was created in 1871 to be used as a forum for information exchange and is one of the largest insurance regulatory databases. Schedule P presents losses and aggregated claims over a 10-year period, which can be arranged into loss triangles. It also provides the unpaid losses and premiums earned for all LOBs.

Each triangle contains data for accident years 1988–1997 and ten development years. The loss triangles of this dataset can be found in Appendix A, in Tables A1 and A2.

Shi and Frees (2011) assumed that the personal auto line follows a log-normal distribution and the commercial auto line follows a gamma distribution. The authors used visual and statistical tests to demonstrate the model fit for the marginals. We work with their conclusion and use the same distributions for each LOB.

Having changed the parametrization, we re-performed the Akaike information criterion (AIC) (see Akaike 1974) and the Kolmogorov–Smirnov (KS) goodness-of-fit test (see Berger and Zhou 2014) for the residuals of the personal auto line with the log-normal distribution, and the commercial auto line with the gamma distribution (log link). If the p-value for the KS test is bigger than the significance level, there is not enough evidence that the data do not come from the given distribution. In fact, Table A3 in Appendix A shows that there is no strong evidence against stating that the personal auto line follows a log-normal distribution and the commercial auto line follows a gamma distribution with a log link, although the fit of the commercial auto is borderline.

As a preliminary assessment of the dependence between the two LOBs, we use Kendall's τ. As shown by Genest et al. (2011), the formula used to calculate Kendall's τ for multiple datasets, such as residuals of multiple LOBs, is given below

$$\tau_{L,m} = \frac{1}{2^{L-1}-1}\left[-1+\frac{2^L}{m(m-1)}\sum_{(i,j)\neq(i*,j*)}\mathbf{1}\left(r^{(1)}_{i*j*}\leq r^{(1)}_{ij},\ldots,r^{(L)}_{i*j*}\leq r^{(L)}_{ij}\right)\right], \quad (13)$$

where L is the number of datasets (LOBs) and m is the data number (loss ratios) in each set.

To investigate the dependence between LOBs, we use Kendall's τ coefficient instead of the correlation coefficient throughout the remainder of this paper. This choice is particularly pertinent in the context of loss triangles because Kendall's τ coefficient effectively isolates and eliminates the influences of the accident year and/or development year effects. Indeed, Kendall's τ is a correlation measure based on ranks, evaluating the degree of association between variables by considering the order of their values rather than their specific numerical values. Furthermore, Kendall's τ offers a more robust assessment of association, as it is less susceptible to the influence of outliers when compared to certain other correlation coefficients, such as Pearson's correlation coefficient. Finally, it is worth noting that Kendall's τ is scale-insensitive, which proves advantageous in our illustration as we compare data and LOBs with varying volumes.

Kendall's τ between personal and commercial auto lines is presented in Table 1, which shows a negative dependence between the two LOBs.

Table 1. Kendall's τ for personal and commercial auto LOBs.

LOBs	Personal and Commercial
Kendall's τ	−0.1556

The results imply that the negative correlation between personal and commercial auto lines should not be dismissed, and the Sarmanov distribution can effectively account for this negative correlation, which is further demonstrated in the upcoming section. This carries significant implications from a risk management standpoint, as elaborated upon in Section 5.

4.1.2. Canadian Insurer Data 1

The second dataset comes from a large P&C Canadian insurer and was also already used in the actuarial literature, see Côté et al. (2016). It consists of two LOBs of an auto insurance product ($\ell = Auto$) and a home insurance product ($\ell = Home$) in all provinces combined. The data are in Appendix B and a descriptive summary of the two LOBs is presented in Table 2 below.

Table 2. Descriptive summary of two LOBs from a Canadian insurance company.

LOB	Region	Product	Coverage
Auto	West	Auto	Bodily injury
Home	Country-wide	Home	Liability

The auto LOB provides bodily injury (BI) coverage in Western Canada. BI coverage offers compensation to policyholders who are injured or killed in automobile accidents caused by uninsured vehicle owners or unidentified vehicles. Western Canada includes the provinces of Manitoba, Saskatchewan, Alberta, and British Columbia, as well as the Northwest Territories, Yukon, and Nunavut.

The Home LOB encompasses the company's nationwide personal and commercial home insurance offerings. Liability insurance within this LOB safeguards policyholders against legal liabilities arising from injuries or damage caused to others.

Côté et al. (2016) showed that both LOBs follow a gamma distribution using the AIC and KS goodness-of-fit test. The results are reproduced in Table A7 in Appendix B.

For the exploratory dependence analysis, we work with Kendall's τ, for the reasons mentioned earlier. The results are presented in Table 3 and show a positive dependence between these LOBs.

Table 3. Kendall's τ for auto and home LOBs.

LOBs	Auto and Home
Kendall's τ	0.2848

The fact that the two LOBs are positively correlated is partly due to exogenous common factors, such as inflation and interest rates. Furthermore, strategic decisions can impact several lines within the insurance product, e.g., the acceleration of payments on all lines of the auto insurance sector could induce some positive dependence across the whole portfolio.

4.1.3. Canadian Insurer Data 2

The third dataset is also sourced from a prominent Canadian property and casualty (P&C) insurer and has been previously utilized in actuarial research, see Côté et al. (2016). We focus on the three automobile insurance LOBs from the province of Ontario and apply both the bivariate and trivariate Sarmanov models. The three LOBs consist of Ontario bodily injuries ($\ell = BI$), Ontario accident benefits excluding disability income ($\ell = AB$), and Ontario accident benefits with disability income ($\ell = DI$). The data are in Appendix C and a descriptive summary of the three LOBs is presented in Table 4 below.

Table 4. Descriptive summary of three LOBs from a Canadian insurance company.

LOB	Region	Product	Coverage
BI	Ontario	Auto	Bodily injury
AB	Ontario	Auto	Accident benefits excluding disability income
DI	Ontario	Auto	Accident benefits: disability income only

BI coverage is described earlier; the accident benefits (AB) coverage provides compensation for injury or death involved in a vehicle collision, regardless of fault, if you, your passengers, or pedestrians are injured or killed due to the accident. Disability income provides compensation if the accident results in a disability and the insured is not able to work at their regular employment anymore.

The three LOBs again follow a gamma distribution, which we showed using the AIC and KS goodness-of-fit test. The results are provided in Table A12 in Appendix C.

For the preliminary investigation of dependence, we work with Kendall's τ, which are presented in Table 5 and show positive dependence between these LOBs.

Table 5. Kendall's τ for the BI, AB, and DI LOBs.

LOB	Line BI and AB	Line BI and DI	AB and DI	BI, AB, and DI
Kendall's τ	0.2444	0.2094	0.2000	0.2180

The positive correlation among the three LOBs can be attributed to shared strategic decisions, external factors discussed in relation to the dependence between auto and home LOBs mentioned earlier, as well as legislative changes within the province of Ontario. Moreover, when we delve into the granular level, the positive relationship between Ontario's AB and BI can be elucidated by the frequent occurrence of the same accidents triggering claims in both coverage types.

4.2. One-Stage Inference Analysis

4.2.1. US Schedule P Data Calibration

As a starting point, we use the one-stage estimation method in this section, to estimate the ω dependence parameter for the bivariate Sarmanov model for the personal and commercial auto lines from the US Schedule P data. The results are shown in Table 6.

Table 6. Estimated omega for the bivariate Sarmanov model with personal and commercial LOBs using the one-stage inference method.

LOB	Estimated Omega	Log-Likelihood
Personal and Commercial	-4.4296	348.6252

As elucidated in Section 3 of Lee (1996), the sign of the ω dependence parameter for Sarmanov distribution is contingent upon the interdependence observed between the LOBs. Notably, Kendall's τ in Table 1 signifies negative dependence between personal and commercial LOBs. Consequently, Table 6 presents a negative dependence parameter ω for the bivariate Sarmanov distribution of these two LOBs, which is corroborated by the negative dependence identified using Kendall's τ.

We use several inference tests to check the significance of the dependence parameter. From the AIC result in Table 7, we can see that it shows that the bivariate Sarmanov model using a one-stage inference method for personal and commercial auto lines is better than the independent case. The smaller AIC is presented in bold. We also use the likelihood-ratio test (see Woolf 1957) to check whether the ω dependence parameter is significant.

Table 7. AIC for the bivariate Sarmanov model with personal and commercial LOBs using the one-stage inference method.

Model for Personal and Commercial	AIC
Independence	-613.1788
Bivariate Sarmanov with one-stage inference	$\mathbf{-615.2503}$

In Table 8, we see that the *p*-value is lower than 0.05, which indicates that we can reject the null (independence) hypothesis at a 5% level. This agrees with the AIC result and shows that the ω dependence parameter of the bivariate Sarmanov model is significant and this model captures the dependence between personal and commercial LOBs.

Table 8. Significance tests for the bivariate Sarmanov model with personal and commercial LOBs using the one-stage inference method.

Significance Tests	Likelihood-Ratio Test
Test statistic	4.0639
p-value	0.0438

After obtaining the estimated parameters for one-stage inference for personal and commercial LOBs, we use them to calculate the estimated total reserve as denoted in (1), and the results are detailed in Table 9.

Table 9. Reserve calculation of the one-stage inference method vs. independence for personal and commercial LOBs.

Models/Reserve	LOB Pers.	LOB Comm.	Total
Independence Pers. and Comm.	6,464,075	490,652	6,954,727
Bivariate Pers. and Comm.	6,465,679	513,622	6,979,302

Notably, Table 9 reveals that the total reserve computed using the one-stage inference method departs from the independent case, violating the linearity property of the mean, as discussed in Section 3.1. From a practical perspective, this also represents an undesirable outcome, as it means that the reserve of one LOB is influenced by the reserve of another.

4.2.2. Canadian Insurer Data 1 Calibration

We now use the second dataset, which consists of the auto and home pair LOBs, described in the previous section, to calibrate the bivariate Sarmanov model with a one-stage inference strategy. Table 10 presents the ω dependence parameter estimation for this pair of LOBs.

Table 10. Estimated omega for the bivariate Sarmanov model with auto and home LOBs using the one-stage inference method.

LOB	Estimated Omega	Log-Likelihood
Auto and Home	256.0006	335.3081

As discussed in Section 4.1.2, the home and auto LOBs are positively dependent. This positive relationship is corroborated by the estimated positive value of the ω dependence parameter within Table 10 for the bivariate Sarmanov model.

Once more, we employ both the Akaike information criterion (AIC) and the likelihood-ratio test (LRT) as the initial tools to evaluate the presence of statistically significant dependence within these pairs.

The AIC results presented in Table 11 indicate that the independence model outperforms the one-stage inference bivariate model for the auto and home pair, as evidenced by its lower AIC value, which is shown in bold. This observation is further substantiated by the LRT results, as demonstrated in Table 12, which reveal that, for the auto and home pair, the null hypothesis of independence cannot be rejected at the 5% significance level. Consequently, we conclude that the bivariate Sarmanov model with one-stage inference falls short in capturing the dependence between the auto and home LOBs.

