Master in Actuarial and Financial Sciences

MASTER THESIS

Title: Box-Cox transformation on the framework of Sarmanov distribution.

Author: Roberto Rodrigo Marqués

Advisor: Catalina Bolancé Losilla

Academic year: 2019-2020

UNIVERSITAT DE BARCELONA
Facultat d'Economia i Empresa

Màster de Ciències Actuarials i Financeres

Faculty of Economics and Business

Universitat de Barcelona

Master thesis

Master in Actuarial and Financial Sciences

Box-Cox transformation on the framework of Sarmanov distribution

Author: Roberto Rodrigo Marqués

Advisor: Catalina Bolancé Losilla

Box-Cox transformation on the framework of Sarmanov distribution

Roberto Rodrigo Marqués

June 9, 2020

Abstract

It is known that in some cases the classical assumption of independence between claim frequency and claim severity does not hold in the collective model. Nowadays exists an increasing interest in models which capture this dependence. In this work we propose to consider the Sarmanov distribution as a bivariate model which captures this kind of dependence. On the other hand, Box-Cox family of transformations are widely used in data analysis to eliminate skewness and other distributional features that complicate analysis, transforming the original data into a Normal distributed sample. We also consider the average claim severity distributed as a Box-Cox back transformed from a Normal distribution in the framework of Sarmanov bivariate distribution. Assuming that the differences between a Normal distribution and claim severity distribution can be explained in terms of a Box-Cox transformation. More over, we propose a maximum likelihood estimation procedure adapted to this Box-Cox transformed bivariate Sarmanov distribution to estimate the parameters of the model.

keywords: dependence, Sarmanov distribution, Box-Cox transformation, frequency, severity, parameters estimation

1 Introduction

1.1 On Box-Cox transformation and Sarmanov distribution

On most models and procedures on data analysis, it must be considered that the observations of the sample are independently Normal distributed with constant variance to meet the requirements of the fitted model. This is usually a so strong assumption that most of data of interest to analyze does not hold. Since Box and Cox (1964), their so called Box-Cox family

of transformations have being widely used on data analysis to take account of this problem. They proposed that under such transformation, one could map the original random variable into an other which is Normal distributed by properly choosing the parameters of the transformation, for example by maximizing the Log-Likelihood function. Then, we can think that this trasformation is capturing all non-Normal features of the original data, as skewness, asimetry or heavy-tails (Wand et al., 1991; Clements et al., 2003; Zhang and King, 2004). By reverse thinking we can also believe that we can add non-Normal behaviour to Normal distributed data by back Box-Cox transform it. This way of thinking let us to define a very general pdf for continous marginals, which may let us to perform data analysis without making any assumption on the particular distribution which data follows.

Moreover, dependence between random variables is not easy to measure and the way to model this dependence is still a field of investigation. In particular, there are diferent approaches to define dependence between discrete and continous random variables (Jonhson et al., 1997; Kotz and Balakrishnan, 2000), as define a general form of the multivariate pdf or through Copulas (Bahraoui et al., 2015). In this work we choose to model dependence through Sarmanov distribution which has been proven to be a very usefull and versatile tool for this purpose (Vernic and Bolancé, 2019; Vernic et al., 2020), among other things due to its ability to join different marginals that can even be discrete and some continuous at the same time.

Putting all these ingredients together, the purpose of this work is to properly define a model for total cost on insurance claims which take acount of claim frequency and claim severity dependence on the framework of Sarmanov distribution, considering claim severity distributed as a Normal Box-Cox back transformed distribution. Also we develop the necesary techniques to fit this model to data and show an example of this fitting on a sample of insurance claim real data.

1.2 Motivation

As we know, the total cost for an insurance company can be written as:

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

Where N is the number of claims and X the average cost per claim. In the collective model these variables are usually considered independent, and furthermore $(N_i, X_i)_{i=1...m}$ i.i.d. bivariate sample from the bivariate random vector (N, X). This independence hypothesis enables us to compute the expected cost as:

$$E[S] = E[NX] = E[N]E[X]. \tag{2}$$

Which is closely related with the risk premium that an insurance company charge its clients.

It has been shown that the independence hypothesis does not hold in certain situations. In auto insurance high frequency can be associated with an urban driving area where costs are low or, on the other hand, high frequency can also be associated with daily journeys on secondary roads where costs usually are bigger. Then, there can be situations when there is dependence between these variables, when the risk premium is not correctly computed without taking into account this dependence.

Vernic et al. (2020), have being studied Sarmanov distribution as a way to model this dependence between marginals. They have proposed Sarmanov distribution to model dependence between marginals due to its ability to join marginals of different types, more precisely, one marginal can be discrete (claim frequency) and the other continuous (average severity). Following their path, the purpose of this work is to study the effect of considering that the average claim severity follows an a priori unknow distribution obtained by Box-Cox back transforming a Normal distribution. We will expect a better adjustment of the model, since we let more flexibility to the model than fixing a certain distribution for claim severity.

2 Theoretical Basis

2.1 Box-Cox back transformed claim severity PDF

The key idea of this work is the assumption that the claim severity follows an unknown distribution. Under this idea one may think that we do not have much to do, but nothing further from reality, in this context we can make the work much easier, and even get certain advantages that we will see later. The Box-Cox transformation, widely used in data analysis and statistics, is known to be a potential type transformation that is capable of normalizing a series of data, that is, it can roughly transform the pdf of a random variable into the pdf of a Normal distribution. This Box-Cox transformation is of the form:

$$g(X) = \begin{cases} \frac{X^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ log(X), & \lambda = 0 \end{cases}, \tag{3}$$

As we said before, by properly choosing the transformation parameter λ, we can transform one random variable into another which roughly follows a Normal distribution. Then, we can work with a general Box-Cox back transformed Normal distribution without making any previous assumption for the continuous marginal, letting Box-Cox transformation to properly define the non-Normal features of this continous random variable.