Table 11. AIC for the bivariate Sarmanov model with auto and home LOBs using the one-stage inference method.

Model for Auto and Home	AIC
Independence	**−590.1085**
Bivariate Sarmanov with one-stage inference	−588.6163

Table 12. Significance tests for the bivariate Sarmanov model with auto and home LOBs using the one-stage inference method.

Significance Tests	Likelihood-Ratio Test
Test statistic	0.5078
p-value	0.4761

Similarly, following the estimation of parameters for the one-stage inference method for home and auto LOBs, we apply these estimates to compute the expected total reserve, as outlined in (1). The resulting insights are presented in Table 13.

Table 13. Reserve calculation of the one-stage inference method vs. independence for auto and home LOBs.

Models/Reserve	LOB Auto	LOB Home	Total
Independence Auto and Home	78,665	98,929	177,594
Bivariate Auto and Home	77,789	101,603	179,392

Analogous to the previous analysis in Section 4.2.1, Table 13 a deviation in the total reserve determined using the one-stage inference method compared to the independent case, thereby contradicting the linearity property of the mean and industry best practices.

4.2.3. Canadian Insurer Data 2 Calibration

We now use the third dataset, from a large Canadian insurer, to calibrate both bivariate and trivariate Sarmanov models with a one-stage inference approach. Table 14 presents the ω dependence parameter estimation for the BI and AB, BI and DI, and AB and DI LOB pairs described in the previous section.

Table 14. Estimated omega for the bivariate Sarmanov model with the BI and AB, BI and DI, and AB and DI LOBs using the one-stage inference method.

LOB	Estimated Omega	Log-Likelihood
BI and AB	436.9040	315.1206
BI and DI	424.5868	397.3058
AB and DI	730.9298	400.3068

In Table 14, the ω dependence parameters of all three pairs agree with Kendall's τ statistics in Section 4.1.3, showing positive dependencies for each pair of LOBs.

Again, we use the AIC as well as the likelihood-ratio test (LRT) to first assess whether there is any significant dependence between these pairs.

The AIC results in Table 15 show that the bivariate Sarmanov model with the one-inference method for the BI and AB pair provides a better fit than the independence case. However, the results for the BI and DI, and AB and DI pairs show that the independence model has a smaller AIC than the one-stage inference bivariate model. We indicate the smaller AICs in bold. These findings are confirmed with the LRT.

Table 15. AIC for the bivariate Sarmanov model with BI and AB, BI and DI, and AB and DI LOBs using the one-stage inference method.

LOB	Model	AIC
BI and AB	Independence Bivariate Sarmanov with one-stage inference	−546.3281 **−548.2413**
BI and DI	Independence Bivariate Sarmanov with one-stage inference	**−714.1499** −712.6117
AB and DI	Independence Bivariate Sarmanov with one-stage inference	**−720.1422** −718.6135

Table 16 shows that, for the BI and AB pair, the null (independence) hypothesis is rejected at the 5% level, but cannot be rejected for BI and DI and AB and DI. We then conclude that the bivariate Sarmanov model with one-stage inference only captures the dependence between the BI and AB LOBs.

Table 16. Significance tests for the bivariate Sarmanov model with BI and AB, BI and DI, AB, and DI LOBs using the one-stage inference method.

LOB/Likelihood-Ratio Test	Test Statistics	*p*-Value
BI and AB	3.91314	0.04791
BI and DI	0.46678	0.49447
AB and DI	0.47131	0.49239

For the trivariate case, we use the three LOBs—BI, AB, and DI—from the Canadian insurer dataset. We first need to estimate the dependence parameters $\omega_{1,2}$, $\omega_{1,3}$, and $\omega_{2,3}$ from (5), using the one-stage inference method. We then maximize the log-likelihood function presented in (6). The results are shown in Table 17.

Table 17. Estimated omega for the trivariate Sarmanov model with BI, AB, and DI LOBs using the one-stage inference method.

Lines BI, AB, and DI	$\omega_{BI,AB}$	$\omega_{BI,DI}$	$\omega_{AB,DI}$
Estimated omega	374.7942	−110.3272	−165.7813
Log-likelihood		556.4291	

Table 17 reveals that not all of the dependence parameters exhibit positive values, which contradicts the initial findings from the dependence analysis presented in Table 5. This suggests that the obtained dependence parameters may not be statistically significant. Therefore, we proceed with the likelihood-ratio test (LRT) to evaluate the significance of these parameters.

Table 18 validates that the three dependence parameters lack statistical significance, with *p*-values exceeding 10% for each of them.

Additionally, we use the AIC and LRT to check if the model is significant.

The results from Table 19 show that the trivariate Sarmanov model using the one-stage inference method is not better than the independence model for lines BI, AB, and DI, i.e., it does not show significant dependence for the LOB triplet (BI, AB, and DI). The smaller AIC is highlighted in bold. This finding is also confirmed by the likelihood-ratio test in Table 20.

Table 18. Significance test for the trivariate Sarmanov model with the BI, AB, and DI lines using the one-step inference method.

Likelihood-Ratio Test	$\omega_{BI,AB}$	$\omega_{BI,DI}$	$\omega_{AB,DI}$
Test statistic	2.2803	−0.1351384	1.061
p-value	0.1310	1.00	0.3030

Table 19. AIC for the trivariate Sarmanov model with BI, AB, and DI LOBs using the one-stage inference method.

Model for Line BI, AB, and DI	AIC
Independence	**−1026.6996**
Trivariate Sarmanov with one-stage inference	−986.8582

Table 20. Significance tests for the trivariate Sarmanov model with BI, AB, and DI LOBs using the one-stage inference method.

Significance Tests	Likelihood-Ratio Test
Test statistic	2.5506
p-value	0.4662

Therefore, we conclude from the results above that the Sarmanov one-stage inference model fails to capture the dependence among the triplet BI, AB, and DI.

Once we obtain the estimated parameters, we use them to compute the predicted total reserve, expressed in (1), and the results are reported in Table 21.

Table 21. Reserve calculation of the one-stage inference method vs. independence for the BI, AB, and DI LOBs.

Models/Reserve	LOB BI	LOB AB	LOB DI	Total for 3 Lines
Independence BI, AB, and DI	132,918	73,220	18,288	224,426
Bivariate BI and AB	129,397	71,457	(18,288)	219,144
Bivariate BI and DI	131,148	(73,220)	18,739	223,107
Bivariate AB and DI	(132,918)	72,144	18,123	223,185
Trivariate BI, AB, and DI	135,061	70,857	18,752	224,671

As the dependence parameter of the bivariate Sarmanov for the BI and DI and AB and DI pairs, as well as the trivariate Sarmanov for the triplet BI, AB, and DI, are not significant, their corresponding total estimated reserve aligns closely with the independence case. However, when the dependence becomes significant, as with the bivariate Sarmanov for the BI and AB pairs, the corresponding total reserve deviates more from the reserve obtained in the independence case.

4.3. Rank-Based Method Analysis

For the rank-based method, we first use Kendall's test to check whether there is any significant dependence between the residuals of the different LOBs.

The calculation of Kendall's τ has already been presented in (13). Under the null hypothesis of multivariate independence, the mean of $\tau_{L,m}$ is 0, and its sample variance can be calculated as follows:

$$Var(\tau_{L,m}) = \frac{m(2^{2L+1} + 2^{L+1} - 4 * 3^L) + 3^L(2^L + 6) - 2^{L+2}(2^L + 1)}{3^L(2^{L-1} - 1)^2 m(m-1)},$$

and the distribution of $\tau_{L,m}$ is assumed to be asymptotically Gaussian. As Kendall's test uses the chi-square test to determine the p-value, the latter is expressed as

$$p = 2 * \left(1 - cdf_{normal}\left(|\tau_{L,m}/\sqrt{Var(\tau_{L,m})}|\right)\right).$$

Next, we report the results for both datasets in the two following subsections, respectively.

4.3.1. US Schedule P Data Calibration

Here, we first start by checking the dependence between the residuals of the personal and commercial auto line from the US Schedule P data.

Based on the p-value of Kendall's test presented in Table 22, we conclude that the null hypothesis of independence is rejected at the 10% level. Therefore, we can say that there exists a significant (but small) negative dependence between the two LOBs. In the case of negative association, it is preferred to work with the anti-ranks (negative of rank of residuals) for the second LOB when estimating $\omega_{1,2}$, as suggested by Côté et al. (2016). Thus, we optimize the following pseudo-likelihood function:

$$\mathcal{L} = \sum_{i=1}^{n} \sum_{j=1}^{n+1-i} log\ h(R_{ij}^{(1)}, -R_{ij}^{(2)}, \omega_{1,2}).$$

This allows us to obtain the estimated $\omega_{1,2}$ in Table 23.

Table 22. Kendall's τ for personal and commercial LOBs.

LOB	Personal and Commercial Auto Line
Kendall's τ	-0.1556
Kendall's test p-value	0.09355

Table 23. Estimated omega for the bivariate Sarmanov model with personal and commercial LOBs using the rank-based method.

LOB	Estimated Omega
Personal and Commercial	-10.14954

The sign of the estimated ω dependence parameter in Table 23 also confirms the negative dependency between personal and commercial LOBs.

When we work with rank-based methods and pseudo-likelihood functions, the diagnostic tools for dependence significance that were used with the one-stage inference methods, such as AIC and LRT, cannot be used anymore. However, bootstrapping can be used to check whether a parameter is significant, as pointed out by Côté et al. (2016). If we simulate and estimate the parameter 5000 times, then we can check if the 95% confidence interval of the 5000 estimation includes 0. If it does not include 0, then the estimated parameter is significant.

We use the bootstrapping method to check whether the ω dependence parameter is significant. We simulate dependent loss triangles from the estimated ω using the rank-based bivariate Sarmanov and re-estimate the corresponding dependence parameter ω^* from each simulated pair of loss triangles. The simulation and bootstrapping procedures are illustrated thoroughly in the next section.

Figure 1 shows the approximate distribution of the ω based on 5000 bootstrap replicates.

The blue line in Figure 1 represents the 95% confidence interval for the parameter and we can see that the confidence interval does not include 0. This indicates that the estimation of $\omega_{P,C}$ is significant in the bivariate Sarmanov model using the rank-based method for personal and commercial auto lines.

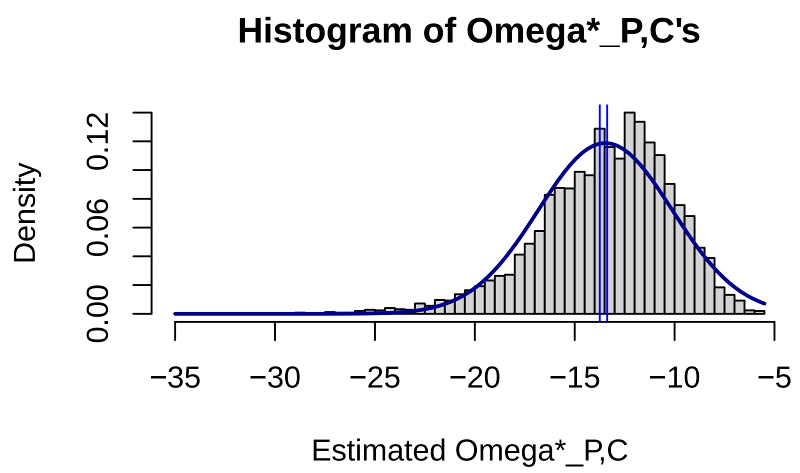

Figure 1. The 5000 $\omega^*_{P,C}$ rank-based bootstrap estimations for bivariate Sarmanov with Pers. and Comm. LOBs.

4.3.2. Canadian Insurer Data 1 Calibration

Now, we perform the same procedure for the first dataset from the Canadian insurer, which consists of auto and home LOBs. Table 24 presents Kendall's τ test between the two LOBs.

Table 24. Kendall's τ for auto and home LOBs.

LOB	Auto and Home Lines
Kendall's τ	0.2848
Kendall's test p-value	0.0021

As depicted in Table 24, a positive dependence exists between the two LOBs. In fact, the p-value obtained from Kendall's test underscores a robust dependency between these two lines of business.

Once more, we use the Sarmanov bivariate model and apply (7) and (8) for the gamma–gamma case to calculate the rank of residuals, which we subsequently insert into (9) to estimate the dependence parameter, denoted as ω. The outcome of the ω estimation through the rank-based method for the auto and home LOBs is presented in Table 25.

Table 25. Estimated omega for the bivariate Sarmanov model with personal and commercial LOBs using the rank-based method.

LOB	Estimated Omega
Auto and Home	155.115

The estimated ω in Table 25 agrees with Kendall's τ test above, showing positive dependencies between auto and home LOBs.

Once again, we can employ the bootstrapping method to assess the significance of the ω values. To achieve this, in a similar manner, we simulate synthetic (dependent) loss triangles using the dependence parameters acquired from Table 25 through the bivariate rank-based Sarmanov model. Subsequently, we re-estimate the new ω^* values for each iteration; the results are presented in Figure 2.

In Figure 2, the blue lines represent the 95% confidence interval. It is evident from the figure that the blue lines do not encompass the value 0 for the parameters. This indicates that the estimation of ω values holds significance in the bivariate Sarmanov model when employing the rank-based method for the auto and home LOBs.

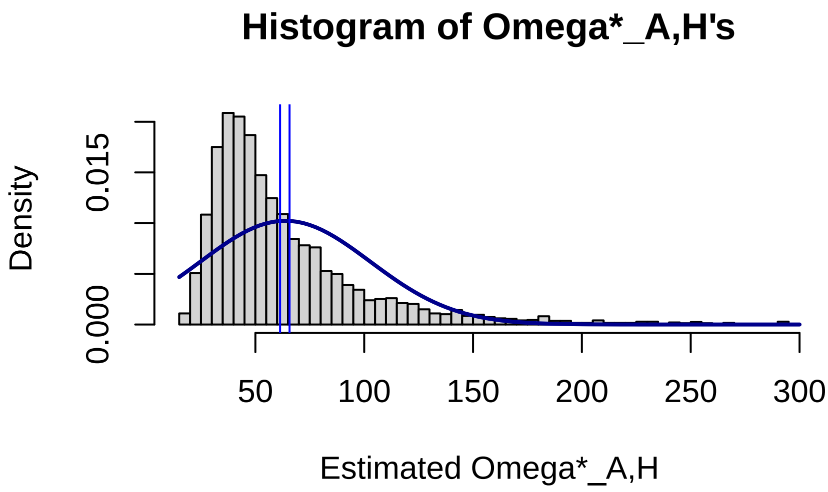

Figure 2. The 5000 $\omega^*_{A,H}$ rank-based bootstrap estimations for bivariate Sarmanov with auto and home LOBs.

4.3.3. Canadian Insurer Data 2 Calibration

Now, we perform the same procedure for the BI, AB, and DI LOBs from the other Canadian insurer dataset. Table 26 presents Kendall's τ tests for all three LOBs.

Table 26. Kendall's τ for the BI, AB, and DI LOBs.

LOB	Line BI and AB	Line BI and DI	Line AB and DI	Line BI, AB, and DI
Kendall's τ	0.2444	0.2094	0.2000	0.2180
Kendall's test p-value	0.0084	0.0240	0.0311	4.7064×10^{-5}

Table 26 shows that the three LOBs are positively correlated. The p-value of Kendall's test shows that there is a strong dependence between the three lines together.

We first consider the bivariate dependence between the BI and AB, BI and DI, and AB and DI pairs, and we examine the trivariate case afterward.

In the bivariate model, we use (7) and (8) in the gamma–gamma case to compute the rank of residuals that we plug in (9) to estimate the omega dependence parameter. Table 27 presents the result of the ω estimation using a rank-based method for the following pairs: BI and AB, BI and DI, and AB and DI.

Table 27. Estimated omega for the bivariate Sarmanov model with the following pairs: BI and AB, BI and DI, AB and DI, using a rank-based method.

LOB	Estimated Omega
BI and AB	24.524
BI and DI	31.482
AB and DI	1374.157

The estimated ω in Table 27 also shows that there are positive dependencies between each pair of LOBs.

Again, we can use the bootstrapping method to check the significance of the ω. As such, similarly, we simulate the synthetic (dependent) loss triangles using the dependence parameters obtained in Table 27 with the bivariate rank-based Sarmanov and re-estimate the new ω^*'s each time.

Similarly, Figures 3–5 present the omega estimates of the bootstrap result, where the blue lines denote the 95% confidence interval, and we can see from the figure that the blue lines do not include 0 for the parameters. This means that the ω estimation is significant in the bivariate Sarmanov model using the rank-based method for the BI and AB, BI and DI, and AB and DI pairs. These figures give some indications that the distribution of the estimated parameters may not be normal, this could be because of the bounds while estimating the ω.

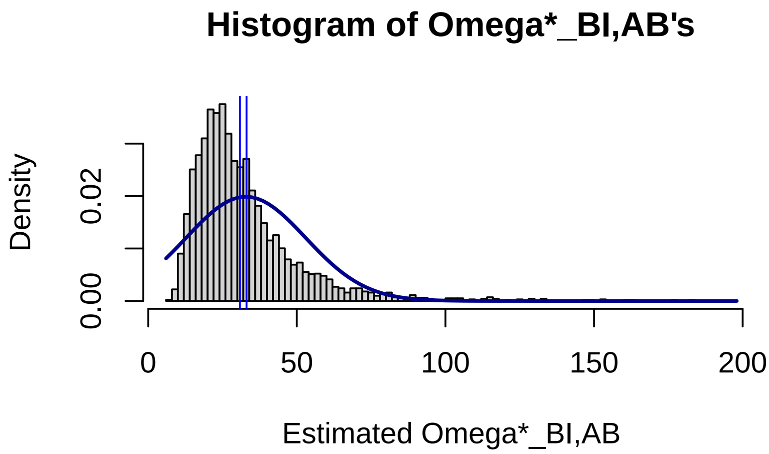

Figure 3. The 5000 $\omega^*_{BI,AB}$ rank-based bootstrap estimations for the bivariate Sarmanov with the BI and AB LOBs.

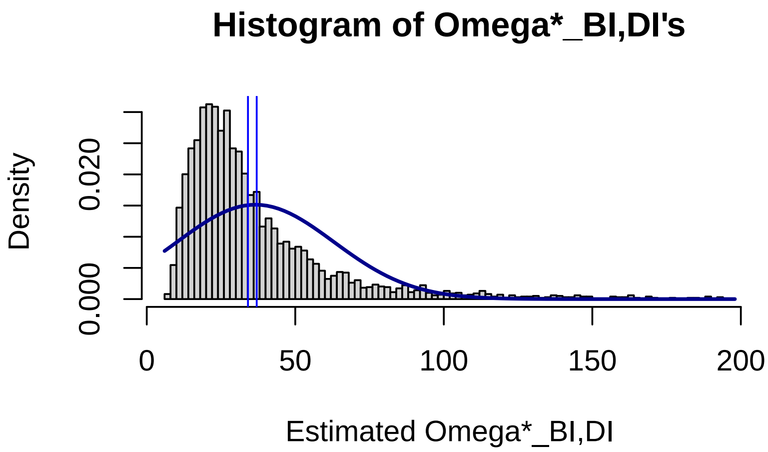

Figure 4. The 5000 $\omega^*_{BI,DI}$ rank-based bootstrap estimations for the bivariate Sarmanov with the BI and DI LOBs.

For the trivariate case, we estimate the three dependence parameters $\omega_{1,2}$, $\omega_{1,3}$, and $\omega_{2,3}$, from (12) after calculating the rank of residuals using (11) and (8). The estimated ωs are presented in Table 28.

As shown in Table 26, Kendall's test shows that the three LOBs are positively dependent on each other, which is confirmed by the signs of the estimated dependence parameter ω in Table 28. We again use the bootstrapping method to check the significance of the three dependence parameters.

Table 28. Estimated omega for the trivariate Sarmanov model with the triplet BI, AB, and DI using the rank-based method.

Lines BI, AB, and DI	$\omega_{BI,AB}$	$\omega_{BI,DI}$	$\omega_{AB,DI}$
Estimated omega	25.2962	30.4092	61.4528

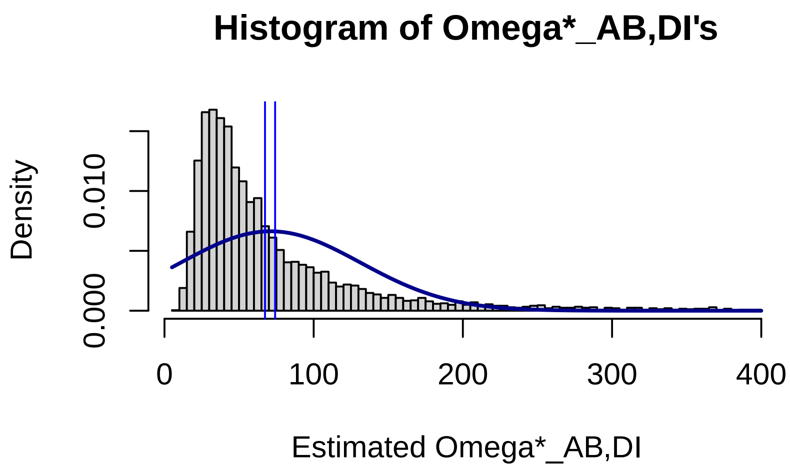

Figure 5. The 5000 $\omega^*_{AB,DI}$ rank-based bootstrap estimations for the bivariate Sarmanov with the AB and DI LOBs.