Let Y a random variable which is distributed as $N(\mu, \sigma)$, then we can write:

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \tag{4}$$

To obtain the back transformed pdf of Y we can just take account of the normalization of f_Y:

$$1 = \int f_Y(y)\, dy = \int f_Y(g(X)) \frac{dg(X)}{dx} dx \tag{5}$$

Therefore the back transformed pdf of Y can be written as:

$$f_X(x) = f_Y(g(X)) \frac{dg(X)}{dx} \tag{6}$$

Applying expresion (6) to our particular case in which Y is Normal distributed as (4) and $g(X)$ is a Box-Cox transformation as (3), we have:

$$f_X(x) = \begin{cases} \frac{x^{\lambda-1}}{\sigma\sqrt{2\pi}}\ exp\left[-\frac{1}{2}\left(\frac{\frac{x^\lambda-1}{\lambda}-\mu}{\sigma}\right)^2\right],\ \lambda \neq 0 \\ \frac{1}{x\sigma\sqrt{2\pi}}\ exp\left[-\frac{1}{2}\left(\frac{log(x)-\mu}{\sigma}\right)^2\right],\ \lambda = 0 \end{cases}, \tag{7}$$

As it is seen on the previous equation (7), the case in which $\lambda = 0$ reproduces the pdf of a Log-Normal distribution.

2.2 Sarmanov dependence

We assume a Sarmanov dependence between N and X as follows

$$f_{X,N}(x,n) = \begin{cases} p(0),\ n = x = 0 \\ p(n) f(x) (1 + \omega\psi(n)\phi(x)),\ n \geq 1,\ x > 0 \end{cases}, \tag{8}$$

Where f is a pdf, ψ and ϕ are bounded non-constant kernel functions and $\omega \in \mathbb{R}$. We call the pdf $f_{X,N}$ mixed because it joins the continuous pdf f and the discrete pmf p. Is important to note that in this way, the marginal pdf of X is defined as a mixed distribution, with a discrete point at $X = 0$ and a continous part for $X > 0$, then we can write X distribution as:

$$\begin{cases} Pr\left(X=0\right)=p\left(0\right), \\ f_X\left(x\right)=\left(1-p\left(0\right)\right)f\left(x\right),\ x>0. \end{cases}, \tag{9}$$

Note that f_X must be normalized to $(1-p\left(0\right))$ instead of 1.

Also, in order to $f_{X,N}$ will be a properly defined pdf (that is non-negative and with unitary norm) we impose the following conditions. First $f_{X,N}$ must be normalized:

$$\begin{aligned} 1 &= \sum_{n\geq 0}\int_{\mathbb{R}^+} f_{X,N}\left(x\right)dx \\ &= p(0)+\sum_{n\geq 1}\int_{\mathbb{R}^+} p(n)f(x)(1+\omega\psi(n)\phi(x))dx \\ &= 1+\omega\sum_{n\geq 1}\int_{\mathbb{R}^+} p(n)f(x)\psi(n)\phi(x)dx \end{aligned} \tag{10}$$

Hence, we must have:

$$\begin{aligned} 0 &= \sum_{n\geq 1}\int_{\mathbb{R}^+} p(n)f(x)\psi(n)\phi(x)dx \\ &= \sum_{n\geq 1} p(n)\psi(n) = \int_{\mathbb{R}^+} f(x)\phi(x)dx \end{aligned} \tag{11}$$

Also, we have to impose no negativity to $f_{X,N}$, so it must hold:

$$1+\omega\psi\left(n\right)\phi\left(x\right)\geq 0,\ \text{for all}\ \ n\geq 1,\ x>0 \tag{12}$$

We propose to use an exponential kernel $\phi(x)$ of the form:

$$\phi(x)=e^{-\gamma x}-k \tag{13}$$

Then putting this kernel into equation (11) we have:

$$0=E[\phi(x)]=\int_{\mathbb{R}^+} f(x)\phi(x)dx=\int_{\mathbb{R}^+} f(x)(e^{-\gamma x}-k)dx=\mathcal{L}_X(\gamma)-k \tag{14}$$

Therefore we have:

$$\phi(x) = e^{-\gamma x} - \mathcal{L}_X(\gamma) \tag{15}$$

Imposing the same functional form for the kernel $\psi(n) = e^{-\delta n} - k$ on equation (11):

$$\begin{aligned} 0 = E[\psi(n)|n \geq 1] &= \sum_{n\geq 1} p(n)e^{-\delta n} - k\sum_{n\geq 1} p(n) \\ &= p(0) + \sum_{n\geq 1} p(n)e^{-\delta n} - p(0) - k\sum_{n\geq 1} p(n) \\ &= \mathcal{L}_N(\delta) - p(0) - k\sum_{n\geq 1} p(n) \end{aligned} \tag{16}$$

Therefore we have:

$$\psi(n) = e^{-\delta n} - \frac{\mathcal{L}_N(\delta) - p(0)}{\sum_{n\geq 1} p(n)} = e^{-\delta n} - \frac{\mathcal{L}_N(\delta) - p(0)}{1 - p(0)} \tag{17}$$

Equation (12) impose upper and lower limits to ω, $\omega \leq \frac{-1}{\phi(x)\psi(n)}$ when $\phi(x)\psi(n) \geq 0$ and $\omega \geq \frac{-1}{\phi(x)\psi(n)}$ when $\phi(x)\psi(n) \leq 0$. Letting $m_1 = \inf_{n\geq 1} \psi(n)$, $m_2 = \inf_{x>0} \phi(x)$, $M_1 = \sup_{n\geq 1} \psi(n)$ and $M_2 = \sup_{x>0} \phi(x)$, condition (12) restricts ω to the following interval:

$$\max\left\{-\frac{1}{m_1 m_2}, -\frac{1}{M_1 M_2}\right\} \leq \omega \leq \min\left\{-\frac{1}{m_1 M_2}, -\frac{1}{M_1 m_2}\right\} \tag{18}$$

2.3 Total cost pdf and its moments

On this subsection we expose the derivation of total cost pdf based on Sarmanov's joined pdf, and we derive its firsts moments, $E[S]$ and $V[S]$. S and N joined pdf, can be obtained just performing the change of variable $s = nx$, $ds = ndx$ to equation (8) as in (5), then we have:

$$f_{S,N}(s,n) = \begin{cases} Pr(S=0) = p(0), & n = s = 0 \\ \frac{p(n)}{n} f\left(\frac{s}{n}\right)\left(1 + \omega\psi(n)\phi\left(\frac{s}{n}\right)\right), & n \geq 1,\ s > 0 \end{cases}, \tag{19}$$

Then total cost pdf is obtained summing previous equation (19) over all posible values of N.

$$f_S(s) = \begin{cases} Pr(S=0) = p(0), & s = 0 \\ \sum_{n\geq 1} \frac{p(n)}{n} f\left(\frac{s}{n}\right)\left(1 + \omega\psi(n)\phi\left(\frac{s}{n}\right)\right), & s > 0 \end{cases}, \tag{20}$$

We can compute $E[S]$ and $V[S]$ from equation (20) or directly from equation (8), just taking into account relation (1). Then, for $E[S]$ we have:

$$\begin{aligned} E[S] = E[NX] &= \sum_{n\geq 1} \int_{\mathbb{R}^+} nxp\,(n)\, f\,(x)\,(1+\omega\psi\,(n)\,\phi\,(x))\,dx \\ &= \sum_{n\geq 1} np(n) \int_{\mathbb{R}^+} xf(x)dx + \omega \sum_{n\geq 1} np(n)\psi(n) \int_{\mathbb{R}^+} xf(x)\phi(x)dx \\ &= E[N]E[X] + \omega E[N\psi(N)]E[X\phi(X)] \\ &= E[S|\omega = 0] + \mathcal{O}(\omega) \end{aligned} \tag{21}$$

As it shows previous equation (21), we can see that under the framework of Sarmanov dependence the expected value of total cost can be computed by the sum of two terms. One term is simply the expected cost with no dependence and the other, $\mathcal{O}(\omega)$, is the contribution of the dependence to the expceted cost which is linear on Sarmanov's dependence parameter, ω. This second term also depends on the expected values of the product of our random variables and their kernels. Note that in the framework of this kind of dependence modelization the dependence efect on the expected value of total cost is easily interpretable as a deviation from the expected value without dependence, $E[S|\omega = 0]$.

To compute $V[S]$, lets first compute $E[S^2]$. Then we have:

$$\begin{aligned} E[S^2] = E[N^2X^2] &= \sum_{n\geq 1} \int_{\mathbb{R}^+} n^2x^2p\,(n)\, f\,(x)\,(1+\omega\psi\,(n)\,\phi\,(x))\,dx \\ &= \sum_{n\geq 1} n^2p(n) \int_{\mathbb{R}^+} x^2f(x)dx + \omega \sum_{n\geq 1} n^2p(n)\psi(n) \int_{\mathbb{R}^+} x^2f(x)\phi(x)dx \\ &= E[N^2]E[X^2] + \omega E[N^2\psi(N)]E[X^2\phi(X)] \\ &= E[S^2|\omega = 0] + \mathcal{O}(\omega) \end{aligned} \tag{22}$$

Equation (22) has the same functional form as equation (21), then it let us to interpret the efect of dependence on $E[S^2]$ as in the previous case; a deviation from $E[S^2|\omega = 0]$ with no dependence, by the adition of a linear term on ω, $\mathcal{O}(\omega)$. This second term also depends on the expected values of the product of the square of our random variables and their kernels.

Finally we are on possition to compute $V[S]$, as $V[S] = E[S^2] - E^2[S]$. Therefore we have:

$$\begin{aligned} V[S] &= E[S^2] - E^2[S] = E[N^2]E[X^2] + \omega E[N^2\psi(N)]E[X^2\phi(X)] - E^2[N]E^2[X] \\ &- \omega^2 E^2[N\psi(N)]E^2[X\phi(X)] - 2\omega E[N]E[X]E[N\psi(N)]E[X\phi(X)] \\ &= (E[N^2]E[X^2] - E^2[N]E^2[X]) + \omega(E[N^2\psi(N)]E[X^2\phi(X)] \\ &- 2E[N]E[X]E[N\psi(N)]E[X\phi(X)]) - \omega^2 E^2[N\psi(N)]E^2[X\phi(X)] \\ &= V[S|\omega = 0] + \mathcal{O}(\omega) + \mathcal{O}(\omega^2) \end{aligned} \tag{23}$$

On this case we can see that this kind of dependence introduce two aditional terms to $V[S]$, one linear on ω, $\mathcal{O}(\omega)$, and the second with cuadratic dependence on ω, $\mathcal{O}(\omega^2)$. These results, as well as the proper definition of Sarmanov dependence (8), make us think that this kind of dependence modelization is just a Taylor expansion centered on $\omega = 0$ of the multivariate pdf, assuming that dependence is governed by a single parameter, ω, which takes account of the strenght of the dependence. Also kernels $\psi(n)$ and $\phi(x)$ model the physical way in which N and X interact with each other.