The estimated omegas of bootstrap results are presented in Figures 6–8.

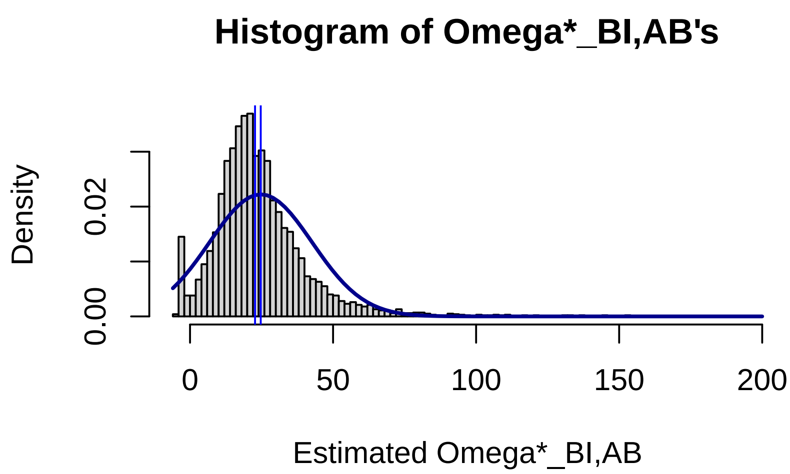

Figure 6. The 5000 $\omega^*_{BI,AB}$ rank-based bootstrap estimations for bivariate Sarmanov with BI, AB, and DI LOBs.

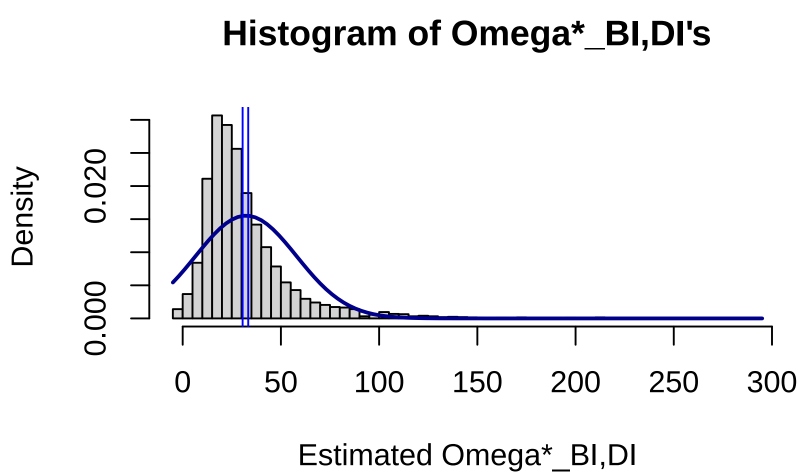

Figure 7. The 5000 $\omega^*_{BI,DI}$ rank-based bootstrap estimations for bivariate Sarmanov with BI, AB, and DI LOBs.

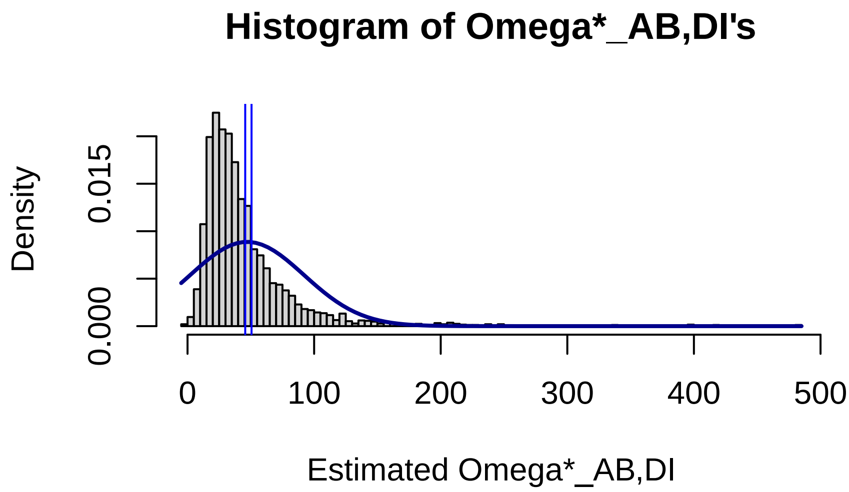

Figure 8. The 5000 $\omega^*_{AB,DI}$ rank-based bootstrap estimations for bivariate Sarmanov with BI, AB, and DI LOBs.

From Figures 6–8, we can conclude that the dependence parameters are all significant for the trivariate Sarmanov distribution using the rank-based method, as the 95% confidence interval (blue lines) does not include 0 for each figure. Interestingly, this trivariate dependence was not captured with the classical one-stage inference method.

4.4. Models Estimation Summary

Table 29 displays a summary of the comparison between the Sarmanov rank-based model and the classical one-stage inference model for all seven LOBs, considering both bivariate and trivariate cases. The results clearly demonstrate that the rank-based method more effectively captures the dependencies among the LOBs. This enhanced understanding of dependencies leads to a more comprehensive risk capital analysis and greater diversification benefits, which are elaborated in the following section.

Table 29. Summary table for the comparison of the one-stage inference method and rank-based method.

Significance of Models/Methods	One-Stage Inference Method	Rank-Based Method
Bivariate Personal and Commercial	✓	✓
Bivariate Auto and Home	×	✓
Bivariate BI and AB	✓	✓
Bivariate BI and DI	×	✓
Bivariate AB and DI	×	✓
Trivariate BI, AB, and DI	×	✓

5. Risk Capital Implications

In addition to reserves, companies also need to set aside additional funds as a buffer in case of potential losses caused by adverse scenarios or extreme events; it is called risk capital. It represents the amount of money that the companies can lose without causing significant harm to the financial situation. In practice, companies calculate their risk capital by summing up the risk capital of each LOB separately. This is called the "Silo" method; it was introduced by Ajne (1994). However, this method implicitly assumes that risks are perfectly correlated, and does not allow any forms of diversification.

Therefore, we address this issue by using a dependence model through the Sarmanov family of multivariate distributions, with both the one-stage inference and rank-based methods. This section then examines and compares both approaches and assesses their impacts on the risk capital and diversification benefits.

In order to calculate the risk capital, risk measures, such as the value-at-risk (VaR) and tail value-at-risk ($TVaR$), are used. VaR_k is calculated as the $100(1-k)$ percentile of the loss distribution, where $k \in (0,1)$ is the risk tolerance.

$TVaR$ is the expected loss, given that the loss is greater than the VaR level. Namely, we have

$$TVaR_k(S) = E[S|S > VaR_k(S)],$$

where S is the total unpaid loss for the portfolio.

In our case, we use the $TVaR$, which is a coherent risk measure, unlike the VaR for which the sub-additive property is, in general, not guaranteed. The capital allocation approach determines the share of the risk capital to be allocated to each LOB. It was first introduced by Tasche (1999) and is summarized by Bargès et al. (2009).

5.1. Simulation Procedure

To calculate the risk capital, we need a predictive distribution of reserves, which can be obtained by simulation, as these distributions cannot be obtained explicitly.

The simulation algorithm is the same for both one-stage inference and rank-based methods. In fact, to generate realizations from the Sarmanov distribution, we use the inversion method, based on the conditional cumulative distribution function, as described by Pelican and Vernic (2013). The simulation method has the following steps for both the bivariate and trivariate cases:

- Generate $y_{ij}^{(1)}$ from the marginal distribution of the first LOB: $Y_{ij}^{(1)} \sim \mathcal{G}(\alpha_1, \tau_1)$ or $Y_{ij}^{(1)} \sim \mathcal{LN}(a_1, b_1)$.
- Generate $y_{ij}^{(2)}$ from the conditional cumulative distribution function $F_{Y_{ij}^{(2)}|Y_{ij}^{(1)}}$ of a random variable $(Y_{ij}^{(2)}|Y_{ij}^{(1)} = y_{ij}^{(1)})$, as below:

$$F_{Y_{ij}^{(2)}|Y_{ij}^{(1)}}(y) = F(y) + \omega_{1,2}\psi^{(1)}(y_{ij}^{(1)}) \int_{-\infty}^{y} f^{(2)}(y_{ij}^{(2)})\, \psi^{(2)}(y_{ij}^{(2)})\, dy_{ij}^{(2)}.$$

For the trivariate Sarmanov, the simulation procedure continues as follows:

- Generate $y_{ij}^{(3)}$ from the conditional cumulative distribution function $F_{Y_{ij}^{(3)}|Y_{ij}^{(1)},Y_{ij}^{(2)}}$ of a random variable $(Y_{ij}^{(3)}|Y_{ij}^{(1)} = y_{ij}^{(1)}, Y_{ij}^{(2)} = y_{ij}^{(2)})$, expressed as below:

$$\begin{aligned} &F_{Y_{ij}^{(3)}|Y_{ij}^{(1)},Y_{ij}^{(2)}}(y) \\ &= F(y) + \frac{\omega_{1,3}\psi^{(1)}(y_{ij}^{(1)}) \int_{-\infty}^{y} f(y_{ij}^{(3)})\psi^{(3)}(y_{ij}^{(3)})\, dy_{ij}^{(3)}}{1 + \omega_{1,2}\psi^{(1)}(y_{ij}^{(1)})\psi^{(2)}(y_{ij}^{(2)})} + \frac{\omega_{2,3}\psi^{(2)}(y_{ij}^{(2)}) \int_{-\infty}^{y} f(y_{ij}^{(3)})\psi^{(3)}(y_{ij}^{(3)})\, dy_{ij}^{(3)}}{1 + \omega_{1,2}\psi^{(1)}(y_{ij}^{(1)})\psi^{(2)}(y_{ij}^{(2)})}. \end{aligned}$$

Once we estimate the parameters from both the one-stage inference and rank-based methods, as described in Sections 2.3 and 3, we simulate the 45 observations of the lower part of the triangle $y_{i,j}^{(\ell)}$, with $2 \le i \le 10$, and $i \le j \le 10$, using the simulation procedure described above. Then we calculate the reserve and estimate the risk measure from the simulated lower part of the triangle, as follows.

For each simulation and LOB ℓ, we compute the total unpaid loss:

$$X^{(\ell)} = \sum_{i=1}^{n} \sum_{j=1}^{n+1-i} p_i^{(\ell)} y_{ij}^{(\ell)}$$

as well as $S = \sum_{\ell} X^{(\ell)}$, the total unpaid loss for the whole portfolio. Here, the $TVaR$-based capital allocation is used and can be written as

$$\widehat{TVaR}_k(X^{(\ell)};S)$$
$$= \frac{1}{N(1-k)}\left[\sum_{j=1}^{N} X_j^{(\ell)}\mathbf{1}(S_j > \widehat{VaR}_k(S)) + \frac{F_N(\widehat{VaR}_k(S)) - k}{\frac{1}{n}\sum_{i=1}^{N}\mathbf{1}(S_i = \widehat{VaR}_k(S))}\sum_{j=1}^{N} X_j^{(\ell)}\mathbf{1}(S_j = \widehat{VaR}_k(S))\right],$$

where F_N is the empirical cumulative distribution function of S and N is the number of simulations. The total $TVaR$-based capital allocation can be written as

$$\widehat{TVaR}_k(S) = \frac{1}{1-k}\left[\frac{1}{N}\sum_{j=1}^{N} S_j\mathbf{1}(S_j > \widehat{VaR}_k(S)) + \widehat{VaR}_k(S)\Big(F_N(\widehat{VaR}_k(S)) - k\Big)\right].$$

The risk capital is defined as the difference between the risk measure and the value of liability (see, e.g., Dhaene et al. 2006). To replicate what is usually being done in practice, the risk measure is used at a high-risk tolerance, say 99%, while the value of liability (reserve) is usually assumed to be equal to the risk measure, but at a lower risk tolerance, generally between 60% and 80%, according to the risk appetite. Here, we set the risk tolerance at 60% for the reserve in our risk capital analysis. Mathematically, the risk capital associated with a risk R, noted by $RC(R)$, is then calculated as follows:

$$RC(R) = TVaR_{99\%}(R) - TVaR_{60\%}(R).$$

We then compute the gain of the dependence model compared to the silo method below:

$$Gain = \Big(RC_{Silo}(R) - RC_{Sarmanov}(R)\Big)/RC_{Silo}(R).$$

First, we apply the aforementioned procedures to the personal and commercial auto LOBs utilizing data from the US Schedule P. We compute the $TVaR_k$ for various risk thresholds, where $k \in \{60\%, 90\%, 95\%, 99\%\}$. Subsequently, we determine the risk capital and gains using the rank-based method and proceed to compare them against both the silo and one-stage inference methods. The results of these comparisons, based on 50,000 simulations, are presented in Table 30. We present the lowest $TVaR$, risk capital and highest gain for each risk level in bold.

Table 30. The $TVaR$ and risk capital comparison based on 50,000 simulations for personal and commercial LOBs.

		TVaR		
Model	60%	90%	95%	99%
Silo	7,176,965	7,370,308	7,450,181	7,613,205
Sarmanov with one-stage inference	7,170,180	7,335,921	7,403,829	7,541,347
Sarmanov with rank-based method	**7,137,733**	**7,297,020**	**7,362,307**	**7,494,715**
		Risk Capital		
Model	60%	90%	95%	99%
Silo	-	193,343	273,216	436,240
Sarmanov with one-stage inference	-	165,741	233,649	371,167
Sarmanov with rank-based method	-	**159,286**	**224,574**	**356,982**
		Gain		
Model	60%	90%	95%	99%
Sarmanov with one-stage inference	-	14.28%	14.48%	14.92%
Sarmanov with rank-based method	-	**17.61%**	**17.80%**	**18.17%**

Unsurprisingly, both one-stage inference and rank-based Sarmanov methods provide lower risk measures and risk capital than the silo method. This confirms and highlights the importance of the diversification benefit when modeling dependence between these two negatively dependent LOBs.

Significantly, it is evident from Table 30 that the rank-based method surpasses the one-stage inference method in terms of gain when compared to the Silo method. Specifically, we note a reduced risk measure and an increased gain for the Sarmanov rank-based method. This highlights that the diversification benefit achieved through the rank-based method is greater than that attained with the one-stage inference method.

Subsequently, we reevaluate the implications of risk capital using two other datasets from the Canadian insurer data. The initial dataset includes the auto and home LOBs, and we proceed to compare the risk capital obtained through the rank-based Sarmanov method against that obtained via the traditional one-stage inference approach.

Table 31 demonstrates and corroborates that the bivariate Sarmanov model, when utilizing the rank-based method, yields lower risk measures and greater risk capital gains in comparison to both the silo and one-stage inference methods. This observation is further substantiated in the subsequent section through the application of bootstrapping. The lowest $TVaR$, risk capital and highest gain for every risk level are indicated in bold below.

Table 31. The $TVaR$ and risk capital comparison based on 50,000 simulations for the auto and home LOBs.

		TVaR		
Model	60%	90%	95%	99%
Silo	187,326	195,737	199,111	205,934
Sarmanov with one-stage inference	186,983	193,446	195,984	201,120
Sarmanov with rank-based method	**185,138**	**191,560**	**194,083**	**199,211**
		Risk Capital		
Model	60%	90%	95%	99%
Silo	-	8411	11,785	18,608
Sarmanov with one-stage inference	-	6463	9000	14,136
Sarmanov with rank-based method	-	**6422**	**8945**	**14,073**
		Gain		
Model	60%	90%	95%	99%
Sarmanov with one-stage inference	-	23.16%	23.63%	24.03%
Sarmanov with rank-based method	-	**23.65%**	**24.10%**	**24.37%**

For the second dataset from the Canadian Insurer, we compare the risk capital for the bivariate case with the following pairs: BI and AB, BI and DI, and AB and DI, as well as for the trivariate case with the triplet BI, AB, and DI. Here, only models with significant dependence shown in Section 4 are illustrated.

Table 32 demonstrates and validates that the bivariate Sarmanov model, employing the rank-based method, yields lower risk capital and higher gains when contrasted with both the silo and one-stage inference methods. The lowest risk capital and highest gain of the total of three LOBs are highlighted in bold.

In the trivariate scenario, we observe that the risk capital allocations are lower than in the bivariate case. Furthermore, the gains are higher, underscoring the additional risk diversification potential enabled by the rank-based trivariate Sarmanov method in the presence of multivariate dependence.

Table 32. The risk capital comparison based on 50,000 simulations for the BI, AB, and DI LOBs.

Model	Line BI	Line AB	Line DI	Total	Gain
Silo BI and AB	16,163	11,301	-	27,464	-
Bivariate BI and AB one-stage inference	13,474	5972	-	19,446	29.19%
Bivariate BI and AB rank-based method	13,549	5820	-	19,369	29.47%
Silo BI and DI	16,163	-	2455	18,618	-
Bivariate BI and DI rank-based method	16,007	-	400	16,407	11.88%
Silo AB and DI	-	11,301	2455	13,756	-
Bivariate AB and DI rank-based method	-	11,075	619	11,694	14.99%
Silo BI, AB, and DI	16,163	11,301	2455	29,920	-
Trivariate BI, AB, and DI rank-based method	13,458	5800	246	**19,505**	**34.81%**

5.2. Bootstrap Procedure

The results from the simulation procedure section above do not incorporate parameter uncertainty, as the model is assumed to be correct. As such, a parametric bootstrap can be used in order to quantify estimation error and tackle potential model over-fitting. Therefore, in order to calculate the predictive distribution of reserves and risk capital, we also use the bootstrapping method to generate sample data and estimate the parameters. We use the same bootstrap algorithm as Taylor and McGuire (2007), which is also shown in work by Shi and Frees (2011) and Abdallah et al. (2016b). The following are the steps included in the bootstrapping method for bivariate or multivariate cases after estimating parameters using the methods described in Sections 2.3 and 3.

- Simulate 55 pseudo-responses $y^{*(\ell)}_{i,j}$, $(1 \leq i \leq 10,\ 1 \leq j \leq 11 - i)$ from the Sarmanov model using the estimated parameters $\vec{\omega}, \alpha_1, \tau_1, \ldots, \alpha_\ell, \tau_\ell, a_1, b_1, \ldots, a_\ell, b_\ell$, with $\ell \geq 2$.
- Estimate the parameters $\vec{\omega}^*, \alpha^*_1, \tau^*_1, \ldots, \alpha^*_\ell, \tau^*_\ell, a^*_1, b^*_1, \ldots, a^*_\ell, b^*_\ell$ from the new simulated (synthetic) data $y^{*(\ell)}_{i,j}$, based on the different models.
- Simulate the lower part (45 observations) of the triangle $y^{*(\ell)}_{i,j}$, where $2 \leq j \leq 10$ and $12 - j \leq i \leq 10$, using the new estimated parameters $\vec{\omega}^*, \alpha^*_1, \tau^*_1, \ldots, \alpha^*_\ell, \tau^*_\ell, a^*_1, b^*_1, \ldots, a^*_\ell, b^*_\ell$ obtained above.
- Calculate the reserve and estimate the risk measures from the simulated lower part of the triangle.

We apply the bootstrap method to the three datasets. We first use the Kolmogorov–Smirnov test to check whether the simulation procedure produces adequate datasets (i.e., loss triangles), as shown in Table 33. We observe that the null hypothesis is not rejected for all models, i.e., there is not enough evidence that the simulated data do not come from the same distribution of the original loss data for each LOB.

Table 33. KS test for simulated vs. original data.

Model/p-Value	1st Line	2nd Line	3rd Line
Bivariate personal and commercial one-stage inference	0.9989	0.9031	-
Bivariate personal and commercial rank-based	0.9989	0.9031	-
Bivariate auto and home one-stage inference	0.7695	0.9789	-
Bivariate auto and home rank-based	0.7695	0.9789	-
Bivariate BI and AB one-stage inference	0.9031	0.9789	-
Bivariate BI and AB rank-based	0.9031	0.9789	-
Bivariate BI and DI rank-based	0.9031	0.9789	-
Bivariate AB and DI rank-based	0.9789	0.9789	-
Trivariate BI, AB, and DI rank-based	0.9031	0.9789	0.9789

For the personal and commercial auto lines based on the US Schedule P Data, Table 34 displays the $TVaR_k$, $k \in \{60\%, 90\%, 95\%, 99\%\}$, as well as the corresponding risk capital estimates and gains obtained through 5000 bootstrap simulations. As the bootstrap is more computationally intensive, a reduced number of simulations is used for this section. The lowest $TVaR$, risk capital and highest gain for each risk level are given in bold below.

Table 34. Comparison of $TVaR$ and risk capital based on 5000 bootstrap samples for the personal and commercial LOBs.

TVaR				
Model	60%	90%	95%	99%
Silo	7,437,291	7,863,068	8,039,379	8,399,543
Sarmanov with one-stage inference	7,392,765	7,750,422	7,898,561	8,208,176
Sarmanov with rank-based method	**7,357,154**	**7,702,067**	**7,846,936**	**8,138,880**
Risk Capital				
Model	60%	90%	95%	99%
Silo	-	425,777	602,088	962,252
Sarmanov with one-stage inference	-	357,657	505,796	815,411
Sarmanov with rank-based method	-	**344,914**	**489,782**	**781,727**
Gain				
Model	60%	90%	95%	99%
Sarmanov with one-stage inference	-	16.00%	15.99%	15.26%
Sarmanov with rank-based method	-	**19.00%**	**18.65%**	**18.76%**

The findings presented in Table 34 corroborate the results obtained through simulations. Specifically, they demonstrate that, once again, the bivariate Sarmanov model employing the rank-based method yields lower risk measures compared to both the silo and one-stage inference methods. This reaffirms the conclusion that rank-based methods consistently outperform both models when applied to the personal and commercial auto LOBs from the US Schedule P dataset.