At this point is important to say, that all these expected values will be computed by numerical integration in the empirical implementation. Box-Cox back transformed pdf has a really complicated analytical representation which prevent us to compute analyticaly those integrals.

2.4 Computing Log-Likelihoods

The empirical implementation of the fitting of this model will be based on Log-Likelihood maximization. Hence, first of all we have to derive its functional form. In general we can write the Likelihood as:

$$\begin{aligned} L[f_{X,N}] &= \prod_{j=1}^{m} f_{X,N} = \prod_{j=1}^{m_0} p(0) \prod_{j=1}^{m_1} p(n) f(x)(1 + \omega\phi(x)\psi(n)) \\ &= L[p_N]L[f_X]L[1 + \omega\phi(x)\psi(n)] \end{aligned} \tag{24}$$

Where m_0 is the number of observations with $n = 0$ and m_1 the number of observations with $n \geq 1$. Therefore, it has been shown that the Likelihood of the mixed pdf is the product of the Likelihoods of the involved marginals and the Likelihood of the mixing term.

The Log-Likelihood is simply the logarithm of the Likelihood, then we have:

$$l[f_{X,N}] = log(L[f_{X,N}]) = log(L[p_N]) + log(L[f_X]) + log(L[1 + \omega\phi(x)\psi(n)]) \tag{25}$$

In our particular case, we will asume a Negative-Binomial behaviour of N marginal, and as we have mentioned before, X marginal will be a Box-Cox back transformed Normal distribution. Now we compute the partial Log-Likelihoods for the marginals of our study.

The partial Log-Likelihood of a Negative-binomial distribution is given by:

$$\begin{aligned} l[p_N(n)] &= \sum_{j=1}^{m} log\left[\frac{\Gamma(r+n_j)}{n_j!\Gamma(r)}p^r(1-p)^{n_j}\right] = m_0 r log(p) + \sum_{j=1}^{m_1} log\left[\frac{\Gamma(r+n_j)}{n_j!\Gamma(r)}p^r(1-p)^{n_j}\right] \\ &= m_0 r log(p) + m_1 r log(p) - m_1 log(\Gamma(r)) + \sum_{j=1}^{m_1} log(\Gamma(r+n_j)) + \\ &+ log(1-p)\sum_{j=1}^{m_1} n_j - \sum_{j=1}^{m_1} log(n_j!) \end{aligned} \tag{26}$$

The partial Log-Likelihood of a Box-Cox back transformed Normal distribution, for $\lambda \neq 0$, is given by:

$$l[f_X(x)] = m_1 log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + (\lambda-1)\sum_{j=1}^{m_1} log(x_j) - \frac{1}{2}\sum_{j=1}^{m_1}\left(\frac{\frac{x_j^\lambda - 1}{\lambda} - \mu}{\sigma}\right)^2 \tag{27}$$

The partial Log-Likelihood of a Box-Cox back transformed Normal distribution, for $\lambda = 0$ (Lognormal distribution), is given by:

$$l[f_X(x)] = m_1 log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{j=1}^{m_1} log(x_j) - \frac{1}{2}\sum_{j=1}^{m_1}\left(\frac{log(x_j)-\mu}{\sigma}\right)^2 \tag{28}$$

The partial Log-Likelihood of the mixing term of a Sarmanov distribution is given by:

$$l[1+\omega\phi(x)\psi(n)] = \sum_{j=1}^{m_1} log(1+\omega\phi(x_j)\psi(n_j)) \tag{29}$$

Also we compute the partial Log-Likelihood for the case of a Gamma distribution, as we will use it for the development of further sections in order to compare some results.

The partial Log-Likelihood of a Gamma distribution is given by:

$$l[f_X(x)] = m_1\alpha log(\beta) - m_1 log(\Gamma(\alpha)) + (\alpha-1)\sum_{j=1}^{m_1} log(x_j) - \beta\sum_{j=1}^{m_1} x_j \tag{30}$$

3 Empirical Implementation

3.1 Procedure

The empirical implementation of this procedure is straight forward. First of all, we have to estimate parameters of marginal distributions. For the discrete pfm we will use the moments method, which will be computed on the sample. For the continuous pdf this method can not be applied, notice that the distribution of X is quite complicated and not even with analytical calculation software we have been able to get a general expression of the moments of X. Therefore, we have been forced to compute numerically all integrals involving X pdf. In this case, we first of all perform a Box-Cox fitting to X, obtaining optimal λ by maximum Log-Likelihood criterion (boxcox() R function). Then, we compute the new transformed random variable as (3), Y, using this λ, and we fit it to a Normal distribution, obtaining the best μ, σ by maximum Log-Likelihood criterion (fitdistr() R function). Once obtained these first approximation for marginal parameters, we could compute the limits on ω for this set of parameters by applying equation (18). Then, we fit the best ω within these limits by maximum Log-Likelihood criterion (optim() R function). With this ω, we now have to reoptimize parameters by maximum Log-Likelihood criterion. Then, with this new set of parameters we actualize ω limits and reoptimize ω, and so on. We iterate this procedure until the Log-Likelihood of ω remains constant. It is important to say that we also need to set a proper interval for marginal parameters optimization, to set them we find approximately their order of magnitude and let them one more order. Then, we set this bounds as $b = p \pm \frac{\alpha}{2}$, where b is a vector of upper and lower parameter bounds, p is the vector of parameters and α is the vector with the order of magnitude of parameters. Also to let more freedom to the fitting process, we set $\alpha = \alpha(1 + \delta)$, where delta allows us to slightly modify α. More over, we force this bounds to actualize in each fitting iteration, centering eachselves about the previous optimal set of parameters, p, and continously decreasing their bandwidth, in order to narrow their optimal solution. We find that a proper decreasing function for the bandwidth in this problem is $\Lambda(i) = \frac{2}{i+i^2}$, where i is the number of each iteration. Note that this process is not trivial, as we seek for a function not to fast decreasing in order to not to let out these bounds the optimal solution and also we want a fast decreasing to narrow as faster as we can the optimal solution. When the process has converged, we have all the parameters of the single marginals, but what it is more important, we will have a measure of ω which takes account of dependence between N and X on the framework of Sarmanov distribution.

On next section we expose the results we have obtained aplying this procedure. We fit three kinds of models. First, we fit the model allowing freedom to λ. On second place, we assume a Log-Normal behaviour of X and we fit the model assuming $\lambda = 0$. Finally, we assume a Gamma behaviour of X. On this last case, we generate a Gamma random sample (10^6 elements) based

on marginal and bivariate estimated parameters, and we perform on these samples a Box-Cox fitting to find the best λ, μ and σ, to compare initial and final results with Free λ model.

3.2 Numerical example

As we said before, our aim is to fit the bivariate Sarmanov distribution taking into account different assumptions for Box-Cox transformation parameter λ, that is; total freedom for λ and fixing $\lambda = 0$ for Lognormal distribution. We also fit the bivariate model assuming a perfect Gamma distribution for X marginal, then we estimate initial and final optimal Box-Cox parameters over this Gamma distribution and compare with other fitted models.

We now analyze a data set of auto insurance policyholders of an iternational company. This data set contains a sample of $m = 99,978$ Spanish insureds. For each individual we have information on the number and average cost of claims.

On next Table 1, it can be seen the marginal parameters fitted without dependence between N and X, which will be the starting point to fit the bivariate dependence models. Obviously, there are no diferences between models on N pfm parameters as far as this part remains unchanged among the three models. On X pdf we can see that between Gamma and Free λ model there are small differences, this makes us think that X is nearly distributed as a Gamma distribution. The differences on μ and σ between this two models and Lognormal model are quite significant; about 55% on both parameters.

Table 1: Initial parameters estimated on marginal distributions for all considered models

	Free λ	Lognormal	Gamma
r	0.3171	0.3171	0.3171
p	0.7814	0.7814	0.7814
μ	6.2333	4.0854	6.2161*
σ	5.1434	3.3003	5.1283*
λ	0.1169	-	0.1162*
α	-	-	0.1935
β	-	-	0.0003

* Estimated Box-Cox transformation parameters over Gamma with estimated marginal parameters.

On Tables 2 and 3, we show the basic stats for the continous part of X marginal ($X > 0$) and a table of frequencies for observed and Negative Binomial adjusted N pfm.

Table 2: Basic stats of average claim severity X sample.*

Mean	Median	STDEV	Skewness	Kurtosis
694.98	444.52	1579.79	15.74	445.39

* Note that this stats are computed for the continous part of X marginal, i.e. $X > 0$.

Table 3: Observed and Negative Binomial adjusted frequencies of number of claims N

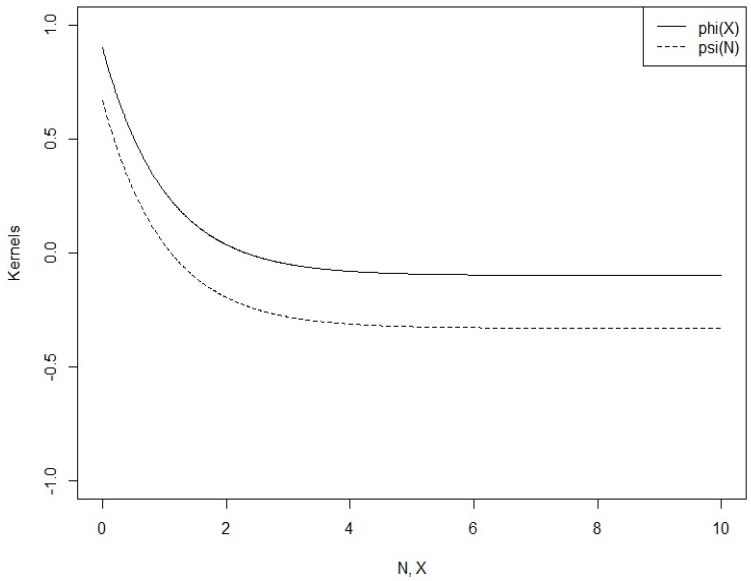
N	Observed	Negative Binomial	% difference
0	92538.00	92447.74	0.1
1	6166.00	6410.43	3.9
2	1122.00	923.06	17.7
3	125.00	155.88	24.7
4	18.00	28.27	57.0
5	3.00	6.63	121.0

Results of previous Table 2 reinforce our thinking that X is nearly Gamma distributed, as we can compute $E[X] = \alpha/\beta \approx 645$ and $V^{\frac{1}{2}}[X] = (\alpha/\beta^2)^{\frac{1}{2}} \approx 1,466$. Also, the sample has positive skewness as by definition must be Gamma.