We next implement the bootstrap method on the Canadian Insurer Data 1, with the results presented in Table 35. The lowest $TVaR$, risk capital and highest gain for each risk level are written in bold. These results reaffirm the conclusions drawn in Section 5.1, specifically that the bivariate Sarmanov distribution using the rank-based method consistently delivers the lowest risk capital allocations and the highest risk capital gains when compared to the one-stage inference model.

Table 35. Comparison of $TVaR$ and risk capital based on 5000 bootstrap samples for the auto and home LOBs.

TVaR				
Model	60%	90%	95%	99%
Silo	230,444	249,837	257,779	273,682
Sarmanov with one-stage inference	199,250	217,121	224,593	239,702
Sarmanov with rank-based method	**197,857**	**215,206**	**222,418**	**235,385**
Risk Capital				
Model	60%	90%	95%	99%
Silo	-	19,393	27,335	43,238
Sarmanov with one-stage inference	-	17,870	25,343	40,542
Sarmanov with rank-based method	-	**17,349**	**24,561**	**37,528**

Table 35. *Cont.*

Gain				
Model	60%	90%	95%	99%
Sarmanov with one-stage inference	-	7.85%	7.29%	6.24%
Sarmanov with rank-based method	-	**10.54%**	**10.15%**	**13.21%**

Finally, we apply the bootstrap method to the Canadian Insurer Data 2, and the results are shown in Table 36. We highlighted the lowest risk capital and highest gain for the total three LOBs in bold. The findings from Section 5.1 are again confirmed, i.e., the trivariate Sarmanov distribution with the rank-based method provides the smallest risk capital allocations and the largest risk capital gain among all models.

It is worth noting that the risk measures obtained through bootstrapping are significantly higher for all models compared to those reported through simulation. This emphasizes the significance of accounting for parameter uncertainty.

Table 36. Risk capital comparison based on 5000 bootstrap samples for the BI, AB, and DI LOBs.

Model	Line BI	Line AB	Line DI	Total	Gain
Silo BI and AB	35,471	26,899	-	62,370	-
Bivariate BI and AB one-stage inference	28,233	18,320	-	46,553	25.36%
Bivariate BI and AB rank-based method	24,717	17,978	-	42,695	31.55%
Silo BI and DI	35,471	-	5563	41,034	-
Bivariate BI and DI rank-based method	35,021	-	713	35,734	16.30%
Silo AB and DI	-	26,899	5563	32,462	-
Bivariate AB and DI rank-based method	-	26,515	1323	27,838	14.24%
Silo BI, AB, and DI	35,471	26,899	5563	67,934	-
Trivariate BI, AB, and DI rank-based method	24,548	17,970	1591	**44,110**	**35.07%**

6. Summary and Concluding Remarks

In this paper, we introduced rank-based techniques to enhance the modeling of the Sarmanov family of multivariate distributions within the context of loss-reserving. Our findings demonstrate that these rank-based methods not only more effectively capture the inter-dependencies between different LOBs when compared to one-stage inference but also yield superior outcomes in terms of risk capital allocation.

The dependence structure has also been extended to more than two LOBs with the trivariate case, which provides the largest risk capital gains and diversification benefits among all models. We provided comprehensive explanations and descriptions for estimations, reserve calculations, as well as simulation and bootstrap procedures for all the models utilized in this paper.

The methods were calibrated and validated on seven LOBs from real-world data and led to the same conclusions that, namely, the robust rank-based estimation method outperforms the classical one-stage inference approach for both bivariate and trivariate Sarmanov models. Indeed, the rank-based Sarmanov model effectively captures the interdependence among LOBs in cases where the one-stage inference model falls short (see the summary in Table 29). Moreover, as demonstrated in the preceding section, the proposed rank-based Sarmanov model not only yields lower risk measures but also produces a more substantial diversification benefit when compared to the one-stage inference model.

The challenge in aggregate loss reserving lies in dealing with over-parameterization due to the limited dataset available within the loss triangle. Although rank-based methods partially alleviate this problem by fixing the marginal parameters, future research could explore the application of rank-based Sarmanov methods at the micro-level of reserving, where more (detailed) data are accessible.

Furthermore, to enhance the accuracy of residuals, we can also work on improving the fit of the marginal model. In this regard, future investigations may consider utilizing the generalized partial linear model (GPLM), which incorporates both linear and nonlinear components. This approach provides greater flexibility in capturing intricate relationships between the response variable and predictor variables. Such flexibility proves particularly valuable when dealing with non-linear relationships, a common occurrence in real-world datasets (see, for example, He et al. 2005; Yousof and Gad 2015).

The Sarmanov distribution family offers numerous advantages over alternative dependence models, such as copulas. Its flexible structure renders it a promising tool for effectively capturing dependencies among LOBs. This methodology can be readily extended to encompass more than three LOBs, as well as broader risk considerations. Furthermore, its applicability extends beyond LOBs and can be effectively employed in other domains of actuarial science, including the valuation of premiums and the development of pricing strategies.

For industry professionals, this research also carries tangible and pragmatic significance. The rank-based multivariate Sarmanov method offers a more comprehensive understanding of dependence structures and portfolio dynamics. Consequently, it can be a valuable resource for P&C insurance companies, aiding them in meeting the International Financial Reporting Standard (IFRS 17) regulations while enhancing their solvency risk assessment. This, in turn, will result in positive economic and societal impacts by improving the insurance company's solvency ratio. Furthermore, the proposed model aligns harmoniously with industry best practices, as it encourages actuaries to avoid adjusting the estimated reserve of one LOB based on another. Instead, it places a strong emphasis on integrating the impact of correlated LOBs into risk management and tail dependence evaluations. This approach aims to harness diversification benefits and provide valuable insights to inform strategic decisions.

Author Contributions: Conceptualization, A.A.; Methodology, L.W.; Software, L.W.; Validation, A.A.; Formal analysis, L.W.; Investigation, L.W.; Resources, A.A.; Writing—original draft, L.W.; Writing—review and editing, A.A.; Supervision, A.A.; Funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC were funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 20016011].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found in Appendices A–C below.

Acknowledgments: This research was also enabled in part by the support provided by SHARCNET (www.sharcnet.ca) (accessed on 19 October 2022) and the Digital Research Alliance of Canada (alliancecan.ca) (accessed on 19 October 2022).

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

Appendix A. US Schedule P Data

Tables A1 and A2 present the net earned premiums and the incremental paid losses for accident years 1988–1997, inclusive, for personal and commercial auto lines developed over ten years, from the US Schedule P Data. Table A3 presents the AIC and KS goodness-of-fit test used for determining the distribution of marginals. Table A4 presents the parameters of the GLMs of personal and commercial auto lines for the independence case and one-stage inference bivariate model. The corresponding reserve for each model and LOB are also provided.

Table A1. Incremental paid losses for the personal auto line.

Year	Premium	1	2	3	4	5	6	7	8	9	10
1988	4,711,333	1,376,384	1,211,168	535,883	313,790	168,142	79,972	39,235	15,030	10,865	4086
1989	5,335,525	1,576,278	1,437,150	652,445	342,694	188,799	76,956	35,042	17,089	12,507	
1990	5,947,504	1,763,277	1,540,231	678,959	364,199	177,108	78,169	47,391	25,288		
1991	6,354,197	1,779,698	1,498,531	661,401	321,434	162,578	84,581	53,449			
1992	6,738,172	1,843,224	1,573,604	613,095	299,473	176,842	106,296				
1993	7,079,444	1,962,385	1,520,298	581,932	347,434	238,375					
1994	7,254,832	2,033,371	1,430,541	633,500	432,257						
1995	7,739,379	2,072,061	1,458,541	727,098							
1996	8,154,065	2,210,754	1,517,501								
1997	8,435,918	2,206,886									

Table A2. Incremental paid losses for commercial auto line.

Year	Premium	1	2	3	4	5	6	7	8	9	10
1988	267,666	33,810	45,318	46,549	35,206	23,360	12,502	6602	3373	2373	778
1989	274,526	37,663	51,771	40,998	29,496	12,669	11,204	5785	4220	1910	
1990	268,161	40,630	56,318	56,182	32,473	15,828	8409	7120	1125		
1991	276,821	40,475	49,697	39,313	24,044	13,156	12,595	2908			
1992	270,214	37,127	50,983	34,154	25,455	19,421	5728				
1993	280,568	41,125	53,302	40,289	39,912	6650					
1994	344,915	57,515	67,881	86,734	18,109						
1995	371,139	61,553	132,208	20,923							
1996	323,753	112,103	33,250								
1997	221,448	37,554									

Table A3. Fit statistics and goodness-of-fit test of marginals for personal and commercial auto.

LOB	AIC		p-Value of the Kolmogorov–Smirnov Test
	Log-Normal	Gamma	
Personal	−395	−384	0.8732 (Log-normal)
Commercial	−214	−218	0.0159 (Gamma)

Table A4. Parameter and reserve estimations for the independence and one-stage inference bivariate model for personal and commercial auto.

Model		Independence		One-Stage Inference Bivariate Model	
LOB ℓ		Personal	Commercial	Personal	Commercial
GLM		Log-normal	Gamma(log)	Log-normal	Gamma(log)
$u^{(\ell)}$		−1.137	−1.670	−1.113	−1.585
Accident Year	2	−0.033	−0.129	−0.039	−0.196
	3	−0.028	−0.142	−0.033	−0.258
	4	−0.131	−0.289	−0.132	−0.403
	5	−0.175	−0.272	−0.178	−0.384
	6	−0.174	−0.252	−0.170	−0.360
	7	−0.173	−0.124	−0.179	−0.207
	8	−0.223	−0.089	−0.256	−0.137
	9	−0.244	0.135	−0.272	0.158
	10	−0.204	−0.104	−0.186	−0.248

Table A4. *Cont.*

Model		**Independence**		**One-Stage Inference Bivariate Model**	
Dev. Lag	2	−0.224	0.196	−0.248	0.238
	3	−1.047	−0.023	−1.070	−0.013
	4	−1.644	−0.409	−1.668	−0.395
	5	−2.254	−1.048	−2.277	−1.041
	6	−3.013	−1.463	−3.041	−1.474
	7	−3.671	−2.089	−3.689	−2.080
	8	−4.493	−2.783	−4.488	−2.811
	9	−4.911	−3.111	−4.965	−2.995
	10	−5.913	−4.171	−5.888	−4.369
sd or scale		0.089	10.093	0.089	9.712
Dependence parameters				−4.430	
Reserve		6,464,075	490,652	6,465,679	513,622

Appendix B. Canadian Insurer Data 1

Tables A5 and A6 display the incremental paid losses and net earned premiums for accident years 2003–2012, inclusive, for both auto and home lines of business (LOBs) developed over a span of ten years, sourced from the Canadian insurer data. Table A7 presents the AIC and KS goodness-of-fit test, showing that both LOBs follow a gamma distribution. Table A8 presents the parameters of the GLMs of auto and home LOBs for the independence case and one-stage inference bivariate model, as well as the corresponding reserve for each model and LOB.

Table A5. Cumulative paid losses for LOB Auto.

Accident	**Development Lag (in Months)**										
Year	**12**	**24**	**36**	**48**	**60**	**72**	**84**	**96**	**108**	**120**	**Premiums**
2003	2279	8683	15,136	21,603	27,650	30,428	32,004	32,592	33,009	34,140	76,620
2004	2139	7077	13,159	16,435	20,416	22,598	24,171	25,034	25,714		65,691
2005	1420	4888	8762	12,184	14,482	15,633	17,089	17,710			55,453
2006	1510	5027	10,763	15,799	19,269	22,504	24,807				54,006
2007	1693	5175	8216	12,263	16,918	20,792					55,425
2008	2097	7509	10,810	15,673	19,791						59,100
2009	2094	5174	8062	12,389							54,438
2010	1487	4789	7448								53,483
2011	1868	6196									52,978
2012	2080										57,879

Table A6. Cumulative paid losses for LOB Home.

Accident	**Development Lag (in Months)**										
Year	**12**	**24**	**36**	**48**	**60**	**72**	**84**	**96**	**108**	**120**	**Premiums**
2003	4157	9558	13,131	17,460	19,608	21,124	21,900	23,360	23,377	23,575	55,484
2004	4158	9956	14,860	18,024	20,397	22,068	23,312	24,555	25,137		65,705
2005	3989	10,519	15,877	20,274	23,428	26,495	30,974	31,580			73,879
2006	4012	10,904	16,141	19,643	21,954	26,215	28,095				91,473
2007	4322	10,814	16,086	20,186	24,157	27,222					87,212
2008	6379	14,524	19,058	24,108	28,329						89,455
2009	5291	14,620	20,799	25,131							90,341
2010	4946	12,956	18,007								89,212
2011	5674	15,026									91,606
2012	5478										99,982

Table A7. Fit statistics and goodness-of-fit test of marginals for auto and home.

LOB	AIC		p-Value of the
	Log-Normal	Gamma	Kolmogorov–Smirnov Test
Auto	−323	−324	0.397 (Gamma)
Home	−259	−267	0.019 (Gamma)

Table A8. Parameter and reserve estimations for the independence and one-stage inference bivariate model for auto and home.

Model		Independence		One-Stage Inference Bivariate Model	
LOB ℓ		Auto	Home	Auto	Home
GLM		Gamma	Gamma	Gamma	Gamma
$u^{(\ell)}$		−3.501	−2.872	−3.495	−2.889
Accident Year	2	0.053	0.101	0.059	0.112
	3	−0.156	0.163	−0.153	0.162
	4	0.238	−0.136	0.254	−0.112
	5	0,137	−0.024	0.146	0.012
	6	0.120	0.095	0.127	0.126
	7	0.003	0.069	0.003	0.138
	8	−0.160	−0.017	−0.153	−0.001
	9	0.169	0.131	0.138	0.171
	10	0.175	−0.032	0.168	−0.011
Dev. Lag	2	0.815	0.420	0.808	0.396
	3	0.817	0.076	0.813	0.066
	4	0.849	−0.095	0.833	−0.094
	5	0.717	−0.406	0.713	−0.373
	6	0.283	−0.481	0.254	−0.473
	7	−0.115	−0.757	−0.131	−0.720
	8	−1.001	−1.215	−1.004	−1.195
	9	−1.375	−2.612	−1.385	−2.601
	10	−0.715	−2.764	−0.711	−2.736
sd or scale		24.046	8.021	25.087	8.104
Dependence parameters				256.001	
Reserve		78,665	98,929	77,789	101,603

Appendix C. Canadian Insurer Data 2

Tables A9–A11 display the net earned premiums and cumulative paid losses for accident years 2003–2012, inclusive, for each LOB (BI, AB, DI) developed over, a maximum of ten years, using data from a large Canadian insurer. To preserve confidentiality, all figures were multiplied by a constant. Table A12 displays the AIC and KS goodness-fit of test results, used to determine the distribution of each marginal. Table A13 displays the parameters of the GLMs of three LOBs for the independence model, one-stage inference bivariate model, and one-stage inference trivariate model, accompanied by the respective reserve for each model and LOB.

Table A9. Cumulative paid losses for the BI LOB.

Accident	Development Lag (in Months)										
Year	12	24	36	48	60	72	84	96	108	120	Premiums
2003	3488	14,559	27,249	37,979	49,561	55,957	58,406	60,862	63,280	63,864	85,421
2004	1169	12,781	20,550	31,547	42,808	47,385	50,251	50,978	51,272		98,579
2005	1478	10,788	25,499	34,279	43,057	49,360	52,329	52,544			103,062
2006	1186	11,852	22,913	32,537	41,824	48,005	52,542				108,412
2007	1737	13,881	25,521	38,037	43,684	47,755					111,176
2008	1571	12,153	27,329	41,832	51,779						112,050
2009	1199	17,077	29,876	44,149							112,577
2010	1263	16,073	28,249								113,707
2011	986	10,003									126,442
2012	683										130,484

Table A10. Cumulative paid losses for LOB AB.

Accident	Development Lag (in Months)										
Year	12	24	36	48	60	72	84	96	108	120	Premiums
2003	13,714	24,996	31,253	38,352	44,185	46,258	47,019	47,894	48,334	48,902	116,491
2004	6883	16,525	24,796	29,263	32,619	33,383	34,815	35,569	35,612		111,467
2005	7933	22,067	32,801	38,028	44,274	44,948	46,507	46,665			107,241
2006	7052	18,166	25,589	31,976	36,092	38,720	39,914				105,687
2007	10,463	23,982	31,621	36,039	38,070	41,260					105,923
2008	9697	28,878	41,678	47,135	50,788						111,487
2009	11,387	37,333	48,452	55,757							113,268
2010	12,150	32,250	40,677								121,606
2011	5348	14,357									110,610
2012	4612										104,304

Table A11. Cumulative paid losses for LOB DI.

Accident	Development Lag (in Months)										
Year	12	24	36	48	60	72	84	96	108	120	Premiums
2003	3043	5656	7505	8593	9403	10,380	10,450	10,812	10,856	10,860	116,491
2004	2070	4662	6690	8253	9286	9724	9942	10,086	10,121		111,467
2005	2001	4825	7344	8918	9824	10,274	10,934	11,155			107,241
2006	1833	4953	7737	9524	10,986	11,267	11,579				105,687
2007	2217	5570	7898	8885	9424	10,402					105,923
2008	2076	5681	8577	10,237	12,934						111,487
2009	2025	6225	9027	10,945							113,268
2010	2024	5888	8196								121,606
2011	1311	3780									110,610
2012	912										104,304

Table A12. Fit statistics and goodness-of-fit test of marginals for BI, AB, and DI.

LOB	AIC		p-Value of the
	Log-Normal	Gamma	Kolmogorov–Smirnov Test
Bodily Injury	−262	−270	0.643 (Gamma)
Accident Benefit	−267	−276	0.135 (Gamma)
Disability Income	−437	−444	0.478 (Gamma)

Table A13. Parameter and reserve estimations for the independence and one-stage inference models.

Model		**Independence**			**Bivariate BI and AB**		**Bivariate BI and DI**		**Bivariate AB and DI**		**Trivariate Model**		
LOB ℓ		BI	AB	DI	BI	AB	BI	DI	AB	DI	BI	AB	DI
GLM		Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma
$u^{(\ell)}$		−3.628	−2.365	−4.064	−3.593	−2.317	−3.605	−4.073	−2.410	−4.035	−3.597	−2.262	−4.086
Accident Year	2	−0.750	−0.413	−0.121	−0.768	−0.409	−0.758	−0.163	−0.380	−0.118	−0.839	−0.483	−0.090
	3	−0.729	−0.196	0.171	−0.771	−0.242	−0.724	0.128	−0.125	0.157	−0.814	−0.288	0.183
	4	−0.651	−0.112	0.129	−0.627	−0.098	−0.659	0.099	−0.046	0.143	−0.683	0.183	0.145
	5	−0.740	−0.095	0.092	−0.744	−0.123	−0.754	0.051	−0.039	0.084	−0.805	−0.186	0.107
	6	−0.574	−0.001	0.396	−0.571	−0.010	−0.575	0.358	0.056	0.377	−0.691	−0.133	0.398
	7	−0.574	0.196	0.254	−0.603	0.122	−0.557	0.223	0.225	0.215	−0.664	0.085	0.265
	8	−0.658	−0.012	0.055	−0.697	−0.091	−0.684	0.052	0.022	0.060	−0.723	−0.147	0.076
	9	−1.147	−0.628	−0.259	−1.168	−0.713	−1.186	−0.295	−0.635	−0.285	−1.168	−0.767	−0.210
	10	−1.625	−0.754	−0.676	−1.675	−0.756	−1.621	−0.649	−0.751	−0.696	−1.694	−0.791	−0.628
Dev. Lag	2	2.061	0.450	0.419	2.047	0.436	2.055	0.480	0.463	0.381	2.119	0.443	0.440
	3	2.065	−0.055	0.114	2.064	−0.066	2.051	0.165	−0.035	0.107	2.107	−0.070	0.120
	4	2.018	−0.507	−0.358	1.994	−0.504	1.983	−0.318	−0.505	−0.366	2.073	−0.501	−0.312
	5	1.818	−0.759	−0.582	1.778	−0.796	1.785	−0.543	−0.758	−0.607	1.884	−0.773	−0.545
	6	1.297	−1.580	−1.154	1.243	−1.631	1.286	−1.101	−1.582	−1.176	1.374	−1.642	−1.143
	7	0.772	−1.899	−1.870	0.729	−1.884	0.757	−1.806	−1.902	−1.898	0.792	−1.943	−1.863
	8	−0.493	−2.670	−2.102	−0.526	−2.713	−0.510	−2.064	−2.629	−2.131	−0.475	−2.752	−2.150
	9	−0.429	−3.762	−3.849	−0.452	−3.801	−0.453	−3.805	−3.720	−3.862	−0.405	−3.874	−3.849
	10	−1.358	−2.960	−6.255	−1.353	−3.037	−1.418	−6.260	−2.927	−6.313	−1.438	−3.154	−6.190
sd or scale		10.699	8.037	10.078	10.749	8.413	10.758	10.118	7.924	9.973	10.213	8.233	10.143
Dependence parameters					$\omega_{BI,AB}$ 436.904		$\omega_{BI,DI}$ 424.587		$\omega_{AB,DI}$ 730.930		$\omega_{BI,AB}$ 374.794	$\omega_{BI,DI}$ −110.327	$\omega_{AB,DI}$ −165.781
Reserve		132,918	73,220	18,289	129,397	71,457	131,148	18,739	72,144	18,123	135,061	70,857	18,752

References

Abdallah, Anas, Jean-Philippe Boucher, and Hélène Cossette. 2015. Modeling dependence between loss triangles with hierarchical Archimedean copulas. *ASTIN Bulletin: The Journal of the IAA* 45: 577–99. [CrossRef]

Abdallah, Anas, Jean-Philippe Boucher, and Hélène Cossette. 2016a. Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insurance. Mathematics & Economics* 68: 120–33.