As it is seen on the previous Table 3, most of observations have null claim frequency, which leads to a high probabiliy of non having claims of about 0.926. Comparing observed frequency with Negative Binomial adjusted one, we can see that diferences increase in percentage as N increase, i.e. on the tail of N distribution. Despite of this differences we find that Negative Binomial distribution models in an appropriately N pfm.

Since we have already defined the initial parameters (marginal parameters) of our bivariate models, we are on possition of asking ourselves what the functional form of dependence we assume for N and X will looks like, previously defined on equations (15) and (17).

Figure 1: Exponential kernels to model X and N dependence



In view of Figure 1, it is clear that both X and N contributions to dependence exponentially increase as each variable decrease. In the context of auto insurance this can be interpreted as if you have an observation with a low claim frequency the most problable outcome is that average claim severity will be also low. Insureds with a high number of claims tend to have little adversion to risk what make them to be more susceptible to have more claims and with a greater severity in terms of fatality of the accident and indeed implying a higher cost for the company.

On next Figure 2, we show the bivariate form of the dependence on a 3D-plot. It is seen what we have mentioned before for Figure 1, we have a peak on dependece for simultaneous small values of X and N. Also it is important to note that for small values of N and high values of X the sign of dependence is inverted, reinforcing our previous arguments.

Figure 2: 3D-Plot of bivariate dependence, $\psi(N)\phi(X)$

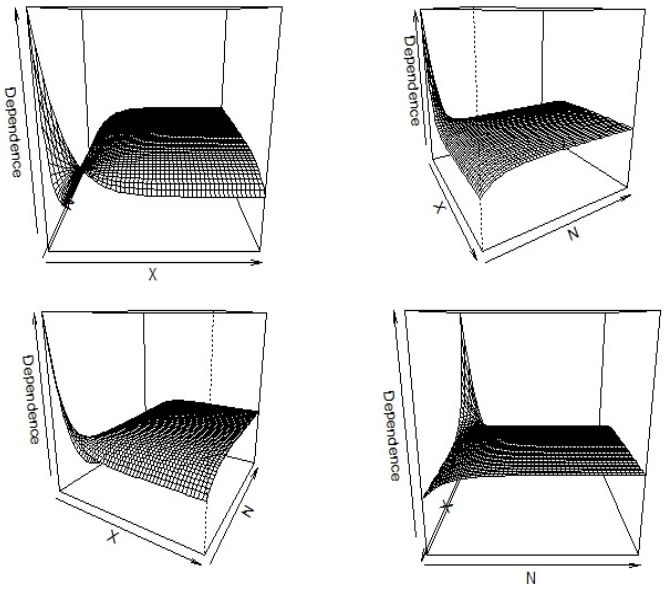


Table 4 shows the final result we have after performing the fitting on the bivariate model. Here we can see all estimated parameters as well as ω boundaries imposed by marginal parameters for all considered models. It is also shown the AIC of each model. We can appreciate that once again, all parameters are quite similar between Gamma and Free λ model. On the other hand, the diferences between these two models and the Lognormal model are quite significant. Its important to note that despite on these diferences on marginal parameters, all three models have similar dependence parameter ω. Looking at AIC, we can determine that the best model is Gamma model, as far as it has the lowest AIC. We expected free λ model to be the best, as far as it allows more freedom to the model to adapt itself to the model, but when we were fitting it we could see that tiny changes on λ lead to great changes on the Log-likelihood function of ω which added some oscillation behaviour on the process making the fitting more complicated.

Table 4: Estimated parameters for all considered bivariate Sarmanov models

	Free λ	Lognormal	Gamma
r	0.2897	0.3017	0.2894
p	0.7655	0.7727	0.7654
μ	6.2333	4.0854	7.2003*
σ	5.1434	3.3003	5.6099*
λ	0.1169	-	0.1519*
α	-	-	0.2826
β	-	-	0.0004
ω	3.3745	3.3398	3.4176
min ω	-28.5475	-29.0678	-27.5439
max ω	3.3745	3.3398	3.4176
AIC	159774.6951	160283.5448	158494.1754

*Estimated Box-Cox parameters for bivariate Sarmanov Gamma distribution parameters

Finally, are shown the p-values of the pararameters of each model. We can see that all parameters are statistically significatives in all models, having all of them a p-value less than 10^{-4}.

Table 5: P-values of bivariate models estimated parameters

	Free λ	Lognormal	Gamma
r	0.00	0.00	0.00
p	0.00	0.00	0.00
μ	0.00	0.00	0.00
σ	0.00	0.00	0.00
λ	0.00	-	-
α	-	-	0.00
β	-	-	0.00
ω	$11.65\ 10^{-6}$	$12.73\ 10^{-6}$	$9.30\ 10^{-6}$

Once fitted our models, we may ask ourselves how look like the estimated pdf's. On next Figures 3 and 4, we show the histogram of X and estimated X pdf's in the framework of Sarmanow bivariate distribution.

Figure 3: Histogram of X and estiamted probability densities

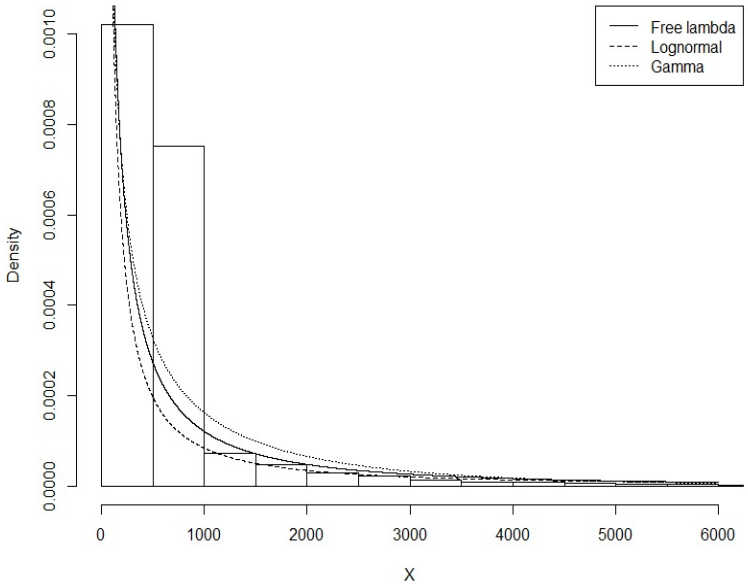


Figure 4: Tail zoom of histogram of X and estiamted probability densities

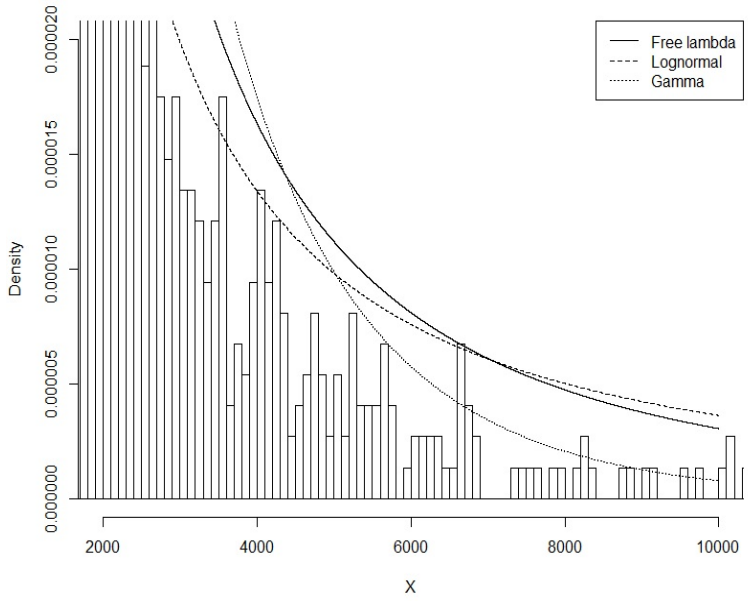


Figure 3 reflects our conclusions of the fitting result on Table 4, Gamma model seems to be the best model to fit the overall empirical density of X, followed by Free λ model, and Logonormal model ($\lambda = 0$) being the worst of all three considered models. On the other hand, if we look at the tail zoom for $X \in [2000, 10000]$ (Figure 4), we can appreciate that results are not the same as in Figure 3. In this case we can see a better adjustment of Free λ model, being Gamma model the worst, underestimating risk for extreme values. Then, we can conclude that the freedom introduced by Box-Cox transformation allows to capture the essence of heavy tails quite well as we have mentioned on previous sections. Then, we think that this model can be really appropiate to quantify risk measures as VaR or $TVaR$.

Finally, we show on Figures 5 and 6 a plot of Log-Likelihood function of Free λ model as a function of λ and ω, respectively.

Figure 5: Log-Likelihood of Free λ model for bivariate estimated parameters as a function of λ

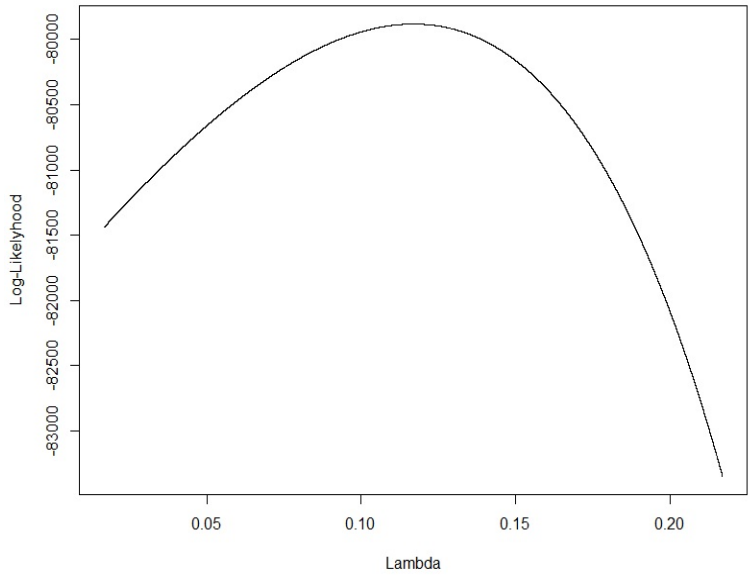
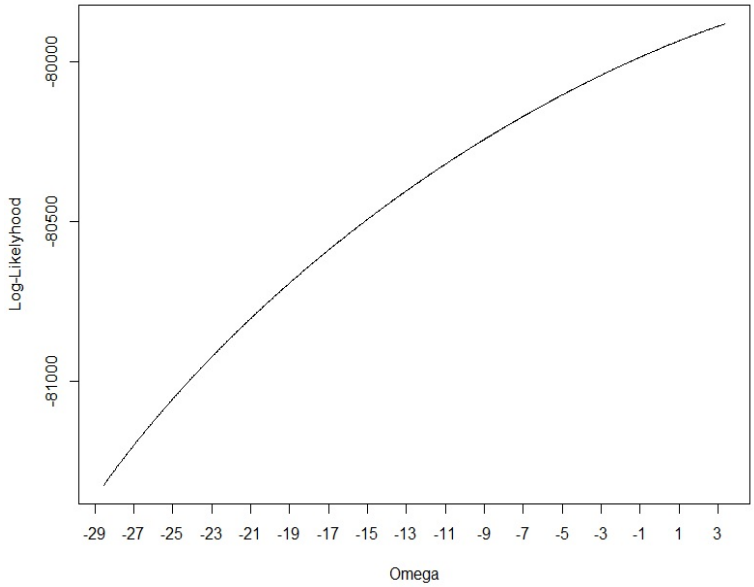