Abdallah, Anas, Jean-Philippe Boucher, Hélène Cossette, and Julien Trufin. 2016b. Sarmanov family of bivariate distributions for multivariate loss reserving analysis. *North American Actuarial Journal* 20: 184–200. [CrossRef]

Ajne, Björn. 1994. Additivity of chain-ladder projections. *ASTIN Bulletin: The Journal of the IAA* 24: 311–18. [CrossRef]

Akaike, Hirotugu. 1974. A new look at the statistical model identification problem. *IEEE Transactions on Automatic Control* 19: 716. [CrossRef]

Araiza Iturria, Carlos Andrés, Frédéric Godin, and Mélina Mailhot. 2021. Tweedie double GLM loss triangles with dependence within and across business lines. *European Actuarial Journal* 11: 619–53. [CrossRef]

Avanzi, Benjamin, Greg Taylor, Phuong Anh Vu, and Bernard Wong. 2016. Stochastic loss reserving with dependence: A flexible multivariate Tweedie approach. *Insurance: Mathematics and Economics* 71: 63–78. [CrossRef]

Badounas, Ioannis, and Georgios Pitselis. 2020. Loss reserving estimation with correlated run-off triangles in a quantile longitudinal model. *Risks* 8: 14. [CrossRef]

Bahraoui, Zuhair, Catalina Bolancé, Elena Pelican, and Raluca Vernic. 2015. On the bivariate Sarmanov distribution and copula. An application on insurance data using truncated marginal distributi. *SORT* 39: 209–30.

Bairamov, Ismihan, Banu Altinsoy, and G. Jay Kerns. 2011. On generalized Sarmanov bivariate distributions. *TWMS Journal of Applied and Engineering Mathematics* 1: 86–97.

Bargès, Mathieu, Hélène Cossette, and Etienne Marceau. 2009. TVaR-based capital allocation with copulas. *Insurance: Mathematics and Economics* 45: 348–61. [CrossRef]

Berger, Vance W., and Yanyan Zhou. 2014. Kolmogorov–smirnov test: Overview. In *Wiley statsref: Statistics Reference Online*. New York: John Wiley and Sons, Ltd. [CrossRef]

Bolancé, Catalina, and Raluca Vernic. 2019. Multivariate count data generalized linear models: Three approaches based on the Sarmanov distribution. *Insurance: Mathematics and Economics* 85: 89–103.

Bolancé, Catalina, Montserrat Guillen, and Albert Pitarque. 2020. A Sarmanov distribution with beta marginals: An application to motor insurance pricing. *Mathematics* 8: 2020. [CrossRef]

Braun, Christian. 2004. The prediction error of the chain ladder method applied to correlated run-off triangles. *ASTIN Bulletin: The Journal of the IAA* 34: 399–423. [CrossRef]

Brehm, Paul. 2002. Correlation and the aggregation of unpaid loss distributions. *CAS Forum* 2: 1–23.

Chen, Yiqing, Jiajun Liu, and Yang Yang. 2023. Ruin under Light-Tailed or Moderately Heavy-Tailed Insurance Risks Interplayed with Financial Risks. *Methodology and Computing in Applied Probability* 25: 14.

Cohen, Leon. 1984. Probability distributions with given multivariate marginals. *Journal of Mathematical Physics* 25: 2402–3. [CrossRef]

Côté, Marie-Pier, Christian Genest, and Anas Abdallah. 2016. Rank-based methods for modelling dependence between loss triangles. *European Actuarial Journal* 6: 377–408. [CrossRef]

Danaher, Peter J., and Michael Stanley Smith. 2011. Modelling multivariate distributions using copulas: Applications in marketing. *Marketing Science* 30: 4–21. [CrossRef]

De Jong, Piet. 2012. Modelling dependence between loss triangles. *North American Actuarial Journal* 16: 74–86. [CrossRef]

De Jong, Piet, and Gillian Z. Heller. 2008. *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press.

Dhaene, Jan, Steven Vanduffel, Marc J. Goovaerts, Rob Kaas, Qihe Tang, and David Vyncke. 2006. Risk measures and comonotonicity: A review. *Stochastic Models* 22: 573–606. [CrossRef]

Drouet Mari, Dominique, and Samuel Kotz. 2001. *Correlation and Dependence*. Singapore: World Scientific Publishing.

Genest, Christian, and Anne-Catherine Favre. 2007. Everything you always wanted to know about copula modelling but were afraid to ask. *Journal of Hydrologic Engineering* 12: 347–68. [CrossRef]

Genest, Christian, and Johanna Nešlehová. 2014. Copulas and copula models. In *Encyclopedia of Environmetrics*, 2nd ed. Edited by A. H. El-Shaarawi and W. W. Piegorsch. Chichester: Wiley, vol. 2, pp. 541–53.

Genest, Christian, Johanna Nešlehová, and Noomen Ben Ghorbal. 2011. Estimators based on Kendall's tau in multivariate copula models. *Australian & New Zealand Journal of Statistics* 53: 157–77.

Genest, Christian, Kilani Ghoudi, and Louis-Paul Rivest. 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82: 543–52. [CrossRef]

Guo, Fenglong, Dingcheng Wang, and Hailiang Yang. 2017. Asymptotic results for ruin probability in a two-dimensional risk model with stochastic investment returns. *Journal of Computational and Applied Mathematics* 325: 198–221. [CrossRef]

He, Xuming, Wing Kin Fung, and Zhongyi Zhu. 2005. Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association* 100: 1176–84. [CrossRef]

Hernández-Bastida, Agustín, and Mª del Pilar Fernández-Sánchez. 2012. A Sarmanov family with beta and gamma marginal distributions: An application to the Bayes premium in a collective risk model. *Statistical Methods & Applications* 21: 391–409.

Johnson, Norman L., and Samuel Kott. 1975. On some generalized farlie-gumbel-morgenstern distributions. *Communications in Statistics-Theory and Methods* 4: 415–27.

Kirschner, Gerald S., Colin Kerley, and Belinda Isaacs. 2002. Two approaches to calculating correlated reserve indications across multiple lines of business. In *Casualty Actuarial Society Forum*. Fall Forum on Reserving Call Papers, pp. 211–46. Available online: https://www.casact.org/sites/default/files/database/forum_02fforum_02ff211.pdf (accessed on 8 August 2023).

Lally, Nathan, and Brian Hartman. 2018. Estimating loss reserves using hierarchical Bayesian Gaussian process regression with input warping. *Insurance: Mathematics and Economics* 82: 124–40. [CrossRef]

Lee, Mei-Ling Ting. 1996. Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics-Theory and Methods* 25: 1207–22.

McCullagh, Peter, and John Ashworth Nelder. 1989. *Generalized Linear Models*, 2nd ed. Boca Raton: Chapman and Hall. ISBN 0-4123176-0-5.

Merz, Michael, and Mario Valentin Wüthrich. 2008. Prediction error of the multivariate chain ladder reserving method. *North American Actuarial Journal* 12: 175–97. [CrossRef]

Merz, Michael, Mario Valentin Wüthrich, and Enkelejd Hashorva. 2013. Dependence modelling in multivariate claims run-off triangles. *Annals of Actuarial Science* 7: 3–25. [CrossRef]

Miravete, Eugenio J. 2009. *Multivariate Sarmanov Count Data Models*. Discussion Paper No. DP7463. London: Centre for Economic Policy Research.

Pelican, Elena, and Raluca Vernic. 2013. Maximum-likelihood estimation for the multivariate Sarmanov distribution: Simulation study. *International Journal of Computer Mathematics* 90: 1958–70. [CrossRef]

Ratovomirija, Gildas, Maissa Tamraz, and Raluca Vernic. 2017. On some multivariate Sarmanov mixed Erlang reinsurance risks: Aggregation and capital allocation. *Insurance: Mathematics and Economics 74*: 197–209. [CrossRef]

Sarmanov, Oleg Vasil'evich. 1966. Generalized normal correlation and two-dimensional Fréchet classes. In *Doklady Akademii Nauk*. Moscow: Russian Academy of Sciences, vol. 168, pp. 32–35.

Schmidt, Klaus D. 2006. Optimal and Additive Loss Reserving for Dependent Lines of Business. Paper presented at 2006 CAS Casualty Loss Reserve Seminar, Atlanta, GA, USA, September 11–12, pp. 319–51.

Schweidel, David A., Peter S. Fader, and Eric T. Bradlow. 2008. A bivariate timing model of customer acquisition and retention. *Marketing Science* 27: 829–43. [CrossRef]

Shi, Peng. 2017. A multivariate analysis of intercompany loss triangles. *Journal of Risk and Insurance* 84: 717–37. [CrossRef]

Shi, Peng, and Edward W. Frees. 2011. Dependent loss reserving using copulas. *ASTIN Bulletin: The Journal of the IAA* 41: 449–86.

Shi, Peng, Sanjib Basu, and Glenn G. Meyers. 2012. A Bayesian log-normal model for multivariate loss reserving. *North American Actuarial Journal* 16: 29–51. [CrossRef]

Tank, Fatih, and Omer L. Gebizlioglu. 2004. Sarmanov distribution class for dependent risks and its applications. *Belgian Actuarial Bulletin* 4: 50–52.

Tasche, Dirk. 1999. *Risk Contributions and Performance Measurement*. Report of the Lehrstuhl für mathematische Statistik. Munich: Munich University of Technology.

Taylor, Greg, and Gráinne McGuire. 2007. A synchronous bootstrap to account for dependencies between lines of business in the estimation of loss reserve prediction error. *North American Actuarial Journal* 11: 70–88. [CrossRef]

Vernic, Raluca, Catalina Bolancé, and Ramon Alemany. 2022. Sarmanov distribution for modeling dependence between the frequency and the average severity of insurance claims. *Insurance: Mathematics and Economics* 102: 111–25. [CrossRef]

Woolf, Barnet. 1957. The log likelihood ratio test (the G-test). *Annals of Human Genetics* 21: 397–409. [CrossRef] [PubMed]

Yang, Yang, and Kam Chuen Yuen. 2016. Finite-time and infinite-time ruin probabilities in a two-dimensional delayed renewal risk model with Sarmanov dependent claims. *Journal of Mathematical Analysis and Applications* 442: 600–26. [CrossRef]

Yousof, Haitham M., and Ahmed M. Gad. 2015. Bayesian estimation and inference for the generalized partial linear model. *International Journal of Probability and Statistics* 4: 51–64.