Figure 6: Log-Likelihood of Free λ model for bivariate estimated parameters as a function of ω



Note that ω domain is restricted by equation (18), then the Log-Likelihood as a function of ω may does not have a local maximum on this allowed interval, as it is the case. This graph justifies the fact that we have obtained as optimal ω the upper limit. On the other hand, we can see on Figure 5 that the Log-Likelihood has a local maximum in the considered interval which also correspond to its global maximum, being the result obtained on the optimization preocedure as well.

3.3 Quantification of dependence effect on premiums

As we know, the pure premium is computed as the expected total cost, $E[S] = E[XN]$. By the assumption of independence between N and X this expresion yields to equation (2), but when dependence is taken into account this relation yields to equation (21) on the framework of Sarmanov distribution, then we expect some diferences which would directly impact on the premium that an insurance company charges its clients. Also this premium has a contribution of the so called risk premium which takes account of the disperison of S, then by the assumption of the standard deviation criteria we can define risk premium as:

$$\rho_R = E[S] + \delta V^{\frac{1}{2}}[S] \tag{31}$$

Now, to quantify this dependence effect we computed pure and risk premiums according to equations (31), (21) and (23), assuming $\omega = 0$ to impose independence and setting optimal ω computed on the previous section for dependence case. Also, we set $\delta = 0$ to evaluate the impact only on pure premium and $\delta = 1$ to see the effect on risk premium. Results are shown on the table bellow.

Table 6: Pure and risk premium computed on free λ model for $\omega = 0$ and $\omega \neq 0$

	$\omega = 0$	$\omega \neq 0$
$\delta = 0$	102.67	103.91
$\delta = 1$	1734.66	1761.30

As it is seen on previous Table 6, the dependence effect on pure premium implies an increment of 1.2%, while on risk premium implies an increment of about 1.54%. In other words, this means that our model estimates that with a not too high positive dependence if we assume totally independence we are understimating the inherent risk involving this insurance operations, increasing in this way the ruin probability of the company.

3.4 Conclusions

In this paper, we have shown that Sarmanov distribution is a very usefull, powerfull and also flexible tool to model dependence among random variables, it let us to mix discrete and continuous distributions in a really intuitive and simple way. We also have shown the utility of Box-Cox transformation to adapt a Normal distribution pdf to an a priori unknown distribution, allowing us to not pre establish any assumption about claim severity pdf, futhermore this technique has been tested with the case of a Gamma distribution. Also, it was showed that

back transformed Box-Cox model performs very good at fitting heavy tails. More over, it was proposed a maximum likelihood estimation method for this Box-Cox back transformed Sarmanov distribution.

We tryed our model to estimate dependence between the frequency and severity of claims in the collective model for real auto insurance data. Results obtained shown a very good performance of the model, being all estimated parameters statistically significant. Also we obtained positive dependence in all tryed models, despite the strenght of this dependence was not to high.

Finally, we used our model to quantify the effect of dependence in pure and risk premiums on insurance, showing that in the case of positive dependence we are underestimating the inherent risk if we don't take into account this dependence, increasing the ruin probability of the company.

References

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

Box, G., Cox, D., 1964. An analysis of transformations. Journal of the Royal Statistical Society 26, 211–252.

Clements, A., Hurn, S., Lindsay, K., 2003. Mobius-like mappings and their use in kernel density estimation. Journal of the American Statistical Association 98, 993.

Jonhson, N., Kotz, S., Balakrishnan, N., 1997. Discrete multivariate distributions. Wiley.

Kotz, S., Balakrishnan, N. Jonhson, N., 2000. Continuous Multivariate Distributions. Vol.1: Models and Applications. Wiley.

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

Vernic, R., Bolancé, C., M., Alemany, R., 2020. Sarmanov distribution for modeling dependence between the frequency and the average severity of insurance claims. Unpublished manuscript .

Wand, M.P., Marron, J.S., Ruppert, D., 1991. Transformations in density estimation. Journal of the American Statistical Association 86, 343–353.

Zhang, X., King, M.L., 2004. Box-cox stochastic volatility models with heavy-tails and correlated errors. Figshare .